

A REDUCED BASIS METHOD FOR PARABOLIC PARTIAL DIFFERENTIAL EQUATIONS WITH PARAMETER FUNCTIONS AND APPLICATION TO OPTION PRICING

ANTONIA MAYERHOFER AND KARSTEN URBAN

ABSTRACT. We consider the Heston model as an example of a parameterized parabolic partial differential equation. A space-time variational formulation is derived that allows for parameters in the coefficients (for calibration) as well as choosing the initial condition (for option pricing) as a parameter function. A corresponding discretization in space and time and initial condition is introduced and shown to be stable. Finally, a Reduced Basis Method (RBM) is introduced that is able to use parameter functions also for the initial condition. Corresponding numerical results are shown.

1. INTRODUCTION

Calibration and pricing are two standard tasks in numerical finance. Using a PDE-model based upon the Feynman-Kac theorem (see also Proposition 2.1 below), calibration amounts fitting unknown parameters in the partial differential equation (PDE) to historical market data in a least-squares sense e.g. by some optimization scheme. This usually requires solving the same PDE with many different values of the parameters (given by the iterative optimization method). Such a problem is also termed ‘multi-query’.

We consider the Heston model, which allows for a non-constant volatility, [20]. Having in mind that every model is wrong to a certain extend, constant volatilities are of course a severe limitation, which has been our motivation for considering the Heston model here. The Heston model leads to a system of parabolic diffusion-convection-reaction PDEs with non-constant coefficients, which makes the model also interesting from a numerical point of view.

Moreover, as we shall see below, the Heston model already contains 4 parameters, i.e., calibration requires solving a 4-dimensional PDE-constrained optimization problem. Of course, by extending the model, the number of parameters can easily be enlarged leading to truly high-dimensional parameter spaces.

In the PDE-framework, pricing amounts solving the same kind of parabolic problem with different initial conditions determined by different payoff functions. The payoff is usually determined by the specific design of the financial product, so that a flexibility in the choice of the payoff function enables efficient product design.

Date: August 12, 2014.

2010 Mathematics Subject Classification. 35K85,49J40,65M15,91B25, 91G80.

Key words and phrases. Option pricing, parabolic problems, reduced basis method, error estimates.

The authors have been supported by the Deutsche Forschungsgemeinschaft (DFG) under GrK 1100. We are grateful to Robert Stelzer for helpful discussions and comments.

Both problems, calibration and pricing need to be solved rapidly, often ‘online’ e.g. in ‘real time’ as for risk management.

Such kind of rapid online multi-query problems are in the scope of Reduced Basis Methods (RBM), that have been developed and widely investigated for parametric PDEs, in particular of elliptic and parabolic type, see, e.g. [14, 25] and references therein. Roughly speaking, the RBM involves an offline and an online computing phase. In the offline phase, a detailed, but costly numerical model (sometimes called ‘truth’), is used to construct a reduced space. The construction is based upon a reliable and effective error estimate which in addition can be evaluated efficiently. Moreover, this error estimate allows to compute an a posteriori bound, which makes RB-approximations certified.

Similar to [30, 31], we suggest a space-time variational formulation of the parabolic PDE-system. This leads to a Petrov-Galerkin method containing space and time. Obviously, the time is added as additional dimension, but it allows for sharp error bounds for a corresponding discretization in space and time. Moreover, similar to [30, 31] we show that a specific choice of trial and test spaces in the Petrov-Galerkin method lead to a Crank-Nicholson scheme also for the case of varying initial conditions. This requires also a discretization of the initial condition so that we end up with a detailed discrete problem in space and time and initial condition, which is used as ‘truth’ for a RBM. It turns out that the space-time approach transfers an initial condition parameter function into a parameter within the PDE-coefficients.

Even though RBM have already been used in numerical finance, see e.g. [5, 16, 28], and even though the current paper is clearly only a first step, we think to the best of our knowledge that we introduce some novel aspects: (a) We use the initial condition of an evolution problem as a parameter within a RBM, which has not been considered so far; (b) The initial condition is a parameter *function*, which in particular means that the parameter space is infinite-dimensional. We solve this problem by expanding the initial condition in an appropriate basis, which is somehow similar to e.g. [17, 22]. (c) A space-time variational formulation is used to treat a non-homogeneous initial condition by a two-step method, namely first approximating the initial condition and then using this as a parameter in the right-hand side for determining the evolution. We note, that this is *not* restricted to linear problems only, see Remark 3.3 and §3.3.5 below.

The remainder of this paper is organized as follows. In Section 2, we collect preliminaries on the Heston model and Reduced Basis Methods (RBM). Section 3 is devoted to the review of the space-time variational formulation, in particular for parabolic parametric partial differential equations. This includes the error estimate from [30, 31], which is the basis for the RB error estimate. Next, we introduce a discretization. In space and time, we follow [30, 31] and use finite elements. In addition, we introduce a discretization for the initial condition and show that we can decompose the numerical computations into (1) a Galerkin-type problem for the approximation of the initial condition and (2) a time-stepping scheme for the evolution.

The next step is the introduction of a Reduced Basis Method (RBM) in particular for parameter functions in Section 4. Again, we derive a two-step method, here for the construction of the reduced basis. We derive stable RB trial and test spaces

for the reduced Petrov-Galerkin scheme and show an error estimate. We present numerical results in Section 5 and close with a summary and outlook in Section 6.

2. PRELIMINARIES

2.1. Calibration and pricing within the Heston Model. The Heston model is a well-known model for option pricing allowing also a non-constant volatility of the underlying. It was invented 1993 by Heston, [20]. The Stochastic Differential Equations (SDEs) for the asset price S_t and the volatility ν_t are assumed to be

$$(2.1) \quad dS_t = rS_t dt + \sqrt{\nu_t} S_t dz_1(t), \quad d\nu_t = \kappa[\theta - \nu_t] dt + \sigma \sqrt{\nu_t} dz_2(t),$$

where z_1, z_2 are Wiener processes with correlation ρ , r is the return rate of the asset, κ the mean reversion rate to the long term variance θ and σ is the volatility of the volatility. In particular, the instantaneous variance ν_t is modeled as a CIR (Cox-Ingersoll-Ross) process, [6]. Finally, the model implies that $z_1 = \sqrt{1 - \rho^2} dz_3 + \rho dz_2$ with independent Brownian motions z_2, z_3 .

The following partial differential equation (PDE) representation is well-known.

Proposition 2.1 (Feynman-Kac formula, [24]). *Let X_t be an n -dimensional Itô diffusion, i.e., $dX_t = b(X_t)dt + \sigma(X_t)dB_t$, $f \in C^2(\mathbb{R}^n)$ be a compactly supported payoff and $q \in C(\mathbb{R}^n)$ be lower bounded. For $u(t, x) := \mathbb{E}[e^{-\int_0^t q(X_s)ds} f(X_t) | X_t = x]$ it holds for $x \in \mathbb{R}^n$ that $u(0, x) = f(x)$ and for all $t \in (0, T]$, $x \in \mathbb{R}^n$ that*

$$\frac{\partial u(t, x)}{\partial t} = \frac{1}{2} \sum_{i,j=1}^n (\sigma(x)\sigma^T(x))_{i,j} \frac{\partial^2 u(t, x)}{\partial x_i \partial x_j} + \sum_{i=1}^n b_i(x) \frac{\partial u(t, x)}{\partial x_i} - q(x) u(t, x). \quad \square$$

For applying the Feynman-Kac formula to the Heston model, we consider the process $X_t := (y_t, \nu_t)^T = (\log(S_t), \nu_t)^T$, $n = 2$. Then, we can rewrite the two SDEs in (2.1) as a system

$$\begin{aligned} dX_t &= d \begin{pmatrix} y_t \\ \nu_t \end{pmatrix} = \begin{pmatrix} r - \frac{1}{2}\nu_t \\ \kappa[\theta - \nu_t] \end{pmatrix} dt + \begin{pmatrix} \sqrt{\nu_t} \sqrt{1 - \rho^2} & \sqrt{\nu_t} \rho \\ 0 & \sigma \sqrt{\nu_t} \end{pmatrix} \begin{pmatrix} dz_3 \\ dz_2 \end{pmatrix} \\ &=: b(X_t) dt + \sigma(X_t) dB_t. \end{aligned}$$

Using Proposition 2.1 then yields the PDE (in x) for the Heston model of the form

$$(2.2) \quad \frac{\partial u(t)}{\partial t} - \text{div}(\underline{\alpha}(t) \nabla u(t)) + \underline{\beta}(t) \nabla u(t) + \gamma(t) u(t) = 0 \text{ for } t \in (0, T], \quad u(0) = u_0,$$

where u_0 is the payoff and the coefficient functions are given by

$$(2.3) \quad \underline{\alpha}(t) := \frac{1}{2} \begin{pmatrix} \nu_t & \nu_t \sigma \rho \\ \nu_t \sigma \rho & \nu_t \sigma^2 \end{pmatrix}, \quad \underline{\beta}(t) := - \begin{pmatrix} r - \frac{1}{2}\nu_t - \frac{1}{2}\sigma \rho \\ \kappa \theta - \kappa \nu_t - \frac{1}{2}\sigma^2 \end{pmatrix}, \quad \gamma(t) := r.$$

Truncating $x \in \mathbb{R}^2$ to some bounded domain $\Omega \subset \mathbb{R}^2$ and posing appropriate (truncation) boundary conditions yields

$$(2.4a) \quad \frac{d}{dt} u(t, x) + \mathcal{A}(t) u(t, x) = 0 \text{ in } (0, T] \times \Omega,$$

$$(2.4b) \quad u(t, x) = 0 \text{ on } [0, T] \times \partial\Omega,$$

$$(2.4c) \quad u(0, x) = u_0(x) \text{ on } \Omega,$$

where the differential operator is given as $\mathcal{A}(t)u := -\text{div}(\underline{\alpha}(t) \nabla u) + \underline{\beta}(t) \nabla u + \gamma(t)u$, i.e., $\mathcal{A}(t) \in \mathcal{L}(V, V')$ with $V := H_0^1(\Omega)$. It is readily seen that $\mathcal{A}(t)$ is elliptic if and only if $\nu_t > 0$ and $\rho \in (-1, 1)$, which we will always assume.

It is well-known that the bilinear form associated to $\mathcal{A}(t)$, i.e., $\langle \mathcal{A}(t)u, v \rangle_{V' \times V} = a(t; u, v)$ reads

$$a(t; u, v) := \int_{\Omega} \{ \underline{\alpha}(t) \nabla u(x) \cdot \nabla v(x) + \underline{\beta}(t) \cdot \nabla u(x) v(x) + \gamma(t) u(x) v(x) \} dx.$$

If $(\underline{\alpha})_{ij}(t), \underline{\beta}_i(t), \gamma(t) \in L_{\infty}(\Omega)$, $i, j = 1, 2$, and $u_0 \in L_2(\Omega) =: H$, then there exist constants $0 < M_a, \alpha_a < \infty$ and $0 \leq \lambda_a < \infty$ such that for all $u, v \in V$ it holds

$$(2.5a) \quad |a(t; u, v)| \leq M_a \|u\|_V \|v\|_V,$$

$$(2.5b) \quad a(t; u, u) + \lambda_a \|u\|_H^2 \geq \alpha_a \|u\|_V^2,$$

if $\mathcal{A}(t)$ is elliptic, see, e.g. [10, Thm. 2, p. 318]. Obviously, (2.5a) means boundedness and (2.5b) is a Gårding inequality.

Parameters. In order to calibrate the model, one has to determine the model parameters, i.e., $\mu_1 := (\rho, \kappa, \theta, \sigma)$ based upon market data. Moreover, in particular for pricing, we would like to determine prices for different payoff functions, i.e., we consider the initial value $\mu_0 := u_0$ as a parameter function.

2.2. Reduced Basis Methods (RBMs). Now, we recall the main features of Reduced Basis Methods (RBMs) for parameterized partial differential equations (PPDEs) that are needed here. For more details, we refer to the surveys [14, 25] and the references therein. We assume that the variational form of the PDE is given by

$$(2.6) \quad u(\mu) \in \mathbb{X} : \quad b(\mu; u(\mu), v) = f(\mu; v) \quad \forall v \in \mathbb{Y},$$

where \mathbb{X}, \mathbb{Y} are (infinite-dimensional) Hilbert spaces, $b : \mathcal{D} \times \mathbb{X} \times \mathbb{Y} \rightarrow \mathbb{R}$ a parametric bilinear form and $f : \mathcal{D} \times \mathbb{Y} \rightarrow \mathbb{R}$ is a parametric functional, $f(\mu; \cdot) \in \mathbb{Y}'$. The parameter space is denoted by \mathcal{D} . Well-posedness of (2.6) is always assumed, which in particular means that an inf-sup-condition

$$(2.7) \quad \inf_{w \in \mathbb{X}} \sup_{v \in \mathbb{Y}} \frac{b(\mu; w, v)}{\|w\|_{\mathbb{X}} \|v\|_{\mathbb{Y}}} \geq \beta_b > 0$$

is satisfied for all $\mu \in \mathcal{D}$.

The next step is the assumption that a stable finite discretization $\mathbb{X}^{\mathcal{N}} \subset \mathbb{X}$, $\mathbb{Y}^{\mathcal{N}} \subset \mathbb{Y}$ in the sense

$$(2.8) \quad \inf_{w^{\mathcal{N}} \in \mathbb{X}^{\mathcal{N}}} \sup_{v^{\mathcal{N}} \in \mathbb{Y}^{\mathcal{N}}} \frac{b(\mu; w^{\mathcal{N}}, v^{\mathcal{N}})}{\|w^{\mathcal{N}}\|_{\mathbb{X}} \|v^{\mathcal{N}}\|_{\mathbb{Y}}} \geq \beta^{\mathcal{N}} > 0$$

is available that is sufficiently fine so that the discrete Petrov-Galerkin solution $u^{\mathcal{N}}(\mu)$ is a sufficiently good approximation to $u(\mu)$. Hence, $u^{\mathcal{N}}(\mu)$ is called *detailed* or *truth* solution. We assume that $u^{\mathcal{N}}(\mu)$ can be computed with \mathcal{N} operations and that this number \mathcal{N} is too large to be acceptable for realtime or multi-query computations, so that a reduced model is necessary.

In order to derive this reduced model, the idea is to split the computations into an offline and an online phase. During the offline phase, one may afford to use the detailed model in order to compute *snapshots* $u^i := u^{\mathcal{N}}(\mu^i)$, for $i = 1, \dots, N$ and $N \ll \mathcal{N}$ is much smaller so that the reduced model is feasible online even if the complexity is of the order $\mathcal{O}(N^3)$ since the stiffness matrix is in general densely populated.

The reduced solution $u_N(\mu)$ is then determined as the (Petrov-)Galerkin solution in the space $\mathbb{X}_N := \text{span}\{u^i : i = 1, \dots, N\}$ with a stable reduced space $\mathbb{Y}_N \subset \mathbb{Y}^{\mathcal{N}}$ of dimension N :

$$(2.9) \quad u_N(\mu) \in \mathbb{X}_N : \quad b(\mu; u_N(\mu), v_N) = f(\mu; v_N) \quad \forall v_N \in \mathbb{Y}_N.$$

Stability is to be understood in a uniform inf-sup-sense, i.e.,

$$(2.10) \quad \inf_{w_N \in \mathbb{X}_N} \sup_{v_N \in \mathbb{Y}_N} \frac{b(\mu; w_N, v_N)}{\|w_N\|_{\mathbb{X}} \|v_N\|_{\mathbb{Y}}} \geq \beta_{\text{LB}} > 0,$$

with β_{LB} independent of N as $N \rightarrow \infty$ as well as of μ .

The snapshots defining the reduced space are determined by the (offline-)selection of parameter samples $S_N := \{\mu^i : i = 1, \dots, N\}$ and those samples are computed by maximizing a computable error estimate $\Delta_N(\mu)$ w.r.t. the parameter μ . This can be done e.g. by nonlinear optimization or a greedy method w.r.t. a finite so-called *training set* $\Xi^{\text{train}} \subset \mathcal{D}$. Such an error estimate can e.g. be derived as follows.

Proposition 2.2. *Let $\mathbb{X}_N \subset \mathbb{X}^{\mathcal{N}}$ and $u_N(\mu) \in \mathbb{X}_N$ be the solution of (2.9). Defining the residual by $r_N(\mu; v) := f(\mu; v) - b(\mu; u_N(\mu), v) = b(\mu; u^{\mathcal{N}}(\mu) - u_N(\mu), v)$ for all $v \in \mathbb{Y}^{\mathcal{N}}$, we obtain the following error estimate*

$$(2.11) \quad \|u^{\mathcal{N}}(\mu) - u_N(\mu)\|_{\mathbb{X}^{\mathcal{N}}} \leq \frac{1}{\beta^{\mathcal{N}}} \|r_N(\mu)\|_{(\mathbb{Y}^{\mathcal{N}})'} =: \Delta_N(\mu),$$

where $\beta^{\mathcal{N}}$ denotes the inf-sup-constant (2.8) of $b(\mu; \cdot, \cdot)$ on $\mathbb{X}^{\mathcal{N}} \times \mathbb{Y}^{\mathcal{N}}$ (possibly with discrete norms $\|\cdot\|_{\mathbb{X}^{\mathcal{N}}}$, $\|\cdot\|_{\mathbb{Y}^{\mathcal{N}}}$). \square

The ultimate efficiency aim is to realize an online complexity that is independent of \mathcal{N} . The key for that realization is the assumption that bilinear form and right-hand side functional are separable w.r.t. the parameter, i.e.,

$$(2.12) \quad b(\mu; w, v) = \sum_{q=1}^{Q_b} \vartheta_q^b(\mu) b_q(w, v), \quad f(\mu; v) = \sum_{q=1}^{Q_f} \vartheta_q^f(\mu) f_q(v),$$

with functions $\vartheta_q^b, \vartheta_q^f : \mathcal{D} \rightarrow \mathbb{R}$ and parameter-independent forms $b_q : \mathbb{X} \times \mathbb{Y} \rightarrow \mathbb{R}$ and $f_q : \mathbb{Y} \rightarrow \mathbb{R}$. In the RB literature this is often –a bit misleading– called *affine decomposition*.

Let $\{v^j : 1 \leq j \leq N\}$ be a basis of the space \mathbb{Y}_N , then (2.9) reads

$$(2.13) \quad \mathbf{u}_N = (u_{N,i})_{i=1, \dots, N} \in \mathbb{R}^N : \quad \sum_{i=1}^N u_{N,i} b(\mu; u^i, v^j) = f(\mu; v^j) \quad \forall j = 1, \dots, N.$$

Furthermore, let $u^i = \sum_{n=1}^{\mathcal{N}} \alpha_n^i \varphi_n^{\mathcal{N}}$ be the representation of the snapshots in a basis $\{\varphi_n^{\mathcal{N}} : n = 1, \dots, \mathcal{N}\}$ of $\mathbb{X}^{\mathcal{N}}$ and correspondingly $v^j = \sum_{n=1}^{\mathcal{N}} \beta_n^j \psi_n^{\mathcal{N}}$ in a basis $\{\psi_n^{\mathcal{N}} : n = 1, \dots, \mathcal{N}\}$ of $\mathbb{Y}^{\mathcal{N}}$. If (2.12) holds, the computation of the stiffness matrix and the right-hand side of the reduced linear system can be split in an offline/online fashion as follows (here only for the stiffness matrix)

$$\begin{aligned} b(\mu; u^i, v^j) &= \sum_{n, n'=1}^{\mathcal{N}} \alpha_n^i \beta_{n'}^j b(\mu; \varphi_n^{\mathcal{N}}, \psi_{n'}^{\mathcal{N}}) = \sum_{n, n'=1}^{\mathcal{N}} \alpha_n^i \beta_{n'}^j \sum_{q=1}^{Q_b} \vartheta_q^b(\mu) b_q(\varphi_n^{\mathcal{N}}, \psi_{n'}^{\mathcal{N}}) \\ &= \sum_{q=1}^{Q_b} \vartheta_q^b(\mu) \sum_{n, n'=1}^{\mathcal{N}} \alpha_n^i \beta_{n'}^j b_q(\varphi_n^{\mathcal{N}}, \psi_{n'}^{\mathcal{N}}) =: \sum_{q=1}^{Q_b} \vartheta_q^b(\mu) (b_q)_{i,j}, \end{aligned}$$

where the terms $(b_q)_{i,j}$ are μ -independent and can thus be computed offline. Online, for a new parameter $\mu \in \mathcal{D}$, the stiffness matrix of the RB-system is obtained by

$$(\mathbf{B}_N(\mu))_{i,j} = b(\mu; u^i, v^j) = \sum_{q=1}^{Q_b} \vartheta_q^b(\mu) (\mathbf{B}_q)_{i,j}, \quad i, j = 1, \dots, N,$$

which requires $\mathcal{O}(Q_b N^2)$ operations independent of \mathcal{N} . The same procedure can be done for the right-hand side and also for the error estimator $\Delta_N(\mu)$ since one can derive a separation like in (2.12) also for the residual $\|r_N(\mu)\|_{(\mathbb{Y}^N)'}.$

Starting from an elliptic PPDE, there are several RB methods for parabolic problems using a usual time-stepping approach, e.g. [13, 15]. We do not follow this path here, in particular since we aim at using the initial value as a parameter function. Standard RB error estimates for time-stepping-based methods involve sums of residuals of each time step, so that an initial error is possibly heavily amplified. This is one of the reasons, why we consider a space-time variational formulation of parabolic PPDEs that will lead us to a problem of the form (2.6), where \mathbb{X}, \mathbb{Y} will be Bochner spaces, i.e. involve space and time.

3. SPACE-TIME FORMULATION OF PARABOLIC PPDES

We now review the space-time variational formulation and its discretization.

3.1. Parabolic PPDEs. The Heston model problem yields a parameterized parabolic initial-boundary value problem of the following form: Let $I := (0, T)$ be the (open) time interval and $V \hookrightarrow H \hookrightarrow V'$ be a Gelfand triple of Hilbert spaces (e.g. $V = H_0^1(\Omega)$, $H = L_2(\Omega)$ for a bounded domain $\Omega \subset \mathbb{R}^d$). The parameters are assumed to take the form $\mu = (\mu_0, \mu_1) \in \mathcal{D} := \mathcal{D}_0 \times \mathcal{D}_1$, where $\mathcal{D}_0 \subset H$ is a set of possible initial values and $\mathcal{D}_1 \subset \mathbb{R}^P$ is a parameter space of finite (or infinite) dimension. Then, given some $g(\mu; t) \in V'$, $t \in I$ a.e., we look for $u(\mu; t) \in V$, $t \in I$ a.e., such that

$$(3.1a) \quad \langle \dot{u}(\mu; t), \phi \rangle_{V' \times V} + a(\mu_1; u(\mu; t), \phi) = \langle g(\mu_1; t), \phi \rangle_{V' \times V} \quad \forall \phi \in V, t \in I \text{ a.e.},$$

$$(3.1b) \quad u(\mu; 0) = \mu_0 \text{ in } H,$$

where $a(\mu_1; \cdot, \cdot) : V \times V \rightarrow \mathbb{R}$ is a bounded bilinear form. Note, that $a(\cdot; \cdot, \cdot)$ and $g(\cdot; \cdot)$ are assumed to depend only on μ_1 , *not* on μ_0 . As described above, μ_0 is the initial value parameter, whereas μ_1 contains the parameters occurring within the coefficients of bilinear form and right-hand side, e.g. parameters to be calibrated. We can easily extend our findings to time-dependent bilinear forms $a(\mu; t, \cdot, \cdot)$ as well.

As already motivated in Section 2.2, we assume the following separability w.r.t. the parameter $\mu_1 \in \mathcal{D}_1$ for $\phi, \psi \in V$ and $t \in I$

$$(3.2) \quad a(\mu_1; \phi, \varphi) = \sum_{q=1}^{Q_a} \vartheta_q^a(\mu_1) a_q(\phi, \varphi), \quad g(\mu_1; t) = \sum_{q=1}^{Q_g} \vartheta_q^g(\mu_1) g_q(t),$$

where $\vartheta_q^a, \vartheta_q^g : \mathcal{D}_1 \rightarrow \mathbb{R}$, $a_q(\cdot, \cdot) : V \times V \rightarrow \mathbb{R}$ and $g_q(t) \in V'$ are given.

3.2. Variational Form. For existence and uniqueness of a solution to (3.1) we assume that there exist constants $M_a, \alpha_a > 0$ and $\lambda_a \in \mathbb{R}$ such that for all $\phi, \psi \in V$ and all $\mu_1 \in \mathcal{D}_1$

$$(3.3) \quad |a(\mu_1; \phi, \psi)| \leq M_a \|\phi\|_V \|\psi\|_V \quad (\text{boundedness}),$$

$$(3.4) \quad a(\mu_1; \psi, \psi) + \lambda_a \|\psi\|_H^2 \geq \alpha_a \|\psi\|_V^2 \quad (\text{Gårding inequality}),$$

see (2.5). Of course, in general the constants M_a, λ_a and α_a depend on μ_1 – for simplicity we consider upper respectively lower bounds that are parameter-independent. For the space-time variational form of (3.1) we define $\mathbb{Z} := L_2(I; V)$ and consider as in [29] the following trial space

$$(3.5) \quad \mathbb{X} := \{w \in \mathbb{Z} : \dot{w} \in \mathbb{Z}'\} = L_2(I; V) \cap H^1(I; V'),$$

with the norm $\|w\|_{\mathbb{X}}^2 := \|w\|_{\mathbb{Z}}^2 + \|\dot{w}\|_{\mathbb{Z}'}^2 + \|w(T)\|_H^2$, $w \in \mathbb{X}$. The test space is $\mathbb{Y} := \mathbb{Z} \times H$ and for every $v = (z, h)$ in \mathbb{Y} the norm is defined by $\|v\|_{\mathbb{Y}}^2 := \|z\|_{\mathbb{Z}}^2 + \|h\|_H^2$. For $w \in \mathbb{X}$ and $v = (z, h) \in \mathbb{Y}$ we define

$$(3.6) \quad \begin{aligned} b(\mu_1; w, v) &:= \int_I \langle \dot{w}(t), z(t) \rangle_{V' \times V} dt + \int_I a(\mu_1; w(t), z(t)) dt + (w(0), h)_H \\ &=: b_1(\mu_1; w, z) + (w(0), h)_H \end{aligned}$$

and the right-hand side as

$$(3.7) \quad f(\mu; v) := \int_I \langle g(\mu_1; t), z(t) \rangle_{V' \times V} dt + (\mu_0, h)_H =: g_1(\mu_1; z) + (\mu_0, h)_H.$$

The space-time variational formulation of (3.1) is of the form (2.6), i.e.,

$$(3.8) \quad \text{find } u(\mu) \in \mathbb{X} \text{ such that } b(\mu_1; u(\mu), v) = f(\mu; v) \quad \forall v \in \mathbb{Y}.$$

Due to the separability of $a(\mu_1; \cdot, \cdot)$ and $g(\mu_1; \cdot)$ w.r.t. the parameter μ_1 , we get a corresponding separation for the bilinear form as well ($v = (z, h)$):

$$(3.9) \quad \begin{aligned} b(\mu_1; w, v) &= \int_I \langle \dot{w}(t), z(t) \rangle_{V' \times V} dt + \sum_{q=1}^{Q_a} \vartheta_q^a(\mu_1) \int_I a_q(w(t), z(t)) dt + (w(0), h)_H \\ &=: \sum_{q=1}^{Q_b} \vartheta_q^b(\mu_1) b_q(w, v) \end{aligned}$$

with $Q_b = Q_a + 1$, $\vartheta_q^b = \vartheta_q^a$, $1 \leq q \leq Q_a$, $\vartheta_{Q_b}^b \equiv 1$ as well as $b_q(w, v) = \int_I a_q(w(t), z(t)) dt$, $1 \leq q \leq Q_a$, and $b_{Q_b}(w, v) := (w(0), h)_H + \int_I \langle \dot{w}(t), z(t) \rangle_{V' \times V} dt$. The situation is slightly different for the right-hand side. Let $\Psi := \{\psi_m : m \in \mathbb{N}\}$ be a (Riesz-)basis of H , then $\mu_0 = \sum_{m \in \mathbb{N}} \mu_{0,m} \psi_m$, so that

$$f(\mu; v) = \sum_{q=1}^{Q_g} \vartheta_q^g(\mu_1) \int_I \langle g_q(t), z(t) \rangle_{V' \times V} dt + \sum_{m=1}^{\infty} \mu_{0,m} (\psi_m, h)_H, \quad v = (z, h),$$

which is a separation w.r.t. the parameters, but with infinitely many terms. If μ_0 has some finite expansion (even in terms of a different set of functions or also obtained by an approximation), we would get an obvious separation with $Q_f = Q_g + L$, where L denotes the number of terms in such a finite expansion of μ_0 . We will come back to this point later, in particular w.r.t. an efficient offline-online treatment of the right-hand side.

We define space-dependent quantities^a

$$M_e := \sup_{w \in \mathbb{X} \setminus \{0\}} \frac{\|w(0)\|_H}{\|w\|_{\mathbb{X}}}, \quad \varrho := \sup_{0 \neq \phi \in V} \frac{\|\phi\|_H}{\|\phi\|_V}$$

and form-dependent ones (recall (2.7))

$$\beta_a^* := \inf_{\mu_1 \in \mathcal{D}_1} \inf_{\phi \in V} \sup_{\psi \in V} \frac{a(\mu_1; \psi, \phi)}{\|\phi\|_V \|\psi\|_V}, \quad \beta_b := \inf_{\mu_1 \in \mathcal{D}_1} \inf_{w \in \mathbb{X}} \sup_{v \in \mathbb{Y}} \frac{b(\mu_1; w, v)}{\|w\|_{\mathbb{X}} \|v\|_{\mathbb{Y}}},$$

as well as lower bounds for the inf-sup-constant of the bilinear form b :

$$(3.10a) \quad \beta_{\text{coer}}^{\text{LB}} = \beta_{\text{coer}}^{\text{LB}}(\alpha_a, \lambda_a, M_a) := \frac{\min\{\min\{1, M_a^{-2}\}(\alpha_a - \lambda_a \varrho^2), 1\}}{\sqrt{2 \max\{1, (\beta_a^*)^{-2}\} + M_e^2}},$$

$$(3.10b) \quad \beta_{\text{time}}^{\text{LB}}(T) = \beta_{\text{time}}^{\text{LB}}(T, \alpha_a, \lambda_a, M_a) := \frac{\beta_{\text{coer}}^{\text{LB}}(\alpha_a, 0, M_a + \lambda_a \varrho^2) e^{-2\lambda_a T}}{\sqrt{\max\{2, 1 + 2\lambda_a^2 \varrho^4\}}}.$$

Proposition 3.1 ([31, Prop. 2.2, Cor. 2.7], [29, Thm. 5.1]). *Let $a(\mu_1; \cdot, \cdot)$ satisfy (3.3) and (3.4). Then, we obtain the inf-sup lower bound $\beta_b \geq \beta_b^{\text{LB}} := \max\{\beta_{\text{coer}}^{\text{LB}}, \beta_{\text{time}}^{\text{LB}}(T)\}$. \square*

Remark 3.2. (a) *Note that $\beta_{\text{coer}}^{\text{LB}}$ does not depend on time. However, this estimate is only meaningful (i.e., positive) if $\alpha_a - \lambda_a \varrho^2 > 0$ which means that $a(\mu_1; \cdot, \cdot)$ is coercive.*

(b) *In case of homogeneous initial conditions using*

$$(3.11) \quad \mathbb{W} := \{w \in \mathbb{X} : w(0) = 0\},$$

the above estimate holds for the form b_1 introduced in (3.6) with $M_e = 0$.

(c) *As an example, let us consider the heat equation, i.e., $A(\mu_1) \equiv -\Delta$ (no parameter dependence), $V = H_0^1(\Omega)$, $H = L_2(\Omega)$. Hence, in this case, parameters only appear in the right-hand side, i.e., $a(\mu_1; \cdot, \cdot) \equiv a(\cdot, \cdot)$ as well as $b(\mu_1; \cdot, \cdot) \equiv b(\cdot, \cdot)$ and $f(\mu; v) = f(\mu_0; v) = \int_I (g(t), z(t))_{V' \times V} dt + (\mu_0, h)_H$, $v = (z, h)$. Then, we use $\|\phi\|_V^2 = a(v, v)$, i.e., $M_a = 1$, $\lambda_a = 0$, $\alpha_a = \beta_a^* = 1$. As in [31, Cor. 2.5], we have $\beta_b \geq 1$.*

Remark 3.3. *At least for the numerical realization, we will also consider (quadratically) nonlinear problems, even though well-posedness will not be investigated here. In this case $a(\mu_1; \cdot, \cdot, \cdot) : V^3 \rightarrow \mathbb{R}$ is a trilinear form and in (3.1a) we would have the term $a(\mu_1; u(\mu; t), u(\mu; t), \phi)$ instead. The reason for this consideration is twofold: (1) Several financial models involve nonlinearities; (2) Our subsequent numerical approach using the space-time variational form particularly allows the treatment of polynomial nonlinearities, see §3.3.5 below.*

3.3. Discretization. For a linear problem, one could reduce (3.1), in particular (3.1b) to a homogeneous initial condition. However, since we aim at considering the initial condition (also) as a parameter and also have nonlinear problems in mind, we keep the inhomogeneous initial condition, so that we need to modify what has been proposed in [31], see also [1, 2].

We use (also) finite elements to construct finite dimensional subspaces $\mathbb{X}^{\mathcal{N}} \subset \mathbb{X}$, $\mathbb{Y}^{\mathcal{N}} \subset \mathbb{Y}$ and aim at determining an approximation $u^{\mathcal{N}}(\mu) \in \mathbb{X}^{\mathcal{N}}$ of the solution $u(\mu) \in \mathbb{X}$ of (3.8). The parameter \mathcal{N} will contain dimension parameters for time, space and initial value discretization spaces. Concerning notation, a (calligraphic)

^aTypically, one can bound $M_e \leq \sqrt{3}$.

superscript will always denote a large (offline, ‘truth’, detailed) dimension, whereas a reduced dimension will later be indicated with an (non-calligraphic) index. Moreover, spaces written with ‘blackboard bold’ letters (\mathbb{X}) contain space-time functions, whereas ‘normal’ letters are used for functions in space *or* time.

In order to introduce the discretization, we basically follow [29, 31] and note that

$$(3.12) \quad \mathbb{X} = H^1(I) \otimes V, \quad \mathbb{Y} = \mathbb{Z} \times H := L_2(I; V) \times H = (L_2(I) \otimes V) \times H.$$

Let $\mathcal{T}_{\text{space}}^{\mathcal{J}}$ be a triangulation of the underlying space Ω . For discretizing the function space V we consider the finite subspace $V^{\mathcal{J}}$ spanned by a nodal basis $\{\phi_1, \dots, \phi_{\mathcal{J}}\}$ with respect to the triangulation $\mathcal{T}_{\text{space}}^{\mathcal{J}}$.

For the finite-dimensional temporal subspaces $E^{\mathcal{K}} \subset H_{\{0\}}^1(I) := \{\phi \in H^1(I) : \phi(0) = 0\}$ and $F^{\mathcal{K}} \subset L_2(I)$ consider the discretization of the interval I given by $\mathcal{T}_{\text{time}}^{\mathcal{K}} := \{t^k = k \Delta t : 0 \leq k \leq \mathcal{K}, \Delta t := \frac{T}{\mathcal{K}}\}$. The trial space $E^{\mathcal{K}}$ is spanned by the piecewise linear functions $\{\sigma^1, \dots, \sigma^{\mathcal{K}}\}$ w.r.t. $\mathcal{T}_{\text{time}}^{\mathcal{K}}$. For every $1 \leq k \leq \mathcal{K} - 1$, we choose σ^k as the hat function with nodes t^{k-1} , t^k and t^{k+1} and the remaining ones are defined by $\sigma^{\mathcal{K}} := \frac{t-t^{\mathcal{K}-1}}{\Delta t} \chi_{[t^{\mathcal{K}-1}, t^{\mathcal{K}}]}$ and $\sigma^0 := \frac{t^1-t}{\Delta t} \chi_{[0, t^1]}$ (which will be needed later). The test space $F^{\mathcal{K}}$ is chosen as $\text{span}\{\tau^1, \dots, \tau^{\mathcal{K}}\}$ with respect to $\mathcal{T}_{\text{time}}^{\mathcal{K}}$ where $\tau^k \equiv \chi_{I^k}$ is the characteristic function on $I^k := (t^{k-1}, t^k]$.

Finally, the trial (and test) space $H^{\mathcal{M}} \subset H$ for the initial condition is denoted by $H^{\mathcal{M}} := \text{span}\{\psi_1, \dots, \psi_{\mathcal{M}}\} \subset V \subset H^a$. We keep $H^{\mathcal{M}}$ and in particular \mathcal{M} arbitrary here and will detail possible choices later. In particular, we also allow for the case $H^{\mathcal{M}} = V^{\mathcal{J}}$ ($\mathcal{M} = \mathcal{J}$) and discuss advantages and disadvantages of this choice.

With these preparations at hand, the discrete approximation subspaces of \mathbb{X} and \mathbb{Y} are defined as ($\langle \sigma^0 \rangle := \text{span}\{\sigma^0\}$)

$$(3.13a) \quad \mathbb{X}^{\mathcal{N}} := (\langle \sigma^0 \rangle \otimes H^{\mathcal{M}}) \oplus (E^{\mathcal{K}} \otimes V^{\mathcal{J}}) =: \mathbb{Q}^{\mathcal{M}} \oplus \mathbb{W}^{\mathcal{I}}, \quad \mathcal{I} = \mathcal{K} \cdot \mathcal{J},$$

$$(3.13b) \quad \mathbb{Y}^{\mathcal{N}} := (F^{\mathcal{K}} \otimes V^{\mathcal{J}}) \times H^{\mathcal{M}} =: \mathbb{Z}^{\mathcal{I}} \times H^{\mathcal{M}b}.$$

Since $\dim(E^{\mathcal{K}}) = \dim(F^{\mathcal{K}}) = \mathcal{K}$, $\dim(H^{\mathcal{M}}) = \mathcal{M}$ and $\dim(V^{\mathcal{J}}) = \mathcal{J}$, we have $\dim(\mathbb{X}^{\mathcal{N}}) = \mathcal{M} + \mathcal{K}\mathcal{J} =: \mathcal{N} = \mathcal{I} + \mathcal{M} = \dim(\mathbb{Y}^{\mathcal{N}})$. In case the discretized versions of \mathbb{X} and \mathbb{Y} would have different dimensions, a least squares method has to be used, [1]. Finally, note that $\mathbb{W}^{\mathcal{I}} \subset \mathbb{W}$, see (3.11). We obtain a discrete variational formulation of (3.8): Find $u^{\mathcal{N}}(\mu) \in \mathbb{X}^{\mathcal{N}}$ such that

$$(3.14) \quad b(\mu_1; u^{\mathcal{N}}(\mu), v^{\mathcal{N}}) = f(\mu; v^{\mathcal{N}}) \quad \forall v^{\mathcal{N}} \in \mathbb{Y}^{\mathcal{N}}, \quad \mu = (\mu_0, \mu_1) \in \mathcal{D},$$

which corresponds to a linear system that can be detailed as

$$b_1(\mu_1; u^{\mathcal{N}}(\mu), z^{\mathcal{I}}) + ((u^{\mathcal{N}}(\mu))(0), h^{\mathcal{M}})_H = g(\mu_1; z^{\mathcal{I}}) + (\mu_0, h^{\mathcal{M}})_H$$

for all $v^{\mathcal{N}} = (z^{\mathcal{I}}, h^{\mathcal{M}}) \in \mathbb{Y}^{\mathcal{N}} = \mathbb{Z}^{\mathcal{I}} \times H^{\mathcal{M}}$.

3.3.1. Time-stepping. Similar to [31], it is not difficult to see that the above discretization is equivalent to a time-stepping scheme. In fact, we have the splitting $u^{\mathcal{N}} = q^{\mathcal{M}} + w^{\mathcal{I}} \in \mathbb{Q}^{\mathcal{M}} \oplus \mathbb{W}^{\mathcal{I}} = \mathbb{X}^{\mathcal{N}}$, in particular $u^{\mathcal{N}}(0) = q^{\mathcal{M}}(0)$ and $w^{\mathcal{I}}(0) = 0$.

^aThe reason to impose $\psi_{\ell} \in V$ will also become clear a little later.

^bFor the test space, we could also replace $H^{\mathcal{M}}$ by some other space, possibly also with different dimension. Here, we want to keep this issue simple.

In terms of the respective bases, we get the representations

$$(3.15) \quad u^{\mathcal{N}} := q^{\mathcal{M}} + w^{\mathcal{I}} = \sum_{m=1}^{\mathcal{M}} q_m (\sigma^0 \otimes \psi_m) + \sum_{k=1}^{\mathcal{K}} \sum_{i=1}^{\mathcal{J}} w_i^k (\sigma^k \otimes \phi_i) \in \mathbb{X}^{\mathcal{N}},$$

with the coefficient vectors $\mathbf{q}_{\mathcal{M}} := (q_m)_{m=1, \dots, \mathcal{M}}$ for the initial value as well as $\mathbf{w}_{\mathcal{I}} := (w_i^k)_{i=1, \dots, \mathcal{J}, k=1, \dots, \mathcal{K}} =: (\mathbf{w}_{\mathcal{I}}^k)_{k=1, \dots, \mathcal{K}}$ and (recall $\mathcal{N} = \mathcal{M} + \mathcal{I}$)

$$(3.16) \quad v^{\mathcal{N}} = (z^{\mathcal{I}}, h^{\mathcal{M}}) = \left(\sum_{\ell=1}^{\mathcal{K}} \sum_{j=1}^{\mathcal{J}} z_j^\ell (\tau^\ell \otimes \phi_j), \sum_{m=1}^{\mathcal{M}} h_m \psi_m \right) \in \mathbb{Y}^{\mathcal{N}},$$

with the coefficient vectors $\mathbf{z}_{\mathcal{I}} := (z_j^\ell)_{j=1, \dots, \mathcal{J}, \ell=1, \dots, \mathcal{K}} = (\mathbf{z}_{\mathcal{I}}^\ell)_{\ell=1, \dots, \mathcal{K}}$ and $\mathbf{h}_{\mathcal{M}} := (h_m)_{m=1, \dots, \mathcal{M}}$. With these notations, we obtain

$$\begin{aligned} b_1(\mu_1; u^{\mathcal{N}}, z^{\mathcal{I}}) &= \int_I \langle \dot{u}^{\mathcal{N}}(t), z^{\mathcal{I}}(t) \rangle_{V' \times V} + a(\mu_1; u^{\mathcal{N}}(t), z^{\mathcal{I}}(t)) dt \\ &= \sum_{m=1}^{\mathcal{M}} \sum_{\ell=1}^{\mathcal{K}} \sum_{j=1}^{\mathcal{J}} q_m z_j^\ell \int_I \langle \dot{\sigma}^0(t) \psi_m, \tau^\ell(t) \phi_j \rangle_{V' \times V} + a(\mu_1; \sigma^0(t) \psi_m, \tau^\ell(t) \phi_j) dt \\ &\quad + \sum_{k=1}^{\mathcal{K}} \sum_{\ell=1}^{\mathcal{K}} \sum_{i,j=1}^{\mathcal{J}} w_i^k z_j^\ell \int_I \langle \dot{\sigma}^k(t) \phi_i, \tau^\ell(t) \phi_j \rangle_{V' \times V} + a(\mu_1; \sigma^k(t) \phi_i, \tau^\ell(t) \phi_j) dt \\ &= \sum_{m=1}^{\mathcal{M}} \sum_{\ell=1}^{\mathcal{K}} \sum_{j=1}^{\mathcal{J}} q_m z_j^\ell \left\{ (\dot{\sigma}^0, \tau^\ell)_{L_2(I)}(\psi_m, \phi_j)_H + (\sigma^0, \tau^\ell)_{L_2(I)} a(\mu_1; \psi_m, \phi_j) \right\} \\ &\quad + \sum_{k, \ell=1}^{\mathcal{K}} \sum_{i,j=1}^{\mathcal{J}} w_i^k z_j^\ell \left\{ (\dot{\sigma}^k, \tau^\ell)_{L_2(I)}(\phi_i, \phi_j)_H + (\sigma^k, \tau^\ell)_{L_2(I)} a(\mu_1; \phi_i, \phi_j) \right\}. \end{aligned}$$

Note, that $a(\mu_1; \psi_m, \phi_j)$ is well-defined since we have assumed that $\psi_m \in V$. For $k \geq 0$ and $\ell \geq 1$ we have $(\dot{\sigma}^k, \tau^\ell)_{L_2(I)} = \delta_{k, \ell} - \delta_{k+1, \ell}$ and $(\sigma^k, \tau^\ell)_{L_2(I)} = \frac{\Delta t}{2} (\delta_{k, \ell} + \delta_{k+1, \ell})$, in particular $(\dot{\sigma}^0, \tau^\ell)_{L_2(I)} = -\delta_{1, \ell}$ and $(\sigma^0, \tau^\ell)_{L_2(I)} = \frac{\Delta t}{2} \delta_{1, \ell}$.

Similar to [31], we set

$$(3.17) \quad \mathbf{B}^{\mathcal{I}}(\mu_1) := \mathbf{N}_{\text{time}}^{\mathcal{K}} \otimes \mathbf{M}_{\text{space}}^{\mathcal{J}} + \mathbf{M}_{\text{time}}^{\mathcal{K}} \otimes \mathbf{A}_{\text{space}}^{\mathcal{J}}(\mu_1) \in \mathbb{R}^{\mathcal{K}\mathcal{J} \times \mathcal{K}\mathcal{J}} = \mathbb{R}^{\mathcal{I} \times \mathcal{I}},$$

where the temporal matrices read $\mathbf{N}_{\text{time}}^{\mathcal{K}} := ((\dot{\sigma}^k, \tau^\ell)_{L_2(I)})_{k, \ell=1, \dots, \mathcal{K}}$, $\mathbf{M}_{\text{time}}^{\mathcal{K}} := ((\sigma^k, \tau^\ell)_{L_2(I)})_{k, \ell=1, \dots, \mathcal{K}}$ and the spatial ones are $\mathbf{M}_{\text{space}}^{\mathcal{J}} := ((\phi_i, \phi_j)_H)_{i, j=1, \dots, \mathcal{J}}$ and $\mathbf{A}_{\text{space}}^{\mathcal{J}}(\mu_1) = (a(\mu_1; \phi_i, \phi_j))_{i, j=1, \dots, \mathcal{J}}$. The matrix $\mathbf{B}^{\mathcal{I}}(\mu_1)$ was used in [31, (2.14)]

to describe and analyze the discretization in the case of homogeneous initial conditions. Then, we obtain

$$\begin{aligned}
b_1(\mu_1; u^{\mathcal{N}}, z^{\mathcal{I}}) &= \sum_{k,\ell=1}^{\mathcal{K}} \sum_{i,j=1}^{\mathcal{J}} w_i^k z_j^\ell (\mathbf{B}^{\mathcal{I}}(\mu_1))_{(k,i),(\ell,j)} \\
&\quad + \sum_{m=1}^{\mathcal{M}} \sum_{\ell=1}^{\mathcal{K}} \sum_{j=1}^{\mathcal{J}} q_m z_j^\ell \left\{ (\dot{\sigma}^0, \tau^\ell)_{L_2(I)}(\psi_m, \phi_j)_H + (\sigma^0, \tau^\ell)_{L_2(I)} a(\mu_1; \psi_m, \phi_j) \right\} \\
&= \mathbf{w}_{\mathcal{I}}^T \mathbf{B}^{\mathcal{I}}(\mu_1) \mathbf{z}_{\mathcal{I}} + \sum_{m=1}^{\mathcal{M}} \sum_{j=1}^{\mathcal{J}} q_m z_j^1 \left\{ -(\psi_m, \phi_j)_H + \frac{\Delta t}{2} a(\mu_1; \psi_m, \phi_j) \right\} \\
&= \mathbf{w}_{\mathcal{I}}^T \mathbf{B}^{\mathcal{I}}(\mu_1) \mathbf{z}_{\mathcal{I}} + \mathbf{q}_{\mathcal{M}}^T \left(-\mathbf{M}_{i/s}^{\mathcal{M},\mathcal{J}} + \frac{\Delta t}{2} \mathbf{A}_{i/s}^{\mathcal{M},\mathcal{J}}(\mu_1) \right) \mathbf{z}_{\mathcal{I}}^1 \\
(3.18) \quad &=: \mathbf{w}_{\mathcal{I}}^T \mathbf{B}^{\mathcal{I}}(\mu_1) \mathbf{z}_{\mathcal{I}} + \mathbf{q}_{\mathcal{M}}^T \mathbf{C}_{i/s}^{\mathcal{M},\mathcal{J}}(\mu_1) \mathbf{z}_{\mathcal{I}}^1,
\end{aligned}$$

where the involved matrices $\mathbf{M}_{i/s}^{\mathcal{M},\mathcal{J}}, \mathbf{A}_{i/s}^{\mathcal{M},\mathcal{J}}(\mu_1) \in \mathbb{R}^{\mathcal{M} \times \mathcal{J}}$ are defined for $m = 1, \dots, \mathcal{M}, j = 1, \dots, \mathcal{J}$ as

$$(3.19) \quad (\mathbf{M}_{i/s}^{\mathcal{M},\mathcal{J}})_{m,j} := (\psi_m, \phi_j)_H, \quad (\mathbf{A}_{i/s}^{\mathcal{M},\mathcal{J}}(\mu_1))_{m,j} := a(\mu_1; \psi_m, \phi_j)_H.$$

We split the coefficient vector for the unknown $\mathbf{u}_{\mathcal{N}} = (\mathbf{q}_{\mathcal{M}}, \mathbf{w}_{\mathcal{I}})^T$ and set $\mathbf{u}_{\mathcal{N}}^0 := \mathbf{q}_{\mathcal{M}}$ as well as $\mathbf{u}_{\mathcal{N}}^k := \mathbf{w}_{\mathcal{I}}^k$, $k = 1, \dots, \mathcal{K}$ in order to formulate the time-stepping scheme. Then, we obtain for fixed $\ell \geq 1$ and any $j \in \{1, \dots, n_h\}$

$$\begin{aligned}
b_1(\mu_1; u^{\mathcal{N}}, \tau^\ell \otimes \phi_j) &= \\
&= \begin{cases} [\mathbf{M}_{\text{space}}^{\mathcal{J}} \mathbf{u}_{\mathcal{N}}^1 - (\mathbf{M}_{i/s}^{\mathcal{M},\mathcal{J}})^T \mathbf{u}_{\mathcal{N}}^0 + \frac{\Delta t}{2} (\mathbf{A}_{\text{space}}^{\mathcal{J}}(\mu_1) \mathbf{u}_{\mathcal{N}}^1 + (\mathbf{A}_{i/s}^{\mathcal{M},\mathcal{J}}(\mu_1))^T \mathbf{u}_{\mathcal{N}}^0)]_j & \text{if } \ell = 1, \\ \Delta t [\mathbf{M}_{\text{space}}^{\mathcal{J}} \frac{1}{\Delta t} (\mathbf{u}_{\mathcal{N}}^\ell - \mathbf{u}_{\mathcal{N}}^{\ell-1}) + \mathbf{A}_{\text{space}}^{\mathcal{J}}(\mu_1) \frac{1}{2} (\mathbf{u}_{\mathcal{N}}^\ell + \mathbf{u}_{\mathcal{N}}^{\ell-1})]_j, & \text{if } \ell > 1 \end{cases} \\
(3.20) \quad &=: \Delta t \left[\frac{1}{\Delta t} \mathbf{M}_{\mathcal{N}}^\ell (\mathbf{u}_{\mathcal{N}}^\ell - \mathbf{u}_{\mathcal{N}}^{\ell-1}) + \mathbf{A}_{\mathcal{N}}^\ell(\mu_1) \mathbf{u}_{\mathcal{N}}^{\ell-1/2} \right]_j,
\end{aligned}$$

where $\mathbf{u}_{\mathcal{N}}^{\ell-1/2} := \frac{1}{2}(\mathbf{u}_{\mathcal{N}}^\ell + \mathbf{u}_{\mathcal{N}}^{\ell-1})$. On the right-hand side we use a trapezoidal approximation

$$\begin{aligned}
(3.21) \quad g(\mu_1; \tau^\ell \otimes \phi_j) &= \int_I \langle g(\mu_1; t), \tau^\ell \otimes \phi_j(t, \cdot) \rangle_{V' \times V} dt = \int_I \langle g(\mu_1; t), \tau^\ell(t) \phi_j \rangle_{V' \times V} dt \\
&\approx \frac{\Delta t}{2} \langle g(\mu_1; t^{\ell-1}) + g(\mu_1; t^\ell), \phi_j \rangle_{V' \times V} =: \Delta t (\mathbf{g}_{\mathcal{N}}^{\ell-1/2}(\mu_1))_j.
\end{aligned}$$

3.3.2. Initial value approximation. Let us now discuss the approximation $\mathbf{u}_{\mathcal{N}}^0$ (given by $\mathbf{q}_{\mathcal{M}}$) of the coefficients of the initial value $u(0) = \mu_0$. Since $u^{\mathcal{N}}(0) = q^{\mathcal{M}}(0)$, $q^{\mathcal{M}} \in \mathbb{Q}^{\mathcal{M}}$, we have $u^{\mathcal{N}}(0) = \sum_{m=1}^{\mathcal{M}} q_m (\sigma^0 \otimes \psi_m)(0) = \sum_{m=1}^{\mathcal{M}} q_m \psi_m \in H^{\mathcal{M}}$, so that for $h^{\mathcal{M}} = \sum_{m=1}^{\mathcal{M}} h_m \psi_m \in H^{\mathcal{M}}$, we get

$$(u^{\mathcal{N}}(0), h^{\mathcal{M}})_H = \sum_{m,m'=1}^{\mathcal{M}} q_{m'} h_m (\psi_{m'}, \psi_m)_H = \mathbf{q}_{\mathcal{M}}^T \mathbf{M}_{\text{init}}^{\mathcal{M}} \mathbf{h}_{\mathcal{M}},$$

where $\mathbf{h}_{\mathcal{M}} := (h_m)_{1 \leq m \leq \mathcal{M}}$ and $\mathbf{M}_{\text{init}}^{\mathcal{M}} := ((\psi_{m'}, \psi_m)_H)_{m',m=1,\dots,\mathcal{M}}$. The right-hand side of the discretization of (3.1b) for obtaining an approximation of the initial

condition reads for the same $h^{\mathcal{M}} \in H^{\mathcal{M}}$

$$(\mu_0, h^{\mathcal{M}})_H = \sum_{m=1}^{\mathcal{M}} h_m (\mu_0, \psi_m)_H,$$

which is only computable if μ_0 admits a finite expression. We refer to Remark 3.5 below for possible choices.

At this point we will only assume that a discretization of \mathcal{D}_0 is given by $\mathcal{D}_0^{\mathcal{L}} = \text{span}\{B_1, \dots, B_{\mathcal{L}}\}$, i.e., we consider $\mu_0^{\mathcal{L}} = \sum_{\ell=1}^{\mathcal{L}} \mu_0^{\ell} B_{\ell}$ and obtain

$$(\mu_0^{\mathcal{L}}, h^{\mathcal{M}})_H = \sum_{\ell=1}^{\mathcal{L}} \sum_{m=1}^{\mathcal{M}} \mu_0^{\ell} h_m (B_{\ell}, \psi_m)_H = \boldsymbol{\mu}_{0,\mathcal{L}}^T \mathbf{N}^{\mathcal{L},\mathcal{M}} \mathbf{h}_{\mathcal{M}},$$

where $\mathbf{N}^{\mathcal{L},\mathcal{M}} := ((B_{\ell}, \psi_m)_H)_{\ell=1,\dots,\mathcal{L}; m=1,\dots,\mathcal{M}}$, $\boldsymbol{\mu}_{0,\mathcal{L}} = (\mu_0^{\ell})_{\ell=1,\dots,\mathcal{L}}$.

3.3.3. Crank-Nicolson scheme. Putting everything together, we obtain the following Crank-Nicolson scheme for computing $\mathbf{u}_{\mathcal{N}} = (\mathbf{u}_{\mathcal{N}}^0, \dots, \mathbf{u}_{\mathcal{N}}^{\mathcal{K}})$ (recall (3.20)):

$$(3.22a) \quad (\mathbf{M}_{\text{init}}^{\mathcal{M}})^T \mathbf{u}_{\mathcal{N}}^0 = (\mathbf{N}^{\mathcal{L},\mathcal{M}})^T \boldsymbol{\mu}_{0,\mathcal{L}},$$

$$(3.22b) \quad \frac{1}{\Delta t} \mathbf{M}_{\mathcal{N}}^k (\mathbf{u}_{\mathcal{N}}^k - \mathbf{u}_{\mathcal{N}}^{k-1}) + \mathbf{A}_{\mathcal{N}}^k(\mu_1) \mathbf{u}_{\mathcal{N}}^{k-1/2} = \mathbf{g}_{\mathcal{N}}^{k-1/2}(\mu_1), \quad k = 1, \dots, \mathcal{K}.$$

Note, that $\mathbf{M}_{\text{init}}^{\mathcal{M}}$ is regular, so that $\mathbf{u}_{\mathcal{N}}^0 = \mathbf{u}_{\mathcal{N}}^0(\mu_0)$ is uniquely defined and the discrete problem (3.22) is obviously well-posed.

Remark 3.4. Let us now discuss some relevant special cases.

(a) Let $H^{\mathcal{M}} = V^{\mathcal{J}}$, $\mathcal{M} = \mathcal{J}$. In this case, we get in (3.19) that $\mathbf{M}_{\text{i/s}}^{\mathcal{M},\mathcal{J}} \equiv \mathbf{M}_{\text{space}}^{\mathcal{J}}$ and $\mathbf{A}_{\text{i/s}}^{\mathcal{M},\mathcal{J}} \equiv \mathbf{A}_{\text{space}}^{\mathcal{J}}$. This means that we do not need to distinguish the cases $\ell = 1$ and $\ell > 1$ in (3.20), i.e., we get a standard Crank-Nicolson scheme with initial value $\mathbf{u}_{\mathcal{N}}^0$. Moreover, $\mathbf{M}_{\text{init}}^{\mathcal{M}} = \mathbf{M}_{\text{space}}^{\mathcal{J}}$ (which is symmetric and positive definite (s.p.d.)) and $\mathbf{N}^{\mathcal{L},\mathcal{M}} = \mathbf{N}^{\mathcal{L},\mathcal{J}} := ((B_{\ell}, \phi_j)_H)_{1 \leq \ell \leq \mathcal{L}; 1 \leq j \leq \mathcal{J}}$, i.e., we obtain $\mathbf{u}_{\mathcal{N}}^0 = (\mathbf{M}_{\text{space}}^{\mathcal{J}})^{-1} (\mathbf{N}^{\mathcal{L},\mathcal{J}})^T \boldsymbol{\mu}_{0,\mathcal{L}} \in \mathbb{R}^{\mathcal{J}}$, which can be used as initial value for the Crank-Nicolson scheme (3.22b).

(b) If μ_0 can be represented as (or approximated by) $\mu_0 = \mu_0^{\mathcal{M}} = \sum_{m=1}^{\mathcal{M}} \mu_0^m \psi_m \in H^{\mathcal{M}} \neq V^{\mathcal{J}}$, $\boldsymbol{\mu}_{0,\mathcal{M}} = (\mu_0^m)_{m=1,\dots,\mathcal{M}}$, then $\mathcal{L} = \mathcal{M}$, $\mathcal{D}_0^{\mathcal{L}} = H^{\mathcal{M}}$, $\mathbf{N}^{\mathcal{L},\mathcal{M}} = \mathbf{M}_{\text{init}}^{\mathcal{M}}$, which is s.p.d., so that $\mathbf{u}_{\mathcal{N}}^0 = \boldsymbol{\mu}_{0,\mathcal{M}} \in \mathbb{R}^{\mathcal{M}}$. In this case, we need to modify the first step of the Crank-Nicolson scheme as we do not need to solve a linear system in (3.22a).

Remark 3.5. In view of Remark 3.4 (b) above, let us describe further scenarios for the approximation of $\mu_0 \approx \mu_0^{\mathcal{L}} \in \mathcal{D}_0^{\mathcal{L}} \subset H$ that we have in mind:

(1) If $\Xi = \{\xi_{\ell} : \ell \in \mathbb{N}\}$ is a Riesz basis for H , then μ_0 has an expansion in that basis, i.e., $\mu_0 = \sum_{\ell \in \mathbb{N}} \mu_0^{\ell} \xi_{\ell}$. Then, $\Xi_{\mathcal{L}} := \{\xi_1, \dots, \xi_{\mathcal{L}}\}$, $\mathcal{D}_0^{\mathcal{L}} := \text{span}(\Xi_{\mathcal{L}})$ may be selected as the ‘most significant’ parts of the infinite expansion, e.g. by an adaptive approximation. The above approximation is then obtained using the corresponding expansion coefficients $\mu_0^1, \dots, \mu_0^{\mathcal{L}}$ or approximations of them (if they cannot be computed exactly).

(2) Sometimes, the specific structure of possible initial values is known from the particular application. Then, it might be realistic (as in fact for some payoff

functions in option pricing) that $\mu_0 \in \text{span}\{B_1, \dots, B_{\mathcal{L}}\} = \mathcal{D}_0^{\mathcal{L}} \neq H^{\mathcal{M}}$. In this situation we can directly write $\mu_0 = \sum_{\ell=1}^{\mathcal{L}} \beta_{\ell} B_{\ell}$ and $\mathbf{u}_0^{\mathcal{M}}$ is given as in (3.22a).

(3) As mentioned in Remark 3.4 (b), our approach particularly allows to choose $\mathcal{D}_0^{\mathcal{L}}$ as $H^{\mathcal{M}}$ – even though this results in an immediate smoothing of $\mu_0 \in H$, since $H^{\mathcal{M}} \subset V \subsetneq H$.

3.3.4. Parameter Separation. For later reference, we now detail the specific parameter-dependence of the discrete variational formulation. Obviously, $\mathbf{u}_{\mathcal{N}}^0$ can be computed by (3.22a) in dependency of μ_0 (or its approximation $\mu_0^{\mathcal{L}}$) in the sense that $\mu_0^{\mathcal{L}}$ uniquely determines $\mathbf{u}_{\mathcal{N}}^0$ – μ_1 is not required. This can be formulated as follows. Recall from (3.13a) that $\mathbb{X}^{\mathcal{N}} = \mathbb{Q}^{\mathcal{M}} \oplus \mathbb{W}^{\mathcal{I}}$, in particular

$$\mathbb{W}^{\mathcal{I}} = \{w_{\mathcal{N}} \in \mathbb{X}^{\mathcal{N}} : w_{\mathcal{N}}(0) = 0\} = \text{span}\{\sigma^k \otimes \phi_i : k = 1, \dots, \mathcal{K}, i = 1, \dots, \mathcal{J}\},$$

compare (3.11). Then, (3.14) can be divided as follows:

$$(3.23a) \quad \xi^{\mathcal{M}}(\mu_0) \in H^{\mathcal{M}} : \quad (\xi^{\mathcal{M}}(\mu_0), h^{\mathcal{M}})_H = (\mu_0, h^{\mathcal{M}})_H \quad \forall h^{\mathcal{M}} \in H^{\mathcal{M}},$$

$$(3.23b) \quad w^{\mathcal{I}}(\mu) \in \mathbb{W}^{\mathcal{I}} : \quad b_1(\mu_1; w^{\mathcal{I}}(\mu), z^{\mathcal{I}}) = \check{f}(q^{\mathcal{M}}(\mu_0), \mu_1; z^{\mathcal{I}}) \quad \forall z^{\mathcal{I}} \in \mathbb{Z}^{\mathcal{I}},$$

with the extension of the initial value $q^{\mathcal{M}}(\mu_0) := \sigma^0 \otimes \xi^{\mathcal{M}}(\mu_0) \in \mathbb{Q}^{\mathcal{M}}$ and the modified right-hand side $\check{f}(q^{\mathcal{M}}(\mu_0), \mu_1; z^{\mathcal{I}}) := g_1(\mu_1; z^{\mathcal{I}}) - b_1(\mu_1; q^{\mathcal{M}}(\mu_0), z^{\mathcal{I}})$. In matrix-vector form as (3.18) the second equation (3.23b) reads

$$\begin{aligned} \mathbf{B}^{\mathcal{I}}(\mu_1)^T \mathbf{w}_{\mathcal{I}}(\mu) &= \check{\mathbf{f}}(\mathbf{u}_{\mathcal{N}}^0(\mu_0), \mu_1) \\ &:= (g_1(\mu_1; \phi_i))_{i=1, \dots, \mathcal{J}} - (\mathbf{C}_{i/s}^{\mathcal{M}, \mathcal{J}}(\mu_1)^T \mathbf{u}_{\mathcal{N}}^0(\mu_0), 0, \dots, 0)^T. \end{aligned}$$

The arising coefficient vectors define functions

$$w^{\mathcal{I}}(\mu) = \sum_{k=1}^{\mathcal{K}} \sum_{i=1}^{\mathcal{J}} w_i^k(\mu) (\sigma^k \otimes \phi_i) \in \mathbb{W}^{\mathcal{I}}, \quad q^{\mathcal{M}}(\mu_0) = \sum_{m=1}^{\mathcal{M}} q_m(\mu_0) (\sigma^0 \otimes \psi_m) \in \mathbb{Q}^{\mathcal{M}},$$

so that $w^{\mathcal{N}}(\mu) := q^{\mathcal{M}}(\mu_0) + w^{\mathcal{I}}(\mu) \in \mathbb{X}^{\mathcal{N}}$ is the desired approximate solution.

We stress the fact that (3.23) can also be interpreted as a separation. In fact, (3.23a) determines an approximation of the initial value μ_0 – independent of μ_1 , whereas the evolution is determined in (3.23b) and – as we have seen – could be realized for example in terms of a Crank-Nicolson scheme.

3.3.5. Nonlinear equations. Note that the above mentioned separation is *not* a consequence of the fact that a *linear* parabolic problem allows one to reduce non-homogeneous initial conditions to homogeneous ones. In fact, if $a(\mu; \cdot, \cdot, \cdot)$ would be a trilinear form that induces a space-time trilinear form $b_1(\mu_1; \cdot, \cdot, \cdot)$, the analogue of (3.23b) would read

$$\begin{aligned} b_1(\mu_1; w^{\mathcal{I}}(\mu), w^{\mathcal{I}}(\mu), z^{\mathcal{I}}) + b_1(\mu_1; q^{\mathcal{M}}(\mu_0), w^{\mathcal{I}}(\mu), z^{\mathcal{I}}) + b_1(\mu_1; w^{\mathcal{I}}(\mu), q^{\mathcal{M}}(\mu_0), z^{\mathcal{I}}) \\ = g(\mu_1; z^{\mathcal{I}}) - b_1(\mu_1; q^{\mathcal{M}}(\mu_0), q^{\mathcal{M}}(\mu_0), z^{\mathcal{I}}), \end{aligned}$$

i.e., the quadratic term is supplemented by two linear terms since $q^{\mathcal{M}}(\mu_0)$ is known at this stage. This also shows how to extend this approach to polynomial nonlinearities. The only difference is that the parameter induced by the initial value approximation is only in the right-hand side for a linear problem, but also appears within the coefficients of the PDE for polynomial nonlinearities.

3.4. Stability of the discretization. We have to show well-posedness of the discretized system, i.e., continuity, surjectivity and an inf-sup-condition for the linear operator induced by the bilinear form $b(\mu_1; \cdot, \cdot)$ on $\mathbb{X}^{\mathcal{N}} \times \mathbb{Y}^{\mathcal{N}}$, see (2.8). Continuity and surjectivity are (more or less) readily seen.

For the case of homogeneous initial conditions, i.e. for $b_1(\mu_1; \cdot, \cdot)$, inf-sup-stability was investigated in [31] under the condition that the bilinear form $a(\mu_1; \cdot, \cdot)$ satisfies a Gårding inequality (3.4), see also Proposition 3.1. However, the inf-sup-constant may deteriorate in the presence of strong advective terms, which is the case in the considered option pricing models. In that case one might require some sort of (known) stabilization.

In the general case, using the splitting (3.23), we could deduce the stability of the discrete problem (3.14) from known results for both sub-problems (initial value approximation and Crank-Nicolson iteration). However, as Proposition 2.2 shows, we need an explicit estimate for the discrete inf-sup-constant, which is *not* easily obtained from the combination of the sub-problems.

For the special case presented in Remark 3.4 (a) (i.e., $H^{\mathcal{M}} = V^{\mathcal{J}}$), the inf-sup stability was investigated in [1, 2]. In that case and for our chosen time discretization we only have to ensure a standard CFL condition (cf. [2, Prop. 2]) in order to obtain inf-sup-stability and also an estimate for $\beta^{\mathcal{N}}$ in (2.8).

4. A REDUCED BASIS METHOD (RBM) FOR PARAMETER FUNCTIONS

Now, we consider a Reduced Basis (RB) approximation for the Crank-Nicolson interpretation of the discrete space-time problem. Recall that the space-time variational formulation leads to a Petrov-Galerkin problem so that the reduced problem takes the form (2.9), where the bilinear form $b(\mu_1; \cdot, \cdot)$ only depends on μ_1 , whereas the right-hand side $f(\mu; \cdot)$ depends on the full parameter $\mu = (\mu_0, \mu_1)$.

As already pointed out, the parameter μ_0 is a function. We are now going to describe a method to handle this challenge. For $\mu := (\mu_0, \mu_1) \in \mathcal{D}_0 \times \mathcal{D}_1$ the residual reads

$$\begin{aligned} r_N(\mu; v) &= f(\mu; v) - b(\mu; u_N(\mu), v) \\ &= g_1(\mu_1; z) + (\mu_0, h)_H - b_1(\mu_1; u_N(\mu), z) + (u_N(\mu)(0), h)_H \\ &= g_1(\mu_1; z) - b_1(\mu_1; u_N(\mu), z) + (\mu_0 - (u_N(\mu))(0), h)_H \\ &=: r_{N,1}(\mu; z) + r_{N,0}(\mu; h), \end{aligned}$$

for any $v = (z, h) \in \mathbb{Y}^{\mathcal{N}}$. Recall, that we need to construct a reduced basis that ensures a small residuum for the full parameter space \mathcal{D} . In order to do so, we need an efficient online computation of the error estimator $\Delta_N(\mu)$ in (2.11), which requires a separation of the residual $r_N(\mu; v)$ into parts that depend only on μ and others depending only on v . This is no problem for $r_{N,1}(\mu; z)$ due to the separation properties of g_1 and b_1 . However, the term $(\mu_0, h)_H$ is an issue since the inner product involves the parameter $\mu_0 \in \mathcal{D}_0$ and would be needed to be computed online e.g. in terms of a possibly costly quadrature.

4.1. A two-step greedy method. To construct a reduced basis, first assume that μ_0 is, or can at least be well approximated by, a finite sum, i.e., $\mu_0 \approx \mu_0^{\mathcal{L}} \in \mathcal{D}_0^{\mathcal{L}} = \text{span}\{B_1, \dots, B_{\mathcal{L}}\}$. Then, we would get an usual separation of the form $(\mu_0, h)_H = \sum_{\ell=0}^{\mathcal{L}} \beta_{\ell}(\mu_0) (B_{\ell}, h)_H$, where the terms $(B_{\ell}, h)_H$ can be precomputed offline. Constructing a basis with a standard greedy procedure would be possible

by considering an \mathcal{L} -dimensional parameter space containing the coefficients β_ℓ . However, if \mathcal{L} is large, this is infeasible, also since the coefficients cannot easily be bounded, so that we would need to work with a \mathcal{L} -dimensional hypercube of large ‘side lengths’. Consequently, determining parameter samples for snapshots e.g. by a greedy method might be extremely costly. Moreover, the use of $\mathcal{D}_0^{\mathcal{L}}$ might be a severe restriction to possible choices of the initial value μ_0 .

Hence, we need an alternative and consider again the residuum. The following estimate is immediate

$$\begin{aligned} \|r_N(\mu)\|_{\mathbb{Y}'} &= \sup_{v \in \mathbb{Y}} \frac{r_N(\mu; v)}{\|v\|_{\mathbb{Y}}} = \sup_{(z, h) \in \mathbb{Y}^{\mathcal{N}}} \frac{r_{N,1}(\mu; z) + r_{N,0}(\mu; h)}{(\|z\|_{\mathbb{Z}}^2 + \|h\|_H^2)^{1/2}} \\ &= \|r_{N,1}(\mu)\|_{\mathbb{Z}'} + \sup_{h \in \mathcal{H}} \frac{r_{N,0}(\mu; h)}{\|h\|_H} \leq \|r_{N,1}(\mu)\|_{\mathbb{Z}'} + \|\mu_0 - (u_N(\mu))(0)\|_H \\ &=: R_{N,1}(\mu) + R_{N,0}(\mu_0). \end{aligned}$$

At a first glance it seems that $r_{N,1}$ (and $R_{N,1}$) only depends on μ_1 . However, the RB solution $u_N(\mu)$ involves both μ_0 and μ_1 so that both parameters enter. As already said earlier, the approximation of the initial value (and hence $R_{N,0}$), however, depends only on μ_0 .

The above form of the error estimate suggests to compute parameter samples (and snapshots) in a two-stage-method, namely first to determine samples μ_0^i for the initial value by maximizing $R_{N,0}(\mu_0)$ w.r.t. μ_0 and second to consider the evolution and compute samples μ^j by maximizing $R_{N,1}(\mu^j)$ using the before-computed snapshots $h^{\mathcal{M}}(\mu_0^i)$. This corresponds to the separated computations already introduced in §3.3.4. Let us now describe the two parts in detail.

We remark that even though we describe a greedy method, one could also use a different method to determine appropriate parameter samples e.g. by using non-linear optimization w.r.t. the error estimate, [3, 32]. The separation approach is independent of the particular maximization strategy.

Initial value greedy. The first step is to generate a reduced basis for the initial value, i.e., we need the solution at $t = 0$, which only depends on the parameter function $\mu_0 \in \mathcal{D}_0$, as we have seen in §3.3.4. For a given tolerance $\text{tol}_0 > 0$, we are looking for N_0 samples $S_{N_0}^0 := \{\mu_0^1, \dots, \mu_0^{N_0}\}$ and corresponding snapshots $\{h^1, \dots, h^{N_0}\}$, such that $q^i := \sigma^0 \otimes h^i \in \mathbb{Q}^{\mathcal{M}}$ and $u_0^i := q^i + 0 \in \mathbb{X}^{\mathcal{N}}$ (i.e., $u_0^i := q^i + w^i$, $w^i \in \mathbb{W}^{\mathcal{I}}$, $w^i \equiv 0$, $u_0^i(0) = h^i$) is the corresponding snapshot.

Given a specific value of the parameter $\mu_0 \in \mathcal{D}_0$, the corresponding (detailed) snapshot $h^{\mathcal{M}}(\mu_0) \in H^{\mathcal{M}}$ is determined by

$$(4.1) \quad (h^{\mathcal{M}}(\mu_0), \psi_m)_H = (\mu_0, \psi_m)_H, \quad 1 \leq m \leq \mathcal{M},$$

where $\{\psi_1, \dots, \psi_{\mathcal{M}}\}$ is the chosen basis for $H^{\mathcal{M}}$. Set $u_0^i(0) = h^i := h^{\mathcal{M}}(\mu_0^i) = \sum_{n=1}^{\mathcal{M}} \alpha_{0,n}(\mu_0) h^n$.

Given h^1, \dots, h^N (where we should have $N \leq N_0 \ll \mathcal{M}$) computed as snapshots corresponding to S_N^0 , a corresponding RB initial value approximation $h_N(\mu_0)$ of $\mu_0 \in \mathcal{D}_0$ is determined by solving the linear equation system corresponding to

$$(4.2) \quad (h_N(\mu_0), h^i)_H = (\mu_0, h^i)_H, \quad 1 \leq i \leq N,$$

provided that the inner products $(\mu_0, h^i)_H$ can be computed online efficient (i.e., with complexity independent of \mathcal{M}). Then, the error contribution reads

$$R_{N,0}(\mu_0) = \|\mu_0 - h_N(\mu_0)\|_H.$$

One option to determine the reduced basis h^1, \dots, h^{N_0} could be as follows: Choose $\eta^1, \dots, \eta^{\tilde{N}} \in \mathcal{D}_0$, $\tilde{N} > N_0$, arbitrary, compute the Gramian $\mathbf{M}_H^{\tilde{N}} := ((\eta^n, \eta^{n'})_H)_{n, n'=1, \dots, \tilde{N}}$ and choose $\mu_0^1, \dots, \mu_0^{N_0}$ as the (orthogonalized) eigenfunctions corresponding to the N_0 largest eigenvalues of $\mathbf{M}_H^{\tilde{N}}$. This corresponds to a proper orthogonal decomposition (POD). If the η^n are chosen well, this approach results in the best H -orthogonal choice. The obvious drawback is the strong dependence on the choice of the η^n .

For a greedy procedure, one chooses a training set $M_{\text{train}}^0 \subset \mathcal{D}_0$ and determines parameter samples by maximizing $R_{N,0}(\mu_0)$ over $\mu_0 \in M_{\text{train}}^0$. We obtain the greedy scheme in Algorithm 1. The most crucial part may be to find a good training set M_{train}^0 .

Algorithm 1 Initial value greedy

- 1: Let $M_{\text{train}}^0 \subset \mathcal{D}_0$ be the training set of initial values, $\text{tol}_0 > 0$ a given tolerance.
 - 2: Choose $\mu_0^1 \in M_{\text{train}}^0$, $S_1^0 := \{\mu_0^1\}$, compute $h^{\mathcal{M}}(\mu_0^1)$ as in (4.1), $\Xi_1^0 := \{h^{\mathcal{M}}(\mu_0^1)\}$
 - 3: **for** $j = 1, \dots, N_0^{\max}$ **do**
 - 4: $\mu_0^{j+1} = \arg \max_{\mu_0 \in M_{\text{train}}^0} R_{j,0}(\mu_0)$
 - 5: **if** $R_{j,0}(\mu_0^{j+1}) < \text{tol}_0$ **then** $N_0 := j$, $H_{N_0} := \text{span}(\Xi_{N_0}^0)$; **Stop end if**
 - 6: Compute $h^{\mathcal{M}}(\mu_0^{j+1}) \in H^{\mathcal{M}}$ as in (4.1).
 - 7: $S_{j+1}^0 := S_j^0 \cup \{\mu_0^{j+1}\}$, $\Xi_{j+1}^0 := \Xi_j^0 \cup \{h^{\mathcal{M}}(\mu_0^{j+1})\}$, orthogonalize Ξ_{j+1}^0 .
 - 8: **end for**
-

It remains to discuss the efficient computation of the error term $R_{j,0}(\mu_0)$ for a given parameter $\mu_0 \in \mathcal{D}_0$. Note, that we obtain a set of orthonormal functions $\Xi_{N_0}^0$ as an output of Algorithm 1. Hence, the RB approximation $h_{N_0}(\mu_0)$ coincides with the H -orthogonal projection of μ_0 to H_{N_0} . This means that $R_{N_0,0}(\mu_0)$ is the error of the best approximation of μ_0 in H_{N_0} . There are different possibilities to compute this error:

- (1) If μ_0 is given as formula, then an efficient quadrature may be used.
- (2) If μ_0 has a finite expression (like $\mu_0^{\mathcal{L}}$ above) in terms of a stable basis $\{B_1, \dots, B_{\mathcal{L}}\}$, one may either use an efficient quadrature or transform $h_N(\mu_0)$ into that basis and use the coefficients of the difference.
- (3) One could compute an orthonormal basis for the complement $H^{\mathcal{M}} \ominus H_{N_0}$ and approximate $R_{N,0}(\mu_0)$ by computing coefficients of μ_0 w.r.t. that complement basis (e.g. in terms of wavelets).

Evolution greedy. The next step is to find a basis for the part of the solution u in $\mathbb{W}^{\mathcal{I}}$, given the already determined reduced space H_{N_0} . Given a parameter $\mu = (\mu_0, \mu_1) \in \mathcal{D}$ and an approximation $h^{\mathcal{M}}(\mu_0)$, the evolution part $w^{\mathcal{I}}(\mu) \in \mathbb{W}^{\mathcal{I}}$ is computed as

$$(4.3) \quad b_1(\mu_1; w^{\mathcal{I}}(\mu), z^{\mathcal{I}}) = g_1(\mu_1; z^{\mathcal{I}}) - b_1(\mu_1; \sigma^0 \otimes h^{\mathcal{M}}(\mu_0), z^{\mathcal{I}}) \quad \forall z^{\mathcal{I}} \in \mathbb{Z}^{\mathcal{I}}.$$

For a reduced basis approximation corresponding to $\mu = (\mu_0, \mu_1) \in \mathcal{D}$, first compute

$$h_{N_0}(\mu_0) = \sum_{n=1}^{N_0} \alpha_{0,n}(\mu_0) h^n$$

as above. Then, given parameter samples $S_{N_1}^1 = \{\mu^1, \dots, \mu^{N_1}\} \in \mathcal{D} = \mathcal{D}_0 \times \mathcal{D}_1$ (to be determined e.g. by a second greedy described below) and corresponding snapshots $w^i := w^{\mathcal{I}}(\mu^i) \in \mathbb{W}^{\mathcal{I}}$, a reduced basis approximation $w_{N_1}(\mu) \in \mathbb{W}_{N_1} = \text{span}\{w^i : 1 \leq i \leq N_1\}$ is determined by

$$(4.4) \quad b_1(\mu_1; w_{N_1}(\mu), z_{N_1}) = \check{f}(\mu; z_{N_1}) := g_1(\mu_1; z_{N_1}) - b_1(\mu_1; \sigma^0 \otimes h_{N_0}(\mu_0), z_{N_1}),$$

for all $z_{N_1} \in \mathbb{Z}_{N_1}$, where \mathbb{Z}_{N_1} is a stable reduced space corresponding to \mathbb{W}_{N_1} w.r.t. the inner product b_1 in the sense that

$$(4.5) \quad \inf_{w_{N_1} \in \mathbb{W}_{N_1}} \sup_{z_{N_1} \in \mathbb{Z}_{N_1}} \frac{b_1(\mu_1; w_{N_1}, z_{N_1})}{\|w_{N_1}\|_{\mathbb{W}} \|z_{N_1}\|_{\mathbb{Z}}} \geq \beta_1(\mu_1) > 0$$

independent of $N_1 \rightarrow \infty$, see §4.2 below. Here β_1 is the inf-sup constant of the bilinear form b_1 .

It is readily seen that the right-hand side of (4.4) admits a separation w.r.t. the parameter. In fact, recalling (3.2) and (3.9) (where here, as opposed to (3.2) we set $b_{Q_b}(w, v) := \int_{\mathcal{I}} \langle \dot{w}(t), z(t) \rangle_{V' \times V} dt$), we have

$$(4.6) \quad \begin{aligned} \check{f}(\mu; z_{N_1}) &:= g_1(\mu_1; z_{N_1}) - b_1(\mu_1; \sigma^0 \otimes h_{N_0}(\mu_0), z_{N_1}) \\ &= \sum_{q=1}^{Q_g} \vartheta_q^g(\mu_1) g_q(z_{N_1}) + \sum_{q=1}^{Q_b} \vartheta_q^b(\mu_1) b_q(\sigma^0 \otimes h_{N_0}(\mu_0), z_{N_1}) \\ &= \sum_{q=1}^{Q_g} \vartheta_q^g(\mu_1) g_q(z_{N_1}) + \sum_{q=1}^{Q_b} \sum_{n=1}^{N_0} \vartheta_q^b(\mu_1) \alpha_{0,n}(\mu_0) b_q(\sigma^0 \otimes h^n, z_{N_1}) \\ &=: \sum_{q=1}^{Q_g+N_0} \vartheta_q^{\check{f}}(\mu) \check{f}(z_{N_1}) \end{aligned}$$

with obvious definitions of the involved terms. Hence, we obtain an efficient offline-online splitting both for the computation of the reduced basis approximation $w_{N_1}(\mu)$ and of the residual $r_{N,1}(\mu; z) = g_1(\mu_1; z) - b_1(\mu_1; u_N(\mu), z)$, where we set $u_N(\mu) := q_{N_0}(\mu_0) + w_{N_1}(\mu) = \sigma^0 \otimes h_{N_0}(\mu_0) + w_{N_1}(\mu)$, which means that

$$\begin{aligned} r_{N,1}(\mu; z) &= g_1(\mu_1; z) - b_1(\mu_1; u_N(\mu), z) \\ &= g_1(\mu_1; z) - b_1(\mu_1; \sigma^0 \otimes h_{N_0}(\mu_0), z) - b_1(\mu_1; w_{N_1}(\mu), z) \\ &= \check{f}(\mu; z) - b_1(\mu_1; w_{N_1}(\mu), z), \end{aligned}$$

which coincides with the residual of (4.4). Recalling that $\mathcal{D}_1 \subset \mathbb{R}^P$, (4.4) is a reduced problem with a $(P + N_0)$ -dimensional parameter space since H_{N_0} is the RB initial value space. Such a dimension might be a challenge. The error estimator is given by

$$\Delta_{N_1}^1(\mu) := \frac{\|r_{N,1}(\mu)\|_{\mathbb{Z}'}}{\beta_{\text{LB}}} = \frac{R_{N,1}(\mu)}{\beta_{\text{LB}}},$$

where β_{LB} is a lower bound of the inf-sup constant of the bilinear form b , and we obtain a – more or less – standard greedy scheme described in Algorithm 2.

Algorithm 2 Evolution greedy

-
- 1: Let $M_{\text{train}} \subset \mathcal{D}$ be the training set, $\text{tol}_1 > 0$ a given tolerance.
 - 2: Choose $\mu^1 \in M_{\text{train}}$, $\mu^1 := (\mu_0^1, \mu_1^1)$, $S_1^1 := \{\mu^1\}$
 - 3: Compute the RB approximation $h_{N_0}(\mu_0^1) \in S_{N_0}^0$ as in (4.2)
 - 4: Compute $w^{\mathcal{I}}(\mu^1) \in \mathbb{W}^{\mathcal{I}}$ as in (4.3), $\Xi_1^1 = \{w^{\mathcal{I}}(\mu^1)\}$
 - 5: **for** $j = 1, \dots, N_1^{\max}$ **do**
 - 6: $\mu^{j+1} = \arg \max_{\mu \in M_{\text{train}}} \Delta_j^1(\mu)$
 - 7: **if** $\Delta_j^1(\mu^{j+1}) < \text{tol}_1$ **then** $N_1 := j$, $\mathbb{W}_{N_1} := \text{span}(\Xi_{N_1}^1)$; **Stop end if**
 - 8: Compute the RB approximation $h_{N_0}(\mu_0^{j+1}) \in S_{N_0}^0$ as in (4.2)
 - 9: Compute $w^{\mathcal{I}}(\mu^{j+1}) \in \mathbb{W}^{\mathcal{I}}$ as in (4.3)
 - 10: $S_{j+1}^1 := S_j^1 \cup \{\mu^{j+1}\}$, $\Xi_{j+1}^1 := \Xi_j^1 \cup \{w^{\mathcal{I}}(\mu^{j+1})\}$
 - 11: **end for**
-

An obvious question arises how to choose the training set M_{train} in Algorithm 2, in particular the training samples for the initial value parameter. Possible choices for the subset of \mathcal{D}_0 include M_{train}^0 from Algorithm 1 or –much smaller– $S_{N_0}^0$, which might be a reasonable choice after performing already a greedy search for the initial value. We will come back to this point in our numerical experiments in Section 5.

4.2. Stable RB test spaces. It remains to construct a stable test space \mathbb{Z}_{N_1} in the sense of (4.5). It is well-known that it might be beneficial to construct this space by so-called *supremizers* in an efficient offline-online manner, [7, 12, 26].

Let $\{w^1, \dots, w^{N_1}\}$ be the basis of \mathbb{W}_{N_1} and fix $\mu_1 \in \mathcal{D}_1$. Then, the supremizer $s^n(\mu_1) \in \mathbb{Z}^{\mathcal{I}}$, $1 \leq n \leq N_1$, is defined by the relation

$$s^n(\mu_1) := \arg \sup_{z^{\mathcal{I}} \in \mathbb{Z}^{\mathcal{I}}} \frac{b_1(\mu_1; w^n, z^{\mathcal{I}})}{\|z^{\mathcal{I}}\|_{\mathbb{Z}}}.$$

In order to compute this quantity, recall that $\{\zeta_i := \tau^k \otimes \phi_j : 1 \leq k \leq \mathcal{K}, 1 \leq j \leq \mathcal{J}, i = (k, j)\}$ is the basis of $\mathbb{Z}^{\mathcal{I}}$, $\mathcal{I} = \mathcal{K} \cdot \mathcal{J}$, and note that

$$\begin{aligned} \|z^{\mathcal{I}}\|_{\mathbb{Z}}^2 &= \|z^{\mathcal{I}}\|_{L_2(I; V)}^2 = \sum_{k, k'=1}^{\mathcal{K}} \sum_{j, j'=1}^{\mathcal{J}} z_j^k z_{j'}^{k'} (\tau^k, \tau^{k'})_{L_2(I)} (\phi_j, \phi_{j'})_V \\ &= z_{\mathcal{I}}^T (\mathbf{I}_{\text{time}}^{\mathcal{K}} \otimes \mathbf{G}_{\text{space}}^{\mathcal{J}}) z_{\mathcal{I}} =: z_{\mathcal{I}}^T \mathbf{Z}^{\mathcal{I}} z_{\mathcal{I}} \end{aligned}$$

with the Gramian matrices $\mathbf{G}_{\text{space}}^{\mathcal{J}} = ((\phi_i, \phi_j)_V)_{i, j=1, \dots, \mathcal{J}}$ for $V^{\mathcal{J}}$ (w.r.t. the V -inner product) and $\mathbf{I}_{\text{time}}^{\mathcal{K}} = ((\tau^k, \tau^{k'})_{L_2(I)})_{k, k'=1, \dots, \mathcal{K}} = (\Delta t) \mathbf{I} \mathbf{d} \in \mathbb{R}^{\mathcal{K} \times \mathcal{K}}$. Next, let the expansion of w^n in terms of the full basis $\{\varpi_i := \sigma^k \otimes \phi_j : i = (k, j), 1 \leq k \leq \mathcal{K}, 1 \leq j \leq \mathcal{J}\}$ of $\mathbb{W}^{\mathcal{I}}$ be denoted by

$$w^n = \sum_{i=1}^{\mathcal{I}} \omega_i^n \varpi_i, \quad \boldsymbol{\omega}_{\mathcal{I}}^n := (\omega_i^n)_{i=1, \dots, \mathcal{I}}.$$

Then, setting $z_i := z_j^k$, $i = (k, j)$, we get

$$b_1(\mu_1; w^n, z^{\mathcal{I}}) = \sum_{i, i'=1}^{\mathcal{I}} \omega_i^n z_{i'} b_1(\mu_1; \varpi_i, \zeta_{i'}) = (\boldsymbol{\omega}_{\mathcal{I}}^n)^T \mathbf{B}^{\mathcal{I}}(\mu_1) \mathbf{z}_{\mathcal{I}} = (\mathbf{z}_{\mathcal{I}})^T (\mathbf{B}^{\mathcal{I}}(\mu_1))^T \boldsymbol{\omega}_{\mathcal{I}}^n.$$

The vector $\mathbf{s}^n(\mu_1)$ containing the expansion coefficients of $s^n(\mu_1)$ is given by

$$\mathbf{s}^n(\mu_1) = (\mathbf{Z}^{\mathcal{I}})^{-1}(\mathbf{B}^{\mathcal{I}}(\mu_1))^T \mathbf{w}_{\mathcal{I}}^n,$$

In view of the separation of $b_1(\mu_1; \cdot, \cdot)$ w.r.t. the parameter μ_1 (see (3.9) with b_{Q_b} replaced by $b_{Q_b}(w, v) := \int_I \langle \dot{w}(t), z(t) \rangle_{V' \times V} dt$), we have $\mathbf{B}^{\mathcal{I}}(\mu_1) = \sum_{q=1}^{Q_b} \vartheta_q^b(\mu_1) \mathbf{B}_q^{\mathcal{I}}$ with parameter-independent matrices $\mathbf{B}_q^{\mathcal{I}}$. Hence, we obtain the representation $\mathbf{s}^n(\mu_1) = \sum_{q=1}^{Q_b} \vartheta_q^b(\mu_1) (\mathbf{Z}^{\mathcal{I}})^{-1} (\mathbf{B}_q^{\mathcal{I}})^T \mathbf{w}_{\mathcal{I}}^n$ and the terms $\mathbf{z}_q^n := (\mathbf{Z}^{\mathcal{I}})^{-1} (\mathbf{B}_q^{\mathcal{I}})^T \mathbf{w}_{\mathcal{I}}^n$ can be computed offline (as they are parameter-independent). Since the μ_1 -dependent supremizers can be build by linear combinations (with μ_1 -dependent coefficients) of the functions $z_q^n \in \mathbb{Z}^{\mathcal{I}}$ corresponding to the coefficient vectors \mathbf{z}_q^n , $1 \leq n \leq N_1$, $1 \leq q \leq Q_b$ we choose for every $\mu_1 \in \mathcal{D}_1$

$$\mathbb{Z}_{N_1}(\mu_1) := \text{span}\{s^1(\mu_1), \dots, s^{N_1}(\mu_1)\}$$

as reduced test space, where $s^n(\mu_1) = \sum_{q=1}^{Q_b} \vartheta_q^b(\mu_1) z_q^n$.

5. NUMERICAL RESULTS

5.1. Heston Model. We consider the Heston model as described in §2.1 above. The initial value μ_0 corresponds to the payoff function u_0 , see (2.4c). For pricing problems, one aims at rapidly changing the payoff, which is the motivation to use the parameter function.

In order to ensure well-posedness of the PDE, we require the natural assumptions $\nu_t > 0$ for the volatility and $\rho \in (0, 1)$ for the correlation. Since we do not transform the initial conditions to homogeneous ones but work in Bochner spaces using the space-time variational approach, we do not need additional conditions for the parameter spaces as e.g. in [21].

Just for the ease of implementation, we choose homogeneous Dirichlet conditions. Unfortunately one has to work with a large domain Ω and with a fine discretization to get good results for the Crank Nicolson solution in comparison to the closed form solution of the Heston model (cf. [20]) that one can use for validation. One could further improve the results by using e.g. the boundary conditions proposed in [33].

5.2. Initial condition parameter function. As already said earlier, we want to use the initial condition as a parameter function. Since payoff functions are not completely arbitrary, but have certain shapes, we introduce a model using Bernstein polynomials that allows for a small parametric representation of the payoff functions, which are usually continuous, piecewise smooth, convex and are composed of linear functions.

5.2.1. Bernstein Polynomials. Bernstein polynomials are H^1 -functions, preserve convexity and can be adapted locally, see, e.g. [27]. We briefly recall the main properties that will be needed here. Consider an interval $\Delta = [v_0, v_1] \subset \mathbb{R}$, where $v_0 < v_1$. Any $x \in \Delta$ has the unique representation

$$x = u_0(x|\Delta) v_0 + u_1(x|\Delta) v_1,$$

where $u_j(\cdot|\Delta) \in \mathbb{P}_1$ (a linear polynomial) is nonnegative on Δ and $u_0(x|\Delta) + u_1(x|\Delta) = 1$ (convex combination, partition of unity). Denoting by

$$\Gamma_n := \{\alpha = (\alpha_0, \alpha_1) \in \mathbb{Z}^2 : \alpha_0, \alpha_1 \geq 0, \alpha_0 + \alpha_1 = n\}$$

the set of all homogeneous multi-indices of length n , the α -th *Bernstein-Bézier basis polynomial* of degree n is defined for $\alpha \in \Gamma_n$ as

$$B_\alpha(x|\Delta) := \binom{n}{\alpha} u_0^{\alpha_0}(x|\Delta) u_1^{\alpha_1}(x|\Delta).$$

The *Bézier surface* for a given function $f : \Delta \rightarrow \mathbb{R}$ is defined by

$$B_n f(x|\Delta) = \sum_{\alpha \in \Gamma_n} f(x_\alpha) B_\alpha(x|\Delta), \quad \text{where } x_\alpha \in \Delta \text{ s.t. } u(x_\alpha|\Delta) = \frac{\alpha}{n}.$$

5.2.2. Payoff approximation for the Heston model. For the Heston model, we consider a domain $\Omega = \Omega_1 \times \Omega_2 \subset \mathbb{R}^2$, where Ω_1 models the logarithmic asset price $\log(S)$ and Ω_2 the volatility ν . The payoff depends only on S . Divide the interval $\exp(\Omega_1)$ into subintervals $I_1, \dots, I_{\mathcal{L}-1}$, $I_\ell = [v_\ell^-, v_\ell^+]$, $\ell = 1, \dots, \mathcal{L} - 1$, $v_{\mathcal{L}-1}^+ = v_\ell^-$ for $\ell = 2, \dots, \mathcal{L} - 1$, and choose Δ as each of these I_ℓ .

We now model the initial value μ_0 on each I_ℓ as a Bézier surface of degree 1, i.e.,

$$\begin{aligned} B_1 \mu_0(x|I_\ell) &= \mu_0(x_{(1,0)}) B_{(1,0)}(x|I_\ell) + \mu_0(x_{(0,1)}) B_{(0,1)}(x|I_\ell) \\ &= \mu_0(v_\ell^-) u_0^1(x|I_\ell) u_1^0(x|I_\ell) + \mu_0(v_\ell^+) u_0^0(x|I_\ell) u_1^1(x|I_\ell) \\ &= \mu_0(v_\ell^-) u_0(x|I_\ell) + \mu_0(v_\ell^+) u_1(x|I_\ell). \end{aligned}$$

The approximation of μ_0 is then defined by

$$\begin{aligned} \mu_0^\mathcal{L}(x) &:= \sum_{\ell=1}^{\mathcal{L}-1} \{ \mu_0(v_\ell^-) u_0(x|I_\ell) + \mu_0(v_\ell^+) u_1(x|I_\ell) \} \\ &= \mu_0(v_1) u_0(x|I_1) + \sum_{\ell=2}^{\mathcal{L}-1} \mu_0(v_\ell) (u_1(x|I_{\ell-1}) + u_0(x|I_\ell)) + \mu_0(v_\mathcal{L}) u_1(x|I_{\mathcal{L}-1}), \end{aligned}$$

where we have used the fact that $\mu_0(v_{\ell-1}^+) = \mu_0(v_\ell^-)$ and renamed v_ℓ^- by v_ℓ as well as $v_\mathcal{L} := v_{\mathcal{L}-1}^+$. Since $u_0(v_\ell|I_\ell) = u_1(v_{\ell+1}|I_\ell) = 1$ and $u_0(v_{\ell+1}|I_\ell) = u_1(v_\ell|I_\ell) = 0$ the following equality is obvious:

$$u_1(x|I_{\ell-1}) + u_0(x|I_\ell) = \frac{x - v_{\ell-1}}{v_\ell - v_{\ell-1}} \chi_{I_{\ell-1}}(x) + \frac{x - v_{\ell+1}}{v_\ell - v_{\ell+1}} \chi_{I_\ell}(x) =: B_\ell(x),$$

so that $\mu_0^\mathcal{L}(x) = \sum_{\ell=1}^{\mathcal{L}} \mu_0(v_\ell) B_\ell(x)$.

5.3. Numerical results. We now present our numerical results.

5.3.1. Data. We used the following data for our simulation:

- We use the correlation as calibration parameter, i.e., $\mu_1 := \rho \in (0, 1) =: \mathcal{D}_1$;
- $\Omega_1 := \log([10^{-8}, 200])$ for the asset price, $\Omega_2 := (0, 1]$ for the volatility, detailed dimension $\mathcal{J} = 14, 271$;
- $\kappa = 0.8, \sigma = 0.6, \theta = 0.2$ and $r \equiv 0.001$;
- $T = 0.25$ (3 months), $\mathcal{K} = 25$;
- $H^\mathcal{M} := V^\mathcal{J}$ and $\mathcal{N} = \mathcal{K} \cdot \mathcal{J} + \mathcal{J} = \mathcal{J}(\mathcal{K} + 1) = 371, 046$;
- Knots $\{v_1, \dots, v_8\} = \ln(\{0, 70, 80, 90, 100, 110, 200\})$ for the Bézier polynomials, which implies that payoff functions with specific strike price $K \in \{70, 80, 90, 100, 110\}$ are represented exactly.

All experiments have been performed on an Intel Core i7-3770 (3.40 GHz), with 16 GB RAM and using Matlab (8.0.0.783 (R2012b)). All RB calculations were implemented in RBmatlab, see <http://www.morepas.org>.

5.3.2. *Initial value reduced approximation.* For determining the RB approximation of the initial value, we used a POD method based upon the linear Bernstein polynomials (in the notation of §4.1 $\eta^\ell := B_\ell$, $\ell = 1, \dots, 7 =: \tilde{N}$). We show the decay of the eigenvalues of the Gramian $\mathbf{M}_H^{\tilde{N}}$ in Figure 5.1. Choosing five basis functions h^1, \dots, h^5 (i.e., a RB space H_5) results in a relative error of 0.0314. We will investigate later how the choice of only 5 POD basis functions influences the reduced solution of the whole problem. The first two orthogonal POD functions are shown in Figure 5.2.

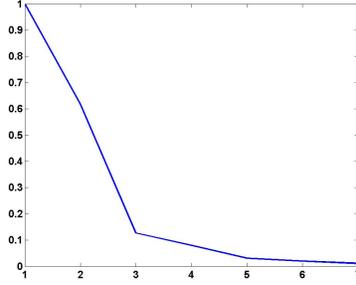


FIGURE 5.1. Eigenvalue decay for the Gramian matrix

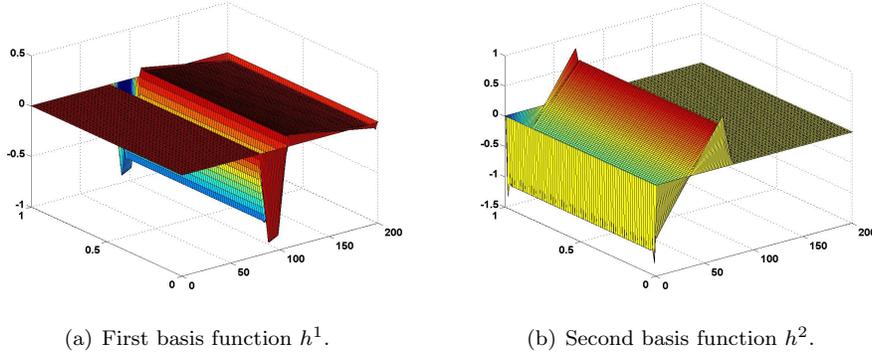


FIGURE 5.2. First two POD eigenfunctions projected to $V^{\mathcal{J}}$.

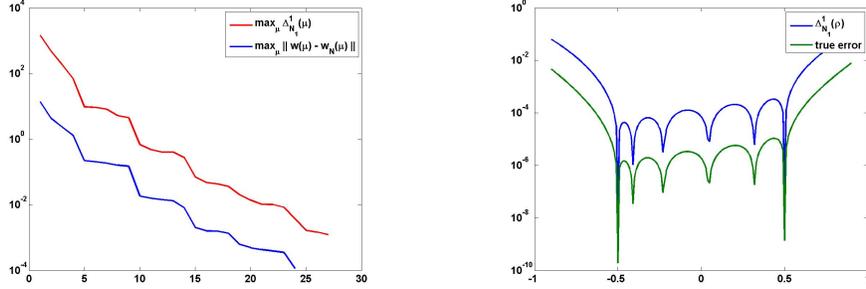
For functions that can be modeled by a small number of basis functions as it is the case for call and put options using Bézier polynomials, it might be advisable to skip this first step and directly enter the evolution greedy using the small basis as a part of the training set.

5.3.3. *Evolution greedy.* In the second step we perform the evolution greedy in Algorithm 2 to compute the RB space \mathbb{W}_{N_1} , where we first use $M_{\text{train}} := \{h^1, \dots, h^5\} \times \{x_k = -0.5 + k \Delta s : k = 0, \dots, 11, \Delta s = \frac{1}{11}\} \subset \text{span}\{B_1, \dots, B_7\} \times (-1, 1)$ as a training set, i.e., $\#M_{\text{train}} = 60$. As in [31] we use a natural discrete space-time norm given by

$$\|w\|_{\mathbb{X}}^2 := \|\bar{w}\|_{L_2(I;V)}^2 + \|\dot{w}\|_{L_2(I;V')}^2 + \|w(T)\|_H^2, \quad w \in \mathbb{X}^{\mathcal{N}},$$

where $\bar{w}^k := (\Delta t)^{-1} \int_{I^k} w(t) dt \in V$ and $\bar{w} := \sum_{k=1}^{\mathcal{K}} \chi_{I^k} \otimes \bar{w}^k \in L_2(I; V)$. Using this norm, the discrete problem appearing in the evolution greedy is well-posed. Another reason for this choice is that the evolution greedy uses the same bilinear form as was used in [31] to treat homogeneous initial conditions.

We compare the evolution greedy with the proposed error estimate $R_{N_1,1}$ with the so called strong (evolution) greedy using the exact error (instead of the estimate) which is determined using a detailed solution. Doing so, we can investigate the performance of the error estimator. As we see in Figure 5.3 (a), the error estimate behaves similar to the true error. An error tolerance of 10^{-3} is reached by 28 basis functions using the estimate, whereas the strong greedy shows that 24 basis functions suffice to reach a tolerance of 10^{-4} . As expected, the error bound overestimates the error. One reason is that (for simplicity) we use a pessimistic lower bound of the inf-sup condition ($\beta_{\text{LB}} = 0.005$).



(a) Maximum error over iterations of the strong evolution greedy and using the error estimate $R_{N_1}^1$

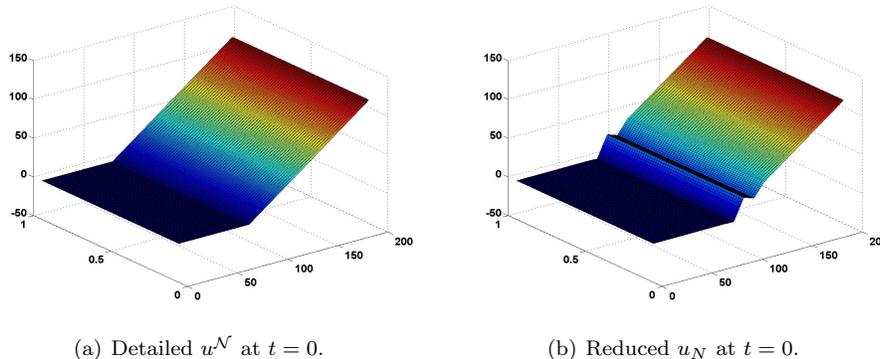
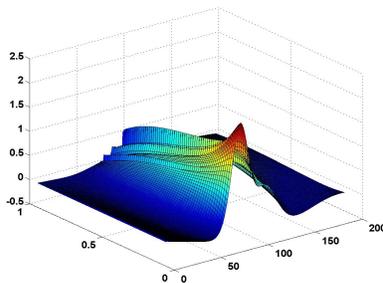
(b) Error estimator and true error for $\rho \in (-1, 1)$

FIGURE 5.3. Performance of error estimators

5.3.4. Error propagation. Now, we investigate how different training sets influence the results. First, we choose $u_0 = h^1$ (i.e., an exact initial condition). As we see in Figure 5.3 (b) the error for $\rho \in [-0.5, 0.5]$ (the part that is covered by the training set) is below 10^{-4} and the error estimation $R_{28,1}$ is below 10^{-3} . For $\rho \in (-1, 1) \setminus [-0.5, 0.5]$ the approximation is –as expected– worse. Note that we plotted on a grid for ρ with step size 0.01. Furthermore, we can see that $(h^1, -0.5)$, $(h^1, -0.4091)$, $(h^1, -0.2273)$, $(h^1, 0.0455)$, $(h^1, 0.3182)$ and $(h^1, 0.5)$ are in the sample set for \mathbb{W}_{28} .

Next, we choose a call payoff function $\mu_0(y, \nu) = \max(\exp(y) - K, 0)$ with strike price $K = 70$ and $\rho = 0.3$. The resulting approximation error using only five POD basis functions can clearly be seen in Figure 5.4. Obviously, the space-time error is large on the domain Ω , in particular due to the errors near the boundaries. However, usually one has to enlarge the domain in advance due to the chosen boundary conditions, hence one is only interested in a smaller part of Ω .

For option pricing the *absolute error*, i.e., the difference of the prices, is also of interest, because the resulting RB price is to be actually paid by the customer. As we can see in Figure 5.5, the absolute error at the final time T is actually about 2, which is clearly too large. The reason is the POD truncation of the initial value based upon a *relative error* basis.

FIGURE 5.4. Detailed and reduced initial condition ($u|_{[10^{-8}, 190] \times [0.05, 0.95]}$).FIGURE 5.5. Absolute error $u^{\mathcal{N}}(T) - u_N(T)$ at maturity ($u|_{[10^{-8}, 190] \times [0.05, 0.95]}$).

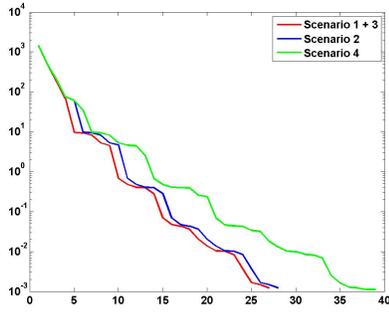
As mentioned before, we can choose the training set M_{train} in the evolution greedy independent of $S_{N_0}^0$. We compare different combinations of reduced basis spaces in Table 5.1. For $\rho \in [-0.5, 0.5]$ with $u_0 = h^1$ (so again an exact initial condition) the true errors of the different approximations are shown in Figure 6(b). We can see, that for the exact initial value all RB approximations are evenly good. In contrast, by looking at the last column of Table 5.1, we note that if we choose a poor initial value approximation resp. a poor training set for \mathcal{D}_0 , the RB approximation is not acceptable. In particular for PDEs with zero right-hand side and a linear operator (as we have for option pricing), we can show that for every new parameter $\bar{\mu}$, where the function part $\bar{\mu}_0$ lies in the span of the training set M_{train}^1 , we get $\Delta_{N_1}^1(\bar{\mu}) < \text{tol}_1$. That explains why we do not get a larger N_1 resp. a better RB approximation by extending M_{train}^1 in Scenario 2 in comparison to Scenario 1. Of course, in this situation, this is also explained by Gronwall's lemma.

6. SUMMARY AND OUTLOOK

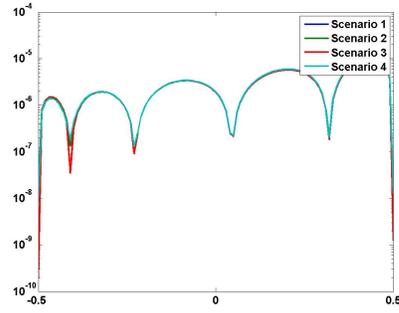
We have introduced a Reduced Basis space-time variational approach for parametric parabolic partial differential equations having coefficient parameters and a

TABLE 5.1. Comparison of different sets for H_{N_0} and the training set of the evolution greedy ($M_{\text{train}} = M_{\text{train}}^0 \times M_{\text{train}}^1$, M_{train}^1 as before, $\text{tol}_1 = 10^{-3}$, $\mu(y, \nu) = (\max(e^y - K, 0), 0.3)$ and with $u|_{[10^{-8}, 190] \times [0.05, 0.95]}$).

Scenario	H_{N_0}	M_{train}^0	$\#M_{\text{train}}$	N_1	$\ u^{\mathcal{N}}(\mu) - u_N(\mu)\ _{\bar{\mathbb{X}}}$
1	$\text{span}\{h^1, \dots, h^5\}$	$\{\mu_0^1, \dots, \mu_0^5\}$	60	28	9.1994
2	$\text{span}\{h^1, \dots, h^5\}$	$\{\mu_0^1, \dots, \mu_0^7\}$	84	29	9.1994
3	$\text{span}\{h^1, \dots, h^7\}$	$\{\mu_0^1, \dots, \mu_0^5\}$	60	28	7.9801
4	$\text{span}\{h^1, \dots, h^7\}$	$\{\mu_0^1, \dots, \mu_0^7\}$	84	40	$1.7162e - 04$



(a) Maximum errors over iterations using the error estimate $R_{N_1}^1$.



(b) True errors for $\rho \in [-0.5, 0.5]$.

FIGURE 5.6. Comparison of the different scenarios presented in Table 5.1.

variable initial condition. Feasibility and efficiency have been demonstrated. Obviously, a whole variety of further questions arises, just to mention some of them that we aim to consider in the near future:

- Extension of the Bernstein representation of the initial condition to an adaptive wavelet approximation.
- In [4], an alternative space-time variational formulation has been considered which transfers the essential initial condition to a natural one.
- In case of a basket of options, the dimension of the problem of course grows, which calls for a specific treatment e.g. by the Hierarchical Tucker format as e.g. in [9, 23]. More general, the choice of trial and test spaces within the (offline) phase can be further investigated and possibly optimized.
- For calibration purposes, all model parameters have to be taken into account. This type of high dimensionality needs particular treatment as e.g. in [8], [18] or [19]. Moreover, numerical stabilization techniques have to be investigated.
- Extensions to other financial models, American options, etc.

REFERENCES

- [1] R. Andreev. *Stability of space-time Petrov-Galerkin discretizations for parabolic evolution equations*. PhD thesis, ETH Zürich, no. 20842 (2012).

- [2] R. Andreev. Stability of sparse space-time finite element discretizations of linear parabolic evolution equations. *IMA J. Numer. Anal.* **33**, no. 1, 242–260 (2013).
- [3] T. Bui-Thanh, K. Willcox, and O. Ghattas. Model reduction for large-scale systems with high-dimensional parametric input space. *SIAM J. Sci. Comput.*, **30** no. 6, 3270–3288 (2008).
- [4] N.G. Chegini and R. Stevenson. Adaptive wavelet schemes for parabolic problems: sparse matrices and numerical results. *SIAM J. Numer. Anal.* **49**, no. 1, 182–212 (2011).
- [5] R. Cont, Rama, N. Lantos, and O. Pironneau. A reduced basis for option pricing. *SIAM J. Financial Math.* **2**, no. 1, 287–316 (2011).
- [6] J.C. Cox, J.E. Ingersoll, and S.A. Ross. A theory of the term structure of interest rates. *Econometrica* **53**, 385–407 (1985).
- [7] W. Dahmen, C. Plesken, and G. Welper. Double greedy algorithms: Reduced basis methods for transport dominated problems. arXiv preprint arXiv:1302.5072, 2013.
- [8] D. Devaud, A. Manzoni and G. Rozza. A combination between the reduced basis method and the ANOVA expansion: On the computation of sensitivity indices. *C.R. Acad. Sci. Paris Series I*, **351**, no. 51, 593–598 (2013).
- [9] S. Dolgov, and B.N. Khoromskij. Simultaneous state-time approximation of the chemical master equation using tensor product formats. arXiv:1311.3143, 2013, <http://arxiv.org/abs/1311.3143>, NLAA, 2014, to appear.
- [10] L.C. Evans. *Partial Differential Equations*. Oxford University Press, Graduate Studies in Mathematics, vol. 19, 1998.
- [11] J. Da Fonseca, M. Grasselli and C. Tebaldi. A multifactor volatility Heston model. *Quant. Finance* **8**, no. 6, 591–604 (2008).
- [12] A.-L. Gerner and K. Veroy. Certified reduced basis methods for parametrized saddle point problems. *SIAM J. Sci. Comp.*, **34** no. 5, A2812–A2836 (2012).
- [13] M. Grepl and A.T. Patera. A posteriori error bounds for reduced-bias approximations of parametrized parabolic partial differential equations. *M2AN Math. Model. Numer. Anal.* **39**, no. 1, 157–181 (2005).
- [14] B. Haasdonk. Reduced Basis Methods for Parametrized PDEs — A Tutorial Introduction for Stationary and Instationary Problems. In: *Reduced Order Modelling*. Luminy Book series, 2014, to appear
- [15] B. Haasdonk and M. Ohlberger. Reduced basis method for finite volume approximations of parametrized linear evolution equations. *ESAIM: Mathematical Modeling and Numerical Analysis*, **42**, no. 2, 277–302 (2008).
- [16] B. Haasdonk, J. Salomon and B. Wohlmuth. A Reduced Basis Method for the Simulation of American Options. In: *ENUMATH 2011 Proceedings*, 2012.
- [17] B. Haasdonk, K. Urban, and B. Wieland. Reduced Basis Methods for Parametrized Partial Differential Equations with Stochastic Influences using the Karhunen Loeve Expansion. *SIAM/ASA Journal on Uncertainty Quantification* **1**, 79–105 (2013).
- [18] J. Hesthaven, B. Stamm and S. Zhang. Efficient greedy algorithms for high-dimensional parameter spaces with applications to empirical interpolation and reduced basis methods. *ESAIM: Mathematical Modeling and Numerical Analysis*, **48**, no. 1, 259–283 (2014).
- [19] J. Hesthaven and S. Zhang. On the use of ANOVA expansions in reduced basis methods for high-dimensional parametric partial differential equations. Scientific Computing Group, Brown University, 2011-31 (2011), http://www.dam.brown.edu/scicomp/media/report_files/BrownSC-2011-31.pdf.
- [20] S.L. Heston. A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options. *The Review of Financial Studies* **6**, no. 2, 327–343 (1993).
- [21] N. Hilber. *Stabilized wavelet methods for option pricing in high dimensional stochastic volatility models*. PhD thesis, ETH Zürich, 2009.
- [22] V.H. Hoang and C. Schwab. Sparse tensor Galerkin discretization of parametric and random parabolic PDEs—analytic regularity and generalized polynomial chaos approximation. *SIAM J. Math. Anal.* **45**, no. 5, 3050–3083 (2013).
- [23] R. Kiesel, A. Rupp, and K. Urban. Valuation of Structured Financial Products by Adaptive Multiwavelet Methods in High Dimensions. In: S. Dahlke et al. (eds.), *Extraction of Quantifiable Information from Complex Systems*, Lecture Notes in Comp. Sci. Eng. 102, Springer Switzerland, Ch. 16, 279–303 (2014).
- [24] B. Øksendal. *Stochastic Differential Equations: An Introduction with Applications*. Springer-Verlag, Berlin Heidelberg, 6th ed. 2014.

- [25] G. Rozza and A.T. Patera. *Reduced Basis Approximation and A Posteriori Error Estimation for Parametrized Partial Differential Equations*. Massachusetts Institute of Technology (MIT), 2006. <http://augustine.mit.edu>.
- [26] G. Rozza and K. Veroy. On the stability of the reduced basis method for stokes equations in parametrized domains. *Comp. Meth. Appl. Mech. Eng.*, **196** no. 7, 1244–1260 (2007).
- [27] T. Sauer. Ein algorithmischer Zugang zu Polynomen und Splines. (German). *Math. Semesterber.* **43**, no. 2, 169–189 (1996).
- [28] E. Sachs and M. Schu. A priori error estimates for reduced order models in finance. *ESAIM Math. Model. Numer. Anal.* **47**, no. 2 (2013), 449–469.
- [29] C. Schwab and R. Stevenson. Space-time adaptive wavelet methods for parabolic evolution problems. *Math. Comp.* **78**, 1293–1318 (2009).
- [30] K. Urban and A.T. Patera. A New Error Bound for Reduced Basis Approximation of Parabolic Partial Differential Equations. *C.R. Acad. Sci. Paris Series I*, 350(3-4), 203–207 (2012).
- [31] K. Urban and A.T. Patera. An Improved Error Bound for Reduced Basis Approximation of Linear Parabolic Problems. *Math. Comput.*, online first (2013), doi: <http://dx.doi.org/10.1090/S0025-5718-2013-02782-2>.
- [32] K. Urban, S. Volkwein and O. Zeeb. Greedy sampling using nonlinear optimization. In: *Reduced Order Methods for modeling and computational reduction*, A. Quarteroni and G. Rozza (eds.), Springer 2014, 137–157.
- [33] G. Winkler, T. Apel, U. Wystup. Valuation of options in Heston’s stochastic volatility model using finite element methods. *Foreign Exchange Risk*, 283–303, (2001).

ANTONIA MAYERHOFER, UNIVERSITY OF ULM, GRADUIERTENKOLLEG 1100, HELMHOLTZSTRASSE 22, D-89069 ULM, GERMANY

E-mail address: antonia.mayerhofer@uni-ulm.de

KARSTEN URBAN, UNIVERSITY OF ULM, INSTITUTE FOR NUMERICAL MATHEMATICS, HELMHOLTZSTRASSE 20, D-89069 ULM, GERMANY

E-mail address: karsten.urban@uni-ulm.de