Weierstraß-Institut
für Angewandte Analysis und Stochastik
Leibniz-Institut im Forschungsverbund Berlin e. V.

Preprint ISSN 2198-5855

Low-dimensional approximations of high-dimensional asset price models

Martin Redmann[1], Christian Bayer[2], Pawan Goyal[3]

submitted: March 27, 2020

[1] Martin Luther University Halle-Wittenberg
Institute of Mathematics
Theodor-Lieser-Str. 5
06120 Halle (Saale)
Germany
E-Mail: martin.redmann@mathematik.uni-halle.de

[2] Weierstrass Institute
Mohrenstr. 39
10117 Berlin
Germany
E-Mail: christian.bayer@wias-berlin.de

[3] Max Planck Institute for Dynamics of Complex Technical Systems
Sandtorstr. 1
39106 Magdeburg
Germany
E-Mail: goyalp@mpi-magdeburg.mpg.de

No. 2706
Berlin 2020

2010 *Mathematics Subject Classification.* 91G20, 91G60, 93A15, 60H10, 65D32.

Key words and phrases. Model order reduction, Black Scholes model, Heston model, option pricing.

Edited by
Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS)
Leibniz-Institut im Forschungsverbund Berlin e. V.
Mohrenstraße 39
10117 Berlin
Germany

Fax: +49 30 20372-303
E-Mail: preprint@wias-berlin.de
World Wide Web: http://www.wias-berlin.de/

Low-dimensional approximations of high-dimensional asset price models

Martin Redmann, Christian Bayer, Pawan Goyal

ABSTRACT. We consider high-dimensional asset price models that are reduced in their dimension in order to reduce the complexity of the problem or the effect of the curse of dimensionality in the context of option pricing. We apply model order reduction (MOR) to obtain a reduced system. MOR has been previously studied for asymptotically stable controlled stochastic systems with zero initial conditions. However, stochastic differential equations modeling price processes are uncontrolled, have non-zero initial states and are often unstable. Therefore, we extend MOR schemes and combine ideas of techniques known for deterministic systems. This leads to a method providing a good pathwise approximation. After explaining the reduction procedure, the error of the approximation is analyzed and the performance of the algorithm is shown conducting several numerical experiments. Within the numerics section, the benefit of the algorithm in the context of option pricing is pointed out.

1. INTRODUCTION

In finance we often encounter high-dimensional models, since the underlying markets are usually high-dimensional. For instance, in equity, take all stocks comprising the S & P 500 index (SPX). Fixed income markets exhibit a myriad of different relevant interest rates. All these are, obviously, only small snapshots of even larger markets. Of course, in many situations, we are only interested in a tiny fraction of these markets, which can be adequately modeled by a low-dimensional stochastic process. Moreover, if we are interested in derivatives on SPX, for example, then we may just model the index itself, disregarding the fine structure. On the other hand, if we consider a larger portfolio, this may not be possible without introducing inconsistencies in the model.

From a numerical perspective, high-dimensional models pose severe difficulties. Indeed, many traditional computational tools suffer from the *curse of dimensionality*, which essentially states that the computational work required to compute the relevant quantity of interest up to a prescribed error tolerance grows exponentially in the dimension n of the model. Most methods for discretizing partial differential equations (such as finite element and finite difference methods) suffer from the curse of dimensionality, as do Fourier based methods. Even many deterministic sampling methods (i.e., tree methods, quasi Monte Carlo) suffer from the curse of dimensionality in one way or another.[1] The notable exception is, of course, Monte Carlo simulation.

One way to overcome the numerical burden in high-dimensional models is *Model order reduction* (MOR) [1, 2, 9]. MOR is a technique in numerical analysis in order to construct low

[1] It is worth pointing out, that the effective dimension for QMC and MC method is, in fact, n multiplied by the number of time-steps, if time discretization is needed.

DOI 10.20347/WIAS.PREPRINT.2706 Berlin 2020

dimensional surrogate models that allow to approximate the quantity of interest with the desired accuracy. As MOR takes the specific quantity of interest into account, reduced models for different options will generally be different. (The specific method introduced later will, however, not depend, e.g., on the specific strike price.)

To fix ideas, suppose that we are given an n dimensional stochastic volatility model with asset price processes $S(t) \in \mathbb{R}^n$ and the corresponding stochastic variance processes v – which will be one-dimensional in our numerical examples. Consider an option with payoff $g(CS)$, where $C \in \mathbb{R}^{p\times n}$ and $g : \mathbb{R}^p \to \mathbb{R}$ possibly non-linear, where we assume that $p \ll n$. For instance, we have $p = 1$ for basket options. Our goal is to construct a Markov process $\tilde{x}$ taking values in $\mathbb{R}^{\tilde{n}}$ – with $\tilde{n} \ll n$ – and a matrix $C_1 \in \mathbb{R}^{p\times\tilde{n}}$ such that the processes CS and $C_1\tilde{x}$ are close in L^2. This usually ensures a good approximation of the payoff, i.e., $g(CS) \approx g(C_1\tilde{x})$. In this paper, we present a general strategy for identifying such processes $\tilde{x}$. We also provide numerical evidence of successful MOR in several financial applications, in the sense that relative errors of the order of 10^{-4} are regularly achieved with very small $\tilde{n}$ even when $n \geq 100$. It should be noted here that we only propose a MOR technique for the asset process S in this paper, but not for the variance process v. This is due to the generally non-linear dynamics of the variance process, which would require more complicated MOR strategies and will be explored in future work.

Before explaining the MOR strategy in detail, some conclusions can already be made based on the fundamental idea. First note that MOR should not be confused with *Markovian projection*, see [11, 17]. The underlying problem is, of course, that the process CS itself is a natural candidate for a reduced model, but it generally lacks the Markov property. There is, however, a Markov process $\hat{x}$ taking values in $\mathbb{R}^p$ such that $CS(t)$ and $\hat{x}(t)$ have the same distribution for every t. This means that *European* option prices based on $\hat{x}$ correspond exactly to the prices in the full model. The coefficients of $\hat{x}$ are, however, not trivial to obtain. Nonetheless, there has been continuous interest in the financial community in applications of Markovian projections, see, for instance, [18] and [3] for two recent examples.

In contrast, the surrogate model $\tilde{x}$ is often easier to construct than the Markovian projection $\hat{x}$. Moreover, our construction provides that $\tilde{x}$ is close to CS on path-space, which directly allows the application to American option pricing. This comes at the price of being only an approximation, though. Moreover, MOR may provide good low dimensional surrogate models even in situations when there is no natural low-dimensional intermediate process, i.e., when $p \approx n$ as of above.

Remark 1. *Generally, the surrogate model $\tilde{x}$ does not have any specific financial interpretation. Hence, its only justification is the approximation quality with respect to the quantity of interest.*

Outline of the paper. After setting the stage in Section 2, we introduce techniques to provide quantitative estimates of "importance" of projections of the state for the dynamics of the process in Section 3. These quantitative estimates are then used in Section 4 to identify

especially efficient reduced models. We continue to provide error bounds in Section 5. Numerical experiments are reported in Section 6, followed by concluding remarks in Section 7. Some general definitions and technical proofs are presented in an appendix.

2. Setting and covariance functions

Let $W = (W_1, \dots, W_q)^T$ be an $\mathbb{R}^q$-valued with mean zero Wiener process with covariance matrix $\mathbf{K} = (k_{ij})$, i.e., $\mathbb{E}[W(t)W^T(t)] = \mathbf{K}t$ for $t \in [0, T]$, where $T > 0$ is the terminal time. Suppose that W and all stochastic process appearing in this paper are defined on a filtered probability space $\big(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P}\big)$[2]. In addition, we assume W to be $(\mathcal{F}_t)_{t\in[0,T]}$-adapted and the increments $W(t + h) - W(t)$ to be independent of $\mathcal{F}_t$ for $t, h \geq 0$. We consider the following large-scale Heston type model:

$$dx(t) = Ax(t)dt + \sum_{i=1}^{q} \sqrt{v(t)} N_i x(t) dW_i(t), \quad x(0) = x_0 = Bz, \tag{1a}$$

$$y(t) = Cx(t), \quad t \in [0, T], \tag{1b}$$

where $A, N_i \in \mathbb{R}^{n\times n}$ and $C \in \mathbb{R}^{p\times n}$. Moreover, the set of initial conditions, in which we are interested, is spanned by the columns of a matrix $B \in \mathbb{R}^{n\times m}$, i.e., there is a vector $z \in \mathbb{R}^m$ such that $x_0 = Bz$. This assumption allows to construct a reduced-order system that performs well for several initial states. However, there are many financial applications, where only a single x_0 is of interest. Then, we have $B = x_0$ and $z = 1$. The scalar $(\mathcal{F}_t)_{t\in[0,T]}$-adapted stochastic process $(v(t))_{t\in[0,T]}$ is non-negative, $\mathbb{P}$-a.s. bounded from above by a constant $c > 0$ and called variance process. The variance process is assumed to be bounded for theoretical considerations below. Practically, boundedness is less relevant. The state dimension n is assumed to be large and the quantity of interest y is rather low-dimensional, i.e., $p \ll n$.

Below, the dependence of the state variable on x_0 is sometimes indicated by writing $x(t; x_0)$, $t \in [0, T]$, for the solution to (1a). Furthermore, we write $M_1 \leq M_2$ for two symmetric matrices M_1 and M_2 if $M_2 - M_1$ is symmetric positive semidefinite. In order to identity the important states in system (1), the covariance function and an upper bound for the covariance will be of interest. Therefore, we formulate the following lemmas.

Lemma 2.1. *The matrix-valued function* $\mathbb{E}\left[x(t; x_0)x^T(t; x_0)\right]$, $t \in [0, T]$, *is a solution to*

$$\dot{X}(t) \leq AX(t) + X(t)A^T + c\sum_{i,j=1}^{q} N_i X(t) N_j^T k_{ij}, \quad X(0) = x_0 x_0^T, \tag{2}$$

where k_{ij} *is the* ij*th entry of the covariance matrix* $\mathbf{K}$.

Proof. The proof is given in Appendix B.1. □

[2] $(\mathcal{F}_t)_{t\in[0,T]}$ shall be right continuous and complete.

We denote the solution to (1a) by x_c if the process v is replaced by its upper bound c ($v \equiv c$). We call (1) Black Scholes model in case the volatility is constant. The covariance function of x_c can be derived through the identity given in the following lemma.

Lemma 2.2. *The matrix-valued function* $\mathbb{E}\left[x_c(t;x_0)x_c^T(t;x_0)\right]$, $t \in [0,T]$, *satisfies*

$$\dot{X}_c(t) = AX_c(t) + X_c(t)A^T + c\sum_{i,j=1}^{q} N_i X_c(t) N_j^T k_{ij}, \quad X_c(0) = x_0 x_0^T, \tag{3}$$

Proof. The statement of this lemma is a special case of [25, Lemma 2.1] □

We now formulate a Gronwall type lemma for matrix differential inequalities involving resolvent positive operators. We refer to Appendix A for a definition of these operators.

Lemma 2.3. *Suppose that* L *is a resolvent positive operator on the space of symmetric matrices. Let the matrix-valued function* $X(t) \geq 0$, $t \in [0,T]$, *satisfy*

$$\dot{X}(t) \leq L(X(t)) \tag{4}$$

and let $Z(t) \geq 0$, $t \in [0,T]$, *be the solution to the matrix differential equation*

$$\dot{Z}(t) = L(Z(t)). \tag{5}$$

If $X(0) \leq Z(0)$, *we have that* $X(t) \leq Z(t)$ *for all* $t \in [0,T]$.

Proof. The proof of this theorem for a special resolvent positive operator is given in [26, Lemma 3.3]. In order to render this paper as self-contained as possible, the proof is stated in Appendix B.2 using the same arguments. □

Lemma 2.3 together with Lemmas 2.1 and 2.2 implies that

$$\mathbb{E}\left[x(t;x_0)x^T(t;x_0)\right] \leq \mathbb{E}\left[x_c(t;x_0)x_c^T(t;x_0)\right], \tag{6}$$

since $L(X) := AX + XA^T + c\sum_{i,j=1}^{q} N_i X N_j^T k_{ij}$ defines a resolvent positive operator on the space of symmetric matrices, see Appendix A. This means that the covariance function of a suitable Black Scholes model dominates the one of a Heston model in case the volatility function is bounded.

Remark 2. *We can use the same approach if we allow for a different volatility* v_i *(*$i = 1,\dots,q$*) in every summand of the diffusion in (1a). Then, boundedness has to be understood in a more general sense, i.e., we need the existence of a positive semidefinite matrix* $\mathbf{C} = (c_{ij})_{i,j=1,\dots,q}$ *such that*

$$\left(v_1^{\frac{1}{2}}(t),\dots,v_q^{\frac{1}{2}}(t)\right)^T \left(v_1^{\frac{1}{2}}(t),\dots,v_q^{\frac{1}{2}}(t)\right) \leq \mathbf{C}$$

for all $t \in [0,T]$. *The operator* L *in Lemma 2.1 then becomes* $L(X) = AX + XA^T + \sum_{i,j=1}^{q} N_i X N_j^T c_{ij} k_{ij}$. *The associated Black Scholes model that guarantees the identity as in Lemma 2.2, is given by setting* $v_i \equiv 1$ *and replacing the Wiener process with covariance*

matrix $\mathbf{K}$ by a Wiener process with covariance $\mathbf{K} \circ \mathbf{C}$, where $\cdot \circ \cdot$ denotes the component-wise product of two matrices. Notice that $\mathbf{K} \circ \mathbf{C}$ is positive semidefinite again due to Schur's product theorem [30].

3. Characterization of dominant states

We are interested in the dominant subspace of system (1) meaning that we aim to identify states that are less important in both equation (1a) and (1b). Those can be neglected in the system dynamics, leading us to an approximation of the system in a lower dimension.

The objects that we choose to identify unimportant states are related to matrices that are used in deterministic control theory. In linear deterministic control systems, the so-called reachability Gramian characterizes the minimal energy that is needed to steer a system from zero to some desired state at time T. Moreover, the observability Gramian determines the energy that is caused by the observations of an unknown initial state on the time interval $[0, T]$, see, e.g., [1]. Consequently, these Gramians can be used to identify states that require a large amount of energy to be reached and states that produce only very little observation energy. Those are less relevant in a control system.

We use these ideas and extend them to the stochastic uncontrolled framework considered here. The matrices identifying the dominant subspaces of system (1) will also be called Gramians due to the link between the concepts.

3.1. **Dominant subspaces of (1a).**

We introduce the fundamental solution to (1a) as an $\mathbb{R}^{n\times n}$-valued stochastic process Φ solving

$$\Phi(t) = I + \int_0^t A\Phi(s)ds + \sum_{i=1}^{q} \int_0^t \sqrt{v(s)} N_i \Phi(s) dW_i(s), \quad t \in [0, T]. \tag{7}$$

If $v \equiv c$, the fundamental solution is denoted by Φ_c. It is not hard to see that the solution to (1a) is given by

$$x(t; x_0) = \Phi(t)x_0 = \Phi(t)Bz, \tag{8}$$

because we assumed that the initial state is spanned by the columns of B. Below, $\langle \cdot, \cdot \rangle_2$ denotes the Euclidean inner product and $\|\cdot\|_2^2$ is the corresponding norm.

Based on (8), let us now identify the states in (1a) that play a minor role. We obtain

$$\begin{aligned} \mathbb{E}\left|\langle x(t; x_0), \tilde{x}\rangle_2\right|^2 &= \mathbb{E}\left|\langle \Phi(t)Bz, \tilde{x}\rangle_2\right|^2 = \mathbb{E}\left|\langle z, B^T\Phi^T(t)\tilde{x}\rangle_2\right|^2 \\ &\leq \tilde{x}^T \mathbb{E}\left[\Phi(t)BB^T\Phi^T(t)\right]\tilde{x}\, \|z\|_2^2 \end{aligned} \tag{9}$$

for a given vector $\tilde{x} \in \mathbb{R}^n$ and using Cauchy's inequality. Since $\mathbb{E}\left[\Phi(t)BB^T\Phi^T(t)\right]$ might not be available from the computational point of view, we find an estimate based on Φ_c in the following proposition.

Proposition 3.1. *Let Φ be the fundamental solution to (1a) and suppose that Φ_c is the fundamental solution to (1a) for the special case $v \equiv c$. Then, we have*

$$\mathbb{E}\left[\Phi(t)BB^T\Phi^T(t)\right] \leq \mathbb{E}\left[\Phi_c(t)BB^T\Phi_c^T(t)\right].$$

Proof. We denote the ith column of the matrix B by b_i, allowing us to write $\Phi(t)B = \left[\ x(t;b_1),\dots,x(t;b_m)\ \right]$. Hence, we have

$$\mathbb{E}\left[\Phi(t)BB^T\Phi^T(t)\right] = \sum_{k=1}^{m}\mathbb{E}\left[x(t;b_k)x^T(t;b_k)\right]. \tag{10}$$

Applying (6) to (10) yields

$$\mathbb{E}\left[\Phi(t)BB^T\Phi^T(t)\right] \leq \sum_{k=1}^{m}\mathbb{E}\left[x_c(t;b_k)x_c^T(t;b_k)\right] = \mathbb{E}\left[\Phi_c(t)BB^T\Phi_c^T(t)\right].$$

This concludes the proof. □

Combining (9) with Proposition 3.1, we find

$$\mathbb{E}\left|\langle x(t;x_0),\tilde{x}\rangle_2\right|^2 \leq \tilde{x}^T F(t)\tilde{x}\,\|z\|_2^2, \tag{11}$$

where $F(t) := \mathbb{E}\left[\Phi_c(t)BB^T\Phi_c^T(t)\right]$. We define $P_T := \int_0^T F(t)dt$ and call P_T (time-limited) reachability Gramian. Integrating both sides of (11) over $[0,T]$ yields

$$\int_0^T \mathbb{E}\left|\langle x(t;x_0),\tilde{x}\rangle_2\right|^2 dt \leq \tilde{x}^T P_T\tilde{x}\,\|z\|_2^2. \tag{12}$$

Consequently, the Gramian P_T characterizes the relevant subspaces as we see in the next proposition.

Proposition 1. *Let $x(\cdot;x_0)$ be the solution to (1a) with initial state $x_0 = Bz$, i.e., it is spanned by the columns of B. Then, it holds that*

$$x(t;x_0) \in \operatorname{im} P_T \quad \mathbb{P}\otimes dt\text{-a.s. on } \Omega\times[0,T].$$

Proof. If $\tilde{x} \in \ker P_T$, then the left-side of (12) is zero, which implies that $\langle x(t;x_0),\tilde{x}\rangle_2 = 0$ $\mathbb{P}\otimes dt$-a.s. Since P_T is symmetric positive semidefinite, this yields the claim. □

Thus, the states that are not in $\operatorname{im} P_T$ are not important in equation (1a). However, it is also important to identify the states that play a minor role. Therefore, we turn our attention to states that lie in $\operatorname{im} P_T$ but that are nevertheless less important. We can choose an orthonormal basis of eigenvectors $(p_k)_{k=1,\dots,n}$ of P_T with associated eigenvalues $(\lambda_k)_{k=1,\dots,n}$. Then, the following representation

$$x(t;x_0) = \sum_{k=1}^{n}\langle x(t;x_0),p_k\rangle_2\, p_k$$

holds. Setting $\tilde{x} = p_k$ in (12) leads to

$$\int_0^T \mathbb{E}\left|\langle x(t;x_0), p_k\rangle_2\right|^2 dt \leq \lambda_k \left\|z\right\|_2^2. \tag{13}$$

Consequently, $x(t;x_0)$ is small in the direction of p_k if λ_k is small. Hence, states with a large component in the direction of such a p_k are less relevant. This means that that eigenspaces of P_T corresponding to small eigenvalues λ_k play a minor role in the system dynamics.

Remark 3. *P_T is related to the Gramian used in* [4]*. However, they choose $\lim_{T\to\infty} P_T$ in some asymptotically stable deterministic setting , i.e., $N_i = 0$ and $\lambda(A) \subset \mathbb{C}_-$, where $\lambda(\cdot)$ denotes the spectrum of a matrix. In the stochastic case the respective stability condition were $\mathbb{E}\left\|x_c(t;x_0)\right\|_2^2 \to 0$ for $t \to \infty$ and all initial conditions x_0 (mean square asymptotic stability), see, e.g.,* [12, 19, 25]*. Stability is not assumed in this paper such that $\lim_{T\to\infty} P_T$ does not exist in general. Moreover, the motivation to use the reachability Gramian P_T is different from the motivation given in* [4]*.*

We conclude this section by a discussion on how to compute P_T which allows to identify redundant information in the system. Using the representation of F in (10) and applying Lemma 2.2 to every summand of its right-side, we see that F satisfies

$$\dot{F}(t) = AF(t) + F(t)A^T + c\sum_{i,j=1}^{q} N_i F(t) N_j^T k_{ij}, \quad F(0) = BB^T. \tag{14}$$

Integrating both sides of (14) yields

$$F(T) - BB^T = AP_T + P_T A^T + c\sum_{i,j=1}^{q} N_i P_T N_j^T k_{ij}. \tag{15}$$

This means that the large-scale generalized Lyapunov equation (15) needs to be solved to derive P_T. This can be done also in a large-scale setting for a given left-side. However, the left-side of (15) depends on $F(T)$, a matrix that needs to be computed beforehand. For dimensions n of a few hundreds, this can be done directly by vectorizing (14). Defining $f(t) := \operatorname{vec}(F(t))$, we then obtain

$$\dot{f}(t) = \mathcal{K} f(t), \quad f(0) = \operatorname{vec}(BB^T), \tag{16}$$

where $\operatorname{vec}(\cdot)$ is the vectorization of a matrix, $\cdot \otimes \cdot$ denotes the Kronecker product of two matrices and

$$\mathcal{K} := I \otimes A + A \otimes I + c\sum_{i,j=1}^{q} N_i \otimes N_j k_{ij}.$$

Consequently, deriving $F(T)$ relies on the efficient computation of a matrix exponential, since

$$f(T) = \mathrm{e}^{\mathcal{K}T} \operatorname{vec}(BB^T). \tag{17}$$

A discussion on how to determine a matrix exponential efficiently can be found in [21] and references therein. However, we need to assume that $0 \notin \lambda(\mathcal{K})$. This guarantees a unique

solution of (15) which we suppose to have below. If system (1) were mean square asymptotically stable in the spirit of Remark 3, then we could take $T \to \infty$ in (15) and $F(T)$ would disappear in the limit which makes the computation of the (infinite) reachability Gramian much simpler. Such type of (infinite) reachability Gramians are, e.g., used to characterize reachability energies in mean square asymptotically stable controlled stochastic systems having time-invariant coefficients [10, 25], a setting that differs significantly from the one considered here.

More advanced approaches need to be used to solve for $F(T)$ if n is very large. There are relevant examples in which $F(T)$ and hence P_T can be derived explicitly as we will see in Section 6.

3.2. **Dominant subspace of (1b).** We now characterize the importance of an initial state x_0 in the output equation. The initial condition is not relevant if the corresponding output $y(\cdot; x_0)$ has zero energy and is of low relevance if the output energy is small, since those initial states barely contribute to the quantity of interest. We begin with an estimate for y based on the result of Section 2.

Proposition 2. *Suppose that y is given by (1b) and let us assume that y_c is the output associated with the solution to the solution of (1a) if $v \equiv c$, i.e., $y_c(t) = Cx_c(t)$. Then, we have*

$$\mathbb{E}\int_0^T \|y(t)\|_2^2\,dt \leq \mathbb{E}\int_0^T \|y_c(t)\|_2^2\,dt. \tag{18}$$

Proof. We use the linearity of the trace to obtain

$$\mathbb{E}\int_0^T \|y(t)\|_2^2\,dt = \mathbb{E}\int_0^T \operatorname{tr}(Cx(t)x^T(t)C^T)dt = \int_0^T \operatorname{tr}(C\mathbb{E}[x(t)x^T(t)]C^T)dt.$$

Using that (6) is preserved when the trace is applied yields

$$\mathbb{E}\int_0^T \|y(t)\|_2^2\,dt \leq \int_0^T \operatorname{tr}(C\mathbb{E}[x_c(t)x_c^T(t)]C^T)dt = \mathbb{E}\int_0^T \|y_c(t)\|_2^2\,dt$$

which concludes the proof of this proposition. □

Now, the goal is to find a bound for the energy of y_c. Therefore, we introduce Q_T as the solution to

$$G(T) - C^TC = A^TQ_T + Q_TA + c\sum_{i,j=1}^{q} N_i^TQ_TN_jk_{ij}, \tag{19}$$

an equation that can be solved for large n once the left-side is given. We refer to Q_T as the observability Gramian since it characterizes the observation energy as we will see below.

$G(t)$, $t \in [0,T]$, entering in (19) satisfies

$$\dot{G}(t) = A^T G(t) + G(t)A + c\sum_{i,j=1}^{q} N_i^T G(t) N_j k_{ij}, \quad G(0) = C^T C, \tag{20}$$

i.e., $Q_T = \int_0^T G(s)ds$. Notice that if n is not too large, $G(T)$ can be computed analogously to (17) meaning that $g(T) := \mathrm{vec}(G(T))$ is given by

$$g(T) = \mathrm{e}^{\mathcal{K}^T T}\,\mathrm{vec}(C^T C).$$

Below, we distinguish between two cases. We first discuss the case in which the system matrices commute.

Commuting matrices. We find a representation for Q_T and subsequently an energy estimate for y_c in case all the matrices in (1a) commute. For that purpose, we establish the following result.

Proposition 3.2. *Let us assume that all matrices $A, N_1, \ldots, N_q$ commute. Hence, we have that these matrices commute with the fundamental solution Φ_c, i.e.,*

$$A\Phi_c(t) = \Phi_c(t)A \quad \textit{and} \quad N_i\Phi_c(t) = \Phi_c(t)N_i \tag{21}$$

for all $t \in [0,T]$ and $i = 1, \ldots, q$.

Proof. These identities hold since the left and the right-sides satisfy the same different equation, e.g., one can multiply (7) with A from the left to obtain the equation for $A\Phi_c$ and with A from the right to get the one for $\Phi_c A$. Since all system matrices commute, the equations coincide. Similarly, one finds the indents for the matrices N_i. □

The example considered in Section 6 satisfies the assumption of Proposition 3.2. Furthermore, notice that in the deterministic case ($N_i = 0$), we have $\Phi_c(t) = \mathrm{e}^{At}$ and hence (21) is always given. Based on Proposition 3.2, a representation of Q_T can be found.

Proposition 3.3. *Under the assumptions of Proposition 3.2, we have*

$$Q_T = \int_0^T \mathbb{E}\left[\Phi_c^T(t) C^T C \Phi_c(t)\right] dt.$$

Proof. We apply Ito's product rule to $\Phi_c^T(t)C^T C\Phi_c(t)$ and take the correlation of the noise processes into account. This yields

$$\begin{aligned}
&d\left(\Phi_c^T(t)C^T C\Phi_c(t)\right)\\
&= d\left(\Phi_c^T(t)\right) C^T C\Phi_c(t) + \Phi_c^T(t)C^T C d\left(\Phi_c(t)\right) + d\left(\Phi_c^T(t)\right) C^T C d\left(\Phi_c(t)\right)\\
&= d\left(\Phi_c^T(t)\right) C^T C\Phi_c(t) + \Phi_c^T(t)C^T C d\left(\Phi_c(t)\right) + c\sum_{i,j=1}^{q} \Phi_c^T(t) N_i^T C^T C N_j k_{ij} \Phi_c(t) dt.
\end{aligned}$$

Above, we plug in (7) for the case when $v \equiv c$ and take the expected value on both sides. Hence, using that the Ito integral has mean zero, we have

$$\begin{aligned} &d\left(\mathbb{E}\left[\Phi_c^T(t)C^TC\Phi_c(t)\right]\right) \\ &= \mathbb{E}\left[\Phi_c^T(t)\left(A^TC^TC + C^TCA + c\sum_{i,j=1}^{q} N_i^TC^TCN_jk_{ij}\right)\Phi_c(t)\right]dt. \end{aligned}$$

Due to (21) we see that $\mathbb{E}\left[\Phi_c^T(t)C^TC\Phi_c(t)\right]$, $t \in [0,T]$, solves (20) and thus $Q_T = \int_0^T \mathbb{E}\left[\Phi_c^T(t)C^TC\Phi_c(t)\right]dt$. □

Inequality (18) and Proposition 3.3 now imply that

$$\mathbb{E}\int_0^T \|y(t)\|_2^2\,dt \le \mathbb{E}\int_0^T \|C\Phi_c(t)x_0\|_2^2\,dt = x_0^TQ_Tx_0. \tag{22}$$

Initial states that are spanned by eigenvectors of Q_T belonging to the small eigenvalues lead to a small right-side in (22) and consequently yield a small output y. Hence, we know that eigenspaces of Q_T corresponding to the small eigenvalues are less relevant in (1b).

General case. We find another bound on the energy of y_c and hence also for y in the general case.

Proposition 3.4. *If y is the quantity of interest in system (1) and Q_T the solution of (19), then*

$$\mathbb{E}\int_0^T \|y(t)\|_2^2\,dt \le x_0^TQ_Tx_0 + \mathcal{R}(T), \tag{23}$$

where $\mathcal{R}(T) := \mathbb{E}\int_0^T x_c^T(t;x_0)G(T)x_c(t;x_0)dt - \mathbb{E}\left[x_c^T(T;x_0)Q_Tx_c(T;x_0)\right]$ with G solving (20).

Proof. We make use of

$$\mathbb{E}\left[x_c^T(T;x_0)Q_Tx_c(T;x_0)\right] = \operatorname{tr}(Q_T\mathbb{E}\left[x_c(T;x_0)x_c^T(T;x_0)\right]). \tag{24}$$

We obtain from Lemma 2.2 that

$$\begin{aligned} Q_T\mathbb{E}\left[x_c(T;x_0)x_c^T(T;x_0)\right] = Q_Tx_0x_0^T &+ \mathbb{E}\int_0^T Q_TAx_c(t;x_0)x_c^T(t;x_0)dt \\ &+ \mathbb{E}\int_0^T Q_Tx_c(t;x_0)x_c^T(t;x_0)A^Tdt \\ &+ c\sum_{i,j=1}^{q}\mathbb{E}\int_0^T Q_TN_ix_c(t;x_0)x_c^T(t;x_0)N_j^Tk_{ij}dt. \end{aligned}$$

Using properties of the trace, (24) becomes

$$\mathbb{E}\left[x_c^T(T;x_0)Q_T x_c(T;x_0)\right]$$
$$= x_0^T Q_T x_0 + \mathbb{E}\int_0^T x_c^T(t;x_0)\left(A^T Q_T + Q_T A + c\sum_{i,j=1}^{q} N_i^T Q_T N_j q_{ij}\right) x_c(t;x_0)ds.$$

We insert equation (19) into the above identity to get

$$\mathbb{E}\left[x_c^T(T;x_0)Q_T x_c(T;x_0)\right] = x_0^T Q_T x_0 + \mathbb{E}\int_0^T x_c^T(t;x_0)(G(T) - C^T C)x_c(t;x_0)dt$$

(25)

$$= x_0^T Q_T x_0 - \mathbb{E}\int_0^T \|y_c(t)\|_2^2\, dt + \mathbb{E}\int_0^T x_c^T(t;x_0)G(T)x_c(t;x_0)dt$$

This, together with (18), gives us the result. □

Assuming that the remainder term $\mathcal{R}(T)$ is not too large, the same conclusions as below (22) can be made. The eigenspaces that belong to the small eigenvalues of Q_T are unimportant. If the system is mean square asymptotic stable, then $Q_\infty := \lim_{T\to\infty} Q_T$ exists and $\mathcal{R}(T) \to 0$ as $T \to \infty$. Taking the limit of $T \to \infty$ in (23) would then lead to a characterization of the output energy by Q_∞ without a remainder term. Q_∞ is also easier to determine than Q_T since it solves (19) with $G(T) = 0$. Energy estimates based on Q_∞ are shown in [6, 10, 25] if the variance v is constant.

We can also get to a more explicit bound by applying Gronwall's lemma to (25) if Q_T is regular. Defining $\alpha(T) := x_0^T Q_T x_0 - \mathbb{E}\int_0^T \|y_c(t)\|_2^2\, dt$ (25) becomes

$$\mathbb{E}\left[\left\|Q_T^{\frac{1}{2}} x_c(t;x_0)\right\|_2^2\right] = \alpha(T) + \mathbb{E}\int_0^T \left\|G^{\frac{1}{2}}(T)x_c(t;x_0)\right\|_2^2 dt$$
$$\leq \alpha(T) + k_T\mathbb{E}\int_0^T \left\|Q_T^{\frac{1}{2}} x_c(t;x_0)\right\|_2^2 dt$$

where $k_T := \left\|G^{\frac{1}{2}}(T)Q_T^{-\frac{1}{2}}\right\|_2^2$. Gronwall's lemma leads to

$$0 \leq \mathbb{E}\left[\left\|Q_T^{\frac{1}{2}} x_c(t;x_0)\right\|_2^2\right] \leq \alpha(T) + \mathbb{E}\int_0^T \alpha(t)k_T\, \mathrm{e}^{k_T(T-t)}\, dt.$$

With a few more steps, we find

$$\mathbb{E}\int_0^T \|y(t)\|_2^2\, dt \leq x_0^T Q_T x_0\, \mathrm{e}^{k_T T},$$

but this bound can not be expected to be tight.

4. State-space transformation and reduced-order model

Balancing related MOR like balanced truncation were initially invented for controlled linear deterministic systems that are asymptotically stable and have zero initial states [23]. Balanced truncation has been extended to stochastic systems with similar properties [6, 10]. Subsequently, this scheme was studied for deterministic and stochastic systems with non-zero initial conditions [4, 5]. However, all these methods are restricted to stable systems. A method for deterministic equations called time-limited balanced truncation aiming to create a good reduced system on a finite time interval only was introduced in [15]. As pointed out in [21], this method has some potential in the context of unstable systems.

The method explained below is a combination of all the methods mentioned above. It follows the same concept which is simultaneously diagonalizing system Gramians. Here, the Gramians are P_T and Q_T solving (15) and (19), respectively. We have shown the relevance of these Gramians in Section 3. Diagonalizing both P_T and Q_T means that we create a system in which the important states in equations (1a) and (1b) are the same. Hence, the unimportant ones can be easily identified and thus truncated.

The unimportant ones can then be easily truncated.

Let $S \in \mathbb{R}^{n\times n}$ be a regular matrix. We do a coordinate transformation by introducing

$$\hat{x}(t) = Sx(t).$$

Based on (1) the associated system is

$$d\hat{x}(t) = \hat{A}\hat{x}(t)dt + \sum_{i=1}^{q} \sqrt{v(t)}\hat{N}_i x(t) dW_i(t), \quad \hat{x}(0) = Sx_0 = SBz, \tag{26a}$$

$$y(t) = \hat{C}\hat{x}(t), \quad t \in [0,T], \tag{26b}$$

where $\hat{A} = SAS^{-1}$, $\hat{B} = SB$, $\hat{C} = CS^{-1}$ and $\hat{N}_i = SN_iS^{-1}$. Notice that the quantity of interest does not change with this transformation. However, the matrices characterizing the importance of states in (26a) and (26b) are different ones. For the transformed system (26), these become

$$\hat{P}_T = SP_TS^T \quad \text{and} \quad \hat{Q}_T = S^{-T}Q_TS^{-1}.$$

The above relation is obtained by multiplying (14) with S from the left and with S^T. Moreover, (20) needs to be multiplied with S^{-T} from the left and with S^{-1} from the right.

We now choose S such that $\hat{P}_T = \hat{Q}_T = \Sigma_T = \operatorname{diag}(\sigma_1, \dots, \sigma_n)$, where $\sigma_1 \geq \dots \geq \sigma_n > 0$ are called Hankel singular values (HSVs) and given by $\sigma_i = \sqrt{\lambda_i(P_TQ_T)}$, where $\lambda_i(\cdot)$ denotes the ith eigenvalue of the matrix and $i = 1, \dots, n$. Such a system is called balanced. A transformation like this always exists if $P_T, Q_T > 0$. It is, together with its inverse, derived the following way:

$$S = \Sigma_T^{-\frac{1}{2}} U^T L_Q^T \quad \text{and} \quad S^{-1} = K_P V \Sigma_T^{-\frac{1}{2}}.$$

The above matrices are computed from factorizations $P_T = K_P K_P^T$ and $Q_T = L_Q L_Q^T$ as well as from the singular value decomposition of $K_P^T L_Q = V \Sigma U^T$.

In a balanced system, it is easy to identify the unimportant states. They are the ones corresponding to the small HSVs of the system and represented by x_2 given by the partition of the balanced state variable

$$\hat{x}(t) = Sx(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix},$$

where $x_1(t) \in \mathbb{R}^{\tilde{n}}$ represents the relevant states in the system dynamics. Furthermore, we partition the balanced realization as follows:

$$SAS^{-1} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad SB = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}, \quad CS^{-1} = \begin{bmatrix} C_1 & C_2 \end{bmatrix}, \quad SN_iS^{-1} = \begin{bmatrix} N_{i,11} & N_{i,12} \\ N_{i,21} & N_{i,22} \end{bmatrix},$$

where $A_{11} \in \mathbb{R}^{\tilde{n} \times \tilde{n}}$ etc. With this, system (26) becomes

$$\begin{bmatrix} dx_1 \\ dx_2 \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} dt + \sum_{i=1}^{q} \sqrt{v} \begin{bmatrix} N_{i,11} & N_{i,12} \\ N_{i,21} & N_{i,22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} dW_i, \quad \hat{x}(0) = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} z, \tag{27}$$

$$y(t) = \begin{bmatrix} C_1 & C_2 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}, \quad t \in [0, T]. \tag{28}$$

The time dependence is omitted in (27) to shorten the notation. The reduced system of dimension $\tilde{n} \ll n$ is now obtained by neglecting x_2, i.e., the second line in (27) is truncated the remaining x_2 variables are set zero in both the first line of (27) and in (28). The reduced-order model then is

$$d\tilde{x}(t) = A_{11}\tilde{x}(t)dt + \sum_{i=1}^{q} \sqrt{v(t)} N_{i,11}\tilde{x}(t)dW_i(t), \quad \tilde{x}(0) = B_1 z, \tag{29a}$$

$$\tilde{y}(t) = C_1\tilde{x}(t), \quad t \in [0, T]. \tag{29b}$$

where $A_{11}, N_{i,11} \in \mathbb{R}^{\tilde{n} \times \tilde{n}}$, $B_1 \in \mathbb{R}^{\tilde{n} \times m}$ and $C_1 \in \mathbb{R}^{p \times \tilde{n}}$.

5. Error bound analysis

We introduce an error system by combining (1a) and (29a) with an output equation that represents the error between (1b) and (29b). The error system is

$$\begin{aligned} dx^e(t) &= A^e x^e(t)dt + \sum_{i=1}^{q} \sqrt{v(t)} N_i^e x^e(t)dW_i(t), \quad x^e(0) = B^e z, \\ y^e(t) &= C^e x^e(t), \quad t \in [0, T], \end{aligned} \tag{30}$$

where the error state x^e and the error matrices (A^e, B^e, C^e, N_i^e) are

$$x^e = \begin{bmatrix} x \\ \tilde{x} \end{bmatrix}, \; A^e = \begin{bmatrix} A & 0 \\ 0 & A_{11} \end{bmatrix}, \; B^e = \begin{bmatrix} B \\ B_1 \end{bmatrix}, \; C^e = \begin{bmatrix} C & -C_1 \end{bmatrix}, \; N_i^e = \begin{bmatrix} N_i & 0 \\ 0 & N_{i,11} \end{bmatrix}. \tag{31}$$

Let us again assume that an index c indicates that v in (30) is replaced by c. We obtain

$$\mathbb{E}\int_0^T \|y^e(t)\|_2^2\,dt \le \mathbb{E}\int_0^T \|y_c^e(t)\|_2^2\,dt$$

the same way as in (18). Based on the fundamental solution Φ^e (or Φ_c^e), the quantity of interest given a constant volatility function, is represented by $y_c^e(t) = C^e\Phi_c^e(t)B^e z$. Plugging this into the above inequality yields

$$\begin{aligned}\mathbb{E}\int_0^T \|y^e(t)\|_2^2\,dt \le& \mathbb{E}\int_0^T \|C^e\Phi_c^e(t)B^e z\|_2^2\,dt \le \mathbb{E}\int_0^T \|C^e\Phi_c^e(t)B^e\|_F^2\,dt\,\|z\|_2^2\\ &= \operatorname{tr}(C^e P_T^e (C^e)^T)\,\|z\|_2^2\,, \end{aligned} \tag{32}$$

where we set $P_T^e := \int_0^T F^e(t)dt$ with $F^e(t) = \Phi_c^e(t)B^e(B^e)^T(\Phi_c^e)^T(t)$. Analogue to (14), F^e solves

$$\dot{F}^e(t) = A^eF^e(t) + F^e(t)(A^e)^T + c\sum_{i,j=1}^q N_i^eF^e(t)(N_j^e)^Tk_{ij},\quad F^e(0) = B^e(B^e)^T. \tag{33}$$

We partition the solution to (33) as follows

$$F^e(t) = \begin{bmatrix} F_{11}(t) & F_{12}(t)\\ F_{12}^T(t) & F_{22}(t)\end{bmatrix} \tag{34}$$

and see that $F_{11}(t) = F(t)$ solving (14) as well as $F_{12}(t) = \bar{F}(t)$ and $F_{22}(t) = \tilde{F}(t)$ that are the solutions to

$$\dot{\bar{F}}(t) = A\bar{F}(t) + \bar{F}(t)A_{11}^T + c\sum_{i,j=1}^q N_i\bar{F}(t)N_{j,11}^Tk_{ij},\quad \bar{F}(0) = BB_1^T \tag{35}$$

$$\dot{\tilde{F}}(t) = A_{11}\tilde{F}(t) + \tilde{F}(t)A_{11}^T + c\sum_{i,j=1}^q N_{i,11}\tilde{F}(t)N_{j,11}^Tk_{ij},\quad \tilde{F}(0) = B_1B_1^T \tag{36}$$

using the partitions in (31). From (32) and (34), we obtain

$$\begin{aligned}&\mathbb{E}\int_0^T \|y(t) - \tilde{y}(t)\|_2^2\,dt = \mathbb{E}\int_0^T \|y^e(t)\|_2^2\,dt\\ &\le \left(\operatorname{tr}(CP_TC^T) - 2\operatorname{tr}(C\bar{P}_TC_1^T) + \operatorname{tr}(C_1\tilde{P}_TC_1^T)\right)\|z\|_2^2\,,\end{aligned} \tag{37}$$

where $\bar{P}_T := \int_0^T \bar{F}(t)dt$ and $\tilde{P}_T := \int_0^T \tilde{F}(t)dt$. By Integrating both (35) and (36), the equations for these two matrices are

$$\bar{F}(T) - BB_1^T = A\bar{P}_T + \bar{P}_TA_{11}^T + c\sum_{i,j=1}^q N_i\bar{P}_TN_{j,11}^Tk_{ij}, \tag{38}$$

$$\tilde{F}(T) - B_1B_1^T = A_{11}\tilde{P}_T + \tilde{P}_TA_{11}^T + c\sum_{i,j=1}^q N_{i,11}\tilde{P}_TN_{j,11}^Tk_{ij}. \tag{39}$$

Since P_T is already known from the balancing procedure explained in Section 4, the bound for the absolute output error in (37) requires only the computation of $\bar{P}_T$ and $\tilde{P}_T$. Since the reduced dimension $\tilde{n}$ is rather small, the corresponding equations (38) and (39) can often be solved directly through vectorization. Hence, we have

$$\operatorname{vec}(\bar{F}(T)) - \operatorname{vec}(BB_1^T) = \bar{\mathcal{K}} \operatorname{vec}(\bar{P}_T),$$
$$\operatorname{vec}(\tilde{F}(T)) - \operatorname{vec}(B_1B_1^T) = \tilde{\mathcal{K}} \operatorname{vec}(\tilde{P}_T)$$

with $\bar{F}(T) = \mathrm{e}^{\bar{\mathcal{K}}T} \operatorname{vec}(BB_1^T)$ and $\tilde{F}(T) = \mathrm{e}^{\tilde{\mathcal{K}}T} \operatorname{vec}(B_1B_1^T)$, where

$$\bar{\mathcal{K}} := \left(I_n \otimes A_{11} + A \otimes I_{\tilde{n}} + c \sum_{i,j=1}^{q} N_i \otimes N_{j,11} k_{ij} \right),$$
$$\tilde{\mathcal{K}} := \left(I_{\tilde{n}} \otimes A_{11} + A_{11} \otimes I_{\tilde{n}} + c \sum_{i,j=1}^{q} N_{i,11} \otimes N_{j,11} k_{ij} \right).$$

Above, the identity matrices are equipped with an index indicating the respective dimension. With (37) a bound for the absolute error of reducing system (1) was found. However, the relative error is more interesting to be analyzed. Therefore, we need a computable lower bound for the L^2-norm of y. This task is relatively simple because the inequality of Cauchy Schwartz yields

$$\int_0^T \|\mathbb{E}[y(t)]\|_2^2 \, dt \leq \mathbb{E} \int_0^T \|y(t)\|_2^2 \, dt.$$

Now, $\mathbb{E}[x(t)]$, $t \in [0,T]$, solves equation (1a) with $c = 0$ which can be seen easily by applying the expected value to both sides of (1a) and by exploiting that the Ito integrals have zero mean. We use that $\Phi_{c=0}(t) = \mathrm{e}^{At}$ such that $\mathbb{E}[y(t)] = C \, \mathrm{e}^{At} \, Bz$. We plug this into the above estimate leading to

$$\mathbb{E} \int_0^T \|y(t)\|_2^2 \, dt \geq \int_0^T \left\| C \, \mathrm{e}^{At} \, Bz \right\|_2^2 dt = z^T B^T Q_{T,0} B z, . \tag{40}$$

where $Q_{T,0} := \int_0^T \mathrm{e}^{A^T t} \, C^T C \, \mathrm{e}^{At} \, dt$. According to Subsection 3.2, $Q_{T,0}$ solves (19) with $c = 0$, i.e.,

$$\mathrm{e}^{A^T T} \, C^T C \, \mathrm{e}^{AT} - C^T C = A^T Q_{T,0} + Q_{T,0} A, \tag{41}$$

an equation that can be solved in a large-scale setting, since there are efficient methods to determine e^{AT} for large n. We summarize the results of this section in a theorem below, where we set $\|y\|_{L^2}^2 := \mathbb{E} \int_0^T \|y(t)\|_2^2 \, dt$.

Theorem 5.1. *Let y be the output of the original system (1) and let $\tilde{y}$ be the output of the reduced model (29). Then, the relative L^2-error between y and $\tilde{y}$ is bounded as follows:*

$$\frac{\|y - \tilde{y}\|_{L^2}}{\|y\|_{L^2}} \leq \frac{\left(\operatorname{tr}(CP_TC^T) - 2 \operatorname{tr}(C\bar{P}_TC_1^T) + \operatorname{tr}(C_1\tilde{P}_TC_1^T) \right)^{\frac{1}{2}} \|z\|_2}{(z^T B^T Q_{T,0} B z)^{\frac{1}{2}}} \tag{42}$$

where P_T, $\bar{P}_T$, $\tilde{P}_T$ and $Q_{T,0}$ are the solutions to (15), (38), (39) and (41), respectively.

Proof. The result follows from (37) and (40). □

The bound in Theorem 5.1 provides a good a priori error estimate, an indicator for the quality of the reduced system. Error bounds for related methods in a deterministic framework can be found in [16, 27, 28].

Remark 4. *Notice that if we aim to reduce a Black-Scholes model with output y_c instead of a Heston model with output y, the energy $\|y_c\|_{L^2}$ is explicitly known according to Section 3.2 given that $A, N_1, \dots, N_q$ commute. Then, it holds that $\|y_c\|_{L^2} = x_0^T Q_T x_0$ such that we can replace $Q_{T,0}$ by Q_T in Theorem 5.1.*

6. Numerical experiments

We apply the MOR technique motivated in Section 3 and explained in Section 4. The goal is to accurately approximate payoff functions associated with the large asset price model (1a) (these are functions of the quantity of interest in (1b)) by payoff functions of the reduced system (29). This type of problem is of particular interest if we price European options with an underlying high-dimensional Heston model because computational complexity can be reduced. Moreover, since the reduced system shows good pathwise approximations, it can be of interest in the context of Bermudan options because regression based methods [22, 31] suffer from the curse of dimensionality which makes them inaccurate in a large-scale setting. Below, we consider a particular Heston model (1) and illustrate the quality of the reduction in dependence of the covariance matrix $\mathbf{K}$ of the noise process.

We consider the following linear stochastic differential equation that represents an asset price model:

$$dx_i(t) = \mathrm{r}x_i(t)dt + \xi_i\sqrt{v(t)}x_i(t)dW_i(t), \quad x_i(0) = x_{0,i}, \tag{43}$$

where x_i denotes the ith component of a price process x ($i = 1, \dots, n$). Moreover, we assume that $\mathrm{r} = 0.02$ is the fixed interest rate and $\xi_i \in [0.2, 0.7]$ are volatility parameter sampled from a uniform distribution. Now, we can rewrite equation (43) in order to guarantee the form given in (1a). The respective matrices are

$$A = \mathrm{r}I, \quad N_i = \xi_i e_i e_i^T, \quad B = x_0 \quad \text{and} \quad z = 1, \tag{44}$$

where e_i is the ith unit vector in $\mathbb{R}^n$, $q = n$, and assuming that we are interested in a single initial value x_0 only. In this particular situation, the matrices $A, N_1, \dots, N_n$ commute and are symmetric. The variance process is $v(t) = \min\{\bar{v}(t), c\}$, $t \in [0, T]$, where $\bar{v}$ is the solution to the following stochastic differential equation:

$$d\bar{v}(t) = a(b - \bar{v}(t))dt + \bar{\sigma}\sqrt{\bar{v}(t)}d\bar{B}(t), \quad v(0) = v_0, \tag{45}$$

where $b = 0.2$ is the long run average variance, $a = 0.2$ is the rate characterizing the speed of convergence of the average variance and $\bar{\sigma} = 0.15$ is the volatility of the volatility process.

Furthermore, $\bar{B}$ is a standard Brownian motion with respect to the filtration $(\mathcal{F}_t)_{t\in[0,T]}$ that negatively correlated with the other standard Brownian motions W_i, i.e., $\mathbb{E}[\bar{B}(t)W_i(t)] = \rho_i t$, where $\rho_i < 0$. The parameters $\xi_i, a, b, \bar{\sigma}$ are chosen to have an average volatility around 0.2, i.e.,

$$\mathbb{E}\left[\sqrt{v(t)}\right]\frac{1}{n}\sum_{i=1}^{n}\xi_i \approx 0.2.$$

Notice that in order to fit the theory, the process v is bounded by a constant c since generally $\bar{v}$ is unbounded. Practically, we simulate a certain number of paths of $\bar{v}$ and choose a c that represents a bound of these simulated paths such that those coincide with the respective paths of v.

Suppose that the quantity of interest is now some one dimensional partial information y of the price process x that is the form

$$y(t) = Cx(t), \tag{46}$$

where the output matrix is $C = [1,\ 1, \dots 1]$. We now determine a reduced system (29). To do so, the matrices P_T and Q_T need to be computed in order to conduct the balancing procedure described in Section 4. Fortunately, these matrices can be derived explicitly from (14) and (20). Plugging in (44) into these equations, we obtain

$$\dot{F}(t) = 2\mathrm{r}IF(t) + c\sum_{i,j=1}^{n} e_i e_i^T F(t) e_j e_j^T \xi_i \xi_j k_{ij}, \quad F(0) = x_0 x_0^T,$$

$$\dot{G}(t) = 2\mathrm{r}IG(t) + c\sum_{i,j=1}^{n} e_i e_i^T G(t) e_j e_j^T \xi_i \xi_j k_{ij}, \quad G(0) = C^T C.$$

By multiplying the above equations with e_i^T from the left and with e_j from the right, we can see that these equations can be solved component-wise. The entries of $F(t) = (f_{ij})_{i,j=1,\dots n}$ satisfy

$$\dot{f}_{ij}(t) = (2\mathrm{r} + c\xi_i\xi_j k_{ij}) f_{ij}(t), \quad f_{ij}(0) = x_{0,i}x_{0,j},$$

such that $f_{ij}(t) = \mathrm{e}^{h_{ij}t}\, x_{0,i}x_{0,j}$, where $h_{ij} := (2\mathrm{r} + c\xi_i\xi_j k_{ij})$. Integrating f_{ij} over $[0,T]$, we find that $P_T = (p_{ij})_{i,j=1,\dots n}$ is given by

$$p_{ij}(t) = \frac{\mathrm{e}^{h_{ij}T} - 1}{h_{ij}} x_{0,i}x_{0,j}.$$

Analogously, it holds that $Q_T = (q_{ij})_{i,j=1,\dots n}$ is represented by

$$q_{ij}(t) = \frac{\mathrm{e}^{h_{ij}T} - 1}{h_{ij}} e_i^T C^T C e_j.$$

Since P_T and Q_T are given explicitly, the reduced system (29) with output $\tilde{y}$ comes basically for free in terms of computational time. This also means that we are able to derive a reduced

model if the number of assets is very large. We investigate the reduction quality for three different covariance matrices. First we consider a matrix with both small and large correlations between the noise processes. We first choose $\mathbf{K} = \mathbf{K}_0 := \tau\tau^T$ according to [13], where $\tau = [\tau_1,\ \tau_2, \dots \tau_n]$ has columns τ_i generated by a vector $s = (s_i)_{i=1,\dots,n-1}$ of samples s_i of independent uniformly distributed random variables with values in $[0.8, 1]$:

$$\tau_1 = \begin{pmatrix} 1 \\ \mathrm{cp}(s) \end{pmatrix}, \quad \tau_2 = \sqrt{1 - s_1^2} \begin{pmatrix} 0 \\ 1 \\ \mathrm{cp}(s_{2:n-1}) \end{pmatrix}, \dots, \tau_n = \sqrt{1 - s_{n-1}^2} \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}.$$

Above, we set $s_{\ell:n-1} := (s_\ell, s_{\ell+1}, \dots, s_{n-1})^T$ and

$$\mathrm{cp}(s) := \left[s_1, s_1 s_2, \dots, s_1 s_2 \dots s_{n-1}\right]^T.$$

Further, we study the two extreme cases of $\mathbf{K} = I$ (independent noise processes) and $\mathbf{K} = \mathbf{1}\mathbf{1}^T$ (perfect correlation), where $\mathbf{1}$ is an n-dimensional vector of ones. We choose $n = 100, T = 1, c = 0.6$ and the initial conditions $x_{0,i} \in [0, 1.25]$ are generated randomly, where $1 \mathrel{\hat{=}} 100$ Dollar. For $\mathbf{K} = \mathbf{K}_0$ we choose $\rho_i \in [-0.9, 0)$, in case of $\mathbf{K} = I$ we have $\rho_i = -0.09$ and we fix $\rho_i = -0.5$ for $\mathbf{K} = \mathbf{1}\mathbf{1}^T$. In all the numerical experiments below, $2\mathrm{e}06$ samples are generated.

6.1. **Approximation error in the quantity of interest.** We begin with analyzing the error between the output of the full system y and the output of the reduced system $\tilde{y}$. The first question is how to choose the dimension $\tilde{n}$ of system (29). The HSVs, i.e., $\sigma_i = \sqrt{\lambda_i(P_T Q_T)}$ are a very good indicator for a suitable choice since the smaller σ_i, the less important the ith state component $\hat{x}_i$ in the balanced system (26) according to what we have derived in Section 3. We can see these values in logarithmic scale for different covariance matrices $\mathbf{K}$ in Figure 1. In each case, we observe that one variable dominates the dynamics meaning that no matter how $\mathbf{K}$ is chosen, a scalar reduced-order model already leads to a relatively good approximation. Moreover, we see that the reduction is expected to be least efficient if all noise processes are independent due to a slow decay of the HSVs. However, with perfect correlation ($\mathbf{K} = \mathbf{1}\mathbf{1}^T$), we only have four non zero HSVs meaning that the 100-dimensional model can be perfectly approximated by a system of four variables. In the case of $\mathbf{K} = \mathbf{K}_0$ the performance is in between the independent and perfectly correlated scenarios. This fits to our general observation that the higher the correlation, the better the algorithm works.

We conclude this subsection by a discussion on the error between the outputs y and $\tilde{y}$ of systems (1) and (29) for this particular example. We determine the relative L^2-error and the corresponding error bound in Theorem 5.1. This bound that is the right-side of (42) is denoted by $\mathcal{EB}$ here. The error bound is relatively tight for the example. In most of the cases it estimates the exact error by factor of three, see Tables 1, 2 and 3. As supposed from the HSVs, very good results are obtained for very small $\tilde{n}$ if $\mathbf{K} = \mathbf{1}\mathbf{1}^T$, compare with Figure 2. However, in the situation of $\mathbf{K} = \mathbf{K}_0$, a reduced-order between 4 and 10 shows a low error, too.

$\tilde{n}$	$\|y-\tilde{y}\|_{L^2} / \|y\|_{L^2}$	$\mathcal{EB}$
1	5.10e−03	1.55e−02
4	2.02e−03	6.18e−03
10	7.01e−04	2.17e−03
15	3.95e−04	1.23e−03
20	2.39e−04	7.46e−04
25	1.62e−04	5.06e−04
50	3.56e−05	1.13e−04

TABLE 1. Relative L^2-error and error bound for $\mathbf{K} = \mathbf{K}_0$.

$\tilde{n}$	$\|y-\tilde{y}\|_{L^2} / \|y\|_{L^2}$	$\mathcal{EB}$
1	2.56e−03	7.79e−03
4	2.23e−03	6.79e−03
10	1.69e−03	5.12e−03
15	1.46e−03	4.42e−03
20	1.27e−03	3.84e−03
25	1.08e−03	3.28e−03
50	5.01e−04	1.51e−03

TABLE 2. Relative L^2-error and error bound for $\mathbf{K} = I$.

$\tilde{n}$	$\|y-\tilde{y}\|_{L^2} / \|y\|_{L^2}$	$\mathcal{EB}$
1	2.40e−03	5.48e−03
2	3.25e−06	1.34e−05
3	2.94e−09	2.98e−08
4	3.76e−12	2.76e−09

TABLE 3. Relative L^2-error and error bound for $\mathbf{K} = \mathbf{1}\mathbf{1}^T$.

6.2. **Approximation error in the payoff.** We have seen in Section 6.1 that the quantity of interest y of system (1), that we specified in (46), can be well approximated by the output $\tilde{y}$ of the reduced system (29) in a path-wise sense on some interval $[0, T]$. However, it is often of interest to consider weak errors instead. Therefore, we consider the following payoff function

$$f(y) = \max\{y - K, 0\}$$

that plays a role in the context of European call options, where K denotes the strike price. We compare the expected payoff $\mathbb{E}f(y(T))$ at time T with the one associated with the reduced system, which is $\mathbb{E}f(\tilde{y}(T))$, for $K = \langle \mathbf{1}, x_0 \rangle_2$ in Figure 3. The relative errors in the expected payoff are larger if the correlations between the noise processes are small. The approximation works best if $\mathbf{K} = \mathbf{1}\mathbf{1}^T$. As displayed in Figure 3, the error is around $5\text{e}{-06}$ for $\tilde{n} = 2$ and our simulations also show an error of $5\text{e}{-13}$ already for $\tilde{n} = 4$. If smaller correlations are involved, a larger $\tilde{n}$ needs to be chosen. However, selecting $4 \leq \tilde{n} \leq 10$ for $\mathbf{K} = \mathbf{K}_0$ already leads to a good estimate of the original payoff.

Looking at Table 4 we observe that weak error (error in the expected payoff) is of the same or of smaller error than the strong error in Table 1. Moreover, it can be seen that the approximation in the payoff is better if the strike price is below the value of the basket $\langle \mathbf{1}, x_0 \rangle_2$ at time zero and it is worse if the strike price is above $\langle \mathbf{1}, x_0 \rangle_2$.

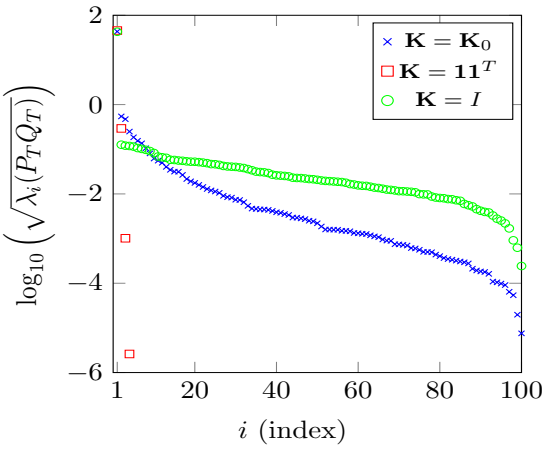

FIGURE 1. Logarithmic HSVs of the large-scale asset model.

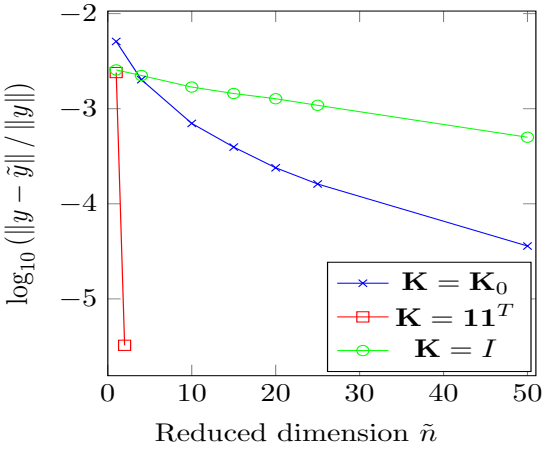

FIGURE 2. Logarithmic relative L^2-error between y and $\tilde{y}$ for $\tilde{n} \in \{1, 4, 10, 15, 20, 25, 50\}$ if $\mathbf{K} = \mathbf{K}_0, I$ and for $\tilde{n} = 1, 2$ if $\mathbf{K} = \mathbf{1}\mathbf{1}^T$.

So far, the weak error has not yet been analyzed concerning error bounds etc. We believe that it requires advanced techniques to succeed in this direction.

Remark 5. *We have seen a good path-wise performance of our method in Section 6.1 and an even better approximation in the payoff in this section. Therefore, we see the potential of pricing high-dimensional Bermudan options with the help of MOR. Classical regression based schemes as in* [22, 31] *cannot accurately determine values of Bermudan options*

	$\lvert\mathbb{E}f(y) - \mathbb{E}f(\tilde{y})\rvert \,/\, \lvert\mathbb{E}f(y)\rvert$		
$\tilde{n}$	$K = 0.9\,\langle \mathbf{1}, x_0\rangle_2$	$K = \langle \mathbf{1}, x_0\rangle_2$	$K = 1.1\,\langle \mathbf{1}, x_0\rangle_2$
1	9.54e−04	1.10e−03	1.98e−02
4	8.81e−04	6.92e−04	5.28e−03
10	2.78e−04	4.13e−04	8.80e−04
15	3.41e−05	8.63e−05	2.22e−04
20	8.62e−06	1.79e−05	2.25e−05
25	2.56e−06	7.22e−06	2.19e−05
50	1.12e−07	2.22e−07	7.09e−07

TABLE 4. Relative error in the payoff function for $\mathbf{K} = \mathbf{K}_0$ and different strike prices K.

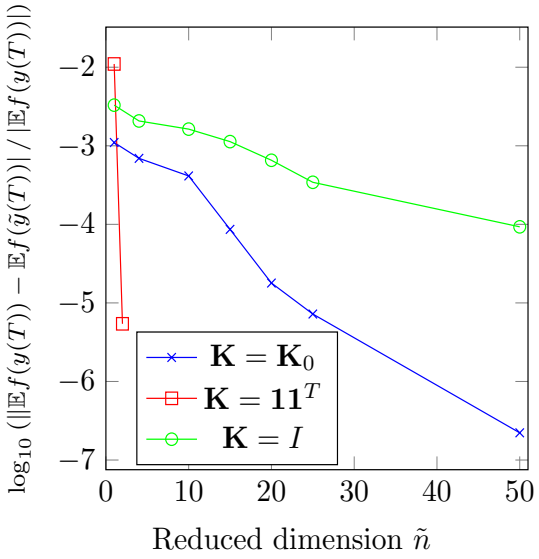

FIGURE 3. Logarithmic relative error expected payoff for $K = \langle \mathbf{1}, x_0\rangle_2$ and different reduced dimensions.

with high-dimensional underlying asset models due to the curse of dimensionality. However, reducing the asset price model in its dimension and subsequently applying the methods in [22, 31] *can be promising and is an interesting topic for further research.*

7. CONCLUSIONS AND OUTLOOK

In this paper, we have shown that model order reduction (MOR) can be an effective technique to construct lower-dimensional surrogate models of large-scale financial models. These surrogate models can be tackled by higher-order computational methods than Monte Carlo simulation. We construct a specific path-wise MOR method, and test it in a multi-dimensional

Heston model. The MOR turns out to work very well for European basket option pricing, especially when the individual assets are strongly correlated (a very realistic scenario). For instance, in a common Doust-type correlation regime, for $n = 100$ assets, a reduced model with dimension $\tilde{n} = 1$ was able to capture the price of an ITM basket option up to a relative error of 10^{-3}, whereas for an OTM option we obtained the same error bound with $\tilde{n} = 10$, which is still a very significant dimension reduction.

Of course, this paper only scratches the surface of applications of MOR in finance. In particular, we identify two very relevant extensions that will be highly beneficial in a financial context. On the one hand, consider that we have restricted ourselves to linear dynamics and essentially linear payoff functions – in the sense that the payoff is assumed to be a non-linear function of a low-dimensional projection of the full price process. Both restrictions can be quite relevant in finance. Allowing non-linear dynamics opens up the possibility of including the stochastic variance process in the model order reduction, as well as having local volatility components. Techniques for MOR in non-linear dynamics have already been developed in the deterministic case [7, 8, 20], and have been extended to stochastic differential equations in some special cases [24]. General non-linear payoff functions are also relevant in finance, think of max-call options. One strategy already available in our framework is to choose C to be the identity matrix.

On the other hand, note that the MOR framework developed in this paper is strong in the probabilistic sense, i.e., we try to approximate the process itself. In many financial application, we are interested in weak approximations, i.e., we want to approximate the distribution of the process. As this is a much weaker concept, even better MOR techniques are conceivable. However, developing an appropriate framework does not seem obvious, and it is unclear how to proceed in this direction.

Appendix A. Resolvent positive operators

Let $(H^n, \langle \cdot, \cdot \rangle_F)$ be the Hilbert space of symmetric $n \times n$ matrices, where $\langle M_1, M_2 \rangle_F := \operatorname{tr}(M_1^T M_2)$ is the Frobenius inner product of two matrices M_1 and M_2. The corresponding norm is defined by $\|M_1\|_F^2 := \langle M_1, M_1 \rangle_F$. Moreover, let H_+^n be the subset of symmetric positive semidefinite matrices. We now define positive and resolvent positive operators on H^n.

Definition A.1. *A linear operator $L : H^n \to H^n$ is called positive if $L(H_+^n) \subset H_+^n$. It is resolvent positive if there is an $\alpha_0 \in \mathbb{R}$ such that for all $\alpha > \alpha_0$ the operator $(\alpha I - L)^{-1}$ is positive.*

The operator $\mathcal{L}(X) := AX + XA^T$ is resolvent positive for $A \in \mathbb{R}^{n \times n}$ which is, e.g., shown in [12]. Moreover, $\Pi(X) := c \sum_{i,j=1}^q N_i X N_j^T k_{ij}$ is positive for $N_i \in \mathbb{R}^{n \times n}$ by [25, Proposition 5.3]. This implies that the generalized Lyapunov operator $\mathcal{L} + \Pi$ is resolvent positive. We now state an equivalent characterization for resolvent positive operators in the following. It can be found in a more general form in [12, 14, 29].

Theorem A.2. *A linear operator $L : H^n \to H^n$ is resolvent positive if and only if $\langle V_1, V_2\rangle_F = 0$ implies $\langle LV_1, V_2\rangle_F \geq 0$ for $V_1, V_2 \in H^n_+$.*

APPENDIX B. PENDING PROOFS

We prove Lemmas 2.1 and 2.3 in the following two subsections.

B.1. **Proof of Lemma 2.1.** We apply Ito's product rule to $x(t)x^T(t)$ and obtain

$$d\left(x(t)x^T(t)\right) = dx(t)x^T(t) + x(t)dx^T(t) + dx(t)dx^T(t).$$

Inserting (1a) yields

(47)
$$\begin{aligned} dx(t)x^T(t) + x(t)dx^T(t) =& Ax(t)x^T(t)dt + \sum_{i=1}^{q} \sqrt{v(t)}N_i x(t)x^T(t)dW_i(t) \\ &+ x(t)x^T(t)A^T dt + \sum_{i=1}^{q} \sqrt{v(t)}x(t)x^T(t)N_i^T dW_i(t). \end{aligned}$$

With (1a) and using that $dW_i(t)dW_j(t) = k_{ij}dt$, we find

$$dx(t)dx^T(t) = v(t)\sum_{i,j=1}^{q} N_i x(t)x^T(t)N_j^T k_{ij}dt. \tag{48}$$

Let e_i denote the ith unit vector. Then, we have

$$k_{ij} = e_i^T \mathbf{K}^{\frac{1}{2}}\mathbf{K}^{\frac{1}{2}}e_j = \sum_{k=1}^{q}\langle \mathbf{K}^{\frac{1}{2}}e_i, e_k\rangle_2 \langle \mathbf{K}^{\frac{1}{2}}e_j, e_k\rangle_2.$$

Using this fact, we obtain that

$$\sum_{i,j=1}^{q} N_i x(t)x^T(t)N_j^T k_{ij} = \sum_{k=1}^{q}\left(\sum_{i=1}^{q} N_i x(t)\langle \mathbf{K}^{\frac{1}{2}}e_i, e_k\rangle_2\right)\left(\sum_{j=1}^{q} N_j x(t)\langle \mathbf{K}^{\frac{1}{2}}e_j, e_k\rangle_2\right)^T \geq 0$$

is a positive semidefinite matrix. Hence, we can enlarge the right-side of (48) by replacing v by its bound c. This leads to

$$dx(t)dx^T(t) \leq c\sum_{i,j=1}^{q} N_i x(t)x^T(t)N_j^T k_{ij}dt. \tag{49}$$

We apply the expected value to both sides of (47) and (49). Since the Ito integrals have mean zero, we have

$$\frac{d}{dt}\mathbb{E}\left[x(t)x^T(t)\right] \leq A\mathbb{E}\left[x(t)x^T(t)\right] + \mathbb{E}\left[x(t)x^T(t)\right]A^T + c\sum_{i,j=1}^{q} N_i\mathbb{E}\left[x(t)x^T(t)\right]N_j^T k_{ij},$$

which concludes the proof.

B.2. **Proof of Lemma 2.3.** We combine (4) with (5) and obtain

$$\dot{Y}(t) \geq L(Y(t)),$$

where $Y := Z - X$. We define the difference function $D(t) := \dot{Y}(t) - L(Y(t)) \geq 0$ and consider the following perturbed differential equation

$$\dot{Y}_\epsilon(t) = L(Y_\epsilon(t)) + D(t) + \epsilon I$$

with parameter $\epsilon \geq 0$ and initial state $Y_\epsilon(0) = Y(0) + \epsilon I$. We see that $Y_0(t) = Y(t)$ for all $t \in [0, T]$ since $Y_0 - Y$ solves (5) with initial condition zero. Since Y_ϵ continuously depends on ϵ and the initial data, we have $\lim_{\epsilon \to 0} Y_\epsilon(t) = Y_0(t) = Y(t)$ for all $t \in [0, T]$.

We want to prove that $Y_\epsilon(t)$ is positive definite for all t and all $\epsilon > 0$. To do so, let us assume the converse, i.e., there is a $\tilde{u} \neq 0$ and a $\tilde{t} > 0$ such that $\tilde{u}^T Y_\epsilon(\tilde{t})\tilde{u} \leq 0$. We know that $f_\epsilon(u, t) := u^T Y_\epsilon(t) u$ is positive at $t = 0$ for all $u \in \mathbb{R}^n \setminus \{0\}$ since $Y(0) \geq 0$ by assumption. Since f_ϵ is non-positive in some point $(\tilde{u}, \tilde{t})$ and due to the continuity of $t \mapsto Y_\epsilon(t)$, there is a point $t_0 \in (0, \tilde{t}]$ for which

$$u_0^T Y_\epsilon(t_0) u_0 = 0 \quad \text{and} \quad u_0^T Y_\epsilon(t) u_0 > 0, \quad t < t_0, \tag{50}$$

for some $u_0 \neq 0$, whereas $u^T Y_\epsilon(t_0) u \geq 0$ for all other $u \in \mathbb{R}^n$. Since L is resolvent positive, $0 = u_0^T Y_\epsilon(t_0) u_0 = \langle Y_\epsilon(t_0), u_0 u_0^T \rangle_F$ implies $0 \leq \langle L(Y_\epsilon(t_0)), u_0 u_0^T \rangle_F = u_0^T L(Y_\epsilon(t_0)) u_0$ by Theorem A.2. Hence, we have

$$u_0^T \dot{Y}_\epsilon(t) u_0 \Big|_{t=t_0} = u_0^T L(Y_\epsilon(t_0)) u_0 + u_0^T D(t_0) u_0 + \epsilon \left\| u_0 \right\|_2^2 > 0.$$

Consequently, we know that there are $t < t_0$ close to t_0 for which $u_0^T Y_\epsilon(t) u_0 < 0$. This contradicts (50) and hence our assumption is wrong such that $Y_\epsilon(t)$ is positive definite for all $t \in [0, T]$ and $\epsilon > 0$. Taking the limit of $\epsilon \to 0$, we obtain $Y(t) \geq 0$ for all $t \in [0, T]$ which concludes the proof.

REFERENCES

[1] A. C. Antoulas. *Approximation of Large-Scale Dynamical Systems*, volume 6 of *Adv. Des. Control*. SIAM Publications, Philadelphia, PA, 2005.

[2] A. C. Antoulas, C. A. Beattie, and S. Gugercin. *Interpolatory methods for model reduction*. SIAM, 2020.

[3] C. Bayer, J. Häppölä, and R. Tempone. Implied stopping rules for American basket options from Markovian projection. *Quantitative Finance*, 19(3):371–390, 2019.

[4] C. A. Beattie, S. Gugercin, and V. Mehrmann. Model reduction for systems with inhomogeneous initial conditions. *Syst. Control Lett.*, 99:99–106, 2017.

[5] S. Becker, C. Hartmann, M. Redmann, and L. Richter. Feedback control theory & Model order reduction for stochastic equations. arXiv preprint 1912.06113, 2019.

[6] P. Benner and T. Damm. Lyapunov equations, energy functionals, and model order reduction of bilinear and stochastic systems. *SIAM J. Control Optim.*, 49(2):686–711, 2011.

[7] P. Benner and P. Goyal. Balanced truncation model order reduction for quadratic-bilinear systems. Technical Report 1705.00160, arXiv, 2017.

[8] P. Benner and P. Goyal. Interpolation-based model order reduction for polynomial parametric systems. e-prints 1904.11891, arXiv, 2019.

[9] P. Benner, M. Ohlberger, A. Cohen, and K. Willcox, editors. *Model Reduction and Approximation: Theory and Algorithms*. SIAM, Philadelphia, PA, 2017.

[10] P. Benner and M. Redmann. Model Reduction for Stochastic Systems. *Stoch PDE: Anal Comp*, 3(3):291–338, 2015.

[11] G. Brunick and S. Shreve. Mimicking an Itô process by a solution of a stochastic differential equation. *The Annals of Applied Probability*, 23(4):1584–1628, 2013.

[12] T. Damm. *Rational Matrix Equations in Stochastic Control.* Lecture Notes in Control and Information Sciences 297. Berlin: Springer, 2004.

[13] P. Doust. Modelling discrete probabilities. Quantitative Anal. Group, RBS, 2007.

[14] L. Elsner. Quasimonotonie und Ungleichungen in halbgeordneten Räumen. *Linear Algebra Appl.*, 8(3):249–261, 1974.

[15] W. Gawronski and J. Juang. Model reduction in limited time and frequency intervals. *Int. J. Syst. Sci.*, 21(2):349–376, 1990.

[16] S. Gugercin and A. C. Antoulas. A survey of model reduction by balanced truncation and some new results. *Internat. J. Control*, 77(8):748–766, 2004.

[17] I. Gyöngy. Mimicking the one-dimensional marginal distributions of processes having an Itô differential. *Probability theory and related fields*, 71(4):501–516, 1986.

[18] B. Hambly, M. Mariapragassam, and C. Reisinger. A forward equation for barrier options under the Brunick & Shreve Markovian projection. *Quantitative Finance*, 16(6):827–838, 2016.

[19] R. Z. Khasminskii. Stochastic stability of differential equations. Monographs and Textbooks on Mechanics of Solids and Fluids. Mechanics: Analysis, 7. Alphen aan den Rijn, The Netherlands; Rockville, Maryland, USA: Sijthoff & Noordhoff., 1980.

[20] B. Krämer and K. Willcox. Balanced Truncation Model Reduction for Lifted Nonlinear Systems. *arXiv preprint: 1907.12084*, 2019.

[21] P. Kürschner. Balanced truncation model order reduction in limited time intervals for large systems. *Adv. Comput. Math.*, pages 1–24, 2018.

[22] F. A. Longstaff and E.S. Schwartz. Valuing American options by simulation: a simple least-squares approach. *Review of Financial Studies*, 14(1):113–147, 2001.

[23] B. C. Moore. Principal component analysis in linear systems: controllability, observability, and model reduction. *IEEE Trans. Autom. Control*, AC-26(1):17–32, 1981.

[24] M. Redmann. Energy estimates and model order reduction for stochastic bilinear systems. *Internat. J. Control*, 2018.

[25] M. Redmann. Type II singular perturbation approximation for linear systems with Lévy noise. *SIAM J. Control Optim.*, 56(3):2120–2158., 2018.

[26] M. Redmann. The missing link between the output and the $\mathcal{H}_2$-norm of bilinear systems. arXiv preprint:1910.14427, 2019.

[27] M. Redmann. An L_T^2-error bound for time-limited balanced truncation. *Syst. Control Lett.*, 136, 2020.

[28] M. Redmann and P. Kürschner. An output error bound for time-limited balanced truncation. *Syst. Control Lett.*, 121:1–6, 2018.

[29] H. Schneider and M. Vidyasagar. Cross-Positive Matrices. *SIAM J. Numer. Anal.*, 7(4):508–519, 1970.

[30] I. Schur. Bemerkungen zur Theorie der beschränkten Bilinearformen mit unendlich vielen Veränderlichen. *Journal für die reine und angewandte Mathematik*, 140:1–28, 1911.

[31] J. Tsitsiklis and B. Van Roy. Regression methods for pricing complex American style options. *IEEE Trans. Neural. Net.*, 12(14):694–703, 2001.