Mathematics of Control, Signals, and Systems (2022) 34:855–881
https://doi.org/10.1007/s00498-022-00328-z

ORIGINAL ARTICLE

Gramian-based model reduction for unstable stochastic systems

Martin Redmann[1] · **Nahid Jamshidi**[1]

Received: 30 September 2021 / Accepted: 10 May 2022 / Published online: 10 July 2022

Abstract

This paper considers large-scale linear stochastic systems representing, e.g., spatially discretized stochastic partial differential equations. Since asymptotic stability can often not be ensured in such a stochastic setting (e.g., due to larger noise), the main focus is on establishing model order reduction (MOR) schemes applicable to unstable systems. MOR is vital to reduce the dimension of the problem in order to lower the enormous computational complexity of for instance sampling methods in high dimensions. In particular, a new type of Gramian-based MOR approach is proposed in this paper that can be used in very general settings. The considered Gramians are constructed to identify dominant subspaces of the stochastic system as pointed out in this work. Moreover, they can be computed via Lyapunov equations. However, covariance information of the underlying systems enters these equations which is not directly available. Therefore, efficient sampling-based methods relying on variance reduction techniques are established to derive the required covariances and hence the Gramians. Alternatively, an ansatz to compute the Gramians by deterministic approximations of covariance functions is investigated. An error bound for the studied MOR methods is proved yielding an a priori criterion for the choice of the reduced system dimension. This bound is new and beneficial even in the deterministic case. The paper is concluded by numerical experiments showing the efficiency of the proposed MOR schemes.

Keywords Model order reduction · Linear stochastic systems · Unstable systems · Stochastic processes · Error analysis

✉ Martin Redmann
martin.redmann@mathematik.uni-halle.de

Nahid Jamshidi
Nahid.Jamshidi@mathematik.uni-halle.de

[1] Institute of Mathematics, Martin Luther University Halle-Wittenberg, Theodor-Lieser-Str. 5, 06120 Halle, Germany

1 Introduction

Let $w = \left(w_1, \ldots, w_q\right)^\top$ be an $\mathbb{R}^q$-valued mean zero Wiener process with covariance matrix $\mathbf{K} = (k_{ij})$, i.e., $\mathbb{E}[w(t)w^\top(t)] = \mathbf{K}t$ 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 $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P}\right)$[1]. 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 controlled linear stochastic differential equation

$$dx(t) = [Ax(t) + Bu(t)]\mathrm{d}t + \sum_{i=1}^{q} N_i x(t)\mathrm{d}w_i(t), \quad x(0) = x_0, \tag{1a}$$

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

where $A, N_i \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$ and $C \in \mathbb{R}^{p\times n}$. The state dimension n is assumed to be large and the quantity of interest y is often low-dimensional, i.e., $p \ll n$, but we also discuss the case of a large p. By $x(t; x_0, u)$, we denote the state in dependence on the initial state x_0 and the control u, for which we assume that it is $(\mathcal{F}_t)_{t\in[0,T]}$-adapted and $\|u\|^2_{L^2_T} := \mathbb{E}\int_0^T \|u(s)\|_2^2\, ds < \infty$ with $\|\cdot\|_2$ representing the Euclidean norm.

The goal is to construct a system with state $\bar{x}$ and quantity of interest $\bar{y}$ having the same structure as (1) but a much smaller state dimension $r \ll n$. At the same time, it is aimed to ensure $y \approx \bar{y}$. Such a reduced-order model (ROM) is particularly beneficial if many evaluations (1) for several controls u are required (e.g., in an optimal control problem) combined with need of generating many samples of y for each individual u. Now, a ROM shall be achieved under very general conditions such as the absence of mean square asymptotic stability, i.e., $\mathbb{E}\,\|x(t; x_0, 0)\|_2^2 \to 0$ (as $t \to \infty$) is not given. Methods involving such a stability condition are intensively studied in the literature [3, 4, 13, 16] since it is often guaranteed if (1a) results from a spatial discretization of a stochastic partial differential equation (SPDE) such as

$$\frac{\partial \mathcal{X}(t,\zeta)}{\partial t} = \Delta\mathcal{X}(t,\zeta) + \mathcal{B}u(t) + \sum_{i=1}^{q} \mathcal{N}_i \mathcal{X}(t,\zeta)\frac{\partial w_i(t)}{\partial t}. \tag{2}$$

We refer to [6] for more details on the theory of such equations. The solution $\mathcal{X}(t, \cdot)$ to the heat equation (2) is viewed as a stochastic process taking values in a Hilbert space and shall be approximated by x. In this context, A can be seen as a discretized version of the Laplacian Δ and B, N_i represent discretizations of the linear bounded operators $\mathcal{B}, \mathcal{N}_i$. Moreover, w_i can be interpreted as Fourier coefficients corresponding to a truncated series of space-time noise. Further explanations on different schemes for a spatial discretization can, for example, be found in [2, 10]. However, even in a setting like in (2), mean square asymptotic stability can be violated since the noise can easily cause instabilities (e.g., if it is sufficiently large).

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

Such a scenario is of interest in this paper. We establish generalizations of balancing related model order reduction (MOR) schemes in order to make them applicable to general systems (1). These MOR methods rely on matrices called Gramians that can be used to identify the dominant subspaces of (1). Based on this characterization of the relevance of different state directions, less important information in the dynamics is removed leading to the desired ROM. This step can be interpreted as an optimization procedure applied to spatially discretized SPDE. In an unstable setting, Gramians need to be defined that generally exist in contrast to previous approaches. We consider generalized time-limited Gramians in this work. Such type of Gramians have been used in deterministic frameworks [9, 11, 15]. Although such an ansatz is beneficial for the setting we want to cover, the analysis of MOR methods based on generalized time-limited Gramians is much more challenging. Furthermore, the question of how to compute these Gramians in practice is very difficult but vital since they are required to derive the ROM.

In this paper, we introduce time-limited Gramian in the stochastic setting studied here. We point out the relation between these Gramians and the dominant subspaces of (1) and show their relation to matrix (differential) equations. Subsequently, we discuss two different MOR techniques based on these Gramians and analyze the respective error. In particular, an error bound is established that allows us to identify situations in which the approaches work well. It is important to mention that this bound is more than just a generalization of the deterministic case [15]. The new type of representation links the truncated Hankel singular values of the system or the truncated eigenvalues of the reachability Gramian, respectively, to the error of the approximation without needing asymptotic stability and is hence beneficial also in unstable settings. Moreover, we discuss different strategies that can be used to compute the proposed Gramians. They are solutions to Lyapunov equations. However, in a time-limited scenario, covariance information at the terminal time enters these Lyapunov equations which is not immediately available. Since direct methods only work in moderate high dimensions, we focus on sampling based approaches to estimate the required covariances. In order to increase the efficiency of such procedures we apply variance reduction methods in this context leading to an efficient way of solving for the time-limited Gramians. Apart from this empirical procedure, a second strategy to approximate covariance functions and hence the Gramians is investigated, where potentially expensive sampling is not required. The paper is concluded by several numerical experiments showing the efficiency of the MOR methods.

2 Gramian-based MOR

2.1 Gramians and characterization of dominant subspaces

Identifying the effective dimensionality of system (1) requires the study of the fundamental solution to the homogeneous stochastic state equation. It is defined as the matrix valued stochastic process Φ solving

$$\Phi(t) = I + \int_0^t A\Phi(s)\mathrm{d}s + \sum_{i=1}^{q} \int_0^t N_i \Phi(s)\mathrm{d}w_i(s), \quad t \in [0, T], \tag{3}$$

where I denotes the identity matrix. Multiplying (3) with x_0 from the right, we obtain the solution to (1a) if $u \equiv 0$. Based on Φ we define two Gramians by

$$P_T := \mathbb{E} \int_0^T \Phi(s)BB^\top \Phi^\top(s)\mathrm{d}s \tag{4}$$

$$Q_T := \mathbb{E} \int_0^T \Phi^\top(s)C^\top C\Phi(s)\mathrm{d}s, \tag{5}$$

where P_T and Q_T are supposed to identify the less relevant states in (1a) and (1b), respectively. P_T and Q_T can be viewed as generalizations of deterministic time-limited Gramians which are obtained by setting $N_i = 0$ for all $i = 1, \dots, q$ resulting in $\Phi(t) = \mathrm{e}^{At}$. MOR schemes based on such Gramians in a deterministic framework are investigated, e.g., in [9, 11, 15]. P_T and Q_T generally exist in contrast to their limits $\lim_{T\to\infty} P_T$ and $\lim_{T\to\infty} Q_T$ which require mean square asymptotic stability. MOR methods based on these limits are, e.g., considered in [3, 4, 13, 16] and are already analyzed in detail. However, the necessary stability condition is often not satisfied in practice.

Let us briefly sketch the relation between P_T and dominant subspaces in (1a) for the case of zero initial data. Suppose that $(p_k)_{k=1,\dots,n}$ is an orthonormal basis of $\mathbb{R}^n$ consisting of eigenvectors of P_T. We can then write the state as

$$x(t; 0, u) = \sum_{k=1}^{n} \langle x(t; 0, u), p_k\rangle_2 \, p_k.$$

Given $x_0 = 0$, the expansion coefficient can be bound from above as follows

$$\sup_{t\in[0,T]} \mathbb{E} \left|\langle x(t, 0, u), p_k\rangle_2\right| \leq \sqrt{\lambda_k} \, \|u\|_{L^2_T} , \tag{6}$$

see [13, Section 3], where λ_k is the eigenvalue corresponding to p_k. If λ_k is small, the same is true for $\langle x(\cdot, 0, u), p_k\rangle_2$ and hence p_k is a less relevant direction that can be neglected. This implies that the eigenspaces of P_T belonging to the small eigenvalues can be removed from the system. On the other hand, we aim to find state directions that have a low impact on the quantity of interest y. We therefore look at the initial state x_0 since it determines the dynamics of the state variable. We expand

$$x_0 = \sum_{k=1}^{n} \langle x_0, q_k\rangle_2 \, q_k,$$

where $(q_k)_{k=1,\dots,n}$ is an orthonormal basis of eigenvectors of Q_T with associated eigenvalues $(\mu_k)_{k=1,\dots,n}$. Using the solution representation of the state variable, we obtain

$$\begin{aligned} y(t; x_0, u) &= C\Phi(t)x_0 + C\int_0^t \Phi(t,s)Bu(s)\mathrm{d}s \\ &= \sum_{k=1}^{n} \langle x_0, q_k\rangle_2 \, C\Phi(t)q_k + C\int_0^t \Phi(t,s)Bu(s)\mathrm{d}s \end{aligned}$$

with $t \in [0, T]$ and $\Phi(t,s) := \Phi(t)\Phi^{-1}(s)$. Consequently, neglecting q_k has a low impact on y if $C\Phi(\cdot)q_k$ is small on $[0, T]$. It now follows that

$$\mathbb{E}\int_0^T \|C\Phi(t)q_k\|_2^2 \,\mathrm{d}t = q_k^\top Q_T q_k = \mu_k, \tag{7}$$

telling us that the eigenspaces of Q_T are unimportant for which the associated eigenvalues μ_k are small. Knowing both the less relevant state directions in (1a) and (1b) from (6) and (7) it is aimed to remove them. This can be done by diagonalizing P_T such that less important variables in (1a) can be easily identified and truncated. Another, but computationally more expensive, approach is based on simultaneously diagonalizing P_T and Q_T which allows to remove more redundant information from the system. Both strategies are discussed in Sect. 2.2.

Below, we point out the relation between the Gramians and linear matrix differential equations. To do so, we introduce two operators $\mathcal{L}_A(X) = AX + XA^\top$ and $\Pi(X) = \sum_{i,j=1}^{q} N_i X N_j^\top k_{ij}$ on the space of symmetric matrices endowed with the Frobenius inner product $\langle\cdot,\cdot\rangle_F$. $\mathcal{L}_A$ is a Lyapunov operator and Π is positive in the sense that $\Pi(X)$ is a positive semidefinite matrix if X is positive semidefinite. The corresponding adjoint operators are $\mathcal{L}_A^*(X) = A^\top X + XA$ and $\Pi^*(X) = \sum_{i,j=1}^{q} N_i^\top X N_j k_{ij}$.

The equations related to P_T and Q_T will be helpful to compute these Gramians that are needed in order to derive the reduced system. By Ito's product rule [12], we can show that $F(t) = \mathbb{E}[\Phi(t)BB^\top\Phi^\top(t)]$, $t \in [0, T]$, solves

$$\dot{F}(t) = \mathcal{L}_A\left(F(t)\right) + \Pi\left(F(t)\right), \quad F(0) = BB^\top. \tag{8}$$

Integrating both sides of (8) yields

$$F(T) - BB^\top = \mathcal{L}_A\left(P_T\right) + \Pi\left(P_T\right), \tag{9}$$

see [7, 13, 14].

Remark 2.1 The generalized Lyapunov operator $\mathcal{L}_A + \Pi$ is linked to the Kronecker matrix

$$\mathcal{K} = A \otimes I + I \otimes A + \sum_{i,j=1}^{q} N_i \otimes N_j k_{ij}, \tag{10}$$

where $\cdot \otimes \cdot$ is the Kronecker product between two matrices. Let $\operatorname{vec}(\cdot)$ be the vectorization of a matrix. Then, it holds that $\operatorname{vec}\left(\left(\mathcal{L}_A + \Pi\right)(X)\right) = \mathcal{K}\operatorname{vec}(X)$.

The link between Q_T and the corresponding matrix equation is established in a different way. We formulate this result in the following proposition.

Proposition 2.2 *Let $C^\top C$ be contained in the eigenspace of the Lyapunov operator $\mathcal{L}_A^* + \Pi^*$. Then, $G(t) = \mathbb{E}[\Phi^\top(t) C^\top C \Phi(t)]$, $t \in [0, T]$, satisfies*

$$\dot{G}(t) = \mathcal{L}_A^*\left(G(t)\right) + \Pi^*\left(G(t)\right), \quad G(0) = C^\top C. \tag{11}$$

Proof Since $C^\top C$ is contained in the eigenspace of the Lyapunov operator, there exist $\alpha_1, \ldots, \alpha_{n^2} \in \mathbb{C}$ such that $C^\top C = \sum_{k=1}^{n^2} \alpha_k \mathcal{V}_k$, where $(\mathcal{V}_k)$ are eigenvectors of $\mathcal{L}_A^* + \Pi^*$ corresponding to the eigenvalues (β_k). Then, we have $\mathbb{E}[\Phi^\top(t) C^\top C \Phi(t)] = \sum_{k=1}^{n^2} \alpha_k \mathbb{E}[\Phi^\top(t) \mathcal{V}_k \Phi(t)]$. Let us apply Ito's product rule, see [12], to $\Phi^\top(t)\mathcal{V}_k\Phi(t)$ resulting in

$$\mathrm{d}\left(\Phi^\top(t)\mathcal{V}_k\Phi(t)\right) = \mathrm{d}\left(\Phi^\top(t)\right)\mathcal{V}_k\Phi(t) + \Phi^\top(t)\mathcal{V}_k\mathrm{d}\left(\Phi(t)\right) + \mathrm{d}\left(\Phi^\top(t)\right)\mathcal{V}_k\mathrm{d}\left(\Phi(t)\right).$$

We insert the stochastic differential of Φ above, compare with (3), leading to

$$\begin{aligned}
\mathrm{d}\left(\Phi^\top(t)\mathcal{V}_k\Phi(t)\right) &= \left(\Phi^\top(t)A^\top \mathrm{d}t + \sum_{i=1}^{q} \Phi^\top(t) N_i^\top \mathrm{d}w_i(t)\right)\mathcal{V}_k\Phi(t) \\
&\quad + \Phi^\top(t)\mathcal{V}_k\left(A\Phi(t)\mathrm{d}t + \sum_{i=1}^{q} N_i\Phi(t)\mathrm{d}w_i(t)\right) \\
&\quad + \Phi^\top(t)\sum_{i,j=1}^{q} N_i^\top\mathcal{V}_k N_j k_{ij}\Phi(t)\mathrm{d}t \\
&= \Phi^\top(t)\left(A^\top\mathcal{V}_k + \mathcal{V}_k A + \sum_{i,j=1}^{q} N_i^\top\mathcal{V}_k N_j k_{ij}\right)\Phi(t)\mathrm{d}t \\
&\quad + \sum_{i=1}^{q}\Phi^\top(t)\left(N_i^\top\mathcal{V}_k + \mathcal{V}_k N_i\right)\Phi(t)\mathrm{d}w_i(t).
\end{aligned}$$

We apply the expected value to both sides of the above identity and exploit that Ito integrals have mean zero (see, e.g., [12]). Hence, we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\Phi^\top(t)\mathcal{V}_k\Phi(t)] = \mathbb{E}[\Phi^\top(t)(\mathcal{L}_A^* + \Pi^*)(\mathcal{V}_k)\Phi(t)] = \beta_k\mathbb{E}[\Phi^\top(t)\mathcal{V}_k\Phi(t)].$$

This implies that $\mathbb{E}[\Phi^\top(t)\mathcal{V}_k\Phi(t)] = \mathrm{e}^{\beta_k t}\mathcal{V}_k$ providing $\mathbb{E}[\Phi^\top(t)C^\top C\Phi(t)] = \sum_{k=1}^{n^2}\alpha_k \mathrm{e}^{\beta_k t}\mathcal{V}_k$. Consequently, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\Phi^\top(t)C^\top C\Phi(t)] = \sum_{k=1}^{n^2}\alpha_k \mathrm{e}^{\beta_k t}\beta_k\mathcal{V}_k = \sum_{i=1}^{n^2}\alpha_k \mathrm{e}^{\beta_k t}(\mathcal{L}_A^* + \Pi^*)(\mathcal{V}_k)$$
$$= (\mathcal{L}_A^* + \Pi^*)(\mathbb{E}[\Phi^\top(t)C^\top C\Phi(t)])$$

using the linearity of $\mathcal{L}_A^* + \Pi^*$. This concludes the proof. □

Remark 2.3 The assumption of Proposition 2.2 is always true if $\mathcal{K}$ is diagonalizable over $\mathbb{C}$ because in that case there is a basis of $\mathbb{C}^{n^2}$ consisting of eigenvectors of $\mathcal{K}^\top$. Hence, $\mathrm{vec}(C^\top C)$ can be spanned by these eigenvectors which are of the form $\mathrm{vec}(\mathcal{V}_k)$ with $\mathcal{V}_k$ being an eigenvector of $\mathcal{L}_A^* + \Pi^*$ providing that $C^\top C$ is in the eigenspaces of this operator. Therefore, from the computational point of view, the assumption of Proposition 2.2 does not restrict the generality since the set of diagonalizable $n^2 \times n^2$ matrices is dense in $\mathbb{C}^{n^2\times n^2}$.

In fact, we can find a stochastic representation of the solution to (11) different from $\mathbb{E}[\Phi^\top(t)C^\top C\Phi(t)]$, $t \in [0, T]$. Introducing the fundamental solution Φ_d by the equation $\Phi_d(t) = I + \int_0^t A^\top\Phi_d(s)\mathrm{d}s + \sum_{i=1}^q \int_0^t N_i^\top\Phi_d(s)\mathrm{d}w_i(s)$, we see that $G(t) = \mathbb{E}[\Phi_d(t)C^\top C\Phi_d^\top(t)]$. This is a direct consequence of the relation between $\mathbb{E}[\Phi(t)BB^\top\Phi^\top(t)]$ and the solution of (8) when (A, B, N_i) is replaced by $(A^\top, C^\top, N_i^\top)$. Therefore, $\mathbb{E}[\Phi_d(t)C^\top C\Phi_d^\top(t)]$, $t \in [0, T]$, solves (11) and hence coincides with $\mathbb{E}[\Phi^\top(t)C^\top C\Phi(t)]$, $t \in [0, T]$, given the assumption of Proposition 2.2.

Generally, we have $\Phi_d(t) \neq \Phi^\top(t)$. In case all matrices $A, N_1, \ldots, N_q$ commute, we know that A and N_i commute with Φ (see, e.g., [14]). Hence, $\Phi_d(t) = \Phi^\top(t)$ which can be seen be transposing (3) and subsequently exploiting the commutative property. This is particularly given in the deterministic case where $N_i = 0$ for all $i = 1, \ldots, q$.

Under the assumption of Proposition 2.2, it holds that

$$G(T) - C^\top C = \mathcal{L}_A^*(Q_T) + \Pi^*(Q_T), \tag{12}$$

exploiting (11). In fact, we need to compute P_T and Q_T within the MOR procedure described later. Lyapunov equations (9) and (12) are used to do so. However, one needs to have access to $F(T)$ and $G(T)$ which are the terminal values of the matrix-differential equations (8) and (11). This is indeed very challenging in a framework, where $n \gg 100$. We will address possible approaches for computing P_T and Q_T for such settings in Sect. 4.

2.2 Reduced-order modeling by transformation of Gramians

In this work, we address MOR techniques that rely on a change of basis. In particular, one seeks for a suitable regular matrix S that defines $x_S(t) = Sx(t)$. Inserting this into (1) yields

$$\mathrm{d}x_S(t) = [A_S x_S(t) + B_S u(t)]\mathrm{d}t + \sum_{i=1}^{q} N_{i,S} x_S(t)\mathrm{d}w_i(t), \quad y(t) = C_S x_S(t), \quad t \in [0, T], \tag{13}$$

where $(A_S, B_S, C_S, N_{i,S}) = (SAS^{-1}, SB, CS^{-1}, SN_iS^{-1})$. System (13) has the same input–output behavior as (1) but the fundamental solution and hence the Gramians are different. The fundamental solution of (13) is $\Phi_S(t) = S\Phi(t)S^{-1}$ which can be observed by multiplying (3) with S from the left and with S^{-1} from the right. Consequently, the new Gramians are

$$P_{T,S} = \mathbb{E}\int_0^T \Phi_S(s)B_S B_S^\top \Phi_S^\top(s)\mathrm{d}s = SP_TS^\top,$$
$$Q_{T,S} = \mathbb{E}\int_0^T \Phi_S^\top(s)C_S^\top C_S\Phi_S(s)\mathrm{d}s = S^{-\top}Q_TS^{-1}.$$

The idea is to diagonalize at least one of these Gramians, since in a system with diagonal Gramians, the orthonormal bases (p_k) and (q_k) are canonical unit vectors (columns of the identity matrix). Thus, unimportant directions can be identified easily by (6) and (7) and are associated with the small diagonal entries of the new Gramians. For the first approach, we set $S = S_1$, where S_1 is part of the eigenvalue decomposition $P_T = S_1^\top \Sigma_T^{(1)} S_1$. This leads to $P_{T,S} = \Sigma_T^{(1)}$ with $\Sigma_T^{(1)}$ being the diagonal matrix of eigenvalues of P_T. Notice that $S^\top = S^{-1}$ holds in this case. If (1a) is mean square asymptotically stable, P_T can be replaced by $\lim_{T\to\infty} P_T$. This method based on the limit is investigated in [16].

The second approach uses $S = S_2$, which leads to $P_{T,S} = Q_{T,S} = \Sigma_T^{(2)}$, where $\Sigma_T^{(2)}$ is the diagonal matrix of the square roots of eigenvalues of P_TQ_T. Those are called Hankel singular values (HSVs). Given $P_T, Q_T > 0$, the transformation S_2 and its inverse are obtained by

$$S_2 = \Sigma_T^{(2)^{-\frac{1}{2}}} U^\top L^\top, \quad S_2^{-1} = KV\Sigma_T^{(2)^{-\frac{1}{2}}}, \tag{14}$$

where the ingredients of (14) are computed by the factorizations $P_T = KK^\top$, $Q_T = LL^\top$ and the singular value decomposition of $K^\top L = V\Sigma_T^{(2)}U^\top$. The same procedure can be conducted for the limits of the Gramians (as $T \to \infty$) if mean square asymptotic stability is given [4]. However, such a stability condition is generally too restrictive in practice. We introduce the matrix

$$\Sigma_T = \mathrm{diag}(\sigma_{T,1}, \ldots, \sigma_{T,n}) = \Sigma_T^{(i)}, \quad i \in \{1, 2\}, \tag{15}$$

as the diagonal matrix of either eigenvalues of P_T or of HSVs of system (1). For $S = S_1$ or $S = S_2$ the coefficients of (13) are partitioned as follows

$$\begin{aligned} A_S &= \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \quad B_S = \begin{pmatrix} B_1 \\ B_2 \end{pmatrix}, \quad C_S = \begin{pmatrix} C_1 & C_2 \end{pmatrix}, \\ N_{i,S} &= \begin{pmatrix} N_{i,11} & N_{i,12} \\ N_{i,21} & N_{i,22} \end{pmatrix}, \quad x_S(t) = \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix}, \quad \Sigma_T = \begin{pmatrix} \Sigma_{T,1} & \\ & \Sigma_{T,2} \end{pmatrix}, \end{aligned} \tag{16}$$

where $x_1(t) \in \mathbb{R}^r$, $A_{11} \in \mathbb{R}^{r \times r}$, $B_1 \in \mathbb{R}^{r \times m}$, $C_1 \in \mathbb{R}^{p \times r}$, $N_{i,11} \in \mathbb{R}^{r \times r}$ and $\Sigma_{T,1} \in \mathbb{R}^{r \times r}$, etc. The variables x_2 are associated with the matrix $\Sigma_{T,2}$ of small diagonal entries of Σ_T and are the less relevant ones. A reduced system is now obtained by truncating the equations of x_2 in (13). Additionally, we set $x_2 \equiv 0$ in the equations for x_1 leading to a reduced system

$$\mathrm{d}\bar{x}(t) = [A_{11}\bar{x}(t) + B_1 u(t)]\mathrm{d}t + \sum_{i=1}^{q} N_{i,11}\bar{x}(t)\mathrm{d}w_i(t), \quad \bar{x}(0) = \bar{x}_0, \tag{17a}$$

$$\bar{y}(t) = C_1\bar{x}(t), \quad t \in [0, T], \tag{17b}$$

approximating (1). Below, we give another interpretation for (17). Let us decompose the transformation

$$S = \begin{pmatrix} W^\top \\ \star \end{pmatrix}, \quad S^{-1} = \begin{pmatrix} V & \star \end{pmatrix} \tag{18}$$

where $W^\top$ and V are the first r rows and columns of S and S^{-1}, respectively. Notice that $W^\top V = I$ and hence $VW^\top$ is a projection. Furthermore, we have $W = V$ if $S = S_1$. Consequently, (17) can be seen as a projection-based model with $A_{11} = W^\top AV$, $B_1 = W^\top B$, $C_1 = CV$ and $N_{i,11} = W^\top N_i V$ which is obtained by the state approximation $x(t) \approx V\bar{x}(t)$. Inserting this approximation into (1) and subsequently multiplying the state equation with $W^\top$ to enforce the remainder term to be zero then results in (17).

3 Output error bound

In this section, we prove a bound for the error between (1) and (17). Below, we assume zero initial conditions, i.e., $x_0 = 0$ and $\bar{x}_0 = 0$. We begin with a general bound following the steps of [4, 13]. The solutions $x(t)$ and $\bar{x}(t)$, $t \in [0, T]$, to (1) and (17) can be expressed using their fundamental matrices $\Phi(t)$ and $\bar{\Phi}(t)$, respectively, see [13]. Therefore, we have

$$x(t; 0, u) = \int_0^t \Phi(t, s) B u(s)\mathrm{d}s, \quad \bar{x}(t; 0, u) = \int_0^t \bar{\Phi}(t, s) B_1 u(s)\mathrm{d}s,$$

where $\Phi(t,s)=\Phi(t)\Phi^{-1}(s)$ and $\bar{\Phi}(t,s)=\bar{\Phi}(t)\bar{\Phi}^{-1}(s)$. Consequently, representations for the outputs are

$$\begin{aligned} y(t) &= Cx(t;0,u) = C\int_0^t \Phi(t,s)Bu(s)\mathrm{d}s,\\ \bar{y}(t) &= C_1\bar{x}(t;0,u) = C_1\int_0^t \bar{\Phi}(t,s)B_1u(s)\mathrm{d}s, \end{aligned} \tag{19}$$

where $t\in[0,T]$. Then, we find

$$\begin{aligned} \mathbb{E}\|y(t)-\bar{y}(t)\|_2 &= \mathbb{E}\left\|C\int_0^t \Phi(t,s)Bu(s)\mathrm{d}s - C_1\int_0^t \bar{\Phi}(t,s)B_1u(s)\mathrm{d}s\right\|_2\\ &\le \mathbb{E}\int_0^t \left\|\left(C\Phi(t,s)B - C_1\bar{\Phi}(t,s)B_1\right)u(s)\right\|_2\mathrm{d}s\\ &\le \mathbb{E}\int_0^t \left\|C\Phi(t,s)B - C_1\bar{\Phi}(t,s)B_1\right\|_F\|u(s)\|_2\mathrm{d}s. \end{aligned} \tag{20}$$

Here, $\|\cdot\|_F$ denotes the Frobenius norm. Using Cauchy's inequality, it holds that

$$\begin{aligned} \mathbb{E}\|y(t)-\bar{y}(t)\|_2 &\le \left(\mathbb{E}\int_0^t \left\|C\Phi(t,s)B - C_1\bar{\Phi}(t,s)B_1\|_F^2\mathrm{d}s\right)^{\frac{1}{2}}\left(\mathbb{E}\int_0^t \|u(s)\|_2^2\mathrm{d}s\right)^{\frac{1}{2}}\\ &= \left(\mathbb{E}\int_0^t \left\|C^e\Phi^e(t,s)B^e\right\|_F^2\mathrm{d}s\right)^{\frac{1}{2}}\left(\mathbb{E}\int_0^t \|u(s)\|_2^2\mathrm{d}s\right)^{\frac{1}{2}}, \end{aligned}$$

where $B^e=\left(\begin{smallmatrix}B\\B_1\end{smallmatrix}\right)$, $C^e=\left(\begin{smallmatrix}C & -C\end{smallmatrix}\right)$ and $\Phi^e=\left(\begin{smallmatrix}\Phi & 0\\ 0 & \bar{\Phi}\end{smallmatrix}\right)$ is the fundamental solution to the system with coefficients $A^e=\left(\begin{smallmatrix}A & 0\\ 0 & A_{11}\end{smallmatrix}\right)$ and $N_i^e=\left(\begin{smallmatrix}N_i & 0\\ 0 & N_{i,11}\end{smallmatrix}\right)$.

Applying the arguments that are used in [4, 13], we know that

$$\mathbb{E}[\Phi^e(t,s)B^eB^{e\top}\Phi^{e\top}(t,s)] = \mathbb{E}[\Phi^e(t-s)B^eB^{e\top}\Phi^{e\top}(t-s)]. \tag{21}$$

For $t\in[0,T]$, the identity in (21) yields

$$\begin{aligned} \mathbb{E}\int_0^t \left\|C^e\Phi^e(t,s)B^e\right\|_F^2\mathrm{d}s &= \mathbb{E}\int_0^t \mathrm{tr}(C^e\Phi^e(t,s)B^eB^{e\top}\Phi^{e\top}(t,s)C^{e\top})\mathrm{d}s\\ &= \mathbb{E}\int_0^t \mathrm{tr}(C^e\Phi^e(s)B^eB^{e\top}\Phi^{e\top}(s)C^{e\top})\mathrm{d}s\\ &\le \mathrm{tr}\left(C^e\int_0^T F^e(s)\mathrm{d}s\, C^{e\top}\right) \end{aligned} \tag{22}$$

with $F^e(t)=\mathbb{E}\left[\Phi^e(t)B^eB^{e\top}\Phi^{e\top}(t)\right]$ exploiting Fubini's theorem as well as the fact that the trace and C^e are linear operators. Since $F(t)=\mathbb{E}\left[\Phi(t)BB^\top\Phi^\top(t)\right]$ is a stochastic representation for equation (8), see Sect. 2.1, F^e satisfies

$$\dot{F}^e(t) = A^e F^e(t) + F^e(t) A^{e\top} + \sum_{i,j=1}^{q} N_i^e F^e(t) N_j^{e\top} k_{ij}, \quad F^e(0) = B^e B^{e\top}, \tag{23}$$

using the same arguments. From (23), it can be seen that the left upper $n \times n$ block of F^e is F which solves (8). On the other hand, the right lower $r \times r$ block $\bar{F}$ and the right upper $n \times r$ block $\tilde{F}$ of F^e satisfy

$$\dot{\bar{F}}(t) = A_{11}\bar{F}(t) + \bar{F}(t)A_{11}^\top + \sum_{i,j=1}^{q} N_{i,11}\bar{F}(t)N_{j,11}^\top k_{ij}, \quad \bar{F}(0) = B_1 B_1^\top, \tag{24}$$

$$\dot{\tilde{F}}(t) = A\tilde{F}(t) + \tilde{F}(t)A_{11}^\top + \sum_{i,j=1}^{q} N_i\tilde{F}(t)N_{j,11}^\top k_{ij}, \quad \tilde{F}(0) = BB_1^\top, \tag{25}$$

with stochastic representations

$$\bar{F}(t) = \mathbb{E}[\bar{\Phi}(t)B_1B_1^\top\bar{\Phi}^\top(t)], \quad \tilde{F}(t) = \mathbb{E}[\Phi(t)BB_1^\top\bar{\Phi}^\top(t)]. \tag{26}$$

Consequently, using (22) with the partition $F^e = \begin{pmatrix} F & \tilde{F} \\ \tilde{F}^\top & \bar{F} \end{pmatrix}$, we find

$$\mathbb{E}\int_0^t \left\| C^e\Phi^e(t,s)B^e \right\|_F^2 \mathrm{d}s \leq \operatorname{tr}\Big(CP_T\, C^\top\Big) + \operatorname{tr}\Big(C_1\bar{P}_T C_1^\top\Big) - 2\operatorname{tr}\Big(C\tilde{P}_T C_1^\top\Big),$$

where $\bar{P}_T = \int_0^T \bar{F}(t)\mathrm{d}t$ and $\tilde{P}_T = \int_0^T \tilde{F}(t)\mathrm{d}t$ solve

$$\bar{F}(T) - B_1B_1^\top = A_{11}\bar{P}_T + \bar{P}_T A_{11}^\top + \sum_{i,j=1}^{q} N_{i,11}\bar{P}_T N_{j,11}^\top k_{ij}, \tag{27}$$

$$\tilde{F}(T) - BB_1^\top = A\tilde{P}_T + \tilde{P}_T A_{11}^\top + \sum_{i,j=1}^{q} N_i\tilde{P}_T N_{j,11}^\top k_{ij}. \tag{28}$$

Summing up, we obtain that

$$\sup_{t\in[0,T]} \mathbb{E}\|y(t) - \bar{y}(t)\|_2 \leq \Big(\operatorname{tr}(CP_TC^\top) + \operatorname{tr}(C_1\bar{P}_TC_1^\top) - 2\operatorname{tr}(C\tilde{P}_TC_1^\top)\Big)^{\frac{1}{2}} \|u\|_{L_T^2}. \tag{29}$$

The bound in (29) is very useful in order to check for the quality of a reduced system. Since P_T has to be computed to obtain (17), the actual cost to determine the bound lies in solving the low-dimensional matrix equations (27) and (28). However, (29) is only an a posteriori estimate which is computed after the reduced order model is derived. Therefore, we discuss the role of $\Sigma_{T,2} = \operatorname{diag}(\sigma_{T,r+1}, \ldots, \sigma_{T,n})$ which is either the matrix of neglected eigenvalues of P_T or HSVs of the system. $\Sigma_{T,2}$ is associated with the truncated state variables x_2 of (13), compare with (16). By (6) and (7), it is already known that such variables x_2 are less relevant if $\sigma_{T,r+1}, \dots, \sigma_{T,n}$ are small. This makes the values σ_i a good a priori criterion for the choice of r. In the following, we want to investigate how the truncated values $\sigma_{T,r+1}, \dots, \sigma_{T,n}$ characterize the error of the approximation. For that reason, we prove an error bound depending on $\Sigma_{T,2}$. As we will see, $\Sigma_{T,2}$ is not the only factor having an impact on the bound that is structurally independent of whether we choose $S = S_1$ or $S = S_2$.

Theorem 3.1 *Let y be the output of (1) and $\bar{y}$ be the one of (17). Suppose that $S = S_1, S_2$, where S_1 is the factor of the eigenvalue decomposition of the Gramian P_T and S_2 is the balancing transformation defined in (14). Using partition (16) of the realization $(A_S, B_S, C_S, N_{i,S})$, we have*

$$\begin{aligned}
&\sup_{t\in[0,T]} \mathbb{E}\|y(t) - \bar{y}(t)\|_2 \\
&\quad\leq \left(\operatorname{tr}\left(\Sigma_{T,2}\left[C_2^\top C_2 + 2A_{12}^\top \tilde{Q}_2 + \sum_{i,j=1}^{q} N_{i,12}^\top\left(2\tilde{Q}\begin{pmatrix} N_{j,12} \\ N_{j,22}\end{pmatrix} - \bar{Q}N_{j,12}\right)k_{ij}\right]\right)\right. \\
&\quad\quad \left. + 2\operatorname{tr}\left(\tilde{Q}\begin{pmatrix}\tilde{F}_1 - F_{11} \\ \tilde{F}_2 - F_{21}\end{pmatrix}\right) + \operatorname{tr}\left(\bar{Q}(F_{11} - \bar{F})\right)\right)^{\frac{1}{2}} \|u\|_{L_T^2},
\end{aligned}$$

where $\bar{Q}$ and $\tilde{Q} = \begin{pmatrix}\tilde{Q}_1 & \tilde{Q}_2\end{pmatrix}$ and are the unique solutions to

$$A_{11}^\top \bar{Q} + \bar{Q}A_{11} + \sum_{i,j=1}^{q} N_{i,11}^\top \bar{Q} N_{j,11} k_{ij} = -C_1^\top C_1, \tag{30}$$

$$A_{11}^\top \tilde{Q} + \tilde{Q}A_S + \sum_{i,j=1}^{q} N_{i,11}^\top \tilde{Q} N_{j,S} k_{ij} = -C_1^\top C_S. \tag{31}$$

Moreover, the above bound involves $F_S(T) := SF(T)S^\top = \left(\begin{smallmatrix} F_{11} & F_{12} \\ F_{21} & F_{22}\end{smallmatrix}\right)$ and $\tilde{F}_S(T) := S\tilde{F}(T) = \left(\begin{smallmatrix}\tilde{F}_1 \\ \tilde{F}_2\end{smallmatrix}\right)$, where $F(T)$, $\bar{F} = \bar{F}(T)$ and $\tilde{F}(T)$ are the terminal values of (8), (24) and (25), respectively.

The terms in the bound of Theorem 3.1 that do not directly depend on $\Sigma_{T,2}$ are related to the covariance error of the dimension reduction at the terminal time T (with $u \equiv 0$). To see this, let V be the matrix introduced in (18). As explained below (18), the state of the reduced system (17) can be interpreted as an approximation of the original state in the subspace spanned by the columns of V. By the stochastic representations of $F(T)$, $\tilde{F}(T)$ and $\bar{F}(T)$ (see above (8) and (26)), we can view $F(T)$ and $\bar{F}(T)$ as covariances of the original and reduced model at time T, whereas $\tilde{F}(T)$ describes the correlations between both systems. Let us now assume that

$$F(T) \approx \tilde{F}(T)V^\top, \tag{32}$$

$$F(T) \approx V\bar{F}(T)V^\top, \tag{33}$$

i.e., the covariance at T is well-approximated in the reduced system. This is, e.g., given if the uncontrolled state is well-approximated in the range of V at time T, i.e., $\Phi(T)B \approx V\bar{\Phi}(T)B_1$. Now, multiplying (32) with S from the left and with W (defined in (18)) from the right, we obtain that $\left(\begin{smallmatrix}\tilde{F}_1 - F_{11}\\ \tilde{F}_2 - F_{21}\end{smallmatrix}\right)$ is small. Multiplying (33) with $W^\top$ from the left and with W from the right provides a low deviation between $F_{11} = W^\top F(T)W$ and $\bar{F}$. Although we additionally have these terms related to the covariance error, looking at $\Sigma_{T,2}$ is still suitable for getting an intuition concerning the error and hence a first idea for the choice of r. This is because a small $\Sigma_{T,2}$ goes along with a small error between $\Phi(T)B$ and its approximation $V\bar{\Phi}(T)B_1$ in the range of V. This observation can be made due to

$$\mathbb{E}\int_0^T \|\left(\Phi(t)B\right)^\top z_T\|_2^2 \mathrm{d}t = z_T^\top P_T z_T = 0,$$

where $z_T \in \ker P_T$. Since $t \mapsto \Phi(t)$ is $\mathbb{P}$-almost surely continuous, we have $\left(\Phi(t)B\right)^\top z_T = 0$ $\mathbb{P}$-almost surely for all $t \in [0, T]$. Choosing $t = T$, we therefore know that the columns of $\Phi(T)B$ are orthogonal to $\ker P_T$. This means that $\Phi(T)B \in \operatorname{im} P_T$ since P_T is symmetric. Hence, there is a matrix Z_T such that

$$\Phi(T)B = P_T Z_T = S^{-1}\Sigma_T S^{-\top} Z_T = \begin{pmatrix} V & \star \end{pmatrix} \begin{pmatrix} \Sigma_{T,1} & \\ & \Sigma_{T,2} \end{pmatrix} \begin{pmatrix} V^\top \\ \star \end{pmatrix} Z_T \approx V\Sigma_{T,1}V^\top Z_T,$$

i.e., the columns of $\Phi(T)B$ lie almost in the span of V if $\Sigma_{T,2}$ is small. Therefore, a good approximation can be expected if one truncates states with associated small values $\sigma_{T,r+1}, \ldots, \sigma_{T,n}$. This can be confirmed by computing the representation in (29) after a reduced order dimension r was chosen based on the values $\sigma_{T,i}$.

Remark 3.2 Notice that the covariance $F(T)$ vanishes in the limit as $T \to \infty$ if (1) is mean square asymptotically stable. In this context, the deviations in (32) and (33) can be expected to be small for sufficiently large T since the covariance error disappears at ∞. If the system is unstable, we have $\|F(T)\| \to \infty$ as $T \to \infty$. In this case, the covariance error might be large and dominant if T is very large such that the approximation quality is lower. The role of T is additionally discussed in Sect. 5.

We are now ready to prove the error bound in the following:

Proof of Theorem 3.1 Since $S = S_1, S_2$ diagonalizes P_T, we have

$$A_S\Sigma_T + \Sigma_T A_S^\top + \sum_{i,j=1}^{q} N_{i,S}\Sigma_T N_{j,S}^\top k_{ij} = -B_S B_S^\top + F_S(T). \tag{34}$$

We set $\tilde{Y}_T := S\tilde{P}_T$ and obtain the corresponding equation by multiplying (28) with S from the left resulting in

$$A_S\tilde{Y}_T + \tilde{Y}_T A_{11}^\top + \sum_{i,j=1}^{q} N_{i,S}\tilde{Y}_T N_{j,11}^\top k_{ij} = -B_S B_1^\top + \tilde{F}_S(T). \tag{35}$$

Now, we analyze the trace expression $\epsilon^2 := (\operatorname{tr}(CP_TC^\top) + \operatorname{tr}(C_1\bar{P}_TC_1^\top) - 2\operatorname{tr}(C\tilde{P}_TC_1^\top))$ in (29). We see that

$$\begin{aligned}\epsilon^2 &= \left(\operatorname{tr}(C_S\Sigma_TC_S^\top) + \operatorname{tr}(C_1\bar{P}_TC_1^\top) - 2\operatorname{tr}(C_S\tilde{Y}_TC_1^\top)\right)\\ &= \left(\operatorname{tr}(C_1\Sigma_{T,1}C_1^\top) + \operatorname{tr}(C_2\Sigma_{T,2}C_2^\top) + \operatorname{tr}(C_1\bar{P}_TC_1^\top) - 2\operatorname{tr}(C_S\tilde{Y}_TC_1^\top)\right).\end{aligned} \tag{36}$$

Exploiting (31) yields

$$\begin{aligned}-\operatorname{tr}(C_S\tilde{Y}_TC_1^\top) = -\operatorname{tr}(\tilde{Y}_TC_1^\top C_S) &= \operatorname{tr}\left(\tilde{Y}_T\left[A_{11}^\top\tilde{Q} + \tilde{Q}A_S + \sum_{i,j=1}^{q} N_{i,11}^\top\tilde{Q}N_{j,S}k_{ij}\right]\right)\\ &= \operatorname{tr}\left(\tilde{Q}\left[A_S\tilde{Y}_T + \tilde{Y}_TA_{11}^\top + \sum_{i,j=1}^{q} N_{i,S}\tilde{Y}_TN_{j,11}^\top k_{ij}\right]\right).\end{aligned}$$

Comparing (31) and (35), we find that

$$-\operatorname{tr}(C_S\tilde{Y}_TC_1^\top) = -\operatorname{tr}(\tilde{Q}B_SB_1^\top) + \operatorname{tr}(\tilde{Q}\tilde{F}_S(T)). \tag{37}$$

Using the partition in (16), the first r columns of (34) are

$$\begin{aligned}&\begin{pmatrix}A_{11}\\A_{21}\end{pmatrix}\Sigma_{T,1} + \begin{pmatrix}\Sigma_{T,1}A_{11}^\top\\\Sigma_{T,2}A_{12}^\top\end{pmatrix} + \sum_{i,j=1}^{q}\left(\begin{pmatrix}N_{i,11}\\N_{i,21}\end{pmatrix}\Sigma_{T,1}N_{j,11}^\top + \begin{pmatrix}N_{i,12}\\N_{i,22}\end{pmatrix}\Sigma_{T,2}N_{j,12}^\top\right)k_{ij}\\ &= -B_SB_1^\top + \begin{pmatrix}F_{11}\\F_{21}\end{pmatrix}.\end{aligned} \tag{38}$$

We insert (38) into (37) and obtain

$$\begin{aligned}-\operatorname{tr}(C_S\tilde{Y}_TC_1^\top) &= \operatorname{tr}\left(\tilde{Q}\begin{pmatrix}\tilde{F}_1 - F_{11}\\\tilde{F}_2 - F_{21}\end{pmatrix}\right) + \operatorname{tr}\left(\tilde{Q}\left[\begin{pmatrix}A_{11}\\A_{21}\end{pmatrix}\Sigma_{T,1} + \begin{pmatrix}\Sigma_{T,1}A_{11}^\top\\\Sigma_{T,2}A_{12}^\top\end{pmatrix}\right.\right.\\ &\quad\left.\left. + \sum_{i,j=1}^{q}\left(\begin{pmatrix}N_{i,11}\\N_{i,21}\end{pmatrix}\Sigma_{T,1}N_{j,11}^\top + \begin{pmatrix}N_{i,12}\\N_{i,22}\end{pmatrix}\Sigma_{T,2}N_{j,12}^\top\right)k_{ij}\right]\right)\\ &= \operatorname{tr}\left(\tilde{Q}\begin{pmatrix}\tilde{F}_1 - F_{11}\\\tilde{F}_2 - F_{21}\end{pmatrix}\right) + \operatorname{tr}\left(\Sigma_{T,2}\left[A_{12}^\top\tilde{Q}_2 + \sum_{i,j=1}^{q} N_{i,12}^\top\tilde{Q}\begin{pmatrix}N_{j,12}\\N_{j,22}\end{pmatrix}k_{ij}\right]\right)\\ &\quad + \operatorname{tr}\left(\Sigma_{T,1}\left[\tilde{Q}\begin{pmatrix}A_{11}\\A_{21}\end{pmatrix} + A_{11}^\top\tilde{Q}_1 + \sum_{i,j=1}^{q} N_{i,11}^\top\tilde{Q}\begin{pmatrix}N_{j,11}\\N_{j,21}\end{pmatrix}k_{ij}\right]\right).\end{aligned}$$

Using the partition of the balanced realization in (16), we observe that the last term of above equation is the first r columns of (31). So, we can say that

$$\begin{aligned} -\mathrm{tr}(C_S\tilde{Y}_T C_1^\top) &= \mathrm{tr}\left(\tilde{Q}\begin{pmatrix}\tilde{F}_1 - F_{11}\\ \tilde{F}_2 - F_{21}\end{pmatrix}\right) + \mathrm{tr}\left(\Sigma_{T,2}\left[A_{12}^\top\tilde{Q}_2 + \sum_{i,j=1}^{q} N_{i,12}^\top\tilde{Q}\begin{pmatrix}N_{j,12}\\ N_{j,22}\end{pmatrix}k_{ij}\right]\right)\\ &\quad - \mathrm{tr}(\Sigma_{T,1}C_1^\top C_1). \end{aligned} \tag{39}$$

Inserting (39) into (36), we have

$$\begin{aligned} \epsilon^2 &= \mathrm{tr}\left(\Sigma_{T,2}\left[C_2^\top C_2 + 2A_{12}^\top\tilde{Q}_2 + 2\sum_{i,j=1}^{q} N_{i,12}^\top\tilde{Q}\begin{pmatrix}N_{j,12}\\ N_{j,22}\end{pmatrix}k_{ij}\right]\right)\\ &\quad + 2\mathrm{tr}\left(\tilde{Q}\begin{pmatrix}\tilde{F}_1 - F_{11}\\ \tilde{F}_2 - F_{21}\end{pmatrix}\right) + \mathrm{tr}\left((\bar{P}_T - \Sigma_{T,1})C_1^\top C_1\right). \end{aligned} \tag{40}$$

Equation (30) now yields

$$\begin{aligned} &\mathrm{tr}\left((\bar{P}_T - \Sigma_{T,1})C_1^\top C_1\right)\\ &\quad = -\mathrm{tr}\left(\bar{Q}\left[A_{11}(\bar{P}_T - \Sigma_{T,1}) + (\bar{P}_T - \Sigma_{T,1})A_{11}^\top + \sum_{i,j=1}^{q} N_{i,11}(\bar{P}_T - \Sigma_{T,1})N_{j,11}^\top k_{ij}\right)\right] \end{aligned}$$

The combination of (27) and the left upper block of (34) gives

$$\begin{aligned} &A_{11}(\bar{P}_T - \Sigma_{T,1}) + (\bar{P}_T - \Sigma_{T,1})A_{11}^\top + \sum_{i,j=1}^{q} N_{i,11}(\bar{P}_T - \Sigma_{T,1})N_{j,11}^\top k_{ij}\\ &\quad = \sum_{i,j=1}^{q} N_{i,12}\Sigma_{T,2}N_{j,12}^\top k_{ij} + (\bar{F} - F_{11}). \end{aligned}$$

Consequently, we have

$$\mathrm{tr}\left((\bar{P}_T - \Sigma_{T,1})C_1^\top C_1\right) = -\mathrm{tr}\left(\Sigma_{T,2}\left[\sum_{i,j=1}^{q} N_{i,12}^\top\bar{Q}N_{j,12}k_{ij}\right]\right) + \mathrm{tr}\left(\bar{Q}(F_{11} - \bar{F})\right).$$

So, we obtain that

$$\begin{aligned} \epsilon^2 &= \mathrm{tr}\left(\Sigma_{T,2}\left[C_2^\top C_2 + 2A_{12}^\top\tilde{Q}_2 + \sum_{i,j=1}^{q} N_{i,12}^\top\left(2\tilde{Q}\begin{pmatrix}N_{j,12}\\ N_{j,22}\end{pmatrix} - \bar{Q}N_{j,12}\right)k_{ij}\right]\right)\\ &\quad + 2\mathrm{tr}\left(\tilde{Q}\begin{pmatrix}\tilde{F}_1 - F_{11}\\ \tilde{F}_2 - F_{21}\end{pmatrix}\right) + \mathrm{tr}\left(\bar{Q}(F_{11} - \bar{F})\right), \end{aligned}$$

which concludes the proof of this theorem. □

Notice that the estimate in Theorem 3.1 is also beneficial if $N_i = 0$ for all $i = 1, \dots, q$, since it improves the deterministic bound [15] in the sense that we can generally deduce the relation between the truncated HSVs and the actual approximation error here. It is important to notice that, in the deterministic case, "improvement" is not meant in terms of accuracy. The error bound representation in [15] just has the drawback that it allows to make similar conclusions only if the underlying system is asymptotically stable. Moreover, the result of Theorem 3.1 is a generalization of the bounds for mean square asymptotically stable stochastic systems [4, 16], where the covariance related terms vanish as $T \to \infty$.

4 Computation of Gramians

In this section, we discuss how to compute P_T and Q_T which allow us to identify redundant information in the system. These matrices are solutions of Lyapunov equations (9) and (12) with left-hand sides depending on $F(T)$ and $G(T)$, respectively. Given $F(T)$ and $G(T)$ it is therefore required to solve generalized Lyapunov equations

$$L = \mathcal{L}_A(X) + \Pi(X) \tag{41}$$

efficiently, where L is a symmetric matrix of suitable dimension. According to Remark 2.1 this can be done by vectorization, i.e., one can try to solve $\text{vec}\,(L) = \mathcal{K}\text{vec}(X)$ with the Kronecker matrix $\mathcal{K}$ defined in (10). Since $\mathcal{K}$ is of order n^2, the complexity of deriving $\text{vec}(X)$ from this linear system of equations is $\mathcal{O}(n^6)$ making this procedure infeasible for $n \gg 100$.

However, more efficient techniques have been developed in order to solve (41), see, e.g., [8], where a sequence of standard Lyapunov equations ($\Pi = 0$) is solved to find X. Such standard Lyapunov equations can either be tackled by direct methods, such as Bartels-Stewart [1], which cost $\mathcal{O}(n^3)$ operations, or by iterative methods such as ADI or Krylov subspace methods [17], which have a much smaller complexity than the Bartels–Stewart algorithm, in particular, when the left-hand side is of low rank or structured (complexity of $\mathcal{O}(n^2)$ or less).

Solving for P_T and Q_T now relies on having access to $F(T)$ and $G(T)$ which are the terminal values of the matrix-differential equations (8) and (11). The remainder of this section will deal with strategies to compute these terminal values.

4.1 Exact methods

One solution to overcome the issue of unknown $F(T)$ and $G(T)$ is to use vectorizations of (8) and (11) for dimensions n of a few hundreds. If we define $f(t) := \text{vec}(F(t))$ and $g(t) = \text{vec}(G(t))$, then

$$\dot{f}(t) = \mathcal{K}f(t), \quad f(0) = \text{vec}(BB^\top), \quad \dot{g}(t) = \mathcal{K}^\top g(t), \quad g(0) = \text{vec}(C^\top C),$$

where $\mathcal{K}$ is defined in (10). Therefore, obtaining $F(T)$ and $G(T)$ relies on the efficient computation of a matrix exponential, since

$$f(T) = \mathrm{e}^{\mathcal{K}T}\operatorname{vec}(BB^\top), \quad g(T) = \mathrm{e}^{\mathcal{K}^\top T}\operatorname{vec}(C^\top C).$$

One can find a discussion on how to determine a matrix exponential efficiently in [11] and references therein. Alternatively, one might think of discretizing the matrix differential equations (8) and (11) to find an approximation of $F(T)$ and $G(T)$. However, as stated above, these equations are equivalent to ordinary differential equations of order n^2. Solving such extremely large scale systems is usually not feasible. In addition, only implicit schemes would allow for a reasonable step size in the discretization making the problem even more complex. For that reason, we discuss more suitable numerical approximations in the following.

4.2 Sampling-based approaches

We aim to derive an approximation of the terminal value $F(T) = \mathbb{E}[\Phi(T)BB^\top\Phi^\top(T)]$ of (8) by different stochastic representations. This alternative approach is required since computing $\mathrm{e}^{\mathcal{K}T}$ is not feasible if $n \gg 100$ knowing that $\mathcal{K} \in \mathbb{R}^{n^2\times n^2}$. Therefore, we discuss sampling-based approaches in the following. Let $\Phi^i(T)$, $i \in \{1, \dots, M\}$, be i.i.d. copies of $\Phi(T)$. Then, we have $\frac{1}{M}\sum_{i=1}^M \Phi^i(T)BB^\top\Phi^i(T)^\top \approx F(T)$ if M is sufficiently large. This requires to sample the random variable $\Phi(T)B$ possibly many times. $\Phi(T)B$ is the terminal value of the stochastic differential equation

$$\mathrm{d}x_B(t) = Ax_B(t)\mathrm{d}t + \sum_{i=1}^q N_i x_B(t)\mathrm{d}w_i(t), \quad x_B(0) = B, \tag{42}$$

with $x_B(t) \in \mathbb{R}^{n\times m}$. System (42) can be seen as a matrix-valued homogeneous version of (1a) ($u \equiv 0$) with initial state B. If (1) needs to be evaluated for many different controls u and additionally a large number of samples are required for each fixed u, it even pays off to generate many samples of the solution to (42). In particular, this is true if the number of columns of B is low. However, we want to avoid evaluating (42) too often. The number of samples M required for a good estimate of $F(T)$ depends on the variance of $\Phi(T)BB^\top\Phi^\top(T)$. Therefore, we want to reduce the variance by finding a better stochastic representation than $\mathbb{E}[\Phi(T)BB^\top\Phi^\top(T)]$. In the spirit of variance reduction techniques, we find the zero variance unbiased estimator first. To do so, we apply Ito's product rule (see, e.g., [12]) to obtain

$$\begin{aligned}\mathrm{d}\left(x_B(t)x_B^\top(t)\right) &= \mathrm{d}\left(x_B(t)\right)x_B^\top(t) + x_B(t)\mathrm{d}\left(x_B^\top(t)\right) + \mathrm{d}\left(x_B(t)\right)\mathrm{d}\left(x_B^\top(t)\right)\\ &= \left(Ax_B(t)\mathrm{d}t + \sum_{i=1}^q N_i x_B(t)\mathrm{d}w_i(t)\right)x_B^\top(t)\end{aligned}$$

$$
\begin{aligned}
&+ x_B(t)\left(x_B^\top(t)A^\top \mathrm{d}t + \sum_{i=1}^q x_B^\top(t)N_i^\top \mathrm{d}w_i(t)\right)\\
&+ \sum_{i,j=1}^q N_i x_B(t)x_B^\top(t)N_j^\top k_{ij}\mathrm{d}t\\
&= (\mathcal{L}_A + \Pi)\left(x_B(t)x_B^\top(t)\right)\mathrm{d}t + \sum_{i=1}^q \mathcal{L}_{N_i}\left(x_B(t)x_B^\top(t)\right)\mathrm{d}w_i(t).
\end{aligned}
$$

This stochastic differential is now exploited to find

$$
\begin{aligned}
d\left(\mathrm{e}^{\mathcal{K}(T-t)}\mathrm{vec}(x_B(t)x_B^\top(t))\right) &= -\mathrm{e}^{\mathcal{K}(T-t)}\mathcal{K}\mathrm{vec}(x_B(t)x_B^\top(t))\mathrm{d}t + \mathrm{e}^{\mathcal{K}(T-t)}d\\
&\quad \left(\mathrm{vec}(x_B(t)x_B^\top(t))\right)\\
&= \sum_{i=1}^q \mathrm{e}^{\mathcal{K}(T-t)}\mathrm{vec}\left(\mathcal{L}_{N_i}\left(x_B(t)x_B^\top(t)\right)\right)\mathrm{d}w_i(t)
\end{aligned}
$$

using that $\mathrm{vec}\left((\mathcal{L}_A + \Pi)\left(x_B(t)x_B^\top(t)\right)\right) = \mathcal{K}\mathrm{vec}(x_B(t)x_B^\top(t))$. Hence, we have

$$
\mathrm{vec}\left(x_B(T)x_B^\top(T)\right) = \mathrm{e}^{\mathcal{K}T}\mathrm{vec}(BB^\top) + \sum_{i=1}^q \int_0^T \mathrm{e}^{\mathcal{K}(T-t)}\mathrm{vec}\left(\mathcal{L}_{N_i}\left(x_B(t)x_B^\top(t)\right)\right)\mathrm{d}w_i(t).
$$

Devectorizing this equation yields

$$
F(T) = x_B(T)x_B^\top(T) - \sum_{i=1}^q \int_0^T F\left(T-t, \mathcal{L}_{N_i}\left(x_B(t)x_B^\top(t)\right)\right)\mathrm{d}w_i(t), \tag{43}
$$

where the second argument in F represents the initial condition of (8). The right-hand side of (43) now is unbiased zero variance estimator of $F(T)$. However, this estimator depends on F which is not available. Therefore, given a symmetric matrix X_0, we approximate $F(t, X_0)$ by a computable matrix function $\mathcal{F}(t, X_0)$ that we specify later. This leads to the unbiased estimator

$$
E_{\mathcal{F}}(T) := x_B(T)x_B^\top(T) - \sum_{i=1}^q \int_0^T \mathcal{F}\left(T-t, \mathcal{L}_{N_i}\left(x_B(t)x_B^\top(t)\right)\right)\mathrm{d}w_i(t) \tag{44}
$$

for $F(T)$. The hope is that a few samples of $E_{\mathcal{F}}(T)$ can give an accurate approximation of $F(T)$. Of course, $E_{\mathcal{F}}(T)$ can only be simulated by further discretizing the above Ito integrals, e.g., by a Riemann–Stieltjes sum approximation. The variance of $E_{\mathcal{F}}(T)$ is

$$
\mathbb{E}\left\| E_{\mathcal{F}}(T) - F(T)\right\|_F^2 = \mathbb{E}\left\| \sum_{i=1}^q \int_0^T F\left(T-t, X_i(t)\right) - \mathcal{F}\left(T-t, X_i(t)\right)\mathrm{d}w_i(t)\right\|_F^2
$$

$$= \sum_{i,j=1}^{q} \mathbb{E} \int_0^T \Big\langle F\left(T-t, X_i(t)\right) - \mathcal{F}\left(T-t, X_i(t)\right), \\ F\left(T-t, X_j(t)\right) - \mathcal{F}\left(T-t, X_j(t)\right) \Big\rangle_F k_{ij} \mathrm{d}t$$

setting $X_i(t) = N_i x_B(t) x_B^\top(t) + x_B(t) x_B^\top(t) N_i^\top$ and exploiting Ito's isometry, see [12]. Consequently, the benefit of the variance reduction depends on the difference $F(t, X_0) - \mathcal{F}(t, X_0)$.

We conclude this section by discussing suitable approximations $\mathcal{F}(t, X_0)$ of $F(t, X_0)$. For that reason, we establish the following theorem.

Theorem 4.1 *Let $F(t, X_0)$, $t \in [0, T]$, be the solution to*

$$\dot{F}(t) = \mathcal{L}_A\left(F(t)\right) + \Pi\left(F(t)\right), \quad F(0) = X_0,$$

where the initial data X_0 is a symmetric matrix. Then, there exist constants $\underline{c}$ and $\overline{c}$ such that

$$e^{At} X_0 e^{A^\top t} + \underline{c} \int_0^t e^{As} \Pi\left(I\right) e^{A^\top s} \mathrm{d}s \leq F(t) \leq e^{At} X_0 e^{A^\top t} + \overline{c} \int_0^t e^{As} \Pi\left(I\right) e^{A^\top s} \mathrm{d}s.$$

Proof Exploiting the product rule, it can be seen that F is implicitly given by

$$F(t) = \mathrm{e}^{At} X_0 \mathrm{e}^{A^\top t} + \int_0^t \mathrm{e}^{A(t-s)} \Pi\left(F(s)\right) \mathrm{e}^{A^\top (t-s)} ds. \tag{45}$$

The solution $t \mapsto F(t)$ is continuous and $F(t)$ is a symmetric matrix for all $t \in [0, T]$. Consequently, exploiting [5, Corollary VI.1.6], there exist continuous and real functions $\lambda_1, \ldots, \lambda_n$ such that $\lambda_1(t), \ldots, \lambda_n(t)$ represent the eigenvalues of $F(t)$ for each fixed t. We now define continuous functions by $\underline{\lambda} := \min\{\lambda_1, \ldots, \lambda_n\}$ and $\overline{\lambda} := \max\{\lambda_1, \ldots, \lambda_n\}$. Symmetric matrices can be estimated from below and above by their smallest and largest eigenvalue, respectively, leading to $\underline{\lambda}(t) I \leq F(t) \leq \overline{\lambda}(t) I$. Therefore, given an arbitrary vector in $v \in \mathbb{R}^n$, we have

$$\begin{aligned} v^\top \Pi\left(F(t)\right) v &= \sum_{i,j=1}^{q} (N_i v)^\top F(t) N_j v k_{ij} = \sum_{i,j=1}^{q} (N_i v)^\top F(t) N_j v e_i^\top \mathbf{K}^{\frac{1}{2}} \mathbf{K}^{\frac{1}{2}} e_j \\ &= \sum_{i,j=1}^{q} (N_i v)^\top F(t) N_j v \sum_{k=1}^{q} \langle \mathbf{K}^{\frac{1}{2}} e_i, e_k \rangle_2 \langle \mathbf{K}^{\frac{1}{2}} e_j, e_k \rangle_2 \\ &= \sum_{k=1}^{q} \Bigg(\sum_{i=1}^{q} N_i v \langle \mathbf{K}^{\frac{1}{2}} e_i, e_k \rangle_2 \Bigg)^\top F(t) \underbrace{\Bigg(\sum_{j=1}^{q} N_j v \langle \mathbf{K}^{\frac{1}{2}} e_j, e_k \rangle_2 \Bigg)}_{=: v_k} \\ &\begin{cases} \leq \overline{\lambda}(t) \sum_{k=1}^{q} v_k^\top I v_k \\ \geq \underline{\lambda}(t) \sum_{k=1}^{q} v_k^\top I v_k \end{cases} \end{aligned}$$

resulting in $\underline{\lambda}(t)\Pi(I) \leq \Pi(F(t)) \leq \overline{\lambda}(t)\Pi(I)$, where e_i is the canonical basis of $\mathbb{R}^q$. Since $\underline{\lambda}, \overline{\lambda}$ are continuous on $[0, T]$, they can be bounded from below and above by some suitable constants. Applying this to (45), we obtain the result by substitution. □

Of course, the constants in Theorem 4.1 are generally unknown. However, this result gives us the intuition that $F(t, X_0)$ can be approximated by

$$\mathcal{F}(t, X_0) = \mathrm{e}^{At} X_0 \mathrm{e}^{A^\top t} + c \int_0^t \mathrm{e}^{As} \Pi(I)\, \mathrm{e}^{A^\top s} ds, \tag{46}$$

where $c \in [\underline{c}, \overline{c}]$ is a real number. From the proof of Theorem 4.1, we further know that $\underline{c}, \overline{c} \geq 0$ if X_0 is positive semidefinite. We cannot generally expect a reduction of the variance for all choices of c. However, a good candidate will reduce the computational complexity. A general strategy how to find such a candidate is an interesting question for future research.

Remark 4.2 Besides generating (a few) samples of x_B from (42), we require the matrix exponentials e^{At_i} on a grid $0 = t_0 < t_1 < \cdots < t_{n_g} = T$ to determine the estimator (44) with $\mathcal{F}$ as in (46). Here, n_g is the number of grid points when discretizing the Ito integral in (44). If the points t_i are equidistant with step size h, one first computes e^{Ah}. The other exponentials are then powers of e^{Ah} such that a certain number of matrix multiplications (depending on n_g) have to be conducted.

The Gramian Q_T can be computed from (12) requiring to determine $G(T)$. According to Remark 2.3, we know that $G(T) = \mathbb{E}[x_C(T)x_C^\top(T)]$, where

$$\mathrm{d}x_C(t) = A^\top x_C(t)\mathrm{d}t + \sum_{i=1}^q N_i^\top x_C(t)\mathrm{d}w_i(t), \quad x_C(0) = C^\top,$$

with $x_C(t) \in \mathbb{R}^{n\times p}$. Exploiting the above consideration regarding $F(T)$, we can see that

$$E_{\mathcal{G}}(T) := x_C(T)x_C^\top(T) - \sum_{i=1}^q \int_0^T \mathcal{G}\left(T - t, \mathcal{L}_{N_i}^*\left(x_C(t)x_C^\top(t)\right)\right) \mathrm{d}w_i(t) \tag{47}$$

is a possible unbiased estimator for $G(T)$. The approximation $\mathcal{G}$ of G can be chosen as in (46) replacing $(A, N_i) \mapsto (A^\top, N_i^\top)$.

4.3 Gramians based on deterministic approximations of *F*(*T*) and *G*(*T*)

Based on Theorem 4.1, an estimation of $F(T)$ (and also $G(T)$) is given in (46). Instead of using these approximations in a variance reduction procedure like in Sect. 4.2, we exploit it directly in (9) and (12). This leads to matrices $\mathcal{P}_T$ and $\mathcal{Q}_T$ solving

$$\mathcal{F}(T, BB^\top) - BB^\top = \mathcal{L}_A(\mathcal{P}_T) + \Pi(\mathcal{P}_T), \tag{48}$$

$$\mathcal{G}(T, C^\top C) - C^\top C = \mathcal{L}_A^* \left(\mathcal{Q}_T\right) + \Pi^* \left(\mathcal{Q}_T\right), \tag{49}$$

where the left hand sides are defined by

$$\mathcal{F}(T, BB^\top) = \mathrm{e}^{AT} BB^\top \mathrm{e}^{A^\top T} + c_F \int_0^T \mathrm{e}^{As} \Pi\left(I\right) \mathrm{e}^{A^\top s} \mathrm{d}s, \quad c_F \in \mathbb{R}, \tag{50}$$

$$\mathcal{G}(T, C^\top C) = \mathrm{e}^{A^\top T} C^\top C \mathrm{e}^{AT} + c_G \int_0^T \mathrm{e}^{A^\top s} \Pi^*\left(I\right) \mathrm{e}^{As} \mathrm{d}s, \quad c_G \in \mathbb{R}. \tag{51}$$

Certainly, the choice of the constants c_F and c_G determine how well P_T and Q_T are approximated by $\mathcal{P}_T$ and $\mathcal{Q}_T$, e.g., in terms of the characterization of the respective dominant subspaces of system (1). Notice that for $N_i = 0$, $\mathcal{F}(T, BB^\top)$ and $\mathcal{G}(T, C^\top C)$ yield the exact values for $F(T, BB^\top)$ and $G(T, C^\top C)$. At this point, it is important to mention that the Gramian approximation of this section is computationally less complex than the one in Sect. 4.2. First of all, we do not need to sample from (42) and secondly no Ito integral as in (44) has to be discretized. Calculating $\mathcal{F}$ and $\mathcal{G}$ might also require to compute matrix exponentials on a partition of $[0, T]$, compare with Remark 4.2. However, less grid points than for the sampled Gramians of Sect. 4.2 have to be considered since an ordinary integral can be discretized with a larger step size compared to an Ito integral. Alternatively, the integrals in (50) and (51) can also be determined without a discretization since it holds that

$$\mathcal{L}_A\left(\int_0^T \mathrm{e}^{As} \Pi\left(I\right) \mathrm{e}^{A^\top s} ds\right) = -\Pi\left(I\right) + \mathrm{e}^{AT} \Pi\left(I\right) \mathrm{e}^{A^\top T},$$

$$\mathcal{L}_A^*\left(\int_0^T \mathrm{e}^{A^\top s} \Pi^*\left(I\right) \mathrm{e}^{As} ds\right) = -\Pi^*\left(I\right) + \mathrm{e}^{A^\top T} \Pi^*\left(I\right) \mathrm{e}^{AT}.$$

This approach has the advantage that only the matrix exponential e^{AT} at the terminal time is needed.

5 Numerical experiments

In order to indicate the benefit of the model reduction method presented in Sect. 2, we consider a linear controlled SPDE as in (2). In addition, we emphasize the applicability to unstable systems by rescaling and shifting the Laplacian. The concrete example of interest is

$$\begin{aligned} \frac{\partial \mathcal{X}(t,\zeta)}{\partial t} &= (\alpha\Delta + \beta I)\, \mathcal{X}(t,\zeta) + 1_{[\frac{\pi}{4}, \frac{3\pi}{4}]^2}(\zeta) u(t) + \gamma \mathrm{e}^{-|\zeta_1 - \frac{\pi}{2}| - \zeta_2} \mathcal{X}(t,\zeta) \frac{\partial w(t)}{\partial t}, \\ &\quad t \in [0,1], \quad \zeta \in [0,\pi]^2, \\ \mathcal{X}(t,\zeta) &= 0, \quad t \in [0,1], \quad \zeta \in \partial[0,\pi]^2, \quad \text{and} \quad \mathcal{X}(0,\zeta) \equiv 0, \end{aligned}$$

where $\alpha, \beta > 0$, $\gamma \in \mathbb{R}$ and w is an one-dimensional Wiener process. $\mathcal{X}(t, \cdot)$, $t \in [0, T]$, is interpreted as a process taking values in $H = L^2([0, \pi]^2)$. The input operator

$\mathcal{B}$ in (2) is characterized by $1_{[\frac{\pi}{4},\frac{3\pi}{4}]^2}(\cdot)$ and the noise operator $\mathcal{N}_1 = \mathcal{N}$ is defined trough $\mathcal{N}\mathcal{X} = \mathrm{e}^{-|\cdot-\frac{\pi}{2}|-\cdot}\mathcal{X}$ for $\mathcal{X} \in L^2([0,\pi]^2)$. Since the Dirichlet Laplacian generates a C_0-semigroup and its eigenfunctions $(h_k)_{k\in\mathbb{N}}$ represent a basis of H, the same is true for $\alpha\Delta+\beta I$. Therefore, we interpret the solution of the above SPDE in the mild sense. For more information to SPDEs and the mild solution concept, we refer to [6]. The quantity of interest is the average temperature on the non-controlled area, i.e.,

$$\mathcal{Y}(t) = \mathcal{C}\mathcal{X}(t,\cdot) := \frac{4}{3\pi^2}\int_{[0,\pi]^2\setminus[\frac{\pi}{4},\frac{3\pi}{4}]^2} \mathcal{X}(t,\zeta)\mathrm{d}\zeta.$$

In order to solve this SPDE numerically, a spatial discretization can be considered as a first step. Here, we choose a spectral Galerkin method relying on the global basis of eigenfunctions $(h_k)_{k\in\mathbb{N}}$. The idea is to construct an approximation $\mathcal{X}_n$ to $\mathcal{X}$ taking values in the subspace $H_n = \operatorname{span}\{h_1,\cdots,h_n\}$ and which converges to the SPDE solution with $n\to\infty$. For more detailed information on this discretization scheme, we refer to [10]. The vector of Fourier coefficients $x(t) = (\langle\mathcal{X}_n(t),h_1\rangle_H,\cdots,\langle\mathcal{X}_n(t),h_n\rangle_H)^\top$ is a solution of a system like (1) with $q=1$ and discretized operators

- $A = \alpha\operatorname{diag}(-\lambda_1,\cdots,-\lambda_n)+\beta I, \quad B = (\langle\mathcal{B},h_k\rangle_H)_{k=1\cdots n}, \quad C = (\mathcal{C}h_k)_{k=1\cdots n}$,
- $N_1 = (\langle\mathcal{N}h_i,h_k\rangle_H)_{k,i=1\cdots n}$ and $x_0 = 0$,

where $(-\lambda_k)_{k\in\mathbb{N}}$ are the ordered eigenvalues of Δ. We refer to [4], where a similar example was studied. There, more details are provided on how this system with its matrices is derived. Now, a small α and a larger β yield an unstable A, i.e., $\sigma(A)\not\subset\mathbb{C}$ which already violates asymptotic mean square stability of (1), i.e., $\mathbb{E}\,\|x(t;x_0,0)\|_2^2 \nrightarrow 0$ as $t\to\infty$. Moreover, a larger γ (larger noise) causes further instabilities. For that reason, we pick $\alpha=0.4$, $\beta=3$ and $\gamma=2$ in order to demonstrate the MOR procedure for a relatively unstable system. Notice that enlarging β or γ (or making α smaller) leads to a higher degree of instability. This affects the approximation quality in the reduced system given T is fixed. The intuition is that the less stable a system is the stronger the dominant subspaces are expanding in time. This is because some variables in unstable systems are strongly growing such that initially redundant directions become more relevant from a certain point of time. This can also be observed in numerical experiments.

Below, we fix a normalized control $u(t) = c_u\mathrm{e}^{-0.1t}$, $t\in[0,T]$, (the constant c_u ensures $\|u\|_{L^2_T}=1$) and apply the MOR method to the spatially discretized SPDE that is based on the balancing transformation $S=S_2$ described in Sect. 2.2. In Sect. 5.1, we compare the approximation quality of the ROMs using either the exact Gramian or inexact Gramians introduced in Sect. 4. Subsequently, Sect. 5.2 shows the reduced model accuracy in higher state space dimension, where solely inexact Gramians are available. We conclude the numerical experiments by discussing the impact of the terminal time T and the covariance matrix K in Sect. 5.3.

5.1 Simulations for $n = 100$ and $T = 1$

We compare the associated ROM (17) with the original system in dimension $n = 100$ first since this choice allows to determine $F(T)$, $G(T)$ and hence the Gramians P_T, Q_T exactly according to Sect. 4.1. As a consequence, we can compare the MOR scheme involving the exact Gramians with the same type of scheme relying on the approximated Gramians that are computed exploiting the approaches in Sects. 4.2 and 4.3. In particular, we first approximate $F(T)$ and $G(T)$ based on a Monte Carlo simulation using 10 realizations of the estimators (44) and (47), respectively. The functions $\mathcal{F}$ and $\mathcal{G}$ entering these estimators are chosen as in (46) with $c = 0$. We refer to the resulting matrices as Sect. 4.2 Gramians. At this point, we want to emphasize that these sampling based Gramians do not necessarily have to be accurate approximations of the exact Gramians in a component-wise sense. It is more important that the dominant subspaces of the system (eigenspaces of the Gramians) are captured in the approximation. Notice that the dominant subspace characterization is not improved if the number of samples is enlarged to 1000. Second, we determine the approximations $\mathcal{P}_T$ and $\mathcal{Q}_T$ according to Sect. 4.3 and call them Sect. 4.3 Gramians. The associated constants are chosen to be $c_F = c_G = 0$.

In Fig. 1, the HSVs $\sigma_{T,i}$, $i = \{1, \ldots, 50\}$, of system (1) are displayed. By Theorem 3.1 and the explanations below this theorem, it is known that small truncated $\sigma_{T,i}$ go along with a small reduction error of the MOR scheme. Due to the rapid decay of these values, we can therefore conclude that small error can already be achieved for small reduced dimensions r. For instance, we observe that $\sigma_{T,i} < 3.5\text{e}{-06}$ for $i \geq 8$ indicating a very high accuracy in the ROM for $r \geq 7$. This is confirmed by the error plot in Fig. 2 and the second column of Table 1. Moreover, Fig. 2 shows the tightness of the error bound in (29) that was specified in Theorem 3.1. The bound differs from the exact error only by a factor between 2.5 and 4.6 for the reduced dimensions considered in Fig. 2 and is hence a good indicator for the expected performance. Notice that the error is only exact up to deviations occurring due to the semi-implicit Euler–Maruyama discretization of (1) and (17) as well as the Monte Carlo approximation of the expected value using 10 000 paths. Besides the MOR error based on P_T and Q_T, Table 1 states the errors in case the approximating Gramians of Sects. 4.2 and 4.3 are used. It can be seen that both approximations perform roughly the same and that one looses an order of accuracy compared to the exact Gramian approach. However, one can lower the reduction error by an optimization with respect to the constants c, c_F, c_G. Moreover, we see that the accuracy is very good for the estimators of the covariances $F(T)$ and $G(T)$ used here.

5.2 Simulations for $n = 1000$ and $T = 1$

We repeat the simulations of Sect. 5.1 for $n = 1000$. This is a scenario, where the exact Gramians are not available anymore. Therefore, we conduct the balancing MOR scheme using the Sects. 4.2 and 4.3 Gramians only. In the context of Sect. 4.2 Gramians, it is important to mention that in higher dimensions it is required to use very efficient discretizations of the Ito integrals in (44) and (47). Otherwise, a very small

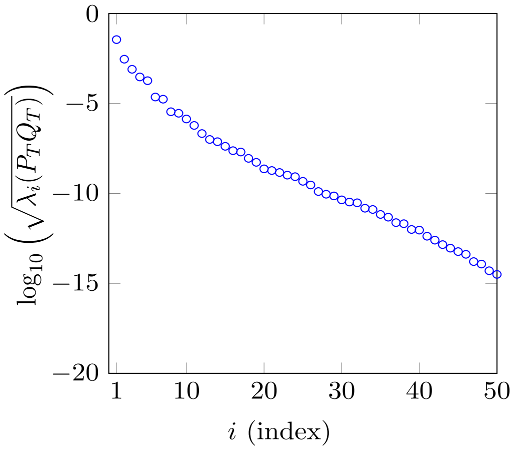

Fig. 1 Decay of first 50 logarithmic HSVs of system (1) based on time-limited Gramians P_T and Q_T

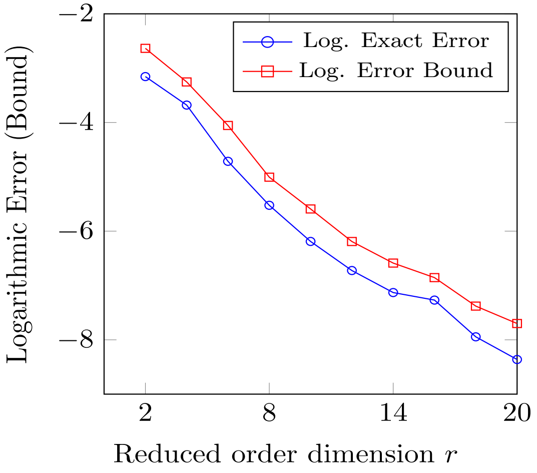

Fig. 2 $\log_{10}\left(\sup_{t\in[0,1]}\mathbb{E}\|y(t)-\bar{y}(t)\|_2\right)$ and logarithmic bound in (29) for $r \in \{2, 4, 6, 8, 10, 12, 14, 16, 18, 20\}$

Table 1 Error between the output y of (1) with $n = 100$ and the reduced output $\bar{y}$ of (17) using different Gramians to compute the balancing transformation $S = S_2$

Reduced dimension r	Error $\sup_{t\in[0,1]}\mathbb{E}\|y(t)-\bar{y}(t)\|_2$ of MOR using		
	Exact Gramians P_T, Q_T	Sect. 4.2 Gramians	Sect. 4.3 Gramians
2	7.00e−04	2.61e−03	1.75e−03
4	2.09e−04	1.82e−03	8.61e−04
8	2.99e−06	2.63e−05	4.51e−05
16	5.38e−08	1.31e−06	1.55e−06

Table 2 Error between the output y of (1) with $n = 1000$ and the reduced output $\bar{y}$ of (17) using Sects. 4.2 and 4.3 Gramians to compute the balancing transformation $S = S_2$

Reduced dimension r	Error $\sup_{t\in[0,1]} \mathbb{E}\|y(t) - \bar{y}(t)\|_2$ of MOR using	
	Sect. 4.2 Gramians	Sect. 4.3 Gramians
2	1.43e−03	1.72e−03
4	2.07e−03	8.57e−04
8	5.18e−05	9.26e−05
16	2.13e−06	4.88e−06

Table 3 Error between the output y of (1) and the reduced output $\bar{y}$ of (17) using the exact Gramians: $n = 100$, $S = S_2$ and $T = 0.5, 1, 2, 3$

Reduced dimension r	Error $\sup_{t\in[0,T]} \mathbb{E}\|y(t) - \bar{y}(t)\|_2$ of MOR for			
	$T = 0.5$	$T = 1$	$T = 2$	$T = 3$
2	3.98e−04	7.00e−04	2.17e−02	3.13e−02
4	1.46e−05	2.09e−04	2.86e−04	6.86e−04
8	2.82e−07	2.99e−06	7.80e−06	2.23e−05
16	5.46e−09	5.38e−08	1.12e−07	2.90e−07

step size is needed such that from the computational point of view it is better to omit these Ito integrals within the estimators, i.e., just x_B and x_C are supposed to be sampled to approximate $F(T)$ and $G(T)$. Table 2 shows that the balancing related MOR technique based on the approximated Gramians of Sects. 4.2 and 4.3 is beneficial in high dimensions. A very small reduction error can be observed and in the majority of the cases the sampling-based approach seems slightly more accurate than the approach of Sect. 4.3 given the same type of approximations for $F(T)$ and $G(T)$ for each ansatz.

5.3 Relevance of *T* and *K*

As in Sect. 5.1, let us fix $n = 100$ to be able to compute the Gramians exactly. We begin with deriving reduced systems on different intervals $[0, T]$. Second, we extend our model to a stochastic differential equation with noise dimension $q = 2$ and investigate the effect of different correlations between the two Wiener processes.

Relevance of the terminal time Let us study the scenario of Sect. 5.1 with $T = 0.5, 1, 2, 3$ using the exact Gramians to illustrate that dominant subspaces are changing in time. Indeed, we observe in Table 3 that for a fixed reduced dimension r the error gets bigger the larger the interval $[0, T]$ is. This means that with increasing T the reduced dimension has to be enlarged to ensure a certain desired approximation error. This is also intuitive in the sense that it is generally harder to find a good approximation on a larger interval in comparison with a smaller one.

Relevance the the covariance structure Let us extend the SPDE discretization by introducing $N_2 := N_1^{\frac{6}{5}}$ so that we have a system of the form (1) with $q = 2$ and standard

Table 4 Error between the output y of (1) and the reduced output $\bar{y}$ of (17) using the exact Gramians: $n = 100$, $S = S_2$, $T = 1$, $q = 2$ and different correlations $\rho = 0, 0.5, 1$

Reduced dimension r	Error $\sup_{t\in[0,1]} \mathbb{E}\|y(t) - \bar{y}(t)\|_2$ of MOR for		
	$\rho = 0$	$\rho = 0.5$	$\rho = 1$
2	1.10e−03	1.43e−03	1.79e−03
4	2.44e−04	2.34e−04	3.24e−04
8	5.71e−06	8.95e−06	1.34e−05
16	1.64e−07	2.37e−07	3.36e−07

Wiener processes w_1 and w_2. The goal is to investigate how the correlation between w_1 and w_2 influences the MOR error. For that reason, we choose the following three scenarios: $\mathbb{E}[w_1(t)w_2(t)] = \rho t$ with $\rho = 0, 0.5, 1$. Table 4 states the MOR errors for these correlations. In this example, we can observe that a higher correlation between the processes yields a larger error. A different observation was made in numerical examples studied in [14], where systems with high correlations in the noise processes gave a smaller reduction error. However, [14] studies different types of stochastic differential equations in the context of asset price models which do not have control inputs.

Funding Open Access funding enabled and organized by Projekt DEAL.

References

1. Bartels RH, Stewart GW (1972) Solution of the matrix equation $AX + XB = C$. Commun ACM 15(9):820–826
2. Barth A (2010) A finite element method for martingale-driven stochastic partial differential equations. Commun Stoch Anal 4(3):355–375
3. Becker S, Hartmann C (2019) Infinite-dimensional bilinear and stochastic balanced truncation with error bounds. Math Control Signals Syst 31:1–37
4. Benner P, Redmann M (2015) Model reduction for stochastic systems. Stoch PDE Anal Comp 3(3):291–338
5. Bhatia R (1997) Matrix analysis, vol 169. Springer, Berlin
6. Da Prato G, Zabczyk J (1992) Stochastic equations in infinite dimensions. Encyclopedia of mathematics and its applications, vol 44. Cambridge University Press, Cambridge
7. Damm T (2004) Rational matrix equations in stochastic control. Lecture notes in control and information sciences, vol 297. Springer, Berlin
8. Damm T (2008) Direct methods and ADI-preconditioned Krylov subspace methods for generalized Lyapunov equations. Numer Linear Algebra Appl 15(9):853–871

9. Gawronski W, Juang J (1990) Model reduction in limited time and frequency intervals. Int J Syst Sci 21(2):349–376
10. Hausenblas E (2003) Approximation for semilinear stochastic evolution equations. Potential Anal 18(2):141–186
11. Kürschner P (2018) Balanced truncation model order reduction in limited time intervals for large systems. Adv Comput Math 44(6):1821–1844
12. Øksendal B (2013) Stochastic differential equations (6th edition): an introduction with application. Springer, Berlin
13. Redmann M (2018) Type II singular perturbation approximation for linear systems with Lévy noise. SIAM J Control Optim 56(3):2120–2158
14. Redmann M, Bayer C, Goyal P (2021) Low-dimensional approximations of high-dimensional asset price models. SIAM J Financ Math 12(1):1–28
15. Redmann M, Kürschner P (2018) An output error bound for time-limited balanced truncation. Syst Control Lett 121:1–6
16. Redmann M, Pontes Duff I (2022) Full state approximation by Galerkin projection reduced order models for stochastic and bilinear systems. Appl Math Comput 420:126561
17. Simoncini V (2016) Computational methods for linear matrix equations. SIAM Rev 58(3):377–441