ORIGINAL ARTICLE

A new type of singular perturbation approximation for stochastic bilinear systems

Martin Redmann[1]

Abstract

Model order reduction (MOR) techniques are often used to reduce the order of spatially discretized (stochastic) partial differential equations and hence reduce computational complexity. A particular class of MOR techniques is balancing related methods which rely on simultaneously diagonalizing the system Gramians. This has been extensively studied for deterministic linear systems. The balancing procedure has already been extended to bilinear equations, an important subclass of nonlinear systems. The choice of Gramians in Al-Baiyat and Bettayeb (In: Proceedings of the 32nd IEEE conference on decision and control, 1993) is the most frequently used approach. A balancing related MOR scheme for bilinear systems called singular perturbation approximation (SPA) has been described that relies on this choice of Gramians. However, no error bound for this method could be proved. In this paper, we extend SPA to stochastic systems with bilinear drift and linear diffusion term. However, we propose a slightly modified reduced order model in comparison to previous work and choose a different reachability Gramian. Based on this new approach, an L^2-error bound is proved for SPA which is the main result of this paper. This bound is new even for deterministic bilinear systems.

Keywords Model order reduction · Singular perturbation approximation · Nonlinear stochastic systems · Lévy process

Mathematics Subject Classification Primary 93A15 · 93C10 · 93E03; Secondary 15A24 · 60J75

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

1   Institute of Mathematics, Martin Luther University Halle-Wittenberg, Theodor-Lieser-Str. 5, 06120 Halle (Saale), Germany

1 Introduction

Many phenomena in real life can be described by partial differential equations (PDEs). For an accurate mathematical modeling of these real-world applications, it is often required to take random effects into account. Uncertainties in a PDE model can, for example, be represented by an additional noise term leading to stochastic PDEs (SPDEs) [11,15,28,29].

It is often necessary to numerically approximate time-dependent SPDEs since analytic solutions do not exist in general. Discretizing in space can be considered as a first step. This can, for example, be done by spectral Galerkin [17,19,20] or finite element methods [2,21,22]. This usually leads to large-scale SDEs. Solving such complex SDE systems causes large computational cost. In this context, model order reduction (MOR) is used to save computational time by replacing high-dimensional systems by systems of low order in which the main information of the original system should be captured.

1.1 Literature review

Balancing related MOR schemes were developed for deterministic linear systems first. Famous representatives of this class of methods are balanced truncation (BT) [3,26,27] and singular perturbation approximation (SPA) [14,23].

BT was extended in [5,8] and SPA was generalized in [33] to stochastic linear systems. With this first extension, however, no L^2-error bound can be achieved [6,12]. Therefore, an alternative approach based on a different reachability Gramian was studied for stochastic linear systems leading to an L^2-error bound for BT [12] and for SPA [32].

BT [1,5] and SPA [18] were also generalized to bilinear systems, which we refer to as the standard approach for these systems. Although bilinear terms are very weak nonlinearities, they can be seen as a bridge between linear and nonlinear systems. This is because many nonlinear systems can be represented by bilinear systems using a so-called Carleman linearization. Applications of these equations can be found in various fields [10,25,34]. A drawback of the standard approach for bilinear systems is that no L^2-error bound could be shown so far. A first error bound for the standard ansatz was recently proved in [4], where an output error bound in L^∞ was formulated for infinite dimensional bilinear systems. Based on the alternative choice of Gramians in [12], a new type of BT for bilinear systems was considered [31] providing an L^2-error bound under the assumption of a possibly small bound on the controls.

A more general setting extending both the stochastic linear and the deterministic bilinear case is investigated in [30]. There, BT was studied and an L^2-error bound was proved overcoming the restriction of bounded controls in [31]. In this paper, we consider SPA for the same setting as in [30] in order to generalize the work in [18]. Moreover, we modify the reduced order model (ROM) in comparison to [18] and show an L^2-error bound which closes the gap in the theory in this context.

For further extensions of balancing related MOR techniques to other nonlinear systems, we refer to [7,35].

1.2 Outline of the paper

This work on SPA for stochastic bilinear systems, see (1), can be interpreted as a generalization of the deterministic bilinear case [18]. This extension builds a bridge between stochastic linear systems and stochastic nonlinear systems such that SPA can possibly be used for many more stochastic equations and applications.

In Sect. 2, the procedure of SPA is described and the ROM is stated. With this, we provide an alternative to [30], where BT was studied for the same kind of systems. We also extend the work of [18] combined with a modification of the ROM and the choice of a new Gramian, compare with (3). Based on this, we obtain an error bound in Sect. 3 that was not available even for the deterministic bilinear case. Its proof requires new techniques that cannot be found in the literature so far and this is the main result of this paper. The efficiency of the error bound is shown in Sect. 5. There, the proposed version of SPA is compared with the one in [18] and with BT [30].

2 Setting and ROM

Let every stochastic process appearing in this paper be defined on a filtered probability space $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P}\right)$.[1] Suppose that $M = (M_1, \ldots, M_v)^T$ is an $(\mathcal{F}_t)_{t \geq 0}$-adapted and $\mathbb{R}^v$-valued mean zero Lévy process with $\mathbb{E}\|M(t)\|_2^2 = \mathbb{E}\left[M^T(t)M(t)\right] < \infty$ for all $t \geq 0$. Moreover, we assume that for all $t, h \geq 0$ the random variable $M(t+h) - M(t)$ is independent of $\mathcal{F}_t$.

We consider a large-scale stochastic control system with bilinear drift that can be interpreted as a spatially discretized SPDE. We investigate the system

$$\mathrm{d}x(t) = [Ax(t) + Bu(t) + \sum_{k=1}^{m} N_k x(t)u_k(t)]\mathrm{d}t + \sum_{i=1}^{v} H_i x(t-)\mathrm{d}M_i(t), \quad (1a)$$

$$y(t) = Cx(t), \quad t \geq 0. \quad (1b)$$

We assume that $A, N_k, H_i \in \mathbb{R}^{n \times n}$ ($k \in \{1, \ldots, m\}$ and $i \in \{1, \ldots, v\}$), $B \in \mathbb{R}^{n \times m}$ and $C \in \mathbb{R}^{p \times n}$. Moreover, we define $x(t-) := \lim_{s \uparrow t} x(s)$. The control $u = (u_1, \ldots, u_m)^T$ is assumed to be deterministic and square integrable, i.e.,

$$\|u\|_{L_T^2}^2 := \int_0^T \|u(t)\|_2^2 \, \mathrm{d}t < \infty$$

for every $T > 0$. By [28, Theorem 4.44] there is a matrix $K = \left(k_{ij}\right)_{i, j=1,\ldots,v}$ such that $\mathbb{E}[M(t)M^T(t)] = Kt$. K is called covariance matrix of M.

In this paper, we study SPA to obtain a ROM. SPA is a balancing related method and relies on defining a reachability Gramian P and an observability Gramian Q. These two matrices are selected such that P characterizes the states that barely contribute to the dynamics in (1a) and Q identifies the less important states in (1b), see [30] for

---

[1] We assume that $(\mathcal{F}_t)_{t \geq 0}$ is right-continuous and $\mathcal{F}_0$ contains all sets A with $\mathbb{P}(A) = 0$.

estimates on the reachability and observability energy functionals. The estimates in [30] are global, whereas the standard choice of Gramians leads to results being valid in a small neighborhood of zero only [5,16].

In order to ensure the existence of these Gramians, throughout the paper it is assumed that

$$\lambda \left(A \otimes I + I \otimes A + \sum_{k=1}^{m} N_k \otimes N_k + \sum_{i,j=1}^{v} H_i \otimes H_j k_{ij} \right) \subset \mathbb{C}_-. \quad (2)$$

Here, $\lambda(\cdot)$ denotes the spectrum of a matrix. The reachability Gramian P and the observability Gramian Q are, according to [30], defined as the solutions to

$$A^T P^{-1} + P^{-1} A + \sum_{k=1}^{m} N_k^T P^{-1} N_k + \sum_{i,j=1}^{v} H_i^T P^{-1} H_j k_{ij} \leq -P^{-1} B B^T P^{-1}, \quad (3)$$

$$A^T Q + Q A + \sum_{k=1}^{m} N_k^T Q N_k + \sum_{i,j=1}^{v} H_i^T Q H_j k_{ij} \leq -C^T C, \quad (4)$$

where the existence of a positive definite solution to (3) goes back to [12,32] and is ensured if (2) holds.

We approximate the large scale system (1) by a system which has a much smaller state dimension $r \ll n$. This ROM is supposed be chosen, such that the corresponding output y_r is close to the original one, i.e., $y_r \approx y$ in some metric. In order to be able to remove both the unimportant states in (1a) and (1b) simultaneously, the first step of SPA is a state space transformation

$$(A, B, C, H_i, N_k) \mapsto (\tilde{A}, \tilde{B}, \tilde{C}, \tilde{H}_i, \tilde{N}_k) := (SAS^{-1}, SB, CS^{-1}, SH_iS^{-1}, SN_kS^{-1}),$$

where $S = \Sigma^{-\frac{1}{2}} X^T L_Q^T$ and $S^{-1} = L_P Y \Sigma^{-\frac{1}{2}}$. The ingredients of the balancing transformation are computed by the Cholesky factorizations $P = L_P L_P^T, Q = L_Q L_Q^T$, and the singular value decomposition $X \Sigma Y^T = L_Q^T L_P$. This transformation does not change the output y of the system, but it guarantees that the new Gramians are diagonal and equal, i.e., $SPS^T = S^{-T} Q S^{-1} = \Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_n)$ with $\sigma_1 \geq \ldots \geq \sigma_n$ being the Hankel singular values (HSVs) of the system.

We partition the balanced coefficients of (1) as follows:

$$\tilde{A} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \ \tilde{B} = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}, \ \tilde{N}_k = \begin{bmatrix} N_{k,11} & N_{k,12} \\ N_{k,21} & N_{k,22} \end{bmatrix}, \ \tilde{H}_i = \begin{bmatrix} H_{i,11} & H_{i,12} \\ H_{i,21} & H_{i,22} \end{bmatrix}, \ \tilde{C} = [\, C_1 \ C_2 \,],$$
$$(5)$$

where $A_{11}, N_{k,11}, H_{i,11} \in \mathbb{R}^{r \times r}$ ($k \in \{1, \ldots, m\}$ and $i \in \{1, \ldots, v\}$), $B_1 \in \mathbb{R}^{r \times m}$ and $C_1 \in \mathbb{R}^{p \times r}$ etc. Furthermore, we partition the state variable $\tilde{x}$ of the balanced system

and the diagonal matrix of HSVs

$$\tilde{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \text{ and } \Sigma = \begin{bmatrix} \Sigma_1 & \\ & \Sigma_2 \end{bmatrix}, \tag{6}$$

where x_1 takes values in $\mathbb{R}^r$ (x_2 accordingly), Σ_1 is the diagonal matrix of large HSVs and Σ_2 contains the small ones.

Remark 1 The balancing procedure requires the computation of the Gramians from (3) and (4). Practically, one always computes the solution of the equation in (4). The reason why an inequality is considered is that the proof of the error bound in Theorem 3 does not need an equality in (4). However, it is essential that we consider an inequality in (3). In contrast to the equation that may not have a solution, the inequality always has a solution under the given assumptions, but some regularization is needed to enforce uniqueness. In particular, one solves an optimization problem like, e.g., minimize $\mathrm{tr}(P)$ subject to (3). The reason why the trace is minimized is that one wants to achieve small HSVs, because this ensures a small error according to Theorem 3. We refer to Sect. 5 for more details on the computation of P.

Based on the balanced full model (1) with matrices as in (5), the ROM is obtained by neglecting the state variables x_2 corresponding to the small HSVs. The ROM using SPA is obtained by setting $\mathrm{d}x_2(t) = 0$ and furthermore neglecting the diffusion and bilinear term in the equation related to x_2. Note that the condition $\mathrm{d}x_2(t) = 0$ is almost surely false. However, we enforce it since it leads to a ROM with remarkable properties. After setting $\mathrm{d}x_2(t) = 0$ it is no simplification to not take the diffusion into account since it would follow automatically that it is zero within the resulting algebraic constraint due to the consideration in [33, Section 2]. Assuming that the bilinear term is equal to zero in the equation is needed, so that the matrices of the ROM below do not depend on the control u. The dependence on u is something that is not desired. With this simplification, one can solve for x_2 in the algebraic constraint. This leads to $x_2(t) = -A_{22}^{-1}(A_{21}x_1(t) + B_2u(t))$. Inserting this expression into the equation for x_1 and into the output equation, the reduced system is

$$\mathrm{d}\bar{x} = \left[\bar{A}\bar{x} + \bar{B}u + \sum_{k=1}^{m}(\bar{N}_k\bar{x} + \bar{E}_ku)u_k \right] \mathrm{d}t + \sum_{i=1}^{v}(\bar{H}_i\bar{x} + \bar{F}_iu)\mathrm{d}M_i, \tag{7a}$$

$$\bar{y}(t) = \bar{C}\bar{x}(t) + \bar{D}u(t), \quad t \geq 0, \tag{7b}$$

with matrices defined by

$$\bar{A} := A_{11} - A_{12}A_{22}^{-1}A_{21}, \quad \bar{B} := B_1 - A_{12}A_{22}^{-1}B_2, \quad \bar{C} := C_1 - C_2A_{22}^{-1}A_{21},$$
$$\bar{D} := -C_2A_{22}^{-1}B_2, \quad \bar{E}_k := -N_{k,12}A_{22}^{-1}B_2, \quad \bar{F}_i := -H_{i,12}A_{22}^{-1}B_2,$$
$$\bar{H}_i := H_{i,11} - H_{i,12}A_{22}^{-1}A_{21}, \quad \bar{N}_k := N_{k,11} - N_{k,12}A_{22}^{-1}A_{21},$$

where $\bar{x}(0) = 0$ and the time dependence in (7a) is omitted to shorten the notation. This straightforward ansatz is based on observations from the deterministic case ($N_k =$

$H_i = 0$), where x_2 represents the fast variables, i.e., $\dot{x}_2(t) \approx 0$ after a short time, see [23].

This ansatz for stochastic systems might, however, be false, no matter how small the HSVs corresponding to x_2 are. Despite the fact that, for the motivation, a possibly less convincing argument is used, this leads to a viable MOR method for which an error bound can be proved. An averaging principle would be a mathematically well-founded alternative to this naive approach. Averaging principles for stochastic systems have for example been investigated in [36,37]. A further strategy to derive a ROM in this context can be found in [9].

Moreover, notice that system (7) is not a bilinear system anymore due to the quadratic term in the control u. This is an essential difference to the ROM proposed in [18]. One can think about a structure preserving version by setting $B_2 = 0$ in (7). This would lead to a generalized variant of the ROMs considered in [18,33]. The reason why this simplified method is not studied is, because the error bound in Theorem 3 could not be achieved. We refer to a further discussion below Theorem 3 and to Sect. 5 where (7) is compared numerically with the version obtained by choosing $B_2 = 0$.

Remark 2 Notice that if $\sigma_r \neq \sigma_{r+1}$, then (2) implies

$$\lambda \left(A_{ll} \otimes I + I \otimes A_{ll} + \sum_{k=1}^{m} N_{k,ll} \otimes N_{k,ll} + \sum_{i,j=1}^{v} H_{i,ll} \otimes H_{j,ll} k_{ij} \right) \subset \mathbb{C}_-$$

for $l = 1, 2$ due to considerations in [6]. This implies $\lambda(A_{ll}) \subset \mathbb{C}_-$ and hence guarantees the existence of A_{ll}^{-1} for $l = 1, 2$.

3 L^2-error bound for SPA

The proof of the main result (Theorem 3) is divided into two parts. We first investigate the error that we encounter by removing the smallest HSV from the system in Sect. 3.1. In this reduction step, the structure from the full model (1) to the ROM (7) changes. Therefore, when removing the other HSVs from the system, another case needs to be studied in Sect. 3.2. There, an error bound between two ROM is achieved which are neighboring, i.e., the larger ROM has exactly one HSV more than the smaller one. The results of Sects. 3.1 and 3.2 are then combined in Sect. 3.3 in order to prove the general error bound.

For simplicity, let us from now on assume that system (1) is already balanced and has a zero initial condition ($x_0 = 0$). Thus, (3) and (4) become

$$A^T \Sigma^{-1} + \Sigma^{-1} A + \sum_{k=1}^{m} N_k^T \Sigma^{-1} N_k + \sum_{i,j=1}^{v} H_i^T \Sigma^{-1} H_j k_{ij} \leq -\Sigma^{-1} B B^T \Sigma^{-1},$$

$$\tag{8}$$

$$A^T \Sigma + \Sigma A + \sum_{k=1}^{m} N_k^T \Sigma N_k + \sum_{i,j=1}^{v} H_i^T \Sigma H_j k_{ij} \leq -C^T C, \tag{9}$$

i.e., $P = Q = \Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_n) > 0$.

3.1 Error bound of removing the smallest HSV

We consider the balanced system (1) with partitions as in (5) and (6). As mentioned before, the reduced state equation is obtained by approximating x_2 through $\bar{x}_2 := -A_{22}^{-1}(A_{21}\bar{x} + B_2 u)$ in the differential equation for x_1 and within the output Eq. (1b). The ROM (7) can hence be rewritten as

$$d\bar{x} = \left[A_1 \begin{bmatrix} \bar{x} \\ \bar{x}_2 \end{bmatrix} + B_1 u + \sum_{k=1}^{m} N_{k,1} \begin{bmatrix} \bar{x} \\ \bar{x}_2 \end{bmatrix} u_k \right] dt + \sum_{i=1}^{v} H_{i,1} \begin{bmatrix} \bar{x} \\ \bar{x}_2 \end{bmatrix} dM_i, \quad (10a)$$

$$\bar{y}(t) = C \begin{bmatrix} \bar{x}(t) \\ \bar{x}_2(t) \end{bmatrix}, \quad t \geq 0, \quad (10b)$$

where we define

$$A_1 = [\, A_{11} \; A_{12} \,], \quad H_{i,1} = [\, H_{i,11} \; H_{i,12} \,], \quad N_{k,1} = [\, N_{k,11} \; N_{k,12} \,].$$

We want to be able to subtract the ROM (10) from (1). Therefore, we add the following zero line to (10a)

$$d0 = \left[[\, A_{21} \; A_{22} \,] \begin{bmatrix} \bar{x} \\ \bar{x}_2 \end{bmatrix} + B_2 u - c_0 + \sum_{k=1}^{m} [\, N_{k,21} \; N_{k,22} \,] \begin{bmatrix} \bar{x} \\ \bar{x}_2 \end{bmatrix} u_k \right] dt \quad (11)$$

$$+ \sum_{i=1}^{v} \left[[\, H_{i,21} \; H_{i,22} \,] \begin{bmatrix} \bar{x} \\ \bar{x}_2 \end{bmatrix} - c_i \right] dM_i$$

with the compensation terms $c_0(t) := \sum_{k=1}^{m} [\, N_{k,21} \; N_{k,22} \,] \begin{bmatrix} \bar{x}(t) \\ \bar{x}_2(t) \end{bmatrix} u_k(t)$ and $c_i(t) :=$ $[\, H_{i,21} \; H_{i,22} \,] \begin{bmatrix} \bar{x}(t) \\ \bar{x}_2(t) \end{bmatrix}$ for $i = 1, \ldots, v$. Subtracting (10) from (1) together with (11) yields the following error system

$$dx_- = \left[A\mathbf{x}_- + \begin{bmatrix} 0 \\ c_0 \end{bmatrix} + \sum_{k=1}^{m} N_k \mathbf{x}_- u_k \right] dt + \sum_{i=1}^{v} \left[H_i \mathbf{x}_- + \begin{bmatrix} 0 \\ c_i \end{bmatrix} \right] dM_i, \quad (12a)$$

$$y_-(t) = C\mathbf{x}_-(t) = y(t) - \bar{y}(t), \quad t \geq 0, \quad (12b)$$

where we introduce the variables $\mathbf{x}_- = \begin{bmatrix} x_1 - \bar{x} \\ x_2 - \bar{x}_2 \end{bmatrix}$ and $x_- = \begin{bmatrix} x_1 - \bar{x} \\ x_2 \end{bmatrix}$. Moreover, adding (1a) and (10a) combined with (11) leads to

$$dx_+ = \left[A\mathbf{x}_+ + 2Bu - \begin{bmatrix} 0 \\ c_0 \end{bmatrix} + \sum_{k=1}^{m} N_k \mathbf{x}_+ u_k \right] dt + \sum_{i=1}^{v} \left[H_i \mathbf{x}_+ - \begin{bmatrix} 0 \\ c_i \end{bmatrix} \right] dM_i, \quad (13)$$

setting $\mathbf{x}_+ = \begin{bmatrix} x_1 + \bar{x} \\ x_2 + \bar{x}_2 \end{bmatrix}$ and $x_+ = \begin{bmatrix} x_1 + \bar{x} \\ x_2 \end{bmatrix}$.

Before we prove an error bound based on (12a) and (13), we need to introduce the vector u^0 of control components with a nonzero N_k. This is given by

$$u^0 = (u_1^0, \ldots, u_m^0)^T, \quad \text{where} \quad u_k^0 \equiv \begin{cases} 0 & \text{if } N_k = 0, \\ u_k & \text{else.} \end{cases} \tag{14}$$

The proof of the error bound when only one HSV is removed can be reduced to the task of finding suitable estimates for $\mathbb{E}[x_-^T(t)\, \Sigma\, x_-(t)]$ and $\mathbb{E}[x_+^T(t)\, \Sigma^{-1} x_+(t)]$. In particular, the rough idea is to apply Ito's lemma to $\mathbb{E}[x_-^T(t)\, \Sigma\, x_-(t)]$ and subsequently the lemma of Gronwall. Then, a first estimate for (12b) is obtained, i.e.,

$$\mathbb{E}[x_-^T(t)\, \Sigma\, x_-(t)] \leq -\mathbb{E}\, \|y - \bar{y}\|_{L_t^2}^2 + \sigma^2 f(t),$$

where f depends on the control u, the state x, the compensation terms $c_0, \ldots, c_v$ and Σ_2^{-1} assuming $\Sigma_2 = \sigma I$. Of course, the dependence of the above estimate on the state and the compensation terms is not desired. That is why another inequality is derived by using Ito's and Gronwall's lemma for $\mathbb{E}[x_+^T(t)\, \Sigma^{-1} x_+(t)]$ which yields

$$\mathbb{E}[x_+^T(t)\, \Sigma^{-1} x_+(t)] \leq -f(t) + 4\, \|u\|_{L_t^2}^2 \exp\left(\int_0^t \left\|u^0(s)\right\|_2^2 \mathrm{d}s\right).$$

Combining both inequalities, the result of the next theorem is obtained. A similar idea was also used to determine an error bound for BT [30]. However, the proof for SPA requires different techniques to find the estimates sketched above.

Theorem 1 *Let y be the output of the full model (1) with $x(0) = 0$, $\bar{y}$ be the output of the ROM (7) with $\bar{x}(0) = 0$ and $\Sigma_2 = \sigma I$, $\sigma > 0$, in (6). Then, the following holds:*

$$\left(\mathbb{E}\, \|y - \bar{y}\|_{L_T^2}^2\right)^{\frac{1}{2}} \leq 2\sigma\, \|u\|_{L_T^2} \exp\left(0.5 \left\|u^0\right\|_{L_T^2}^2\right).$$

Proof In order to improve the readability of this paper, the proof is given in Sect. 4.1
$\square$

We proceed with the study of an error bound between two ROM that are neighboring.

3.2 Error bound for neighboring ROMs

In this section, we investigate the output error between two ROMs, in which the larger ROM has exactly one more HSV than the smaller one. This concept of neighboring ROMs was first introduced in [32] but in the much simpler stochastic linear setting.

The reader might wonder why a second case is considered besides the one in Sect. 3.1 since one might just start with a full model that has the same structure as the ROM (7). The reason is that it is not clear how the Gramians need to be chosen for (7). In order to investigate the error between two ROMs by SPA, a finer partition than

the one in (5) is required. We partition the matrices of the balanced full system (1) as follows:

$$
A = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix}, \quad B = \begin{bmatrix} B_1 \\ B_2 \\ B_3 \end{bmatrix}, \quad C = [\, C_1 \; C_2 \; C_3 \,], \tag{15a}
$$

$$
H_i = \begin{bmatrix} H_{i,11} & H_{i,12} & H_{i,13} \\ H_{i,21} & H_{i,22} & H_{i,23} \\ H_{i,31} & H_{i,32} & H_{i,33} \end{bmatrix}, \quad N_k = \begin{bmatrix} N_{k,11} & N_{k,12} & N_{k,13} \\ N_{k,21} & N_{k,22} & N_{k,23} \\ N_{k,31} & N_{k,32} & N_{k,33} \end{bmatrix}. \tag{15b}
$$

The partitioned balanced solution to (1a) and the Gramians are then of the form

$$
x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \text{ and } \Sigma = \begin{bmatrix} \Sigma_1 & & \\ & \Sigma_2 & \\ & & \Sigma_3 \end{bmatrix}. \tag{16}
$$

We introduce the ROM of truncating Σ_3 first. According to the procedure described in Sect. 2, the reduced system is obtained by setting dx_3 equal to zero, neglecting the bilinear and the diffusion term in this equation. The solution $\bar{x}_3$ of the resulting algebraic constraint is an approximation for x_3. One can solve for this approximating variable and obtains $\bar{x}_3 = -A_{33}^{-1}(A_{31}x_1 + A_{32}x_2 + B_3 u)$. Inserting this result for x_3 in the equations for x_1, x_2 and into the output Eq. (1b) leads to

$$
d \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \left[\hat{A} \begin{bmatrix} x_1 \\ x_2 \\ \bar{x}_3 \end{bmatrix} + \hat{B}u + \sum_{k=1}^{m} \hat{N}_k \begin{bmatrix} x_1 \\ x_2 \\ \bar{x}_3 \end{bmatrix} u_k \right] dt + \sum_{i=1}^{v} \hat{H}_i \begin{bmatrix} x_1 \\ x_2 \\ \bar{x}_3 \end{bmatrix} dM_i, \tag{17a}
$$

$$
\bar{y}(t) = C \begin{bmatrix} x_1(t) \\ x_2(t) \\ \bar{x}_3(t) \end{bmatrix}, \quad t \geq 0, \tag{17b}
$$

where $\begin{bmatrix} x_1(0) \\ x_2(0) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ and

$$
\hat{A} = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \end{bmatrix}, \quad \hat{B} = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}, \quad \hat{H}_i = \begin{bmatrix} H_{i,11} & H_{i,12} & H_{i,13} \\ H_{i,21} & H_{i,22} & H_{i,23} \end{bmatrix}, \quad \hat{N}_k = \begin{bmatrix} N_{k,11} & N_{k,12} & N_{k,13} \\ N_{k,21} & N_{k,22} & N_{k,23} \end{bmatrix}.
$$

We aim to determine the error between this ROM and the reduced system of neglecting Σ_2 and Σ_3. This is

$$
d\bar{x}_r = \left[\hat{A}_r \begin{bmatrix} \bar{x}_r \\ h_1 \\ h_2 \end{bmatrix} + B_1 u + \sum_{k=1}^{m} \hat{N}_{r,k} \begin{bmatrix} \bar{x}_r \\ h_1 \\ h_2 \end{bmatrix} u_k \right] dt + \sum_{i=1}^{v} \hat{H}_{r,i} \begin{bmatrix} \bar{x}_r \\ h_1 \\ h_2 \end{bmatrix} dM_i, \tag{18a}
$$

$$
\bar{y}_r(t) = [\, C_1 \; C_2 \; C_3 \,] \begin{bmatrix} \bar{x}_r(t) \\ h_1(t) \\ h_2(t) \end{bmatrix}, \quad t \geq 0, \tag{18b}
$$

where $\bar{x}_r(0) = 0$,

$$
\hat{A}_r = [\, A_{11} \; A_{12} \; A_{13} \,], \quad \hat{H}_{r,i} = [\, H_{i,11} \; H_{i,12} \; H_{i,13} \,], \quad \hat{N}_{r,k} = [\, N_{k,11} \; N_{k,12} \; N_{k,13} \,]
$$

and we define

$$h(t) = \begin{bmatrix} h_1(t) \\ h_2(t) \end{bmatrix} = - \begin{bmatrix} A_{22} & A_{23} \\ A_{32} & A_{33} \end{bmatrix}^{-1} \left(\begin{bmatrix} A_{21} \\ A_{31} \end{bmatrix} \bar{x}_r(t) + \begin{bmatrix} B_2 \\ B_3 \end{bmatrix} u(t) \right). \tag{19}$$

In order to find a bound for the error between (17b) and (18b), state variables analogous to $\mathbf{x}_-$ and $\mathbf{x}_+$ in Sect. 3.1 are constructed in the following and corresponding equations are derived. For simplicity, we use a similar notation again and define

$$\hat{\mathbf{x}}_- = \begin{bmatrix} x_1 - \bar{x}_r \\ x_2 - h_1 \\ \bar{x}_3 - h_2 \end{bmatrix} \text{ and } \hat{\mathbf{x}}_+ = \begin{bmatrix} x_1 + \bar{x}_r \\ x_2 + h_1 \\ \bar{x}_3 + h_2 \end{bmatrix}.$$

Now, we find the differential equations for $\hat{\mathbf{x}}_-$ and $\hat{\mathbf{x}}_+$. Using (19), we find that

$$\begin{aligned} \begin{bmatrix} A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix} \begin{bmatrix} \bar{x}_r \\ h_1 \\ h_2 \end{bmatrix} &= \begin{bmatrix} A_{21} \\ A_{31} \end{bmatrix} \bar{x}_r + \begin{bmatrix} A_{22} & A_{23} \\ A_{32} & A_{33} \end{bmatrix} h \\ &= \begin{bmatrix} A_{21} \\ A_{31} \end{bmatrix} \bar{x}_r - \begin{bmatrix} A_{22} & A_{23} \\ A_{32} & A_{33} \end{bmatrix} \begin{bmatrix} A_{22} & A_{23} \\ A_{32} & A_{33} \end{bmatrix}^{-1} \left(\begin{bmatrix} A_{21} \\ A_{31} \end{bmatrix} \bar{x}_r + \begin{bmatrix} B_2 \\ B_3 \end{bmatrix} u \right) \\ &= - \begin{bmatrix} B_2 \\ B_3 \end{bmatrix} u. \end{aligned} \tag{20}$$

Applying the first line of (20), we obtain the following equation

$$d0 = \left[\begin{bmatrix} A_{21} & A_{22} & A_{23} \end{bmatrix} \begin{bmatrix} \bar{x}_r \\ h_1 \\ h_2 \end{bmatrix} + B_2 u - \hat{c}_0 + \sum_{k=1}^{m} \begin{bmatrix} N_{k,21} & N_{k,22} & N_{k,23} \end{bmatrix} \begin{bmatrix} \bar{x}_r \\ h_1 \\ h_2 \end{bmatrix} u_k \right] \mathrm{d}t$$
$$+ \sum_{i=1}^{v} \left[\begin{bmatrix} H_{i,21} & H_{i,22} & H_{i,23} \end{bmatrix} \begin{bmatrix} \bar{x}_r \\ h_1 \\ h_2 \end{bmatrix} - \hat{c}_i \right] \mathrm{d}M_i, \tag{21}$$

where $\hat{c}_0 = \sum_{k=1}^{m} \begin{bmatrix} N_{k,21} & N_{k,22} & N_{k,23} \end{bmatrix} \begin{bmatrix} \bar{x}_r \\ h_1 \\ h_2 \end{bmatrix} u_k$ and $\hat{c}_i = \begin{bmatrix} H_{i,21} & H_{i,22} & H_{i,23} \end{bmatrix} \begin{bmatrix} \bar{x}_r \\ h_1 \\ h_2 \end{bmatrix}$ for $i = 1, \ldots, v$. We add the zero line (21) to the state Eq. (18a) and subtract the resulting system from (17). Hence, we obtain

$$\mathrm{d}\hat{x}_- = \left[\hat{A}\hat{\mathbf{x}}_- + \begin{bmatrix} 0 \\ \hat{c}_0 \end{bmatrix} + \sum_{k=1}^{m} \hat{N}_k \hat{\mathbf{x}}_- u_k \right] \mathrm{d}t + \sum_{i=1}^{v} \left[\hat{H}_i \hat{\mathbf{x}}_- + \begin{bmatrix} 0 \\ \hat{c}_i \end{bmatrix} \right] \mathrm{d}M_i, \tag{22a}$$

$$\hat{y}_-(t) = C\hat{\mathbf{x}}_-(t) = \bar{y}(t) - \bar{y}_r(t), \quad t \geq 0, \tag{22b}$$

where $\hat{x}_- = \begin{bmatrix} x_1 - \bar{x}_r \\ x_2 \end{bmatrix}$. One can see that the output of (22) is the output error that we aim to analyze. The sum of (17a) and (18a) together with (21) leads to

$$\mathrm{d}\hat{x}_+ = \left[\hat{A}\hat{\mathbf{x}}_+ + 2\hat{B}u - \begin{bmatrix} 0 \\ \hat{c}_0 \end{bmatrix} + \sum_{k=1}^{m} \hat{N}_k \hat{\mathbf{x}}_+ u_k \right] \mathrm{d}t + \sum_{i=1}^{v} \left[\hat{H}_i \hat{\mathbf{x}}_+ - \begin{bmatrix} 0 \\ \hat{c}_i \end{bmatrix} \right] \mathrm{d}M_i, \tag{23}$$

where $\hat{x}_+ = \begin{bmatrix} x_1 + \bar{x}_r \\ x_2 \end{bmatrix}$. We now state the output error between the systems (17) and (18) for the case that the ROM are neighboring, i.e., the larger model has exactly one HSV more than the smaller one. Similarly as in Theorem 1, the proof relies on finding suitable estimates for $\mathbb{E}\left[\hat{x}_-^T(t)\hat{\Sigma}\hat{x}_-(t)\right]$ and $\mathbb{E}\left[\hat{x}_+^T(t)\hat{\Sigma}\hat{x}_+(t)\right]$, where $\hat{\Sigma} = \begin{bmatrix} \Sigma_1 & \\ & \Sigma_2 \end{bmatrix}$ is a submatrix of Σ in (16).

Theorem 2 *Let $\bar{y}$ be the output of the ROM (17), $\bar{y}_r$ be the output of the ROM (18) and $\Sigma_2 = \sigma I, \sigma > 0$, in (16). Then, it holds that*

$$\left(\mathbb{E}\|\bar{y} - \bar{y}_r\|_{L_T^2}^2\right)^{\frac{1}{2}} \leq 2\sigma \|u\|_{L_T^2} \exp\left(0.5 \left\|u^0\right\|_{L_T^2}^2\right).$$

Proof In order to improve the readability of this paper, the proof is presented later, see Sect. 4.2.   □

3.3 Main result

In the following, the main result of this paper is formulated. It is a consequence of Theorems 1 and 2.

Theorem 3 *Let y be the output of the full model (1) with $x(0) = 0$ and $\bar{y}$ be the output of the ROM (7) with zero initial state. Then, for all $T > 0$, it holds that*

$$\left(\mathbb{E}\|y - \bar{y}\|_{L_T^2}^2\right)^{\frac{1}{2}} \leq 2(\sigma_{r+1} + \sigma_{r+2} + \ldots + \sigma_n) \|u\|_{L_T^2} \exp\left(0.5 \left\|u^0\right\|_{L_T^2}^2\right),$$

where $\sigma_{r+1}, \sigma_{r+2}, \ldots, \sigma_n$ are the diagonal entries of Σ_2 and $u^0 = (u_1^0, \ldots, u_m^0)^T$ is the control vector with components defined by $u_k^0 \equiv \begin{cases} 0 & \text{if } N_k = 0, \\ u_k & \text{else.} \end{cases}$

Proof We apply the results in Theorems 1 and 2. We remove the HSVs step by step and exploit the triangle inequality in order to bound the error between the outputs y and $\bar{y}$. We introduce $\bar{y}_\ell$ as the output of ROM (7) with state space dimension $\ell = r, r+1, \ldots, n-1$. Notice that $\bar{y}_r$ coincides with $\bar{y}$. Moreover, we set $\bar{y}_n := y$. We then have

$$\left(\mathbb{E}\|y - \bar{y}\|_{L_T^2}^2\right)^{\frac{1}{2}} \leq \sum_{\ell=r+1}^{n} \left(\mathbb{E}\|\bar{y}_\ell - \bar{y}_{\ell-1}\|_{L_T^2}^2\right)^{\frac{1}{2}}.$$

In the reduction step from y to $\bar{y}_{n-1}$ only the smallest HSV σ_n is removed from the system. Hence, by Theorem 1, we have

$$\left(\mathbb{E}\|y - \bar{y}_{n-1}\|_{L_T^2}^2\right)^{\frac{1}{2}} \leq 2\sigma_n \|u\|_{L_T^2} \exp\left(0.5 \left\|u^0\right\|_{L_T^2}^2\right).$$

The ROMs of the outputs $\bar{y}_\ell$ and $\bar{y}_{\ell-1}$ are neighboring according to Sect. 3.2, i.e., only the HSV σ_ℓ is removed in the reduction step. By Theorem 2, we obtain

$$\left(\mathbb{E} \left\| \bar{y}_\ell - \bar{y}_{\ell-1} \right\|_{L_T^2}^2 \right)^{\frac{1}{2}} \leq 2\sigma_\ell \left\| u \right\|_{L_T^2} \exp\left(0.5 \left\| u^0 \right\|_{L_T^2}^2 \right)$$

for $\ell = r + 1, \ldots, n - 1$. This provides the claimed result.   □

The result in Theorem 3 tells us that the ROM (7) yields a very good approximation if the truncated HSVs (diagonal entries of Σ_2) are small and the vector u^0 of control components with a nonzero N_k is not too large. The exponential in the error bound can be an indicator that SPA performs badly if u^0 is very large.

We conclude this section by a discussion on the result in Theorem 3. It is important to notice that the dependence of the error bound on the system matrices and the covariance matrix of M is hidden in the HSVs. Those are given by $\sigma_\ell = \sqrt{\lambda_\ell(PQ)}$. We can see from (3) and (4) that the Gramians depend on (A, B, C, H_i, N_k, K) and hence $\sigma_\ell = \sigma_\ell(A, B, C, H_i, N_k, K)$. Consequently, changing the system matrices or the covariance matrix will change the HSVs too. Now, if K is large, the terms related to H_i in (3) and (4) become more dominant which results in larger HSVs. According to Theorem 3 a worse approximation can then be expected. We also observe that the exponential term in Theorem 3 is related to the bilinearity in the drift. Setting $N_k = 0$ for all $k = 1, \ldots, m$ the exponential becomes a one. This results in the bound that is known from the stochastic linear case [32]. Choosing $H_i = 0$ for all $i = 1, \ldots, v$ yields a bound for the deterministic bilinear case. Notice that in this case, the state variables of the full and reduced system are not random anymore such that the expected value is redundant and can hence be omitted. Finally, considering $H_i = N_k = 0$ leads to the bound obtained for the deterministic linear case [23].

Since the ROM (7) has a different structure than (1), it is an important question why we do not use a structure preserving generalization of SPA considered in [18,33]. This variant of SPA is obtained by setting $B_2 = 0$ in (7). Conducting the proof of the error bound for this simplified method, one would obtain an additional term in the error bound that depends on Σ_1, a possibly large matrix. This indicates that SPA with $B_2 = 0$ probability performs worse than the version stated in (7). We refer to Sect. 5 where both versions of SPA are compared numerically.

4 Proofs of Theorems 1 and 2

In this section, we present the pending proofs of Theorems 1 and 2.

4.1 Proof of Theorem 1

We derive a suitable upper bound for $\mathbb{E}[x_-^T(t)\Sigma x_-(t)]$ first applying Ito's formula. Hence, Lemma 1 ("Appendix") and Eq. (12a) yield

$$\mathbb{E}\left[x_-^T(t)\Sigma x_-(t)\right] = 2\int_0^t \mathbb{E}\left[x_-^T \Sigma \left(Ax_- + \sum_{k=1}^m (N_k\mathbf{x}_- u_k) + \begin{bmatrix} 0 \\ c_0 \end{bmatrix}\right)\right]\mathrm{d}s$$

$$+ \int_0^t \sum_{i,j=1}^v \mathbb{E}\left[\left(H_i\mathbf{x}_- + \begin{bmatrix} 0 \\ c_i \end{bmatrix}\right)^T \Sigma \left(H_j\mathbf{x}_- + \begin{bmatrix} 0 \\ c_j \end{bmatrix}\right)\right]k_{ij}\mathrm{d}s.$$

$$(24)$$

We find an estimate for the terms related to N_k, that is

$$\sum_{k=1}^m 2x_-^T(s)\Sigma N_k\mathbf{x}_-(s)u_k(s) = \sum_{k=1}^m 2\left\langle \Sigma^{\frac{1}{2}}x_-(s)u_k(s), \Sigma^{\frac{1}{2}}N_k\mathbf{x}_-(s)\right\rangle_2$$

$$\leq \sum_{k=1}^m \left\| \Sigma^{\frac{1}{2}}x_-(s)u_k^0(s)\right\|_2^2 + \left\| \Sigma^{\frac{1}{2}}N_k\mathbf{x}_-(s)\right\|_2^2$$

$$= x_-^T(s)\Sigma x_-(s)\left\| u^0(s)\right\|_2^2 + \sum_{k=1}^m \mathbf{x}_-^T(s)N_k^T \Sigma N_k\mathbf{x}_-(s),$$

$$(25)$$

where u^0 is defined in (14). Moreover, adding a zero, we rewrite

$$2x_-^T(s)\Sigma Ax_-(s) = 2\mathbf{x}_-^T(s)\Sigma Ax_-(s) + 2\begin{bmatrix} 0 \\ \bar{x}_2(s) \end{bmatrix}^T \Sigma Ax_-(s)$$

$$= \mathbf{x}_-^T(s)(A^T\Sigma + \Sigma A)\mathbf{x}_-(s) + 2\begin{bmatrix} 0 \\ \bar{x}_2(s) \end{bmatrix}^T \Sigma Ax_-(s). \quad (26)$$

With (25) and (26), (24) becomes

$$\mathbb{E}\left[x_-^T(t)\Sigma x_-(t)\right] \leq \mathbb{E}\int_0^t \mathbf{x}_-^T\left(A^T\Sigma + \Sigma A + \sum_{k=1}^m N_k^T\Sigma N_k + \sum_{i,j=1}^v H_i^T\Sigma H_j k_{ij}\right)\mathbf{x}_-\mathrm{d}s$$

$$+ \mathbb{E}\int_0^t 2x_-^T\Sigma \begin{bmatrix} 0 \\ c_0 \end{bmatrix} + \sum_{i,j=1}^v \left(2H_i\mathbf{x}_- + \begin{bmatrix} 0 \\ c_i \end{bmatrix}\right)^T \Sigma \begin{bmatrix} 0 \\ c_j \end{bmatrix}k_{ij}\mathrm{d}s$$

$$+ \int_0^t \mathbb{E}\left[x_-^T\Sigma x_-\right]\left\| u^0\right\|_2^2 \mathrm{d}s + \mathbb{E}\int_0^t 2\begin{bmatrix} 0 \\ \bar{x}_2 \end{bmatrix}^T \Sigma Ax_-\mathrm{d}s. \quad (27)$$

Taking the partitions of x_- and Σ into account, we see that $x_-^T\Sigma \begin{bmatrix} 0 \\ c_0 \end{bmatrix} = x_2^T\Sigma_2 c_0$. Furthermore, the partitions of $\mathbf{x}_-$ and H_i yield

$$\left(2H_i\mathbf{x}_- + \begin{bmatrix} 0 \\ c_i \end{bmatrix}\right)^T \Sigma \begin{bmatrix} 0 \\ c_j \end{bmatrix} = \left(2H_i\mathbf{x}_- + \begin{bmatrix} 0 \\ c_i \end{bmatrix}\right)^T \begin{bmatrix} 0 \\ \Sigma_2 c_j \end{bmatrix}$$

$$= \left(2\begin{bmatrix} H_{i,21} & H_{i,22} \end{bmatrix}\left(x - \begin{bmatrix} \bar{x} \\ \bar{x}_2 \end{bmatrix}\right) + c_i\right)^T \Sigma_2 c_j$$

$$= \left(2\begin{bmatrix} H_{i,21} & H_{i,22} \end{bmatrix}x - c_i\right)^T \Sigma_2 c_j, \quad (28)$$

since $[\,H_{i,21}\ H_{i,22}\,]\begin{bmatrix}\bar{x}\\\bar{x}_2\end{bmatrix}=c_i$. Using the partition of A, it holds that

$$
\begin{aligned}
2\begin{bmatrix}0\\\bar{x}_2\end{bmatrix}^T \Sigma A\mathbf{x}_- &= 2\begin{bmatrix}0\ \bar{x}_2^T \Sigma_2\end{bmatrix}A\mathbf{x}_- = 2\bar{x}_2^T \Sigma_2[\,A_{21}\ A_{22}\,]\left(x-\begin{bmatrix}\bar{x}\\\bar{x}_2\end{bmatrix}\right)\\
&= 2\bar{x}_2^T \Sigma_2([\,A_{21}\ A_{22}\,]x + B_2 u),
\end{aligned}
\tag{29}
$$

because $[\,A_{21}\ A_{22}\,]\begin{bmatrix}\bar{x}\\\bar{x}_2\end{bmatrix}=-B_2 u$. We insert (9) and (12b) into inequality (27) and exploit the relations in (28) and (29). Hence,

$$
\begin{aligned}
\mathbb{E}\left[x_-^T(t)\Sigma x_-(t)\right] &\le -\mathbb{E}\,\|y-\bar{y}\|_{L_t^2}^2 + \int_0^t \mathbb{E}\left[x_-^T \Sigma x_-\right]\left\|u^0\right\|_2^2 \mathrm{d}s\\
&\quad + \mathbb{E}\int_0^t 2x_2^T \Sigma_2 c_0 + \sum_{i,j=1}^v (2[\,H_{i,21}\ H_{i,22}\,]x - c_i)^T \Sigma_2 c_j k_{ij}\mathrm{d}s\\
&\quad + \mathbb{E}\int_0^t 2\bar{x}_2^T \Sigma_2([\,A_{21}\ A_{22}\,]x + B_2 u)\mathrm{d}s.
\end{aligned}
$$

We define the function $\alpha_-(t) := \mathbb{E}\int_0^t 2x_2^T \Sigma_2 c_0 + \sum_{i,j=1}^v (2[\,H_{i,21}\ H_{i,22}\,]x - c_i)^T$ $\Sigma_2 c_j k_{ij}\mathrm{d}s + \mathbb{E}\int_0^t 2\bar{x}_2^T \Sigma_2([\,A_{21}\ A_{22}\,]x + B_2 u)\mathrm{d}s$ and apply Lemma 3 ('Appendix') implying

$$
\begin{aligned}
\mathbb{E}\left[x_-^T(t)\Sigma x_-(t)\right] &\le \alpha_-(t) - \mathbb{E}\,\|y-\bar{y}\|_{L_t^2}^2\\
&\quad + \int_0^t (\alpha_-(s) - \mathbb{E}\,\|y-\bar{y}\|_{L_s^2}^2)\left\|u^0(s)\right\|_2^2 \exp\left(\int_s^t \left\|u^0(w)\right\|_2^2 \mathrm{d}w\right)\mathrm{d}s.
\end{aligned}
$$

Since Σ is positive definite, we obtain an upper bound for the output error by

$$
\mathbb{E}\,\|y-\bar{y}\|_{L_t^2}^2 \le \alpha_-(t) + \int_0^t \alpha_-(s)\left\|u^0(s)\right\|_2^2 \exp\left(\int_s^t \left\|u^0(w)\right\|_2^2 \mathrm{d}w\right)\mathrm{d}s.
$$

Defining the term $\alpha_+(t) := \mathbb{E}\int_0^t 2x_2^T \Sigma_2^{-1} c_0 + \sum_{i,j=1}^v (2[\,H_{i,21}\ H_{i,22}\,]x - c_i)^T$ $\Sigma_2^{-1} c_j k_{ij}\mathrm{d}s + \mathbb{E}\int_0^t 2\bar{x}_2^T \Sigma_2^{-1}([\,A_{21}\ A_{22}\,]x + B_2 u)\mathrm{d}s$ and exploiting the assumption that $\Sigma_2 = \sigma I$, leads to

$$
\mathbb{E}\,\|y-\bar{y}\|_{L_t^2}^2 \le \sigma^2\left[\alpha_+(t) + \int_0^t \alpha_+(s)\left\|u^0(s)\right\|_2^2 \exp\left(\int_s^t \left\|u^0(w)\right\|_2^2 \mathrm{d}w\right)\mathrm{d}s\right].
\tag{30}
$$

The remaining step is to find a bound for the right side of (30) that does not depend on α_+ anymore. For that reason, a bound for the expression $\mathbb{E}[x_+^T(t)\Sigma^{-1}x_+(t)]$ is derived next using Ito's lemma again. From (13) and Lemma 1 ('Appendix'), we obtain

$$\mathbb{E}\left[x_+^T(t)\Sigma^{-1}x_+(t)\right] = 2\int_0^t \mathbb{E}\left[x_+^T\Sigma^{-1}\left(Ax_+ + 2Bu + \sum_{k=1}^m (N_k\mathbf{x}_+ u_k) - \left[\begin{smallmatrix}0\\c_0\end{smallmatrix}\right]\right)\right]ds$$
$$+ \int_0^t \sum_{i,j=1}^v \mathbb{E}\left[\left(H_i\mathbf{x}_+ - \left[\begin{smallmatrix}0\\c_i\end{smallmatrix}\right]\right)^T \Sigma^{-1}\left(H_j\mathbf{x}_+ - \left[\begin{smallmatrix}0\\c_j\end{smallmatrix}\right]\right)\right]k_{ij}ds.$$

(31)

Analogously to (25), it holds that

$$\sum_{k=1}^m 2x_+^T(s)\Sigma^{-1}N_k\mathbf{x}_+(s)u_k(s) \le x_+^T(s)\Sigma^{-1}x_+(s)\left\|u^0(s)\right\|_2^2 + \sum_{k=1}^m \mathbf{x}_+^T(s)N_k^T\Sigma^{-1}N_k\mathbf{x}_+(s).$$

Additionally, we rearrange the term related to A as follows

$$2x_+^T(s)\Sigma^{-1}A\mathbf{x}_+(s) = 2\mathbf{x}_+^T(s)\Sigma^{-1}A\mathbf{x}_+(s) - 2\left[\begin{smallmatrix}0\\\tilde{x}_2(s)\end{smallmatrix}\right]^T \Sigma^{-1}A\mathbf{x}_+(s)$$
$$= \mathbf{x}_+^T(s)(A^T\Sigma^{-1} + \Sigma^{-1}A)\mathbf{x}_+(s) - 2\left[\begin{smallmatrix}0\\\tilde{x}_2(s)\end{smallmatrix}\right]^T \Sigma^{-1}A\mathbf{x}_+(s).$$

Moreover, we have

$$4x_+^T(s)\Sigma^{-1}Bu(s) = 4\mathbf{x}_+^T(s)\Sigma^{-1}Bu(s) - 4\left[\begin{smallmatrix}0\\\tilde{x}_2(s)\end{smallmatrix}\right]^T \Sigma^{-1}Bu(s).$$

We plug in the above results into (31) which gives us

$$\mathbb{E}\left[x_+^T(t)\Sigma^{-1}x_+(t)\right]$$
$$\le \mathbb{E}\int_0^t \mathbf{x}_+^T\left(A^T\Sigma^{-1} + \Sigma^{-1}A + \sum_{k=1}^m N_k^T\Sigma^{-1}N_k + \sum_{i,j=1}^v H_i^T\Sigma^{-1}H_j k_{ij}\right)\mathbf{x}_+ds$$
$$- \mathbb{E}\int_0^t 2x_+^T\Sigma^{-1}\left[\begin{smallmatrix}0\\c_0\end{smallmatrix}\right] + \sum_{i,j=1}^v \left(2H_i\mathbf{x}_+ - \left[\begin{smallmatrix}0\\c_i\end{smallmatrix}\right]\right)^T \Sigma^{-1}\left[\begin{smallmatrix}0\\c_j\end{smallmatrix}\right]k_{ij}ds$$
$$- \mathbb{E}\int_0^t 2\left[\begin{smallmatrix}0\\\tilde{x}_2\end{smallmatrix}\right]^T \Sigma^{-1}(A\mathbf{x}_+ + 2Bu)ds + \mathbb{E}\int_0^t 4\mathbf{x}_+^T\Sigma^{-1}Bu ds$$
$$+ \int_0^t \mathbb{E}\left[x_+^T\Sigma^{-1}x_+\right]\left\|u^0\right\|_2^2 ds.$$

(32)

From inequality (8) and the Schur complement condition on definiteness, it follows that

$$\begin{bmatrix} A^T\Sigma^{-1} + \Sigma^{-1}A + \sum_{k=1}^m N_k^T\Sigma^{-1}N_k + \sum_{i,j=1}^v H_i^T\Sigma^{-1}H_j k_{ij} & \Sigma^{-1}B \\ B^T\Sigma^{-1} & -I \end{bmatrix} \le 0.$$

(33)

We multiply (33) with $\left[\begin{smallmatrix} \mathbf{x}_+ \\ 2u \end{smallmatrix}\right]^T$ from the left and with $\left[\begin{smallmatrix} \mathbf{x}_+ \\ 2u \end{smallmatrix}\right]$ from the right. Hence,

$$
4\|u\|_2^2 \geq \mathbf{x}_+^T \left(A^T \Sigma^{-1} + \Sigma^{-1} A + \sum_{k=1}^m N_k^T \Sigma^{-1} N_k + \sum_{i,j=1}^v H_i^T \Sigma^{-1} H_j k_{ij} \right) \mathbf{x}_+
$$
$$
+ 4\mathbf{x}_+^T \Sigma^{-1} Bu. \tag{34}
$$

Applying this result to (32) yields

$$
\mathbb{E}\left[x_+^T(t)\Sigma^{-1}x_+(t)\right] \leq 4\|u\|_{L_t^2}^2 + \int_0^t \mathbb{E}\left[x_+^T \Sigma^{-1} x_+\right] \|u^0\|_2^2 \, ds
$$
$$
- \mathbb{E} \int_0^t 2 \left[\begin{smallmatrix} 0 \\ \bar{x}_2 \end{smallmatrix}\right]^T \Sigma^{-1}(A\mathbf{x}_+ + 2Bu)ds
$$
$$
- \mathbb{E} \int_0^t 2x_+^T \Sigma^{-1} \left[\begin{smallmatrix} 0 \\ c_0 \end{smallmatrix}\right] + \sum_{i,j=1}^v \left(2H_i\mathbf{x}_+ - \left[\begin{smallmatrix} 0 \\ c_i \end{smallmatrix}\right]\right)^T \Sigma^{-1} \left[\begin{smallmatrix} 0 \\ c_j \end{smallmatrix}\right] k_{ij} ds. \tag{35}
$$

We first of all see that $x_+^T \Sigma^{-1} \left[\begin{smallmatrix} 0 \\ c_0 \end{smallmatrix}\right] = x_2^T \Sigma_2^{-1} c_0$ using the partitions of x_+ and Σ. With the partition of H_i, we moreover have

$$
\left(2H_i\mathbf{x}_+ - \left[\begin{smallmatrix} 0 \\ c_i \end{smallmatrix}\right]\right)^T \Sigma^{-1} \left[\begin{smallmatrix} 0 \\ c_j \end{smallmatrix}\right] = \left(2H_i\mathbf{x}_+ - \left[\begin{smallmatrix} 0 \\ c_i \end{smallmatrix}\right]\right)^T \left[\begin{smallmatrix} 0 \\ \Sigma_2^{-1} c_j \end{smallmatrix}\right]
$$
$$
= \left(2\left[\, H_{i,21}\ H_{i,22}\,\right]\left(x + \left[\begin{smallmatrix} \bar{x} \\ \bar{x}_2 \end{smallmatrix}\right]\right) - c_i\right)^T \Sigma_2^{-1} c_j
$$
$$
= \left(2\left[\, H_{i,21}\ H_{i,22}\,\right] x + c_i\right)^T \Sigma_2^{-1} c_j.
$$

In addition, it holds that

$$
-2\left[\begin{smallmatrix} 0 \\ \bar{x}_2 \end{smallmatrix}\right]^T \Sigma^{-1}(A\mathbf{x}_+ + 2Bu) = -2\left[\, 0\ \bar{x}_2^T \Sigma_2^{-1}\,\right](A\mathbf{x}_+ + 2Bu)
$$
$$
= -2\bar{x}_2^T \Sigma_2^{-1}\left(\left[\, A_{21}\ A_{22}\,\right]\left(x + \left[\begin{smallmatrix} \bar{x} \\ \bar{x}_2 \end{smallmatrix}\right]\right) + 2B_2u\right)
$$
$$
= -2\bar{x}_2^T \Sigma_2^{-1}\left(\left[\, A_{21}\ A_{22}\,\right] x + B_2u\right).
$$

Plugging the above relations into (35) leads to

$$
\mathbb{E}\left[x_+^T(t)\Sigma^{-1}x_+(t)\right] \leq 4\|u\|_{L_t^2}^2 + \int_0^t \mathbb{E}\left[x_+^T \Sigma^{-1} x_+\right] \left\|u^0\right\|_2^2 \, ds
$$
$$
- \mathbb{E} \int_0^t 2\bar{x}_2^T \Sigma_2^{-1}(\left[\, A_{21}\ A_{22}\,\right] x + B_2u)ds
$$
$$
- \mathbb{E} \int_0^t 2x_2^T \Sigma_2^{-1} c_0 + \sum_{i,j=1}^v \left(2\left[\, H_{i,21}\ H_{i,22}\,\right] x + c_i\right)^T \Sigma_2^{-1} c_j k_{ij} ds. \tag{36}
$$

We add $2\mathbb{E}\int_0^t \sum_{i,j=1}^v c_i^T \Sigma_2^{-1} c_j k_{ij} \mathrm{d}s$ to the right side of (36) and preserve the inequality since this term is nonnegative due to Lemma 2 ('Appendix'). This results in

$$\mathbb{E}\left[x_+^T(t)\Sigma^{-1}x_+(t)\right] \le 4\left\|u\right\|_{L_t^2}^2 - \alpha_+(t) + \int_0^t \mathbb{E}\left[x_+^T(s)\Sigma^{-1}x_+(s)\right]\left\|u^0(s)\right\|_2^2 \mathrm{d}s.$$

Gronwall's inequality in Lemma 3 ("Appendix") yields

$$\mathbb{E}\left[x_+^T(t)\Sigma^{-1}x_+(t)\right]$$
$$\le 4\left\|u\right\|_{L_t^2}^2 - \alpha_+(t) + \int_0^t \left(4\left\|u\right\|_{L_s^2}^2 - \alpha_+(s)\right)\left\|u^0(s)\right\|_2^2 \exp\left(\int_s^t \left\|u^0(w)\right\|_2^2 \mathrm{d}w\right)\mathrm{d}s.$$
$$(37)$$

We find an estimate for the following expression:

$$\int_0^t \left\|u\right\|_{L_s^2}^2 \left\|u^0(s)\right\|_2^2 \exp\left(\int_s^t \left\|u^0(w)\right\|_2^2 \mathrm{d}w\right)\mathrm{d}s$$
$$\le \left\|u\right\|_{L_t^2}^2 \left[-\exp\left(\int_s^t \left\|u^0(w)\right\|_2^2 \mathrm{d}w\right)\right]_{s=0}^t$$
$$= \left\|u\right\|_{L_t^2}^2 \left(\exp\left(\int_0^t \left\|u^0(s)\right\|_2^2 \mathrm{d}s\right) - 1\right). \qquad (38)$$

Combining (37) with (38), we obtain

$$\alpha_+(t) + \int_0^t \alpha_+(s)\left\|u^0(s)\right\|_2^2 \exp\left(\int_s^t \left\|u^0(w)\right\|_2^2 \mathrm{d}w\right)\mathrm{d}s$$
$$\le 4\left\|u\right\|_{L_t^2}^2 \exp\left(\int_0^t \left\|u^0(s)\right\|_2^2 \mathrm{d}s\right). \qquad (39)$$

Comparing this result with (30) implies

$$\left(\mathbb{E}\left\|y - \bar{y}\right\|_{L_t^2}^2\right)^{\frac{1}{2}} \le 2\sigma \left\|u\right\|_{L_t^2} \exp\left(0.5\left\|u^0\right\|_{L_t^2}^2\right). \qquad (40)$$

4.2 Proof of Theorem 2

We make use of Eqs. (22a) and (23) in order to prove this bound. We set $\hat{\Sigma} = \begin{bmatrix} \Sigma_1 & \\ & \Sigma_2 \end{bmatrix}$ as a submatrix of Σ in (16). Lemma 1 ('Appendix') now yields

$$\mathbb{E}\left[\hat{x}_-^T(t)\hat{\Sigma}\hat{x}_-(t)\right] = 2\int_0^t \mathbb{E}\left[\hat{x}_-^T\hat{\Sigma}\left(\hat{A}\hat{\mathbf{x}}_- + \sum_{k=1}^m (\hat{N}_k\hat{\mathbf{x}}_- u_k) + \begin{bmatrix} 0 \\ \hat{c}_0 \end{bmatrix}\right)\right] \mathrm{d}s$$

$$+ \int_0^t \sum_{i,j=1}^v \mathbb{E}\left[\left(\hat{H}_i\hat{\mathbf{x}}_- + \begin{bmatrix} 0 \\ \hat{c}_i \end{bmatrix}\right)^T \hat{\Sigma}\left(\hat{H}_j\hat{\mathbf{x}}_- + \begin{bmatrix} 0 \\ \hat{c}_j \end{bmatrix}\right)\right] k_{ij}\mathrm{d}s. \tag{41}$$

We see that the right side of (41) contains the submatrices $\hat{A}$, $\hat{B}$, $\hat{H}$, $\hat{N}$ and $\hat{\Sigma}$. In order to be able to refer to the full matrix inequality (9), we find upper bounds for certain terms in the following involving the full matrices A, B, H, N and Σ. With the same estimate as in (25) and the control vector u^0 defined in (14), we have

$$\sum_{k=1}^m 2\hat{x}_-^T(s)\hat{\Sigma}\hat{N}_k\hat{\mathbf{x}}_-(s)u_k(s) \le \hat{x}_-^T(s)\hat{\Sigma}\hat{x}_-(s)\left\|u^0(s)\right\|_2^2 + \sum_{k=1}^m \hat{\mathbf{x}}_-^T(s)\hat{N}_k^T\hat{\Sigma}\hat{N}_k\hat{\mathbf{x}}_-(s).$$

Adding the term $\sum_{k=1}^m \left(\begin{bmatrix} N_{k,31} & N_{k,32} & N_{k,33} \end{bmatrix}\hat{\mathbf{x}}_-(s)\right)^T \Sigma_3 \begin{bmatrix} N_{k,31} & N_{k,32} & N_{k,33} \end{bmatrix}\hat{\mathbf{x}}_-(s)$ to the right side of this inequality results in

$$\sum_{k=1}^m 2\hat{x}_-^T(s)\hat{\Sigma}\hat{N}_k\hat{\mathbf{x}}_-(s)u_k(s) \le \hat{x}_-^T(s)\hat{\Sigma}\hat{x}_-(s)\left\|u^0(s)\right\|_2^2 + \sum_{k=1}^m \hat{\mathbf{x}}_-^T(s)N_k^T\Sigma N_k\hat{\mathbf{x}}_-(s). \tag{42}$$

Moreover, it holds that

$$\hat{\mathbf{x}}_-^T(A^T\Sigma + \Sigma A)\hat{\mathbf{x}}_- = 2\hat{\mathbf{x}}_-^T\Sigma A\hat{\mathbf{x}}_-$$

$$= 2\begin{bmatrix} x_1 - \bar{x}_r \\ x_2 - h_1 \end{bmatrix}^T \hat{\Sigma}\hat{A}\hat{\mathbf{x}}_- + 2(\bar{x}_3 - h_2)^T \Sigma_3 \begin{bmatrix} A_{31} & A_{32} & A_{33} \end{bmatrix}\hat{\mathbf{x}}_-.$$

We derive $\begin{bmatrix} A_{31} & A_{32} & A_{33} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ \bar{x}_3 \end{bmatrix} = -B_3 u$ by the definition of $\bar{x}_3$. Moreover, it can be seen from the second line of (20) that $\begin{bmatrix} A_{31} & A_{32} & A_{33} \end{bmatrix}\hat{\mathbf{x}}_- = 0$. Hence,

$$\hat{\mathbf{x}}_-^T(A^T\Sigma + \Sigma A)\hat{\mathbf{x}}_- = 2\hat{x}_-^T\hat{\Sigma}\hat{A}\hat{\mathbf{x}}_- - 2\begin{bmatrix} 0 \\ h_1 \end{bmatrix}^T \hat{\Sigma}\hat{A}\hat{\mathbf{x}}_-. \tag{43}$$

It remains to find a suitable upper bound related to the expression depending on $\hat{H}_i$. We first of all see that

$$\sum_{i,j=1}^v \left(\hat{H}_i\hat{\mathbf{x}}_- + \begin{bmatrix} 0 \\ \hat{c}_i \end{bmatrix}\right)^T \hat{\Sigma}\left(\hat{H}_j\hat{\mathbf{x}}_- + \begin{bmatrix} 0 \\ \hat{c}_j \end{bmatrix}\right) k_{ij}$$

$$= \hat{\mathbf{x}}_-^T \sum_{i,j=1}^v \hat{H}_i^T \hat{\Sigma}\hat{H}_j k_{ij}\hat{\mathbf{x}}_- + \sum_{i,j=1}^v \left(2\hat{H}_i\hat{\mathbf{x}}_- + \begin{bmatrix} 0 \\ \hat{c}_i \end{bmatrix}\right)^T \hat{\Sigma}\begin{bmatrix} 0 \\ \hat{c}_j \end{bmatrix} k_{ij}.$$

The term $\sum_{i,j=1}^{v}\left([\,H_{i,31}\ H_{i,32}\ H_{i,33}\,]\,\hat{\mathbf{x}}_-(s)\right)^T \Sigma_3\,[\,H_{j,31}\ H_{j,32}\ H_{j,33}\,]\,\hat{\mathbf{x}}_-(s)k_{ij}$ is nonnegative by Lemma 2 ('Appendix'). Adding this term to the right side of the above equation yields

$$
\sum_{i,j=1}^{v}\left(\hat{H}_i\hat{\mathbf{x}}_- + \begin{bmatrix}0\\\hat{c}_i\end{bmatrix}\right)^T \hat{\Sigma}\left(\hat{H}_j\hat{\mathbf{x}}_- + \begin{bmatrix}0\\\hat{c}_j\end{bmatrix}\right)k_{ij}
$$

$$
\leq \hat{\mathbf{x}}_-^T \sum_{i,j=1}^{v} H_i^T \Sigma H_j k_{ij}\hat{\mathbf{x}}_- + \sum_{i,j=1}^{v}\left(2\hat{H}_i\hat{\mathbf{x}}_- + \begin{bmatrix}0\\\hat{c}_i\end{bmatrix}\right)^T \hat{\Sigma}\begin{bmatrix}0\\\hat{c}_j\end{bmatrix}k_{ij}. \qquad (44)
$$

Applying (42), (43) and (44) to (41), results in

$$
\mathbb{E}\left[\hat{x}_-^T(t)\hat{\Sigma}\hat{x}_-(t)\right] \leq \mathbb{E}\int_0^t \hat{\mathbf{x}}_-^T\left(A^T\Sigma + \Sigma A + \sum_{k=1}^{m}N_k^T\Sigma N_k + \sum_{i,j=1}^{v}H_i^T\Sigma H_j k_{ij}\right)\hat{\mathbf{x}}_-\,ds
$$

$$
+ \mathbb{E}\int_0^t 2\hat{x}_-^T\hat{\Sigma}\begin{bmatrix}0\\\hat{c}_0\end{bmatrix} + \sum_{i,j=1}^{v}\left(2\hat{H}_i\hat{\mathbf{x}}_- + \begin{bmatrix}0\\\hat{c}_i\end{bmatrix}\right)^T \hat{\Sigma}\begin{bmatrix}0\\\hat{c}_j\end{bmatrix}k_{ij}\,ds
$$

$$
+ \int_0^t \mathbb{E}\left[\hat{x}_-^T\hat{\Sigma}\hat{x}_-\right]\|u^0\|_2^2\,ds + \mathbb{E}\int_0^t 2\begin{bmatrix}0\\h_1\end{bmatrix}^T \hat{\Sigma}\hat{A}\hat{\mathbf{x}}_-\,ds. \qquad (45)
$$

Using that $\hat{c}_i = [\,H_{i,21}\ H_{i,22}\ H_{i,23}\,]\begin{bmatrix}\bar{x}_r\\h_1\\h_2\end{bmatrix}$, we have

$$
\left(2\hat{H}_i\hat{\mathbf{x}}_- + \begin{bmatrix}0\\\hat{c}_i\end{bmatrix}\right)^T \hat{\Sigma}\begin{bmatrix}0\\\hat{c}_j\end{bmatrix} = \left(2\hat{H}_i\hat{\mathbf{x}}_- + \begin{bmatrix}0\\\hat{c}_i\end{bmatrix}\right)^T\begin{bmatrix}0\\\Sigma_2\hat{c}_j\end{bmatrix}
$$

$$
= \left(2\,[\,H_{i,21}\ H_{i,22}\ H_{i,23}\,]\left(\begin{bmatrix}x_1\\x_2\\\bar{x}_3\end{bmatrix} - \begin{bmatrix}\bar{x}_r\\h_1\\h_2\end{bmatrix}\right) + \hat{c}_i\right)^T \Sigma_2\hat{c}_j
$$

$$
= \left(2\,[\,H_{i,21}\ H_{i,22}\ H_{i,23}\,]\begin{bmatrix}x_1\\x_2\\\bar{x}_3\end{bmatrix} - \hat{c}_i\right)^T \Sigma_2\hat{c}_j. \qquad (46)
$$

It can be seen further that

$$2 \begin{bmatrix} 0 \\ h_1 \end{bmatrix}^T \hat{\Sigma} \hat{A} \hat{x}_- = 2 \begin{bmatrix} 0 & h_1^T \Sigma_2 \end{bmatrix} \hat{A} \hat{x}_- = 2 h_1^T \Sigma_2 \begin{bmatrix} A_{21} & A_{22} & A_{23} \end{bmatrix} \left(\begin{bmatrix} x_1 \\ x_2 \\ \bar{x}_3 \end{bmatrix} - \begin{bmatrix} \bar{x}_r \\ h_1 \\ h_2 \end{bmatrix} \right)$$

$$= 2 h_1^T \Sigma_2 \left(\begin{bmatrix} A_{21} & A_{22} & A_{23} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \bar{x}_3 \end{bmatrix} + B_2 u \right) \tag{47}$$

taking the first line of (20) into account. Inserting (46) and (47) into (45) and using the fact that $2\hat{x}_-^T \hat{\Sigma} \begin{bmatrix} 0 \\ \hat{c}_0 \end{bmatrix} = 2x_2 \Sigma_2 \hat{c}_0$ leads to

$$\mathbb{E} \left[\hat{x}_-^T(t) \hat{\Sigma} \hat{x}_-(t) \right] \leq \int_0^t \mathbb{E} \left[\hat{x}_-^T \hat{\Sigma} \hat{x}_- \right] \| u^0 \|_2^2 \, ds + \hat{\alpha}_-(t)$$

$$+ \mathbb{E} \int_0^t \hat{x}_-^T \left(A^T \Sigma + \Sigma A + \sum_{k=1}^m N_k^T \Sigma N_k + \sum_{i,j=1}^v H_i^T \Sigma H_j k_{ij} \right) \hat{x}_- ds, \tag{48}$$

where we set $\hat{\alpha}_-(t) := \mathbb{E} \int_0^t 2x_2^T \Sigma_2 \hat{c}_0 + \left(2 \begin{bmatrix} H_{i,21} & H_{i,22} & H_{i,23} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \bar{x}_3 \end{bmatrix} - \hat{c}_i \right)^T \Sigma_2 \hat{c}_j ds +$ $\mathbb{E} \int_0^t 2 h_1^T \Sigma_2 (\begin{bmatrix} A_{21} & A_{22} & A_{23} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \bar{x}_3 \end{bmatrix} + B_2 u) ds$. With (9) and (22b), we obtain

$$\mathbb{E} \left[\hat{x}_-^T(t) \hat{\Sigma} \hat{x}_-(t) \right] \leq \int_0^t \mathbb{E} \left[\hat{x}_-^T \hat{\Sigma} \hat{x}_- \right] \left\| u^0 \right\|_2^2 ds + \hat{\alpha}_-(t) - \mathbb{E} \| \bar{y} - \bar{y}_r \|_{L_t^2}^2 .$$

Applying Lemma 3 ('Appendix') to this inequality yields

$$\mathbb{E} \left[\hat{x}_-^T(t) \hat{\Sigma} \hat{x}_-(t) \right] \leq \hat{\alpha}_-(t) - \mathbb{E} \| \bar{y} - \bar{y}_r \|_{L_t^2}^2$$

$$+ \int_0^t \hat{\alpha}_-(s) \left\| u^0(s) \right\|_2^2 \exp \left(\int_s^t \left\| u^0(w) \right\|_2^2 dw \right) ds.$$

Since the above left side of the inequality is positive, we obtain

$$\mathbb{E} \| \bar{y} - \bar{y}_r \|_{L_t^2}^2$$
$$\leq \hat{\alpha}_-(t) + \int_0^t \hat{\alpha}_-(s) \left\| u^0(s) \right\|_2^2 \exp \left(\int_s^t \left\| u^0(w) \right\|_2^2 dw \right) ds.$$

We exploit that $\Sigma_2 = \sigma I$. Hence, we have

$$\mathbb{E} \| \bar{y} - \bar{y}_r \|_{L_t^2}^2$$
$$\leq \sigma^2 \left(\hat{\alpha}_+(t) + \int_0^t \hat{\alpha}_+(s) \left\| u^0(s) \right\|_2^2 \exp \left(\int_s^t \left\| u^0(w) \right\|_2^2 dw \right) ds \right), \tag{49}$$

where we set $\hat{\alpha}_+(t) := \mathbb{E} \int_0^t 2x_2^T \Sigma_2^{-1} \hat{c}_0 + \left(2 \begin{bmatrix} H_{i,21} & H_{i,22} & H_{i,23} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \bar{x}_3 \end{bmatrix} - \hat{c}_i \right)^T \Sigma_2^{-1} \hat{c}_j ds +$ $\mathbb{E} \int_0^t 2 h_1^T \Sigma_2^{-1} (\begin{bmatrix} A_{21} & A_{22} & A_{23} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \bar{x}_3 \end{bmatrix} + B_2 u) ds$. In order to find a suitable bound for the

right side of (49), Ito's lemma is applied to $\mathbb{E}[\hat{x}_+^T(t)\hat{\Sigma}^{-1}\hat{x}_+(t)]$. Due to (23) and Lemma 1 ('Appendix'), we obtain

$$\mathbb{E}\left[\hat{x}_+^T(t)\hat{\Sigma}^{-1}\hat{x}_+(t)\right] = 2\int_0^t \mathbb{E}\left[\hat{x}_+^T\hat{\Sigma}^{-1}\left(\hat{A}\hat{\mathbf{x}}_+ + 2\hat{B}u + \sum_{k=1}^m (\hat{N}_k\hat{\mathbf{x}}_+u_k) - \left[\begin{smallmatrix} 0 \\ \hat{c}_0 \end{smallmatrix}\right]\right)\right]ds$$

$$+ \int_0^t \sum_{i,j=1}^v \mathbb{E}\left[\left(\hat{H}_i\hat{\mathbf{x}}_+ - \left[\begin{smallmatrix} 0 \\ \hat{c}_i \end{smallmatrix}\right]\right)^T\hat{\Sigma}^{-1}\left(\hat{H}_j\hat{\mathbf{x}}_+ - \left[\begin{smallmatrix} 0 \\ \hat{c}_j \end{smallmatrix}\right]\right)\right]k_{ij}ds. \tag{50}$$

Analogously to (42), it holds that

$$\sum_{k=1}^m 2\hat{x}_+^T(s)\hat{\Sigma}^{-1}\hat{N}_k\hat{\mathbf{x}}_+(s)u_k(s)$$

$$\leq \hat{x}_+^T(s)\hat{\Sigma}^{-1}\hat{x}_+(s)\left\|u^0(s)\right\|_2^2 + \sum_{k=1}^m \hat{\mathbf{x}}_+^T(s)\hat{N}_k^T\hat{\Sigma}^{-1}\hat{N}_k\hat{\mathbf{x}}_+(s)$$

$$\leq \hat{x}_+^T(s)\hat{\Sigma}^{-1}\hat{x}_+(s)\left\|u^0(s)\right\|_2^2 + \sum_{k=1}^m \hat{\mathbf{x}}_+^T(s)N_k^T\Sigma^{-1}N_k\hat{\mathbf{x}}_+(s). \tag{51}$$

Furthermore, we see that

$$\hat{\mathbf{x}}_+^T(A^T\Sigma^{-1} + \Sigma^{-1}A)\hat{\mathbf{x}}_+ + 4\hat{\mathbf{x}}_+^T\Sigma^{-1}Bu = 2\hat{\mathbf{x}}_+^T\Sigma^{-1}(A\hat{\mathbf{x}}_+ + 2Bu)$$

$$= 2\left[\begin{smallmatrix} x_1 + \bar{x}_r \\ x_2 + h_1 \end{smallmatrix}\right]^T\hat{\Sigma}^{-1}(\hat{A}\hat{\mathbf{x}}_+ + 2\hat{B}u) + 2(\bar{x}_3 + h_2)^T\Sigma_3^{-1}\left(\left[\begin{smallmatrix} A_{31} & A_{32} & A_{33} \end{smallmatrix}\right]\hat{\mathbf{x}}_+ + 2B_3u\right).$$

Since $\left[\begin{smallmatrix} A_{31} & A_{32} & A_{33} \end{smallmatrix}\right]\left[\begin{smallmatrix} x_1 \\ x_2 \\ \bar{x}_3 \end{smallmatrix}\right] = \left[\begin{smallmatrix} A_{31} & A_{32} & A_{33} \end{smallmatrix}\right]\left[\begin{smallmatrix} \bar{x}_r \\ h_1 \\ h_2 \end{smallmatrix}\right] = -B_3u$ by the definition of $\bar{x}_3$ and the second line of (20), we obtain $\left[\begin{smallmatrix} A_{31} & A_{32} & A_{33} \end{smallmatrix}\right]\hat{\mathbf{x}}_+ = -2B_3u$. Thus,

$$\hat{\mathbf{x}}_+^T(A^T\Sigma^{-1} + \Sigma^{-1}A)\hat{\mathbf{x}}_+ + 4\hat{\mathbf{x}}_+^T\Sigma^{-1}Bu$$

$$= 2\hat{x}_+^T\hat{\Sigma}^{-1}(\hat{A}\hat{\mathbf{x}}_+ + 2\hat{B}u) + 2\left[\begin{smallmatrix} 0 \\ h_1 \end{smallmatrix}\right]^T\hat{\Sigma}^{-1}(\hat{A}\hat{\mathbf{x}}_+ + 2\hat{B}u). \tag{52}$$

Finally, we see that

$$\sum_{i,j=1}^v \left(\hat{H}_i\hat{\mathbf{x}}_+ - \left[\begin{smallmatrix} 0 \\ \hat{c}_i \end{smallmatrix}\right]\right)^T\hat{\Sigma}^{-1}\left(\hat{H}_j\hat{\mathbf{x}}_+ - \left[\begin{smallmatrix} 0 \\ \hat{c}_j \end{smallmatrix}\right]\right)k_{ij}$$

$$= \hat{\mathbf{x}}_+^T\sum_{i,j=1}^v \hat{H}_i^T\hat{\Sigma}^{-1}\hat{H}_jk_{ij}\hat{\mathbf{x}}_+ - \sum_{i,j=1}^v \left(2\hat{H}_i\hat{\mathbf{x}}_+ - \left[\begin{smallmatrix} 0 \\ \hat{c}_i \end{smallmatrix}\right]\right)^T\hat{\Sigma}^{-1}\left[\begin{smallmatrix} 0 \\ \hat{c}_j \end{smallmatrix}\right]k_{ij}$$

$$\leq \hat{\mathbf{x}}_+^T\sum_{i,j=1}^v H_i^T\Sigma^{-1}H_jk_{ij}\hat{\mathbf{x}}_+ - \sum_{i,j=1}^v \left(2\hat{H}_i\hat{\mathbf{x}}_+ - \left[\begin{smallmatrix} 0 \\ \hat{c}_i \end{smallmatrix}\right]\right)^T\hat{\Sigma}^{-1}\left[\begin{smallmatrix} 0 \\ \hat{c}_j \end{smallmatrix}\right]k_{ij} \tag{53}$$

applying Lemma 2 ('Appendix'). With (51), (52) and (53) inequality (50) becomes

$$\mathbb{E}\left[\hat{x}_+^T(t)\hat{\Sigma}^{-1}\hat{x}_+(t)\right]$$

$$\leq \mathbb{E}\int_0^t \hat{\mathbf{x}}_+^T\left(A^T\Sigma^{-1} + \Sigma^{-1}A + \sum_{k=1}^m N_k^T\Sigma^{-1}N_k + \sum_{i,j=1}^v H_i^T\Sigma^{-1}H_jk_{ij}\right)\hat{\mathbf{x}}_+\mathrm{d}s$$

$$- \mathbb{E}\int_0^t 2\hat{x}_+^T\hat{\Sigma}^{-1}\begin{bmatrix}0\\\hat{c}_0\end{bmatrix} + \sum_{i,j=1}^v\left(2\hat{H}_i\hat{\mathbf{x}}_+ - \begin{bmatrix}0\\\hat{c}_i\end{bmatrix}\right)^T\hat{\Sigma}^{-1}\begin{bmatrix}0\\\hat{c}_j\end{bmatrix}k_{ij}\mathrm{d}s$$

$$- \mathbb{E}\int_0^t 2\begin{bmatrix}0\\h_1\end{bmatrix}^T\hat{\Sigma}^{-1}(\hat{A}\hat{\mathbf{x}}_+ + 2\hat{B}u)\mathrm{d}s + \mathbb{E}\int_0^t 4\hat{\mathbf{x}}_+^T\Sigma^{-1}Bu\,\mathrm{d}s$$

$$+ \int_0^t \mathbb{E}\left[\hat{x}_+^T\hat{\Sigma}^{-1}\hat{x}_+\right]\left\|u^0\right\|_2^2\mathrm{d}s. \tag{54}$$

Similarly to (34), we obtain

$$4\|u\|_2^2 \geq \hat{\mathbf{x}}_+^T\left(A^T\Sigma^{-1} + \Sigma^{-1}A + \sum_{k=1}^m N_k^T\Sigma^{-1}N_k + \sum_{i,j=1}^v H_i^T\Sigma^{-1}H_jk_{ij}\right)\hat{\mathbf{x}}_+ + 4\hat{\mathbf{x}}_+^T\Sigma^{-1}Bu.$$

This leads to

$$\mathbb{E}\left[\hat{x}_+^T(t)\hat{\Sigma}^{-1}\hat{x}_+(t)\right]$$

$$\leq 4\|u\|_{L_t^2}^2 + \int_0^t \mathbb{E}\left[\hat{x}_+^T\hat{\Sigma}^{-1}\hat{x}_+\right]\left\|u^0\right\|_2^2\mathrm{d}s$$

$$- \mathbb{E}\int_0^t 2\hat{x}_+^T\hat{\Sigma}^{-1}\begin{bmatrix}0\\\hat{c}_0\end{bmatrix} + \sum_{i,j=1}^v\left(2\hat{H}_i\hat{\mathbf{x}}_+ - \begin{bmatrix}0\\\hat{c}_i\end{bmatrix}\right)^T\hat{\Sigma}^{-1}\begin{bmatrix}0\\\hat{c}_j\end{bmatrix}k_{ij}\mathrm{d}s$$

$$- \mathbb{E}\int_0^t 2\begin{bmatrix}0\\h_1\end{bmatrix}^T\hat{\Sigma}^{-1}(\hat{A}\hat{\mathbf{x}}_+ + 2\hat{B}u)\mathrm{d}s. \tag{55}$$

In the following (55) is expressed by terms depending on Σ_2. We obtain $\hat{x}_+^T\hat{\Sigma}^{-1}\begin{bmatrix}0\\\hat{c}_0\end{bmatrix} = x_2^T\Sigma_2^{-1}\hat{c}_0$ exploiting the partitions of $\hat{x}_+$ and $\hat{\Sigma}$. The terms depending on $\hat{H}_i$ become

$$- \sum_{i,j=1}^v\left(2\hat{H}_i\hat{\mathbf{x}}_+ - \begin{bmatrix}0\\\hat{c}_i\end{bmatrix}\right)^T\hat{\Sigma}^{-1}\begin{bmatrix}0\\\hat{c}_j\end{bmatrix}k_{ij}$$

$$= - \sum_{i,j=1}^v\left(2\hat{H}_i\hat{\mathbf{x}}_+ - \begin{bmatrix}0\\\hat{c}_i\end{bmatrix}\right)^T\begin{bmatrix}0\\\Sigma_2^{-1}\hat{c}_j\end{bmatrix}k_{ij}$$

$$= -\sum_{i,j=1}^{v} \left(2 \left[H_{i,21} \; H_{i,22} \; H_{i,23} \right] \left(\begin{bmatrix} x_1 \\ x_2 \\ \bar{x}_3 \end{bmatrix} + \begin{bmatrix} \bar{x}_r \\ h_1 \\ h_2 \end{bmatrix} \right) - \hat{c}_i \right)^T \Sigma_2^{-1} \hat{c}_j k_{ij}$$

$$= -\sum_{i,j=1}^{v} \left(2 \left[H_{i,21} \; H_{i,22} \; H_{i,23} \right] \begin{bmatrix} x_1 \\ x_2 \\ \bar{x}_3 \end{bmatrix} + \hat{c}_i \right)^T \Sigma_2^{-1} \hat{c}_j k_{ij}$$

$$\le -\sum_{i,j=1}^{v} \left(2 \left[H_{i,21} \; H_{i,22} \; H_{i,23} \right] \begin{bmatrix} x_1 \\ x_2 \\ \bar{x}_3 \end{bmatrix} - \hat{c}_i \right)^T \Sigma_2^{-1} \hat{c}_j k_{ij} \qquad (56)$$

adding $2\sum_{i,j=1}^{v} \hat{c}_i^T \Sigma_2^{-1} \hat{c}_j k_{ij}$ which is positive due to Lemma 2 ('Appendix'). Furthermore, using the first line of (20), it holds that

$$-2 \begin{bmatrix} 0 \\ h_1 \end{bmatrix}^T \hat{\Sigma}^{-1} (\hat{A}\hat{\mathbf{x}}_+ + 2\hat{B}u) = -2 \left[0 \; h_1^T \Sigma_2^{-1} \right] (\hat{A}\hat{\mathbf{x}}_+ + 2\hat{B}u)$$

$$= -2 h_1^T \Sigma_2^{-1} \left(\left[A_{21} \; A_{22} \; A_{23} \right] \left(\begin{bmatrix} x_1 \\ x_2 \\ \bar{x}_3 \end{bmatrix} + \begin{bmatrix} \bar{x}_r \\ h_1 \\ h_2 \end{bmatrix} \right) + 2B_2 u \right)$$

$$= -2 h_1^T \Sigma_2^{-1} \left(\left[A_{21} \; A_{22} \; A_{23} \right] \begin{bmatrix} x_1 \\ x_2 \\ \bar{x}_3 \end{bmatrix} + B_2 u \right). \qquad (57)$$

We insert (56) and (57) into (55) and obtain

$$\mathbb{E}\left[\hat{x}_+^T(t) \hat{\Sigma}^{-1} \hat{x}_+(t) \right] \le 4 \|u\|_{L_t^2}^2 + \int_0^t \mathbb{E}\left[\hat{x}_+^T \hat{\Sigma}^{-1} \hat{x}_+ \right] \left\| u^0 \right\|_2^2 ds - \hat{\alpha}_+(t).$$

With Lemma 3 ('Appendix'), analogously to (39), we find

$$\hat{\alpha}_+(t) + \int_0^t \hat{\alpha}_+(s) \left\| u^0(s) \right\|_2^2 \exp\left(\int_s^t \left\| u^0(w) \right\|_2^2 dw \right) ds$$

$$\le 4 \|u\|_{L_t^2}^2 \exp\left(\int_0^t \left\| u^0(s) \right\|_2^2 ds \right). \qquad (58)$$

The relations (49) and (58) yield the claim.

5 Numerical experiments

We conduct a numerical experiment in order to compare several MOR techniques and to check the performance of the error bound in Theorem 3. We determine three different ROMs. One is by SPA stated in (7). The corresponding output is denoted by $\bar{y}_{SPA}$. Moreover, we study a structure preserving version of SPA that is obtained by setting $B_2 = 0$ in (7), i.e., $(\bar{B}, \bar{D}, \bar{E}_k, \bar{F}_i) = (B_1, 0, 0, 0)$. This technique is denoted by SPA2 and its output is written as $\bar{y}_{SPA2}$. Notice that this method is a generalization of the one in [18]. Finally, we deal with BT [30], another structure preserving scheme. The respective output is $\bar{y}_{BT}$.

In particular, we apply the different MOR variants to a heat transfer problem that was proposed in [30]. We consider a heat equation on $[0, 1]^2$:

$$\frac{\partial}{\partial t} X(t, \zeta) = \Delta X(t, \zeta), \quad \zeta \in [0, 1]^2, \quad t \in [0, T],$$

with Dirichlet and noisy Robin boundary conditions

$$X(t, \zeta) = u(t) \quad \text{on} \quad \Gamma_L := \{0\} \times (0, 1),$$
$$\frac{\partial}{\partial \mathbf{n}} X(t, \zeta) = \frac{1}{\sqrt{2}} (u(t) + \dot{w}(t)) X(t, \zeta) \quad \text{on} \quad \Gamma_R := \{1\} \times [0, 1],$$
$$X(t, \zeta) = 0 \quad \text{on} \quad \partial [0, 1]^2 \setminus (\Gamma_L \cup \Gamma_R), \quad t \in [0, T],$$

where $u \in L_T^2$ is a scalar deterministic input and w denotes a scalar standard Wiener process. We discretize the heat equation with a finite difference scheme on an equidistant $\tilde{n} \times \tilde{n}$-mesh. This leads to an $n = \tilde{n}^2$-dimensional stochastic bilinear system

$$dx(t) = \left[Ax(t) + Bu(t) + \frac{1}{\sqrt{2}} Nx(t)u(t) \right] dt + \frac{1}{\sqrt{2}} Nx(t) dw(t),$$
$$y(t) = Cx(t), \quad t \in [0, T], \tag{59}$$

where $C = \frac{1}{n}[1 \ 1 \ \dots \ 1]$, i.e., the average temperature is considered. We refer to [5] or [12] for more details on the matrices A, B and N. There, a similar example was investigated for deterministic bilinear systems and linear stochastic systems, respectively.

According to (3) and (4), the associated Gramians are the solutions to

$$A^T P^{-1} + P^{-1} A + N^T P^{-1} N \leq -P^{-1} B B^T P^{-1},$$
$$A^T Q + Q A + N^T Q N = -C^T C. \tag{60}$$

We multiply (60) with P from the left and the right. Applying the Schur complement condition on definiteness, (60) can then be equivalently written as the following linear matrix inequality:

$$\begin{bmatrix} AP + PA^T + BB^T & PN^T \\ NP & -P \end{bmatrix} \leq 0, \tag{61}$$

see also [12, Remark III.2]. The matrix inequality (61) now is solved using the LMI-solver YALMIP [24] minimizing $\text{tr}(P)$. LMI-solver are generally not suitable in a large-scale setting. Therefore, we choose $\tilde{n} = 10$ implying $n = 100$.

As in [30], set $T = 2$ and choose two different controls $u(t) = \tilde{u}(t), \hat{u}(t)$, where $\tilde{u}(t) = \cos(\pi t)$ and $\hat{u}(t) = \frac{1}{\sqrt{2}}$, $t \in [0, T]$. We derive the ROMs using SPA, a modified structure preserving version SPA2 and BT based on Q and P. We determine the reduced systems for $r = 3, 6, 9$. We have an error bound for

Table 1 L_T^2-error of SPA with ROM defined in (7) and the respective error bound for different reduced order dimensions r, $T = 2$ and outputs $u = \tilde{u}, \hat{u}$

r	$\sqrt{\mathbb{E}\,\|y - \bar{y}_{SPA}\|_{L_T^2}^2}, u = \tilde{u}$	$\sqrt{\mathbb{E}\,\|y - \bar{y}_{SPA}\|_{L_T^2}^2}, u = \hat{u}$	$\mathcal{EB}_r$
3	1.79e−04	1.36e−04	6.01e−03
6	9.45e−07	6.53e−07	7.57e−05
9	1.64e−08	1.15e−08	2.96e−06

Table 2 L_T^2-error of BT studied in [30] and the respective error bound for different reduced order dimensions r, $T = 2$ and outputs $u = \tilde{u}, \hat{u}$

r	$\sqrt{\mathbb{E}\,\|y - \bar{y}_{BT}\|_{L_T^2}^2}, u = \tilde{u}$	$\sqrt{\mathbb{E}\,\|y - \bar{y}_{BT}\|_{L_T^2}^2}, u = \hat{u}$	$\mathcal{EB}_r$
3	9.75e−04	1.04e−03	6.01e−03
6	6.46e−06	3.46e−06	7.57e−05
9	1.25e−07	7.28e−08	2.96e−06

Table 3 L_T^2-error of SPA2 with ROM that is obtained by setting $B_2 = 0$ in (7) for different reduced order dimensions r, $T = 2$ and outputs $u = \tilde{u}, \hat{u}$

r	$\sqrt{\mathbb{E}\,\|y - \bar{y}_{SPA2}\|_{L_T^2}^2}, u = \tilde{u}$	$\sqrt{\mathbb{E}\,\|y - \bar{y}_{SPA2}\|_{L_T^2}^2}, u = \hat{u}$
3	2.34e−03	2.42e−03
6	1.73e−05	1.86e−05
9	1.52e−07	1.57e−07

SPA (Theorem 3) and BT [30] but none for SPA2. The bound for BT and SPA is $\mathcal{EB}_r := 2\left(\sum_{i=r+1}^{100} \sigma_i\right) \|u\|_{L_T^2} \exp\left(0.5 \|u\|_{L_T^2}^2\right)$. Notice that $u^0 \equiv u$ in this example. We compute $\sqrt{\mathbb{E}\,\|y - \bar{y}_l\|_{L_T^2}^2}$ for $l = SPA, BT, SPA2$ in Tables 1, 2 and 3.

We can see by looking at Tables 1 and 3 that SPA performs clearly better than the structure preserving variant SPA2. This tells us that it is worth to allow a structure change since this can lead to better approximations. We can also see that the error bound for SPA is relatively tight. It is tighter for BT, compare with Table 2. However, this also means that BT performs worse than SPA. Consequently, SPA is the best choice for the example considered here.

6 Conclusions

In this paper, we investigated a large-scale stochastic bilinear system. In order to reduce the state space dimension, a model order reduction technique called singular perturbation approximation was extended to this setting. This method is based on

Gramians proposed in [30] that characterize how much a state contributes to the system dynamics or to the output of a system. This choice of Gramians as well as the structure of the reduced system is different than in [18]. With this modification, we provided a new L^2-error bound that can be used to point out the cases in which the reduced order model by singular perturbation approximation delivers a good approximation to the original model. This error bound is new even for deterministic bilinear systems. Its quality was tested in a numerical experiment.

A Supporting Lemmas

In this appendix, we state three important results and the corresponding references that we frequently use throughout this paper.

Lemma 1 *Let $a, b_1, \ldots, b_v$ be $\mathbb{R}^d$-valued processes, where a is $(\mathcal{F}_t)_{t\geq 0}$-adapted and almost surely Lebesgue integrable and the functions b_i are integrable with respect to the mean zero square integrable Lévy process $M = (M_1, \ldots, M_v)^T$ with covariance matrix $K = (k_{ij})_{i,j=1,\ldots,v}$. If the process x is given by*

$$\mathrm{d}x(t) = a(t)\mathrm{d}t + \sum_{i=1}^{v} b_i(t)\mathrm{d}M_i,$$

then, we have

$$\frac{d}{\mathrm{d}t}\mathbb{E}\left[x^T(t)x(t)\right] = 2\mathbb{E}\left[x^T(t)a(t)\right] + \sum_{i,j=1}^{v}\mathbb{E}\left[b_i^T(t)b_j(t)\right]k_{ij}.$$

Proof We refer to [32, Lemma 5.2] for a proof of this lemma. □

Lemma 2 *Let $A_1, \ldots, A_v$ be $d_1 \times d_2$ matrices and $K = (k_{ij})_{i,j=1,\ldots,v}$ be a positive semidefinite matrix, then*

$$\tilde{K} := \sum_{i,j=1}^{v} A_i^T A_j k_{ij}$$

is also positive semidefinite.

Proof The proof can be found in [32, Proposition 5.3].   □

Lemma 3 (Gronwall lemma) *Let $T > 0$, $z, \alpha : [0, T] \to \mathbb{R}$ be measurable bounded functions and $\beta : [0, T] \to \mathbb{R}$ be a nonnegative integrable function. If*

$$z(t) \leq \alpha(t) + \int_0^t \beta(s)z(s)\mathrm{d}s,$$

then it holds that

$$z(t) \leq \alpha(t) + \int_0^t \alpha(s)\beta(s) \exp\left(\int_s^t \beta(w)\mathrm{d}w\right) \mathrm{d}s$$

for all $t \in [0, T]$.

Proof The result is shown as in [13, Proposition 2.1].   □

References

1. Al-Baiyat SA, Bettayeb M (1993) A new model reduction scheme for k–power bilinear systems. In: Proceedings of the 32nd IEEE conference on decision and control, pp 22–27
2. Allen EJ, Novosel SJ, Zhang Z (1998) Finite element and difference approximation of some linear stochastic partial differential equations. Stochast Stochast Rep 64(1–2):117–142
3. Antoulas AC (2005) Approximation of large-scale dynamical systems. Advances in design and control 6. SIAM, Philadelphia
4. Becker S, Hartmann C (2019) Infinite-dimensional bilinear and stochastic balanced truncation with error bounds. Math Control Signals Syst 31(2):1–37
5. Benner P, Damm T (2011) Lyapunov equations, energy functionals, and model order reduction of bilinear and stochastic systems. SIAM J Control Optim 49(2):686–711
6. Benner P, Damm T, Rodriguez Cruz YR (2017) Dual pairs of generalized Lyapunov inequalities and balanced truncation of stochastic linear systems. IEEE Trans Autom Contr 62(2):782–791
7. Benner P, Goyal P (2017) Balanced truncation model order reduction for quadratic-bilinear control systems. Technical report. arXiv preprint arXiv:1705.00160
8. Benner P, Redmann M (2015) Model reduction for stochastic systems. Stoch PDE Anal Comp 3(3):291–338
9. Berglund N, Gentz B (2006) Noise-induced phenomena in slow-fast dynamical systems. A sample-paths approach. Springer, London
10. Bruni C, DiPillo G, Koch G (1971) On the mathematical models of bilinear systems. Automatica 2(1):11–26
11. Da Prato G, Zabczyk J (1992) Stochastic equations in infinite dimensions. Encyclopedia of mathematics and its applications, vol 44. Cambridge University Press, Cambridge
12. Damm T, Benner P (2014) Balanced truncation for stochastic linear systems with guaranteed error bound. In: Proceedings of MTNS–2014, Groningen, The Netherlands, pp 1492–1497
13. Emmrich E (1999) Discrete versions of Gronwall's lemma and their application to the numerical analysis of parabolic problems. Preprint No. 637, TU Berlin
14. Fernando KV, Nicholson H (1982) Singular perturbational model reduction of balanced systems. IEEE Trans Automat Control 27:466–468
15. Gawarecki L, Mandrekar V (2011) Stochastic differential equations in infinite dimensions with applications to stochastic partial differential equations. Springer, Berlin
16. Gray WS, Mesko J (1998) Energy functions and algebraic gramians for bilinear systems. Proc Fourth IFAC Nonlinear Control Syst Des Symp 31(17):101–106

17. Grecksch W, Kloeden PE (1996) Time-discretised Galerkin approximations of parabolic stochastic PDEs. Bull Aust Math Soc 54(1):79–85
18. Hartmann C, Schäfer-Bung B, Thöns-Zueva A (2013) Balanced averaging of bilinear systems with applications to stochastic control. SIAM J Control Optim 51(3):2356–2378
19. Hausenblas E (2003) Approximation for semilinear stochastic evolution equations. Pot Anal 18(2):141–186
20. Jentzen A, Kloeden PE (2009) Overcoming the order barrier in the numerical approximation of stochastic partial differential equations with additive space-time noise. Proc R Soc A 2009(465):649–667
21. Kovács M, Larsson S, Saedpanah F (2010) Finite element approximation of the linear stochastic wave equation with additive noise. SIAM J Numer Anal 48(2):408–427
22. Kruse R (2014) Strong and weak approximation of semilinear stochastic evolution equations. Lecture notes in mathematics, vol 2093. Springer, Berlin
23. Liu Y, Anderson BDO (1989) Singular perturbation approximation of balanced systems. Int J Control 50(4):1379–1405
24. Löfberg J (2004) YALMIP: a toolbox for modeling and optimization in MATLAB. In: Proceedings of the CACSD conference, Taipei, Taiwan
25. Mohler RR (1973) Bilinear control processes. Academic Press, New York
26. Moore BC (1981) Principal component analysis in linear systems: controllability, observability, and model reduction. IEEE Trans Autom Control 26:17–32
27. Obinata G, Anderson BDO (2001) Model reduction for control system design. Communications and control engineering series. Springer, London
28. Peszat S, Zabczyk J (2007) Stochastic partial differential equations with Lévy noise. An evolution equation approach. Encyclopedia of mathematics and its applications, vol 113. Cambridge University Press, Cambridge
29. Prévôt C, Röckner M (2007) A concise course on stochastic partial differential equations. Lecture notes in mathematics, vol 1905. Springer, Berlin
30. Redmann M (2018) Energy estimates and model order reduction for stochastic bilinear systems. Int J Control. https://doi.org/10.1080/00207179.2018.1538568
31. Redmann M (2018) Type II balanced truncation for deterministic bilinear control systems. SIAM J Control Optim 56(4):2593–2612
32. Redmann M (2018) Type II singular perturbation approximation for linear systems with Lévy noise. SIAM J Control Optim 56(3):2120–2158
33. Redmann M, Benner P (2018) Singular perturbation approximation for linear systems with Lévy noise. Stochast Dyn 18(4):1850033
34. Rugh WJ (1981) Nonlinear system theory. The Johns Hopkins University Press, Baltimore
35. Scherpen JMA (1993) Balancing for nonlinear systems. Syst Control Lett 21:143–153
36. Thompson WF, Kuske RA, Monahan AH (2014) Stochastic averaging of dynamical systems with multiple time scales forced with α-stable noise. SIAM Multisc Model Simul 13(4):1194–1223
37. Xu Y, Duan J, Xu W (2011) An averaging principle for stochastic dynamical systems with Lévy noise. Phys D Nonlinear Phen 240(17):1395–1401