# Model order reduction for bilinear systems with non-zero initial states – different approaches with error bounds

Martin Redmann* Igor Pontes Duff †

July 12, 2021

**Abstract**

In this paper, we consider model order reduction for bilinear systems with non-zero initial conditions. We discuss choices of Gramians for both the homogeneous and the inhomogeneous parts of the system individually and prove how these Gramians characterize the respective dominant subspaces of each of the two subsystems. Proposing different, not necessarily structure preserving, reduced order methods for each subsystem, we establish several strategies to reduce the dimension of the full system. For all these approaches, error bounds are shown depending on the truncated Hankel singular values of the subsystems. Besides the error analysis, stability is discussed. In particular, a focus is on a new criterion for the homogeneous subsystem guaranteeing the existence of the associated Gramians and an asymptotically stable realization of the system.

**Keywords:** Model order reduction, bilinear systems, error bounds, stability analysis
**MSC classification:** 65L05, 93A15, 93C10, 93D20

## 1 Introduction

In this paper, we study model order reduction (MOR) techniques for the following system with non-zero initial states:

$$\dot{x}(t) = Ax(t) + Bu(t) + \sum_{k=1}^{m} N_k x(t) u_k(t), \quad x(0) = x_0 = X_0 v_0, \tag{1a}$$

$$y(t) = Cx(t), \quad t \geq 0, \tag{1b}$$

where $A, N_k \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, and $C \in \mathbb{R}^{p\times n}$. Moreover, $x$ is the state vector, $y$ the quantity of interest and the columns of $X_0 \in \mathbb{R}^{n\times q}$ span all initial states $x_0$ that are considered here, i.e., there exist $v_0 \in \mathbb{R}^q$ such that $x_0 = X_0 v_0$. We assume that the

*Martin Luther University Halle-Wittenberg, Institute of Mathematics, Theodor-Lieser-Str. 5, 06120 Halle (Saale), Germany, Email: martin.redmann@mathematik.uni-halle.de.

†Max Planck Institute for Dynamics of Complex Technical Systems, Magdeburg, Germany, Email: pontes@mpi-magdeburg.mpg.de.

matrix $A$ is Hurwitz, meaning that $\sigma(A) \subset \mathbb{C}_- = \{z \in \mathbb{C} : \quad \Re(z) < 0\}$, where $\sigma(\cdot)$ denotes the spectrum of a matrix and $\Re(\cdot)$ represents the real part of a complex number. Furthermore, let $u \in L^2$, i.e.,

$$\|u\|_{L^2}^2 := \int_0^\infty \|u(s)\|_2^2 \, ds = \int_0^\infty u^\top(s)u(s)ds < \infty.$$

There exist many different MOR techniques for bilinear systems when $x_0 = 0$, e.g., methods that are balancing related [1, 14, 19, 21], optimization/interpolation based [5, 13] and data-driven [2]. However, many applications involve non-zero initial states such that a study for MOR for (1) is essential. Several approaches in this context have been established for linear systems [3, 4, 12, 15, 25]. There is no straightforward generalization of these techniques to (1), since the study of transfer functions and fundamental solutions is much more involved for bilinear systems. In this work, choose an approach that relies on estimates for fundamental solutions of bilinear systems that originate in [22]. These estimates enable a detailed theoretical analysis for an ansatz that conceptionally extends the one used in [4]. The general idea is to split (1) into two subsystems. System

$$\dot{x}_{x_0}(t) = Ax_{x_0}(t) + \sum_{k=1}^m N_k x_{x_0}(t)u_k(t), \quad x_{x_0}(0) = X_0 v_0, \tag{2a}$$

$$y_{x_0}(t) = Cx_{x_0}(t). \tag{2b}$$

involves the initial condition and

$$\dot{x}_B(t) = Ax_B(t) + Bu(t) + \sum_{k=1}^m N_k x_B(t)u_k(t), \quad x_B(0) = 0, \tag{3a}$$

$$y_B(t) = Cx_B(t) \tag{3b}$$

captures the inhomogeneous part of (1). Consequently, we have $x = x_{x_0} + x_B$ and $y = y_{x_0} + y_B$. The above splitting was considered in [16], where the authors discuss a balancing approach to produce reduced order models (ROMs). Additionally, MOR of (1) based on a different splitting was proposed in [9]. However, theoretical questions remain open for this approaches such as the error analysis.

Notice that the need for MOR of bilinear systems with non-zero initial states is higher than for linear systems since there is an essential difference between both cases. For linear systems, it is required that several initial states are of interest in order to motivate applying MOR to the homogeneous equation. However, the homogeneous bilinear system (2) is control dependent such that MOR can already pay off for a single initial condition ($X_0 = x_0$ and $v_0 = 1$) if system evaluations for multiple controls are desired. The individual reduction of (2) and (3) has several advantages. As for linear systems, one subsystem can have a higher reduction potential than the other. Hence, reduced order dimensions can be chosen differently, but the actual benefit of the splitting goes beyond this degree of freedom. In addition, it turns out that using different Gramians and different structures of the reduced systems can be beneficial.

In this work, we discuss several Gramian based approaches in which subsystems (2) and (3) are reduced separately. This leads to reduced order models

$$\dot{\tilde{x}}_{x_0}(t) = \tilde{A}_{x_0}\tilde{x}_{x_0}(t) + \sum_{k=1}^{m} \tilde{N}_{x_0,k}\tilde{x}_{x_0}(t)u_k(t), \quad \tilde{x}_{x_0}(0) = \tilde{X}_0 v_0, \quad \tilde{y}_{x_0}(t) = \tilde{C}_{x_0}\tilde{x}_{x_0}(t) \tag{4}$$

approximating (2) and to reduced systems

$$\begin{aligned} \dot{\tilde{x}}_B(t) &= \tilde{A}_B\tilde{x}_B(t) + \tilde{B}u(t) + \sum_{k=1}^{m} \left(\tilde{N}_{B,k}\tilde{x}_B(t) + \tilde{E}_k u(t)\right) u_k(t), \quad \tilde{x}_B(0) = 0, \\ \tilde{y}_B(t) &= \tilde{C}_B\tilde{x}_B(t) + \tilde{D}u(t) \end{aligned} \tag{5}$$

approximating (3) with $\tilde{x}_{x_0}(t) \in \mathbb{R}^{r_{x_0}}$ and $\tilde{x}_B(t) \in \mathbb{R}^{r_B}$, where $r_{x_0}, r_B \ll n$ and all above matrices are of suitable dimension. The goal is to choose (4) and (5) such that $y \approx \tilde{y}_{x_0} + \tilde{y}_B$.

In this paper, we provide estimates that explain how the considered Gramians characterize dominant subspaces in both (2) and (3). Such a result for (2) has not even been established in the linear case. These estimates give a motivation for different Gramian based MOR techniques proposed in this paper without directly using control concepts such as reachability or observability. Moreover, we prove error bounds for all methods studied within this paper, closing a gap in the analysis of such schemes. However, the main focus is on analyzing (4), since different results on properties of (5) already exist in the literature.

## 2 Solution representation and fundamental solutions

The fundamental solution to (1a) represents a basis for the solution to the homogeneous state equation ($B = 0$). Its precise definition is as follows:

**Definition 2.1.** *Given that $s \leq t$, the fundamental solution to* (1a) *is a matrix-valued function $\Phi$ satisfying*

$$\Phi(t,s) = I + \int_s^t A\Phi(v,s)dv + \sum_{k=1}^{m} \int_s^t N_k\Phi(v,s)u_k(v)dv.$$

*If $s = 0$, we set $\Phi(t) := \Phi(t,0)$.*

This fundamental solution can now be used to derive an explicit representation for the state variable.

**Lemma 2.2.** *The solution to* (1a) *for $0 \leq t_0 \leq t$ is given by*

$$x(t) = \Phi(t,t_0)x(t_0) + \int_{t_0}^{t} \Phi(t,s)Bu(s)ds.$$

*Proof.* Using that $\Phi(t,s) = \Phi(t)\Phi^{-1}(s)$, the result follows by applying the product rule to $\Phi(t)g(t)$, where $g(t) := \Phi^{-1}(t_0)x(t_0) + \int_{t_0}^{t} \Phi^{-1}(s)Bu(s)ds$. □

In the context of MOR and associated error estimates the solution representation in Lemma 2.2 is vital. However, the fundamental solution is control dependent and hence $\Phi(t,s) \neq \Phi(t-s)$ (no semigroup property of $\Phi(\cdot)$). Therefore, an estimate on $\Phi$ is needed in order to extract the dependence on $u$. Given two symmetric matrices $M_1$ and $M_2$ we write $M_1 \leq M_2$ below if $M_2 - M_1$ is symmetric positive semidefinite.

**Lemma 2.3.** *Let $\Phi$ be the fundamental solution according to Definition 2.1, $K \geq 0$ and $\gamma > 0$. Then,*

$$\Phi(t,s)K\Phi^\top(t,s) \leq \exp\left\{\int_s^t \left\|\gamma u^0(v)\right\|_2^2 dv\right\} Z_\gamma(t-s),$$

*where $Z_\gamma(t)$, $t \geq 0$, satisfies the matrix differential equation*

$$\dot{Z}_\gamma(t) = AZ_\gamma(t) + Z_\gamma(t)A^\top + \frac{1}{\gamma^2}\sum_{k=1}^{m} N_k Z_\gamma(t) N_k^\top, \quad Z_\gamma(0) = K, \tag{6}$$

*and $u^0$ is the vector of control functions entering the bilinear part*

$$u^0 = (u_1^0\ u_2^0\ \dots\ u_m^0)^\top \quad \textit{with} \quad u_k^0 \equiv \begin{cases} 0, & \textit{if } N_k = 0 \\ u_k, & \textit{else.} \end{cases} \tag{7}$$

*Proof.* We factorize $K = FF^\top$. Let $f_i$ be the $i$th column of the matrix $F$ and $x_{f_i}(\cdot,s)$ denote the solution to (2a) with initial state $f_i$ and initial time $s$. Then, we have

$$\Phi(t,s)F = [x_{f_1}(t,s), x_{f_2}(t,s), \dots, x_{f_d}(t,s)],$$

where $d$ is the number of columns of $F$. Using the scaling $\gamma > 0$, $x_{f_i}(\cdot,s)$ can be interpreted as the solution to $\dot{x}_{f_i}(t) = Ax_{f_i}(t) + \sum_{k=1}^{m} \frac{1}{\gamma} N_k x_{f_i}(t)\gamma u_k(t)$. Applying the results of [22, Section 2] on a bound for $x_{f_i}(t,s)x_{f_i}^\top(t,s)$, we obtain

$$\begin{aligned}\Phi(t,s)K\Phi^\top(t,s) = \sum_{k=1}^{m} x_{f_i}(t,s)x_{f_i}^\top(t,s) &\leq \exp\left\{\int_s^t \left\|\gamma u^0(v)\right\|_2^2 dv\right\} \sum_{k=1}^{m} Z_\gamma(t-s, f_i f_i^\top) \\ &= \exp\left\{\int_s^t \left\|\gamma u^0(v)\right\|_2^2 dv\right\} Z_\gamma(t-s, K),\end{aligned}$$

where the second argument in $Z_\gamma$ denotes the respective initial condition. □

Lemma 2.3 is a variation of the results from Lemmas 2.2, 2.3 and 2.4 in [22]. The constant $\gamma$ in Lemma 2.3 is essential to achieve asymptotic stability of (6). Based on this stability, Gramians for (2) will be introduced in Section 3.1. However, a bit less than asymptotic stability is needed, as the following theorem shows. It contains a sufficient condition for the existence of Gramians. This criterion is related to a matrix inequality and can be seen as an extended notion of stability for (6). We will also see later that ROMs (4) based on balancing generally satisfy such a condition.

**Theorem 2.4.** *Let $\gamma > 0$ and $Z_\gamma(\cdot, X_0X_0^\top)$ the solution to (6) with $K = X_0X_0^\top$. If there exists a matrix $X > 0$ such that*

$$AX + XA^\top + \frac{1}{\gamma^2}\sum_{k=1}^m N_k X N_k^\top \leq -X_0X_0^\top. \tag{8}$$

*Then, (6) is stable meaning that*

$$\sigma\left(I \otimes A + A \otimes I + \frac{1}{\gamma^2}\sum_{k=1}^m N_k \otimes N_k\right) \subset \overline{\mathbb{C}_-}. \tag{9}$$

*Moreover, there is a constant $c > 0$ such that $\left\|Z_\gamma(t, X_0X_0^\top)\right\|_2 \lesssim \mathrm{e}^{-ct}$, i.e., the initial condition $K = X_0X_0^\top$ yields exponential decay. In particular, we can construct a matrix $V \in \mathbb{R}^{n\times\tilde{n}}$, $\tilde{n} \leq n$, with $V^\top V = I$ providing a projected system with coefficients $\tilde{A} = V^\top AV$, $\tilde{X}_0 = V^\top X_0$ and $\tilde{N}_k = V^\top N_k V$. This reduced system with fundamental solution $\tilde{\Phi}$ has an asymptotically stable equation (6), i.e., it holds that*

$$\sigma\left(I \otimes \tilde{A} + \tilde{A} \otimes I + \frac{1}{\gamma^2}\sum_{k=1}^m \tilde{N}_k \otimes \tilde{N}_k\right) \subset \mathbb{C}_- \tag{10}$$

*and it has no reduction error in the sense that*

$$\Phi(t)X_0 = V\tilde{\Phi}(t)\tilde{X}_0.$$

*Proof.* Condition (8) implies (9) by [7, Corollary 3.2] or [18, Lemma 6.12]. Now, we can use a stochastic representation for $Z_\gamma(\cdot, X_0X_0^\top)$, see, e.g., [10, 19] which is $Z_\gamma(t, X_0X_0^\top) = \mathbb{E}\left[\Phi_w(t)X_0X_0^\top\Phi_w^\top(t)\right]$. Here, the stochastic fundamental solution $\Phi_w$ satisfies $\Phi_w(t) = I + \int_0^t A\Phi_w(s)ds + \sum_{k=1}^m \int_0^t N_k\Phi_w(s)dw_k(s)$ by definition, where $w_1, \ldots, w_m$ are independent standard Brownian motions. Based on [24, Theorem 4.4, Remark 1], we then find

$$\left\|Z_\gamma(t, X_0X_0^\top)\right\|_2 \leq \mathbb{E}\left\|\Phi_w(t)X_0X_0^\top\Phi_w^\top(t)\right\|_2 \leq \mathbb{E}\left\|\Phi_w(t)X_0\right\|_F^2 \lesssim \mathrm{e}^{-ct}.$$

Let us finally consider

$$\begin{aligned}\left\|\Phi(t)X_0 - V\tilde{\Phi}(t)\tilde{X}_0\right\|_F^2 &= \left\|\begin{bmatrix} I & -V\end{bmatrix}\begin{bmatrix}\Phi(t) & 0\\ 0 & \tilde{\Phi}(t)\end{bmatrix}\begin{bmatrix}X_0\\ \tilde{X}_0\end{bmatrix}\right\|_F^2\\ &= \operatorname{tr}\left(\begin{bmatrix} I & -V\end{bmatrix}\begin{bmatrix}\Phi(t) & 0\\ 0 & \tilde{\Phi}(t)\end{bmatrix}\begin{bmatrix}X_0\\ \tilde{X}_0\end{bmatrix}\begin{bmatrix} X_0^\top & \tilde{X}_0^\top\end{bmatrix}\begin{bmatrix}\Phi^\top(t) & 0\\ 0 & \tilde{\Phi}^\top(t)\end{bmatrix}\begin{bmatrix} I\\ -V^\top\end{bmatrix}\right).\end{aligned}$$

Since $\begin{bmatrix}\Phi(t) & 0\\ 0 & \tilde{\Phi}(t)\end{bmatrix}$ is the fundamental solution to a bilinear system with matrices $\begin{bmatrix}A & 0\\ 0 & \tilde{A}\end{bmatrix}$ and $\begin{bmatrix}N_k & 0\\ 0 & \tilde{N}_k\end{bmatrix}$, we can apply Lemma 2.3 leading to

$$\left\|\Phi(t)X_0 - V\tilde{\Phi}(t)\tilde{X}_0\right\|_F^2 \leq \operatorname{tr}\left(\begin{bmatrix} I & -V\end{bmatrix}Z_\gamma^e(t)\begin{bmatrix} I\\ -V^\top\end{bmatrix}\right)\exp\left\{\int_0^t \left\|\gamma u^0(s)\right\|_2^2 ds\right\},$$

where $Z^e_\gamma$ is the matrix function solving (6) with coefficients $\left[\begin{smallmatrix} A & 0 \\ 0 & \tilde{A} \end{smallmatrix}\right]$, $\left[\begin{smallmatrix} N_k & 0 \\ 0 & \tilde{N}_k \end{smallmatrix}\right]$ and $K = \left[\begin{smallmatrix} X_0 \\ \tilde{X}_0 \end{smallmatrix}\right] \left[\begin{smallmatrix} X_0^\top & \tilde{X}_0^\top \end{smallmatrix}\right]$. We exploit the associated stochastic representation which is $Z^e(t) = \mathbb{E}\left(\left[\begin{smallmatrix} \Phi_w(t) & 0 \\ 0 & \tilde{\Phi}_w(t) \end{smallmatrix}\right] \left[\begin{smallmatrix} X_0 \\ \tilde{X}_0 \end{smallmatrix}\right] \left[\begin{smallmatrix} X_0^\top & \tilde{X}_0^\top \end{smallmatrix}\right] \left[\begin{smallmatrix} \Phi_w^\top(t) & 0 \\ 0 & \tilde{\Phi}_w^\top(t) \end{smallmatrix}\right]\right)$, where $\tilde{\Phi}_w$ is the reduced order stochastic fundamental solution involving the matrices $\tilde{A}$ and $\tilde{N}_k$. Consequently, based on the linearity of the trace and the definition of the Frobenius norm, we have

$$\left\|\Phi(t)X_0 - V\tilde{\Phi}(t)\tilde{X}_0\right\|_F^2 \leq \mathbb{E}\left\|\Phi_w(t)X_0 - V\tilde{\Phi}_w(t)\tilde{X}_0\right\|_F^2 \exp\left\{\int_0^t \left\|\gamma u^0(s)\right\|_2^2 ds\right\}.$$

Due to [24, Corollary 4.5, Remark 1] we know about the existence of $V$ with $V^\top V = I$ such that $\Phi_w(t)X_0 = V\tilde{\Phi}_w(t)\tilde{X}_0$, where $\tilde{\Phi}_w$ decays exponentially in the mean square sense. This decay of $\tilde{\Phi}_w$ is equivalent to (10), see, e.g., [10] which concludes the proof. □

Theorem 2.4 shows that if (8) is satisfied, the bilinear system represented by the matrices $A, N_k$ with initial conditions encoded by the matrix $X_0$ can be always reduced to a asymptotically stable system in the sense of (10) with no reduction error.

**Remark 1.** *If* (6) *is asymptotically stable there exists an* $X > 0$ *such that*

$$AX + XA^\top + \frac{1}{\gamma^2}\sum_{k=1}^m N_k X N_k^\top = Y$$

*given* $Y < 0$*, see [10]. Setting* $Y = -I - X_0X_0^\top$ *now implies* (8).

# 3 Gramians and dominant subspaces

## 3.1 Gramians and dominant subspaces for (2)

We begin with investigating the homogeneous part of (1a) with non-zero initial states. To do so, we study two Gramians for (2) that provide information concerning the dominant subspaces of (2a) and (2b), respectively.

In order to identify the unimportant directions in (2a) a Gramian $P_0$ is introduced below. Let $Z_\gamma = Z_\gamma(t, X_0X_0^\top)$ as in (6) and $K = X_0X_0^\top$. The existence of the Gramians requires the asymptotic stability of (6) which is stronger than $\sigma(A) \subset \mathbb{C}_-$. However, we can enforce this stronger type of stability by a sufficiently large $\gamma > 0$ providing

$$\sigma(A \otimes I + I \otimes A + \frac{1}{\gamma^2}\sum_{k=1}^m N_k \otimes N_k) \subset \mathbb{C}_-. \tag{11}$$

The rescaled matrices $\frac{1}{\gamma}N_k$ in (11) are associated to the following equivalent reformulation of (2a):

$$\dot{x}_{x_0}(t) = Ax_{x_0}(t) + \sum_{k=1}^m \frac{1}{\gamma}N_k x_{x_0}(t)\gamma u_k(t), \quad x_{x_0}(0) = X_0v_0,$$

but it goes along with an enlarged control energy in the bilinearity. Now, we define

$$P_0 := \int_0^\infty Z_\gamma(s, X_0 X_0^\top) ds.$$

The dependence of $P_0$ on $\gamma$ is not explicitly indicated to simplify the notation. By definition of $P_0$ and the asymptotic stability of (6), we can immediately see that $P_0$ solves

$$AP_0 + P_0 A^\top + \frac{1}{\gamma^2} \sum_{k=1}^m N_k P_0 N_k^\top = -X_0 X_0^\top. \tag{12}$$

We are now ready to establish an estimate identifying redundant information in (2a). Therefore, let us introduce an orthonormal basis $(p_{0,i})$ of eigenvectors of $P_0$. Consequently, we can write $x_{x_0}(t) = \sum_{i=1}^n \langle x_{x_0}(t), p_{0,i}\rangle_2 \, p_{0,i}$. The following estimate for $\langle x_{x_0}(t), p_{0,i}\rangle_2$ allows us to find directions $p_{0,i}$ which barely contribute to the dynamics.

**Proposition 3.1.** *Let $x_{x_0}$ denote the solution to* (2a) *and $\gamma > 0$ such that* (11) *holds. Then,*

$$\left\| \langle x_{x_0}(\cdot), p_{0,i}\rangle_2 \right\|_{L^2} \leq \lambda_{0,i}^{\frac{1}{2}} \exp\left\{ 0.5 \left\| \gamma u^0 \right\|_{L^2}^2 \right\} \left\| v_0 \right\|_2, \tag{13}$$

*where $\lambda_{0,i}$ is the eigenvalue associated to $p_{0,i}$.*

*Proof.* Based on Lemma 2.2 we find that

$$x_{x_0}(t) = \Phi(t) x_0 = \Phi(t) X_0 v_0.$$

Exploiting this leads to

$$\begin{aligned} \int_0^t \langle x_{x_0}(s), p_{0,i}\rangle_2^2 ds &= \int_0^t \langle \Phi(s) X_0 v_0, p_{0,i}\rangle_2^2 ds = \int_0^t \langle v_0, X_0^\top \Phi^\top(s) p_{0,i}\rangle_2^2 ds \\ &\leq \left\| v_0 \right\|_2^2 p_{0,i}^\top \int_0^t \Phi(s) X_0 X_0^\top \Phi^\top(s) ds \; p_{0,i} \end{aligned}$$

using the inequality of Cauchy-Schwarz. Using Lemma 2.3, we obtain

$$\begin{aligned} \int_0^t \langle x_{x_0}(s), p_{0,i}\rangle^2 ds &\leq \left\| v_0 \right\|_2^2 \exp\left\{ \int_0^t \left\| \gamma u^0(v) \right\|_2^2 dv \right\} p_{0,i}^\top \int_0^t Z_\gamma(s, X_0 X_0^\top) ds \; p_{0,i} \\ &\leq \left\| v_0 \right\|_2^2 \exp\left\{ \left\| \gamma u^0 \right\|_{L^2}^2 \right\} p_{0,i}^\top P_0 \; p_{0,i} = \left\| v_0 \right\|_2^2 \exp\left\{ \left\| \gamma u^0 \right\|_{L^2}^2 \right\} \lambda_{0,i}. \end{aligned}$$

□

Consequently, $x_{x_0}$ is small in the direction of an eigenvector $p_{0,i} = p_{0,i}(\gamma)$ of $P_0$ associated to a small eigenvalue $\lambda_{0,i} = \lambda_{0,i}(\gamma)$. This means that eigenspaces corresponding to small eigenvalues of $P_0$ are less relevant and hence can be neglected.

Let us now turn our attention to the choice of Gramians and the related dominant subspaces of (2b). We introduce the matrix-valued function $Z_\gamma^* = Z_\gamma^*(t, C^\top C)$ satisfying

$$\dot{Z}_\gamma^*(t) = A^\top Z_\gamma^*(t) + Z_\gamma^*(t)A + \frac{1}{\gamma^2}\sum_{k=1}^m N_k^\top Z_\gamma^*(t)N_k, \quad Z_\gamma^*(0) = C^\top C, \tag{14}$$

where the superscript $*$ indicates that the Lyapunov operator defining the right side of (14) is the adjoint operator of the one entering (6). Let us further assume that (11) holds. Then, we define

$$Q := \int_0^\infty Z_\gamma^*(s, C^\top C)ds. \tag{15}$$

By definition of $Q$ and the asymptotic stability of (14), we have

$$A^\top Q + QA + \frac{1}{\gamma^2}\sum_{k=1}^m N_k^\top Q N_k = -C^\top C. \tag{16}$$

Let $0 \le t_0 < \infty$. We now expand $x_{x_0}(t_0)$ using an orthonormal basis $(q_i)$ of eigenvectors of $Q$, i.e., we write $x_{x_0}(t_0) = \sum_{i=1}^n \langle x_{x_0}(t_0), q_i\rangle_2 q_i$. The goal is to identify the directions $q_i$ which do not contribute significantly to the output $y_{x_0}$ on the interval $(t_0, \infty)$. We exploit the representation in Lemma 2.2 and obtain for $t \ge t_0$ that

$$y_{x_0}(t) = C\Phi(t, t_0)x_{x_0}(t_0) = \sum_{i=1}^n C\Phi(t, t_0)q_i\langle x_{x_0}(t_0), q_i\rangle_2. \tag{17}$$

Eigenvectors $q_i$ can now be neglected if the respective summand in (17) is small in some norm. These summands are now analyzed in the following theorem.

**Proposition 3.2.** *Let $(q_i)$ be an orthonormal basis of eigenvectors of the Gramian $Q$ and $\gamma > 0$ such that (11) holds. Then,*

$$\left(\int_{t_0}^\infty \|C\Phi(t, t_0)q_i\|_2^2\, dt\right)^{\frac{1}{2}} \le \mu_i^{\frac{1}{2}} \exp\left\{0.5\left\|\gamma u^0\right\|_{L^2}^2\right\}, \tag{18}$$

*where $\mu_i$ is the eigenvalue associated to $q_i$.*

*Proof.* With Lemma 2.3, we find

$$\begin{aligned}
\int_{t_0}^\infty \|C\Phi(t, t_0)q_i\|_2^2\, dt &= \int_{t_0}^\infty q_i^\top \Phi(t, t_0)C^\top C\Phi(t, t_0)q_i dt \\
&= \int_{t_0}^\infty \operatorname{tr}\left(C\Phi(t, t_0)q_i q_i^\top \Phi^\top(t, t_0)C^\top\right) dt \\
&\le \int_{t_0}^\infty \operatorname{tr}\left(C\exp\left\{\int_{t_0}^t \left\|\gamma u^0(v)\right\|_2^2\, dv\right\} Z_\gamma(t - t_0, q_i q_i^\top)C^\top\right) dt \\
&\le \exp\left\{\left\|\gamma u^0\right\|_{L^2}^2\right\} \int_0^\infty \operatorname{tr}\left(CZ_\gamma(s, q_i q_i^\top)C^\top\right) ds \\
&= \exp\left\{\left\|\gamma u^0\right\|_{L^2}^2\right\} \operatorname{tr}\left(C^\top C\int_0^\infty Z_\gamma(s, q_i q_i^\top)ds\right).
\end{aligned}$$

$\int_0^\infty Z_\gamma(s, q_i q_i^\top)ds$ solves (12) with right hand side $q_i q_i^\top$. Inserting (16) for $C^\top C$ above, we can see that $\operatorname{tr}\left(C^\top C \int_0^\infty Z_\gamma(s, q_i q_i^\top)ds\right) = q_i^\top Q q_i = \mu_i$. This concludes the proof. □

Estimate (18) now tells us that $q_i = q_i(\gamma)$ is an unimportant direction in $x_{x_0}(t_0)$ for each $t_0 \geq 0$ if $\mu_i = \mu_i(\gamma)$ is small since these vectors have a low impact on the output $y_{x_0}(t)$, $t \geq t_0$. Consequently, eigenspaces of $Q$ corresponding to small eigenvalues can be removed from the system.

### 3.2 Gramians and dominant subspaces for (3)

We introduce a reachability Gramian $P_B$ as a positive definite solution to

$$A^\top P_B^{-1} + P_B^{-1} A + \frac{1}{\gamma^2} \sum_{k=1}^{m} N_k^\top P_B^{-1} N_k \leq -P_B^{-1} B B^\top P_B^{-1}. \tag{19}$$

Such a solution exists given that (11) holds, see [11, Lemma III.1] or more generally [19, Proposition 3.1] Notice that an inequality is considered in (19), since the existence of a positive definite solution of the associated equality is not ensured. $P_B$ identifies directions in the state equation (3a) that can be removed from the system. To see this, let $(p_{B,i})$ an orthonormal basis of eigenvectors of $P_B$, such that $x_B(t) = \sum_{i=1}^n \langle x_B(t), p_{B,i}\rangle_2 \; p_{B,i}$. As in Proposition 3.1 an estimate for $\langle x_B(t), p_{B,i}\rangle_2$ can be found. However, the norm is a different one.

**Proposition 3.3.** *Let $x_B$ denote the solution to* (3a) *and $\gamma > 0$ such that* (11) *holds. Then,*

$$\sup_{t\geq 0} |\langle x_B(t), p_{B,i}\rangle_2| \leq \lambda_{B,i}^{\frac{1}{2}} \|u\|_{L^2} \exp\left(0.5 \left\|\gamma u^0\right\|_{L^2}^2\right), \tag{20}$$

*where $\lambda_{B,i}$ is the eigenvalue associated to $p_{B,i}$.*

*Proof.* The result for $\gamma = 1$ is a special case of [21, Section 2.1]. Rescaling $N_k x_B(t) u_k(t) \mapsto \frac{1}{\gamma} N_k x_B(t) \gamma u_k(t)$ in (3a) immediately provides the desired estimate for general $\gamma$. □

By (20), we can see that $p_{B,i} = p_{B,i}(\gamma)$ is less relevant for the dynamics if $\lambda_{B,i} = \lambda_{B,i}(\gamma)$ is small. For that reason, one is interested in computing a $P_B$ with possibly small eigenvalues since such a solution to (19) characterizes the negligible information best. Therefore, determining $P_B$ becomes an optimization problem of, e.g., minimizing $\operatorname{tr}(P_B)$ subject to (19).

The dominant subspace of (3b) can be found with the same Gramian $Q$, defined in (15) as in the case of $y_{x_0}$. We expand $x_B(t_0) = \sum_{i=1}^n \langle x_B(t_0), q_i\rangle_2 q_i$ for $0 \leq t_0 < \infty$. By Lemma 2.2 we have

$$\begin{aligned} y_B(t) &= C\Phi(t, t_0) x_B(t_0) + \int_{t_0}^t C\Phi(t, s) B u(s) ds \\ &= \sum_{i=1}^n C\Phi(t, t_0) q_i \langle x_B(t_0), q_i\rangle_2 + \int_{t_0}^t C\Phi(t, s) B u(s) ds \end{aligned}$$

for $t \geq t_0$. Therefore, the direction $q_i$ is less relevant if $C\Phi(t, t_0)q_i$ is small. The corresponding estimate for this expression has already been established in Proposition 3.2. Consequently, $q_i$ is also negligible for $y_B$ if the eigenvalue $\mu_i$ is small.

# 4 Gramian-based model order reduction

## 4.1 Balancing of subsystems (2) and (3)

We have seen in Sections 3.1 and 3.2 that the eigenspaces corresponding to small eigenvalues of $P_0$ and $Q$ are not important for subsystem (2) and the ones of $P_B$ and $Q$ are less relevant for subsystem (3). Therefore, we construct a state space transformation ensuring that $P_0$ and $Q$ are diagonal and equal, meaning that $p_{0,i} = q_i = e_i$, where $e_i$ is the $i$th unit vector in $\mathbb{R}^n$. The $i$th diagonal entry of the diagonalized Gramians then determines how much the $i$th component of the state variable contributes to the dynamics. This procedure of simultaneously diagonalizing the Gramians is called balancing. After conducting this procedure for (2), another balancing transformation is constructed for (3), guaranteeing that $P_B$ and $Q$ are diagonal and equal as well. Subsequently, the unimportant information in both subsystems can be removed, leading to the reduced models (4) and (5).

The procedure sketched above now works as follows. Based on the assumption that $P_0, Q > 0$, we can construct the following regular matrices and their inverses

$$\mathcal{S} = \Theta^{-\frac{1}{2}}\mathcal{U}^\top L^\top, \quad \mathcal{S}^{-1} = \mathcal{K}\mathcal{V}\Theta^{-\frac{1}{2}} \quad \text{and} \quad S = \Sigma^{-\frac{1}{2}}U^\top L^\top, \quad S^{-1} = KV\Sigma^{-\frac{1}{2}}, \tag{21}$$

where $\Theta = \operatorname{diag}(\theta_1, \ldots, \theta_n) > 0$ and $\Sigma = \operatorname{diag}(\sigma_1, \ldots, \sigma_n) > 0$ with $\theta_i$ and $\sigma_i$ being the square root of the $i$th eigenvalue of $P_0Q$ and $P_BQ$, respectively. These diagonal entries of $\Theta$ and $\Sigma$ are called Hankel singular values (HSVs) of (2) and (3). The other ingredients in (21) are computed by the factorizations $P_0 = \mathcal{K}\mathcal{K}^\top$, $P_B = KK^\top$, $Q = LL^\top$ and the singular value decompositions of $\mathcal{K}^\top L = \mathcal{V}\Theta\mathcal{U}^\top$ and $K^\top L = V\Sigma U^\top$.

Replacing $(A, X_0, C, N_k)$ by the transformed matrices

$$\mathcal{S}A\mathcal{S}^{-1} = \begin{bmatrix} \mathcal{A}_{11} & \mathcal{A}_{12} \\ \mathcal{A}_{21} & \mathcal{A}_{22} \end{bmatrix}, \quad \mathcal{S}X_0 = \begin{bmatrix} X_{0,1} \\ X_{0,2} \end{bmatrix}, \quad C\mathcal{S}^{-1} = \begin{bmatrix} \mathcal{C}_1 & \mathcal{C}_2 \end{bmatrix}, \quad \mathcal{S}N_k\mathcal{S}^{-1} = \begin{bmatrix} \mathcal{N}_{k,11} & \mathcal{N}_{k,12} \\ \mathcal{N}_{k,21} & \mathcal{N}_{k,22} \end{bmatrix}, \tag{22}$$

in (2) with $\mathcal{A}_{11}, \mathcal{N}_{k,11} \in \mathbb{R}^{r_{x_0} \times r_{x_0}}$, $X_{0,1} \in \mathbb{R}^{r_{x_0} \times q}$, and $\mathcal{C}_1 \in \mathbb{R}^{p \times r_{x_0}}$, we obtain the following system

$$\begin{aligned} \begin{bmatrix} \dot{\mathbf{x}}_1(t) \\ \dot{\mathbf{x}}_2(t) \end{bmatrix} &= \begin{bmatrix} \mathcal{A}_{11} & \mathcal{A}_{12} \\ \mathcal{A}_{21} & \mathcal{A}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{x}_1(t) \\ \mathbf{x}_2(t) \end{bmatrix} + \sum_{k=1}^{m} \begin{bmatrix} \mathcal{N}_{k,11} & \mathcal{N}_{k,12} \\ \mathcal{N}_{k,21} & \mathcal{N}_{k,22} \end{bmatrix} \begin{bmatrix} \mathbf{x}_1(t) \\ \mathbf{x}_2(t) \end{bmatrix} u_k(t), \quad \begin{bmatrix} \mathbf{x}_1(0) \\ \mathbf{x}_2(0) \end{bmatrix} = \begin{bmatrix} X_{0,1} \\ X_{0,2} \end{bmatrix} v_0 \\ y_{x_0}(t) &= \begin{bmatrix} \mathcal{C}_1 & \mathcal{C}_2 \end{bmatrix} \begin{bmatrix} \mathbf{x}_1(t) \\ \mathbf{x}_2(t) \end{bmatrix}, \quad t \geq 0, \end{aligned} \tag{23}$$

having the same output as (2). Above, we set $\mathcal{S}x_{x_0}(t) = \begin{bmatrix} \mathbf{x}_1(t) \\ \mathbf{x}_2(t) \end{bmatrix}$. The Gramian of (23) are

$$\mathcal{S}P_0\mathcal{S}^\top = \mathcal{S}^{-\top}Q\mathcal{S}^{-1} = \Theta = \begin{bmatrix} \Theta_1 & \\ & \Theta_2 \end{bmatrix} \tag{24}$$

with $\Theta_1 \in \mathbb{R}^{r_{x_0} \times r_{x_0}}$ and $\Theta_2 = \operatorname{diag}(\theta_{r_{x_0}+1}, \ldots, \theta_n)$ contains the $n - r_{x_0}$ smallest HSVs of the subsystem.

The same way, $(A, B, C, N_k)$ is replaced by

$$SAS^{-1} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad SB = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}, \quad CS^{-1} = \begin{bmatrix} C_1 & C_2 \end{bmatrix}, \quad SN_kS^{-1} = \begin{bmatrix} N_{k,11} & N_{k,12} \\ N_{k,21} & N_{k,22} \end{bmatrix}, \tag{25}$$

in (3) with $A_{11}, N_{k,11} \in \mathbb{R}^{r_B \times r_B}$, $B_1 \in \mathbb{R}^{r_B \times m}$, and $C_1 \in \mathbb{R}^{p \times r_B}$ such that we have

$$\begin{aligned} \begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{bmatrix} &= \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} u(t) + \sum_{k=1}^{m} \begin{bmatrix} N_{k,11} & N_{k,12} \\ N_{k,21} & N_{k,22} \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} u_k(t), \\ y_B(t) &= \begin{bmatrix} C_1 & C_2 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}, \quad t \geq 0, \end{aligned} \tag{26}$$

where $Sx_B(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}$ and the new Gramians are

$$SP_BS^\top = S^{-\top}QS^{-1} = \Sigma = \begin{bmatrix} \Sigma_1 & \\ & \Sigma_2 \end{bmatrix}$$

with $\Sigma_1 \in \mathbb{R}^{r_B \times r_B}$ and $\Sigma_2 = \operatorname{diag}(\sigma_{r_B+1}, \ldots, \sigma_n)$.

**Remark 2.** *It is important to point out that the balancing transformations $\mathcal{S}$ and $S$ depend on $\gamma$ since the Gramians are functions of this parameter. Consequently, the balancing realizations in* (22) *and* (25)*, as well as the later ROMs, depend on $\gamma$.*

### 4.2 Model order reduction for subsystem (2)

In this section, we discuss two different MOR techniques for (2) that rely on the balancing procedure described in Section 4.1. We already know that the state variables $\mathbf{x}_2$ in the balanced realization (23) are less relevant since they are associated to the small HSVs $\theta_{r_{x_0}+1}, \ldots, \theta_n$. A ROM (4) can now be obtained by neglecting $\mathbf{x}_2$. A first option is to truncate the second line of the state equation in (23) and to set $\mathbf{x}_2(t) = 0$ in the remaining parts of the subsystem. This methods is called balanced truncation and leads to a ROM with

$$\left(\tilde{A}_{x_0}, \tilde{X}_0, \tilde{C}_{x_0}, \tilde{N}_{x_0,k}\right) = \left(\mathcal{A}_{11}, X_{0,1}, \mathcal{C}_1, \mathcal{N}_{k,11}\right). \tag{27}$$

Alternatively, one can argue that due to (13), $\mathbf{x}_2$ is close to its equilibrium (especially if the system is uncontrolled). Hence, it is in a quasi steady state, motivating to set $\dot{\mathbf{x}}_2(t) = 0$ in (23). If we further neglect $\mathcal{N}_{k,21}$ and $\mathcal{N}_{k,22}$ in the resulting algebraic constrain in order to avoid a control dependence of the matrices in the ROM, we obtain $\mathbf{x}_2(t) = -\mathcal{A}_{22}^{-1}\mathcal{A}_{21}\mathbf{x}_1(t)$. Inserting this for $\mathbf{x}_2$ in (23) leads to a ROM with

$$\left(\tilde{A}_{x_0}, \tilde{X}_0, \tilde{C}_{x_0}, \tilde{N}_{x_0,k}\right) = \left(\bar{\mathcal{A}}, X_{0,1}, \bar{\mathcal{C}}, \bar{\mathcal{N}}_k\right), \tag{28}$$

where $\bar{\mathcal{A}} := \mathcal{A}_{11} - \mathcal{A}_{12}\mathcal{A}_{22}^{-1}\mathcal{A}_{21}$, $\bar{\mathcal{C}} := \mathcal{C}_1 - \mathcal{C}_2\mathcal{A}_{22}^{-1}\mathcal{A}_{21}$ and $\bar{\mathcal{N}}_k := \mathcal{N}_{k,11} - \mathcal{N}_{k,12}\mathcal{A}_{22}^{-1}\mathcal{A}_{21}$. It is important to point out that both ROMs (27) and (28) share the same initial condition matrix $\tilde{X}_0$. Notice that the structure preservation in the ROM is also desired here, which is motivated by the existence of an error bound that we prove later. This bound can only be achieved between systems having the same structure. We refer to a related SPA MOR scheme for (3) in [14], where such a reduced system was derived by an averaging principle representing a more detailed motivation than given here.

**Remark 3.** *A result on stability preservation for BT has already been established in [7]. Given $\Theta > 0$ and $\sigma(\Theta_1) \cap \sigma(\Theta_2) = \emptyset$, it was shown that*

$$\sigma(\mathcal{A}_{11} \otimes I + I \otimes \mathcal{A}_{11} + \frac{1}{\gamma^2}\sum_{k=1}^{m} \mathcal{N}_{k,11} \otimes \mathcal{N}_{k,11}) \subset \mathbb{C}_-.$$

*Whether SPA guarantees this type of stability under the same assumption is an open question. However, for SPA it can be proved that the eigenvalue of the above Kronecker matrix involving the matrices in* (28) *are in $\overline{\mathbb{C}_-}$, see [23]. Since $\Theta > 0$ and $\sigma(\Theta_1) \cap \sigma(\Theta_2) = \emptyset$ might not be always given, stability preservation and the existence of Gramians for the two different balancing related methods are discussed in the following, only assuming $\Theta_1 > 0$.*

**Theorem 4.1.** *Let $\mathcal{S}$ be the balanced transformation providing* (24) *with $\Theta_1 > 0$ and consider the associated balanced realization in* (22)*. Given the matrix differential equations*

$$\dot{\tilde{Z}}_\gamma(t) = \tilde{A}_{x_0}\tilde{Z}_\gamma(t) + \tilde{Z}_\gamma(t)\tilde{A}_{x_0}^\top + \frac{1}{\gamma^2}\sum_{k=1}^{m} \tilde{N}_{x_0,k}\tilde{Z}_\gamma(t)\tilde{N}_{x_0,k}^\top, \quad \tilde{Z}_\gamma(0) = \tilde{X}_0\tilde{X}_0^\top,$$

$$\dot{\tilde{Z}}_\gamma^*(t) = \tilde{A}_{x_0}^\top\tilde{Z}_\gamma^*(t) + \tilde{Z}_\gamma^*(t)\tilde{A}_{x_0} + \frac{1}{\gamma^2}\sum_{k=1}^{m} \tilde{N}_{x_0,k}^\top\tilde{Z}_\gamma^*(t)\tilde{N}_{x_0,k}, \quad \tilde{Z}_\gamma^*(0) = \tilde{C}_{x_0}^\top\tilde{C}_{x_0},$$

*the Gramians $\tilde{P} := \int_0^\infty \tilde{Z}_\gamma(s)ds$ and $\tilde{Q} := \int_0^\infty \tilde{Z}_\gamma^*(s)ds$ exist for reduced system* (4) *with coefficients as in* (27) *(BT). If instead the ROM by SPA defined in* (28) *is considered, the existence of $\tilde{Q}$ is ensured.*

*Proof.* Since the Gramians of a balanced system are identical and equal to the diagonal matrix $\Theta$, we have

$$\begin{bmatrix} \mathcal{A}_{11} & \mathcal{A}_{12} \\ \mathcal{A}_{21} & \mathcal{A}_{22} \end{bmatrix}\begin{bmatrix} \Theta_1 & \\ & \Theta_2 \end{bmatrix} + \begin{bmatrix} \Theta_1 & \\ & \Theta_2 \end{bmatrix}\begin{bmatrix} \mathcal{A}_{11}^\top & \mathcal{A}_{21}^\top \\ \mathcal{A}_{12}^\top & \mathcal{A}_{22}^\top \end{bmatrix} + \frac{1}{\gamma^2}\sum_{k=1}^{m}\begin{bmatrix} \mathcal{N}_{k,11} & \mathcal{N}_{k,12} \\ \mathcal{N}_{k,21} & \mathcal{N}_{k,22} \end{bmatrix}\begin{bmatrix} \Theta_1 & \\ & \Theta_2 \end{bmatrix}\begin{bmatrix} \mathcal{N}_{k,11}^\top & \mathcal{N}_{k,21}^\top \\ \mathcal{N}_{k,12}^\top & \mathcal{N}_{k,22}^\top \end{bmatrix}$$
$$= -\begin{bmatrix} X_{0,1} \\ X_{0,2} \end{bmatrix}\begin{bmatrix} X_{0,1}^\top & X_{0,2}^\top \end{bmatrix}, \tag{29}$$

$$\begin{bmatrix} \mathcal{A}_{11}^\top & \mathcal{A}_{21}^\top \\ \mathcal{A}_{12}^\top & \mathcal{A}_{22}^\top \end{bmatrix}\begin{bmatrix} \Theta_1 & \\ & \Theta_2 \end{bmatrix} + \begin{bmatrix} \Theta_1 & \\ & \Theta_2 \end{bmatrix}\begin{bmatrix} \mathcal{A}_{11} & \mathcal{A}_{12} \\ \mathcal{A}_{21} & \mathcal{A}_{22} \end{bmatrix} + \frac{1}{\gamma^2}\sum_{k=1}^{m}\begin{bmatrix} \mathcal{N}_{k,11}^\top & \mathcal{N}_{k,21}^\top \\ \mathcal{N}_{k,12}^\top & \mathcal{N}_{k,22}^\top \end{bmatrix}\begin{bmatrix} \Theta_1 & \\ & \Theta_2 \end{bmatrix}\begin{bmatrix} \mathcal{N}_{k,11} & \mathcal{N}_{k,12} \\ \mathcal{N}_{k,21} & \mathcal{N}_{k,22} \end{bmatrix}$$
$$= -\begin{bmatrix} \mathcal{C}_1^\top \\ \mathcal{C}_2^\top \end{bmatrix}\begin{bmatrix} \mathcal{C}_1 & \mathcal{C}_2 \end{bmatrix}. \tag{30}$$

The left upper blocks of these equations yield

$$\mathcal{A}_{11}\Theta_1 + \Theta_1\mathcal{A}_{11}^\top + \frac{1}{\gamma^2}\sum_{k=1}^m \mathcal{N}_{k,11}\Theta_1\mathcal{N}_{k,11}^\top \leq -X_{0,1}X_{0,1}^\top, \tag{31}$$
$$\mathcal{A}_{11}^\top\Theta_1 + \Theta_1\mathcal{A}_{11} + \frac{1}{\gamma^2}\sum_{k=1}^m \mathcal{N}_{k,11}^\top\Theta_1\mathcal{N}_{k,11} \leq -\mathcal{C}_1^\top\mathcal{C}_1.$$

Therefore, $\tilde{Z}_\gamma$ and $\tilde{Z}_\gamma^*$ decay exponentially by Theorem 2.4 if BT is considered. Consequently, the integrals $\tilde{P}$ and $\tilde{Q}$ exist. We now multiply (30) by

$$\begin{bmatrix}\mathcal{A}_{11} & \mathcal{A}_{12}\\ \mathcal{A}_{21} & \mathcal{A}_{22}\end{bmatrix}^{-1} = \begin{bmatrix}\bar{\mathcal{A}}^{-1} & \star\\ -\mathcal{A}_{22}^{-1}\mathcal{A}_{21}\bar{\mathcal{A}}^{-1} & \star\end{bmatrix} \tag{32}$$

from the right and with its transposed from the left. Evaluating the left upper block of the resulting equation and multiplying it with $\bar{\mathcal{A}}$ from the right and its transposed from the left, we find

$$\bar{\mathcal{A}}^\top\Theta_1 + \Theta_1\bar{\mathcal{A}} + \frac{1}{\gamma^2}\sum_{k=1}^m \bar{\mathcal{N}}_k^\top\Theta_1\bar{\mathcal{N}}_k = -\bar{\mathcal{C}}^\top\bar{\mathcal{C}} - \frac{1}{\gamma^2}\sum_{k=1}^m \bar{\mathcal{N}}_{k,21}^\top\Theta_2\bar{\mathcal{N}}_{k,21} \leq -\bar{\mathcal{C}}^\top\bar{\mathcal{C}}, \tag{33}$$

where $\bar{\mathcal{N}}_{k,21} := \mathcal{N}_{k,21} - \mathcal{N}_{k,22}\mathcal{A}_{22}^{-1}\mathcal{A}_{21}$ providing the existence of $\tilde{Q}$ for SPA using Theorem 2.4. □

Due to Theorem 4.1 technical assumptions like $\theta_{r_{x_0}} \neq \theta_{r_{x_0}+1}$ can be omitted in the error analysis if BT is considered since the reduced Gramians will still exist. Furthermore, given $\Theta_1 > 0$, Theorem 2.4 and (31) tell us that the ROM by BT can always be reduced to a system satisfying (10) without causing an additional error. Whether $\tilde{P}$ generally exists for SPA remains open and is therefore always assumed below. Now, we establish error bounds for the BT and SPA procedures. Firstly, we prove an intermediate lemma in order to show this result.

**Lemma 4.2.** *Let $y_{x_0}$ be the output of (2) and $\tilde{y}_{x_0}$ be the reduced order output of system (4). Then, we have*

$$\|y_{x_0}(t) - \tilde{y}_{x_0}(t)\|_2^2 \leq \operatorname{tr}\left(\begin{bmatrix} C & -\tilde{C}_{x_0}\end{bmatrix} Z_\gamma^e(t) \begin{bmatrix} C^\top \\ -\tilde{C}_{x_0}^\top\end{bmatrix}\right) \exp\left\{\int_0^t \left\|\gamma u^0(v)\right\|_2^2 dv\right\} \|v_0\|_2^2,$$

*where $Z_\gamma^e(t)$, $t \geq 0$, satisfies the matrix differential equation*

$$\dot{Z}_\gamma^e(t) = \begin{bmatrix} A & 0\\ 0 & \tilde{A}_{x_0}\end{bmatrix} Z_\gamma^e(t) + Z_\gamma^e(t)\begin{bmatrix} A^\top & 0\\ 0 & \tilde{A}_{x_0}^\top\end{bmatrix} + \frac{1}{\gamma^2}\sum_{k=1}^m \begin{bmatrix} N_k & 0\\ 0 & \tilde{N}_{x_0,k}\end{bmatrix} Z_\gamma^e(t) \begin{bmatrix} N_k^\top & 0\\ 0 & \tilde{N}_{x_0,k}^\top\end{bmatrix},$$
$$Z_\gamma^e(0) = \begin{bmatrix} X_0\\ \tilde{X}_0\end{bmatrix}\begin{bmatrix} X_0^\top & \tilde{X}_0^\top\end{bmatrix}.$$

*Proof.* By Lemma 2.2, we have that $y_{x_0}(t) = C\Phi(t)X_0v_0$ and $y_{x_0}(t) = \tilde{C}_{x_0}\tilde{\Phi}(t)\tilde{X}_0v_0$, where $\Phi$ and $\tilde{\Phi}$ are the fundamental solutions to the original and the reduced system, respectively, introduced in Definition 2.1. Consequently, we obtain

$$\begin{aligned}\|y_{x_0}(t) - \tilde{y}_{x_0}(t)\|_2^2 &\leq \left\|C\Phi(t)X_0 - \tilde{C}_{x_0}\tilde{\Phi}(t)\tilde{X}_0\right\|_F^2 \|v_0\|_2^2 = \left\|\begin{bmatrix} C & -\tilde{C}_{x_0}\end{bmatrix}\begin{bmatrix}\Phi(t) & 0\\ 0 & \tilde{\Phi}(t)\end{bmatrix}\begin{bmatrix}X_0\\ \tilde{X}_0\end{bmatrix}\right\|_F^2 \|v_0\|_2^2\\ &= \operatorname{tr}\left(\begin{bmatrix} C & -\tilde{C}_{x_0}\end{bmatrix}\begin{bmatrix}\Phi(t) & 0\\ 0 & \tilde{\Phi}(t)\end{bmatrix}\begin{bmatrix}X_0\\ \tilde{X}_0\end{bmatrix}\begin{bmatrix} X_0^\top & \tilde{X}_0^\top\end{bmatrix}\begin{bmatrix}\Phi^\top(t) & 0\\ 0 & \tilde{\Phi}^\top(t)\end{bmatrix}\begin{bmatrix} C^\top\\ -C_{x_0}^\top\end{bmatrix}\right)\|v_0\|_2^2.\end{aligned}$$

Now, $\begin{bmatrix}\Phi(t) & 0\\ 0 & \tilde{\Phi}(t)\end{bmatrix}$ is the fundamental solution to the bilinear system with matrices $\begin{bmatrix}A & 0\\ 0 & \tilde{A}_{x_0}\end{bmatrix}$ and $\begin{bmatrix}N_k & 0\\ 0 & \tilde{N}_{x_0,k}\end{bmatrix}$. Therefore, the result follows by Lemma 2.3 setting $K = \begin{bmatrix}X_0\\ \tilde{X}_0\end{bmatrix}\begin{bmatrix} X_0^\top & \tilde{X}_0^\top\end{bmatrix}$. □

**Theorem 4.3.** *Let $y_{x_0}$ be the output of (2) given that (11) holds for a sufficiently large $\gamma > 0$. Suppose that $\tilde{y}_{x_0}$ is the reduced order output of system (4), where the matrices $\left(\tilde{A}_{x_0}, \tilde{X}_0, \tilde{C}_{x_0}, \tilde{N}_{x_0,k}\right)$ are either given by BT stated in (27) or by SPA defined in (28) given a balancing transformation $\mathcal{S}$ as in (24) with $\Theta_1 > 0$. We assume that the reduced system Gramian $\tilde{P}$ for SPA exists. Defining $Y = \begin{bmatrix}Y_1\\ Y_2\end{bmatrix}$ as the unique solution to*

$$\begin{bmatrix}\mathcal{A}_{11} & \mathcal{A}_{12}\\ \mathcal{A}_{21} & \mathcal{A}_{22}\end{bmatrix}\begin{bmatrix}Y_1\\ Y_2\end{bmatrix} + \begin{bmatrix}Y_1\\ Y_2\end{bmatrix}\tilde{A}_{x_0}^\top + \frac{1}{\gamma^2}\sum_{k=1}^{m}\begin{bmatrix}\mathcal{N}_{k,11} & \mathcal{N}_{k,12}\\ \mathcal{N}_{k,21} & \mathcal{N}_{k,22}\end{bmatrix}\begin{bmatrix}Y_1\\ Y_2\end{bmatrix}\tilde{N}_{x_0}^\top = -\begin{bmatrix}X_{0,1}\\ X_{0,2}\end{bmatrix}X_{0,1}^\top, \tag{34}$$

*using the balanced realization in (22), it holds that*

$$\|y_{x_0} - \tilde{y}_{x_0}\|_{L^2} \leq (\operatorname{tr}(\Theta_2\mathcal{W}_{x_0}))^{\frac{1}{2}} \exp\left\{0.5\left\|\gamma u^0\right\|_{L^2}^2\right\}\|v_0\|_2. \tag{35}$$

*The above weight is*

$$\mathcal{W}_{x_0} = X_{0,2}X_{0,2}^\top + 2Y_2\mathbf{A}_{21}^\top + 2\begin{bmatrix}\mathcal{A}_{21} & \mathcal{A}_{22}\end{bmatrix}Y\bar{\mathbf{A}}_{21}^\top + \frac{1}{\gamma^2}\sum_{k=1}^{m} 2\begin{bmatrix}\mathcal{N}_{k,21} & \mathcal{N}_{k,22}\end{bmatrix}Y\mathbf{N}_{k,21}^\top - \mathbf{N}_{k,21}\tilde{P}\mathbf{N}_{k,21}^\top,$$

*where $(\mathbf{A}_{21}, \bar{\mathbf{A}}_{21}, \mathbf{N}_{k,21}) = (\mathcal{A}_{21}, 0, \mathcal{N}_{k,21})$ for BT and $(\mathbf{A}_{21}, \bar{\mathbf{A}}_{21}, \mathbf{N}_{k,21}) = (0, -\mathcal{A}_{22}^{-1}\mathcal{A}_{21}, \mathcal{N}_{k,21} - \mathcal{N}_{k,22}\mathcal{A}_{22}^{-1}\mathcal{A}_{21})$ for SPA.*

*Proof.* We integrate the result of Lemma 4.2 on $[0, \infty)$ and obtain

$$\|y_{x_0} - \tilde{y}_{x_0}\|_{L^2}^2 \leq \mathcal{E}\exp\left\{\left\|\gamma u^0\right\|_{L^2}^2\right\}\|v_0\|_2^2,$$

where $\mathcal{E} := \operatorname{tr}\left(\begin{bmatrix} C & -\tilde{C}_{x_0}\end{bmatrix}\int_0^\infty Z_\gamma^e(t)dt\begin{bmatrix} C^\top\\ -\tilde{C}_{x_0}^\top\end{bmatrix}\right)$. The left upper and the right lower block of $\int_0^\infty Z_\gamma^e(t)dt$ are $P_0$ and $\tilde{P}$, respectively. Both Gramians exist by assumption and Theorem 4.1. This also implies the existence of the right upper block of $\int_0^\infty Z_\gamma^e(t)dt$ which we denote by $\widehat{P}$. It satisfies

$$A\widehat{P} + \widehat{P}\tilde{A}_{x_0}^\top + \frac{1}{\gamma^2}\sum_{k=1}^{m} N_k\widehat{P}\tilde{N}_{x_0,k}^\top = -X_0X_{0,1}^\top. \tag{36}$$

Let $\mathcal{S}$ be the matrix ensuring (24). Since $\Theta = \mathcal{S}P_0\mathcal{S}^\top$, $Y = \mathcal{S}\widehat{P}$ and $[\,\mathcal{C}_1\ \mathcal{C}_2\,] = C\mathcal{S}^{-1}$, we have

$$\begin{aligned}\mathcal{E} &= \operatorname{tr}(CP_0C^\top) + \operatorname{tr}(\tilde{C}_{x_0}\tilde{P}\tilde{C}_{x_0}^\top) - 2\operatorname{tr}(C\widehat{P}\tilde{C}_{x_0}^\top)\\ &= \operatorname{tr}([\,\mathcal{C}_1\ \mathcal{C}_2\,]\begin{bmatrix}\Theta_1 & \\ & \Theta_2\end{bmatrix}\begin{bmatrix}\mathcal{C}_1^\top\\ \mathcal{C}_2^\top\end{bmatrix}) + \operatorname{tr}(\tilde{C}_{x_0}\tilde{P}_0\tilde{C}_{x_0}^\top) - 2\operatorname{tr}([\,\mathcal{C}_1\ \mathcal{C}_2\,]Y\tilde{C}_{x_0}^\top).\end{aligned}$$

Comparing (29) with (30), we see that $\operatorname{tr}([\,\mathcal{C}_1\ \mathcal{C}_2\,]\Theta\begin{bmatrix}\mathcal{C}_1^\top\\ \mathcal{C}_2^\top\end{bmatrix}) = \operatorname{tr}([\,X_{0,1}^\top\ X_{0,2}^\top\,]\Theta\begin{bmatrix}X_{0,1}\\ X_{0,2}\end{bmatrix})$. Since the same is true for the reduced Gramians, we obtain

$$\mathcal{E} = \operatorname{tr}([\,X_{0,1}^\top\ X_{0,2}^\top\,]\begin{bmatrix}\Theta_1 & \\ & \Theta_2\end{bmatrix}\begin{bmatrix}X_{0,1}\\ X_{0,2}\end{bmatrix}) + \operatorname{tr}(X_{0,1}^\top\tilde{Q}X_{0,1}) - 2\operatorname{tr}([\,\mathcal{C}_1\ \mathcal{C}_2\,]Y\tilde{C}_{x_0}^\top), \tag{37}$$

where $\tilde{Q}$ exists due to Theorem 4.1. Now, it is needed to find an equation for $\tilde{C}_{x_0}^\top[\,\mathcal{C}_1\ \mathcal{C}\,]$ for both BT and SPA in order to analyze the error further. We evaluate the first $r_{x_0}$ rows of (30) to obtain an expression for the case of BT:

$$-\mathcal{C}_1^\top[\,\mathcal{C}_1\ \mathcal{C}_2\,] = [\,\mathcal{A}_{11}^\top\Theta_1\ \mathcal{A}_{21}^\top\Theta_2\,] + [\,\Theta_1\ 0\,]\begin{bmatrix}\mathcal{A}_{11} & \mathcal{A}_{12}\\ \mathcal{A}_{21} & \mathcal{A}_{22}\end{bmatrix} + \frac{1}{\gamma^2}\sum_{k=1}^{m}[\,\mathcal{N}_{k,11}^\top\Theta_1\ \mathcal{N}_{k,21}^\top\Theta_2\,]\begin{bmatrix}\mathcal{N}_{k,11} & \mathcal{N}_{k,12}\\ \mathcal{N}_{k,21} & \mathcal{N}_{k,22}\end{bmatrix}. \tag{38}$$

For SPA we multiply (30) with $\begin{bmatrix}\mathcal{A}_{11}^\top & \mathcal{A}_{21}^\top\\ \mathcal{A}_{12}^\top & \mathcal{A}_{22}^\top\end{bmatrix}^{-1}$ from the left and obtain

$$\begin{aligned}-\begin{bmatrix}\bar{\mathcal{A}}^{-\top}\bar{\mathcal{C}}^\top\\ \star\end{bmatrix}[\,\mathcal{C}_1\ \mathcal{C}_2\,] &= \begin{bmatrix}\Theta_1 & \\ & \Theta_2\end{bmatrix} + \begin{bmatrix}\bar{\mathcal{A}}^{-\top} & -\bar{\mathcal{A}}^{-\top}(\mathcal{A}_{22}^{-1}\mathcal{A}_{21})^\top\\ \star & \star\end{bmatrix}\begin{bmatrix}\Theta_1 & \\ & \Theta_2\end{bmatrix}\begin{bmatrix}\mathcal{A}_{11} & \mathcal{A}_{12}\\ \mathcal{A}_{21} & \mathcal{A}_{22}\end{bmatrix}\\ &\quad + \frac{1}{\gamma^2}\sum_{k=1}^{m}\begin{bmatrix}\bar{\mathcal{A}}^{-\top}\bar{\mathcal{N}}_k^\top & \bar{\mathcal{A}}^{-\top}\bar{\mathcal{N}}_{k,21}^\top\\ \star & \star\end{bmatrix}\begin{bmatrix}\Theta_1 & \\ & \Theta_2\end{bmatrix}\begin{bmatrix}\mathcal{N}_{k,11} & \mathcal{N}_{k,12}\\ \mathcal{N}_{k,21} & \mathcal{N}_{k,22}\end{bmatrix}\end{aligned}$$

using the partition in (32) and setting $\bar{\mathcal{N}}_{k,21} := \mathcal{N}_{k,21} - \mathcal{N}_{k,22}\mathcal{A}_{22}^{-1}\mathcal{A}_{21}$. Multiplying the first $r_{x_0}$ rows of the above equation by $\bar{\mathcal{A}}^\top$ from the left results in

$$-\bar{\mathcal{C}}^\top[\,\mathcal{C}_1\ \mathcal{C}_2\,] = \bar{\mathcal{A}}^\top\Theta_1 + [\,\Theta_1\ -(\mathcal{A}_{22}^{-1}\mathcal{A}_{21})^\top\Theta_2\,]\begin{bmatrix}\mathcal{A}_{11} & \mathcal{A}_{12}\\ \mathcal{A}_{21} & \mathcal{A}_{22}\end{bmatrix} + \frac{1}{\gamma^2}\sum_{k=1}^{m}[\,\bar{\mathcal{N}}_k^\top\Theta_1\ \bar{\mathcal{N}}_{k,21}^\top\Theta_2\,]\begin{bmatrix}\mathcal{N}_{k,11} & \mathcal{N}_{k,12}\\ \mathcal{N}_{k,21} & \mathcal{N}_{k,22}\end{bmatrix}. \tag{39}$$

We summarize (38) and (39) to one equation. That is

$$-\tilde{C}_{x_0}^\top[\,\mathcal{C}_1\ \mathcal{C}_2\,] = [\,\tilde{A}_{x_0}^\top\Theta_1\ \mathbf{A}_{21}^\top\Theta_2\,] + [\,\Theta_1\ \bar{\mathbf{A}}_{21}^\top\Theta_2\,]\begin{bmatrix}\mathcal{A}_{11} & \mathcal{A}_{12}\\ \mathcal{A}_{21} & \mathcal{A}_{22}\end{bmatrix} + \frac{1}{\gamma^2}\sum_{k=1}^{m}[\,\tilde{N}_{x_0,k}^\top\Theta_1\ \mathbf{N}_{k,21}^\top\Theta_2\,]\begin{bmatrix}\mathcal{N}_{k,11} & \mathcal{N}_{k,12}\\ \mathcal{N}_{k,21} & \mathcal{N}_{k,22}\end{bmatrix}.$$

where $\mathbf{A}_{21} \in \{\mathcal{A}_{21}, 0\}$, $\bar{\mathbf{A}}_{21} \in \{0, -\mathcal{A}_{22}^{-1}\mathcal{A}_{21}\}$ and $\mathbf{N}_{k,21} \in \{\mathcal{N}_{k,21}, \bar{\mathcal{N}}_{k,21}\}$. Inserting this

into $\operatorname{tr}([\,\mathcal{C}_1\ \mathcal{C}_2\,]\,Y\tilde{C}_{x_0}^\top)$ yields

$$
\begin{aligned}
&-\operatorname{tr}([\,\mathcal{C}_1\ \mathcal{C}_2\,]\,Y\tilde{C}_{x_0}^\top)\\
&=\operatorname{tr}\left(Y\left[[\,\tilde{A}_{x_0}^\top\Theta_1\ \mathbf{A}_{21}^\top\Theta_2\,]+[\,\Theta_1\ \bar{\mathbf{A}}_{21}^\top\Theta_2\,]\begin{bmatrix}\mathcal{A}_{11}&\mathcal{A}_{12}\\ \mathcal{A}_{21}&\mathcal{A}_{22}\end{bmatrix}+\frac{1}{\gamma^2}\sum_{k=1}^m[\,\tilde{N}_{x_0,k}^\top\Theta_1\ \mathbf{N}_{k,21}^\top\Theta_2\,]\begin{bmatrix}\mathcal{N}_{k,11}&\mathcal{N}_{k,12}\\ \mathcal{N}_{k,21}&\mathcal{N}_{k,22}\end{bmatrix}\right]\right)\\
&=\operatorname{tr}\left(\Theta_1\left[[\,\mathcal{A}_{11}\ \mathcal{A}_{12}\,]\,Y+Y_1\tilde{A}_{x_0}^\top+\frac{1}{\gamma^2}\sum_{k=1}^m[\,\mathcal{N}_{k,11}\ \mathcal{N}_{k,12}\,]\,Y\tilde{N}_{x_0,k}^\top\right]\right)\\
&\quad+\operatorname{tr}\left(\Theta_2\left[Y_2\mathbf{A}_{21}^\top+[\,\mathcal{A}_{21}\ \mathcal{A}_{22}\,]\,Y\bar{\mathbf{A}}_{21}^\top+\frac{1}{\gamma^2}\sum_{k=1}^m[\,\mathcal{N}_{k,21}\ \mathcal{N}_{k,22}\,]\,Y\mathbf{N}_{k,21}^\top\right]\right).
\end{aligned}
$$

The first $r_{x_0}$ rows of (34) give us

$$
\begin{aligned}
-\operatorname{tr}([\,\mathcal{C}_1\ \mathcal{C}_2\,]\,Y\mathcal{C}_1^\top)&=-\operatorname{tr}\left(X_{0,1}^\top\Theta_1X_{0,1}\right)\\
&\quad+\operatorname{tr}\left(\Theta_2\left[Y_2\mathbf{A}_{21}^\top+[\,\mathcal{A}_{21}\ \mathcal{A}_{22}\,]\,Y\bar{\mathbf{A}}_{21}^\top+\frac{1}{\gamma^2}\sum_{k=1}^m[\,\mathcal{N}_{k,21}\ \mathcal{N}_{k,22}\,]\,Y\mathbf{N}_{k,21}^\top\right]\right).
\end{aligned}
$$

Inserting this into (37) leads to

$$
\begin{aligned}
\mathcal{E}=&\operatorname{tr}(X_{0,1}^\top(\tilde{Q}-\Theta_1)X_{0,1})\\
&+\operatorname{tr}\left(\Theta_2\left[X_{0,2}X_{0,2}^\top+2Y_2\mathbf{A}_{21}^\top+2\,[\,\mathcal{A}_{21}\ \mathcal{A}_{22}\,]\,Y\bar{\mathbf{A}}_{21}^\top+\frac{2}{\gamma^2}\sum_{k=1}^m[\,\mathcal{N}_{k,21}\ \mathcal{N}_{k,22}\,]\,Y\mathbf{N}_{k,21}^\top\right]\right),
\end{aligned}
$$

By the left upper $r_{x_0}\times r_{x_0}$ block of (30) (BT) and (33) (SPA), it holds that

$$
\tilde{A}_{x_0}^\top\Theta_1+\Theta_1\tilde{A}_{x_0}+\frac{1}{\gamma^2}\sum_{k=1}^m\tilde{N}_{x_0,k}^\top\Theta_1\tilde{N}_{x_0,k}=-\tilde{C}_{x_0}^\top\tilde{C}_{x_0}-\frac{1}{\gamma^2}\sum_{k=1}^m\mathbf{N}_{k,21}^\top\Theta_2\mathbf{N}_{k,21}
$$

for both reduced order schemes. Subtracting this identity from the equation for the reduced Gramian $\tilde{Q}$, we find

$$
\tilde{A}_{x_0}^\top(\tilde{Q}-\Theta_1)+(\tilde{Q}-\Theta_1)\tilde{A}_{x_0}+\frac{1}{\gamma^2}\sum_{k=1}^m\tilde{N}_{x_0,k}^\top(\tilde{Q}-\Theta_1)\tilde{N}_{x_0,k}=\frac{1}{\gamma^2}\sum_{k=1}^m\mathbf{N}_{k,21}^\top\Theta_2\mathbf{N}_{k,21}.
$$

Hence, exploiting the equation for $\tilde{P}$, we have

$$
\begin{aligned}
\operatorname{tr}(X_{0,1}^\top(\tilde{Q}-\Theta_1)X_{0,1})&=-\operatorname{tr}\left(\left[\tilde{A}_{x_0}\tilde{P}+\tilde{P}\tilde{A}_{x_0}^\top+\frac{1}{\gamma^2}\sum_{k=1}^m\tilde{N}_{x_0,k}\tilde{P}\tilde{N}_{x_0,k}^\top\right](\tilde{Q}-\Theta_1)\right)\\
&=-\operatorname{tr}\left(\tilde{P}\left[\tilde{A}_{x_0}^\top(\tilde{Q}-\Theta_1)\ +(\tilde{Q}-\Theta_1)\ \tilde{A}_{x_0}+\frac{1}{\gamma^2}\sum_{k=1}^m\tilde{N}_{x_0,k}^\top(\tilde{Q}-\Theta_1)\ \tilde{N}_{x_0,k}\right]\right)\\
&=-\operatorname{tr}\left(\Theta_2\frac{1}{\gamma^2}\sum_{k=1}^m\mathbf{N}_{k,21}\tilde{P}\mathbf{N}_{k,21}^\top\right),
\end{aligned}
$$

which concludes the proof of this theorem. □

The result of Theorem 4.3 shows an error bound that depends on the truncated HSVs. Choosing $r_{x_0}$ such that $\Theta_2$ is small therefore ensures a small error and hence a good approximation.

### 4.3 Model order reduction for subsystem (3)

In this section, BT and SPA are studied for (3). As for the methods considered in Section 4.2 they rely on the balancing procedure described in Section 4.1. However, these methods are not necessarily structured preserving and rely on a different type of Gramian $P_B$. In order to find a ROM for (3), state variables $x_2$ in (26) need to be removed. These variables belong to the small HSVs $\sigma_{r_B+1}, \ldots, \sigma_n$ and are hence negligible. A ROM (5) by BT is here obtained by truncating the second line of the state equation in (26) and to set $x_2(t) = 0$ in the remaining parts of the subsystem. This results in

$$\left(\tilde{A}_B, \tilde{B}, \tilde{C}_B, \tilde{D}, \tilde{E}_k, \tilde{N}_{B,k}\right) = (A_{11}, B_1, C_1, 0, 0, N_{k,11})\,. \tag{40}$$

Using similar arguments as in Section 4.2, $\dot{x}_2(t) = 0$ can be set alternatively in (26). Additionally ignoring the terms depending on $N_{k,21}$ and $N_{k,22}$, we obtain $x_2(t) = -A_{22}^{-1}\left(A_{21}x_1(t) + B_2u(t)\right)$. Inserting this for $x_2$ in (26), a ROM (5) is obtained that has a different structure than (3). The associated matrices are

$$\left(\tilde{A}_B, \tilde{B}, \tilde{C}_B, \tilde{D}, \tilde{E}_k, \tilde{N}_{B,k}\right) = \left(\bar{A}, \bar{B}, \bar{C}, \bar{D}, \bar{E}_k, \bar{N}_k\right), \tag{41}$$

where $\bar{A} := A_{11} - A_{12}A_{22}^{-1}A_{21}$, $\bar{B} := B_1 - A_{12}A_{22}^{-1}B_2$, $\bar{C} := C_1 - C_2A_{22}^{-1}A_{21}$, $\bar{D} := -C_2A_{22}^{-1}B_2$, $\bar{E}_k := -N_{k,12}A_{22}^{-1}B_2$ and $\bar{N}_k := N_{k,11} - N_{k,12}A_{22}^{-1}A_{21}$. It is important to mention that the main motivation behind the ROM given by (41) is an error bound based on the sum of truncated HSVs that we state below. This shows the actual benefit of the structure change.

**Remark 4.** *Notice that both balancing related methods above preserve the type of stability given in* (11)*. If* $\tilde{A}_B$ *and* $\tilde{N}_{B,k}$ *are as in* (40) *or* (41) *and if additionally* $\Sigma > 0$ *and* $\sigma(\Sigma_1) \cap \sigma(\Sigma_2) = \emptyset$*, we have*

$$\sigma(\tilde{A}_B \otimes I + I \otimes \tilde{A}_B + \frac{1}{\gamma^2}\sum_{k=1}^{m} \tilde{N}_{B,k} \otimes \tilde{N}_{B,k}) \subset \mathbb{C}_-.$$

*This was proved in [8, Theorem II.2] for BT and shown in [19, Section 4.2] for SPA in the context of stochastic systems.*

**Theorem 4.4.** *Let* $y_B$ *be the output of* (3) *given that* (11) *holds for a sufficiently large* $\gamma > 0$*. Suppose that* $\tilde{y}_B$ *is the reduced order output of system* (5)*, where the matrices* $\left(\tilde{A}_B, \tilde{B}, \tilde{C}_B, \tilde{D}, \tilde{E}_k, \tilde{N}_{B,k}\right)$ *are either given by BT stated in* (40) *or by SPA defined in*

(41). *Then, we have*

$$\|y_B - \tilde{y}_B\|_{L^2} \leq \left( 2 \sum_{i=r_B+1}^{n} \sigma_i \right) \exp\left\{ 0.5 \left\| \gamma u^0 \right\|_{L^2}^2 \right\} \|u\|_{L^2} ,$$

*where* $\sigma_{r_B+1}, \ldots, \sigma_n$ *are the truncated small HSVs of system* (3).

*Proof.* The above results for $\gamma = 1$ are special cases of the ones in [21, Theorem 3.1] (BT) and [20, Theorem 3] (SPA). Rescaling $N_k x_B(t) u_k(t) \mapsto \frac{1}{\gamma} N_k x_B(t) \gamma u_k(t)$ in (3a) provides the formulation of this theorem for general $\gamma$. □

Theorem 4.4 shows that the truncated HSVs determine the error of the approximation. Hence, these values are a good indicator for the expected error and can be chosen to find a suitable reduced order dimension $r_B$. Notice that for the HSVs it holds that $\sigma_i = \sigma_i(\gamma)$ since the underlying Gramians $P_B$ and $Q$ depend on the rescaling factor $\gamma$.

### 4.4 Model order reduction and an error bound for (1)

In this section, we exploit the results of Sections 4.2 and 4.3 in order to find an error bound between the output $y$ of (1) and some reduced output $\tilde{y}$ which we construct as the sum of the outputs $\tilde{y}_{x_0}$ and $\tilde{y}_B$ of subsystems (4) and (5). We discussed BT and SPA for the homogeneous and the inhomogeneous part of the bilinear system in order to obtain $\tilde{y}_{x_0}$ and $\tilde{y}_B$. All approaches are designed to provide an error bound in $L^2$, which enables us to combine them in the following theorem.

**Theorem 4.5.** *Suppose that* (11) *holds for a sufficiently large* $\gamma > 0$. *Let* $y$ *be the output of the original system* (1) *and let us define the reduced output* $\tilde{y} = \tilde{y}_{x_0} + \tilde{y}_B$, *where* $\tilde{y}_{x_0}$ *is the quantity of interest in* (4) *and* $\tilde{y}_B$ *the one of* (5). *We assume that* (4) *is the ROM of either BT stated in* (27) *or by SPA defined in* (28) *with* $\Theta_1 > 0$, *state dimension* $r_{x_0}$ *and an existing reduced Gramian* $\tilde{P}$ *for SPA. Let* (5) *be an* $r_B$*-dimensional ROM computed by BT with matrices* (40) *or by SPA defined through* (41). *Then, there exist a matrix* $\mathcal{W}_{x_0}$ *defined in Theorem 4.3 such that*

$$\|y - \tilde{y}\|_{L^2} \leq \left[ \left( \operatorname{tr}(\Theta_2 \mathcal{W}_{x_0}) \right)^{\frac{1}{2}} \|v_0\|_2 + \left( 2 \sum_{i=r_B+1}^{n} \sigma_i \right) \|u\|_{L^2} \right] \exp\left\{ 0.5 \left\| \gamma u^0 \right\|_{L^2}^2 \right\}$$

*with* $\Theta_2 = \operatorname{diag}(\theta_{r_{x_0}+1}, \ldots, \theta_n)$, *where* $\theta_{r_{x_0}+1}, \ldots, \theta_n$ *and* $\sigma_{r_B+1}, \ldots, \sigma_n$ *are the truncated small HSVs of* (2) *and* (3), *respectively.*

*Proof.* Let us recall that $y$ can be written as $y_{x_0} + y_B$, where $y_{x_0}$ is the output of (2) and $y_B$ the one of (3). Consequently, we have

$$\|y - \tilde{y}\|_{L^2} = \|y_{x_0} + y_B - (\tilde{y}_{x_0} + \tilde{y}_B)\|_{L^2} \leq \|y_{x_0} - \tilde{y}_{x_0}\|_{L^2} + \|y_B - \tilde{y}_B\|_{L^2}$$

using the triangle inequality. The claim now follows by Theorems 4.3 and 4.4. □

Theorem 4.5 indicates that one finds a good approximation $\tilde{y}$ for the output $y$ of the large-scale system (1) with non-zero initial states if each individual subsystem of (1) is reduced such that the associated truncated HSVs are small. Depending on the decay, the number of truncated HSVs can differ in each subsystem leading to reduced order dimensions $r_{x_0} \neq r_B$. The result of Theorem 4.5 is the generalization of the error bound for the linear case proved in [4, Theorem 3.2] if BT is applied to both subsystems (2) and (3). The result for the case of SPA as well as the combination of BT and SPA is new even for linear systems.

The representation of the error bound nicely shows the relation between the error and the truncated HSVs which makes these values a good a-priori criterion for the choice of the reduced order dimensions. However, the first part of the bound is not suitable for practical computation as $\mathcal{W}_{x_0}$ depends on the full balanced realization (22) which is not computed in practice. Instead one can use the general representation $\mathcal{E} = \operatorname{tr}(CP_0C^\top) + \operatorname{tr}(\tilde{C}_{x_0}\tilde{P}\tilde{C}_{x_0}^\top) - 2\operatorname{tr}(C\widehat{P}\tilde{C}_{x_0}^\top)$ from which $\operatorname{tr}(\Theta_2\mathcal{W}_{x_0})$ was derived at the beginning of the proof of Theorem 4.3. $\mathcal{E}$ is easily available since it involves the Gramian $P_0$ (which needs to be computed anyway to derive the reduced system) as well as the reduced Gramian $\tilde{P}$ and the solution $\widehat{P}$ of (36) (both computationally cheap). Of course $\mathcal{E}$ then is an a-posteriori bound but still very powerful in order to determine an estimate for the reduction quality. The representation $\mathcal{E}$ provides another strategy in reducing (2) since $\sqrt{\mathcal{E}}$ is the $\mathcal{H}_2$-distance between systems (3) and (5), where $(B, \tilde{B})$ are replaced by $(X_0, \tilde{X}_0)$, see [26]. Necessary conditions for local minimality have been provided in [26] and a method called bilinear iterative rational Krylov algorithm (BIRKA) was proposed in [5] satisfying these conditions. Therefore, one can also use BIRKA instead of a balancing related scheme in order to keep the first summand of the bound in Theorem 4.5 small.

The second part of the bound in Theorem 4.5 can be calculated once the Gramian $P_B$, satisfying the linear matrix inequality (LMI) (19), is computed. At the moment, LMI solver can only solve problems in moderate high dimensions, which might not be sufficient for some practical applications. However, once efficient computational methods are available, considering a Gramian like $P_B$ is very useful due to the a-priori $L^2$-error bound. In fact, only an $L^2$-bound is compatible with the bound in Theorem 4.3. One might also think of choosing a Gramian like $P_0$ satisfying (12) for subsystem (3) as proposed in [1]. However, an $L^2$-error bound does not exist in such a case indicating the relevance of the new approach of choosing some LMI solution as a Gramian.

**Remark 5.** *The value $\gamma > 0$ was introduced in order to ensure* (11) *which is a condition ensuring the existence of the Gramians. On the other hand, $\gamma$ can also be seen as an optimization parameter since the ROMs, and the HSVs depend on $\gamma$. This value was chosen equally in both subsystems, as one can see in Theorem 4.5 but certainly they can also be different if this leads to a better reduction quality.*

## 5 Numerical results

In this section, we conduct some numerical experiments illustrating the efficiency of the proposed MOR schemes. All the simulations are done on a CPU 2,2 GHz Intel® Core™i7, 16 GB 2400 MHz DDR4, MATLAB® 9.1.0.441655 (R2016b).
We consider a standard test case model representing a 2D boundary controlled heat transfer system as described in [6]. The model is governed by the following boundary value problem

$$\begin{aligned} \partial_t X(t,\xi) &= \Delta X(t,\xi), && \text{if} \quad \xi \in (0,1)\times(0,1), \quad \text{and} \quad t>0,\\ X(t,\xi) &= u(t), && \text{if} \quad \xi \in \Gamma_1,\\ n\cdot\nabla X(t,\xi) &= u(t)X(t,\xi), && \text{if} \quad \xi \in \Gamma_3,\\ X(t,\xi) &= 0, && \text{if} \quad \xi \in \Gamma_2 \cup \Gamma_4, \end{aligned}$$

where $\Gamma_1 = \{0\}\times(0,1)$, $\Gamma_2 = (0,1)\times\{0\}$, $\Gamma_3 = \{1\}\times(0,1)$ and $\Gamma_4 = (0,1)\times\{1\}$. Here the heat transfer term $u$ acting on $\Gamma_1$ and $\Gamma_3$ is the input variable. Moreover, we assume that the initial temperature is positive and constant in space, i.e.,

$$X(0,\xi) = 0.1, \quad \xi \in (0,1)\times(0,1).$$

A semi-discretization in space using finite differences with $k = 10$ equidistant grid points in each direction leads to a bilinear system of dimension $n = k^2 = 100$ having the form

$$\dot{x}(t) = Ax(t) + Nx(t)u(t) + Bu(t), \quad x(0) = X_0 v_0, \tag{42a}$$

$$y(t) = Cx(t), \tag{42b}$$

where $X_0 = \begin{bmatrix} 1 & \dots & 1 \end{bmatrix}^\top$, $v_0 = 0.1$ and $C = \frac{1}{n}\begin{bmatrix} 1 & \dots & 1 \end{bmatrix}$ (see [6] for more details).
Firstly, in order to apply the proposed techniques, one need to compute $P_0$, $Q$ and $P_B$ satisfying equations (12), (16) and (19), respectively. As shown in [11], by applying the Schur complement condition, (19) can be equivalently written as the following linear matrix inequality

$$\begin{bmatrix} AP_B + P_B A^\top + BB^\top & P_B N^\top \\ NP_B & -P_B \end{bmatrix} \leq 0. \tag{43}$$

Hence, we use the YALMIP toolbox [17] to the cost function $\mathrm{tr}(P_B)$ subject to (43) in order to find a good candidate for the Gramian.
Then, we compute the Hankel singular values associated to subsystems (2) and (3) using, respectively, the pair of Gramians $(P_{x_0}, Q)$ and $(P_B, Q)$. The resulting Hankel singular values are depicted in Figure 1. We notice a fast decay of these values, and hence, we expect accurate reduced models already for small reduced order dimensions as a consequence of Theorems 4.3 and 4.4.
For subsystem (2), we construct ROMs of orders 5 and 10 using balanced truncation (referred here as `BT`) and SPA (referred here as `SPA`) based on the Gramians $P_0$ and $Q$. Similarly, for subsystem (3), we construct ROMs of order 5 and 10 using balanced truncation (referred here as `BT_2`) and SPA (referred here as `SPA_2`) based on the Gramians

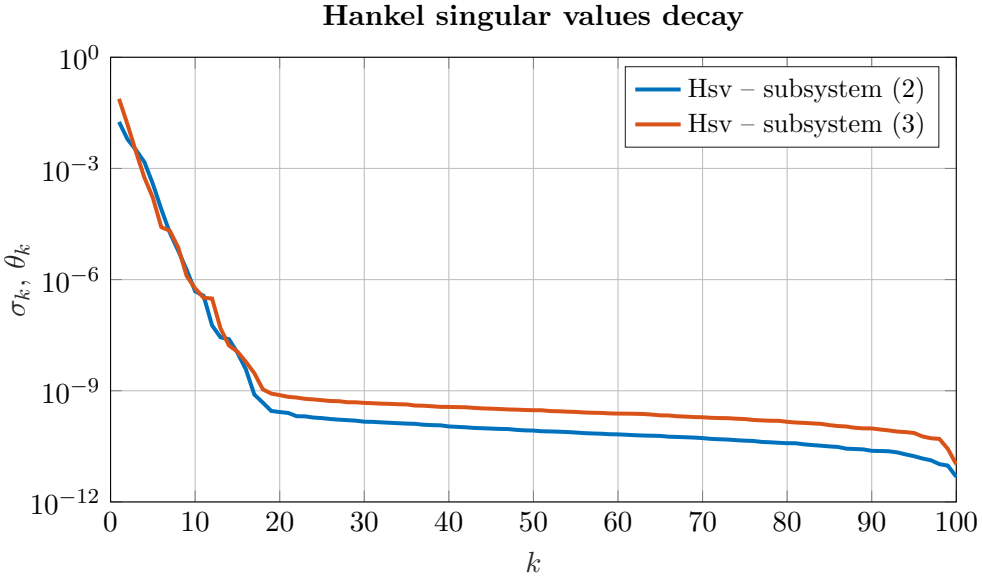


Figure 1: Decay of Hankel singular values for the subsystems (2) and (3).

| Method | Err. bound $r_{x_0} = 5$ | $L^2$-err. $r_{x_0} = 5$ | Err. bound $r_{x_0} = 10$ | $L^2$-err. $r_{x_0} = 10$ |
|---|---|---|---|---|
| `BT` | $7.64 \cdot 10^{-5}$ | $2.80 \cdot 10^{-5}$ | $1.32 \cdot 10^{-6}$ | $6.76 \cdot 10^{-7}$ |
| `SPA` | $1.53 \cdot 10^{-4}$ | $5.33 \cdot 10^{-5}$ | $2.57 \cdot 10^{-6}$ | $1.77 \cdot 10^{-6}$ |

Table 1: $L^2$-errors for subsystem (2): Error bounds in Theorem 4.3 and real errors for `BT` and `SPA` for the simulation presented in Figure 2.

| Method | Err. bound $r_B = 5$ | $L^2$-err. $r_B = 5$ | Err. bound $r_B = 10$ | $L^2$-err. $r_B = 10$ |
|---|---|---|---|---|
| `BT_2` | $1.69 \cdot 10^{-4}$ | $3.25 \cdot 10^{-5}$ | $7.88 \cdot 10^{-7}$ | $5.31 \cdot 10^{-8}$ |
| `SPA_2` | $1.69 \cdot 10^{-4}$ | $3.42 \cdot 10^{-6}$ | $7.88 \cdot 10^{-7}$ | $4.00 \cdot 10^{-9}$ |

Table 2: $L^2$-errors for subsystem (3): Error bound in Theorem 4.4 and real errors for `BT_2` and `SPA_2` for the simulation presented in Figure 3.

$P_B$ and $Q$. In order to compare the quality of ROMs, we simulate the original system and the reduced models using the input $u(t) = e^{-t}\cos(0.5t)$, $t \in [0, 1]$. In Figure 2, the pointwise absolute errors for `BT` and `SPA` are depicted, i.e., the function $\epsilon(t) = |y(t) - \tilde{y}(t)|$ is computed, where $y$ and $\tilde{y}$ are, receptively, the original and reduced outputs. One can observe that the results improve once the reduced order is increased. Both methodologies produce ROMs with a similar accuracy. Similarly, in Figure 3 the pointwise absolute error plots for `BT_2` and `SPA_2` are presented. We notice that, for this example, `SPA_2` produces ROMs with a higher accuracy than `BT_2`. Additionally, in Tables 1 and 2, the $L^2$-error values for the time interval $[0, 1]$ together with the corresponding error bounds for the different methods are shown, respectively, for subsystems (2) and (3).

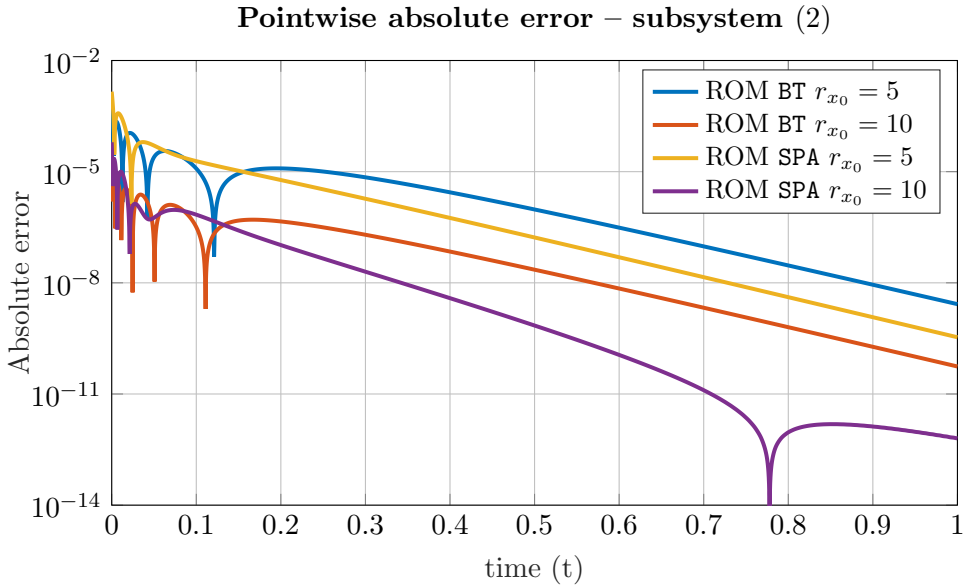


Figure 2: The pointwise absolute error for the approximations of subsystem (2) with input $u(t) = e^{-t}\cos(0.5t)$.

In Figure 4, we depict the decay of normalized $L^2$-error bounds from Theorem 4.3 together with the real $L^2$-errors produced by the time domain simulations for the subsystem (2) using the methods `BT` and `SPA`. Normalized here means to be divided by the $L^2$-norm of the original system output, e.g., the normalized $L^2$-error for a given order is $\|y - \tilde{y}\|_{L^2}/\|y\|_{L^2}$, where $\tilde{y}$ is the reduced output. For this example, `BT` is performing slightly better than `SPA`. Similarly, in Figure 5 we depict the decay of the normalized $L^2$-error bound from Theorem 4.4 together with the normalized $L^2$-errors produced by the time domain simulations for the subsystem (3) using the methods `BT_2` and `SPA_2`. As stated before, `SPA_2` produces ROMs that are more accurate than `BT_2`.

## References


[1] S. A. Al-Baiyat and M. Bettayeb. A new model reduction scheme for k–power bilinear systems. *Proceedings of the 32nd IEEE Conference on Decision and Control*, pages 22–27, 1993.

[2] A. C. Antoulas, I. V. Gosea, and A. C. Ionita. Model Reduction of Bilinear Systems in the Loewner Framework. *SIAM J. Sci. Comput.*, 38(5):B889–B916, 2016.

[3] U. Bauer, P. Benner, and L. Feng. Model Order Reduction for Linear and Nonlinear Systems: A System-Theoretic Perspective. *Arch. Comput. Methods Eng.*, 21(4):331–358, 2014.

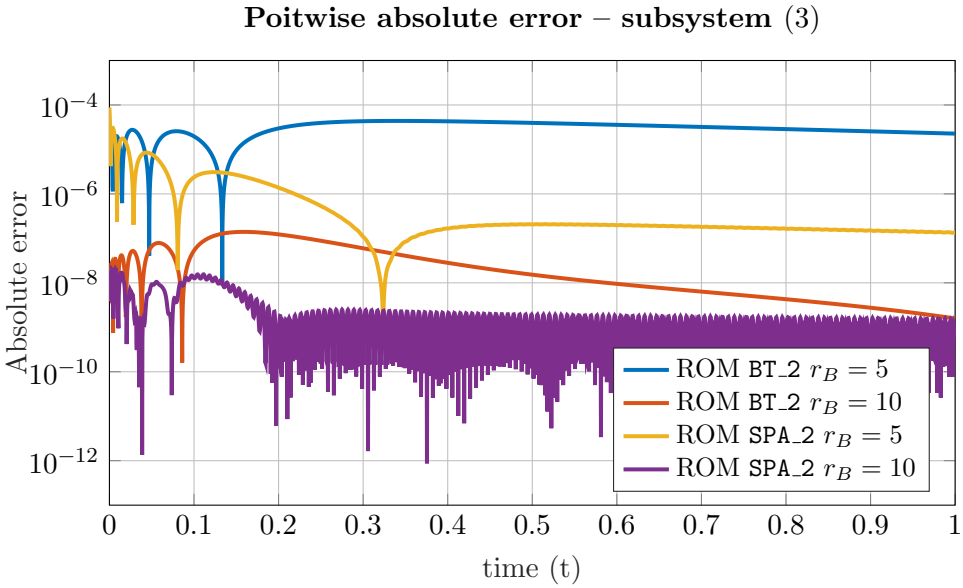


Figure 3: The pointwise absolute error for the approximations of subsystem (3) with input $u(t) = e^{-t}\cos(0.5t)$.


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

[5] P. Benner and T. Breiten. Interpolation-based $\mathcal{H}_2$-model reduction of bilinear control systems. *SIAM J. Matrix Anal. Appl.*, 33(3):859–885, 2012.

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

[7] P. Benner, T. Damm, M. Redmann, and Y. R. Rodriguez Cruz. Positive Operators and Stable Truncation. *Linear Algebra Appl*, 498:74–87, 2016.

[8] P. Benner, T. Damm, and Y. R. Rodriguez Cruz. Dual pairs of generalized Lyapunov inequalities and balanced truncation of stochastic linear systems. *IEEE Trans. Autom. Contr.*, 62(2):782–791, 2017.

[9] X. Cao, P. Benner, I. Pontes Duff, and W. Schilders. Model order reduction for bilinear control systems with inhomogeneous initial conditions. *Int. J. Control*, 2020.

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

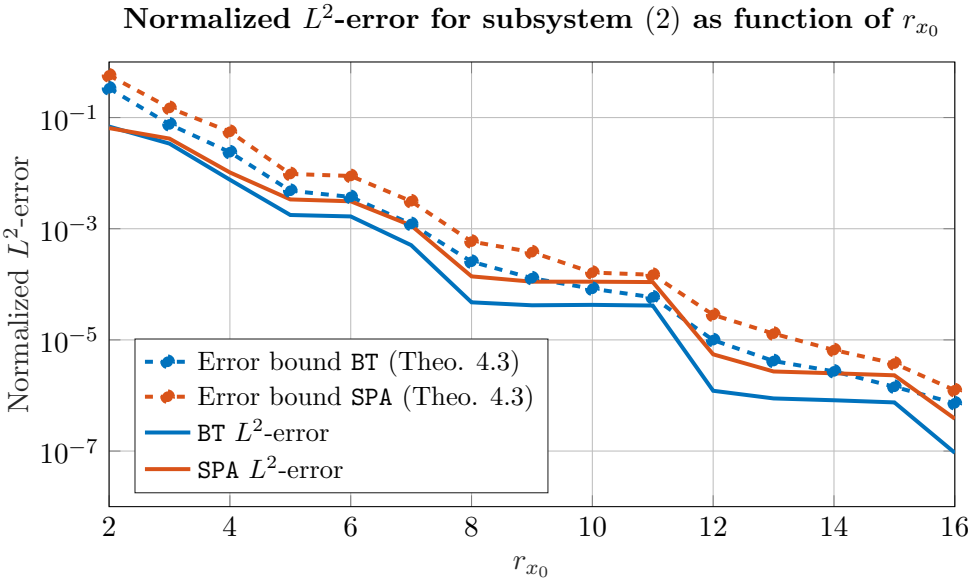


Figure 4: Subsystem (2): Decay of normalized error bound (Err. Bound/$\|y_{x_0}\|_{L^2}$) and normalized $L^2$-error ($\|y_{x_0} - \tilde{y}_{x_0}\|_{L^2}/\|y_{x_0}\|_{L^2}$) produced by `SPA` and `BT` for different orders $r_{x_0} = 1, \ldots, 16$.

.


[11] T. Damm and P. Benner. Balanced truncation for stochastic linear systems with guaranteed error bound. *Proceedings of MTNS–2014, Groningen, The Netherlands*, pages 1492–1497, 2014.

[12] A. Daragmeh, C. Hartmann, and N. Qatanani. Balanced model reduction of linear systems with nonzero initial conditions: Singular perturbation approximation. *Appl. Math. Comput.*, 353:295–307, 2019.

[13] G. Flagg and S. Gugercin. Multipoint Volterra Series Interpolation and $\mathcal{H}_2$ Optimal Model Reduction of Bilinear Systems. *SIAM J. Matrix Anal. Appl*, 36(2):549–579, 2015.

[14] C. Hartmann, B. Schäfer-Bung, and A. Thöns-Zueva. Balanced averaging of bilinear systems with applications to stochastic control. *SIAM J. Control Optim.*, 51(3):2356–2378, 2013.

[15] M. Heinkenschloss, T. Reis, and A. C. Antoulas. Balanced truncation model reduction for systems with inhomogeneous initial conditions. *Automatica*, 47(3):559–564, 2011.

[16] Y. Huang, Y.-L. Jiang, and Z.-Y. Qiu. Splitting model reduction for bilinear control systems. *Asian Journal of Control*, 2021.

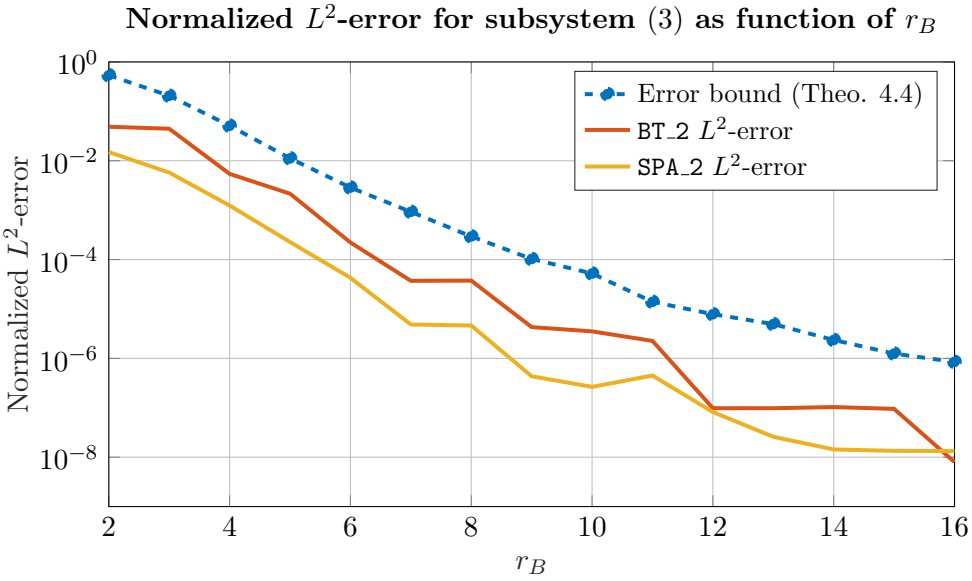


Figure 5: Subsystem (3): Decay of the normalized error bound (Err. Bound/$\|y_B\|_{L^2}$) and normalized $L^2$-error ($\|y_B - \tilde{y}_B\|_{L^2}/\|y_B\|_{L^2}$) produced by `SPA_2` and `BT_2` for different orders $r_B = 1, \ldots, 16$.

.


[17] J. Lofberg. YALMIP: A toolbox for modeling and optimization in MATLAB. In *2004 IEEE international conference on robotics and automation (IEEE Cat. No. 04CH37508)*, pages 284–289. IEEE, 2004.

[18] M. Redmann. *Model Order Reduction Techniques Applied to Evolution Equations with Lévy Noise*. PhD thesis, Otto-von-Guericke-Universität Magdeburg, 2016.

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

[20] M. Redmann. A new type of singular perturbation approximation for stochastic bilinear systems. *Math. Control. Signals, Syst.*, 32:129–156, 2020.

[21] M. Redmann. Energy estimates and model order reduction for stochastic bilinear systems. *Int. J. Control*, 93(8):1954–1963, 2020.

[22] M. Redmann. Bilinear systems – A new link to $\mathcal{H}_2$-norms, relations to stochastic systems and further properties. *SIAM J. Control Optim.*, Accepted 2021.

[23] M. Redmann and P. Benner. Singular Perturbation Approximation for Linear Systems with Lévy Noise. *Stochastics and Dynamics*, 18(4), 2018.

[24] M. Redmann and I. Pontes Duff. Full state approximation by Galerkin projection reduced order models for stochastic and bilinear systems. *arXiv preprint: 2102.07534*, 2021.

[25] C. Schröder and M. Voigt. Balanced Truncation Model Reduction with A Priori Error Bounds for LTI Systems with Nonzero Initial Value. *arXiv preprint: 2006.02495*, 2020.

[26] L. Zhang and J. Lam. On $H_2$ model reduction of bilinear systems. *Automatica*, 38(2):205–216, 2002.