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

Energy estimates and model order reduction for stochastic bilinear systems

Martin Redmann

submitted: May 3, 2018

Weierstrass Institute
Mohrenstr. 39
10117 Berlin
Germany
E-Mail: martin.redmann@wias-berlin.de

Energy estimates and model order reduction for stochastic bilinear systems

Martin Redmann

Abstract

In this paper, we investigate a large-scale stochastic system with bilinear drift and linear diffusion term. Such high dimensional systems appear for example when discretizing a stochastic partial differential equations in space. We study a particular model order reduction technique called balanced truncation (BT) to reduce the order of spatially-discretized systems and hence reduce computational complexity. We introduce suitable Gramians to the system and prove energy estimates that can be used to identify states which contribute only very little to the system dynamics. When BT is applied the reduced system is obtained by removing these states from the original system. The main contribution of this paper is an L^2-error bound for BT for stochastic bilinear systems. This result is new even for deterministic bilinear equations. In order to achieve it, we develop a new technique which is not available in the literature so far.

1 Introduction

Many phenomena in real life can be described by partial differential equations (PDEs). Famous examples are the motion of viscous fluids, the description of water or sound waves and the distribution of heat. 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. This leads to stochastic PDEs (SPDEs) [10, 16, 29, 30].

It is necessary to discretize a time-dependent SPDE in space and time in order to solve it numerically. Discretizing in space can be considered as a first step. This can be done for example by spectral Galerkin [18, 20, 21] or finite element methods [1, 23, 24]. This usually leads to a high dimensional SDE. Solving such complex SDE systems causes large computational cost. In this context, model order reduction (MOR) is used to save computational time by replacing large scale systems by systems of low order in which the main information of the original system should be captured.

1.1 Setting

We consider a large-scale stochastic control system with bilinear drift that can be interpreted as a spatially-discretized SPDE. The corresponding noise process is an $\mathbb{R}^v$-valued Lévy process $M = (M_1, \ldots, M_v)^T$ with mean zero and $\mathbb{E} \|M(t)\|_2^2 = \mathbb{E} \left[M^T(t) M(t) \right] < \infty$ for all $t \geq 0$. We investigate the system

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

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

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.,

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

for every $T > 0$. We denote the covariance matrix of M by $K = (k_{ij})_{i,j=1,\ldots,v}$. It characterizes the covariance function of M, see [29, Theorem 4.44], meaning that $\mathbb{E}[M(t) M^T(t)] = Kt$.

We aim to replace the large scale system (1) by a system of the same structure, but with a much smaller state dimension $r \ll n$. This reduced order model (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 this paper, we consider balanced truncation (BT) for obtaining a ROM. It relies on defining a reachability Gramian P and an observability Gramian Q. These matrices are selected, such that P characterizes the states in (1a) and Q the states in (1b) which barely contribute to the system dynamics (see Section 2 for details). In order to be able to remove both the unimportant states in (1a) and (1b) simultaneously, the first step of BT 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 $S = L_Q X \Sigma^{-\frac{1}{2}}$ 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., $SPS^T = S^{-T} Q S^{-1} = \Sigma = \text{diag}(\sigma_1, \ldots, \sigma_n)$ with $\sigma_1 \geq \ldots \geq \sigma_n$ being the Hankel singular values (HSVs) of the system. The ROM is then obtained by selecting the left upper $r \times r$ blocks of $\tilde{A}, \tilde{H}_i, \tilde{N}_k$, the first r rows of $\tilde{B}$ and the first r columns of $\tilde{C}$, such that the smallest $n - r$ HSVs are removed from the system.

Stochastic linear case ($N_k = 0$)   BT is a method that was developed for deterministic linear systems ($H_i = N_k = 0$) [2, 27, 28]. Basically, two ways of extending BT to stochastic linear systems have been considered so far. The so-called type I approach relies on defining Gramians based on generalized fundamental solutions of the system [5, 7]. The drawback of this generalization is that only an $\mathcal{H}_2$-error bound is available [7] and an $\mathcal{H}_\infty$-error bound cannot be achieved [6, 12]. To overcome this issue the type II ansatz was introduced in [12], where an $\mathcal{H}_\infty$-error bound is proved. There, a different reachability Gramian was considered which is defined as the solution to a matrix inequality. Energy estimates for linear stochastic systems for both the type I and the type II ansatz have recently been given in [32], such that MOR based on both approaches can be justified. As an alternative to BT, we want to refer to the singular perturbation approximation, where the work in [14, 25] was extended to stochastic linear systems in [32, 33].

Deterministic bilinear case ($H_i = 0$)   Although the bilinear term is a very weak non-linearity, deterministic bilinear system can be seen as bridge between linear and nonlinear systems. This is because many deterministic nonlinear systems can be represented by bilinear systems using a so-called Carleman linearization. Applications of these equations can be found in various fields [9, 26, 34]. Apart from balancing related MOR techniques [5, 19], various alternative methods have already been studied for the case of $H_i = 0$ [3, 4, 8, 15]. We, however, only discuss the case of BT below. When considering BT for deterministic bilinear systems, Gramians have to be chosen properly in order to find suitable characterizations for the reachability and observability energy of the system. Concerning the choice of the bilinear Gramians the control components u_k that are multiplied with the state x in (1a) are treated like white noise. Gramians according to the stochastic type I approach were, e.g., considered in [5, 17]. In both references energy estimates can be found, but they are only valid for states being in a possibly very small neighborhood of zero. Moreover, no error bound exists for this approach. Choosing the bilinear Gramian according to the stochastic type II approach has been considered in [31]. To be more precise, perturbed type II Gramians were used there. This has the advantage of finding global energy estimates under the assumption of having bounded controls. Furthermore, an $\mathcal{H}_\infty$-error bound was proved in [31], again, assuming bounded controls. Depending on the underlying system, the bound on the controls can be small which is the drawback of this method.

1.2   Outline of the paper and main result

The work in this paper on BT for system (1) can be interpreted as a generalization of the deterministic bilinear case, where the control u is perturbed by Lévy noise. We see this extension as a first step to find a bridge between stochastic linear systems and stochastic nonlinear systems to open the field of balancing related MOR to many more stochastic equations and applications.

In this paper, the main contributions are energy estimates for the stochastic bilinear system

and an error bound for BT. It is important to notice that these results are not just an extension of existing theory, they are new even for the deterministic bilinear case ($H_i = 0$), since in contrast to [31] no bound on the control is assumed. To be more precise, in Section 2, we propose Gramians P and Q to system (1). We show that the reachability Gramian P provides information about the degree of reachability of a state. Moreover, we establish a bound on the observation energy using the observability Gramian Q. The following result on an L^2-error bound for BT is proved in Section 3. For the proof of this theorem, the existing methods in the literature cannot be applied. Hence, we also provide a new technique to achieve this bound.

Theorem 1.1. *Let y be the output of the full model (1) with $x(0) = 0$ and y_r be the output of the ROM by BT with zero initial state. Then, for all $T > 0$, it holds that*

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

where $\sigma_{r+1}, \sigma_{r+2}, \ldots, \sigma_n$ are the smallest $n - r$ HSVs of system (1).

Theorem 1.1 implies that BT works well for stochastic bilinear systems if the truncated HSVs are small and the control energy is not too large.

2 Energy estimates

BT relies on the idea to create a system (1), in which the dominant reachable and observable states are the same. Afterwards, the unimportant states are removed to obtain an accurate approximation to the original model. In order to find the states that are hardly reachable and observable, a reachability Gramian P and an observability Gramian Q are introduced in this section. We will see that the definitions of the Gramians are meaningful, since they lead to estimates, which allow us to find the states that barely contribute to the system dynamics. This justifies to balance the system based on the proposed Gramians.

2.1 Reachability Gramian

We introduce a reachability Gramian P as a positive definite solution to

$$A^T P^{-1} + P^{-1} A + \sum_{k=1}^{m} N_k^T P^{-1} N_k + \sum_{i,j=1}^{v} H_i^T P^{-1} H_j k_{ij} \leq -P^{-1} B B^T P^{-1}. \quad (2)$$

An inequality is considered in (2), since the existence of a positive definite solution is not ensured when having an equality. The existence of a solution to (2) goes back to [12, 32] and is given if

$$\lambda\left(A \otimes I + I \otimes A + \sum_{k=1}^{m} N_k \otimes N_k + \sum_{i,j=1}^{v} H_i \otimes H_j k_{ij}\right) \subset \mathbb{C}_-, \quad (3)$$

which we assume to hold througout the remainder of the paper. Here, $\lambda\left(\cdot\right)$ denotes the spectrum of a matrix. Condition (3) is called mean square asymptotic stability [11, 22, 32]. It means that if the control components u_k in the bilinear term of (1a) would truely be white noise, then the second moment of the solution would tend to zero in the uncontrolled setting ($B = 0$) if $t \to \infty$.

Let $x(t, x_0, u)$ denote the solution to (1a) for $t \geq 0$, an initial state $x_0 \in \mathbb{R}^n$ and a control $u \in L_T^2$. We choose $(p_k)_{k=1,\dots,n}$ to be an orthonormal basis of $\mathbb{R}^n$ consisting of eigenvectors of P. We denote the corresponding eigenvalues by $(\lambda_k)_{k=1,\dots,n}$. For the Fourier coefficients of $x(t, 0, u)$, we obtain

$$
\langle x(t,0,u), p_k\rangle_2^2 \leq \lambda_k \sum_{i=1}^n \lambda_i^{-1} \langle x(t,0,u), p_i\rangle_2^2 = \lambda_k \left\| \sum_{i=1}^n \lambda_i^{-\frac{1}{2}} \langle x(t,0,u), p_i\rangle_2 \, p_i \right\|_2^2
$$

$$
= \lambda_k \left\| P^{-\frac{1}{2}} x(t,0,u) \right\|_2^2 = \lambda_k \, x^T(t,0,u) P^{-1} x(t,0,u). \tag{4}
$$

We use a shorter notation for the state below, i.e., we write $x(t)$ instead of $x(t, 0, u)$ if required. By Lemma A.1, we have

$$
\mathbb{E}\left[x^T(t)P^{-1}x(t)\right] = 2 \int_0^t \mathbb{E}\left[x^T(s)P^{-1}\left(Ax(s) + Bu(s) + \sum_{k=1}^m N_k x(s)u_k(s)\right)\right] ds \tag{5}
$$

$$
+ \int_0^t \mathbb{E}\left[x^T(s) \sum_{i,j=1}^v H_i^T P^{-1} H_j k_{ij} x(s)\right] ds.
$$

The bilinear term in the above equation can be bounded as follows:

$$
\sum_{k=1}^m 2 \int_0^t x^T(s)P^{-1}N_k x(s)u_k(s)ds = \sum_{k=1}^m 2 \int_0^t \left\langle P^{-\frac{1}{2}}x(s)u_k(s), P^{-\frac{1}{2}}N_k x(s)\right\rangle_2 ds
$$

$$
\leq \sum_{k=1}^m \left(\int_0^t \left\| P^{-\frac{1}{2}}x(s)u_k(s) \right\|_2^2 ds + \int_0^t \left\| P^{-\frac{1}{2}}N_k x(s) \right\|_2^2 ds \right)
$$

$$
= \int_0^t x^T(s)P^{-1}x(s) \left\| u(s) \right\|_2^2 ds + \int_0^t x^T(s) \sum_{k=1}^m N_k^T P^{-1} N_k x(s) ds. \tag{6}
$$

We insert this inequality into (5), such that

$$
\mathbb{E}\left[x^T(t)P^{-1}x(t)\right] \tag{7}
$$

$$
\leq \mathbb{E} \int_0^t x^T(s)(A^T P^{-1} + P^{-1}A + \sum_{k=1}^m N_k^T P^{-1} N_k + \sum_{i,j=1}^v H_i^T P^{-1} H_j k_{ij})x(s)ds
$$

$$
+ \mathbb{E} \int_0^t 2x^T(s)P^{-1}Bu(s)ds + \int_0^t \mathbb{E}\left[x^T(s)P^{-1}x(s)\right] \left\| u(s) \right\|_2^2 ds.
$$

We then plug in (2), which yields

$$
\begin{aligned}
\mathbb{E}\left[x^T(t)P^{-1}x(t)\right] \leq &- \mathbb{E}\int_0^t x^T(s)P^{-1}BB^TP^{-1}x(s)ds \\
&+ \mathbb{E}\int_0^t 2x^T(s)P^{-1}Bu(s)ds + \int_0^t \mathbb{E}\left[x^T(s)P^{-1}x(s)\right]\|u(s)\|_2^2\,ds \\
=&\,\mathbb{E}\int_0^t \|u(s)\|_2^2 - \left\|B^TP^{-1}x(s) - u(s)\right\|_2^2 ds \\
&+ \int_0^t \mathbb{E}\left[x^T(s)P^{-1}x(s)\right]\|u(s)\|_2^2\,ds \\
\leq &\int_0^t \|u(s)\|_2^2\,ds + \int_0^t \mathbb{E}\left[x^T(s)P^{-1}x(s)\right]\|u(s)\|_2^2\,ds
\end{aligned}
$$

The Gronwall inequality, see Lemma A.3, provides

$$
\mathbb{E}\left[x^T(t)P^{-1}x(t)\right] \leq \int_0^t \|u(s)\|_2^2\,ds \exp\left(\int_0^t \|u(s)\|_2^2\,ds\right).
$$

Consequently, by (4), we have

$$
\sup_{t\in[0,T]} \sqrt{\mathbb{E}\langle x(t,0,u), p_k\rangle_2^2} \leq \lambda_k^{0.5}\|u\|_{L_T^2}\exp\left(0.5\|u\|_{L_T^2}^2\right). \tag{8}
$$

Given a state it is not possible to gain information about the corresponding energy from (8). However, given a bound on the energy which is not too large, let us say $\|u\|_{L_T^2} \leq 1$, we can conclude how much a state component contributes to the systems dynamics. If λ_k is small, (8) implies that the Fourier coefficient $\langle x(\cdot,0,u), p_k\rangle_2$ is close to zero on $[0,T]$ for normalized controls u. This means that the state variable takes only very small values in the direction of p_k such that hardly reachable states have a large component in the eigenspaces of P belonging to the small eigenvalues. Inequality (8) has already been pointed out in [31, Remark 1] for the case $H_i = 0$.

2.2 Observability Gramian

We define the observability Gramian to be the solution to

$$
A^TQ + QA + \sum_{k=1}^m N_k^TQN_k + \sum_{i,j=1}^v H_i^TQH_j k_{ij} = -C^TC. \tag{9}
$$

Condition (3) guarantees the existence of a positive semidefinite solution [32], but we will furthermore assume that $Q > 0$ for the rest of the paper. Again, we use a short notation by setting $x_{x_0}(t) := x(t,x_0,u)$. In order to find a suitable estimate for $\mathbb{E}\left[x_{x_0}^T(t)Qx_{x_0}(t)\right]$, it

is only required to replace P^{-1} by Q in (7) and take into account the additional term $x_0^T Q x_0$ that is due to the non-zero initial condition. Hence,

$$\mathbb{E}\left[x_{x_0}^T(t)Qx_{x_0}(t)\right]$$
$$\leq \mathbb{E}\int_0^t x_{x_0}^T(s)(A^TQ + QA + \sum_{k=1}^m N_k^TQN_k + \sum_{i,j=1}^v H_i^TQH_jk_{ij})x_{x_0}(s)ds$$
$$+ \mathbb{E}\int_0^t 2x_{x_0}^T(s)QBu(s)ds + \int_0^t \mathbb{E}\left[x_{x_0}^T(s)Qx_{x_0}(s)\right]\|u(s)\|_2^2 ds + x_0^T Q x_0.$$

Applying (9) gives us

$$\mathbb{E}\left[x_{x_0}^T(t)Qx_{x_0}(t)\right] \leq -\mathbb{E}\int_0^t y^T(s)y(s)ds + \mathbb{E}\int_0^t 2x_{x_0}^T(s)QBu(s)ds + x_0^T Q x_0$$
$$+ \int_0^t \mathbb{E}\left[x_{x_0}^T(s)Qx_{x_0}(s)\right]\|u(s)\|_2^2 ds,$$

where $y(t) = y(t, x_0, u)$. Due to Lemma A.3, we find

$$\mathbb{E}\left[x_{x_0}^T(t)Qx_{x_0}(t)\right] \leq \alpha(t) + \int_0^t \alpha(s)\|u(s)\|_2^2 \exp\left(\int_s^t \|u(w)\|_2^2 dw\right)ds, \quad (10)$$

where we define $\alpha(t) := -\mathbb{E}\int_0^t \|y(s)\|_2^2 ds + \mathbb{E}\int_0^t 2x_{x_0}^T(s)QBu(s)ds + x_0^T Q x_0$. We analyze this inequality further by looking at the terms depending on x_0:

$$x_0^T Q x_0 \int_0^t \|u(s)\|_2^2 \exp\left(\int_s^t \|u(w)\|_2^2 dw\right)ds = x_0^T Q x_0 \left[-\exp\left(\int_s^t \|u(w)\|_2^2 dw\right)\right]_{s=0}^t$$
$$= x_0^T Q x_0 \left(\exp\left(\int_0^t \|u(s)\|_2^2 ds\right) - 1\right).$$

Using this computation and $\mathbb{E}\left[x_{x_0}^T(t)Qx_{x_0}(t)\right] \geq 0$, we get from (10) that

$$\mathbb{E}\int_0^t \|y(s)\|_2^2 ds \leq x_0^T Q x_0 \exp\left(\int_0^t \|u(s)\|_2^2 ds\right) \qquad (11)$$
$$+ f_B(t) + \int_0^t f_B(s)\|u(s)\|_2^2 \exp\left(\int_s^t \|u(w)\|_2^2 dw\right)ds,$$

where the term depending on the input matrix B is $f_B(t) := \mathbb{E}\int_0^t 2x_{x_0}^T(s)QBu(s)ds$. In an observation problem an unknown initial condition x_0 is aimed to be reconstructed from the observations $y(t, x_0, u)$, $t \in [0, T]$. Since the control part Bu does not depend on the unknown initial state, it can be assumed to be known and hence be neglected in the considerations by setting $B = 0$. This assumption is also taken in [5, 17], where the observation energy of deterministic bilinear systems is studied. Now, $B = 0$ implies $f_B \equiv 0$. Applying this to (11) leads to the following bound on the observation energy on $[0, T]$:

$$\mathbb{E}\int_0^T \|y(s, x_0, u)\|_2^2 ds\bigg|_{B=0} \leq x_0^T Q x_0 \exp\left(\int_0^T \|u(s)\|_2^2 ds\right). \qquad (12)$$

If we the energy of the control is sufficiently small, e.g., $\|u\|_{L_T^2} \leq 1$, we can identify states from (12) producing only little observation energy. We see that the energy that is caused by the observations of x_0 is small if the initial state is close to the kernel of Q. These initial states are contained in the eigenspaces of Q corresponding to the small eigenvalues.

3 L^2-error bound for BT

Let us assume that system (1) has a zero initial condition ($x_0 = 0$) and is already balanced. Thus, (2) and (9) become

$$A^T\Sigma^{-1} + \Sigma^{-1}A + \sum_{k=1}^{m} N_k^T\Sigma^{-1}N_k + \sum_{i,j=1}^{v} H_i^T\Sigma^{-1}H_j k_{ij} \leq -\Sigma^{-1}BB^T\Sigma^{-1}, \quad (13)$$

$$A^T\Sigma + \Sigma A + \sum_{k=1}^{m} N_k^T\Sigma N_k + \sum_{i,j=1}^{v} H_i^T\Sigma H_j k_{ij} \leq -C^TC, \quad (14)$$

i.e., $P = Q = \Sigma = \operatorname{diag}(\sigma_1, \ldots, \sigma_n) > 0$. We partition the balanced coefficients of (1) as follows:

$$A = \left[\begin{smallmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{smallmatrix}\right], \ B = \left[\begin{smallmatrix} B_1 \\ B_2 \end{smallmatrix}\right], \ N_k = \left[\begin{smallmatrix} N_{k,11} & N_{k,12} \\ N_{k,21} & N_{k,22} \end{smallmatrix}\right], \ H_i = \left[\begin{smallmatrix} H_{i,11} & H_{i,12} \\ H_{i,21} & H_{i,22} \end{smallmatrix}\right], \ C = \left[\begin{smallmatrix} C_1 & C_2 \end{smallmatrix}\right],$$

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 and the Gramian

$$x = \left[\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right] \text{ and } \Sigma = \left[\begin{smallmatrix} \Sigma_1 & \\ & \Sigma_2 \end{smallmatrix}\right],$$

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. The reduced system by BT is

$$dx_r = [A_{11}x_r + B_1 u + \sum_{k=1}^{m} N_{k,11}x_r u_k]dt + \sum_{i=1}^{v} H_{i,11}x_r dM_i, \quad (15a)$$

$$y_r(t) = C_1 x_r(t), \quad t \geq 0, \quad (15b)$$

where $x_r(0) = 0$ and the time dependence in (15a) is omitted to shorten the notation. In order to find a bound for the approximation through BT, we define

$$x_- = \left[\begin{smallmatrix} x_1 - x_r \\ x_2 \end{smallmatrix}\right] \text{ and } x_+ = \left[\begin{smallmatrix} x_1 + x_r \\ x_2 \end{smallmatrix}\right],$$

and write down the corresponding equations for these variables. The system for x_- is given by

$$dx_- = [Ax_- + \sum_{k=1}^{m} N_k x_- u_k]dt + \left[\begin{smallmatrix} 0 \\ c_0 \end{smallmatrix}\right]dt + \sum_{i=1}^{v} [H_i x_- + \left[\begin{smallmatrix} 0 \\ c_i \end{smallmatrix}\right]]dM_i, \quad (16a)$$

$$y_-(t) = Cx_-(t) = Cx(t) - C_1 x_r(t) = y(t) - y_r(t), \quad t \geq 0, \quad (16b)$$

where $c_0(t) := A_{21}x_r(t) + B_2u(t) + \sum_{k=1}^{m} N_{k,21}x_r(t)u_k(t)$ and $c_i(t) := H_{i,21}x_r(t)$ for $i = 1, \ldots, v$. We derive (16) by comparing the partitioned system (1) with the reduced system (15). The equation for x_+ looks similarly, the difference lies only in the signs for the compensation terms $c_0, \ldots, c_v$ and an additional control term:

$$dx_+ = [Ax_+ + 2Bu + \sum_{k=1}^{m} N_k x_+ u_k]dt - [\begin{smallmatrix} 0 \\ c_0 \end{smallmatrix}]\,dt + \sum_{i=1}^{v}[H_i x_+ - [\begin{smallmatrix} 0 \\ c_i \end{smallmatrix}]]dM_i. \qquad (17)$$

We will see that the proof of the error bound can be reduced to the task of finding suitable estimates for $\mathbb{E}[x_-^T(t)\Sigma x_-(t)]$ and $\mathbb{E}[x_+^T(t)\Sigma^{-1}x_+(t)]$. The next theorem is the main result of this paper.

Theorem 3.1. *Let y be the output of the full model (1) with $x(0) = 0$ and y_r be the output of the ROM (15) with $x_r(0) = 0$. Then, for all $T > 0$, it holds that*

$$\left(\mathbb{E}\left\|y - y_r\right\|_{L_T^2}^2\right)^{\frac{1}{2}} \le 2(\tilde{\sigma}_1 + \tilde{\sigma}_2 + \ldots + \tilde{\sigma}_\kappa)\left\|u\right\|_{L_T^2} \exp\left(0.5\left\|u\right\|_{L_T^2}^2\right),$$

where $\tilde{\sigma}_1, \tilde{\sigma}_2, \ldots, \tilde{\sigma}_\kappa$ are the distinct diagonal entries of $\Sigma_2 = \mathrm{diag}(\sigma_{r+1}, \ldots, \sigma_n) = \mathrm{diag}(\tilde{\sigma}_1 I, \tilde{\sigma}_2 I, \ldots, \tilde{\sigma}_\kappa I)$.

Proof. We compute an upper bound for $\mathbb{E}[x_-^T(t)\Sigma x_-(t)]$ making use of Lemma A.1. Taking (16a) into account then yields

$$\mathbb{E}\left[x_-^T(t)\Sigma x_-(t)\right] = 2\int_0^t \mathbb{E}\left[x_-^T\Sigma\left(Ax_- + \sum_{k=1}^{m}(N_k x_- u_k) + [\begin{smallmatrix} 0 \\ c_0 \end{smallmatrix}]\right)\right]ds \qquad (18)$$

$$+ \int_0^t \sum_{i,j=1}^{v} \mathbb{E}\left[(H_i x_- + [\begin{smallmatrix} 0 \\ c_i \end{smallmatrix}])^T \Sigma\left(H_j x_- + [\begin{smallmatrix} 0 \\ c_j \end{smallmatrix}]\right)\right]k_{ij}ds,$$

where the time dependence of all functions is omitted for a shorter notation. Applying an estimate as in (6) provides

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

Hence, (18) becomes

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

$$+ \mathbb{E}\int_0^t 2x_-^T\Sigma[\begin{smallmatrix} 0 \\ c_0 \end{smallmatrix}] + \sum_{i,j=1}^{v}(2H_i x_- + [\begin{smallmatrix} 0 \\ c_i \end{smallmatrix}])^T\Sigma[\begin{smallmatrix} 0 \\ c_j \end{smallmatrix}]k_{ij}ds \qquad (19)$$

$$+ \int_0^t \mathbb{E}\left[x_-^T\Sigma x_-\right]\left\|u\right\|_2^2 ds.$$

Using the partition for x_- and Σ, we see that $x_-^T \Sigma \begin{bmatrix} 0 \\ c_0 \end{bmatrix} = x_2^T \Sigma_2 c_0$. With the partition of H_i, we additionally obtain

$$
\begin{aligned}
\left(2H_i x_- + \begin{bmatrix} 0 \\ c_i \end{bmatrix}\right)^T \Sigma \begin{bmatrix} 0 \\ c_j \end{bmatrix} &= \left(2H_i x_- + \begin{bmatrix} 0 \\ c_i \end{bmatrix}\right)^T \begin{bmatrix} 0 \\ \Sigma_2 c_j \end{bmatrix} \\
&= \left(2\begin{bmatrix} H_{i,21} & H_{i,22} \end{bmatrix}\left(x - \begin{bmatrix} x_r \\ 0 \end{bmatrix}\right) + c_i\right)^T \Sigma_2 c_j = \left(2\begin{bmatrix} H_{i,21} & H_{i,22} \end{bmatrix} x - c_i\right)^T \Sigma_2 c_j.
\end{aligned}
$$

Inserting (14) and (16b) into inequality (19) and taking the above rearrangements into account leads to

$$
\begin{aligned}
\mathbb{E}\left[x_-^T(t)\Sigma x_-(t)\right] \leq &- \mathbb{E}\left\|y - y_r\right\|_{L_t^2}^2 + \int_0^t \mathbb{E}\left[x_-^T \Sigma x_-\right] \|u\|_2^2 \, ds \\
&+ \mathbb{E}\int_0^t 2x_2^T \Sigma_2 c_0 + \sum_{i,j=1}^{v} \left(2\begin{bmatrix} H_{i,21} & H_{i,22} \end{bmatrix} x - c_i\right)^T \Sigma_2 c_j k_{ij} ds.
\end{aligned}
$$

We define $\alpha_-(t) := \mathbb{E}\int_0^t 2x_2^T \Sigma_2 c_0 + \sum_{i,j=1}^{v} \left(2\begin{bmatrix} H_{i,21} & H_{i,22} \end{bmatrix} x - c_i\right)^T \Sigma_2 c_j k_{ij} ds$. Then, Lemma A.3 implies

$$
\begin{aligned}
\mathbb{E}\left[x_-^T(t)\Sigma x_-(t)\right] \leq &\alpha_-(t) - \mathbb{E}\left\|y - y_r\right\|_{L_t^2}^2 \\
&+ \int_0^t \left(\alpha_-(s) - \mathbb{E}\left\|y - y_r\right\|_{L_s^2}^2\right) \|u(s)\|_2^2 \exp\left(\int_s^t \|u(w)\|_2^2 \, dw\right) ds.
\end{aligned}
$$

Thus, we find

$$
\mathbb{E}\left\|y - y_r\right\|_{L_t^2}^2 \leq \alpha_-(t) + \int_0^t \alpha_-(s) \|u(s)\|_2^2 \exp\left(\int_s^t \|u(w)\|_2^2 \, dw\right) ds.
$$

We assume for the moment that $\Sigma_2 = \sigma I$ and set
$\alpha_+(t) := \mathbb{E}\int_0^t 2x_2^T \Sigma_2^{-1} c_0 + \sum_{i,j=1}^{v} \left(2\begin{bmatrix} H_{i,21} & H_{i,22} \end{bmatrix} x - c_i\right)^T \Sigma_2^{-1} c_j k_{ij} ds$. Hence,

$$
\mathbb{E}\left\|y - y_r\right\|_{L_t^2}^2 \leq \sigma^2 \left[\alpha_+(t) + \int_0^t \alpha_+(s) \|u(s)\|_2^2 \exp\left(\int_s^t \|u(w)\|_2^2 \, dw\right) ds\right]. \tag{20}
$$

Let us turn our attention to the expression $\mathbb{E}[x_+^T(t)\Sigma^{-1}x_+(t)]$ for the further analysis of (20). Due to (17) and Lemma A.1 it holds that

$$
\mathbb{E}\left[x_+^T(t)\Sigma^{-1}x_+(t)\right] = 2\int_0^t \mathbb{E}\left[x_+^T \Sigma^{-1}\left(Ax_+ + 2Bu + \sum_{k=1}^{m}(N_k x_+ u_k) - \begin{bmatrix} 0 \\ c_0 \end{bmatrix}\right)\right] ds \tag{21}
$$

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

As above, for the case of x_-, we use the estimate

$$
\sum_{k=1}^{m} 2x_+^T(s)\Sigma^{-1} N_k x_+(s) u_k(s) \leq x_+^T(s)\Sigma^{-1}x_+(s)\|u(s)\|_2^2 + \sum_{k=1}^{m} x_+^T(s) N_k^T \Sigma^{-1} N_k x_+(s),
$$

which leads to

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

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

$$
-\mathbb{E}\int_0^t 2x_+^T\Sigma^{-1}\left[\begin{smallmatrix}0\\c_0\end{smallmatrix}\right]+\sum_{i,j=1}^v\left(2H_ix_+-\left[\begin{smallmatrix}0\\c_i\end{smallmatrix}\right]\right)^T\Sigma^{-1}\left[\begin{smallmatrix}0\\c_j\end{smallmatrix}\right]k_{ij}ds \tag{22}
$$

$$
+\int_0^t \mathbb{E}\left[x_+^T\Sigma^{-1}x_+\right]\|u\|_2^2\,ds+4\mathbb{E}\int_0^t x_+^T\Sigma^{-1}Bu ds.
$$

From inequality (13) and the Schur complement condition on definiteness it follows that

$$
\begin{bmatrix} A^T\Sigma^{-1}+\Sigma^{-1}A+\sum_{k=1}^m N_k^T\Sigma^{-1}N_k+\sum_{i,j=1}^v H_i^T\Sigma^{-1}H_j k_{ij} & \Sigma^{-1}B \\ B^T\Sigma^{-1} & -I \end{bmatrix}\leq 0. \tag{23}
$$

We multiply (23) with $\left[\begin{smallmatrix}x_+\\2u\end{smallmatrix}\right]^T$ from the left and with $\left[\begin{smallmatrix}x_+\\2u\end{smallmatrix}\right]$ from the right. This leads to

$$
4\|u\|_2^2\geq
$$

$$
x_+^T\left(A^T\Sigma^{-1}+\Sigma^{-1}A+\sum_{k=1}^m N_k^T\Sigma^{-1}N_k+\sum_{i,j=1}^v H_i^T\Sigma^{-1}H_j k_{ij}\right)x_+ +4x_+^T\Sigma^{-1}Bu.
$$

Applying this result to inequality (22) gives

$$
\mathbb{E}\left[x_+^T(t)\Sigma^{-1}x_+(t)\right]\leq 4\|u\|_{L_t^2}^2+\int_0^t \mathbb{E}\left[x_+^T\Sigma^{-1}x_+\right]\|u\|_2^2\,ds \tag{24}
$$

$$
-\mathbb{E}\int_0^t 2x_+^T\Sigma^{-1}\left[\begin{smallmatrix}0\\c_0\end{smallmatrix}\right]+\sum_{i,j=1}^v\left(2H_ix_+-\left[\begin{smallmatrix}0\\c_i\end{smallmatrix}\right]\right)^T\Sigma^{-1}\left[\begin{smallmatrix}0\\c_j\end{smallmatrix}\right]k_{ij}ds.
$$

We further analyze the terms in (24). We find 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

$$
\left(2H_ix_+-\left[\begin{smallmatrix}0\\c_i\end{smallmatrix}\right]\right)^T\Sigma^{-1}\left[\begin{smallmatrix}0\\c_j\end{smallmatrix}\right]=\left(2H_ix_+-\left[\begin{smallmatrix}0\\c_i\end{smallmatrix}\right]\right)^T\left[\begin{smallmatrix}0\\\Sigma_2^{-1}c_j\end{smallmatrix}\right]
$$

$$
=\left(2\left[\begin{smallmatrix}H_{i,21} & H_{i,22}\end{smallmatrix}\right]\left(x+\left[\begin{smallmatrix}x_r\\0\end{smallmatrix}\right]\right)-c_i\right)^T\Sigma_2^{-1}c_j=\left(2\left[\begin{smallmatrix}H_{i,21} & H_{i,22}\end{smallmatrix}\right]x+c_i\right)^T\Sigma_2 c_j.
$$

We plug this into (24), such that

$$
\mathbb{E}\left[x_+^T(t)\Sigma^{-1}x_+(t)\right]\leq 4\|u\|_{L_t^2}^2+\int_0^t \mathbb{E}\left[x_+^T\Sigma^{-1}x_+\right]\|u\|_2^2\,ds \tag{25}
$$

$$
-\mathbb{E}\int_0^t 2x_2^T\Sigma_2^{-1}c_0+\sum_{i,j=1}^v\left(2\left[\begin{smallmatrix}H_{i,21} & H_{i,22}\end{smallmatrix}\right]x+c_i\right)^T\Sigma_2^{-1}c_j k_{ij}ds.
$$

Adding $2\mathbb{E}\int_0^t \sum_{i,j=1}^v c_i^T \Sigma_2^{-1} c_j k_{ij} ds$ to the right side of (25), which is a nonnegative term due to Lemma A.2, we have

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

Lemma A.3 yields

$$\mathbb{E}\left[x_+^T(t)\Sigma^{-1}x_+(t)\right] \le 4\left\|u\right\|_{L_t^2}^2 - \alpha_+(t) \tag{26}$$
$$+ \int_0^t (4\left\|u\right\|_{L_s^2}^2 - \alpha_+(s))\left\|u(s)\right\|_2^2 \exp\left(\int_s^t \left\|u(w)\right\|_2^2 dw\right) ds.$$

Moreover, we have

$$\int_0^t \left\|u\right\|_{L_s^2}^2 \left\|u(s)\right\|_2^2 \exp\left(\int_s^t \left\|u(w)\right\|_2^2 dw\right) ds \le \left\|u\right\|_{L_t^2}^2 \left[-\exp\left(\int_s^t \left\|u(w)\right\|_2^2 dw\right)\right]_{s=0}^t$$
$$= \left\|u\right\|_{L_t^2}^2 \left(\exp\left(\int_0^t \left\|u(s)\right\|_2^2 ds\right) - 1\right). \tag{27}$$

Combining (26) with (27), we get

$$\alpha_+(t) + \int_0^t \alpha_+(s)\left\|u(s)\right\|_2^2 \exp\left(\int_s^t \left\|u(w)\right\|_2^2 dw\right) ds \le 4\left\|u\right\|_{L_t^2}^2 \exp\left(\int_0^t \left\|u(s)\right\|_2^2 ds\right).$$

Comparing this result with (20) implies

$$\left(\mathbb{E}\left\|y - y_r\right\|_{L_t^2}^2\right)^{\frac{1}{2}} \le 2\sigma \left\|u\right\|_{L_t^2} \exp\left(0.5\left\|u\right\|_{L_t^2}^2\right). \tag{28}$$

For the proof of a general Σ_2, we remove the HSVs step by step. We use the triangle inequality to bound the error between the outputs y and y_r:

$$\left(\mathbb{E}\left\|y - y_r\right\|_{L_T^2}^2\right)^{\frac{1}{2}}$$
$$\le \left(\mathbb{E}\left\|y - y_{r_\kappa}\right\|_{L_T^2}^2\right)^{\frac{1}{2}} + \left(\mathbb{E}\left\|y_{r_\kappa} - y_{r_{\kappa-1}}\right\|_{L_T^2}^2\right)^{\frac{1}{2}} + \ldots + \left(\mathbb{E}\left\|y_{r_2} - y_r\right\|_{L_T^2}^2\right)^{\frac{1}{2}},$$

where the dimensions r_i of the corresponding states are defined by $r_{i+1} = r_i + m(\tilde\sigma_i)$ for $i = 1, 2\ldots, \kappa - 1$. The number $m(\tilde\sigma_i)$ denotes the multiplicity of $\tilde\sigma_i$ and $r_1 = r$. In the first step only the smallest HSV $\tilde\sigma_\kappa$ is removed from the system. By inequality (28), we have

$$\left(\mathbb{E}\left\|y - y_{r_\kappa}\right\|_{L_T^2}^2\right)^{\frac{1}{2}} \le 2\tilde\sigma_\kappa \left\|u\right\|_{L_T^2} \exp\left(0.5\left\|u\right\|_{L_T^2}^2\right).$$

The same kind of bound can be established when comparing the reduced order outputs y_{r_κ} and $y_{r_{\kappa-1}}$. Again, only one HSV, namely $\tilde\sigma_{r_{\kappa-1}}$, is removed. Moreover, the matrix inequalities

in the ROM have the same structure as (13) and (14). To be more precise, evaluating the left upper blocks of (13) and (14), we obtain

$$A_{11}^T \Sigma_1^{-1} + \Sigma_1^{-1} A_{11} + \sum_{k=1}^m N_{k,11}^T \Sigma_1^{-1} N_{k,11} + \sum_{i,j=1}^v H_{i,11}^T \Sigma_1^{-1} H_{j,11} k_{ij} \leq -\Sigma_1^{-1} B_1 B_1^T \Sigma_1^{-1},$$

$$A_{11}^T \Sigma_1 + \Sigma_1 A_{11} + \sum_{k=1}^m N_{k,11}^T \Sigma_1 N_{k,11} + \sum_{i,j=1}^v H_{i,11}^T \Sigma_1 H_{j,11} k_{ij} \leq -C_1^T C_1$$

applying Lemma A.2. Thus, by repeatedly applying the above arguments, we have

$$\left(\mathbb{E} \left\| y_{r_j} - y_{r_{j-1}} \right\|_{L_T^2}^2 \right)^{\frac{1}{2}} \leq 2\tilde{\sigma}_{r_{j-1}} \left\| u \right\|_{L_T^2} \exp \left(0.5 \left\| u \right\|_{L_T^2}^2 \right)$$

for $j = 2, \ldots, \kappa$. This concludes the proof. $\qquad\square$

The result in Theorem 3.1 is the first one of that type for deterministic/stochastic systems. In contrast to [31], we assume no bound on the control u which possibly can be small. We pay a price for dealing with general L_T^2 controls, since we obtain an exponential term in Theorem 3.1 which is due to the bilineararity. However, this result is still very meaningful, because it tells us that the ROM (15) yields a very good approximation if the truncated HSVs (diagonal entries of Σ_2) are small and, e.g., a normalized control is used. At the same time, the exponential term in the error bound can be an indicator that BT performs terribly bad if the control energy is large.

4 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 balanced truncation was extended to this setting. To do so, we proposed a reachability and an observability Gramian. We proved energy estimates with the help of these Gramian that allow us to find the unimportant states within the system. The reduced system was then obtained by removing these states from the stochastic bilinear system. Finally, we provided a new error bound that can be used to point out the cases in which the reduced order model by balanced truncation delivers a good approximation to the original model.

A Supporting Lemmas

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

Lemma A.1. *Let* $a, b_1, \ldots, b_v$ *be* $\mathbb{R}^d$*-valued processes, where* a *is 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*

$$dx(t) = a(t)dt + \sum_{i=1}^{v} b_i(t)dM_i,$$

then, we have

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

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

Lemma A.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 A.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*

$$z(t) \leq \alpha(t) + \int_0^t \beta(s)z(s)ds,$$

then it holds that

$$z(t) \leq \alpha(t) + \int_0^t \alpha(s)\beta(s)\exp\left(\int_s^t \beta(w)dw\right)ds \qquad (29)$$

for all $t \in [0, T]$. *Moreover, a non-decreasing function* α *implies that (29) becomes*

$$z(t) \leq \alpha(t)\exp\left(\int_0^t \beta(s)ds\right)$$

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

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

References

[1] E. J. Allen, S. J. Novosel, and Z. Zhang. Finite element and difference approximation of some linear stochastic partial differential equations. *Stochastics Stochastics Rep.*, 64(1-2):117–142, 1998.

[2] A. C. Antoulas. *Approximation of large-scale dynamical systems.* Advances in Design and Control 6. Philadelphia, PA: SIAM, 2005.

[3] P. Benner and T. Breiten. On $\mathcal{H}_2$-model reduction of linear parameter-varying systems. *Proc. Appl. Math. Mech.*, 11:805–806, 2011.

[4] 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.

[5] 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.

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

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

[8] T. Breiten and T. Damm. Krylov subspace methods for model order reduction of bilinear control systems. *Systems Control Lett.*, 59(10):443–450, 2010.

[9] C. Bruni, G. DiPillo, and G. Koch. On the mathematical models of bilinear systems. *Automatica*, 2(1):11–26, 1971.

[10] G. Da Prato and J. Zabczyk. *Stochastic Equations in Infinite Dimensions.* Encyclopedia of Mathematics and Its Applications. 44. Cambridge etc.: Cambridge University Press, 1992.

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

[12] 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.

[13] E. Emmrich. Discrete versions of Gronwall's lemma and their application to the numerical analysis of parabolic problems. Preprint No. 637, TU Berlin, 1999.

[14] K. V. Fernando and H. Nicholson. Singular perturbational model reduction of balanced systems. *IEEE Trans. Automat. Control*, 27:466–468, 1982.

[15] G. Flagg and S. Gugercin. Multipoint Volterra series interpolation and $\mathcal{H}_2$ optimal model reduction of bilinear systems. *SIAM. J. Matrix Anal. and Appl.*, 36(2):549–579, 2015.

[16] L. Gawarecki and V. Mandrekar. *Stochastic Differential Equations in Infinite Dimensions with Applications to Stochastic Partial Differential Equations.* Berlin: Springer, 2011.

[17] W. S. Gray and J. Mesko. Energy Functions and Algebraic Gramians for Bilinear Systems. *Proceedings of the 4th IFAC Nonlinear Control Systems Design Symposium*, 31(17):101–106, 1998.

[18] W. Grecksch and P. E. Kloeden. Time-discretised Galerkin approximations of parabolic stochastic PDEs. *Bull. Aust. Math. Soc.*, 54(1):79–85, 1996.

[19] 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.

[20] E. Hausenblas. Approximation for Semilinear Stochastic Evolution Equations. *Potential Anal.*, 18(2):141–186, 2003.

[21] A. Jentzen and P. E. Kloeden. 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, 2009.

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

[23] M. Kovács, S. Larsson, and F. Saedpanah. Finite element approximation of the linear stochastic wave equation with additive noise. *SIAM J. Numer. Anal.*, 48(2):408–427, 2010.

[24] R. Kruse. *Strong and Weak Approximation of Semilinear Stochastic Evolution Equations*, volume 2093 of *Lecture Notes in Mathematics*. Springer, 2014.

[25] Y. Liu and B. D. Anderson. Singular perturbation approximation of balanced systems. *Int. J. Control*, 50(4):1379–1405, 1989.

[26] R. R. Mohler. *Bilinear Control Processes.* Academic Press, New York, 1973.

[27] B. C. Moore. Principal component analysis in linear systems: Controllability, observability, and model reduction. *IEEE Trans. Autom. Control*, 26:17–32, 1981.

[28] G. Obinata and B. D. Anderson. *Model Reduction for Control System Design.* Communications and Control Engineering Series. London: Springer, 2001.

[29] S. Peszat and J. Zabczyk. *Stochastic Partial Differential Equations with Lévy Noise. An evolution equation approach.* Encyclopedia of Mathematics and Its Applications 113. Cambridge: Cambridge University Press, 2007.

[30] C. Prévôt and M. Röckner. *A Concise Course on Stochastic Partial Differential Equations.* Lecture Notes in Mathematics 1905. Berlin: Springer, 2007.

[31] M. Redmann. Type II balanced truncation for deterministic bilinear control systems. arXiv preprint 1709.05655, 2017.

[32] M. Redmann. Type II singular perturbation approximation for linear systems with Lévy noise. *SIAM J. Control Optim.*, accepted 2018.

[33] M. Redmann and P. Benner. Singular Perturbation Approximation for Linear Systems with Lévy Noise. *Stochastics and Dynamics*, 2017.

[34] W. J. Rugh. *Nonlinear System Theory*. The Johns Hopkins University Press, Baltimore, MD, 1981.