

University of Groningen

## From Data to Reduced-order Models via Generalized Balanced Truncation

Burohman, Azka; Besselink, Bart; Scherpen, Jacquélien M. A.; Camlibel, Kanat

*Published in:*  
IEEE Transactions on Automatic Control

*DOI:*  
[10.1109/TAC.2023.3238856](https://doi.org/10.1109/TAC.2023.3238856)

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

*Document Version*  
Publisher's PDF, also known as Version of record

*Publication date:*  
2023

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*  
Burohman, A., Besselink, B., Scherpen, J. M. A., & Camlibel, K. (2023). From Data to Reduced-order Models via Generalized Balanced Truncation. *IEEE Transactions on Automatic Control*, 68(10), 6160-6175. <https://doi.org/10.1109/TAC.2023.3238856>

### Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).





The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

### Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

*Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.*

# From Data to Reduced-Order Models via Generalized Balanced Truncation

Azka Muji Burohman , Bart Besselink , *Member, IEEE*, Jacquélien M. A. Scherpen , *Fellow, IEEE*, and M. Kanat Camlibel , *Senior Member, IEEE*

## I. INTRODUCTION

**Abstract**—This article proposes a data-driven model reduction approach on the basis of noisy data with a known noise model. First, the concept of data reduction is introduced. In particular, we show that the set of reduced-order models obtained by applying a Petrov–Galerkin projection to all systems explaining the data characterized in a large-dimensional quadratic matrix inequality (QMI) can again be characterized in a lower-dimensional QMI. Next, we develop a data-driven generalized balanced truncation method that relies on two steps. First, we provide necessary and sufficient conditions such that systems explaining the data have common generalized Gramians. Second, these common generalized Gramians are used to construct matrices that allow to characterize a class of reduced-order models via generalized balanced truncation in terms of a lower-dimensional QMI by applying the data reduction concept. Additionally, we present alternative procedures to compute a priori and a posteriori upper bounds with respect to the true system generating the data. Finally, the proposed techniques are illustrated by means of application to an example of a system of a cart with a double-pendulum.

**Index Terms**—Data-driven model reduction, data informativity, error bounds, generalized balancing.

Manuscript received 23 September 2021; revised 18 November 2022; accepted 5 January 2023. Date of publication 23 January 2023; date of current version 27 September 2023. This work was based on research developed in the DSSC Doctoral Training Programme and supported by the Marie Skłodowska-Curie COFUND under Grant DSSC 754315. Recommended by Associate Editor D. Regruto. (*Corresponding author: Azka Muji Burohman.*)

Azka Muji Burohman is with the Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence, University of Groningen, Nijenborgh 9 9747 AG, Groningen, The Netherlands, also with the Engineering and Technology Institute Groningen (ENTEG), University of Groningen, Nijenborgh 4 9747 AG, The Netherlands, and also with the Jan C. Willems Center for Systems and Control, Faculty of Science and Engineering, University of Groningen, The Netherlands (e-mail: a.m.burohman@rug.nl).

Bart Besselink and M. Kanat Camlibel are with the Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence, University of Groningen, Nijenborgh 9 9747 AG, Groningen, The Netherlands, and also with the Jan C. Willems Center for Systems and Control, Faculty of Science and Engineering, University of Groningen, The Netherlands (e-mail: b.besselink@rug.nl; m.k.camlibel@rug.nl).

Jacquélien M. A. Scherpen is with the Engineering and Technology Institute Groningen (ENTEG), University of Groningen, Nijenborgh 4 9747 AG, Groningen, The Netherlands, and also with the Jan C. Willems Center for Systems and Control, Faculty of Science and Engineering, University of Groningen, The Netherlands (e-mail: j.m.a.scherpen@rug.nl).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TAC.2023.3238856>.

Digital Object Identifier 10.1109/TAC.2023.3238856

MODEL reduction refers to the problem of constructing low-dimensional system models that accurately approximate complex high-dimensional systems. Traditionally, model reduction techniques solve this problem by deriving low-dimensional models on the basis of the given high-dimensional model through suitable operations such as projection. In the field of systems and control, roughly two classes of model reduction techniques can be distinguished for linear systems: 1) methods based on energy functions such as balanced truncation [1], [2], [3], [4] and optimal Hankel norm approximation [5]; and 2) methods based on interpolation and/or moment matching [6], [7], [8], [9], sometimes also referred to as Krylov methods. Extensions to nonlinear systems have emerged in the form of nonlinear balancing methods [10], [11] and nonlinear moment matching techniques [9], [12]. We refer the reader to [13], [14], and [15] for details on a variety of existing model reduction methods.

Recently, the problem of *data-driven* model reduction is attracting increasing attention, partly motivated by the widespread availability of measurement data. Here, low-order models are constructed directly on the basis of measurement data, thus not requiring the availability of a high-order model. We emphasize that these data-driven model reduction approaches differ from traditional approaches in which, first, a high-order model is derived using system identification techniques and, second, existing model-based techniques for model reduction are used. Nevertheless, standard (model-based) model reduction techniques have inspired various data-driven techniques.

First, in the class of energy-based methods for linear systems, to which this article belongs, the authors in [16] and [17] proposed a data-driven balanced truncation method from *persistently exciting* data and [18] estimated Gramians from frequency and time-domain data based on their quadrature form. For nonlinear systems, empirical balanced truncation was presented in [19] and [20], whereas data-driven reduction for monotone nonlinear systems was considered in [21]. Second, in the class of interpolatory methods, we begin by mentioning contributions to data-driven reduction methods on the basis of frequency-domain data by the Loewner framework [22]. In this method, noise-free frequency-domain data are formulated to construct Loewner matrix pencils to enable the construction of state-space models. Extensions of this approach aim at constructing reduced-order models preserving stability [23] and achieving

$\mathcal{H}_2$ -optimality [24]. In addition, the use of Loewner methods based on time-domain data and noisy frequency-domain data was pursued in [25] and [26], respectively. Besides the Loewner framework, data-driven moment matching techniques were presented in [9] and [27], where the latter exploits the so-called data informativity framework.

Despite these developments, existing methods for data-driven model reduction do not often allow for guaranteeing system properties such as asymptotic stability and do not provide an error bound, especially when the available data is subject to noise. In this article, we develop a data-driven reduction technique that provides such guarantees on the low-order model, even for noisy data. Specifically, this article has the following contributions.

First, we introduce the concept of *data reduction* via a Petrov–Galerkin projection. Following the data informativity framework of [28], we characterize the class of systems that are consistent with the measurement data for a given noise model in terms of a quadratic matrix inequality (QMI). Then, we define the class of reduced-order systems as the set of systems obtained by applying the Petrov–Galerkin projection to all systems explaining the noisy data. Importantly, we show that this class of reduced-order systems can again be characterized in terms of a quadratic matrix inequality, but one of *lower dimension*. As the relevant matrix variables in this QMI depend only on the measurement data, noise model and matrices corresponding to a projection, this can be regarded as *data reduction*.

The second contribution of this article is the development of a data-driven generalized balanced truncation method. Here, we characterize the set of all reduced-order models obtained by applying generalized balanced truncation to the class of systems explaining the noisy data. This relies on the following two steps. As the first step, we give necessary and sufficient conditions for all systems explaining the data to have a *common* generalized controllability and *common* generalized observability Gramian. In this case, we say that the data are *informative* for generalized Lyapunov balancing. These conditions heavily rely on the so-called matrix S-lemma from [29] and again build on the data informativity framework of [28]. We note that this framework has also been successfully applied in solving various control problems, e.g., data-driven  $\mathcal{H}_2$  and  $\mathcal{H}_\infty$  control [29]. The second step comprises the use of the common generalized Gramians to obtain the Petrov–Galerkin projection that achieves (generalized) balanced truncation (see [30] for details on generalized balanced truncation for a single given system). This allows for the application of the data reduction concept and yields the desired class of reduced-order models in terms of a low-dimensional quadratic matrix inequality.

This data-driven model reduction procedure has various desirable properties by virtue of the inherent advantages of using a balancing-type reduction method. Namely, all reduced-order models are guaranteed to be asymptotically stable and satisfy an a priori error bound. However, the ordinary a priori upper bound from model-based reduction methods, e.g., [13], [30], does not determine the error between a selected reduced-order model (from the class of reduced-order models) to the true system generating the data because the true system is unknown. Therefore, as the final contribution of this article, we provide

two alternative error bounds. First, we compute a uniform a priori upper bound, i.e., an error bound that holds for any chosen high-order system explaining the data and any reduced-order model. The computation of this error bound again exploits the QMI characterization of the class of (reduced-order) systems, together with the bounded real lemma. Second, we also present an a posteriori error bound that is uniform over all systems explaining the data for a *selected* reduced-order system.

The rest of this article is organized as follows. In Section II, we provide appropriate background material on Petrov–Galerkin model reduction and generalized balanced truncation. Section III deals with the data reduction problem through the Petrov–Galerkin projection. The problem formulation of informativity for Lyapunov balancing is given in Section IV, followed by the main results containing necessary and sufficient conditions for generalized Lyapunov balancing, a characterization of the set of reduced-order models and error bounds in Sections IV-A, IV-B, and IV-C, respectively. In Section V, an illustrative example is provided to show how the set of reduced-order models is extracted from data. Finally, Section VI concludes the article. For the sake of completeness, some important results and proofs are presented in the Appendix.

*Notation:* We denote  $M > 0$  ( $M \geq 0$ ) and  $M < 0$  ( $M \leq 0$ ) for positive and negative (semi) definite symmetric matrices, respectively. We denote the number of negative, zero, and positive eigenvalues of a symmetric matrix  $M$  by  $v_-(M)$ ,  $v_0(M)$ , and  $v_+(M)$ , respectively. The *inertia* of  $M$  is denoted by  $\text{In}(M) = (v_-(M), v_0(M), v_+(M))$ . For a symmetric matrix  $M$  partitioned as

$$M = \begin{bmatrix} M_{11} & M_{12} \\ M_{12}^\top & M_{22} \end{bmatrix}$$

its Schur complement with respect to  $M_{22}$  is denoted by  $M|M_{22}$ , i.e.,  $M|M_{22} := M_{11} - M_{12}M_{22}^{-1}M_{12}^\top$ . For a square matrix  $A$ , its spectral radius is denoted by  $\rho(A)$  and the sum of its main diagonal elements is denoted by  $\text{trace}(A)$ . We denote  $\text{blkdiag}(A_1, A_2, \dots, A_n)$  for a block diagonal matrix of the form

$$\begin{bmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_n \end{bmatrix}.$$

The matrix  $I_j$  denotes the identity matrix of size  $j$ . For a discrete-time linear system  $\Sigma$ , its  $\mathcal{H}_\infty$ -norm is denoted by  $\|\Sigma\|_{\mathcal{H}_\infty}$ . For a system  $\Sigma$  having realization  $(A, B, C, D)$  and transfer function  $G(z) = C(zI - A)^{-1}B + D$ , the norm  $\|\Sigma\|_{\mathcal{H}_\infty}$  is defined by  $\|\Sigma\|_{\mathcal{H}_\infty} = \sup_{\omega \in \mathbb{R}} \|G(e^{i\omega})\|$ .

## II. PRELIMINARIES

### A. Model Reduction via a Petrov–Galerkin Projection

Consider the discrete-time input/state/output system

$$\Sigma : \begin{cases} \mathbf{x}(k+1) = A\mathbf{x}(k) + B\mathbf{u}(k), \\ \mathbf{y}(k) = C\mathbf{x}(k) + D\mathbf{u}(k) \end{cases} \quad (1)$$

with input  $\mathbf{u} \in \mathbb{R}^m$ , state  $\mathbf{x} \in \mathbb{R}^n$  and output  $\mathbf{y} \in \mathbb{R}^p$ . Let  $\widehat{W}, \widehat{V} \in \mathbb{R}^{n \times r}$  be matrices such that  $\widehat{W}^\top \widehat{V} = I$  and  $r < n$ . A reduced-order model is obtained by Petrov–Galerkin projection. Here, approximating  $\mathbf{x}(t)$  by  $\widehat{V}\widehat{\mathbf{x}}(t)$  leads to

$$\widehat{V}\widehat{\mathbf{x}}(t+1) = A\widehat{V}\widehat{\mathbf{x}}(t) + B\mathbf{u}(t) + r(t)$$

where  $r(t)$  is a residual that results from this approximation. Next, requiring that  $\widehat{W}^\top r(t) = 0$  leads to the reduced-order model

$$\widehat{\Sigma} : \begin{cases} \widehat{\mathbf{x}}(k+1) = \widehat{W}^\top A\widehat{V}\widehat{\mathbf{x}}(k) + \widehat{W}^\top B\mathbf{u}(k), \\ \widehat{\mathbf{y}}(k) = C\widehat{V}\widehat{\mathbf{x}}(k) + D\mathbf{u}(k) \end{cases} \quad (2)$$

with state  $\widehat{\mathbf{x}} \in \mathbb{R}^r$ . The Petrov–Galerkin projection method provides a general framework for model order reduction. Namely, many reduction techniques, including Gramian- and Krylov-based methods can be regarded as Petrov–Galerkin projections with appropriate choice of  $\widehat{V}$  and  $\widehat{W}$ , see, e.g., [13].

### B. Generalized Lyapunov Balancing

Lyapunov balancing is a popular method for model reduction, which is slightly generalized in the concept of generalized Lyapunov balancing (GLB). We begin our discussion on GLB by introducing the notion of a generalized Gramian. Consider the discrete-time system (1). Then, a matrix  $P = P^\top > 0$  satisfying

$$APA^\top - P + BB^\top < 0$$

and  $Q = Q^\top > 0$  satisfying

$$A^\top QA - Q + C^\top C < 0$$

are called the *generalized controllability Gramian* and *generalized observability Gramian*, respectively. This is a strict version of the definition of generalized Gramians given in [30]. The generalized Gramians are lower bounded by the ordinary Gramians, i.e.,  $P > P_0$  and  $Q > Q_0$ , where  $Q_0$  and  $P_0$  are the solutions of the corresponding Lyapunov equations. Note that asymptotic stability of  $\Sigma$ , i.e.,  $\rho(A) < 1$ , where  $\rho(A)$  represents the spectral radius of  $A$ , is a necessary condition for the generalized Gramians to exist.

These generalized Gramians can be used to obtain a so-called balanced realization, i.e., a realization of (1) for which the corresponding generalized Gramians are equal and diagonal. Specifically, by [13, Lemma 7.3], there exists a nonsingular matrix  $T$ , such that  $TPT^\top = T^{-\top}QT^{-1} = \Sigma_H$ , where  $\Sigma_H$  is a diagonal matrix of the *generalized* Hankel singular values of  $\Sigma$  in (1), i.e.,

$$\Sigma_H := \text{blkdiag}(\sigma_1 I_{m_1}, \sigma_2 I_{m_2}, \dots, \sigma_\kappa I_{m_\kappa}) \quad (3)$$

with  $\sigma_1 > \sigma_2 > \dots > \sigma_\kappa > 0$ , and where  $m_i$  denotes the multiplicity of  $\sigma_i$  for  $i = 1, \dots, \kappa$  satisfying  $n = \sum_{i=1}^{\kappa} m_i$ . In this case, we say that such a realization is *balanced in the sense of generalized Lyapunov balancing*. In particular, the balanced realization is given by

$$A_{\text{bal}} := TAT^{-1}, B_{\text{bal}} := TB, C_{\text{bal}} := CT^{-1}, D_{\text{bal}} := D. \quad (4)$$

Finally, the reduced-order model via balanced truncation is obtained by truncating the balanced system  $(A_{\text{bal}}, B_{\text{bal}}, C_{\text{bal}}, D_{\text{bal}})$ .

Indeed, balanced-truncation model reduction is essentially a Petrov–Galerkin projection. Namely, after introducing the matrix  $\Pi \in \mathbb{R}^{n \times r}$  as

$$\Pi := \begin{bmatrix} I_r \\ 0 \end{bmatrix}$$

the matrices  $\widehat{V} = T^{-1}\Pi$  and  $\widehat{W} = T^\top\Pi$  satisfy  $\widehat{W}^\top\widehat{V} = I$  and the reduced-order model (2) is equal to the model obtained by balanced truncation (to order  $r$ ).

Generalized balanced truncation guarantees the preservation of some relevant system properties, similarly as in ordinary balanced truncation, see [30, Prop. 4.19] for continuous-time systems. Most importantly, asymptotic stability is preserved. The discrete-time version is presented without proof below.

*Proposition 1:* Consider the system  $\Sigma$  given in (1). Let  $\widehat{\Sigma}$  of the form (2) be a reduced-order system of  $\Sigma$  via generalized balanced truncation. Suppose that  $\widehat{\Sigma}$  is of order  $r < n$  where  $r = \sum_{i=1}^{\ell} m_i$  with  $\ell < \kappa$ . Then, it is balanced in the sense of GLB with  $\rho(\widehat{W}^\top A\widehat{V}) < 1$  and

$$\|\Sigma - \widehat{\Sigma}\|_{\mathcal{H}_\infty} \leq 2 \sum_{i=\ell+1}^{\kappa} \sigma_i$$

where the  $\sigma_i$ 's are the neglected generalized Hankel singular values given in (3).

### III. DATA-DRIVEN PETROV–GALERKIN PROJECTION

Consider the linear discrete-time input/state/output system

$$\Sigma_{\text{true}} : \begin{cases} \mathbf{x}(k+1) = A_{\text{true}}\mathbf{x}(k) + B_{\text{true}}\mathbf{u}(k) + \mathbf{w}(k), \\ \mathbf{y}(k) = C_{\text{true}}\mathbf{x}(k) + D_{\text{true}}\mathbf{u}(k) + \mathbf{z}(k) \end{cases} \quad (5)$$

where  $(\mathbf{u}, \mathbf{x}, \mathbf{y}) \in \mathbb{R}^{m+n+p}$  are the input/state/output and  $(\mathbf{w}, \mathbf{z}) \in \mathbb{R}^{n+p}$  are noise terms. Throughout the article, we assume that the system matrices  $(A_{\text{true}}, B_{\text{true}}, C_{\text{true}}, D_{\text{true}})$  and the noise  $(\mathbf{w}, \mathbf{z})$  are *unknown*. What is known instead are a finite number of input/state/output measurements harvested from the true system (5)

$$\begin{aligned} &u(0), u(1), \dots, u(L-1), \\ &x(0), x(1), \dots, x(L), \\ &y(0), y(1), \dots, y(L-1). \end{aligned}$$

We collect these data in the matrices

$$\begin{aligned} X &:= \begin{bmatrix} x(0) & x(1) & \dots & x(L) \end{bmatrix}, \\ X_- &:= \begin{bmatrix} x(0) & x(1) & \dots & x(L-1) \end{bmatrix}, \\ X_+ &:= \begin{bmatrix} x(1) & x(2) & \dots & x(L) \end{bmatrix}, \\ U_- &:= \begin{bmatrix} u(0) & u(1) & \dots & u(L-1) \end{bmatrix}, \\ Y_- &:= \begin{bmatrix} y(0) & y(1) & \dots & y(L-1) \end{bmatrix}. \end{aligned}$$

Now, we can define the set of all systems that *explain* the data as

$$\Sigma := \left\{ (A, B, C, D) : \begin{bmatrix} X_+ \\ Y_- \end{bmatrix} - \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} X_- \\ U_- \end{bmatrix} \in \mathcal{N} \right\}$$

where  $\mathcal{N} \subseteq \mathbb{R}^{(n+p) \times L}$  captures a *noise model* such that

$$(A_{\text{true}}, B_{\text{true}}, C_{\text{true}}, D_{\text{true}}) \in \Sigma. \quad (6)$$

In this article, we work with a noise model that is described by a quadratic matrix inequality as

$$\mathcal{N} := \left\{ Z \in \mathbb{R}^{(n+p) \times L} : \begin{bmatrix} I \\ Z^\top \end{bmatrix}^\top \begin{bmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{12}^\top & \Phi_{22} \end{bmatrix} \begin{bmatrix} I \\ Z^\top \end{bmatrix} \geq 0 \right\} \quad (7)$$

where  $\Phi_{11} = \Phi_{11}^\top \in \mathbb{R}^{(n+p) \times (n+p)}$ ,  $\Phi_{12} \in \mathbb{R}^{(n+p) \times L}$ , and  $\Phi_{22} = \Phi_{22}^\top \in \mathbb{R}^{L \times L}$ .

Throughout the article, we make the following blanket assumption on the set  $\mathcal{N}$ .

*Assumption 1:* The set  $\mathcal{N}$  is bounded and has nonempty interior.

As shown in [31], one can verify this assumption by using the following lemma.

*Lemma 1:* The set  $\mathcal{N}$  given by (7) is bounded and has nonempty interior, if and only if  $\Phi_{22} < 0$  and  $\Phi_{11} - \Phi_{12}\Phi_{22}^{-1}\Phi_{12}^\top > 0$ .

The noise model (7) satisfying Assumption 1 includes various relevant examples, see, e.g., [32]. Two of those, with  $Z_- = [z(0) \ z(1) \ \dots \ z(L-1)]$ , are as follows.

- 1)  $\Phi_{22} = -I$  and  $\Phi_{12} = 0$  give the energy bound  $\sum_{t=0}^{L-1} z(t)z(t)^\top < \Phi_{11}$ .
- 2) if we have the individual noise sample bounds  $\|z(t)\|_2^2 \leq \epsilon$  for all  $t$ , then this satisfies the noise model  $\Phi_{22} = -I$ ,  $\Phi_{12} = 0$ , and  $\Phi_{11} = \epsilon TI$ .

We stress that we do not take a probabilistic perspective, instead we work with noise satisfying the deterministic bound (7).

It is clear from the definition of  $\Sigma$  and (7) that  $(A, B, C, D) \in \Sigma$ , if and only if the following QMI is satisfied

$$\begin{bmatrix} I & 0 \\ 0 & I \\ A^\top & C^\top \\ B^\top & D^\top \end{bmatrix}^\top N \begin{bmatrix} I & 0 \\ 0 & I \\ A^\top & C^\top \\ B^\top & D^\top \end{bmatrix} \geq 0 \quad (8)$$

where

$$N := \begin{bmatrix} I & 0 & X_+ \\ 0 & I & Y_- \\ 0 & 0 & -X_- \\ 0 & 0 & -U_- \end{bmatrix} \begin{bmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{12}^\top & \Phi_{22} \end{bmatrix} \begin{bmatrix} I & 0 & X_+ \\ 0 & I & Y_- \\ 0 & 0 & -X_- \\ 0 & 0 & -U_- \end{bmatrix}^\top. \quad (9)$$

In characterizing the set of systems that explain the data, one may wonder whether the set  $\Sigma$  is bounded and has nonempty

interior. The following proposition provides the required condition, which solely relies on the data.

*Proposition 2:* The set  $\Sigma$  is bounded and has nonempty interior, if and only if there exists  $\bar{S} \in \mathbb{R}^{(n+m) \times (n+p)}$  such that

$$\begin{bmatrix} I \\ \bar{S} \end{bmatrix}^\top N \begin{bmatrix} I \\ \bar{S} \end{bmatrix} > 0 \quad (10)$$

and  $\begin{bmatrix} X_- \\ U_- \end{bmatrix}$  has full row rank.

*Proof:* The proof is presented in Appendix VI.  $\square$

The condition (10) is referred as the *generalized Slater condition*.

In this article, we are not interested in the true system (5) per se. Instead, we would like to find a *reduced-order approximation* of (5) directly on the basis of the available data.

As a first step, we consider a Petrov–Galerkin projection as in Section II and assume that the matrices  $\widehat{W}$  and  $\widehat{V}$  satisfying  $\widehat{W}^\top \widehat{V} = I$  are given. Then, the set of reduced-order models of all systems explaining the data is defined as

$$\Sigma_{\widehat{V}, \widehat{W}} := \left\{ (\widehat{W}^\top A \widehat{V}, \widehat{W}^\top B, C \widehat{V}, D) : (A, B, C, D) \in \Sigma \right\}.$$

The first main result of this article is that the set  $\Sigma_{\widehat{V}, \widehat{W}}$  can itself be represented as a QMI of a similar form as (8). This is formalized in the following theorem, whose proof can be found in Appendix VI-A.

*Theorem 1:* Consider the set  $\Sigma$  of systems explaining the data. Suppose that there exists  $\bar{S}$  such that (10) holds and the matrix  $\begin{bmatrix} X_- \\ U_- \end{bmatrix}$  has full row rank. Let  $\widehat{W}, \widehat{V} \in \mathbb{R}^{n \times r}$  be such that  $\widehat{W}^\top \widehat{V} = I$  and  $r < n$ . Then, the set  $\Sigma_{\widehat{V}, \widehat{W}}$  of reduced-order models of  $\Sigma$  using matrices  $\widehat{W}, \widehat{V}$  satisfies

$$\Sigma_{\widehat{V}, \widehat{W}} = \left\{ (\widehat{A}, \widehat{B}, \widehat{C}, \widehat{D}) : \begin{bmatrix} I & 0 \\ 0 & I \\ \widehat{A}^\top & \widehat{C}^\top \\ \widehat{B}^\top & \widehat{D}^\top \end{bmatrix}^\top N_{V,W} \begin{bmatrix} I & 0 \\ 0 & I \\ \widehat{A}^\top & \widehat{C}^\top \\ \widehat{B}^\top & \widehat{D}^\top \end{bmatrix} \geq 0 \right\}$$

where  $N_{V,W}$  is given by (11) shown at the bottom of the next page:

$$W := \begin{bmatrix} \widehat{W} \\ I_p \end{bmatrix} \text{ and } V := \begin{bmatrix} \widehat{V} \\ I_m \end{bmatrix}.$$

Theorem 1 has a nice interpretation in terms of *data reduction*. Namely, the matrix  $N_{V,W}$  characterizing all reduced-order models depends only on the matrices  $\widehat{V}, \widehat{W}$  and the original data matrix  $N$ . As such,  $N_{V,W}$  is constructed from the data and noise model only. Importantly,  $N_{V,W}$  has a lower dimension than  $N$  and can thus be regarded as a reduced data matrix. Hence, we can characterize all reduced-order models by directly reducing the data matrix  $N$  rather than reducing individual systems  $(A, B, C, D) \in \Sigma$ . It is also worth mentioning that

$$(\widehat{W}^\top A_{\text{true}} \widehat{V}, \widehat{W}^\top B_{\text{true}}, C_{\text{true}} \widehat{V}, D_{\text{true}})$$

i.e., the reduced-order model of the true system, is in  $\Sigma_{\widehat{V}, \widehat{W}}$ .

In this section, we have characterized reduced-order approximations of systems explaining the collected data for *given* matrices  $\widehat{W}$  and  $\widehat{V}$ . We emphasize that up to this point, there are no assumptions on these matrices such that Theorem 1 can be used for arbitrary Petrov–Galerkin projections. In the next section, we will construct the Petrov–Galerkin projection on the basis of the available data by following a generalized balancing framework.

#### IV. DATA-DRIVEN GENERALIZED BALANCED TRUNCATION

In this section, we will introduce the notion of informativity for GLB. Moreover, we give necessary and sufficient conditions for informativity for GLB, followed by the set of reduced-order models obtained from data-driven GLB and their error-bounds.

##### A. Data Informativity for GLB

Based on Section II-B, one can introduce the notion of informativity for GLB as follows.

*Definition 1:* We say that the data  $(U_-, X, Y_-)$  are *informative for generalized Lyapunov balancing (GLB)* if there exist  $P = P^\top > 0$  and  $Q = Q^\top > 0$  such that

$$APA^\top - P + BB^\top < 0 \quad (12)$$

and

$$A^\top QA - Q + C^\top C < 0 \quad (13)$$

for all  $(A, B, C, D) \in \Sigma$ .

*Remark 1:* It is known, see, e.g., [13, Ch. 7], that the satisfaction of (12) or (13) implies that all systems in  $\Sigma$  are asymptotically stable.

From Definition 1,  $P$  and  $Q$  can be regarded as *common* generalized controllability and observability Gramian, respectively, for all systems explaining the data. We thus formalize the following informativity and model reduction problems.

*Problem 1:* Find necessary and sufficient conditions under which the data  $(U_-, X, Y_-)$  are informative for GLB. If the data are informative for GLB, then characterize the reduced-order models via data-driven balanced truncation and provide error bounds with respect to the true system.

Observe that the data are informative for GLB, if and only if QMI (8) implies the existence of positive definite matrices  $P$  and  $Q$  such that (12) and (13) hold. Such QMI implications can be viewed as a generalization of the classical S-lemma [33] and have been investigated in [29]. Based on the results of [29] and [31] (see Appendix VI-A), data informativity for GLB can

be fully characterized in terms of feasibility of certain LMIs as stated next.

*Theorem 2:* Suppose that there exists  $\bar{S}$  such that (10) holds. Define

$$N_C := \begin{bmatrix} I_n & 0 \\ 0 & 0 \\ 0 & I_{n+m} \end{bmatrix}^\top N \begin{bmatrix} I_n & 0 \\ 0 & 0 \\ 0 & I_{n+m} \end{bmatrix}$$

and

$$N_O := \begin{bmatrix} I_n & 0 \\ 0 & 0 \\ 0 & I_{n+p} \end{bmatrix}^\top N^\# \begin{bmatrix} I_n & 0 \\ 0 & 0 \\ 0 & I_{n+p} \end{bmatrix}$$

where

$$N^\# := \begin{bmatrix} 0 & -I_{n+m} \\ I_{n+p} & 0 \end{bmatrix} N^{-1} \begin{bmatrix} 0 & -I_{n+p} \\ I_{n+m} & 0 \end{bmatrix}.$$

Then, the data  $(U_-, X, Y_-)$  are informative for generalized Lyapunov balancing, if and only if

- i)  $\begin{bmatrix} X_- \\ U_- \end{bmatrix}$  has full row rank,
- ii) there exists  $P = P^\top > 0$  and a scalar  $\alpha > 0$  such that

$$\begin{bmatrix} P & 0 & 0 \\ 0 & -P & 0 \\ 0 & 0 & -I_m \end{bmatrix} - \alpha N_C > 0, \quad (14)$$

- iii) there exists  $Q = Q^\top > 0$  and a scalar  $\beta > 0$  such that

$$\begin{bmatrix} Q & 0 & 0 \\ 0 & -Q & 0 \\ 0 & 0 & -I_p \end{bmatrix} - \beta N_O > 0. \quad (15)$$

*Proof:* The proof can be found in Appendix VI-B.  $\square$

A direct consequence of data informativity for GLB is that all systems explaining the data have *common* generalized Gramians  $P$  and  $Q$ . As a result, all systems in  $\Sigma$  are balanced by a common balancing transformation matrix  $T$  satisfying  $TPPT^\top = T^{-\top}QT^{-1} = \Sigma_H$  where  $\Sigma_H$  is a matrix of the form (3), containing the common generalized Hankel singular values. Next, we note that the balanced realizations of all systems explaining the data can be constructed from (4) where  $(A, B, C, D) \in \Sigma$ .

*Remark 2:* Since  $P$  and  $Q$  satisfying (14) and (15), respectively, are lower bounded by the ordinary Gramians of the true system, (see the discussion in Section II-B), smaller  $P$  and  $Q$  are expected to yield a balancing that is “closer” to the ordinary balancing of the true system. Therefore, one may solve LMIs (14) and (15) by minimizing  $\text{trace}(P)$  and  $\text{trace}(Q)$  to expect a better reduced-order approximation.

In the next section, we will use these balanced realizations to obtain reduced-order models directly from data.

$$N_{V,W} := \begin{bmatrix} W^\top (N|N_{22} + N_{12}N_{22}^{-1}V(V^\top N_{22}^{-1}V)^{-1}V^\top N_{22}^{-1}N_{12}^\top)W & W^\top N_{12}N_{22}^{-1}V(V^\top N_{22}^{-1}V)^{-1} \\ \hline (V^\top N_{22}^{-1}V)^{-1}V^\top N_{22}^{-1}N_{12}^\top W & (V^\top N_{22}^{-1}V)^{-1} \end{bmatrix} \quad (11)$$

## B. Reduced-Order Models

By applying the Petrov–Galerkin projection, the reduced-order models of all systems in  $\Sigma$  via generalized balanced truncation are contained in the set

$$\widehat{\Sigma} := \left\{ (\widehat{W}^\top A \widehat{V}, \widehat{W}^\top B, C \widehat{V}, D) : (A, B, C, D) \in \Sigma \right\}$$

where  $\widehat{V} = T^{-1}\Pi$  and  $\widehat{W} = T^\top\Pi$  with  $\Pi$  given by

$$\Pi := \begin{bmatrix} I_r \\ 0 \end{bmatrix}$$

and  $T$  is obtained from the common generalized Gramians  $P$  and  $Q$  for all systems in  $\Sigma$ .

Based on Theorem 1, we can characterize the set  $\widehat{\Sigma}$  in terms of a quadratic matrix inequality. We formalize this fact in the following corollary of Theorem 1.

*Corollary 1:* Suppose that there exists  $\bar{S}$  such that (10) holds and the data  $(U_-, X, Y_-)$  are informative for generalized Lyapunov balancing with  $T$  the corresponding balancing transformation. Then

$$\widehat{\Sigma} = \left\{ (\widehat{A}, \widehat{B}, \widehat{C}, \widehat{D}) : \begin{bmatrix} I & 0 \\ 0 & I \\ \widehat{A}^\top & \widehat{C}^\top \\ \widehat{B}^\top & \widehat{D}^\top \end{bmatrix}^\top N_{V,W} \begin{bmatrix} I & 0 \\ 0 & I \\ \widehat{A}^\top & \widehat{C}^\top \\ \widehat{B}^\top & \widehat{D}^\top \end{bmatrix} \geq 0 \right\}$$

where  $N_{V,W}$  is given by (11) with

$$W = \begin{bmatrix} T^\top \Pi & \\ & I_p \end{bmatrix}, \quad V = \begin{bmatrix} T^{-1} \Pi & \\ & I_m \end{bmatrix} \quad \text{and} \quad \Pi := \begin{bmatrix} I_r \\ 0 \end{bmatrix}.$$

Now, we can see that the set  $\widehat{\Sigma}$  characterizes a *data reduction* by generalized balanced truncation. Namely, the reduced matrix  $N_{V,W}$  depends only on the data matrix  $N$  and matrices  $\widehat{V} = T^{-1}\Pi$  and  $\widehat{W} = T^\top\Pi$ , where now these matrices are also derived from the data only via Theorem 2.

From the definition of  $\widehat{\Sigma}$  above, suppose that  $(\widehat{A}, \widehat{B}, \widehat{C}, \widehat{D}) \in \widehat{\Sigma}$ , then it is always a truncation of a model in  $\Sigma$  by generalized balanced truncation. Therefore, any  $(\widehat{A}, \widehat{B}, \widehat{C}, \widehat{D}) \in \widehat{\Sigma}$  satisfies the guaranteed properties in Proposition 1. Namely,  $\rho(\widehat{A}) < 1$ , i.e., the reduced-order system is asymptotically stable. Moreover, the  $\mathcal{H}_\infty$ -norm error between a system in  $\widehat{\Sigma}$  and its *corresponding* system in  $\Sigma$  is upper bounded by the neglected common generalized Hankel singular values, i.e.,

$$\|\Sigma - \widehat{\Sigma}\|_{\mathcal{H}_\infty} \leq 2 \sum_{i=\ell+1}^{\kappa} \sigma_i \quad (17)$$

where  $\Sigma$  and  $\widehat{\Sigma}$  are systems whose realizations are in  $\Sigma$  and  $\widehat{\Sigma}$ , respectively, and  $\sigma_i$  is defined similarly as in Proposition 1.

However, the bound (17) has little practical relevance as it characterizes the error between *one* reduced-order system in  $\widehat{\Sigma}$

and its *corresponding* high-order system in  $\Sigma$ . Instead, recall that we are interested in a reduced-order approximation of the true system (5). This system is unknown, but is guaranteed to satisfy (6). As also the corresponding reduced-order system is unknown (but in  $\widehat{\Sigma}$ ), a practical relevant error bound should hold for *any* selection of a high-order system (from  $\Sigma$ ) and *any* reduced-order system (from  $\widehat{\Sigma}$ ). The following section provides such bounds.

## C. Distance to the True System

Suppose that we consider a reduced-order system of order  $r < n$  given by  $\widehat{\Sigma}_0$  with realization  $(\widehat{A}_0, \widehat{B}_0, \widehat{C}_0, \widehat{D}_0) \in \widehat{\Sigma}$ . We will use  $\widehat{\Sigma}_0$  as an approximation of the unknown true system  $\Sigma_{\text{true}}$ . To evaluate the quality of this approximation, note that

$$\|\widehat{\Sigma}_0 - \Sigma_{\text{true}}\|_{\mathcal{H}_\infty} \leq \sup \left\{ \|\widehat{\Sigma} - \Sigma\|_{\mathcal{H}_\infty} : \Sigma \in \Sigma, \widehat{\Sigma} \in \widehat{\Sigma} \right\}. \quad (18)$$

Here, we have used the small abuse of notation  $\Sigma \in \Sigma$  to mean  $(A, B, C, D) \in \Sigma$ , where  $(A, B, C, D)$  is a realization of  $\Sigma$ .

In this section, we aim to compute a bound on the right-hand side of (18) on the basis of the available data only. The computation of this is stated in the following result.

*Theorem 3:* The bound

$$\|\widehat{\Sigma} - \Sigma\|_{\mathcal{H}_\infty} < \gamma$$

holds for any  $\Sigma \in \Sigma$  and any  $\widehat{\Sigma} \in \widehat{\Sigma}$ , if and only if there exist a matrix  $K = K^\top > 0$  in  $\mathbb{R}^{(n+r) \times (n+r)}$  partitioned as

$$K = \begin{bmatrix} K_{11} & K_{12} \\ K_{12}^\top & K_{22} \end{bmatrix}, \quad \text{with } K_{11} \in \mathbb{R}^{n \times n}$$

and scalars  $\delta > 0, \eta > 0$  and  $\mu$  such that (16) shown at the bottom of this page, holds, where  $N$  and  $N_{V,W}$  are given by (9) and (11), respectively.

*Proof:* The proof can be found in Appendix VI-B.  $\square$

In order to obtain the smallest upper bound, i.e., the smallest  $\gamma$  such that the conditions in Theorem 3 hold, one may solve the following semidefinite program (SDP) [34, Sec. 6.4]:

$$\min_{K, \delta, \eta, \mu} \gamma \quad (19a)$$

$$\text{subject to } K > 0, \delta > 0, \eta > 0 \text{ and (16)}. \quad (19b)$$

In conclusion, the solution of (19) gives  $\|\widehat{\Sigma}_0 - \Sigma_{\text{true}}\| < \gamma$ . Note that this upper bound is uniform for any  $\widehat{\Sigma}_0$  picked from  $\widehat{\Sigma}$ . Therefore, it can be regarded as an a priori error bound.

## C. A Posteriori Error Bound

Now suppose we pick a *known* system  $\widehat{\Sigma}_0$  from  $\widehat{\Sigma}$ . Let its realization be  $(\widehat{A}_0, \widehat{B}_0, \widehat{C}_0, \widehat{D}_0)$ . An a posteriori error bound is computed to measure the error between this known system  $\widehat{\Sigma}_0$

$$\left[ \begin{array}{c|c} \text{blkdiag}(K_{11}, (\frac{1}{2} - \mu)I_p, -K_{11}, -\gamma^{-2}I_m) & \text{blkdiag}(K_{12}, -\mu I_p, -K_{12}, -\gamma^{-2}I_m) \\ \hline \text{blkdiag}(K_{12}^\top, -\mu I_p, -K_{12}, -\gamma^{-2}I_m) & \text{blkdiag}(K_{22}, (\frac{1}{2} - \mu)I_p, -K_{22}, -\gamma^{-2}I_m) \end{array} \right] - \text{blkdiag}(\delta N, \eta N_{V,W}) > 0 \quad (16)$$

and the true system  $\Sigma_{\text{true}}$ . However, since the true system is unknown, this error cannot be directly computed.

Fortunately, we know that

$$\|\widehat{\Sigma}_0 - \Sigma_{\text{true}}\|_{\mathcal{H}_\infty} \leq \sup \left\{ \|\widehat{\Sigma}_0 - \Sigma\|_{\mathcal{H}_\infty} : \Sigma \in \Sigma \right\}.$$

The following proposition gives the computation of an upper bound of  $\|\widehat{\Sigma}_0 - \Sigma\|_{\mathcal{H}_\infty}$  for any  $\Sigma \in \Sigma$ .

*Proposition 3:* Let  $\widehat{\Sigma}_0$  be a given reduced-order model with realization  $(\widehat{A}_0, \widehat{B}_0, \widehat{C}_0, \widehat{D}_0) \in \widehat{\Sigma}$ . Then, the bound

$$\|\widehat{\Sigma}_0 - \Sigma\|_{\mathcal{H}_\infty} < \gamma_0$$

holds for any  $\Sigma \in \Sigma$ , if and only if there exist  $K = K^\top > 0$  in  $\mathbb{R}^{(n+r) \times (n+r)}$  partitioned as

$$K = \begin{bmatrix} K_{11} & K_{12} \\ K_{12}^\top & K_{22} \end{bmatrix}, \text{ with } K_{11} \in \mathbb{R}^{n \times n},$$

and a scalar  $\delta > 0$  such that (22) shown at the bottom of this page holds.

*Proof:* The proof is presented in Appendix VI-B.  $\square$

Similar to before, one may obtain the smallest upper bound by solving the following SDP:

$$\min_{K, \delta} \gamma_0 \quad (20a)$$

$$\text{subject to } K > 0, \delta > 0, \text{ and (22)}. \quad (20b)$$

Finally, we have  $\|\widehat{\Sigma}_0 - \Sigma_{\text{true}}\|_{\mathcal{H}_\infty} < \gamma_0$ . Note that this a posteriori error bound holds for a specific  $\widehat{\Sigma}_0$ . It follows readily that  $\gamma_0 \leq \gamma$  where  $\gamma$  and  $\gamma_0$  are the solutions of (19) and (20), respectively, since the bound  $\gamma_0$  holds for a specific  $\widehat{\Sigma}_0 \in \widehat{\Sigma}$  while the bound  $\gamma$  holds for any  $\widehat{\Sigma}$  taken from  $\widehat{\Sigma}$ .

The full data-driven balanced truncation approach is summarized in Algorithm 1.

## V. ILLUSTRATIVE EXAMPLE

Consider a continuous-time system of a cart with double pendulum, see [35] for details. After discretizing this model using the zero-order hold method with sampling time 0.5 s, we obtain a true discrete-time system of the form (5) where  $A_{\text{true}}$ ,  $B_{\text{true}}$ ,  $C_{\text{true}}$ , and  $D_{\text{true}}$  are given by

$$\begin{bmatrix} 0.9299 & 0.4160 & 0.7447 & 0.2291 & 0.2452 & 0.0592 \\ -0.1869 & 0.7430 & 0.3318 & 0.7617 & 1.0859 & 0.3560 \\ 0.0380 & 0.0477 & -0.3644 & 0.0647 & 0.1370 & 0.0766 \\ 0.0169 & 0.0549 & -0.0972 & -0.3693 & -0.8685 & 0.0484 \\ 0.0250 & 0.0285 & 0.2741 & 0.1393 & -0.0474 & 0.1615 \\ 0.1108 & 0.1358 & -1.7370 & 0.1855 & -1.8002 & -0.2311 \end{bmatrix}$$

### Algorithm 1: Data-Driven Generalized Lyapunov BT.

**Input:** Input, state, output data  $(U_-, X, Y_-)$  and noise model  $\mathcal{N}$ .

**Output:** The set  $\widehat{\Sigma}$  and error bounds  $\gamma$  and  $\gamma_0$ .

*Initialisation:* Construct data matrix  $N$ .

- 1: Construct matrices  $N_O$  and  $N_C$ .
- 2: **if**  $\begin{bmatrix} X_- \\ U_- \end{bmatrix}$  is full rank, LMIs (14) and (15) are feasible **then**
- 3: Compute  $P$  and  $Q$ .
- 4: **end if**
- 5: Construct balancing transformation  $T$  and  $\Sigma_H$  from  $P$  and  $Q$ .
- 6: Select a reduced dimension  $r$  based on  $\Sigma_H$  to construct matrices  $(\widehat{W}, \widehat{V})$  and reduced-order data matrix  $N_{V,W}$ .
- 7: Solve SDP (19).
- 8: Pick one  $(A_0, B_0, C_0, D_0) \in \widehat{\Sigma}$  and solve SDP (20).
- 9: **return**  $\widehat{\Sigma}$ ,  $\gamma$ , and  $\gamma_0$ .

$$\begin{bmatrix} 0.0701 & 0.1869 & -0.0380 & -0.0169 & -0.0250 & -0.1108 \end{bmatrix}^\top, \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \text{ and } 0, \text{ respectively.}$$

To illustrate the data-driven model reduction from noisy data, we collect input/state/output data of system (5) up to  $L = 200$  for input signal

$$u(k) = 2 \sin(k) + \cos(0.5k) \quad (21)$$

and a random initial condition  $x(0)$  which follows a Gaussian distribution with zero mean and unit variance. In addition, we take the noise  $w$  and  $z$  in (5) to be Gaussian with zero mean and variance  $\sigma^2$ . In this example, we assume knowledge of a bound on the energy of the noise which corresponds to the noise model (7) with  $\Phi_{11} = 1.35\sigma^2 I$ ,  $\Phi_{12} = 0$  and  $\Phi_{22} = -I$ . We verified that this noise model is satisfied by the generated noise sequence. For a realization of this noise, we obtain the data matrices  $U_-, X_-, X_+, Y_-$ .

We simulate the noise with different levels:  $\sigma \in \{0.002, 0.005, 0.01, 0.03, 0.05\}$  and it has been checked that they satisfy the noise model above. Hence, we characterize all systems explaining the data in a QMI of the form (8), where  $N \in \mathbb{R}^{14 \times 14}$ .

First, we check the generalized Slater condition (10) by verifying that  $N$  has 7 positive eigenvalues. Next, we can verify that conditions (i), (ii), and (iii) of Theorem 2 are satisfied, for each  $\sigma \in \{0.002, 0.005, 0.01, 0.03, 0.05\}$ , meaning that the data are

$$\begin{bmatrix} K_{11} & 0 & 0 & 0 & K_{12} \\ 0 & I_p - \widehat{C}_0 K_{22} \widehat{C}_0^\top - \gamma_0^{-2} \widehat{D}_0 \widehat{D}_0^\top & \widehat{C}_0 K_{12}^\top & \gamma_0^{-2} \widehat{D}_0 & \widehat{C}_0 K_{22} \widehat{A}_0^\top + \gamma_0^{-2} \widehat{D}_0 \widehat{B}_0^\top \\ 0 & K_{12} \widehat{C}_0^\top & -K_{11} & 0 & -K_{12} \widehat{A}_0^\top \\ 0 & \gamma_0^{-2} \widehat{D}_0^\top & 0 & -\gamma_0^{-2} I_m & -\gamma_0^{-2} \widehat{B}_0^\top \\ K_{12}^\top & \widehat{A}_0 K_{22} \widehat{C}_0^\top + \gamma_0^{-2} \widehat{B}_0 \widehat{D}_0^\top & -\widehat{A}_0 K_{12}^\top & -\gamma_0^{-2} \widehat{B}_0 & K_{22} - \widehat{A}_0 K_{22} \widehat{A}_0^\top - \gamma_0^{-2} \widehat{B}_0 \widehat{B}_0^\top \end{bmatrix} - \begin{bmatrix} \delta N & 0 \\ 0 & 0 \end{bmatrix} > 0. \quad (22)$$



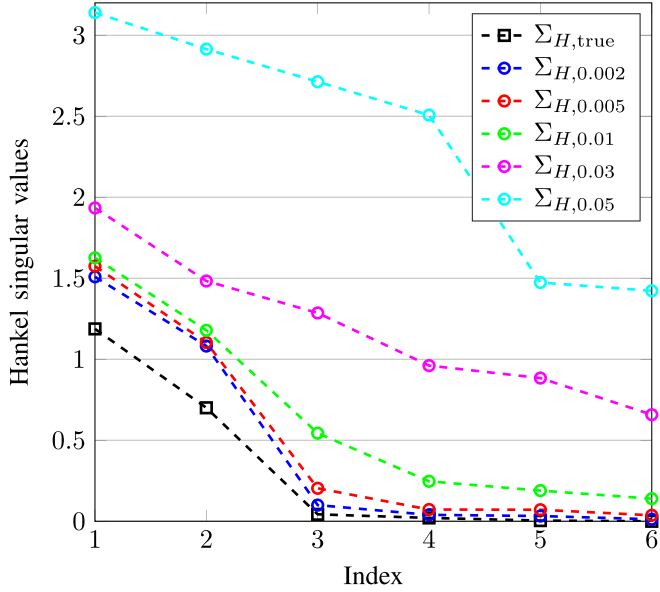


Fig. 1. Hankel singular values of the true system denoted by  $\Sigma_{H,true}$  and generalized Hankel singular values of all systems explaining the data for different noise levels denoted by  $\Sigma_{H,\sigma}$ . The values  $\Sigma_{H,true}$  are computed via the ordinary balancing procedure, i.e., balancing by Lyapunov equations, and using the system matrices of  $\Sigma_{true}$ .

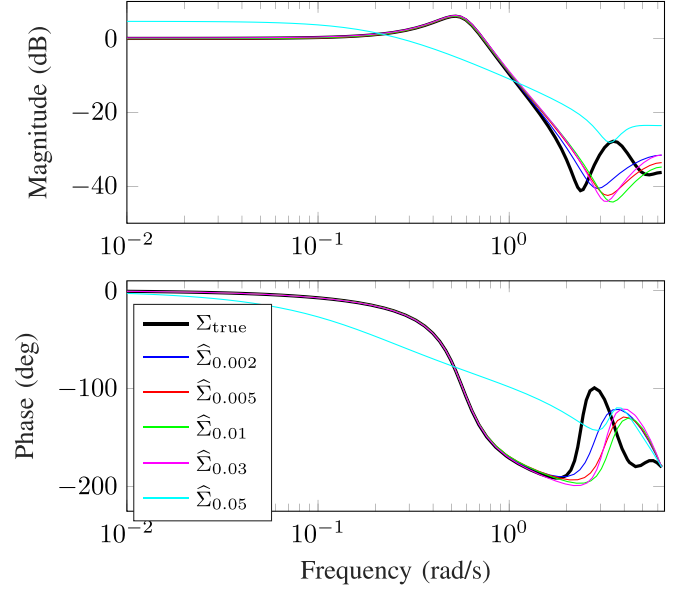


Fig. 2. Bode plots of reduced-order models for five different noise levels. The systems  $\hat{\Sigma}_{0.002}$ ,  $\hat{\Sigma}_{0.005}$ ,  $\hat{\Sigma}_{0.01}$ ,  $\hat{\Sigma}_{0.03}$ , and  $\hat{\Sigma}_{0.05}$  denote reduced-order models of systems explaining the data with noise levels  $\sigma = 0.002, 0.005, 0.01, 0.03$ , and  $0.05$ , respectively, while  $\Sigma_{true}$  denotes the true system. We stress that  $\Sigma_{\sigma}$  is one from infinitely many systems in  $\hat{\Sigma}_{\sigma}$  and  $\Sigma_{true}$  is assumed to be unknown.

informative for generalized Lyapunov balancing. Conditions (ii) and (iii) of Theorem 2 are semidefinite programs that we solve in MATLAB, using Yalmip [36] with SDPT3 [37] as an LMI solver. As a result, we obtain generalized Hankel singular values that are common to all systems explaining the data. They are depicted in Fig. 1 for various noise levels.

Fig. 1 shows that the generalized Hankel singular values (generalized Gramians) obtained from the data indeed bound the ordinary Hankel singular values (Gramians) of the true system, see also the discussion in Section II-B. We stress however that the true system (and, hence, its Gramians) is assumed to be unknown. Additionally, we observe that the generalized Hankel singular values provide less strict bounds on the unknown ordinary Hankel singular values when the noise level is increased.

In the balancing process, we also obtain the data-driven balancing transformation  $T$  as well as the matrices  $\hat{W}$  and  $\hat{V}$  corresponding to balanced truncation. Here, we take reduced models of order  $r = 3$  and therefore we have  $\hat{W}, \hat{V} \in \mathbb{R}^{6 \times 3}$ . By these matrices, the set  $\hat{\Sigma}$ , i.e., the set of reduced-order models via data-driven generalized balanced truncation, can be defined in terms of a QMI as stated in Corollary 1. We note that the set of reduced-order models  $\hat{\Sigma}$  is only characterized by matrix  $N_{V,W} \in \mathbb{R}^{8 \times 8}$ , which is of reduced dimension (with respect to  $N$ ).

In this example, for each noise level which is indicated by  $\sigma$ , we have a different set of reduced-order models denoted by  $\hat{\Sigma}_{\sigma}$ . Then, from each set  $\hat{\Sigma}_{\sigma}$ , we pick a reduced-order model  $\hat{\Sigma}_{\sigma}$ . We stress that for a given noise level,  $\hat{\Sigma}_{\sigma}$  depicts one from infinitely many possible reduced-order models contained in  $\hat{\Sigma}_{\sigma}$ . The Bode diagram of the reduced-order systems  $\hat{\Sigma}_{\sigma}$ 's is depicted in Fig. 2. Additionally, the time-domain output of  $\hat{\Sigma}_{0.03}$  is shown

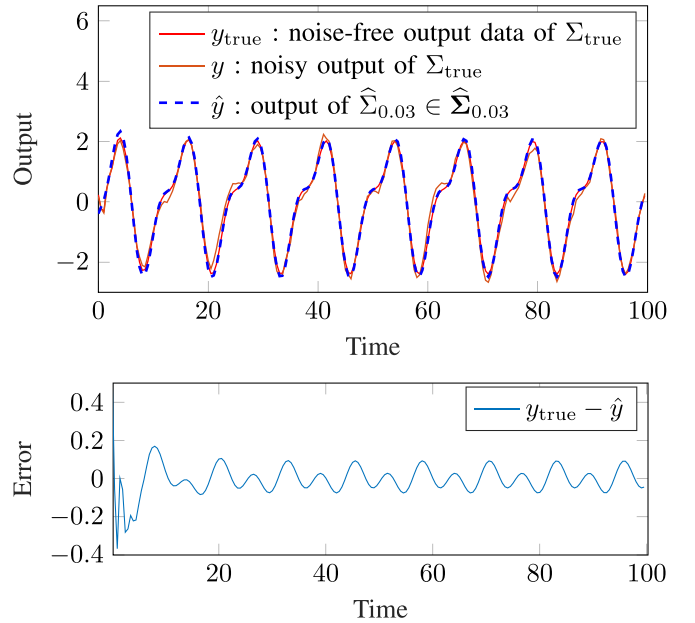


Fig. 3. Time-domain output of a reduced-order model corresponding to the noise level  $\sigma = 0.03$  with input signal (21):  $u(k) = 2 \sin(k) + \cos(0.5k)$  compared to the noise-free and noisy output of the true system with the same input signal.

in Fig. 3. The Bode diagram in Fig. 2 shows that reduced-order models accurately approximate the true system at least up to the noise level  $\sigma = 0.03$ . But, if we increase the noise, e.g.,  $\sigma = 0.05$ , the resulting reduced-order model may not be able to accurately approximate the true system. From Fig. 3, we see

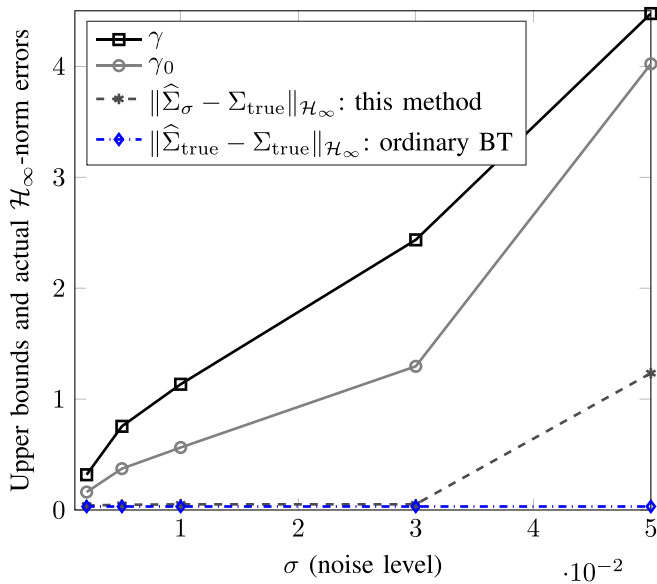


Fig. 4. Comparison of a priori error bounds  $\gamma$  [solution of Problem (19)], a posteriori error bounds  $\gamma_0$  (solution of Problem 20), the actual  $\mathcal{H}_\infty$ -norm of the error between the true system  $\Sigma_{\text{true}}$  and a reduced-order model  $\hat{\Sigma}_\sigma$  using this method and the  $\mathcal{H}_\infty$ -norm of the error between the true system and its reduced-order model via ordinary balanced truncation (ordinary BT).

that the reduced-order model corresponding to noise  $\sigma = 0.03$  is able to reconstruct the output data of the true system.

Next, we will compute the error bounds for the reduced-order models in this framework. We note first that from the result of Theorem 2, all systems in  $\Sigma$  and therefore  $\hat{\Sigma}$  are guaranteed to be asymptotically stable. As a consequence, the LMIs (16) and (22) are guaranteed to be feasible for some large enough  $\gamma$  and  $\gamma_0$ , respectively. Hence, we can solve problems (19) and (20). The results can be found in Fig. 4.

From Fig. 4, the upper bounds on the error with respect to the true system either from a priori or a posteriori upper bounds are getting more conservative when the noise levels are increased. This conservatism is caused by the fact that the only knowledge that is available on the true system is that it is contained in  $\Sigma$ , the set of systems explaining the data. As the set  $\Sigma$  has a larger size for increasing noise level, this leads to more conservative results. It is also clear that the a posteriori upper bound [the solution of problem (20)] is less conservative than the corresponding a priori upper bound [solutions of problem (19)] for each noise level.

In spite of the conservatism, the actual  $\mathcal{H}_\infty$ -norms of the errors between the true system and the reduced-order models selected from  $\hat{\Sigma}_\sigma$  for some small enough noise levels show that this data-driven method performs well. In particular, the  $\mathcal{H}_\infty$ -norm of the errors for noise levels  $\sigma = 0.002, 0.005, 0.01$ , and  $0.03$  which are given by  $0.0405, 0.0470, 0.0507$ , and  $0.0513$ , respectively, are relatively small compared to the error of reduction by the ordinary balanced truncation, which is equal to  $0.0314$ . We stress, however, that computation via the ordinary balanced truncation requires the knowledge of the true system which cannot be achieved on the basis of the data.

## VI. CONCLUSION

In this article, a data-driven procedure to obtain reduced-order models from noisy data is developed. The procedure begins with introducing the concept of data reduction. Based on the noise model introduced by [29], all (higher-order) systems explaining the data can be characterized in a high-dimensional QMI with a special structure. Due to this special structure, the class of reduced-order models obtained by applying a Petrov–Galerkin projection to all systems explaining the data can be characterized in a reduced-order QMI. As these QMIs depend only on the data and matrices corresponding to projection, this can be regarded as a data reduction procedure. Since this concept holds for general Petrov–Galerkin projections, it can potentially be extended to solve data-driven model reduction problems via any projection-based reduction technique.

We then follow up the data reduction concept by constructing specific projection from data. In particular, based on generalized controllability and observability Gramians, we provide necessary and sufficient conditions such that all systems explaining the data have common generalized Gramians. These conditions substitute Lyapunov inequalities by a data-guided linear matrix inequality which can be solved efficiently by modern LMI solvers. Subsequently, a common balancing transformation and therefore common matrices corresponding to projection for generalized balanced truncation (which are in the class of the Petrov–Galerkin projections) are available to apply the data reduction. As such, a set of reduced-order models via generalized balanced truncation can then be characterized in a lower-dimensional QMI. Moreover, all reduced-order models in this set are guaranteed to be asymptotically stable and computable a priori and a posteriori upper bounds on the reduction error with respect to the true system are available.

Beside the extension on exploiting data reduction via any projection-based reduction technique as mentioned above, ideas for future work include several directions. First, we aim at extending this result for input–output noisy data, even though an obstacle in this setting may be the construction of a state sequence. Second, investigating model reduction with preserving specific system properties such as network structure and port-Hamiltonian structure is often desirable.

## APPENDIX

### A. Proof of Proposition 2

*Proof:* To prove the “only if” part, suppose that  $\Sigma$  is bounded and has nonempty interior. Since  $\Sigma$  has nonempty interior, then (10) holds for some  $\bar{S}$ . Next, let  $N$  be defined as in (9) and partitioned as

$$N = \begin{bmatrix} N_{11} & N_{12} \\ N_{12}^\top & N_{22} \end{bmatrix}$$

where  $N_{11} \in \mathbb{R}^{(n+p) \times (n+p)}$ ,  $N_{12} \in \mathbb{R}^{(n+p) \times (n+m)}$ , and  $N_{22} \in \mathbb{R}^{(n+m) \times (n+m)}$ . By Lemma 1, matrix  $N$  is nonsingular with

$N_{11} - N_{12}N_{22}^{-1}N_{12}^\top > 0$  and  $N_{22} < 0$ . As a result, the matrix

$$\begin{bmatrix} I & 0 & X_+ \\ 0 & I & Y_- \\ 0 & 0 & -X_- \\ 0 & 0 & -U_- \end{bmatrix}$$

and therefore  $\begin{bmatrix} X_- \\ U_- \end{bmatrix}$  are full row rank.

To prove the ‘‘if’’ part, observe that it follows from (10) and [38, Fact 5.8.16] that

$$v_+(N) \geq n + p. \quad (23)$$

Note that

$$N_{22} = \begin{bmatrix} X_- \\ U_- \end{bmatrix} \Phi_{22} \begin{bmatrix} X_- \\ U_- \end{bmatrix}^\top.$$

Since  $\Phi_{22} < 0$  and  $\begin{bmatrix} X_- \\ U_- \end{bmatrix}$  has full row rank, we have that  $N_{22} < 0$ . Then, following Haynsworth’s inertia additivity formula, see [38, Fact 6.5.5], we have

$$\text{In}(N) = \text{In}(N_{22}) + \text{In}(N_{11} - N_{12}N_{22}^{-1}N_{12}^\top)$$

and hence

$$v_-(N) \geq n + m. \quad (24)$$

Since  $N \in \mathbb{R}^{(m+2n+p) \times (m+2n+p)}$ ,  $v_-(N) + v_+(N) \leq m + 2n + p$ . This, together with (23) and (24), implies that  $v_+(N) + v_-(N) = m + 2n + p$ . Therefore,  $N$  has no zero eigenvalues. Consequently, it is nonsingular with  $N_{22} < 0$  and  $N_{11} - N_{12}N_{22}^{-1}N_{12}^\top > 0$ . Then, it follows from (8) and Lemma 1 that  $\Sigma$  is bounded and has nonempty interior.  $\square$

## B. Proof of Theorem 1

Before giving the proof of Theorem 1, we will develop some general results on quadratic matrix inequalities.

To this end, consider the set

$$\mathcal{M} := \left\{ Z \in \mathbb{R}^{p \times q} : \begin{bmatrix} I \\ Z^\top \end{bmatrix}^\top \Psi \begin{bmatrix} I \\ Z^\top \end{bmatrix} \geq 0 \right\} \quad (25)$$

where  $\Psi$  admits the partitioning

$$\Psi = \begin{bmatrix} \Psi_{11} & \Psi_{12} \\ \Psi_{12}^\top & \Psi_{22} \end{bmatrix} \quad (26)$$

with  $\Psi_{11} \in \mathbb{R}^{p \times p}$ . We assume throughout this Appendix that  $\Psi_{22} < 0$  and  $\Psi_{11} - \Psi_{12}\Psi_{22}^{-1}\Psi_{12}^\top > 0$ . It is known from Lemma 1 that  $\mathcal{M}$  is bounded and has nonempty interior. As a consequence of this,  $\mathcal{M}$  admits various representations, as stated next.  $\square$

*Lemma A1:* Consider  $\mathcal{M}$  given by (25). Define  $\Psi|\Psi_{22} := \Psi_{11} - \Psi_{12}\Psi_{22}^{-1}\Psi_{12}^\top$  and

$$\Psi^\# := \begin{bmatrix} 0 & -I_q \\ I_p & 0 \end{bmatrix} \Psi^{-1} \begin{bmatrix} 0 & -I_p \\ I_q & 0 \end{bmatrix}. \quad (27)$$

Suppose that  $\Psi|\Psi_{22} > 0$  and  $\Psi_{22} < 0$ . Then,  $\mathcal{M} = \mathcal{M}_1 = \mathcal{M}_2$ , where

$$\mathcal{M}_1 := \left\{ Z : \begin{bmatrix} I \\ Z \end{bmatrix}^\top \Psi^\# \begin{bmatrix} I \\ Z \end{bmatrix} \geq 0 \right\},$$

$$\mathcal{M}_2 := \left\{ Z : \begin{bmatrix} I \\ Z^\top \end{bmatrix}^\top \Psi \begin{bmatrix} I \\ Z^\top \end{bmatrix} = \mathcal{Q} \text{ where } 0 \leq \mathcal{Q} \leq \Psi|\Psi_{22} \right\}.$$

*Proof:* The proof of  $\mathcal{M} = \mathcal{M}_1$  is provided in [31]. We will prove that  $\mathcal{M} = \mathcal{M}_2$ . Clearly,  $\mathcal{M}_2 \subseteq \mathcal{M}$ . Then, it remains to show that the reverse inclusion holds. Let  $Z \in \mathcal{M}$  and let

$$\mathcal{Q} = \begin{bmatrix} I \\ Z^\top \end{bmatrix}^\top \Psi \begin{bmatrix} I \\ Z^\top \end{bmatrix}.$$

Clearly,  $\mathcal{Q} \geq 0$ . Note that

$\mathcal{Q} = \Psi|\Psi_{22} + (Z^\top + \Psi_{22}\Psi_{12}^\top)^\top \Psi_{22} (Z^\top + \Psi_{22}\Psi_{12}^\top) \leq \Psi|\Psi_{22}$  since  $\Psi_{22} < 0$ . Therefore,  $0 \leq \mathcal{Q} \leq \Psi|\Psi_{22}$  and, as a result,  $Z \in \mathcal{M}_2$ .  $\square$

Now, we will consider projections of the elements of the set  $\mathcal{M}$ . Let  $V \in \mathbb{R}^{q \times \hat{q}}$  and  $W \in \mathbb{R}^{p \times \hat{p}}$  be full column rank matrices with  $\hat{p} \leq p$  and  $\hat{q} \leq q$ . We refer to the set

$$\mathcal{M}_{V,W} := \{W^\top ZV : Z \in \mathcal{M}\}$$

as a reduction of  $\mathcal{M}$  using matrices  $W$  and  $V$ . Note that we do not assume that  $W^\top V = I$ .

We will show that elements of  $\mathcal{M}_{V,W}$  themselves satisfy a quadratic matrix inequality. To do so, let  $Z \in \mathcal{M}$  such that we have

$$\begin{bmatrix} I \\ Z^\top \end{bmatrix}^\top \begin{bmatrix} \Psi_{11} & \Psi_{12} \\ \Psi_{12}^\top & \Psi_{22} \end{bmatrix} \begin{bmatrix} I \\ Z^\top \end{bmatrix} \geq 0$$

which can be written as

$$\Psi|\Psi_{22} + (Z + \Psi_{12}\Psi_{22}^{-1})\Psi_{22}(Z^\top + \Psi_{22}^{-1}\Psi_{12}^\top) \geq 0. \quad (28)$$

Using the Schur complement, (28) is equivalent to

$$\begin{bmatrix} \Psi|\Psi_{22} & Z + \Psi_{12}\Psi_{22}^{-1} \\ Z^\top + \Psi_{22}^{-1}\Psi_{12}^\top & -\Psi_{22}^{-1} \end{bmatrix} \geq 0 \quad (29)$$

where we note that the inverse of  $\Psi_{22}$  exists as  $\Psi_{22} < 0$  by assumption. Pre and postmultiplying (29) by  $\text{blkdiag}(W^\top, V^\top)$  and  $\text{blkdiag}(W, V)$ , respectively, gives

$$\begin{bmatrix} W^\top(\Psi|\Psi_{22})W & W^\top(Z + \Psi_{12}\Psi_{22}^{-1})V \\ V^\top(Z^\top + \Psi_{22}^{-1}\Psi_{12}^\top)W & -V^\top\Psi_{22}^{-1}V \end{bmatrix} \geq 0. \quad (30)$$

Let us define  $\hat{Z} = W^\top ZV$ . Using again a Schur complement argument and writing the result as a quadratic matrix inequality, we obtain that (30) is equivalent to

$$\begin{bmatrix} I \\ \hat{Z}^\top \end{bmatrix}^\top \Psi_{V,W} \begin{bmatrix} I \\ \hat{Z}^\top \end{bmatrix} \geq 0 \quad (31)$$

where  $\Psi_{V,W}$  is given by (32) shown at the bottom of the next page. This shows that if  $Z \in \mathcal{M}$  then  $W^\top ZV$  satisfies (31).

Stated differently, we have that  $\mathcal{M}_{V,W} \subseteq \{\hat{Z} : (31) \text{ holds}\}$ . In fact, we have that the equality holds as asserted in the following theorem.

*Theorem A1:* It holds that

$$\mathcal{M}_{V,W} = \left\{ \hat{Z} : \begin{bmatrix} I \\ \hat{Z}^\top \end{bmatrix}^\top \Psi_{V,W} \begin{bmatrix} I \\ \hat{Z}^\top \end{bmatrix} \geq 0 \right\}.$$

Before we present the proof of Theorem A1, we need to state some auxiliary lemmas. First, we recall the following result from linear algebra [38, Fact 5.10.19].

*Lemma A2:* Let  $X, Y \in \mathbb{R}^{p \times q}$ . Then,  $X^\top X = Y^\top Y$ , if and only if  $Y = UX$  where  $U$  is an orthogonal matrix, i.e.,  $U^\top U = I$ .

The next two lemmas are central to prove Theorem A1.

*Lemma A3:* Let  $S = S^\top \in \mathbb{R}^{p \times p}$  such that  $S > 0$ . Let  $V \in \mathbb{R}^{p \times r}$  be a full rank matrix with  $r \leq p$  and  $Q_V = Q_V^\top \in \mathbb{R}^{r \times r}$  such that  $0 \leq Q_V \leq V^\top S V$ . Then, there exists  $Q = Q^\top$  such that  $0 \leq Q \leq S$  and  $V^\top Q V = Q_V$ .

Equations (33) and (34) shown at the bottom of this page.

*Proof:* Since  $V$  is full rank, there exists a matrix  $\tilde{V}$  such that the square matrix  $T_V := [V \ \tilde{V}]$  is nonsingular. Note that

$$T_V^\top S T_V = F^\top \begin{bmatrix} V^\top S V & 0 \\ 0 & (T_V^\top S T_V) \setminus (V^\top S V) \end{bmatrix} F$$

where

$$F := \begin{bmatrix} I & (V^\top S V)^{-1} V^\top S \tilde{V} \\ 0 & I \end{bmatrix}.$$

Let

$$\tilde{Q} := F^\top \begin{bmatrix} Q_V & 0 \\ 0 & \Delta \end{bmatrix} F$$

where  $0 \leq \Delta \leq (T_V^\top S T_V) \setminus (V^\top S V)$ . Clearly, we have  $0 \leq \tilde{Q} \leq T_V^\top S T_V$ . Take  $Q = T_V^{-\top} \tilde{Q} T_V^{-1}$  to guarantee that  $0 \leq Q \leq S$  and  $V^\top Q V = Q_V$ .  $\square$

*Lemma A4:* Consider the matrix  $\Psi \in \mathbb{R}^{(p+q) \times (p+q)}$  partitioned as in (26) such that  $\Psi_{11} - \Psi_{12} \Psi_{22}^{-1} \Psi_{12}^\top > 0$  and  $\Psi_{22} < 0$ . Let  $W \in \mathbb{R}^{p \times r}$  be a full rank matrix with  $r \leq p$ . Suppose that  $Z_W$  satisfies

$$\begin{bmatrix} I \\ Z_W^\top \end{bmatrix}^\top \begin{bmatrix} W^\top \Psi_{11} W & W^\top \Psi_{12} \\ \Psi_{12}^\top W & \Psi_{22} \end{bmatrix} \begin{bmatrix} I \\ Z_W^\top \end{bmatrix} \geq 0. \quad (35)$$

Then, there exists  $Z$  such that  $Z^\top W = Z_W^\top$  and

$$\begin{bmatrix} I \\ Z^\top \end{bmatrix}^\top \Psi \begin{bmatrix} I \\ Z^\top \end{bmatrix} \geq 0.$$

*Proof:* Note that (35) can be written as

$$W^\top (\Psi | \Psi_{22}) W + \bar{Z}_W \Psi_{22} \bar{Z}_W^\top \geq 0$$

where  $\bar{Z}_W = Z_W + W^\top \Psi_{12} \Psi_{22}^{-1}$ . Let

$$Q_W := W^\top (\Psi | \Psi_{22}) W + \bar{Z}_W \Psi_{22} \bar{Z}_W^\top. \quad (36)$$

Then,  $0 \leq Q_W \leq W^\top (\Psi | \Psi_{22}) W$  since  $\Psi_{22} < 0$ . By Lemma A3, there exists  $Q = Q^\top$  such that  $0 \leq Q \leq \Psi | \Psi_{22}$  and  $W^\top Q W = Q_W$ . As  $\Psi | \Psi_{22} - Q \geq 0$ , there exists a matrix  $R$  such that

$$\Psi | \Psi_{22} - Q = R^\top R$$

which implies

$$Q_W = W^\top (\Psi | \Psi_{22}) W - W^\top R^\top R W. \quad (37)$$

By comparing (36) and (37), it follows from Lemma A2 that

$$(-\Psi_{22})^{\frac{1}{2}} \bar{Z}_W^\top = U R W$$

for some orthogonal matrix  $U$ . Note that since  $W$  is full rank, there exists  $\tilde{W}$  such that the square matrix  $[W \ \tilde{W}]$  is nonsingular. Now, let  $\tilde{Z}_W$  be such that

$$(-\Psi_{22})^{\frac{1}{2}} \tilde{Z}_W^\top = U R \tilde{W}.$$

Next, define

$$Z^\top = \begin{bmatrix} \bar{Z}_W^\top & \tilde{Z}_W^\top \end{bmatrix} \begin{bmatrix} W & \tilde{W} \end{bmatrix}^{-1} - \Psi_{22}^{-1} \Psi_{12}^\top$$

which can be easily checked to verify  $Z^\top W = Z_W^\top$ . Moreover

$$\begin{aligned} & \begin{bmatrix} I \\ Z^\top \end{bmatrix}^\top \begin{bmatrix} \Psi_{11} & \Psi_{12} \\ \Psi_{12}^\top & \Psi_{22} \end{bmatrix} \begin{bmatrix} I \\ Z^\top \end{bmatrix} \\ &= \Psi | \Psi_{22} + (Z + \Psi_{12} \Psi_{22}^{-1}) \Psi_{22} (Z + \Psi_{12} \Psi_{22}^{-1})^\top \\ &= \Psi | \Psi_{22} - R^\top R = Q \geq 0 \end{aligned}$$

as desired.

Now, we are ready to prove Theorem A1.

*Proof of Theorem A1:* It is clear that

$$\mathcal{M}_{V,W} \subseteq \left\{ \hat{Z} : (31) \text{ holds} \right\}.$$

$$\Psi_{V,W} := \begin{bmatrix} W^\top (\Psi | \Psi_{22} + \Psi_{12} \Psi_{22}^{-1} V (V^\top \Psi_{22}^{-1} V)^{-1} V^\top \Psi_{22}^{-1} \Psi_{12}^\top) W & W^\top \Psi_{12} \Psi_{22}^{-1} V (V^\top \Psi_{22}^{-1} V)^{-1} \\ \hline (V^\top \Psi_{22}^{-1} V)^{-1} V^\top \Psi_{22}^{-1} \Psi_{12}^\top W & (V^\top \Psi_{22}^{-1} V)^{-1} \end{bmatrix} \quad (32)$$

$$\Psi_V := \begin{bmatrix} \Psi | \Psi_{22} + \Psi_{12} \Psi_{22}^{-1} V (V^\top \Psi_{22}^{-1} V)^{-1} V^\top \Psi_{22}^{-1} \Psi_{12}^\top & \Psi_{12} \Psi_{22}^{-1} V (V^\top \Psi_{22}^{-1} V)^{-1} \\ \hline (V^\top \Psi_{22}^{-1} V)^{-1} V^\top \Psi_{22}^{-1} \Psi_{12}^\top & (V^\top \Psi_{22}^{-1} V)^{-1} \end{bmatrix} \quad (33)$$

$$\Psi_V^\# := \begin{bmatrix} -V^\top (\Psi_{22}^{-1} - \Psi_{22}^{-1} \Psi_{12}^\top (\Psi | \Psi_{22})^{-1} \Psi_{12} \Psi_{22}^{-1}) V & -V^\top \Psi_{22}^{-1} \Psi_{12}^\top (\Psi | \Psi_{22})^{-1} \\ \hline -(\Psi | \Psi_{22})^{-1} \Psi_{12} \Psi_{22}^{-1} V & -(\Psi | \Psi_{22})^{-1} \end{bmatrix}. \quad (34)$$

To prove the reverse inclusion, let  $\widehat{Z}$  be such that (31) holds, i.e.,

$$\begin{bmatrix} I \\ \widehat{Z}^\top \end{bmatrix}^\top \Psi_{V,W} \begin{bmatrix} I \\ \widehat{Z}^\top \end{bmatrix} \geq 0.$$

We will show that there exists  $Z \in \mathcal{M}$  such that  $\widehat{Z} = W^\top ZV$ . By Lemma A4, there exists  $Z_V$  such that  $Z_V^\top W = \widehat{Z}^\top$  and

$$\begin{bmatrix} I \\ Z_V^\top \end{bmatrix}^\top \Psi_V \begin{bmatrix} I \\ Z_V^\top \end{bmatrix} \geq 0 \quad (38)$$

where  $\Psi_V$  is given by (33). From Lemma A1, (38) is equivalent to

$$\begin{bmatrix} I \\ Z_V \end{bmatrix}^\top \Psi_V^\# \begin{bmatrix} I \\ Z_V \end{bmatrix} \geq 0 \quad (39)$$

where  $\Psi_V^\#$  is given by (34). Using Lemma A4 again, (39) implies the existence of  $Z$  such that  $ZV = Z_V$  and

$$\begin{bmatrix} I \\ Z \end{bmatrix}^\top \Psi^\# \begin{bmatrix} I \\ Z \end{bmatrix} \geq 0$$

where  $\Psi^\#$  is given by (27). Therefore, there exists  $Z$  such that  $\widehat{Z} = W^\top ZV$  and, due to Lemma A1,  $Z \in \mathcal{M}$ .

*Proof of Theorem 1:* Since (10) holds for some  $\bar{S}$  and matrix  $\begin{bmatrix} X_- \\ U_- \end{bmatrix}$  has full row rank,  $\Sigma$  is bounded and has nonempty interior. In addition, since  $\widehat{W}^\top \widehat{V} = I$ , then  $W$  and  $V$  are full column rank. Therefore, the claim follows from the result of Theorem A1.  $\square$

### C. Strict Matrix S-Lemma

*Proposition 1* ([29, Th. 11]): Let  $F, G \in \mathbb{R}^{(q+r) \times (q+r)}$  be symmetric matrices. Assume that

$$\Sigma_G := \left\{ V \in \mathbb{R}^{r \times q} : \begin{bmatrix} I \\ V \end{bmatrix}^\top G \begin{bmatrix} I \\ V \end{bmatrix} \geq 0 \right\}$$

is bounded. Consider the statements

i) There exists some matrix  $\bar{V} \in \mathbb{R}^{r \times q}$  such that

$$\begin{bmatrix} I \\ \bar{V} \end{bmatrix}^\top G \begin{bmatrix} I \\ \bar{V} \end{bmatrix} > 0.$$

ii)  $\begin{bmatrix} I \\ V \end{bmatrix}^\top F \begin{bmatrix} I \\ V \end{bmatrix} > 0 \forall V \in \mathbb{R}^{r \times q}$  with  $\begin{bmatrix} I \\ V \end{bmatrix}^\top G \begin{bmatrix} I \\ V \end{bmatrix} \geq 0$ .

iii) There exists a scalar  $\alpha \geq 0$  such that  $F - \alpha G > 0$ .

Then, the following implications hold:

I) i) and ii)  $\Rightarrow$  iii).

II) iii)  $\Rightarrow$  ii).

### D. Proof of Theorem 2

*Proof:* Let us first prove the ‘‘only if’’ statement. Suppose that the data  $(U_-, X, Y_-)$  are informative for GLB. By Definition 1, there exist  $P = P^\top > 0$  and  $Q = Q^\top > 0$  such that (12) and

(13) hold for all  $(A, B, C, D)$  satisfying (8). We begin with statement (i). Let  $\xi \in \mathbb{R}^n$  and  $\eta \in \mathbb{R}^m$  be such that

$$\begin{bmatrix} \xi^\top & \eta^\top \end{bmatrix} \begin{bmatrix} X_- \\ U_- \end{bmatrix} = 0.$$

Moreover, let  $(A, B, C, D) \in \Sigma$  and  $\zeta \in \mathbb{R}^n$  be a nonzero vector. Note that

$$(A + \alpha\zeta\xi^\top, B + \alpha\zeta\eta^\top, C, D) \in \Sigma$$

for every  $\alpha \in \mathbb{R}$ , as can be concluded from (8). Since the data are informative for GLB, there exists  $P = P^\top > 0$  such that

$$P - A_\alpha P A_\alpha^\top - B_\alpha B_\alpha^\top > 0 \quad (40)$$

where  $A_\alpha := A + \alpha\zeta\xi^\top$  and  $B_\alpha := B + \alpha\zeta\eta^\top$ . Note that (40) holds for every  $\alpha \in \mathbb{R}$ . Then, by dividing (40) by  $\alpha^2$  and letting  $\alpha \rightarrow \infty$ , we obtain

$$(-\xi^\top P \xi - \eta^\top \eta) \zeta \zeta^\top \geq 0.$$

Since  $P > 0$  and  $\zeta \neq 0$ , we see that  $\xi = 0$  and  $\eta = 0$ . Therefore,

$\begin{bmatrix} X_- \\ U_- \end{bmatrix}$  has full row rank.

To show (ii) and (iii), we first rewrite the matrix inequalities (12) and (13) as the quadratic matrix inequalities

$$\begin{bmatrix} I \\ A^\top \\ B^\top \end{bmatrix}^\top \begin{bmatrix} P & 0 & 0 \\ 0 & -P & 0 \\ 0 & 0 & -I_m \end{bmatrix} \begin{bmatrix} I \\ A^\top \\ B^\top \end{bmatrix} > 0 \quad (41)$$

and

$$\begin{bmatrix} I \\ A \\ C \end{bmatrix}^\top \begin{bmatrix} Q & 0 & 0 \\ 0 & -Q & 0 \\ 0 & 0 & -I_p \end{bmatrix} \begin{bmatrix} I \\ A \\ C \end{bmatrix} > 0, \quad (42)$$

respectively. Note that we have the QMI (8) characterizing the set of all systems explaining the data. Moreover, since  $\begin{bmatrix} X_- \\ U_- \end{bmatrix}$  has full row rank and we assume that there exists  $\bar{S}$  such that (10) holds, it follows from the proof of Proposition 2 that systems explaining the data are equivalently characterized by

$$\begin{bmatrix} I & 0 \\ 0 & I \\ A & B \\ C & D \end{bmatrix}^\top N^\# \begin{bmatrix} I & 0 \\ 0 & I \\ A & B \\ C & D \end{bmatrix} \geq 0 \quad (43)$$

where

$$N^\# := \begin{bmatrix} 0 & -I_{n+m} \\ I_{n+p} & 0 \end{bmatrix} N^{-1} \begin{bmatrix} 0 & -I_{n+p} \\ I_{n+m} & 0 \end{bmatrix}$$

see Lemma A1. From a projection of (8) and by considering Lemma A4, we have that all  $(A, B)$  satisfying (8) are equivalent to those satisfying

$$\begin{bmatrix} I \\ A^\top \\ B^\top \end{bmatrix}^\top N_C \begin{bmatrix} I \\ A^\top \\ B^\top \end{bmatrix} \geq 0 \quad (44)$$

where  $N_C$  is given by

$$N_C := \begin{bmatrix} I_n & 0 \\ 0 & 0 \\ 0 & I_{n+m} \end{bmatrix}^\top N \begin{bmatrix} I_n & 0 \\ 0 & 0 \\ 0 & I_{n+m} \end{bmatrix}.$$

Similarly, all  $(A, C)$  satisfying (43) are equivalent to those satisfying

$$\begin{bmatrix} I \\ A \\ C \end{bmatrix}^\top N_O \begin{bmatrix} I \\ A \\ C \end{bmatrix} \geq 0 \quad (45)$$

where

$$N_O := \begin{bmatrix} I_n & 0 \\ 0 & 0 \\ 0 & I_{n+p} \end{bmatrix}^\top N^\# \begin{bmatrix} I_n & 0 \\ 0 & 0 \\ 0 & I_{n+p} \end{bmatrix}.$$

Now, we are ready to apply the matrix S-lemma from Appendix VI-A. In particular, by informativity for GLB, (41) holds for all  $(A, B)$  satisfying (44), such that the use of the matrix S-lemma (Proposition 1 in Appendix VI-A) yields (14) and proves (ii). The proof of (iii) is similar, using (42) and (45).

To prove the “if” statement, first suppose that  $\begin{bmatrix} X_- \\ U_- \end{bmatrix}$  has full row rank. Then, under Assumption (10),  $N$  is nonsingular with  $N_{22} < 0$  and  $N_{11} - N_{12}N_{22}^{-1}N_{12}^\top > 0$ . Thus,  $N_O$  is well defined. Now, suppose that statements (ii) and (iii) are satisfied. Then, the matrix S-lemma in Proposition 1 implies that (41) and (42) hold for all  $(A, B)$  and  $(A, C)$  satisfying (44) and (45), respectively. This implies that (12) and (13) hold for all systems explaining the data, i.e., the data are informative for generalized Lyapunov balancing.  $\square$

### E. Proof of Theorem 3

*Proof:* We will prove the upper bound by employing the bounded real lemma. To do so, consider any  $\Sigma \in \Sigma$  and  $\widehat{\Sigma} \in \widehat{\Sigma}$  with realizations  $(A, B, C, D)$  and  $(\widehat{A}, \widehat{B}, \widehat{C}, \widehat{D})$ , respectively. Then, a realization for  $\widehat{\Sigma} - \Sigma$  is given by the quadruplet

$$A_d := \begin{bmatrix} A & 0 \\ 0 & \widehat{A} \end{bmatrix}, B_d := \begin{bmatrix} B \\ \widehat{B} \end{bmatrix}, C_d := \begin{bmatrix} C & -\widehat{C} \end{bmatrix}, D_d := D - \widehat{D}.$$

Let  $\gamma > 0$ . By (the discrete-time version of) the bounded real lemma, e.g., [39, Thm. 4.6.6 (iv)], the matrix  $A_d$  satisfies  $\rho(A_d) < 1$  and  $\|\widehat{\Sigma} - \Sigma\|_{\mathcal{H}_\infty} < \gamma$ , if and only if there exists

$K \in \mathbb{R}^{(n+r) \times (n+r)}$  with  $K = K^\top > 0$  such that

$$\begin{bmatrix} K & 0 \\ 0 & I_p \end{bmatrix} - \begin{bmatrix} A_d & B_d \\ C_d & D_d \end{bmatrix} \begin{bmatrix} K & 0 \\ 0 & \gamma^{-2}I_m \end{bmatrix} \begin{bmatrix} A_d & B_d \\ C_d & D_d \end{bmatrix}^\top > 0. \quad (46)$$

If (46) holds for all  $\Sigma \in \Sigma$  and  $\widehat{\Sigma} \in \widehat{\Sigma}$  (for the same  $K$ ), then clearly the norm  $\|\Sigma - \widehat{\Sigma}\|_{\mathcal{H}_\infty}$  is upper bounded by  $\gamma$  for all choices of systems in  $\Sigma$  and  $\widehat{\Sigma}$ . Note that (46) can be written in the QMI form

$$\begin{bmatrix} I & 0 \\ 0 & I \\ A_d^\top & C_d^\top \\ B_d^\top & D_d^\top \end{bmatrix}^\top \begin{bmatrix} K & 0 & 0 & 0 \\ 0 & I_p & 0 & 0 \\ 0 & 0 & -K & 0 \\ 0 & 0 & 0 & -\gamma^{-2}I_m \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & I \\ A_d^\top & C_d^\top \\ B_d^\top & D_d^\top \end{bmatrix} > 0. \quad (47)$$

We will show the equivalence of (16) and the satisfaction of (47) for all systems in  $\Sigma$  and  $\widehat{\Sigma}$  by using the matrix S-lemma. As a first step, we introduce the notation

$$J := \begin{bmatrix} I & 0 \\ 0 & I \\ A^\top & C^\top \\ B^\top & D^\top \end{bmatrix} \quad \text{and} \quad \widehat{J} := \begin{bmatrix} I & 0 \\ 0 & I \\ \widehat{A}^\top & \widehat{C}^\top \\ \widehat{B}^\top & \widehat{D}^\top \end{bmatrix} \quad (48)$$

such that the data (8) and

$$\begin{bmatrix} I & 0 \\ 0 & I \\ \widehat{A}^\top & \widehat{C}^\top \\ \widehat{B}^\top & \widehat{D}^\top \end{bmatrix}^\top N_{V,W} \begin{bmatrix} I & 0 \\ 0 & I \\ \widehat{A}^\top & \widehat{C}^\top \\ \widehat{B}^\top & \widehat{D}^\top \end{bmatrix} \geq 0$$

can be written as  $J^\top N J \geq 0$  and  $\widehat{J}^\top N_{V,W} \widehat{J} \geq 0$ , respectively. On the other hand, it can be checked that (47) is equivalent to

$$\Gamma^\top \begin{bmatrix} J & 0 \\ 0 & \widehat{J} \end{bmatrix}^\top \begin{bmatrix} \tilde{\Theta}_{11} & \tilde{\Theta}_{12} \\ \tilde{\Theta}_{12}^\top & \tilde{\Theta}_{22} \end{bmatrix} \begin{bmatrix} J & 0 \\ 0 & \widehat{J} \end{bmatrix} \Gamma > 0 \quad (49)$$

where

$$\begin{aligned} \tilde{\Theta}_{12} &:= \text{blkdiag}(K_{12}, 0, -K_{12}, -\gamma^{-2}I_m), \\ \tilde{\Theta}_{ii} &:= \text{blkdiag}(K_{ii}, \frac{1}{2}I_p, -K_{ii}, -\gamma^{-2}I_m) \end{aligned}$$

for  $i = 1, 2$ , and

$$\Gamma := \begin{bmatrix} I_n & 0 & 0 \\ 0 & 0 & I_p \\ 0 & I_r & 0 \\ 0 & 0 & -I_p \end{bmatrix}.$$

Note that (49) means that

$$x^\top \begin{bmatrix} J & 0 \\ 0 & \widehat{J} \end{bmatrix}^\top \begin{bmatrix} \tilde{\Theta}_{11} & \tilde{\Theta}_{12} \\ \tilde{\Theta}_{12}^\top & \tilde{\Theta}_{22} \end{bmatrix} \begin{bmatrix} J & 0 \\ 0 & \widehat{J} \end{bmatrix} x > 0$$

for all  $x \in \text{im}(\Gamma) \setminus \{0\} \subset \mathbb{R}^{n+r+p}$  or, equivalently,  $x \in \ker[0 \ I_p \ 0 \ I_p] \setminus \{0\}$ . Let  $R = [0 \ I_p \ 0 \ I_p]$ , then

by Finsler's lemma [40], (49) is equivalent to

$$\begin{bmatrix} J & 0 \\ 0 & \hat{J} \end{bmatrix}^\top \begin{bmatrix} \tilde{\Theta}_{11} & \tilde{\Theta}_{12} \\ \tilde{\Theta}_{12}^\top & \tilde{\Theta}_{22} \end{bmatrix} \begin{bmatrix} J & 0 \\ 0 & \hat{J} \end{bmatrix} - \mu R^\top R > 0$$

for some  $\mu$ , which can be written as

$$\begin{bmatrix} J & 0 \\ 0 & \hat{J} \end{bmatrix}^\top \begin{bmatrix} \Theta_{11} & \Theta_{12} \\ \Theta_{12}^\top & \Theta_{22} \end{bmatrix} \begin{bmatrix} J & 0 \\ 0 & \hat{J} \end{bmatrix} > 0 \quad (50)$$

with

$$\Theta_{12} := \text{blkdiag}(K_{12}, -\mu I_p, -K_{12}, -\gamma^{-2} I_m)$$

and

$$\Theta_{ii} := \text{blkdiag}(K_{ii}, \left(\frac{1}{2} - \mu\right) I_p, -K_{ii}, -\gamma^{-2} I_m),$$

for  $i = 1, 2$ . The motivation of writing (47) in the form (50) is that the later form can be written in a QMI with the same quadratic variable as  $J^\top N J \geq 0$ . Namely, by using a Schur complement argument, (50) is equivalent to

$$J^\top \left( \Theta_{11} - \Theta_{12} \hat{J} \left( \hat{J}^\top \Theta_{22} \hat{J} \right)^{-1} \hat{J}^\top \Theta_{12}^\top \right) J > 0 \quad (51)$$

and  $\hat{J}^\top \Theta_{22} \hat{J} > 0$ . This form allows us to use the matrix S-lemma in Proposition 1 such that QMI  $J^\top N J \geq 0$  implies (51). Particularly, (51) holds with  $J$  satisfying  $J^\top N J \geq 0$ , if and only if

$$\Theta_{11} - \delta N - \Theta_{12} \hat{J} \left( \hat{J}^\top \Theta_{22} \hat{J} \right)^{-1} \hat{J}^\top \Theta_{12}^\top > 0 \quad (52)$$

for some  $\delta > 0$ . To this end, we assume that  $\hat{J}^\top \Theta_{22} \hat{J} > 0$  holds for all  $\hat{J}$  satisfying  $\hat{J}^\top N_{V,W} \hat{J} \geq 0$ . We will see that this assumption is satisfied after completing the proof.

Next, by using the (backward) Schur complement, (52) together with  $\hat{J}^\top \Theta_{22} \hat{J} > 0$  is equivalent to

$$\begin{bmatrix} \Theta_{11} - \delta N & \Theta_{12} \hat{J} \\ \hat{J}^\top \Theta_{12}^\top & \hat{J}^\top \Theta_{22} \hat{J} \end{bmatrix} > 0.$$

Then, a Schur complement with respect to the block matrix  $\Theta_{11} - \delta N$  results in

$$\hat{J}^\top \left( \Theta_{22} - \Theta_{12}^\top (\Theta_{11} - \delta N)^{-1} \Theta_{12} \right) \hat{J} > 0 \quad (53)$$

and  $\Theta_{11} - \delta N > 0$ . Using Proposition 1 again, (53) holds for  $\hat{J}$  satisfying  $\hat{J}^\top N_{V,W} \hat{J} \geq 0$ , if and only if

$$\Theta_{22} - \eta N_{V,W} - \Theta_{12}^\top (\Theta_{11} - \delta N)^{-1} \Theta_{12} > 0$$

for some  $\eta > 0$ . Finally, a (backward) Schur complement argument implies that this is equivalent to (16) as desired. Here, we have seen that  $\hat{J}$  satisfies  $\hat{J}^\top \Theta_{22} \hat{J} > 0$  for any  $\hat{J}$  satisfying  $\hat{J}^\top N_{V,W} \hat{J} \geq 0$  as an implication of (53).  $\square$

### F. Proof of Proposition 3

*Proof:* Let  $\gamma_0 > 0$ . Then, from the bounded real lemma [39], [41], we have that  $\|\hat{\Sigma}_0 - \Sigma\|_{\mathcal{H}_\infty} < \gamma_0$  if and only if there exists

$K = K^\top > 0$  in  $\mathbb{R}^{(n+r) \times (n+r)}$  such that

$$\begin{bmatrix} K & 0 \\ 0 & I_p \end{bmatrix} - \begin{bmatrix} A_d & B_d \\ C_d & D_d \end{bmatrix} \begin{bmatrix} K & 0 \\ 0 & \gamma_0^{-2} I_m \end{bmatrix} \begin{bmatrix} A_d & B_d \\ C_d & D_d \end{bmatrix}^\top > 0 \quad (54)$$

where

$$A_d := \begin{bmatrix} A & 0 \\ 0 & \hat{A}_0 \end{bmatrix}, B_d := \begin{bmatrix} B \\ \hat{B}_0 \end{bmatrix}, \\ C_d := \begin{bmatrix} C & -\hat{C}_0 \end{bmatrix}, D_d := D - \hat{D}_0.$$

Next, we will show that condition (22) is equivalent to the existence of  $K = K^\top > 0$  such that (54) holds for any system in  $\Sigma$ .

First, we introduce  $J$  to denote the matrix as in (48). This allows to write (8) into  $J^\top N J \geq 0$  and moreover (54) into

$$\begin{bmatrix} J^\top \bar{\Theta}_{11} J & J^\top \bar{\Theta}_{12} \\ \bar{\Theta}_{12}^\top J & \bar{\Theta}_{22} \end{bmatrix} > 0 \quad (55)$$

where

$$\bar{\Theta}_{11} := \begin{bmatrix} K_{11} & 0 & 0 & 0 \\ 0 & I_p - \hat{C}_0 K_{22} \hat{C}_0^\top - \gamma_0^{-2} \hat{D}_0 \hat{D}_0^\top & \hat{C}_0 K_{12}^\top & \gamma_0^{-2} \hat{D}_0 \\ 0 & K_{12} \hat{C}_0^\top & -K_{11} & 0 \\ 0 & \gamma_0^{-2} \hat{D}_0^\top & 0 & -\gamma_0^{-2} I_m \end{bmatrix}, \\ \bar{\Theta}_{12} := \begin{bmatrix} K_{12} \\ \hat{C}_0 K_{22} \hat{A}_0^\top + \gamma_0^{-2} \hat{D}_0 \hat{B}_0^\top \\ -K_{12} \hat{A}_0^\top \\ -\gamma_0^{-2} \hat{B}_0^\top \end{bmatrix},$$

and  $\bar{\Theta}_{22} := K_{22} - \hat{A}_0 K_{22} \hat{A}_0^\top - \gamma_0^{-2} \hat{B}_0 \hat{B}_0^\top$ , which are denoting the block elements of the first matrix in (22). Furthermore, the Schur complement of (55) admits that  $\bar{\Theta}_{22} > 0$  and

$$J^\top (\bar{\Theta}_{11} - \bar{\Theta}_{12} \bar{\Theta}_{22}^{-1} \bar{\Theta}_{12}^\top) J > 0. \quad (56)$$

Hence, by the strict matrix S-lemma in Proposition 1, (56) holds with  $J$  satisfying  $J^\top N J \geq 0$ , if and only if

$$\bar{\Theta}_{11} - \bar{\Theta}_{12} \bar{\Theta}_{22}^{-1} \bar{\Theta}_{12}^\top - \delta N > 0 \quad (57)$$

for some  $\delta > 0$ . Finally, (57) and  $\bar{\Theta}_{22} > 0$  yield (22) via the (backward) Schur complement.  $\square$

### REFERENCES

- [1] B. C. Moore, "Principal component analysis in linear systems: Controllability, observability, and model reduction," *IEEE Trans. Autom. Control*, vol. AC-26, no. 1, pp. 17–32, Feb. 1981.
- [2] L. Pernebo and L. Silverman, "Model reduction via balanced state space representations," *IEEE Trans. Autom. Control*, vol. AC-27, no. 2, pp. 382–387, Apr. 1982.
- [3] D. F. Enns, "Model reduction with balanced realizations: An error bound and a frequency weighted generalization," in *Proc. IEEE 23rd Conf. Decis. Control*, 1984, pp. 127–132.
- [4] P. Benner and T. Damm, "Lyapunov equations, energy functionals, and model order reduction of bilinear and stochastic systems," *SIAM J. Control Optim.*, vol. 49, no. 2, pp. 686–711, 2011.

- [5] K. Glover, "All optimal Hankel-norm approximations of linear multivariable systems and their  $\mathcal{L}_\infty$ -error bounds," *Int. J. Control*, vol. 39, no. 6, pp. 1115–1193, 1984.
- [6] P. Feldmann and R. W. Freund, "Efficient linear circuit analysis by Padé approximation via the Lanczos process," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 14, no. 5, pp. 639–649, May 1995.
- [7] E. J. Grimme, "Krylov projection methods for model reduction," Ph.D. dissertation, Dept. Elect. Eng., Univ. Illinois at Urbana-Champaign, Urbana, IL, 1997.
- [8] K. Gallivan, A. Vandendorpe, and P. Van Dooren, "Model reduction of MIMO systems via tangential interpolation," *SIAM J. Matrix Anal. Appl.*, vol. 26, no. 2, pp. 328–349, 2004.
- [9] A. Astolfi, "Model reduction by moment matching for linear and nonlinear systems," *IEEE Trans. Autom. Control*, vol. 55, no. 10, pp. 2321–2336, Oct. 2010.
- [10] J. M. A. Scherpen, "Balancing for nonlinear systems," *Syst. Control Lett.*, vol. 21, no. 2, pp. 143–153, 1993.
- [11] B. Besselink, N. van de Wouw, J. M. A. Scherpen, and H. Nijmeijer, "Model reduction for nonlinear systems by incremental balanced truncation," *IEEE Trans. Autom. Control*, vol. 59, no. 10, pp. 2739–2753, Oct. 2014.
- [12] T. C. Ionescu and A. Astolfi, "Nonlinear moment matching-based model order reduction," *IEEE Trans. Autom. Control*, vol. 61, no. 10, pp. 2837–2847, Oct. 2015.
- [13] A. C. Antoulas, *Approximation of Large-Scale Dynamical Systems*. Philadelphia, PA: SIAM, 2005.
- [14] A. C. Antoulas, C. A. Beattie, and S. Gugercin, *Interpolatory Methods for Model Reduction*. Philadelphia, PA: Computational Science and Engineering, SIAM, 2020.
- [15] P. Benner, S. Gugercin, and K. Willcox, "A survey of projection-based model reduction methods for parametric dynamical systems," *SIAM Rev.*, vol. 57, no. 4, pp. 483–531, 2015.
- [16] P. Rapisarda and H. L. Trentelman, "Identification and data-driven model reduction of state-space representations of lossless and dissipative systems from noise-free data," *Automatica*, vol. 47, no. 8, pp. 1721–1728, 2011.
- [17] I. Markovsky, J. C. Willems, P. Rapisarda, and B. L. M. De Moor, "Algorithms for deterministic balanced subspace identification," *Automatica*, vol. 41, no. 5, pp. 755–766, 2005.
- [18] I. V. Gosea, S. Gugercin, and C. Beattie, "Data-driven balancing of linear dynamical systems," *SIAM J. Sci. Comput.*, vol. 44, no. 1, pp. A554–A582, 2022.
- [19] S. Lall, J. E. Marsden, and S. Glavaški, "A subspace approach to balanced truncation for model reduction of nonlinear control systems," *Int. J. Robust Nonlinear Control*, vol. 12, no. 6, pp. 519–535, 2002.
- [20] Y. Kawano and J. M. A. Scherpen, "Empirical differential Gramians for nonlinear model reduction," *Automatica*, vol. 127, 2021, Art. no. 109534.
- [21] Y. Kawano, B. Besselink, J. M. A. Scherpen, and M. Cao, "Data-driven model reduction of monotone systems by nonlinear DC gains," *IEEE Trans. Autom. Control*, vol. 65, no. 5, pp. 2094–2106, May 2020.
- [22] A. J. Mayo and A. C. Antoulas, "A framework for the solution of the generalized realization problem," *Linear Algebra Its Appl.*, vol. 425, no. 2–3, pp. 634–662, 2007.
- [23] I. V. Gosea and A. C. Antoulas, "Stability preserving post-processing methods applied in the Loewner framework," in *Proc. IEEE 20th Workshop Signal Power Integrity*, 2016, pp. 1–4.
- [24] C. Beattie and S. Gugercin, "Realization-independent  $\mathcal{H}_2$ -approximation," in *Proc. IEEE 51st Conf. Decis. Control*, 2012, pp. 4953–4958.
- [25] B. Peherstorfer, S. Gugercin, and K. Willcox, "Data-driven reduced model construction with time-domain Loewner models," *SIAM J. Sci. Comput.*, vol. 39, no. 5, pp. A2152–A2178, 2017.
- [26] Z. Drmač and B. Peherstorfer, "Learning low-dimensional dynamical-system models from noisy frequency-response data with Loewner rational interpolation," in *Realization Model Reduction Dyn. Syst.*, 2022, pp. 39–57.
- [27] A. M. Burohman, B. Besselink, J. M. A. Scherpen, and M. K. Camlibel, "From data to reduced-order models via moment matching," 2020, *arXiv:2011.00150*.
- [28] H. J. van Waarde, J. Eising, H. L. Trentelman, and M. K. Camlibel, "Data informativity: A new perspective on data-driven analysis and control," *IEEE Trans. Autom. Control*, vol. 65, no. 11, pp. 4753–4768, Nov. 2020.
- [29] H. J. van Waarde, M. K. Camlibel, and M. Mesbahi, "From noisy data to feedback controllers: Non-conservative design via a matrix S-lemma," *IEEE Trans. Autom. Control*, vol. 67, no. 1, pp. 162–175, Jan. 2020.
- [30] G. E. Dullerud and F. Paganini, *A Course in Robust Control Theory: A Convex Approach*. New York, NY: Springer Science Business Media, 2000, vol. 36.
- [31] H. J. van Waarde, M. K. Camlibel, P. Rapisarda, and H. L. Trentelman, "Data-driven dissipativity analysis: Application of the matrix S-lemma," *IEEE Control Syst. Mag.*, vol. 42, no. 3, pp. 140–149, Jun. 2022.
- [32] H. J. van Waarde, M. K. Camlibel, J. Eising, and H. L. Trentelman, "Quadratic matrix inequalities with applications to data-based control," 2022, *arXiv:2203.12959*.
- [33] V. A. Yakubovich, "S-procedure in nonlinear control theory," *Vestnik Leningrad Univ. Math.*, vol. 4, pp. 73–93, 1977.
- [34] Y. Nesterov and A. Nemirovskii, *Interior-Point Polynomial Algorithms in Convex Programming*, vol. 13. Philadelphia, PA: SIAM, 1994.
- [35] T. C. Ionescu, "Two-sided time-domain moment matching for linear systems," *IEEE Trans. Autom. Control*, vol. 61, no. 9, pp. 2632–2637, Sep. 2015.
- [36] J. Lofberg, "YALMIP : A toolbox for modeling and optimization in MATLAB," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2004, pp. 284–289.
- [37] K. C. Toh, M. J. Todd, and R. H. Tütüncü, "SDPT3-A MATLAB software package for semidefinite programming, version 1.3," *Optim. Methods Softw.*, vol. 11, no. 1–4, pp. 545–581, 1999.
- [38] D. S. Bernstein, *Matrix Mathematics: Theory, Facts, and Formulas*, 2nd ed. Princeton, NJ: Princeton Univ. Press, 2009.
- [39] R. E. Skelton, T. Iwasaki, and K. M. Grigoriadis, *A Unified Algebraic Approach to Linear Control Design*. London, U.K.: Taylor & Francis Ltd., 1997.
- [40] P. Finsler, "Über das vorkommen definitiver und semidefinitiver formen in scharen quadratischer formen," *Commentarii Mathematici Helvetici*, vol. 9, pp. 188–192, 1936.
- [41] C. E. de Souza and L. Xie, "On the discrete-time bounded real lemma with application in the characterization of static state feedback  $\mathcal{H}_\infty$  controllers," *Syst. Control Lett.*, vol. 18, no. 1, pp. 61–71, 1992.



**Azka Muji Burohman** received the B.Sc. degree in mathematics and the M.Sc. degree in instrumentation and control, from Institut Teknologi Bandung, Bandung, Indonesia, in 2013 and 2017, respectively. He is currently working toward the Ph.D. degree in mathematical theory of systems and control with the University of Groningen, Groningen, The Netherlands.

His current research interests include model reduction and control of network systems.



**Bart Besselink** (Member, IEEE) received the M.Sc. (cum laude) degree in mechanical engineering and the Ph.D. degree in mechanical engineering from the Eindhoven University of Technology, Eindhoven, The Netherlands, in 2008 and 2012, respectively.

Since 2016, he has been an Assistant Professor with the Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence, University of Groningen, Groningen, The Netherlands. He was a short-term Visiting Researcher with the Tokyo Institute of Technology, Tokyo, Japan, in 2012. Between 2012 and 2016, he was a Postdoctoral Researcher with the ACCESS Linnaeus Centre and Department of Automatic Control, KTH Royal Institute of Technology, Stockholm, Sweden. His main research interests include mathematical systems theory for large-scale interconnected systems, with emphasis on contract-based design and control, compositional analysis, model reduction, and applications in intelligent transportation systems and neuromorphic computing.

Dr. Besselink was a recipient (with Xiaodong Cheng and Jacquelin Scherpen) of the 2020 Automatica Paper Prize.





**Jacqueliën M. A. Scherpen** (Fellow, IEEE) received the M.Sc. and Ph.D. degrees in 1990 and 1994 from the University of Twente, Enschede, the Netherlands.

She joined Delft University of Technology and became a Professor with the Engineering and Technology Institute Groningen (ENTEG), Faculty of Science and Engineering (FSE), University of Groningen (UG), Groningen, The Netherlands, in 2006. From 2013 till 2019, she was Scientific Director of ENTEG. She is currently

Director of Engineering with the UG, and Captain of Science with the Dutch top sector High Tech Systems and Materials (HTSM), The Netherlands. She has held various visiting research positions, such as with the University of Tokyo, Tokyo, Japan, and Kyoto University, Kyoto, Japan, Université de Compiègne, Compiègne, France, and SUPÉLEC, Gif-sur-Yvette, France, and Old Dominion University, VA, USA. She is with the Editorial Board of a few international journals among which include IEEE TRANSACTIONS ON AUTOMATIC CONTROL, and the *International Journal of Robust and Nonlinear Control*. She received the Automatica Best Paper Prize from 2017–2020. In 2019, she received a royal distinction as Knight in the Order of the Netherlands Lion. She has been active at the International Federation of Automatic Control (IFAC), and is currently a Member of the IFAC council. She was a Member of the Board of Governors of the IEEE Control Systems Society, and was Chair of the IEEE CSS Standing Committee on Women in Control in 2020. From 2020 to 2021, she was President of the European Control Association (EUCA), and currently she is chairing the SIAM Activity Group on Control and Systems Theory. Her current research interests include model reduction methods for networks, nonlinear model reduction methods, nonlinear control methods, modeling and control of physical systems with applications to electrical circuits, electromechanical systems, mechanical systems, smart energy networks and distributed optimal control applications to smart grids.



**M. Kanat Camlibel** (Senior Member, IEEE) received the Ph.D. degree in mathematical theory of systems and control from Tilburg University, Tilburg, The Netherlands, in 2001.

From 2001 to 2005, he held Postdoctoral positions with the University of Groningen, Groningen, The Netherlands, and Tilburg University, Tilburg, The Netherlands. Between 2005 and 2007, he was an Assistant Professor with Eindhoven University of Technology, Eindhoven, The Netherlands. In 2007, he joined the University

of Groningen where he is currently a Full Professor with the Bernoulli Institute for Mathematics, Computer Science, and Artificial Intelligence. His research interests include differential variational inequalities, complementarity problems, optimization, piecewise affine dynamical systems, switched linear systems, constrained linear systems, multi-agent systems, model reduction, geometric theory of linear systems, and data-driven control.

Dr. Camlibel is an Associate Editor for the IEEE TRANSACTIONS ON AUTOMATIC CONTROL, and is the past Associate Editor for *International Journal of Robust and Nonlinear Control*, *SIAM Journal on Control and Optimization*, and *Systems and Control Letters*.