Balanced Realizations, Model Order Reduction, and the Hankel Operator

Scherpen, Jacquelien M.A.

Published in:
The Control Handbook. Control System Advanced Methods

4

Balanced Realizations, Model Order Reduction, and the Hankel Operator

Jacquelien M.A. Scherpen
University of Groningen

4.1 Introduction

In many engineering applications, processes are described by increasingly complex models that are difficult to analyze and difficult to control. Reduction of the order of the model may overcome some of these difficulties, but it is quite possible that model reduction incurs a significant loss of accuracy. Therefore, the system has to be analyzed in a manner that is useful for the application purpose. Simplification of the model based on this analysis usually results in a model of lower complexity which is easier to handle, and in a corresponding simplification of synthesis procedures for control and filtering problems. Furthermore, the simplification decreases the computational effort. Every application has its own demands, and different model reduction methods have different properties.

In [1] two types of approximation methods for large-scale linear systems are discussed, namely singular-value decomposition (SVD) and Krylov methods. The Krylov methods are based on matching the moments of the impulse response of the linear system up to a certain level. This can be interpreted in terms of the series expansions of the transfer function; see, for example, [1] and the references therein. The key to the success of these methods is that they can be implemented iteratively and are able to handle much higher-order systems than the SVD methods. Nevertheless, they are not always resulting in models that are useful for control, for example, stability may not be preserved, and no *a priori* error bounds for the reduced order system can be given. Furthermore, an extension of these methods to nonlinear systems is still largely unexplored. Only recently, the first steps toward such extension is presented by [2]. The

SVD methods based on balanced realizations, on the other hand, offer a clear structure for analysis of the system based on controllability and observability properties, and model reduction by balanced truncation does preserve stability and other properties and does have an *a priori* error bound for the reduced order system. Furthermore, the extension of balancing to nonlinear systems has been studied in the past two decades as well.

Kalman's minimal realization theory (e.g., [27]) offers a clear picture of the structure of linear systems. However, the accompanying algorithms are not very satisfactory, since they are only textbook algorithms, which are numerically deficient. Moore [32] showed that there are very useful tools that may be used to cope with this problem. He used the principal component analysis which was introduced in statistics in the 1930s to analyze a linear system, and specifically to apply it to model reduction. The most important contribution of [32] is the introduction of balancing for stable minimal linear systems. The balancing method offers a tool to measure the contribution of the different state components to the past input and future output energy of the system, which are measures of controllability and observability. The algorithmic methods corresponding to the balancing theory nowadays are standard toolboxes in simulation packages like MATLAB®.

In the theory of continuous-time linear systems, the system Hankel operator plays an important role in a number of realization problems. For example, when viewed as a mapping from past inputs to future outputs, it plays a direct role in the abstract definition of *state*. It also plays a central role in minimality theory, in model reduction problems, and related to these, in linear identification methods. Specifically, the Hankel operator supplies a set of similarity invariants, the so-called Hankel singular values, which can be used to quantify the importance of each state in the corresponding input–output system. The Hankel operator can also be factored into the composition of an observability and controllability operator, from which Gramian matrices can be defined and the notion of a balanced realization follows. The Hankel singular values are most easily computed in a state-space setting using the product of the Gramian matrices, though intrinsically they depend only on the given input–output mapping. For linear systems, the Hankel operator offers an immediate relation with the frequency domain setting of balancing for linear systems, for example, [63].

Furthermore, these methods have proved to be very useful for application purposes. To mention a few older applications, we refer to [16,62]. Reference [62] successfully applies methods based on balanced realizations on the controller design of the Philips CD player. In [16], several balanced controller designs and algorithms for sensor and actuator placement based on balanced realizations are given and demonstrated for the NASA Deep Space Network Antenna.

For nonlinear systems, the first step toward extension of the linear balancing methods has been set in [45], where a balancing formalism is developed for stable nonlinear continuous-time state-space systems based on the idea that state components that correspond to low control costs and high output energy generation (in the sense of L_2 energy in a signal) are important for the description of the dynamics of the input–output behavior of the system, while state components with high control costs and low output energy generation can be left out of the description. Since then, many results on state-space balancing, modifications based on sliding time windows, and modifications based on proper orthogonal decomposition (POD), and computational issues for model reduction and related minimality considerations for nonlinear systems have appeared in the literature; for example, [18,20,28,37,38,58,59,64]. The relations of the original nonlinear balancing method of [45] with minimality are later explored in [47], and a more constructive approach that includes a strong relation with the nonlinear input–output operators, the nonlinear Hankel operator, and the Hankel norm of the system is presented in [10,12,48]. Model reduction based on these studies is recently treated in [13,14].

Here we first review the Hankel operator, balancing and balanced truncation for stable linear systems. Then unstable systems and closed-loop balancing are reviewed for linear systems. Finally, the extension of the linear theory to nonlinear systems is treated, based on balancing procedures in a neighborhood (possibly large, almost global) of an equilibrium point.

4.2 Linear Systems

In this section, we briefly review the well-known linear system definitions of the system Hankel matrix; the Hankel operator; the controllability and observability operators, Gramians and functions; and the balanced realization and the corresponding model reduction procedure.

4.2.1 The Hankel, Controllability, and Observability Operator

Consider a continuous-time, causal linear input–output system $\Sigma : u \to y$, with m-dimensional input space $u \in U$ and p-dimensional output space $y \in Y$, and with impulse response $H(t)$. Let

$$H(t) = \sum_{k=0}^{\infty} H_{k+1} \frac{t^k}{k!}, \quad t \geq 0 \tag{4.1}$$

denote its Taylor series expansion about $t = 0$, where $H_k \in \mathbb{R}^{p \times m}$ for each k. The system Hankel matrix is defined as $\hat{\mathcal{H}} = [\hat{\mathcal{H}}_{i,j}]$, where $\hat{\mathcal{H}}_{i,j} = H_{i+j-1}$ for $i, j \geq 1$. If Σ is also (bounded input bounded output) BIBO stable, then the system Hankel operator is the well-defined mapping

$$\begin{aligned} \mathcal{H} : L_2^m[0, +\infty) &\to L_2^p[0, +\infty), \\ : \hat{u} &\to \hat{y}(t) = \int_0^{\infty} H(t + \tau)\hat{u}(\tau)\, d\tau. \end{aligned} \tag{4.2}$$

mapping the past inputs to the future outputs. If we define the *time flipping* operator as

$$\begin{aligned} \tilde{\mathcal{F}} : L_2^m[0, +\infty) &\to L_2^m(-\infty, 0] \\ : \hat{u} &\to u(t) = \begin{cases} \hat{u}(-t) : t < 0 \\ 0 : t \geq 0, \end{cases} \end{aligned}$$

then clearly $\mathcal{H}(\hat{u}) = (\Sigma \circ \tilde{\mathcal{F}})(\hat{u})$; see the illustration in Figure 4.1. The lower side of the figure depicts the input–output behavior of the original operator Σ. The upper side depicts the input–output behavior of the Hankel operator of Σ, where the signal in the upper left side is the time-flipped signal of the lower left side signal. The flipping operator $\mathcal{F}$ is defined by $\mathcal{F}(u(t)) := u(-t)$. The upper right side signal is the truncated signal (to the space $L_2[0, \infty)$) of the lower left side signal. The corresponding truncation operator is given by

$$\mathcal{T}(y(t)) := \begin{cases} 0 & (t < 0) \\ y(t) & (t \geq 0) \end{cases},$$

and $\tilde{\mathcal{F}} = \mathcal{T} \circ \mathcal{F}$. The definition of a Hankel operator implies that it describes the mapping from the input to the output generated by the state at $t = 0$. Hence, we can analyze the relationship between the state and the input–output behavior of the original operator Σ by investigating its Hankel operator.

When $\mathcal{H}$ is known to be a compact operator, then its (Hilbert) adjoint operator, $\mathcal{H}^*$, is also compact, and the composition $\mathcal{H}^*\mathcal{H}$, is a self-adjoint compact operator with a well-defined spectral decomposition:

$$\mathcal{H}^*\mathcal{H} = \sum_{i=1}^{\infty} \sigma_i^2 \, \langle \cdot, \psi_i \rangle_{L_2} \, \psi_i, \quad \sigma_i \geq 0, \tag{4.3}$$

$$\langle \psi_i, \psi_j \rangle_{L_2} = \delta_{ij}, \quad \langle \psi_i, (\mathcal{H}^*\mathcal{H})(\psi_i) \rangle_{L_2} = \sigma_i^2, \tag{4.4}$$

where σ_i^2 is an eigenvalue of $\mathcal{H}^*\mathcal{H}$ with the corresponding eigenvector ψ_i, ordered as $\sigma_1 \geq \cdots \geq \sigma_n > \sigma_{n+1} = \sigma_{n+2} = 0$, and called the *Hankel singular values* for the input–output system Σ.

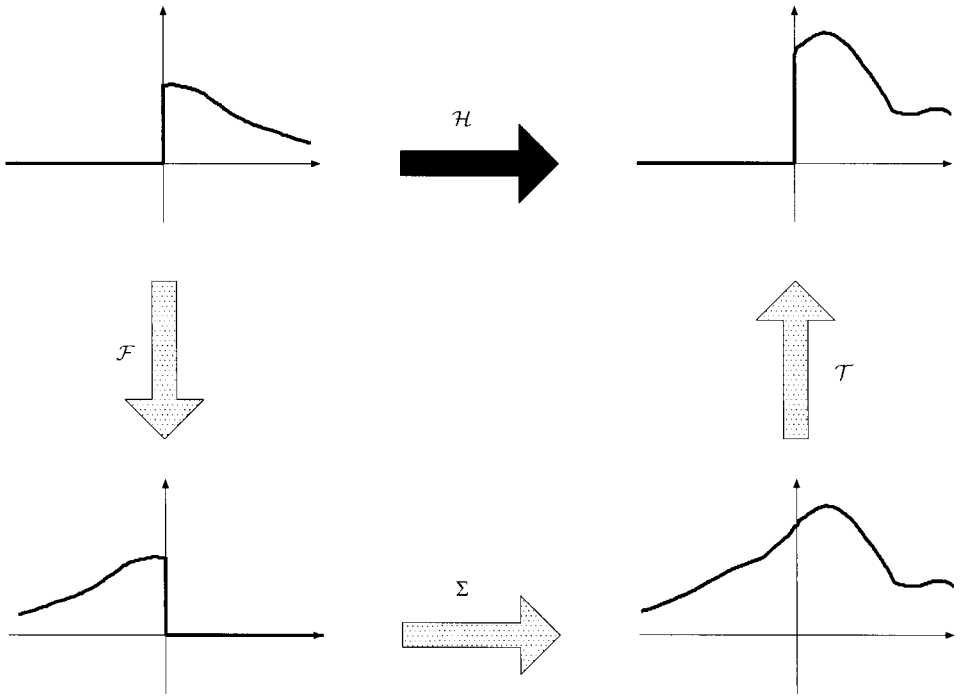

FIGURE 4.1 Hankel operator $\mathcal{H}$ of Σ.

Let (A, B, C) be a state-space realization of Σ with dimension n, that is, consider a linear system:

$$\begin{aligned} \dot{x} &= Ax + Bu, \\ y &= Cx, \end{aligned} \tag{4.5}$$

where $u \in \mathbb{R}^m$, $x \in \mathbb{R}^n$, and $y \in \mathbb{R}^p$. We assume that Equation 4.5 is *stable and minimal*, that is, controllable and observable. Any such realization induces a factorization of the system Hankel matrix into the form $\hat{\mathcal{H}} = \hat{\mathcal{O}}\hat{\mathcal{C}}$, where $\hat{\mathcal{O}}$ and $\hat{\mathcal{C}}$ are the (extended) observability and controllability matrices. If the realization is asymptotically stable, then the Hankel operator can be written as the composition of uniquely determined observability and controllability operators; that is, $\mathcal{H} = \mathcal{O}\mathcal{C}$, where the controllability and observability operators are defined as

$$\begin{aligned} \mathcal{C} &: L_2^m[0, +\infty) \to \mathbb{R}^n : \hat{u} \to \int_0^\infty e^{At} B\hat{u}(t)\, dt, \\ \mathcal{O} &: \mathbb{R}^n \to L_2^p[0, +\infty) : x \to \hat{y}(t) = Ce^{At}x. \end{aligned}$$

Since $\mathcal{C}$ and $\mathcal{O}$ have a finite dimensional range and domain, respectively, they are compact operators; and the composition $\mathcal{O}\mathcal{C}$ is also a compact operator. From the definition of the (Hilbert) adjoint operator, it is easily shown that $\mathcal{C}$ and $\mathcal{O}$ have corresponding adjoints

$$\begin{aligned} \mathcal{C}^* &: \mathbb{R}^n \to L_2^m[0, +\infty) : x \to B^T e^{A^T t} x, \\ \mathcal{O}^* &: L_2^p[0, +\infty) \to \mathbb{R}^n : y \to \int_0^\infty e^{A^T t} C^T y(t)\, dt. \end{aligned}$$

4.2.2 Balanced State-Space Realizations

The above input–output setting can be related with the well-known Gramians that are related to the state-space realization. In order to do so, we consider the energy functions given in the following definition.

Definition 4.1:

The controllability and observability functions of a smooth state-space system are defined as

$$L_c(x_0) = \min_{\substack{u\in L_2(-\infty,0)\\ x(-\infty)=0, x(0)=x_0}} \frac{1}{2}\int_{-\infty}^{0} \| u(t) \|^2 \, dt \tag{4.6}$$

and

$$L_o(x_0) = \frac{1}{2}\int_{0}^{\infty} \| y(t) \|^2 \, dt, \quad x(0) = x_0, \quad u(t) \equiv 0, \quad 0 \le t < \infty, \tag{4.7}$$

respectively.

The value of the controllability function at x_0 is the minimum amount of input energy required to reach the state x_0 from the zero state, and the value of the observability function at x_0 is the amount of output energy generated by the state x_0. The following results are well known:

Theorem 4.1:

Consider the system (Equation 4.5). Then $L_c(x_0) = \frac{1}{2}x_0^T P^{-1} x_0$ *and* $L_o(x_0) = \frac{1}{2}x_0^T Q x_0$, *where* $P = \int_0^\infty e^{At} BB^T e^{A^T t} dt$ *is the controllability Gramian and* $Q = \int_0^\infty e^{A^T t} C^T C e^{At} dt$ *is the observability Gramian. Furthermore, P and Q are symmetric and positive definite, and are unique solutions of the Lyapunov equations*

$$AP + PA^T = -BB^T \tag{4.8}$$

and

$$A^T Q + QA = -C^T C, \tag{4.9}$$

respectively.

From the form of the Gramians in this theorem, it follows immediately that for any $x_1, x_2 \in \mathbb{R}^n$,

$$\begin{aligned} < x_1, \mathcal{C}\mathcal{C}^* x_2 > &= x_1^T \int_0^\infty e^{At} BB^T e^{A^T t} dt \; x_2, \\ &= x_1^T P x_2 \end{aligned} \tag{4.10}$$

$$\begin{aligned} < x_1, \mathcal{O}^* \mathcal{O} x_2 > &= x_1^T \int_0^\infty e^{A^T t} C^T C e^{At} dt \; x_2, \\ &= x_1^T Q x_2 \end{aligned} \tag{4.11}$$

and the relation with the energy functions is given as

$$L_c(x) = \frac{1}{2}x^T P^{-1} x = \frac{1}{2}\langle x, (\mathcal{C}\mathcal{C}^*)^{-1} x \rangle, \tag{4.12}$$

$$L_o(x) = \frac{1}{2}x^T Q x = \frac{1}{2}\langle x, (\mathcal{O}^*\mathcal{O}) x \rangle. \tag{4.13}$$

The following (balancing) theorem is originally due to [32].

Theorem 4.2: [32]

The eigenvalues of QP are similarity invariants, that is, they do not depend on the choice of the sate-space coordinates. There exists a state-space representation where

$$\Sigma := Q = P = \begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_n \end{pmatrix}, \tag{4.14}$$

with $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n > 0$ *the square roots of the eigenvalues of QP. Such representations are called* balanced, *and the system is in* balanced form. *Furthermore, the* σ_i*'s,* $i = 1, \ldots, n$*, equal the Hankel singular values, that is, the singular values of the Hankel operator (Equation 4.2).*

Two other representations that may be obtained from Equation 4.14 by coordinate transformations $x = \Sigma^{-\frac{1}{2}}\tilde{x}$ and $x = \Sigma^{\frac{1}{2}}\tilde{x}$, respectively, follow easily from the above theorem.

Definition 4.2: [32]

A state-space representation is an input-normal/output-diagonal *representation if* $P = I$ *and* $Q = \Sigma^2$*, where* Σ *is given by Equation 4.14. Furthermore, it is an output-normal/input-diagonal representation if* $P = \Sigma^2$ *and* $Q = I$*.*

The largest Hankel singular value is equal to the Hankel norm of the system, that is,

$$\| G \|_H^2 = \max_{\substack{x \in \mathbb{R}^n \\ x \neq 0}} \frac{L_o(x)}{L_c(x)} = \max_{\substack{x \in \mathbb{R}^n \\ x \neq 0}} \frac{x^T Q x}{x^T P^{-1} x} = \max_{\substack{x \in \mathbb{R}^n \\ x \neq 0}} \frac{\tilde{x}^T \Sigma^2 \tilde{x}}{\tilde{x}^T \tilde{x}} = \sigma_1^2, \tag{4.15}$$

where $G = C(sI - A)^{-1}B$ is the transfer matrix of the system. This gives a characterization of the largest Hankel singular value. The other Hankel singular values may be characterized inductively in a similar way; we refer to [6,15].

So far, we have assumed the state-space representation to be minimal. However, if we consider non-minimal state-space realizations, that is, the system is not controllable and/or observable, then we obtain $\sigma_i's$ that are zero, corresponding to the noncontrollable or nonobservable part of the system, and thus, to the nonminimal part of the system. Related to this observation, we have that the minimal realization of a linear input–output system has a dimension n that is equal to the Hankel rank, or in other words, it equals the rank of the Hankel matrix. A well-known result related to the latter is the following theorem for example, [63].

Theorem 4.3:

If (A, B, C) *is asymptotically stable, then the realization is minimal if and only if* $P > 0$ *and* $Q > 0$*.*

4.2.3 Model Reduction

Once the state-space system is in balanced form, an order reduction procedure based on this form may be applied. Thus, in order to proceed, we assume that the system (Equation 4.5) is in balanced form. Then

the controllability and observability function are $\bar{L}_c(\bar{x}_0) = \frac{1}{2}\bar{x}_0^T \Sigma^{-1}\bar{x}_0$ and $\bar{L}_c(\bar{x}_0) = \frac{1}{2}\bar{x}_0^T \Sigma \bar{x}_0$, respectively. For small σ_i, the amount of control energy required to reach the state $\tilde{x} = (0, \ldots, 0, x_i, 0, \ldots, 0)$ is large while the output energy generated by this state $\tilde{x}$ is small. Hence, if $\sigma_k \gg \sigma_{k+1}$, the state components x_{k+1} to x_n are far less important from this energy point of view and may be removed to reduce the number of state components of the model. We partition the system (Equation 4.5) in a corresponding way as follows:

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \quad B = \begin{pmatrix} B_1 \\ B_2 \end{pmatrix}, \quad C = \begin{pmatrix} C_1 & C_2 \end{pmatrix}, \tag{4.16}$$

$$x^1 = (x_1, \ldots, x_k)^T, \quad x^2 = (x_{k+1}, \ldots, x_n)^T, \quad \Sigma = \begin{pmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{pmatrix},$$

where $\Sigma_1 = \text{diag}(\sigma_1, \ldots, \sigma_k)$ and $\Sigma_2 = \text{diag}(\sigma_{k+1}, \ldots, \sigma_n)$.

Theorem 4.4:

Both subsystems (A_{ii}, B_i, C_i), $i = 1, 2$, are again in balanced form, and their controllability and observability Gramians are equal to Σ_i, $i = 1, 2$.

The following result has been proved by [43].

Theorem 4.5:

Assume that $\sigma_k > \sigma_{k+1}$. Then both subsystems (A_{ii}, B_i, C_i), $i = 1, 2$, are asymptotically stable.

The subsystem (A_{11}, B_1, C_1) may be used as an approximation of the full order system (Equation 4.5). The optimality of this approximation in the Hankel and $\mathcal{H}_\infty$-norm has been studied by [17], and an upper bound for the error is given. The $\mathcal{H}_\infty$-norm of $G(s) = C(sI - A)^{-1}B$ is defined as

$$\|G\|_\infty = \sup_{\omega \in \mathbb{R}} \lambda_{\max}^{\frac{1}{2}}(G(-j\omega)^T G(j\omega)),$$

where $\lambda_{\max}^{\frac{1}{2}}(G(-j\omega)^T G(j\omega))$ is the square root of the maximum eigenvalue of $G(-j\omega)^T G(j\omega)$. Denote the transfer matrix of the reduced order system (A_{11}, B_1, C_1) by $\tilde{G}(s) = C_1(sI - A_{11})^{-1}B_1$. The following error bound was originally proved in [17]. A number of proofs can be found in [63].

Theorem 4.6: [17]

$$\|G - \tilde{G}\|_H \leq \|G - \tilde{G}\|_\infty \leq 2(\sigma_{k+1} + \cdots + \sigma_n).$$

Hence, if we remove the state components $x_{k+1}, \ldots, x_n$ that correspond to small Hankel singular values $\sigma_{k+1}, \ldots, \sigma_n$ (small compared to the rest of the singular values, that is, $\sigma_k \gg \sigma_{k+1}$), then the error is small, and the reduced order system (A_{11}, B_1, C_1) constitutes a good approximation in terms of the Hankel norm to the full order system.

The model reduction method that we gave above consists of simply truncating the model. It is also possible to reduce the model in a different way see, for example, [8,21]. Instead of setting

$x^2 = (x_{k+1}, \ldots, x_n) = 0$ we approximate the system by setting $\dot{x}^2 = 0$ (thus interpreting x^2 as a very fast stable state, which may be approximated by a constant function of x^1 and u). The resulting algebraic equation can be solved for x^2 as (note that A_{22}^{-1} exists by Theorem 4.5)

$$x^2 = -A_{22}^{-1}\left(A_{21}x^1 + B_2 u\right).$$

Substitution in Equation 4.5 leads to a reduced order model $(\hat{A}, \hat{B}, \hat{C})$ defined as

$$\begin{aligned}\hat{A} &:= A_{11} - A_{12}A_{22}^{-1}A_{21},\\ \hat{B} &:= B_1 - A_{12}A_{22}^{-1}B_2,\\ \hat{C} &:= C_1 - C_2A_{22}^{-1}A_{21},\end{aligned}$$

The system $(\hat{A}, \hat{B}, \hat{C})$ also gives an approximation to the full order system (Equation 4.5). Theorems 4.4 through 4.6 also hold if we replace the system (A_{11}, B_1, C_1) by $(\hat{A}, \hat{B}, \hat{C})$.

4.2.4 Unstable Systems, Closed-Loop Balancing

4.2.4.1 Linear Quadratic Gaussian Balancing

A major drawback of the original balancing method as described in Section 4.2.2 is that it only applies to stable systems. Furthermore, the method emphasizes the (open-loop) input–output characteristics of the system, while it is *a priori* not clear if it yields good approximations in closed-loop configurations. In this section, we treat (linear quadratic Gaussian) LQG balancing for linear systems, which was introduced by [25,26] (see also [57]). In [41], this concept is further developed. LQG balancing was introduced with the aim of finding a model reduction method for a system (not necessarily stable) together with its corresponding LQG compensator. LQG balancing has been treated from another point of view in [61]. First, we give a review of the formulation of [26,41].

LQG compensation is formulated for a minimal state-space system

$$\begin{aligned}\dot{x} &= Ax + Bu + Bd,\\ y &= Cx + v,\end{aligned} \tag{4.17}$$

where $u \in \mathbb{R}^m$, $x \in \mathbb{R}^n$, $y \in \mathbb{R}^p$, and d and v are independent Gaussian white-noise processes with covariance functions $I\delta(t-\tau)$. The criterion

$$J(x_0, u(\cdot)) = E \lim_{T\to\infty} \frac{1}{T} \int_0^T [x^T(t)C^TCx(t) + u^T(t)u(t)]\, dt \tag{4.18}$$

is required to be minimized. The resulting optimal compensator is given by

$$\begin{aligned}\dot{z} &= Az + Bu + SC^T(y - Cz),\\ u &= -B^TPz.\end{aligned} \tag{4.19}$$

Here S is the stabilizing solution (i.e., $\sigma(A - SC^TC) \subset \mathbb{C}^-$) to the filter algebraic Riccati equation (FARE)

$$AS + SA^T + BB^T - SC^TCS = 0, \tag{4.20}$$

and P is the stabilizing solution (i.e., $\sigma(A - BB^TP) \subset \mathbb{C}^-$) to the control algebraic Riccati equation (CARE)

$$A^TP + PA + C^TC - PBB^TP = 0. \tag{4.21}$$

Theorem 4.7: [26,41]

The eigenvalues of PS are similarity invariants and there exists a state-space representation where

$$M := P = S = \begin{pmatrix} \mu_1 & & 0 \\ & \ddots & \\ 0 & & \mu_n \end{pmatrix}, \tag{4.22}$$

with $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_n > 0$. *This is called an* LQG balanced representation *or* LQG balanced form.

Jonckheere and Silverman [26] and Opdenacker and Jonckheere [41] argue that if $\mu_k \gg \mu_{k+1}$ then the state components x_1 up to x_k are more difficult both to control and to filter than x_{k+1} up to x_n and a synthesis based only on the state components x_1 up to x_k probably retains the essential properties of the system in a closed-loop configuration. Corresponding to the partitioning of the state in the first k components and the last $n-k$ components, the partitioning of the matrices is done as in (4.16), and the reduced order system is

$$\begin{aligned} \dot{x} &= A_{11}x + B_1 u + B_1 d, \\ y &= C_1 x + v. \end{aligned} \tag{4.23}$$

Theorem 4.8: [26,41]

Assume $\mu_k > \mu_{k+1}$. *Then* (A_{11}, B_1, C_1) *is minimal, the reduced order system (Equation 4.23) is again LQG balanced and the optimal compensator for system (Equation 4.23) is the reduced order optimal compensator of the full order system (Equation 4.17).*

As explained in Section 4.2.2, the original idea of balancing stable linear systems, as introduced by [32], considers the Hankel singular values σ_i, $i = 1, \ldots, n$, which are a measure for the importance of a state component in a balanced representation. This balancing technique is based on the input energy which is necessary to reach this state component and the output energy which is generated by this state component. A similar kind of reasoning, using a different pair of energy functions, may be used to achieve the similarity invariants μ_i, $i = 1, \ldots, n$, as above; see [61]. To follow this reasoning, we consider the minimal system (Equation 4.17) *without noise*, that is,

$$\begin{aligned} \dot{x} &= Ax + Bu, \\ y &= Cx, \end{aligned} \tag{4.24}$$

where $u \in \mathbb{R}^m$, $x \in \mathbb{R}^n$, and $y \in \mathbb{R}^p$. We define the energy functions

$$K^-(x_0) := \min_{\substack{u \in L_2(-\infty,0) \\ x(-\infty)=0, x(0)=x_0}} \frac{1}{2} \int_{-\infty}^{0} (\| y(t) \|^2 + \| u(t) \|^2)\, dt,$$

$$K^+(x_0) := \min_{\substack{u \in L_2(0,\infty) \\ x(\infty)=0, x(0)=x_0}} \frac{1}{2} \int_{0}^{\infty} (\| y(t) \|^2 + \| u(t) \|^2)\, dt.$$

$K^-(x_0)$ is called the *past energy* and $K^+(x_0)$ the *future energy* of the system in the state x_0.

Theorem 4.9: [61]

$K^-(x_0) = \frac{1}{2}x_0^T S^{-1}x_0$ and $K^+(x_0) = \frac{1}{2}x_0^T P x_0$, where S and P are the stabilizing solutions of the FARE and the CARE, Equations 4.20 and 4.21, respectively.

For the LQG balanced representation of Theorem 4.7, the past and future energy function are $K^-(x_0) = \frac{1}{2}x_0^T M^{-1}x_0$ and $K^+(x_0) = \frac{1}{2}x_0^T M x_0$, respectively, where M is diagonal. The importance of the state $\tilde{x} = (0, \ldots, 0, x_i, 0, \ldots, 0)$ in terms of past and future energy may be measured by the similarity invariant μ_i. For large μ_i the influence of the state $\tilde{x}$ on the future energy is large while the influence on the past energy is small. Hence, if $\mu_k \gg \mu_{k+1}$, the state components x_{k+1} to x_n are "not important" from this energy point of view and may be removed to reduce the number of state components of the model.

4.2.4.2 Balancing of the Normalized Coprime Representation

In [29,39], balancing of the normalized coprime representation of a linear system is treated. Balancing of the normalized coprime representation was introduced with the aim of finding a model reduction method for unstable linear systems. In [29], balancing of the normalized *right* coprime factorization is treated, while in [39], balancing of the normalized *left* coprime factorization is treated. Here we provide a brief review on this subject.

We consider the system (Equation 4.24) and its transfer function $G(s) = C(sI - A)^{-1}B$. Furthermore, we consider the stabilizing solution P to the CARE (Equation 4.21) and the stabilizing solution S to the FARE (Equation 4.20), leading to the stable matrices $\hat{A} := A - BB^T P$ and $\tilde{A} := A - SC^T C$. First we treat normalized right coprime factorizations and then normalized left coprime factorizations.

We may write any transfer matrix $G(s) = C(sI - A)^{-1}B$ as a right fraction $G(s) = N(s)D(s)^{-1}$ of stable transfer matrices $N(s)$ and $D(s)$. If we choose (e.g., [36]),

$$N(s) := C(sI - \hat{A})^{-1}B,$$
$$D(s) := I - B^T P(sI - \hat{A})^{-1}B,$$

then the *factorization* is *right coprime*, that is, $N(s)$ and $D(s)$ have no common zeros at the same place in the closed right half plane. A state-space realization of the transfer matrix (the so-called graph operator)

$$\begin{pmatrix} N(s) \\ D(s) \end{pmatrix}$$

is

$$\begin{aligned} \dot{x} &= (A - BB^T P)x + Bw, \\ \begin{pmatrix} y \\ u \end{pmatrix} &= \begin{pmatrix} C \\ -B^T P \end{pmatrix} x + \begin{pmatrix} 0 \\ I \end{pmatrix} w, \end{aligned} \tag{4.25}$$

with w a (fictitious) input variable. Furthermore, we are able to find stable transfer matrices $U(s)$ and $V(s)$, such that the Bezout identity

$$U(s)N(s) + V(s)D(s) = I. \tag{4.26}$$

is fulfilled. Indeed, take $U(s) = B^T P(sI - \tilde{A})^{-1}SC^T$ and $V(s) = I + B^T P(sI - \tilde{A})^{-1}B$ (see, e.g., [36,60]). The fact that we are able to find a stable left inverse of the graph operator, that is, we can find the solutions $U(s)$ and $V(s)$ to the Bezout identity (Equation 4.26), is equivalent to the factorization being right coprime.

Furthermore, the graph operator is inner, that is,

$$\| \begin{pmatrix} N \\ D \end{pmatrix} w \|_2 = \|w\|_2$$

or

$$N(-s)^T N(s) + D(-s)^T D(s) = I.$$

Therefore, the factorization is called *normalized*. It is easily checked that the observability Gramian of the system (Equation 4.25) is P. Denote its controllability Gramian by R.

In a similar way, we may write the transfer matrix $G(s)$ as a left fraction $G(s) = \tilde{D}(s)^{-1}\tilde{N}(s)$ of stable transfer matrices $\tilde{D}(s)$ and $\tilde{N}(s)$. If we choose (e.g., [36])

$$\begin{aligned} \tilde{N}(s) &:= C(sI - \tilde{A})^{-1}B, \\ \tilde{D}(s) &= C(sI - \tilde{A})^{-1}SC^T - I, \end{aligned}$$

then this is a *left factorization*. Obviously, $\hat{y}(s) = G(s)\hat{u}(s)$ is equivalent with $0 = \tilde{N}(s)\hat{u}(s) - \tilde{D}(s)\hat{y}(s)$. Moreover, a state-space realization of the transfer matrix

$$\begin{pmatrix} \tilde{N}(s) & \tilde{D}(s) \end{pmatrix}$$

is

$$\begin{aligned} \dot{x} &= (A - SC^T C)x + \begin{pmatrix} B & SC^T \end{pmatrix} \tilde{w}, \\ z &= Cx + \begin{pmatrix} 0 & -I \end{pmatrix} \tilde{w}. \end{aligned} \tag{4.27}$$

If we take

$$\tilde{w} = \begin{pmatrix} u \\ y \end{pmatrix}$$

as the input variable, then the dynamics resulting from setting $z = 0$ in Equation 4.27 is a state-space representation of $G(s)$. We are able to find stable transfer matrices such that the Bezout Identity is fulfilled, that is, there exist stable transfer matrices $\tilde{U}(s)$ and $\tilde{V}(s)$, such that

$$\tilde{N}(s)\tilde{U}(s) + \tilde{D}(s)\tilde{V}(s) = I. \tag{4.28}$$

Indeed, we may take $\tilde{U}(s) = B^T P(sI - \hat{A})^{-1}SC^T$ and $\tilde{V}(s) = I + C(sI - \hat{A})^{-1}SC^T$ (see, e.g., Vid,Ne). This proves that the factorization is *left coprime*. Furthermore $\begin{pmatrix} \tilde{N}(s) & \tilde{D}(s) \end{pmatrix}$ is co-inner, that is,

$$\tilde{N}(s)\tilde{N}(-s)^T + \tilde{D}(s)\tilde{D}(-s)^T = I,$$

which means that the factorization is *normalized*. Hence $\begin{pmatrix} \tilde{N}(s) & \tilde{D}(s) \end{pmatrix}$ represents the normalized left coprime factorization of system (Equation 4.24). The system (Equation 4.27) has as controllability Gramian the positive-definite matrix S and we denote its observability Gramian by the matrix Q. Note that the right factorization

$$\begin{pmatrix} N(s) \\ D(s) \end{pmatrix}$$

can be seen as an *image* representation of $G(s)$, while the left factorization

$$\begin{pmatrix} \tilde{N}(s) & \tilde{D}(s) \end{pmatrix}$$

can be regarded as a *kernel* representation of $G(s)$.

The following result follows rather straightforwardly.

Theorem 4.10:

The Hankel singular values of the right and left factorization (Equations 4.25 and 4.27), respectively, are the same.

Proof. It follows from the Lyapunov Equations 4.8 and 4.9 for the systems (Equations 4.25 and 4.27), that $R = (I + SP)^{-1}S$ and $Q = (I + PS)^{-1}P$. Now, it is easily obtained that PR and SQ have the same eigenvalues.

The Hankel singular values of Equation 4.25 (and, hence, of Equation 4.27) are called the *graph Hankel singular values* of the original system (Equation 4.24). These singular values have the following property:

Theorem 4.11: [29,39]

The graph Hankel singular values of system (Equation 4.24) are strictly less than one.

Denote the graph Hankel singular values by τ_i, $i = 1, \ldots, n$, and assume $\tau_1 \geq \cdots \geq \tau_n$. The relation between τ_i, $i = 1, \ldots, n$, and the similarity invariants μ_i, $i = 1, \ldots, n$, of Theorem 4.7 is given by the following theorem:

Theorem 4.12: [39,61]

$$\mu_i = \tau_i (1 - \tau_i^2)^{-\frac{1}{2}} \quad \text{for } i = 1, \ldots, n.$$

This implies that the reduced model that is obtained by model reduction based on balancing the (left or right) normalized coprime factorization is the same as the reduced model that is obtained by model reduction based on LQG balancing. Consider the normalized right coprime representation (Equation 4.25) and assume that it is in balanced form with

$$\Lambda := P = R = \begin{pmatrix} \tau_1 & & 0 \\ & \ddots & \\ 0 & & \tau_n \end{pmatrix}.$$

Furthermore, assume that $\tau_k > \tau_{k+1}$ and define correspondingly $\Lambda =: \operatorname{diag}\{\Lambda_1, \Lambda_2\}$. It follows, [29], that reducing the order of Equation 4.25 by truncating the system to the first k state components (the partitioning is done corresponding to Equation 4.16) again gives a normalized right coprime representation.

Theorem 4.13: [29]

The reduced order system of Equation 4.25 is of the form

$$\left(\left(A_{11} - B_1 B_1^T D_1 \right), B_1, \begin{pmatrix} C_1 \\ -B_1^T D_1 \end{pmatrix}, \begin{pmatrix} 0 \\ I \end{pmatrix} \right),$$

with controllability and observability Gramian Λ_1. This system is the normalized right coprime representation of the system (A_{11}, B_1, C_1), which is minimal.

4.2.4.3 Extensions to Other Types of Balancing

The methods described above can be put in a more general setting, and extended to the H_∞ case. $H\infty$ balancing for linear systems is introduced in [33–35]. For details, we refer to the latter references. In the H_∞ case, balancing is performed on Q_γ^- and Q_γ^+, [46], which are defined as

$$Q_\gamma^-(x_0) = \min_{u \in L_2(-\infty,0)} \frac{1}{2} \int_{-\infty}^{0} (1-\gamma^{-2})||y(t)||^2 + ||u(t)||^2 \, dt, \quad x(-\infty) = 0, \; x(0) = x_0, \forall \gamma$$

and

$$Q_\gamma^+(x_0) = \min_{u \in L_2(0,\infty)} \frac{1}{2} \int_{0}^{\infty} ||y(t)||^2 + \frac{1}{1-\gamma^{-2}} ||u(t)||^2 \, dt, \quad x(\infty) = 0, \; x(0) = x_0, \quad \text{for } \gamma > 1,$$

while if $\gamma < 1$, then

$$Q_\gamma^+(x_0) = \max_{u \in L_2(0,\infty)} \frac{1}{2} \int_{0}^{\infty} ||y(t)||^2 + \frac{1}{1-\gamma^{-2}} ||u(t)||^2 \, dt, \quad x(\infty) = 0, \; x(0) = x_0.$$

There is an immediate relation with the solutions to the H_∞ Filter and Control Algebraic Riccati equations, see, for example, [46].

Positive real and bounded real balancing can be done by considering dissipativity with respect to a quadratic supply rate that depends on the input and the output of the system:

$$s(u,y) = \frac{1}{2}[u^T \; y^T] J \begin{bmatrix} u \\ y \end{bmatrix}, \tag{4.29}$$

with $u \in \mathbb{R}^m$, $y \in \mathbb{R}^p$, and $J \in \mathbb{R}^{(m+p)\times(m+p)}$, such that $J = J^T$; see, for example, [23,52]. The bounded real balancing case , that is, $J = \begin{bmatrix} I & 0 \\ 0 & -I \end{bmatrix}$ is treated in [1,22,39,42]. The positive real balancing case, that is balancing of strictly passive, asymptotically stable, minimal systems, that is $J = \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix}$, is treated in [1,7,19]. For all these methods it holds that truncation preserves the original balanced structure, and thus truncation based on bounded real or positive real balancing preserves bounded and positive realness, respectively. Furthermore, error bounds for model reduction based on several of the above mentioned methods are available; see, for example, [52].

Additionally, model reduction methods based on balancing methods that preserve some internal structure of the system currently receive a lot of interest due to applications that require this, such as electrical circuit simulators. For some results on Hamiltonian structure preservation, and second-order system structure preservation; see, for example, [30,54,55].

4.3 Balancing for Nonlinear Systems

Balanced realizations and the related model order reduction technique rely on singular-value analysis. The analysis is important since it extracts the gain structure of the operator, that is, it characterizes the largest input-output ratio and the corresponding input [51]. Since linear singular values are defined as eigenvalues of the composition of the given operator and its adjoint, it is natural to introduce a nonlinear version of adjoint operators to obtain a nonlinear counterpart of a singular value. There has been done quite some research on the nonlinear extension of adjoint operators, for example, [3,10,48], and the references therein. Here we do not explicitly use these definitions of nonlinear adjoint operators. We rely on a characterization of singular values for nonlinear operators based on the gain structure as studied in [9]. The balanced realization based on this analysis yields a realization that is based on the singular

values of the corresponding Hankel operator, and results in a method which can be viewed as a complete extension of the linear methods, both from an input–output and a state-space point of view [12].

The related model order reduction technique, nonlinear balanced truncation, preserves several important properties of the original system and the corresponding input–output operator, such as stability, controllability, observability, and the gain structure [11,14].

This section gives a very brief overview of the series of research on balanced realization and the related model order reduction method based on nonlinear singular-value analysis. We refer to [13] for more details.

4.3.1 Basics of Nonlinear Balanced Realizations

This section gives a nonlinear extension of balanced realization introduced in the Section 4.2. Let us consider the following asymptotically stable input-affine nonlinear system

$$\Sigma : \begin{cases} \dot{x} = f(x) + g(x)u \quad x(0) = x^0, \\ y = h(x), \end{cases} \tag{4.30}$$

where $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^m$, and $y(t) \in \mathbb{R}^p$. The controllability operator $\mathcal{C} : U \to X$ with $X = \mathbb{R}^n$ and $U = L_2^m[0, \infty)$, and the observability operator $\mathcal{O} : X \to Y$ with $Y = L_2^p[0, \infty)$ for this system are defined by

$$\mathcal{C} : u \mapsto x^0 : \begin{cases} \dot{x} = -f(x) - g(x)u \quad x(\infty) = 0 \\ x^0 = x(0) \end{cases}$$

$$\mathcal{O} : x^0 \mapsto y : \begin{cases} \dot{x} = f(x) \quad x(0) = x^0 \\ y = h(x) \end{cases}$$

This definition implies that the observability operator $\mathcal{O}$ is a map from the initial condition $x(0) = x^0$ to the output L_2 signal when no input is applied. To interpret the meaning of $\mathcal{C}$, let us consider a time-reversal behavior of the $\mathcal{C}$ operator as

$$\mathcal{C} : u \mapsto x^0 : \begin{cases} \dot{x} = f(x) + g(x)u(-t) \quad x(-\infty) = 0 \\ x^0 = x(0) \end{cases}$$

Then the controllability operator $\mathcal{C}$ can be regarded as a mapping from the input L_2 signal to the terminal state $x(0) = x^0$ when the initial state is $x(-\infty) = 0$. Therefore, as in the linear case, $\mathcal{C}$ and $\mathcal{O}$ represent the input-to-state behavior and the state-to-output behavior, respectively, and the Hankel operator for the nonlinear system Σ in Equation 4.30 is given by the composition of $\mathcal{C}$ and $\mathcal{O}$

$$\mathcal{H} := \mathcal{O} \circ \mathcal{C}. \tag{4.31}$$

To relate the Hankel, controllability and observability operator with the observability and controllability functions defined in Equations 4.6 and 4.7, we first introduce a norm-minimizing inverse $\mathcal{C}^\dagger : X \to U$ of $\mathcal{C}$.

$$\mathcal{C}^\dagger : x^0 \mapsto u := \arg \min_{\mathcal{C}(u)=x^0} \|u\|$$

The operators $\mathcal{C}^\dagger$ and $\mathcal{O}$ yield the definitions of the controllability function $L_c(x)$ and the observability function $L_o(x)$ that are generalizations of the controllability and observability Gramians, respectively

that is,

$$L_c(x^0) := \frac{1}{2} \| C^\dagger(x^0) \|^2 = \min_{\substack{u \in L_2(-\infty,0] \\ x(-\infty)=0, x(0)=x_0}} \frac{1}{2} \int_{-\infty}^{0} \| u(t) \|^2 \, dt, \tag{4.32}$$

$$L_o(x^0) := \frac{1}{2} \| \mathcal{O}(x^0) \|^2 = \frac{1}{2} \int_0^\infty \| y(t) \|^2 \, dt, \quad x(0) = x^0, \quad u(t) \equiv 0, \quad 0 \le t < \infty. \tag{4.33}$$

For linear systems, the relation with the Gramians is given in Theorem 4.1. Here the inverse of P appears in the equation for L_c because $C^\dagger$ appears in the definition (Equation 4.32), whereas C can be used in the linear case. In order to obtain functions $L_c(x)$ and $L_o(x)$, we need to solve a Hamilton–Jacobi equation and a Lyapunov equation.

Theorem 4.14: [45]

Consider the system of Equation 4.30. Suppose that 0 *is an asymptotically stable equilibrium point and that a smooth observability function* $L_o(x)$ *exists. Then* $L_o(x)$ *is the unique smooth solution of*

$$\frac{\partial L_o(x)}{\partial x} f(x) + \frac{1}{2} h(x)^{\mathrm{T}} h(x) = 0,$$

with $L_o(0) = 0$. *Furthermore, assume that a smooth controllability function* $L_c(x)$ *exists. Then* $L_c(x)$ *is the unique smooth solution of*

$$\frac{\partial L_c(x)}{\partial x} f(x) + \frac{1}{2} \frac{\partial L_c(x)}{\partial x} g(x) g(x)^{\mathrm{T}} \frac{\partial L_c(x)}{\partial x}^{\mathrm{T}} = 0,$$

with $L_c(0) = 0$ *such that* 0 *is an asymptotically stable equilibrium point of* $\dot{x} = -f(x) - g(x)g(x)^{\mathrm{T}} (\partial L_c(x)/\partial x)^{\mathrm{T}}$.

Similar to the linear case, the positive definiteness of the controllability and observability functions implies strong reachability and zero-state observability of the system Σ in Equation 4.30, respectively. Combining these two properties, we can obtain the following result on the minimality of the system.

Theorem 4.15: [47]

Consider the system of Equation 4.30, and assume it is analytic. Suppose that

$$0 < L_c(x) < \infty,$$
$$0 < L_o(x) < \infty,$$

hold for all $x \neq 0$. *Then the system is a minimal realization as defined in and under the conditions from, [24].*

$L_c(x)$ and $L_o(x)$ can be used to "measure the minimality" of a nonlinear dynamical system. Furthermore, a basis for nonlinear balanced realizations is obtained as a nonlinear generalization of Definition 4.2 in the linear case. For that, a factorization of $L_o(x)$ into a semiquadratic form needs to be done, that is, in a convex neighborhood of the equilibrium point 0, we can write

$$L_o(x) = \frac{1}{2} x^{\mathrm{T}} M(x) x, \quad \text{with } M(0) = \frac{\partial^2 L_o}{\partial x^2}(0).$$

Now, an input-normal/output-diagonal form can be obtained.

Theorem 4.16: [45]

Consider the system of Equation 4.30 on a neighborhood W of 0. Suppose that 0 is an asymptotically stable equilibrium point, that it is zero-state observable, that smooth controllability and observability functions $L_c(x)$ and $L_o(x)$ exist on W, and that $(\partial^2 L_c/\partial x^2)(0) > 0$ and $(\partial^2 L_o/\partial x^2)(0) > 0$ hold. Furthermore, assume that the number of distinct eigenvalues of $M(x)$ is constant on W. Then there exists coordinates such that the controllability and observability functions $L_c(x)$ and $L_o(x)$ satisfy

$$L_c(x) = \frac{1}{2}\sum_{i=1}^{n} x_i^2, \tag{4.34}$$

$$L_o(x) = \frac{1}{2}\sum_{i=1}^{n} x_i^2 \tau_i(x), \tag{4.35}$$

where $\tau_1(x) \geq \tau_2(x) \geq \cdots \geq \tau_n(x)$.

A state-space realization satisfying the conditions (Equations 4.34 and 4.35) is called an *input-normal form*, and the functions $\tau_i(x)$, $i = 1, 2, \ldots, n$ are called singular-value functions. We refer to [45] for the construction of the coordinate transformation that brings the system in the form of Theorem 4.16. If a singular value function $\tau_i(x)$ is larger than $\tau_j(x)$, then the coordinate axis x_i plays a more important role than the coordinate axis x_j does. Thus this realization is similar to the linear input-normal/output-diagonal realization of Definition 4.2, and it directly yields a tool for model order reduction of a nonlinear systems. However, a drawback of the above realization is that the the singular-value functions $\tau_i(x)$'s and consequently, the corresponding realization are not unique, for example, [18]. For example, if the observability function is given by

$$L_o(x) = \frac{1}{2}(x_1^2\tau_1(x) + x_2^2\tau_2(x)) = \frac{1}{2}(2x_1^2 + x_2^2 + x_1^2x_2^2),$$

with the state-space $x = (x_1, x_2)$, then the corresponding singular-value functions are

$$\tau_1(x) = 2 + kx_2^2,$$
$$\tau_2(x) = 1 + (1-k)x_1^2,$$

with an arbitrary scalar constant k. This example reveals that the singular value function are not uniquely determined by this characterization. To overcome these problems, balanced realizations based on nonlinear singular value analysis is presented in the following section.

4.3.2 Balanced Realizations Based on Singular-Value Analysis of Hankel Operators

In this section, application of singular-value analysis to nonlinear Hankel operators determines a balanced realization with a direct input–output interpretation, whereas the balanced realization of Theorem 4.16 is completely determined based on state-space considerations only. To this end, we consider the Hankel operator $\mathcal{H} : U \to Y$ as defined in Equation 4.31 with $U = L_2^m[0, \infty)$ and $Y = L_2^p[0, \infty)$. Then a singular-value analysis based on the differential form, [12], is given by

$$(\mathrm{d}\mathcal{H}(v))^* \, \mathcal{H}(v) = \lambda \, v, \quad \lambda \in \mathbb{R}, \quad v \in U, \tag{4.36}$$

with λ and v the eigenvalues and corresponding eigenvectors, respectively. Since we consider a singular-value analysis problem on L_2 spaces, we need to find state trajectories of certain Hamiltonian dynamics;

see, for example, [12]. In the linear case, we only need to solve an eigenvalue problem on a finite dimensional space $X = \mathbb{R}^n$ to obtain the singular values and singular vectors of the Hankel operator $\mathcal{H}$. Here we provide the nonlinear counterpart as follows.

Theorem 4.17: [12]

Consider the Hankel operator defined by Equation 4.31. Suppose that the operators $\mathcal{C}^\dagger$ and $\mathcal{O}$ exist and are smooth. Moreover, suppose that $\lambda \in \mathbb{R}$ and $\xi \in X$ satisfy the following equation:

$$\frac{\partial L_o(\xi)}{\partial \xi} = \lambda \frac{\partial L_c(\xi)}{\partial \xi}, \quad \lambda \in \mathbb{R}, \quad \xi \in X. \tag{4.37}$$

Then λ is an eigenvalue of $(d\mathcal{H}(u)^*\mathcal{H}(u)$, and

$$v := \mathcal{C}^\dagger(\xi). \tag{4.38}$$

That is, v defined above is a singular vector of $\mathcal{H}$.

Although the singular-value analysis problem, [13], is a nonlinear problem on an infinite dimensional signal space $U = L_2^m[0, \infty)$, the problem to be solved in the above theorem is a nonlinear algebraic equation on a finite dimensional space $X = \mathbb{R}^n$ which is also related to a nonlinear eigenvalue problem on X; see [9].

In the linear case, Equation 4.37 reduces to

$$\xi^{\mathrm{T}} Q = \lambda \xi^{\mathrm{T}} P^{-1}$$

where P and Q are the controllability and observability Gramians, and λ and ξ are an eigenvalue and an eigenvector of PQ. Furthermore, Equation 4.38 characterizes the relationship between a singular vector v of $\mathcal{H}$ and an eigenvector ξ of PQ. Also, in the linear case, there always exist n independent pairs of eigenvalues and eigenvectors of PQ. What happens in the nonlinear case? The answer is provided in the following theorem.

Theorem 4.18: [12]

Consider the system Σ in Equation 4.30 and the Hankel operator $\mathcal{H}$ in Equation 4.31 with $X = \mathbb{R}^n$. Suppose that the Jacobian linearization of the system has n distinct Hankel singular values. Then Equation 4.37 has n independent solution curves $\xi = \xi_i(s)$, $s \in \mathbb{R}$, $i = 1, 2, \ldots, n$ intersecting to each other at the origin and satisfying the condition

$$\|\xi_i(s)\| = |s|.$$

For linear systems, the solutions of Equation 4.37 are lines (orthogonally) intersecting at the origin. The above theorem shows that instead of these lines, in the nonlinear case n independent curves $x = \xi_i(s)$, $i = 1, 2, \ldots, n$ exist. For instance, if the dimension of the state is $n = 2$, the solution of Equation 4.37 is illustrated in Figure 4.2.

We can relate the solutions $\xi_i(s)$ to the singular values of the Hankel operator $\mathcal{H}$. Let $v_i(s)$ and $\sigma_i(s)$ denote the singular vector and the singular-value parameterized by s corresponding to $\xi_i(s)$. Then we have

$$v_i(s) := \mathcal{C}^\dagger(\xi_i(s)),$$

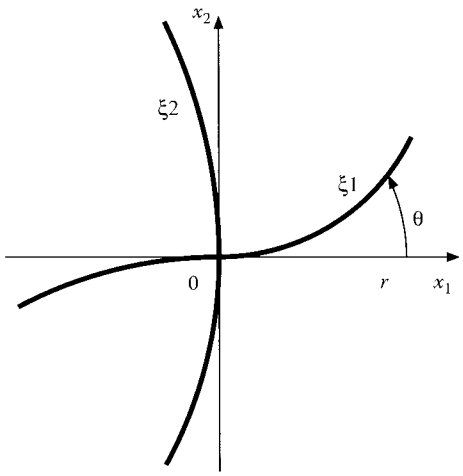

FIGURE 4.2 Configuration of $\xi_1(s)$ and $\xi_2(s)$ in the case $n = 2$.

$$\sigma_i(s) := \frac{\|\mathcal{H}(v_i(s))\|_{L_2}}{\|v_i(s)\|_{L_2}} = \frac{\|\mathcal{O}(\xi_i(s))\|_{L_2}}{\|\mathcal{C}^\dagger(\xi_i(s))\|_{L_2}} = \sqrt{\frac{L_o(\xi_i(s))}{L_c(\xi_i(s))}}.$$

This expression yields an explicit expression of the singular values $\sigma_i(s)$'s of the Hankel operator $\mathcal{H}$. These functions $\sigma_i(s)$'s are called *Hankel singular values*. Without loss of generality we assume that the following equation holds for $i = 1, 2, \ldots, n$ in a neighborhood of the origin

$$\min\{\sigma_i(s), \sigma_i(-s)\} > \max\{\sigma_{i+1}(s), \sigma_{i+1}(-s)\}. \tag{4.39}$$

As in the linear case, the solution curves $\xi_i(s)$ play the role of the coordinate axes of a balanced realization. By applying an isometric coordinate transformation which maps the solution curves $\xi_i(s)$'s into the coordinate axes, we obtain a realization whose (new) coordinate axes x_i are the solution of Equation 4.37, that is,

$$\left.\frac{\partial L_o(x)}{\partial x}\right|_{x=(0,\ldots,0,x_i,0,\ldots,0)} = \lambda \left.\frac{\partial L_c(x)}{\partial x}\right|_{x=(0,\ldots,0,x_i,0,\ldots,0),} \tag{4.40}$$

$$\sigma_i(x_i) = \sqrt{\frac{L_o(0,\ldots,0,x_i,0,\ldots,0)}{L_c(0,\ldots,0,x_i,0,\ldots,0)}}. \tag{4.41}$$

Equation 4.41 implies that the new coordinate axes x_i, $i = 1, \ldots, n$ are the solutions of Equation 4.37 for Hankel singular-value analysis. Therefore, the Hankel norm can be obtained by

$$\|\Sigma\|_H = \sup_{u\neq 0} \frac{\|\mathcal{H}(u)\|_{L_2}}{\|u\|_{L_2}} = \sup_{s\in\mathbb{R}} \max_i \sigma_i(s)$$
$$= \sup_{x_1\in\mathbb{R}} \sqrt{\frac{L_o(x_1,0,\ldots,0)}{L_c(x_1,0,\ldots,0)}},$$

provided the ordering condition (Equation 4.39) holds for all $s \in \mathbb{R}$. Furthermore, apply this coordinate transformation recursively to all lower dimensional subspaces such as $(x_1, x_2, \ldots, x_k, 0, \ldots, 0)$, then we

obtain a state-space realization satisfying Equation 4.41 and

$$x_i = 0 \iff \frac{\partial L_o(x)}{\partial x_i} = 0 \iff \frac{\partial L_c(x)}{\partial x_i} = 0. \tag{4.42}$$

This property is crucial for balanced realization and model order reduction. Using tools from differential topology, for example, [31], we can prove that the obtained realization is diffeomorphic to the following *precise* input-normal/output-diagonal realization.

Theorem 4.19: [11,14]

Consider the system Σ *in Equation 4.30. Suppose that the assumptions in Theorem 4.18 hold. Then there exists coordinates in a neighborhood of the origin such that the system is in input-normal/output-diagonal form satisfying*

$$L_c(x) = \frac{1}{2}\sum_{i=1}^{n} x_i^2,$$

$$L_o(x) = \frac{1}{2}\sum_{i=1}^{n} x_i^2\, \sigma_i(x_i)^2.$$

This realization is more precise than the realization in Theorem 4.16 in the following sense: (a) The solutions of Equation 4.37 coincide with the coordinate axes, that is, Equation 4.40 holds. (b) The ratio of the observability function L_o to the controllability function L_c equals the singular values $\sigma_i(x_i)$'s on the coordinate axes, that is, Equation 4.41 holds. (c) An exact balanced realization can be obtained by a coordinate transformation

$$z_i = \phi_i(x_i) := x_i \sqrt{\sigma_i(x_i)}, \tag{4.43}$$

which is well-defined in a neighborhood of the origin.

Corollary 4.1: [11,14]

The coordinate change (Equation 4.43) transforms the input-normal realization in Theorem 4.19 into the following balanced form:

$$L_c(z) = \frac{1}{2}\sum_{i=1}^{n} \frac{z_i^2}{\sigma_i(z_i)},$$

$$L_o(z) = \frac{1}{2}\sum_{i=1}^{n} z_i^2 \sigma_i(z_i).$$

Since we only use the coordinate transformation (Equation 4.43) preserving the coordinate axes, the realization obtained here also satisfies the properties (a) and (b) explained above. The controllability and observability functions can be written as

$$L_c(z) = \frac{1}{2} z^{\mathrm{T}} \underbrace{\mathrm{diag}(\sigma_1(z_1), \ldots, \sigma_n(z_n))^{-1}}_{P(z)} z,$$

$$L_o(z) = \frac{1}{2} z^{\mathrm{T}} \underbrace{\mathrm{diag}(\sigma_1(z_1), \ldots, \sigma_n(z_n))}_{Q(z)} z.$$

Here $P(z)$ and $Q(z)$ can be regarded as nonlinear counterparts of the balanced controllability and observability Gramians, since

$$P(z) = Q(z) = \mathrm{diag}(\sigma_1(z_1), \sigma_2(z_2), \ldots, \sigma_n(z_n)).$$

The axes of this realization are uniquely determined. We call this state-space realization a *balanced realization* of the nonlinear system Σ in Equation 4.30. As in the linear case, both the relationship between the input-to-state and state-to-output behavior and that among the coordinate axes are balanced.

4.3.3 Model Order Reduction

An important application of balanced realizations is that it is a tool for model order reduction called *balanced truncation*. Here, a model order reduction method preserving the Hankel norm of the original system is proposed. Suppose that the system of Equation 4.30 is balanced in the sense that it satisfies Equations 4.41 and 4.42. Note that the realizations in Theorem 4.19 and Corollary 4.1 satisfy these conditions. Suppose, that

$$\min\{\sigma_k(s), \sigma_k(-s)\} \gg \max\{\sigma_{k+1}(s), \sigma_{k+1}(-s)\}$$

holds with a certain k ($1 \leq k < n$). Divide the state into two vectors $x = (x^a, x^b)$

$$x^a := (x_1, \ldots, x_k) \in \mathbb{R}^k,$$
$$x^b := (x_{k+1}, \ldots, x_n) \in \mathbb{R}^{n-k},$$

and the vector field into two vector fields accordingly

$$f(x) = \begin{pmatrix} f^a(x) \\ f^b(x) \end{pmatrix},$$
$$g(x) = \begin{pmatrix} g^a(x) \\ g^b(x) \end{pmatrix},$$

and truncate the state by substituting $x^b = 0$. Then we obtain a k-dimensional state-space model Σ^a with the state x^a (with a $(n-k)$-dimensional residual model Σ^b with the state x^b).

$$\Sigma^a : \begin{cases} \dot{x}^a = f^a(x^a, 0) + g^a(x^a, 0)u^a \\ y^a = h(x^a, 0) \end{cases} \tag{4.44}$$

$$\Sigma^b : \begin{cases} \dot{x}^b = f^b(0, x^b) + g^b(0, x^b)u^b \\ y^b = h(0, x^b) \end{cases} \tag{4.45}$$

This procedure is called *balanced truncation*. The obtained reduced order models have preserved the following properties.

Theorem 4.20: [13,14]

Suppose that the system Σ satisfies Equations 4.41 and 4.42 and apply the balanced truncation procedure explained above. Then the controllability and observability functions of the reduced order models Σ^a and

Σ^b denoted by L_c^a, L_c^b, L_o^a, and L_o^b, respectively, satisfy the following equations:

$$L_c^a(x^a) = L_c(x^a, 0), \quad L_o^a(x^a) = L_o(x^a, 0),$$
$$L_c^b(x^b) = L_c(0, x^b), \quad L_o^b(x^b) = L_o(0, x^b),$$

which implies

$$\sigma_i^a(x_i^a) = \sigma_i(x_i^a), \quad i = 1, 2, \ldots, k,$$
$$\sigma_i^b(x_i^b) = \sigma_{i+k}(x_i^b), \quad i = 1, 2, \ldots, n-k,$$

with the singular values σ^a's of the system Σ^a and the singular values σ^b of the system Σ^b. In particular, if σ_1 is defined globally, then

$$\|\Sigma^a\|_H = \|\Sigma\|_H.$$

Theorem 4.20 states that the important characteristics of the original system such as represented by the controllability and observability functions and Hankel singular values are preserved. Moreover, by Theorem 4.15, this implies that the controllability, observability, minimality, and the gain property is preserved under the model reduction. These preservation properties hold for truncation of any realization satisfying the conditions (Equations 4.41 and 4.42), such as the realizations in Theorem 4.19 and Corollary 4.1 [13,14]. Furthermore, concerning the stability, (global) Lyapunov stability and local asymptotic stability are preserved with this procedure as well. Note that this theorem is a natural nonlinear counterpart of the linear theory. However, a nonlinear counterpart of the error bound of the reduced order model has not been found yet.

4.3.4 Other Types of Balancing for Nonlinear Systems

As for linear systems, there exist extensions of LQG, and coprime balancing [49], H_∞ or L_2-gain balancing [46], and positive/bounded real and dissipativity-based balancing [23]. In fact, in [23] a direct relation is obtained with Hankel operator analysis for augmented systems.

The presented work treats balanced realizations for nonlinear systems based on balancing in a (possibly large) region around an equilibrium point, where a relation with the Hankel operator, observability and controllability operators and functions, and minimality of the nonlinear system is obtained. A drawback of these methods is the computational effort that is needed to compute the balanced realization. As mentioned in the introduction, other extensions of the linear notion of balancing can be found in for example, [20,28,58,59].

4.4 Concluding Remarks

In this Chapter, balanced realizations for linear and nonlinear systems and model reduction based on these realizations are treated. There exists a vast amount of literature on this topic, and the reference list in this paper is certainly not complete. For example, some of the basics for linear balanced realizations can be found in [40], and we did not pay any attention to the balancing methods treating uncertain, and time- and parameter-varying linear systems; for example, [4,44], behavioral balancing, for example, [53], or the numerical side of balancing, for example, [5].

Recently, a lot of interest is taken in structure preserving order techniques for both linear and nonlinear systems, where the additional structure to be preserved is a physical structure, such as port-Hamiltonian structure, [56], and other physical structures, as mentioned in our linear systems section. For example, for circuit simulators with continuously growing orders of the models, a need for interpreting a reduced order model as a circuit is important, such as, [50]. Due to the explicit interconnection (input/output

like) structure of the circuits, order reduction methods based on balancing are attractive to apply to these circuits. However, structure preservation and circuit interpretation of the corresponding reduced order models is not possible yet, and is one of the motivators for further research to both linear and nonlinear structure preserving order reduction methods.

References

1. A.C. Antoulas, *Approximation of Large-Scale Dynamical Systems*, SIAM, Philadelphia, 2005.
2. A. Astolfi, Model reduction by moment matching, *Proc. 7th IFAC NOLCOS*, Pretoria, South Africa, 95–102, August 2007.
3. J. Batt, Nonlinear compact mappings and their adjoints, *Math. Ann.*, 189, 5–25, 1970.
4. C. L. Beck, J. Doyle, and K. Glover, Model reduction of multi-dimensional and uncertain systems, *IEEE Trans. Automat. Control*, AC-41, 10, 1466–1477, 1996.
5. P. Benner, V. Mehrmann, and D.C. Sorensen, Eds., *Dimension Reduction of i Large-Scale Systems*, Springer-Verlag, Berlin, 2005.
6. R. Courant and D. Hilbert, *Methods of Mathematical Physics*, Vol. 1, Interscience Publishers, New York, 1953.
7. U. B. Desai and D. Pal, A transformation approach to stochastic model reduction, *IEEE Trans. Automat. Control*, 29, 1097–1100, 1984.
8. K.V. Fernando and H. Nicholson, Singular pertubational model reduction of balanced systems, *IEEE Trans. Automat. Control*, AC-27, 466–468, 1982.
9. K. Fujimoto, What are singular values of nonlinear operators?, *Proc. 43rd IEEE Conf. on Decision and Control*, The Bahamas, 1623–1628, 2004.
10. K. Fujimoto, J. M. A. Scherpen, and W. S. Gray, Hamiltonian realizations of nonlinear adjoint operators, *Automatica*, 38, 10, 1769–1775, 2002.
11. K. Fujimoto and J.M.A. Scherpen, Nonlinear balanced realization based on singular value analysis of Hankel operators, *Proc. 42nd IEEE Conf. on Decision and Control*, Maui, HI, 6072–6077, 2003.
12. K. Fujimoto and J.M.A. Scherpen, Nonlinear input-normal realizations based on the differential eigenstructure of Hankel operators, *IEEE Trans. Automat. Control*, AC-50, 1, 2–18, 2005.
13. K. Fujimoto and J.M.A. Scherpen, Singular value analysis and balanced realizations for nonlinear systems, in: *Model Order Reduction: Theory, Research Aspects and Applications*, Eds. W. Schilders, H. van der Vorst, and J. Rommes, Springer-Verlag, Berlin, 251–272, 2008.
14. K. Fujimoto and J.M.A. Scherpen, Model reduction for nonlinear systems based on the balanced realization, *SIAM J. Control Optim.*, 48 (7), 4591–4623, 2010.
15. F.R. Gantmacher, *The Theory of Matrices*, Chelsea, New York, 1960.
16. W. Gawronski, *Balanced Control of Flexible Structures, Lecture Notes in Contr. Inf. Sc.* 211, Springer-Verlag, Berlin, 1996.
17. K. Glover, All optimal Hankel-norm approximations of linear multivariable systems and their L^∞-error bounds, *Int. J. Control*, 39, 1115–1193, 1984.
18. W.S. Gray and J.M.A. Scherpen, State dependent matrices in quadratic forms, *Systems Control Lett.*, 44, 3, 219–232, 2001.
19. S. Gugercin and A. C. Antoulas, A survey of model reduction by balanced truncation and some new results, *Int. J. Control*, 77 (8), 748–766, 2004.
20. J. Hahn and T.F. Edgar, An improved method for nonlinear model reduction using balancing of empirical Gramians, *Comp. Chem. Eng.*, 26, 10, 1379–1397, 2002.
21. P. Heuberger, A family of reduced order models, based on open-loop balancing, *Ident. Model. Contr.*, Delft University Press, 1, 1–10, 1990.
22. J. W. Hoffmann, Normalized coprime factorizations in continuous and discrete time—a joint state-space approach, *IMA J. Math. Contr. Info.* 13, 359–384, 1996.
23. T.C. Ionescu, K. Fujimoto and J.M.A. Scherpen, Dissipativity preserving balancing for nonlinear systems—a Hankel operator approach, *Systems and Control Letters*, 59, 180–194, 2010.
24. A. Isidori, *Nonlinear Control Systems* (3rd edition), Springer-Verlag, Berlin, 1995.
25. E.A. Jonckheere and L.M. Silverman, Singular value analysis of deformable systems, *J. Circ., Syst. Sign. Proc.*, 1, 447–470, 1982.
26. E.A. Jonckheere and L.M. Silverman, A new set of invariants for linear systems—Applications to reduced order compensator design, *IEEE Trans. Automat. Control*, AC-28, 953–964, 1983.
27. T. Kailath, *Linear Systems*, Prentice-Hall, Englewood Cliffs, NJ, 1980.

28. S. Lall, J.E. Marsden, and S. Glavaski, A subspace approach to balanced truncation for model reduction of nonlinear control systems, *Int. J. Robust Nonlinear Control*, 12, 6, 519–535, 2002.
29. D.G. Meyer, Fractional balanced reduction—model reduction via a fractional representation, *IEEE Trans. Automat. Control*, AC-26, 1341–1345, 1990.
30. D.G. Meyer and S. Srinivasan, Balancing and model reduction for second order form linear systems, *IEEE Trans. Automat. Control*, AC-41, 11, 1632–1644, 1996.
31. J. W. Milnor, *Topology from the Differential Viewpoint*, Princeton University Press, Princeton, NJ, 1965.
32. B.C. Moore, Principal component analysis in linear systems: Controllability, observability and model reduction, *IEEE Trans. Automat. Control*, AC-26, 17–32, 1981.
33. D. Mustafa, $\mathcal{H}_\infty$-characteristic values, *Proc. 28th IEEE Conf. on Decision and Control*, Tampa, FL, 1483–1487, 1989.
34. D. Mustafa and K. Glover, *Minimum Entropy $\mathcal{H}_\infty$ Control, Lect. Notes Contr. Inf. Sci.* No. 146, Springer-Verlag, Berlin, 1990.
35. D. Mustafa and K. Glover, Controller reduction by $\mathcal{H}_\infty$-balanced truncation, *IEEE Trans. Automat. Control*, AC-36, 668–682, 1991.
36. C.N. Nett, C.A. Jacobson, and M.J. Balas, A connection between state-space and doubly coprime fractional representations, *IEEE Trans. Autom. Control*, AC-29, 831–832, 1984.
37. A.J. Newman and P.S. Krishnaprasad, Computation for nonlinear balancing, *Proc. 37th IEEE Conf. on Decision and Control*, Tampa, FL, 4103–4104, 1998.
38. A. Newman and P. S. Krishnaprasad, Computing balanced realizations for nonlinear systems, *Proc. Symp. Mathematical Theory of Networks and Systems*, 2000.
39. R. Ober and D. McFarlane, Balanced canonical forms for minimal systems: A normalized coprime factor approach, *Linear Algebra and its Applications*, 122–124, 23–64, 1989.
40. G. Obinata and B.D.O. Anderson, *Model Reduction for Control System Design*, Springer-Verlag, London, 2001.
41. P.C. Opdenacker and E.A. Jonckheere, LQG balancing and reduced LQG compensation of symmetric passive systems, *Int. J. Control*, 41, 73–109, 1985.
42. P. C. Opdenacker and E. A. Jonckheere, A contraction mapping preserving balanced reduction scheme and its infinity norm error bounds, *IEEE Trans. Circuits Systems* 35 (2), 184–189, 1988.
43. L. Pernebo and L.M. Silverman, Model reduction via balanced state space representations, *IEEE Trans. Automat. Control*, AC-27, 382–387, 1982.
44. H. Sandberg and A. Rantzer, Balanced truncation of linear time-varying systems, *IEEE Trans. Automat. Control*, 49, 2, 217—229, 2004.
45. J.M.A. Scherpen, Balancing for nonlinear systems, *Systems Control Lett.*, 21, 143–153, 1993.
46. J.M.A. Scherpen, $\mathcal{H}_\infty$ balancing for nonlinear systems, *Int. J. Robust Nonlinear Control*, 6, 645–668, 1996.
47. J.M.A. Scherpen and W.S. Gray, Minimality and local state decompositions of a nonlinear state space realization using energy functions, *IEEE Trans. Automat. Control*, AC-45, 11, 2079–2086, 2000.
48. J.M.A. Scherpen and W.S. Gray, Nonlinear Hilbert adjoints: Properties and applications to Hankel singular value analysis, *Nonlinear Anal., Theory, Methods Appl.*, 51, 5, 883–901, 2002.
49. J.M.A. Scherpen and A.J. van der Schaft, Normalized coprime factorizations and balancing for unstable nonlinear systems, *Int. J. Control*, 60, 6, 1193–1222, 1994.
50. W.H.A. Schilder, H.A. van der Vorst, and J. Rommes, Eds., *Model Order Reduction: Theory, Research Aspects and Applications*, Mathematics in Industry, Vol. 13, Springer, Berlin, 2008.
51. G.W. Stewart, On the early history of the singular value decomposition, *SIAM Rev.*, 35, 4, 551–566, 1993.
52. H.L. Trentelman, Bounded real and positivey real balanced truncation using Σ normalized coprime factor, *System Control Lett.*, 58, 12, 871–879, 2009.
53. H.L. Trentelman and P. Rapisarda, A behavioral approach to passivity and bounded realness preserving balanced truncation with error bounds, *Proc. 47th IEEE Conf. on Decision and Control*, Shanghai, China, 2009.
54. A.J. van der Schaft, Balancing of lossless and passive systems, *IEEE Trans. Automat. Control*, AC-53, 2153–2157, 2008.
55. A.J. van der Schaft and J.E. Oeloff, Model reduction of linear conservative mechanical systems, *IEEE Trans. Automat. Control*, AC-35, 729–733, 1990.
56. A.J. van der Schaft and R. Polyuga, Structure-preserving model reduction of complex physical systems, *Proc. 48th IEEE Conf. on Decision and Control*, Shanghai, China, December 16–18, 2009.
57. E. Verriest, Low sensitivity design and optimal order reduction for the LQG-problem, *Proc. 24th Midwest Symp. Circ. Syst. Albuquerque*, NM, 1981, 365–369.
58. E.I. Verriest and W.S. Gray, Flow balancing nonlinear systems, *Proc. 2000 Int. Symp. Math. Th. Netw. Syst. (MTNS)*, 2000.

59. E.I. Verriest and W.S. Gray, Balanced nonlinear realizations, *Proc. 43rd IEEE Conf. on Decision and Control*, The Bahamas, 1164–1169, 2004.
60. M. Vidyasagar, *Control System Synthesis: A Factorization Approach*, MIT Press, Cambridge, 1985.
61. S. Weiland, *Theory of Approximation and Disturbance Attenuation for Linear Systems*, Doctoral dissertation, University of Groningen, 1991.
62. P. Wortelboer, *Frequency-Weighted Balanced Reduction of Closed Loop Mechanical Servo Systems: Theory and Tools*, doctoral dissertation, Delft University of Technology, 1994.
63. K. Zhou, J. C. Doyle, and K. Glover, *Robust and Optimal Control*, Prentice-Hall, Inc., Upper Saddle River, NJ, 1996.
64. A. J. Krener, Reduced order modeling of nonlinear control systems, in *Analysis and Design of Nonlinear Control Systems*, A. Astolfi and L. Marconi, eds., Springer-Verlag, 41–62, 2008.