



# **Introduction to Model Order Reduction**

## **Lecture Notes**

Henrik Sandberg

(Version: July 19, 2019)

**Division of Decision and Control Systems  
School of Electrical Engineering and Computer Science  
KTH Royal Institute of Technology, Stockholm, Sweden**



# Preface

This compendium is mainly a collection of formulas, theorems, and exercises from the lectures in the course EL3340 Introduction to Model Order Reduction, given at KTH in the Spring of 2019. References to more complete works are included in each section. Some of the exercises have been borrowed from other sources, and in that case a reference to the original source is provided.

Henrik Sandberg, March 2019.



# CONTENTS

1	Linear Systems, Truncation, and Singular Perturbation . . . . .	7
1.1	Linear State-Space Systems . . . . .	7
1.2	Reduced Order Systems and Approximation Criteria . . . . .	7
1.3	Truncation and Singular Perturbation . . . . .	8
1.4	Truncation $\sim$ Projection . . . . .	9
1.5	Preserving Special Model Structures . . . . .	10
1.6	Recommended Reading . . . . .	12
1.7	Exercises . . . . .	12
1.8	Appendix: Stable systems, $L_2[0, \infty)$ , $H_2$ , and $H_\infty$ . . . . .	14
2	Some Transfer Function Based Heuristics . . . . .	16
2.1	Padé Approximants . . . . .	16
2.2	The Half Rule . . . . .	17
2.3	Pole and Zero Truncation in the Closed Loop . . . . .	17
2.4	Recommended Reading . . . . .	18
2.5	Exercises . . . . .	19
3	Singular Value Decomposition and Principal Component Analysis . . . . .	21
3.1	Singular Value Decomposition . . . . .	21
3.2	Principal Component Analysis (Proper Orthogonal Decomposition [POD]) . .	23
3.3	PCA for Controllability Analysis . . . . .	24
3.4	Recommended Reading . . . . .	24
3.5	Exercises . . . . .	25
4	Gramians and Balanced Realizations . . . . .	26
4.1	Interlude: Optimization in Hilbert Space . . . . .	26
4.2	The Reachability Gramian (Revisited) . . . . .	26
4.3	The Observability Gramian . . . . .	27
4.4	Balanced Realizations . . . . .	28
4.5	Recommended Reading . . . . .	29
4.6	Exercises . . . . .	29
5	Balanced Truncation and Balanced Singular Perturbation . . . . .	32
5.1	Gramians, Reachability and Observability . . . . .	32
5.2	Interlude: Lyapunov Equations . . . . .	32
5.3	Balanced Truncation and Singular Perturbation . . . . .	33
5.4	Suggested Reading . . . . .	35
5.5	Exercises . . . . .	35
6	Frequency-Weighted Balanced Truncation and Controller Reduction . . . . .	37
6.1	Frequency-Weighted Balanced Truncation . . . . .	37
6.2	Controller Reduction . . . . .	39

6.3	Suggested Reading . . . . .	41
6.4	Exercises . . . . .	41
7	Nonlinear Model Order Reduction . . . . .	43
7.1	Suggested Reading . . . . .	44
8	Optimal Model Order Reduction in the Hankel Norm . . . . .	45
8.1	The Hankel Norm and the Hankel Operator . . . . .	45
8.2	State-Space Formulas for Construction of $Q^*$ . . . . .	46
8.3	$H_\infty$ Error Bounds . . . . .	48
8.4	Suggested Reading . . . . .	49
8.5	Exercises . . . . .	49

<b>Bibliography</b>	<b>50</b>
---------------------	-----------

# 1 Linear Systems, Truncation, and Singular Perturbation

## 1.1 Linear State-Space Systems

We consider linear state-space systems

$$\mathbf{G} : \begin{cases} \dot{x} = Ax + Bu, & x(0) = x_0 \\ y = Cx + Du \end{cases} \quad (1.1)$$

with *state*  $x(t) \in \mathbb{R}^n$ , *input*  $u(t) \in \mathbb{R}^m$ , and *output*  $y(t) \in \mathbb{R}^p$ . Knowledge of  $x_0$  and  $u(t)$  in the time interval  $[0, T]$  determines  $x(t)$  and  $y(t)$  uniquely in the same interval by

$$\begin{aligned} x(t) &= e^{At}x_0 + \int_0^t e^{A(t-\tau)}Bu(\tau) d\tau, \\ y(t) &= Cx(t) + Du(t). \end{aligned}$$

If the input-output mapping  $u \mapsto y$  is the main concern, and the initial state is  $x_0 = 0$ , we can identify and represent the system  $\mathbf{G}$  by a linear operator mapping  $y = \mathbf{G}u$  (with some abuse of notation). Often we assume  $A$  is a *Hurwitz matrix*, i.e., all eigenvalues of  $A$  are in the open left complex half plane ( $\mathbb{C}_-$ ) and the system is (asymptotically) stable.

Large classes of models can be written in the form (1.1). Examples include

- Discretized partial differential equations (finite difference/finite element methods etc.), such as diffusion and wave equations; and
- Linearized nonlinear ordinary differential equations.

The linear operator  $\mathbf{G}$  can also be represented by its transfer function (matrix)

$$G(s) = C(sI - A)^{-1}B + D \in \mathbb{C}^{p \times m}$$

for complex frequencies  $s \in \mathbb{C}$ . As a measure of system size, and to measure the distance between two different systems, we regularly use the  $H_\infty$  norm:

$$\begin{aligned} \|G\|_\infty &:= \sup_{s \in \mathbb{C}_+} \|G(s)\| \quad (\mathbb{C}_+ \text{ is the open complex right-half plane}) \\ &= \sup_{\omega} \|G(j\omega)\| \quad (G(s) \text{ has no poles in } \mathbb{C}_+) \end{aligned}$$

which is finite if, and only if, the system  $\mathbf{G}$  is *stable* ( $G(s)$  has no poles in the closed right complex half plane  $\bar{\mathbb{C}}_+$ ). Here  $\|G(s)\|$  denotes the largest singular value of the matrix  $G(s)$  in the MIMO (Multi-Input–Multi-Output) case. In the SISO (Single-Input–Single-Output) case this is equal to the magnitude of the complex number  $G(s)$ . See Appendix at the end of the chapter and Section 3.1 for further details.

**Remark 1.** The  $H_2$  norm is an alternative measure sometimes used in the literature. At this point we note that many of the model-reduction methods we consider give a priori bounds expressed in the  $H_\infty$  norm, but not in the  $H_2$  norm.

## 1.2 Reduced Order Systems and Approximation Criteria

We identify the complexity of the system  $\mathbf{G}$  with its order  $n$ . Some motivation for this definition are

- optimal controllers (LQG/ $H_2$ / $H_\infty$ ) for  $\mathbf{G}$  tend to have order of at least  $n$ , and

- the simulation time of the system (1.1) is strongly correlated to the number  $n$  of differential equations.

A reduced order system ("an approximation") of  $\mathbf{G}$  is a state-space system  $\mathbf{G}_r$

$$\mathbf{G}_r : \begin{cases} \dot{z} = A_r z + B_r u, & z(0) = z_0 \\ y_r = C_r z + D_r u \end{cases} \quad (1.2)$$

such that  $z(t) \in \mathbb{R}^r$ , where  $r < n$ , and  $y_r(t) \in \mathbb{R}^p$ . Note that the same input  $u$  is applied to both  $G$  and  $G_r$ .

Not only should  $\mathbf{G}_r$  be of lower order than  $\mathbf{G}$ , its trajectories should say something about the trajectories of  $\mathbf{G}$ . Otherwise, we can hardly speak of an approximation. The main approximation criterium we will be interested in, in this course, is to make  $\|G - G_r\|_\infty$  small. Motivation for this choice will be given throughout the course. One simple motivation is that it is a measure of the worst-case error. Other criteria, such as the relative criterion  $\|G^{-1}(G - G_r)\|_\infty$  and the frequency-weighted criterion  $\|W_1(G - G_r)W_2\|_\infty$  will also be discussed later in the course.

### 1.3 Truncation and Singular Perturbation

Good approximations  $\mathbf{G}_r$  can often be obtained by means of *truncation* or *singular perturbation (residualization)*. Both methods are done in two steps:

**Step 1:** Transform the state-space coordinates  $x(t)$  into  $\bar{x}(t)$  using  $x(t) = T\bar{x}(t)$ . That is, find a *suitable* invertible matrix  $T \in \mathbb{R}^{n \times n}$ , and transform the state-space model

$$\begin{aligned} \bar{A} &= T^{-1}AT = \begin{pmatrix} \bar{A}_{11} & \bar{A}_{12} \\ \bar{A}_{21} & \bar{A}_{22} \end{pmatrix}, \bar{A}_{11} \in \mathbb{R}^{r \times r}, & \bar{B} &= T^{-1}B = \begin{pmatrix} \bar{B}_1 \\ \bar{B}_2 \end{pmatrix}, \bar{B}_1 \in \mathbb{R}^{r \times m}, \\ \bar{C} &= CT = (\bar{C}_1 \quad \bar{C}_2), \bar{C}_1 \in \mathbb{R}^{p \times r}, & \bar{D} &= D, \end{aligned}$$

where the state-space coordinates are conformably partitioned into  $\bar{x} = (\bar{x}_1, \bar{x}_2) \in \mathbb{R}^n$  and  $\bar{x}_1 \in \mathbb{R}^r$ .

**Step 2:** Depending on method, obtain  $\mathbf{G}_r$  by:

– Truncation:

$$\begin{aligned} A_r &= \bar{A}_{11}, & B_r &= \bar{B}_1 \\ C_r &= \bar{C}_1, & D_r &= \bar{D}. \end{aligned}$$

– Singular perturbation (assuming  $\bar{A}_{22}$  is invertible):

$$\begin{aligned} A_r &= \bar{A}_{11} - \bar{A}_{12}\bar{A}_{22}^{-1}\bar{A}_{21}, & B_r &= \bar{B}_1 - \bar{A}_{12}\bar{A}_{22}^{-1}\bar{B}_2 \\ C_r &= \bar{C}_1 - \bar{C}_2\bar{A}_{22}^{-1}\bar{A}_{21}, & D_r &= \bar{D} - \bar{C}_2\bar{A}_{22}^{-1}\bar{B}_2. \end{aligned}$$

**Remark 2.** Truncation gives an exact system approximation  $\mathbf{G}_r$  if  $\bar{x}_2 = 0$ , and singular perturbation is exact if  $\dot{\bar{x}}_2 = 0$ . One may expect that the methods work well when either of these equalities are approximately correct.

It also holds that

**(P1)** using truncation, we always have  $G(\infty) = G_r(\infty)$ ; and

**(P2)** using singular perturbation, we always have  $G(0) = G_r(0)$ .



Hence, these methods achieve perfect approximation either at zero or at infinite frequency. Depending on how the reduced model will be used, one chooses one method or the other.

For successful application of these methods (usually it is not enough to have a good approximation only at a single frequency), we need to find good coordinate transformations  $T$ , and a suitable approximation order  $r$ . The following lectures will deal with this problem. Mostly, we will use the truncation method in the following, for simplicity.

**Remark 3.** *Truncation and singular perturbation can also be applied to reduce the order of nonlinear systems. If there is a time-scale separation between the transformed variables  $\bar{x}_1$  and  $\bar{x}_2$ , such that  $\dot{\bar{x}}_2 \approx 0$ , it is possible to carry out a rigorous approximation error analysis also in the nonlinear case. See [1] for further details.*

## 1.4 Truncation $\sim$ Projection

The truncation method can also be seen as a projection in the original state space  $\mathbb{R}^n$  to an  $r$ -dimensional subspace corresponding to the reduced state space  $\mathbb{R}^r$ . Consider the matrices

$$\begin{aligned} W^T &= \begin{pmatrix} I_r & 0_{r \times (n-r)} \end{pmatrix} T^{-1} \in \mathbb{R}^{r \times n} \quad ("x \rightarrow z") \\ V &= T \begin{pmatrix} I_r \\ 0_{(n-r) \times r} \end{pmatrix} \in \mathbb{R}^{n \times r}, \quad ("z \rightarrow x") \end{aligned} \quad (1.3)$$

which combine coordinate transformation and truncation. Note that  $P = VW^T \in \mathbb{R}^{n \times n}$  defines a (generally oblique) projection on  $\mathbb{R}^n$  to the subspace spanned by the columns of  $V$ ,

$$P = P^2,$$

because  $W^T V = I_r$ . That is, the columns of  $W$  and  $V$  form a bi-orthogonal system.

Given any projection  $P = VW^T$  (i.e.,  $W^T V = I_r$ ), we can approximate  $\mathbf{G}$  using a so-called a *Petrov-Galerkin projection*. (If  $W^T = V^T$ , the projection is orthogonal and is called a *Galerkin projection*.) A Petrov-Galerkin projection of  $\mathbf{G}$ , yielding  $\mathbf{G}_r$ , is given by the state-space realization

$$\begin{aligned} A_r &= W^T A V, & B_r &= W^T B \\ C_r &= C V, & D_r &= D, \end{aligned}$$

which of course coincides with the truncated model in Section 1.3 when  $W^T$  and  $V$  are chosen as (1.3). To further understand the Petrov-Galerkin projection, we can carry out the following analysis. Assume we want to try to express the solution  $x(t) \in \mathbb{R}^n$  to the model  $\mathbf{G}$  only in  $r$  variables. Such a solution can be written as  $x(t) = Vz(t)$ , where  $z(t) \in \mathbb{R}^r$  and  $V \in \mathbb{R}^{n \times r}$ .  $V$  should be chosen such that its columns span a subspace where we believe the solution  $x(t)$  will lie in. If the believed solution is inserted into the original state-space model  $\mathbf{G}$ , we obtain

$$\dot{x} = V\dot{z} = AVz + Bu + E, \quad (1.4)$$

where  $E$  is the equation error. Now,  $x(t) = Vz(t)$  is a solution to the original problem if, and only if,  $E(t) = 0$  for all  $t$ . There are  $n$  equations in (1.4), but only  $r$  unknowns in  $z$ , and so the system is generally overdetermined. To find a unique solution, we can require that the projection of the residual  $E$  onto the subspace spanned by  $V$  is zero. This projection is given by  $W^T$ . Hence, we add the condition

$$W^T E(t) = 0, \quad \forall t$$

to (1.4). We then obtain the equation

$$\dot{z} = W^T AVz + W^T Bu,$$

which exactly is the Petrov-Galerkin projection of (1.1). From this,  $z(t)$  can be computed, and the projection of the resulting residual  $E$  onto  $V$  is zero.

## 1.5 Preserving Special Model Structures

Although truncation and singular perturbation often can be used to obtain good reduced models, they may destroy important model structures. In applications, sometimes one does not only care about preserving a mapping  $u \mapsto y$ , but also about the reduced system  $\mathbf{G}_r$  retaining a special internal structure. We next give two such examples, and describe how to apply or modify truncation and singular perturbation. In addition to the examples described here, other model structures that may be worth preserving arise from matrix sparsity, second-order systems, Markov chains, positive systems, and gradient systems.

### Port-Hamiltonian Systems

Almost all fundamental dynamical systems in physics can be written as Hamiltonian systems. In order to account for interaction with environment variables ( $u$  and  $y$ ), the class of port-Hamiltonian systems has been introduced. A linear port-Hamiltonian system takes the form

$$\mathbf{G} : \begin{cases} \dot{x} = (J - R)Qx + Bu, & x(0) = x_0 \\ y = B^T Qx, \end{cases} \quad (1.5)$$

here for simplicity stated without a direct feedthrough term  $D$ . The matrices  $J = -J^T$  and  $B$  specify the interconnection constraints of the state variables and the environment,  $R = R^T \geq 0$  (positive definite) is the dissipation matrix, and  $H(x) = \frac{1}{2}x^T Qx$  is the total energy (Hamiltonian) stored in the system. Systems in this form satisfy the basic dissipation inequality

$$\frac{d}{dt}H(x) = u^T y - x^T Q R Q x \leq u^T y.$$

Hence  $u^T y$  has an interpretation as the external power supply to the system, and if  $Q \geq 0$  the port-Hamiltonian system is *passive*. With appropriate choices of matrices, port-Hamiltonian systems can model (linear) mechanical and electrical systems. Conversely, models in this form have a physical interpretation, which explains why a practitioner may want to preserve this particular structure in the reduction process.

Truncation and singular perturbation directly applied to (1.5) will typically destroy the port-Hamiltonian structure. However, two alternative methods exist [2] as described next.

We first note that the coordinate transformation in **Step 1**,  $x = T\bar{x}$ , transforms the state-space matrices as

$$\begin{aligned} \bar{J} &= T^{-1} J T^{-T} (= -\bar{J}^T), & \bar{R} &= T^{-1} R T^{-T} (= \bar{R}^T) \\ \bar{Q} &= T^T Q T (= \bar{Q}^T), & \bar{B} &= T^{-1} B. \end{aligned}$$

For **Step 2**, one of the following methods could be applied. (To simplify the exposition, we do not use the  $\bar{\cdot}$  notation below.)

**Effort constraint method.** Assuming the matrix  $Q$  is invertible,  $e = Qx = \nabla H(x)$  are an alternative set of coordinates, sometimes called the co-energy or effort variables. The co-energy variables often correspond to voltage, current, force, or velocity in physical systems, and it is natural to express system constraints in them. Let  $e = (e_1, e_2) \in \mathbb{R}^n$  and conformably  $x = (x_1, x_2) \in \mathbb{R}^n$ , and one such constraint is

$$0 = e_2 = Q_{21}x_1 + Q_{22}x_2. \quad (1.6)$$

A reduced-order system  $\mathbf{G}_r$  is then given by

$$\mathbf{G}_r : \begin{cases} \dot{x}_1 = (J_{11} - R_{11})e_1 + B_1u = (J_{11} - R_{11})(Q_{11}x_1 + Q_{12}x_2) + B_1u \\ = (J_{11} - R_{11})(Q_{11} - Q_{12}Q_{22}^{-1}Q_{21})x_1 + B_1u \\ y_r = B_1^T(Q_{11} - Q_{12}Q_{22}^{-1}Q_{21})x_1, \end{cases} \quad (1.7)$$

where we have used  $x_2 = -Q_{22}^{-1}Q_{21}x_1$  from (1.6). We note that (1.7) has the required port-Hamiltonian structure, and the reduced-system Hamiltonian is  $H_r(x_1) = \frac{1}{2}x_1^T(Q_{11} - Q_{12}Q_{22}^{-1}Q_{21})x_1$ . By Schur complements, if  $\mathbf{G}$  is passive, so is  $\mathbf{G}_r$ . Passivity may be an essential external model property to preserve in the reduction.

**Flow constraint method.** Let us denote  $F := J - R$ . Note that  $J = \frac{1}{2}(F - F^T)$  and  $-R = \frac{1}{2}(F + F^T)$  are the anti/skew-symmetric and symmetric parts of  $F$ , respectively, and can thus be recovered from  $F$ . Again,  $x = (x_1, x_2) \in \mathbb{R}^n$  and let us for simplicity make the assumption that  $Q$  is conformably block diagonal,  $Q = \begin{bmatrix} Q_{11} & 0 \\ 0 & Q_{22} \end{bmatrix}$  and invertible. The more general case is covered in [2], but note that one can always perform a coordinate transformation to make  $Q$  block diagonal. Let us now impose the constraint

$$0 = \dot{x}_2 = F_{21}Q_{11}x_1 + F_{22}Q_{22}x_2 + B_2u. \quad (1.8)$$

The derivative  $\dot{x}$  is often called the flow of the system, and thus (1.8) is called a flow constraint. A reduced-order system  $\mathbf{G}_r$  is then given by (assuming  $F_{22}$  is invertible)

$$\mathbf{G}_r : \begin{cases} \dot{x}_1 = F_{11}Q_{11}x_1 + F_{12}Q_{22}x_2 + B_1u \\ = (F_{11} - F_{12}F_{22}^{-1}F_{21})Q_{11}x_1 + (B_1 - F_{12}F_{22}^{-1}B_2)u \\ y_r = (B_1^T - B_2^T F_{22}^{-1}F_{21})Q_{11}x_1 - B_2^T F_{22}^{-1}B_2u, \end{cases} \quad (1.9)$$

where we have used  $Q_{22}x_2 = -F_{22}^{-1}F_{21}Q_{11}x_1 - F_{22}^{-1}B_2u$  from (1.8). We note that (1.9) has the required port-Hamiltonian structure if  $(F_{12}F_{22}^{-1})^T = F_{22}^{-1}F_{21}$ , and the reduced-system Hamiltonian is  $H_r(x_1) = \frac{1}{2}x_1^T Q_{11}x_1$ . Note that the reduced model has an additional direct feedthrough term  $-B_2^T F_{22}^{-1}B_2$ , which is positive semi-definite. Hence if  $\mathbf{G}$  is passive, so is  $\mathbf{G}_r$ .

## Consensus Dynamics and Laplacian Matrices

A consensus system takes the form

$$\dot{x} = -\mathcal{L}x, \quad x(0) = x_0,$$

where  $\mathcal{L}$  is a *Laplacian* matrix. For simplicity, we do not include inputs  $u$  and outputs  $y$  in this section, since the state variables themselves are the main interest here. Assume the state variables  $x_i \in \mathbb{R}$ ,  $i = 1, \dots, n$ , each are associated with  $n$  vertices  $v_i \in \mathcal{V} := \{1, 2, \dots, n\}$  in a connected and undirected graph  $\mathcal{G}$ . The vertices  $v_i$  and  $v_j$  may be connected by edges  $e_{ij} \in \mathcal{E}$ , each with a weight  $w_{ij} > 0$ . The (weighted) Laplacian matrix  $\mathcal{L} \in \mathbb{R}^{n \times n}$  is defined as

$$\mathcal{L}_{ij} := \begin{cases} w_i, & \text{if } i = j, \\ -w_{ij}, & \text{if } i \sim j \\ 0, & \text{otherwise.} \end{cases}$$

Here  $i \sim j$  means there is an edge between vertices  $v_i$  and  $v_j$ , and  $w_i = \sum_{i \sim j} w_{ij}$ . The Laplacian has many interesting properties, and has been under intense study. We only name a few properties here:  $\mathcal{L}$  is symmetric positive semi-definite, it has an eigenvector of one entries ( $\mathbf{1} \in \mathbb{R}^n$ ) corresponding

to the smallest eigenvalue  $\lambda_1(\mathcal{L}) = 0$ , and the the second smallest eigenvalue  $\lambda_2(\mathcal{L}) > 0$  reflects the connectivity strength of the system. In fact,  $\lambda_2(\mathcal{L})$  is the worst-case exponential rate of convergence to the consensus state of the system,  $\lim_{t \rightarrow \infty} x(t) = (\frac{1}{n} \mathbf{1}^T x_0) \mathbf{1}$ . Given the special properties of the consensus dynamics, it is clear that it is of interest to retain them during model reduction.

**Remark 4.** A finite-difference approximation of the differential Laplacian operator  $\Delta$  under Neumann boundary conditions attains the form of Laplacian matrix, which motivates the name.

A method to reduce the order of consensus dynamics is to cluster states that reach consensus faster than the system as a whole. Suppose the states in vertices corresponding to  $\mathcal{V}_1 \subset \mathcal{V}$  reach consensus with each other fast. We can then choose a reduced-order state  $z_1$  as

$$x = \mathbf{e}_{\mathcal{V}_1} z_1, \quad z_1 \in \mathbb{R},$$

and  $\mathbf{e}_{\mathcal{V}_1} \in \mathbb{R}^n$  is a vector of zeros except for the entries corresponding to  $\mathcal{V}_1$ , which are set to ones. In this manner, the states can be divided into  $r$  disjoint clusters such that  $\mathcal{V}_1 \cup \mathcal{V}_2 \cup \dots \cup \mathcal{V}_r = \mathcal{V}$ . The reduced-order system with state  $z \in \mathbb{R}^r$  is now defined as

$$\dot{z} = -Q^T \mathcal{L} Q z, \quad Q := [\mathbf{e}_{\mathcal{V}_1} \quad \mathbf{e}_{\mathcal{V}_2} \quad \dots \quad \mathbf{e}_{\mathcal{V}_r}] \in \mathbb{R}^{n \times r}.$$

It is easy to verify that  $Q^T \mathcal{L} Q$  retains the desired Laplacian structure. Note that the matrix  $Q$  is *not* a projection as defined in Section 1.4 (why?). Rather, the reduction here is a singular perturbation of the dynamics in each cluster.

How to choose appropriate vertices to cluster in graphs is a very active area of research. We only mention one method here: The eigenvector  $v_2$  corresponding to the eigenvalue  $\lambda_2(\mathcal{L})$  is the slowest converging mode shape in the system. Fiedler [3] showed that the subgraph corresponding to the positive entries in  $v_2$  is a connected graph. Similarly, the negative entries in  $v_2$  correspond to a connected subgraph. In essence, the eigenvector  $v_2$  splits the graph  $\mathcal{G}$  into two weakly interacting subsystems. If consensus is reached fast in each of these subsystems (which depends on the other eigenvalues of  $\mathcal{L}$ ), a good reduced-order model may be to choose the cluster sets  $\mathcal{V}_1$  and  $\mathcal{V}_2$  corresponding to the subgraphs. If this is not a good reduced model, the procedure can be applied recursively to each subgraph, to create new connected clusters that converge slowly towards each other.

## 1.6 Recommended Reading

Sections 3.2–3.2.3 in *Linear Robust Control* [4] discuss linear systems and the  $H_\infty$  norm. Sections 9.1–9.2.2 in *Linear Robust Control* [4] discuss model truncation and singular perturbation in more detail. The controllability and observability Gramians mentioned on page 315 will be introduced at a later stage in this course.

An introduction to the theory of port-Hamiltonian systems is provided in [5] and aspects of model reduction are discussed in [2]. Consensus dynamics is a very active research area and the survey paper [6] provides an introduction. Clustering-based model reduction in networks is discussed in [7], for example.

## 1.7 Exercises

### EXERCISE 1.1 (Modal representation and truncation)

Assume that  $A$  has  $n$  distinct eigenvalues  $\lambda_i$ . Then there exists a coordinate transformation  $T$  such that

$$\bar{A} = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}, \quad \bar{B} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}, \quad \bar{C} = (c_1 \quad \dots \quad c_n).$$

This is called a modal representation of  $\mathbf{G}$ . The state corresponding to  $\lambda_i$  is often called the  $i$ -th mode of the system. Prove that using truncation on the modal representation we obtain

$$G(s) - G_r(s) = \sum_{i=r+1}^n \frac{c_i b_i}{s - \lambda_i}, \quad (1.10)$$

$$\|G - G_r\|_\infty \leq \sum_{i=r+1}^n \frac{|c_i b_i|}{|\operatorname{Re} \lambda_i|}. \quad (1.11)$$

To show the error bound (1.11) you need that  $A$  is a Hurwitz matrix. Why? Discuss how an error bound like (1.11) can be used. What modes should be truncated?

### EXERCISE 1.2 (Modal truncation)

What modes should be truncated in the following systems if  $\|G - G_r\|_\infty$  should be small?

a)

$$\bar{A} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -3 \end{pmatrix}, \quad \bar{B} = \begin{pmatrix} 1 \\ 0.1 \\ 0.1 \end{pmatrix}, \quad \bar{C} = (1 \quad 1 \quad 0.1).$$

b)

$$\bar{A} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -0.1 & 0 \\ 0 & 0 & -0.101 \end{pmatrix}, \quad \bar{B} = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}, \quad \bar{C} = (1 \quad 1 \quad 1).$$

c)

$$\bar{A} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -0.1 & 0 \\ 0 & 0 & -0.101 \end{pmatrix}, \quad \bar{B} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad \bar{C} = (1 \quad 1 \quad 1).$$

d)

$$G(s) = \frac{(s+2)(s+4)(s+6)(s+8)}{(s+1)(s+3)(s+5)(s+7)}.$$

Relate your choices of truncated states to the error bound (1.11). Plot Bode diagrams of the systems  $G$  and the approximations  $G_r$  you construct.

### EXERCISE 1.3

What are the necessary and sufficient conditions for controllability and observability of the states in a modal representation of  $G$ ?

### EXERCISE 1.4

Prove properties **(P1)** and **(P2)**.

### EXERCISE 1.5

Construct a port-Hamiltonian model of the circuit in Figure 1.1. Use the voltage  $v_{\text{in}}$  as input, the current  $i_{\text{in}}$  as output, and the state vector  $x = (q, \phi)$  where  $q$  is the charge in the capacitor and  $\phi$  is the magnetic flux in the inductor. Here  $L$  is the inductance,  $C$  the capacitance, and  $R_1, R_2$  the resistances.

Apply the effort and flow constraint methods to reduce the inductor state  $\phi$ . Give an electrical interpretation of the reduced model in both cases.

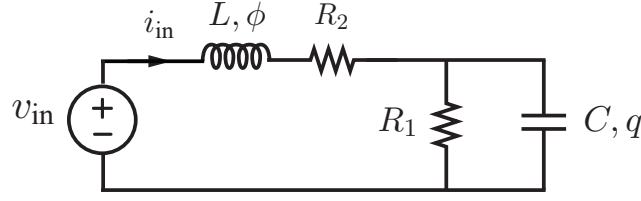


Figure 1.1: The circuit in Exercise 1.5.

## 1.8 Appendix: Stable systems, $L_2[0, \infty)$ , $H_2$ , and $H_\infty$

A short introduction to the function spaces used in the course is provided here. For a more in-depth treatment, see [4, Sections 3.1-3.3], for instance.

The (Hilbert) space  $L_2[0, \infty)$  contains all square-integrable signals, that is

$$L_2[0, \infty) = \{u(t) : \|u\|_2 < \infty\},$$

where

$$\|u\|_2 := \left( \int_0^\infty \|u(t)\|^2 dt \right)^{1/2}$$

is the  $L_2$  norm and  $\|u\| := \sqrt{u^T u}$ . The linear system  $\mathbf{G}$  can be represented by the linear operator mapping  $y = \mathbf{G}u$ , which is (input-output) *stable* if

$$\|\mathbf{G}\| := \sup_{u \in L_2[0, \infty), u \neq 0} \frac{\|\mathbf{G}u\|_2}{\|u\|_2} = \sup_{\|u\|_2 \leq 1} \|\mathbf{G}u\|_2 < \infty.$$

That is, for all input signals  $u \in L_2[0, \infty)$ , the output  $y$  will belong to  $L_2[0, \infty)$ . Put differently,  $\mathbf{G} : L_2[0, \infty) \rightarrow L_2[0, \infty)$ , and

$$\|y\|_2 \leq \|\mathbf{G}\| \|u\|_2.$$

In the (complex) frequency domain, the mapping  $y = \mathbf{G}u$  takes the form  $Y(s) = G(s)U(s)$ , where  $Y(s)$  and  $U(s)$  are the (unilateral) Laplace transforms of  $u(t)$  and  $y(t)$ , respectively. That is,

$$U(s) := \mathcal{L}\{u\}(s) = \int_0^\infty u(t)e^{-st} dt,$$

which is a well defined and analytic function in the complex half plane  $\text{Re } s > a$  if  $u$  is exponentially bounded by a constant times  $e^{at}$  (and analogously for  $y$ ). If we denote the time-domain impulse response of  $\mathbf{G}$  by  $g(t)$ , its transfer function is given by

$$G(s) = \int_0^\infty g(t)e^{-st} dt = C(sI - A)^{-1}B + D,$$

for the complex half plane  $\text{Re } s > \max_i \text{Re } \lambda_i(A)$ . Here  $\lambda_i(A)$  are the eigenvalues of  $A$ . It is a standard result that the cascade connection of the linear systems  $\mathbf{F}$  and  $\mathbf{G}$ , that is  $\mathbf{FG}$ , has transfer function  $F(s)G(s)$ . The parallel connection  $\mathbf{F} + \mathbf{G}$  has transfer function  $F(s) + G(s)$ .

In fact, the Laplace transform is well defined for all signals  $u \in L_2[0, \infty)$  and the resulting  $U(s)$  is analytic in  $s \in \mathbb{C}_+$ , where  $\mathbb{C}_+$  is the open complex right half plane ( $\text{Re } s > 0$ ). Furthermore, the transform belongs to the (Hilbert) space  $H_2$ ,

$$H_2 = \{U(s) : U(s) \text{ is analytic in } \mathbb{C}_+ \text{ and } \|U\|_2 < \infty\}.$$

The norm in  $H_2$  (the  $H_2$  norm) is

$$\|U\|_2 := \left( \sup_{a>0} \frac{1}{2\pi} \int_{-\infty}^{\infty} \|U(a + j\omega)\|^2 d\omega \right)^{1/2},$$

where  $\|U\| := \sqrt{U^*U}$  and  $U^* := \bar{U}^T$  (complex conjugate transpose). If  $U \in H_2$ , a limiting function at the boundary  $\text{Re } s = 0$  exists for almost all  $\omega$ , let us call it  $U(j\omega)$ , and it holds that

$$\|U\|_2 = \left( \frac{1}{2\pi} \int_{-\infty}^{\infty} \|U(j\omega)\|^2 d\omega \right)^{1/2}.$$

This is a formula frequently used in explicit calculations. The *Paley-Wiener theorem* says the spaces  $H_2$  and  $L_2[0, \infty)$  are isomorphic<sup>1</sup> under the Laplace transform. In particular,

$$\|u\|_2 = \|U\|_2.$$

It is now easy to compute  $\|\mathbf{G}\|$  in the frequency domain. By the above discussion, it holds that

$$\|\mathbf{G}\| = \sup_{\|U\|_2 \leq 1} \|\mathbf{G}U\|_2 = \sup_{s \in \mathbb{C}_+} \|G(s)\| =: \|G\|_\infty.$$

If  $G(s)$  is a matrix, then  $\|G(s)\|$  means the largest singular value of  $G(s)$  (see Section 3.1). If  $G(s)$  is a complex number, then  $\|G(s)\| = |G(s)|$ .

The (Banach) space  $H_\infty$  is defined as a set of complex functions  $G(s)$ ,

$$H_\infty = \{G(s) : G(s) \text{ is analytic in } \mathbb{C}_+ \text{ and } \|G\|_\infty < \infty\}.$$

It follows that a linear system  $\mathbf{G}$  is stable in the above sense if, and only if, its transfer function matrix  $G \in H_\infty$ . The  $H_\infty$  norm  $\|\cdot\|_\infty$  gives a measure of the worst-case amplification of a system, and is therefore a good measure of approximation error. For computing  $H_\infty$  norms, see also the *bounded real lemma* [4, Section 3.7]. Important properties of the  $H_\infty$  norm are its sub-multiplicative property  $\|FG\|_\infty \leq \|F\|_\infty \|G\|_\infty$  and the triangle inequality  $\|F + G\|_\infty \leq \|F\|_\infty + \|G\|_\infty$ .

---

<sup>1</sup>There is a bijective linear mapping between the spaces (in this case the Laplace transform) that preserves the inner product. The inner product of two elements  $u, v \in L_2[0, \infty)$  is  $(u, v)_2 := \int_0^\infty u^T v dt$ .

## 2 Some Transfer Function Based Heuristics

In this section, we will give some examples of transfer function based model reduction heuristics that are used in particular applications. While they do generally not come with global approximation guarantees, they often do provide useful reduced models and also intuition about the problem. The chapters following this one will introduce more systematic and rigorous methods, but it is not a bad idea to familiarize oneself with these heuristics first, to properly appreciate what is to come.

### 2.1 Padé Approximants

A locally optimal rational approximation  $G_r(s)$  of a SISO transfer function  $G(s)$  is called a *Padé approximant*. It is a generalization of Maclaurin polynomial approximations and is typically applied around the point  $s = 0$ . That is, we are looking for an approximation

$$G_r(s) = \frac{b_0 + b_1s + \dots + b_ms^m}{1 + a_1s + \dots + a_ns^n}, \quad m, n \text{ fixed},$$

such that it is as close as possible to  $G(s)$  around  $s = 0$ . It turns out that there is a unique Padé approximant such that

$$G_r(s) = G(s) + \mathcal{O}(s^{m+n+1}), \quad s \rightarrow 0.$$

That this

$$\begin{aligned} G_r(0) &= G(0) \\ G'_r(0) &= G'(0) \\ &\vdots \\ G_r^{(m+n)}(0) &= G^{(m+n)}(0). \end{aligned}$$

The approximation is local and tends to work well for "slow" inputs (small  $s$ ), but it can give very poor results for fast inputs. In particular, the impulse response of a Padé approximant  $G_r(s)$  may be unstable even if  $G(s)$  is stable.

Nevertheless, Padé approximants are often applied, especially to construct simple approximations of infinite-dimensional systems, such as (time) delays  $G(s) = e^{-Ls}$ . In particular, we have

$$\begin{aligned} e^{-Ls} &= 1 - Ls + \mathcal{O}(s^2), & (m = 1, n = 0) \\ e^{-Ls} &= \frac{1}{1 + Ls} + \mathcal{O}(s^2), & (m = 0, n = 1) \\ e^{-Ls} &= \frac{e^{-Ls/2}}{e^{Ls/2}} = \frac{1 - Ls/2}{1 + Ls/2} + \mathcal{O}(s^3), & (m = 1, n = 1). \end{aligned}$$

These approximations can be quite useful, especially if they are used in cascade with some low-pass system. The MATLAB command `pade` directly generates delay approximations of desired order.

**Remark 5.** *There are moment-matching-type model-reduction methods that, similar to Padé approximants, locally approximate  $G(s)$  and its derivatives up to a certain order at some fixed set of interpolation points  $\{s_k\}$ . An attractive feature of these methods is that the reduced model is very fast to compute, typically an order of  $\mathcal{O}(r^2n)$  computations are needed. This is achieved by applying Krylov subspaces and other efficient numerical linear algebra algorithms. A drawback is that it is not obvious how to choose the interpolation orders and points, and a globally good (or even stable) approximation is not guaranteed. One should expect to iteratively refine the interpolation points and orders. These lectures will not further cover moment-matching type methods. There are books that covers the topic in great detail, both for the linear [8] and the nonlinear [9] systems.*



## 2.2 The Half Rule

Several tuning rules for PI- and PID-controllers assume the plant to be controlled has a model on the form

$$P(s) = \frac{K}{sT + 1} e^{-sL} \quad \text{or} \quad P(s) = \frac{K}{(sT_1 + 1)(sT_2 + 1)} e^{-sL},$$

where  $K$  is the steady-state gain,  $T, T_1, T_2$  are time constants, and  $L$  the time delay. (See, for instance, Exercise 2.2.) Unfortunately, rarely real plants are this simple, and it is then desirable to find an approximate model of this type. Skogestad [10] has proposed a simple "half rule" to obtain such approximations of systems with real poles and zeros, and time delays. The procedure focuses on obtaining a good match of the initial part of the plant's step response.

The idea is to compensate for neglected time constants by introducing an effective delay, using the following Padé approximants

$$1 - \theta s \approx e^{-\theta s} \quad \text{and} \quad \frac{1}{1 + \theta s} \approx \frac{1}{e^{\theta s}} = e^{-\theta s}.$$

Delays cause large phase loss and to not obtain too conservative approximations, half of the neglected slowest time constant is added to the retained time constant, and the other half is added to the effective delay (hence the "half rule"). We illustrated with an example:

$$P(s) = \frac{k(1 + s\tau_4)(1 - s\tau_5)}{(1 + s\tau_1)(1 + s\tau_2)(1 + s\tau_3)^2} e^{-\theta s},$$

where  $0 < \tau_1 < \tau_2 < \tau_3$ . A first-order plus delay model is now obtained as

$$\begin{aligned} P_r(s) &= \frac{K}{sT + 1} e^{-sL}, \\ K &= k, \\ L &= \theta + \boxed{\tau_2/2} + 2\tau_3 - \tau_4 + \tau_5 && \text{(effective delay)} \\ T &= \tau_1 + \boxed{\tau_2/2} && \text{(effective time constant)}. \end{aligned}$$

The quality of the approximation clearly depends on the exact values of parameters. However, for tuning of PI(D)-controllers in realistic systems, the half rule approximation often gives rise to good control system performance [10].

## 2.3 Pole and Zero Truncation in the Closed Loop

In introductory control systems courses, one of the first control-design methods taught is often pole placement. It is shown how to design a negative feedback controller  $C(s)$  such that the poles of the plant transfer function  $P(s)$  can be arbitrarily placed. Often much less attention is given to the problem of where the poles should be placed. In fact, this is not an easy problem when there are many poles  $n$  to place and several control objectives to fulfill. In practice, it is then often useful to find an approximate plant transfer function  $P_r(s) \approx P(s)$  with fewer poles  $r$ , and design  $C(s)$  for the approximate model. With some care in the approximation, such designs often work well also for the real plant  $P(s)$ . We will here give some heuristic guidelines on which poles and zeros that can be truncated, but we will return to this topic in a more systematic and rigorous manner in Chapter 6.

Two of the most important closed-loop transfer functions are entirely expressed in the loop gain,  $L(s) = P(s)C(s)$ , namely the sensitivity function  $S(s) = \frac{1}{1 + L(s)}$  and the complementary sensitivity function  $T(s) = \frac{L(s)}{1 + L(s)}$ . To ensure robustness and performance in the control system, it is important that the plant is well approximated around all frequencies where  $|L(j\omega)| \approx 1$  (think of Nyquist's

stability criterion). One such frequency is around the desired closed-loop bandwidth  $\omega_b$ , but there may be others as well. Typically poles and zeros that are not significant at such frequencies may be truncated, and thereby reduce the model order. However, once the controller is designed it must be tested on the full-order system. In the following, we assume the transfer function has been factored into its poles and zeros, as

$$P(s) = k \frac{(s + z_1)(s + z_2) \dots (s^2 + 2\zeta_{z,1}\omega_{z,1}s + \omega_{z,1}^2)(s^2 + 2\zeta_{z,2}\omega_{z,2}s + \omega_{z,2}^2) \dots}{(s + p_1)(s + p_2) \dots (s^2 + 2\zeta_{p,1}\omega_{p,1}s + \omega_{p,1}^2)(s^2 + 2\zeta_{p,2}\omega_{p,2}s + \omega_{p,2}^2) \dots},$$

where  $k$  is a constant gain. A pole (zero) is considered *fast* if  $p_i/\omega_{p,i}$  ( $z_i/\omega_{z,i}$ ) is larger than the desired closed-loop bandwidth  $\omega_b$ , and *slow* if is smaller. It is neither fast, nor slow, if the magnitude is around  $\omega_b$ .

The following guidelines are based on the ones given by Bernhardsson and Åström [11]. Poles and zeros are here assumed to be stable (in  $\mathbb{C}_-$ ). In fact, unstable poles and zeros tend to result in fundamental performance limitation and can rarely be neglected. Unstable poles and zeros can never be canceled by the controller in step 5 (why?).

1. Truncate plant poles and zeros that are an order of magnitude faster than  $\omega_b$ , and adjust gain  $k$ .

Example:  $P(s) = \frac{b}{s + a} \Rightarrow P_r(s) = \frac{b}{a}$  if  $\omega_b \lesssim a/10$ .

2. Truncate plant zero and pole pairs that are an order of magnitude slower than  $\omega_b$ .

Example:  $P(s) = \frac{s + b}{s^2} \Rightarrow P_r(s) = \frac{1}{s}$  if  $\omega_b \gtrsim 10b$ .

3. Approximate slow poles by integrators.

Example:  $P(s) = \frac{b}{s + a} \Rightarrow P_r(s) = \frac{b}{s}$  if  $\omega_b \gtrsim 10a$ .

4. Exceptions: Do *not* truncate lightly damped slow zeros (fast poles) unless controller gain sufficiently large (small) at the corresponding frequency. Such poles and zeros may result in  $|L(j\omega)| \approx 1$  also far away from  $\omega_b$ .

Example: Only approximate  $P(s) = \frac{\omega_0^2}{s^2 + 2\zeta\omega_0 s + \omega_0^2}$ ,  $0 \leq \zeta \lesssim 0.1$ , with  $P_r(s) = 1$  if  $\omega_b \lesssim \omega_0/10$  and after assuring the controller has sufficiently small gain at  $\omega_0$ .

5. Cancel fast plant poles and slow plant zeros that cannot be truncated using the controller  $C(s)$ . Design the controller based on simplified model, and add the cancelling factors to the controller after design. Add high-frequency roll-off to  $C(s)$  to gain robustness if necessary.

It is a good exercise to try to justify these guidelines. To do so, remember the Nyquist stability criterion, and that a "good" loop gain  $|L(j\omega)|$  is large for small frequencies  $\omega$ , small for large frequencies, and cannot decay too quickly around the bandwidth frequency (see [12, Chapter 7], for instance). However, these guidelines are not always applicable or may provide conservative results (since it corresponds to truncation of a modal representation, see Exercise 1.1–1.2).

## 2.4 Recommended Reading

Moment-matching and interpolation-type model reduction is considered in detail in [8]. Numerical aspects of model reduction are also discussed.

The half rule and tuning rules for PID-controllers are proposed in [10].

Loop shaping is discussed in [12, Chapters 7–8] in detail. The lectures slides [11] give several examples of model reduction and the design of pole-placement controllers.

## 2.5 Exercises

### EXERCISE 2.1

Compute the Padé approximants for the following systems:

a)  $G(s) = \frac{s+1}{s^2+0.5s+1}$  when  $m=0$  and  $n=1$

b)  $G(s) = \frac{1}{s^3+2s^2+2s+1}$  when  $m=1$  and  $n=2$ .

Are the approximations good in  $H_\infty$  norm (check  $\|G - G_r\|_\infty$ )?

### EXERCISE 2.2

Consider a linear system  $\mathbf{P}$  with transfer function

$$P(s) = \frac{(-0.1s+1)(0.1s+1)}{(2s+1)(s+1)(0.2s+1)(0.01s+1)^2}.$$

- Use the half rule to obtain an approximation  $\mathbf{P}_r$ .
- Use the SIMC rules [10] and the approximation  $\mathbf{P}_r$  to tune a PI-controller with transfer function

$$C(s) = K_P \left( 1 + \frac{1}{sT_I} \right).$$

That is, choose  $K_P = \frac{T}{K} \frac{1}{L+T_c}$  and  $T_I = \min\{T, 4(T_c + L)\}$ , where  $T_c$  is the tunable time constant of the desired closed-loop system.

Evaluate the PI-controller's performance on both  $\mathbf{P}$  and  $\mathbf{P}_r$ . For what values of  $T_c$  do you obtain a good controller for  $\mathbf{P}$ ? (For instance, check how the closed-loop systems respond to steps in the set-point.)

### EXERCISE 2.3

Consider a linear system  $y = \mathbf{P}u$  with transfer function

$$P(s) = \frac{(s+1)(s+3)}{(s^2+0.1s+1)(s+2)}.$$

- Use the guidelines in Section 2.3 to obtain a first-order approximation  $\mathbf{P}_r$ . For what bandwidths should the approximation be valid?
- Consider a PI-controller  $\mathbf{C}$ , given in time domain as

$$u(t) = k_p(b r(t) - y(t)) + k_i \int_0^t r(\tau) - y(\tau) d\tau,$$

with set-point  $r$ , and tuning parameters  $k_p$ ,  $k_i$ , and  $b$  (set-point weight).

Show that the transfer function of the feedback interconnection of  $\mathbf{P}_r$  and  $\mathbf{C}$  (from  $r$  to  $y$ ) takes the form

$$\frac{(2\zeta b s + \omega_0)\omega_0}{s^2 + 2\zeta\omega_0 s + \omega_0^2},$$

in terms of desired damping ratio  $\zeta$  and natural frequency  $\omega_0$ . Also express  $k_p$  and  $k_i$  in terms of these parameters.

- c) Evaluate the PI-controller's performance on  $\mathbf{P}$  and  $\mathbf{P}_r$  for some different values of  $\zeta$ ,  $\omega_0$ , and  $b$ . For what tunings are the closed-loop systems similar?
- d) Use the modal truncation method introduced in Exercise 1.1 to find a reduced model  $\mathbf{P}_r$  that makes  $\|P - P_r\|_\infty$  small. What order is needed, and how is this model different from the one in a)?

### 3 Singular Value Decomposition and Principal Component Analysis

In this chapter, we shall introduce some numerical tools that are useful for computing suitable coordinate transformations for truncation and singular perturbation. However, the tools are of independent interest and are widely applied in engineering and science.

#### 3.1 Singular Value Decomposition

The singular value decomposition (SVD) of a matrix  $A \in \mathbb{C}^{n \times m}$  is arguably the most useful of all the available matrix factorizations. The SVD reveals the complexity of  $A$ , and can be used both when  $A$  is used as a data storage and as a linear mapping. Most modern model reduction techniques use the SVD in one way or another. In this section, the basic properties of the SVD are reviewed.

First, we introduce some notation. We use the induced 2-norm of a matrix and the Frobenius norm,

$$\begin{aligned}\|x\| &:= \sqrt{x^*x}, \quad u \in \mathbb{C}^m \\ \|A\| &:= \sup_x \frac{\|Ax\|}{\|x\|} = \sqrt{\lambda_{\max}(A^*A)}, \quad A \in \mathbb{C}^{n \times m} \\ \|A\|_F &:= \left( \sum_{i=1}^n \sum_{j=1}^m |A_{ij}|^2 \right)^{1/2} = \sqrt{\text{Trace}(A^*A)},\end{aligned}$$

where  $*$  is the complex conjugate transpose. The induced norm has the property  $\|AB\| \leq \|A\|\|B\|$ , for all matrices  $B$  such that the product  $AB$  is defined. For unitary matrices<sup>2</sup>  $U, V$ , we have  $\|UAV\| = \|A\|$  and  $\|UAV\|_F = \|A\|_F$ .

The SVD of  $A$  is defined as follows. For all matrices  $A \in \mathbb{C}^{n \times m}$  there exist unitary matrices

$$\begin{aligned}U &= [u_1 \ \dots \ u_n] \in \mathbb{C}^{n \times n}, \\ V &= [v_1 \ \dots \ v_m] \in \mathbb{C}^{m \times m},\end{aligned}$$

such that

$$A = U\Sigma V^*, \tag{3.1}$$

where

$$\Sigma_1 = \begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_p \end{pmatrix}, \quad \Sigma = \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \in \mathbb{R}^{n \times m},$$

$p = \min\{n, m\}$ , with the *singular values*

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_p \geq 0.$$

The proof of this statement is given in most modern books on linear algebra.

The following properties are useful:

- $\|A\| = \sigma_1 =: \bar{\sigma}$ ;
- $\|A\|_F = (\sum_{i=1}^p \sigma_i^2)^{1/2}$ ;
- if  $\sigma_1 \geq \dots \geq \sigma_k > \sigma_{k+1} = \dots = \sigma_p = 0$ , then  $\text{Rank}(A) = k$ ;
- $\mathcal{N}(A) = \text{Span}\{v_{k+1}, \dots, v_m\}$  (orthonormal basis of the nullspace of  $A$ );

---

<sup>2</sup>The complex square matrix  $U$  is unitary if  $U^* = U^{-1}$ .

- $\mathcal{R}(A) = \text{Span}\{u_1, \dots, u_k\}$  (orthonormal basis of the range space of  $A$ );
- $AA^* = U\Sigma^2U^*$ , and  $A^*A = V\Sigma^2V^*$ .

An alternative way of writing (3.1) is as the *dyadic expansion*

$$A = \sum_{i=1}^k \sigma_i u_i v_i^*. \quad (3.2)$$

One can view (3.2) as a series expansion of  $A$ , with the terms in decreasing order of importance. This follows from the *Schmidt-Mirsky* approximation theorem (see, for instance, [8, Section 3.2.4]), which is stated next. Assume we want to approximate the matrix  $A \in \mathbb{C}^{n \times m}$  that has a rank smaller or equal to  $r$ . Then it holds that

$$\min_{\text{Rank}(B) \leq r} \|A - B\| = \|A - A_r\| = \sigma_{r+1}, \quad (3.3)$$

$$\min_{\text{Rank}(B) \leq r} \|A - B\|_F = \|A - A_r\|_F = \left( \sum_{i=r+1}^p \sigma_i^2 \right)^{1/2}, \quad (3.4)$$

where  $A_r$  is a truncated dyadic expansion of  $A$ ,

$$A_r := \sum_{i=1}^r \sigma_i u_i v_i^*,$$

retaining the  $r$  dominant terms. In the induced norm  $\|\cdot\|$ ,  $A_r$  is not the unique minimizer (see below). This is a fact that is used in optimal Hankel norm approximation (Section 8). In the Frobenius norm  $\|\cdot\|_F$ , the problem is referred to as *total least squares* in applied statistics, and  $A_r$  is the unique minimizer if the singular values are distinct.

That  $A_r$  is not the unique minimizer in  $\|\cdot\|$  norm is seen as follows: A class of minimizers  $A'_r$  is given by

$$A'_r := \sum_{i=1}^r (\sigma_i - \eta_i) u_i v_i^*, \quad |\eta_i| \leq \sigma_{r+1},$$

as can be shown by writing the approximation error  $A - A'_r$  as

$$A - A'_r = \sum_{i=1}^k \sigma_i u_i v_i^* - \sum_{i=1}^r (\sigma_i - \eta_i) u_i v_i^* = \sum_{i=1}^r \eta_i u_i v_i^* + \sum_{i=r+1}^k \sigma_i u_i v_i^*. \quad (3.5)$$

The rightmost term in (3.5) can be interpreted as a singular value decomposition of the error matrix  $A - A'_r$ , where the bounds on  $\eta_i$  ensure that the largest singular value is given by  $\sigma_{r+1}$ . As a result,

$$\|A - A'_r\|_2 = \left\| \sum_{i=1}^r \eta_i u_i v_i^* + \sum_{i=r+1}^k \sigma_i u_i v_i^* \right\|_2 = \sigma_{r+1}, \quad (3.6)$$

and  $A'_r$  also achieves the minimum in (3.3).

### 3.2 Principal Component Analysis (Proper Orthogonal Decomposition [POD])

Principal Component Analysis (PCA) (see, for instance, [13]) can be viewed as an SVD of a multi-dimensional function. PCA is extensively used for model reduction since it is a flexible tool for choosing good coordinate transformations  $T$ , either for truncation or for singular perturbation, see Section 1.3.

Consider a signal<sup>3</sup>  $x \in L_2^n[0, T]$ , where

$$\begin{aligned} L_2^n[0, T] &= \{x : x(t) \in \mathbb{C}^n, \|x\|_2 < \infty\}, \\ (x, y)_2 &= \int_0^T x(t)^* y(t) dt, \\ \|x\|_2 &= \sqrt{(x, x)} = \left( \int_0^T x(t)^* x(t) dt \right)^{1/2}. \end{aligned}$$

We define the *Gramian* of  $x$  by

$$W = \int_0^T x(t)x(t)^* dt \in \mathbb{C}^{n \times n}.$$

The Gramian is a Hermitian positive semidefinite matrix ( $W = W^*$ ). We define the  $n$  singular values of  $x$  by

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n \geq 0, \quad \sigma_i = \sqrt{\lambda_i(W)},$$

with  $\lambda_i(W)$  being the eigenvalues of  $W$ . We can now expand  $x$  as

$$x(t) = \sum_{i=1}^n \sigma_i u_i v_i(t)^*, \quad (3.7)$$

where

- $v_i \in L_2^1[0, T]$ ,  $(v_i, v_j)_2 = 0$  if  $i \neq j$ , and  $\|v_i\|_2 = 1$ ;
- $u_i \in \mathbb{C}^n$ ,  $u_i^* u_j = 0$ , if  $i \neq j$ , and  $\|u_i\| = 1$ ;
- $v_i(t)^* = u_i^* x(t) / \sigma_i$ , and  $\sigma_i^2 u_i = W u_i$ .

The expansion (3.7) should be compared to the dyadic expansion (3.2). As seen, it can be computed from the eigenvalues and the eigenvectors of the Gramian  $W$  (or the SVD). The  $i$ -th *principal component* of  $x$  is defined by  $\sigma_i u_i v_i(t)^*$ , the  $i$ -th *component vector* by  $u_i$ , and the  $i$ -th *component function* by  $v_i(t)$ . The component vectors describe in decreasing order of importance where the energy of the signal  $x$  is found in  $\mathbb{C}^n$ , as seen next. The total energy of the signal  $x$  is given by the singular values,

$$\|x\|_2 = \left( \sum_{i=1}^n \sigma_i^2 \right)^{1/2}.$$

Define the subspace in  $\mathbb{C}^n$  that is spanned by the component vectors  $u_i$  that correspond to the  $k \leq n$  strictly positive singular values of  $x$  ( $\sigma_1 \geq \dots \geq \sigma_k > \sigma_{k+1} = \dots = \sigma_n = 0$ ) as

$$\mathcal{S}_x = \text{Span}\{u_1, \dots, u_k\}.$$

If  $x$  evolves in the entire space  $\mathbb{C}^n$ , then  $\dim \mathcal{S}_x = n$ . Often, however,  $x$  evolves approximately on a subspace of strictly lower dimension. We want to find a subspace of dimension  $r < k \leq n$

---

<sup>3</sup>We use the notation  $L_2^n[0, T]$  here to emphasize the signal takes values in an  $n$ -dimensional space.

that captures as much as possible of the signal energy of  $x$ . This can be done by truncating the expansion (3.7). We have

$$\min_{\dim \mathcal{S}_y \leq r} \|x - y\|_2 = \|x - x_r\|_2 = \left( \sum_{i=r+1}^n \sigma_i^2 \right)^{1/2}, \quad (3.8)$$

where  $x_r(t) := \sum_{i=1}^r \sigma_i u_i v_i(t)^*$ ,  $\dim \mathcal{S}_{x_r} \leq r$ . Thus if there is a significant drop in the magnitude of the singular values after  $\sigma_r$ , then only  $r$  dimensions, spanned by  $u_1, \dots, u_r$ , are needed to model  $x$  accurately.

### 3.3 PCA for Controllability Analysis

Let us apply PCA to analyze how the state-space of a model  $\mathbf{G}$  is excited by an input  $u$ . The input-to-state mapping is given by

$$x(t) = \int_0^t e^{A(t-\tau)} B u(\tau) d\tau. \quad (3.9)$$

To get a unique signal  $x \in L_2^n[0, T]$  to analyze, let us apply an impulse  $u(t) = \delta(t)$ . We then have the impulse response signal

$$x(t) = e^{At} B,$$

which also lies in  $L_2^n[0, \infty)$ , if  $A$  is Hurwitz. In fact, (3.9) is a convolution of  $u$  and the impulse response. The corresponding Gramian is called the *reachability Gramian*, and is given by

$$P(T) = \int_0^T e^{At} B B^T e^{A^T t} dt \in \mathbb{R}^{n \times n}. \quad (3.10)$$

By analyzing the principal components of  $x$  through  $P(T)$ , we can find new coordinate systems  $T\bar{x} = x$  that capture the subspace of  $\mathbb{R}^n$  that is most excited by the input  $u$ .

In practice, it is often convenient to compute the reachability Gramian through the Lyapunov differential equation

$$\dot{P} = AP + PA^T + BB^T, \quad P(0) = 0. \quad (3.11)$$

When  $T \rightarrow \infty$  and  $A$  is Hurwitz,  $P$  can be computed from the algebraic Lyapunov equation

$$AP + PA^T + BB^T = 0. \quad (3.12)$$

As we shall see in the coming lectures,  $P(T)$  contains a lot of information. For example, the system is controllable if, and only if,  $P(T)$  is nonsingular (for arbitrary  $T > 0$ ), and the controllable subspace is spanned by  $u_1, \dots, u_r$ , when  $\text{Rank}(P(T)) = r$ .

### 3.4 Recommended Reading

The SVD is treated in most text books on linear algebra, see [14], for instance. In [4, Section 2.2] many basic properties of the SVD are listed.

The paper [13] by Bruce Moore pioneered the use of PCA for model reduction. Sections I–IV in the paper is recommended reading. Balanced coordinates, which were introduced for model reduction by Moore, will be treated in the next chapter.



### 3.5 Exercises

#### EXERCISE 3.1

- a) Assume that  $A = U\Sigma V^T$  is invertible. What is the SVD of  $A^{-1}$ ?
- b) Compute the SVD of  $A = \begin{pmatrix} 1 & 2 \\ 4 & 1 \end{pmatrix}$ . Use the SVD to illustrate the mapping  $y = Au$ . What direction in  $\mathbb{R}^2$  is amplified most by  $A$ ?

#### EXERCISE 3.2

- a) Compute the principal components of the impulse response  $x(t) = e^{At}B$ ,  $0 \leq t < \infty$ , where

$$A = \begin{pmatrix} -a & 0 \\ 0 & -a - \epsilon \end{pmatrix}, \quad a > 0, \quad a > |\epsilon| \geq 0, \quad B = \begin{pmatrix} b \\ b \end{pmatrix}.$$

Sketch the phase portrait  $(x_1(t), x_2(t))$  and the component vectors as a function of  $\epsilon$ . (You may assume  $\epsilon$  is small and use suitable approximations.)

- b) Compute the principal components of the impulse response  $x(t) = e^{At}B$ ,  $0 \leq t \leq 10$ , where

$$A = \begin{pmatrix} -0.5 & 2 \\ -2 & -0.5 \end{pmatrix}, \quad B = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Sketch the phase portrait  $(x_1(t), x_2(t))$  and the component vectors.

- c) Suggest coordinate transformations  $x = T\bar{x}$  for a)–b) above, and perform model truncations, see Section 1.3. You can use

$$C = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

#### EXERCISE 3.3

- a) Prove that the reachability Gramian  $P$  as defined in (3.10) satisfies (3.11).
- b) How is  $P$  transformed when the coordinates are changed using a linear coordinate transformation  $T$ ,  $T\bar{x} = x$ ?
- c) Show that  $P(t_2) \geq P(t_1)$ ,  $t_2 \geq t_1$ . ( $P(t_2) \geq P(t_1)$  means that  $P(t_2) - P(t_1)$  is positive semidefinite.)

## 4 Gramians and Balanced Realizations

In this lecture, we use an optimization approach to find suitable realizations for truncation and singular perturbation of  $\mathbf{G}$ . It turns out that the recommended realizations coincide with the ones obtained through PCA of the the impulse-to-state maps  $e^{At}B$  (last lecture) and  $e^{A^T t}C^T$  (see [13]). Such realizations are called *balanced realizations*.

It is good to learn both the PCA approach and the optimization approach to balanced realizations, since they yield different reduced models in the nonlinear case. First, we state a useful theorem, of its own independent interest.

### 4.1 Interlude: Optimization in Hilbert Space

Consider the bounded linear operator  $\mathbf{A} : U \rightarrow Y$ , where  $U$  and  $Y$  are Hilbert spaces. If the inverse  $\mathbf{A}^{-1}$  exists, then there is a unique solution  $u$  to the equation  $y = \mathbf{A}u$ , for given  $y$ . Here we are instead interested in the case when the equation is *overdetermined* or *underdetermined*.

The following results are derived in [15, Sections 6.9–6.10].

**Theorem 1.** Suppose  $\mathbf{A} : U \rightarrow Y$  is a bounded linear operator, where  $U$  and  $Y$  are Hilbert spaces.

i) For fixed  $y \in Y$ , the vector  $u \in U$  that minimizes  $\|y - \mathbf{A}u\|_Y$  satisfies the normal equations

$$\mathbf{A}^* \mathbf{A} u = \mathbf{A}^* y.$$

If  $(\mathbf{A}^* \mathbf{A}) : U \rightarrow U$  is invertible, the unique optimal solution is  $u = (\mathbf{A}^* \mathbf{A})^{-1} \mathbf{A}^* y$ .

ii) Suppose  $\mathbf{A}$  has closed range in  $Y$  [ $\overline{\mathcal{R}(\mathbf{A})} = \mathcal{R}(\mathbf{A})$ ]. For example,  $\mathcal{R}(\mathbf{A})$  is finite dimensional. Then the vector  $u \in U$  with the smallest norm  $\|u\|_U$  satisfying  $y = \mathbf{A}u$ , is given by

$$u = \mathbf{A}^* z,$$

for any  $z \in Y$  that satisfies  $\mathbf{A} \mathbf{A}^* z = y$ .

If  $(\mathbf{A} \mathbf{A}^*) : Y \rightarrow Y$  is invertible, the unique optimal solution is  $u = \mathbf{A}^* (\mathbf{A} \mathbf{A}^*)^{-1} y$ .

The operator  $\mathbf{A}^* : Y \rightarrow U$ , is the Hilbert adjoint [15, Section 6.5] of  $\mathbf{A}$ . By definition, it satisfies

$$(\mathbf{A}u, y)_Y = (u, \mathbf{A}^* y)_U.$$

For example, if  $\mathbf{A} = A \in \mathbb{R}^{n \times m}$ , then  $\mathbf{A}^* = A^T \in \mathbb{R}^{m \times n}$  using the standard scalar product  $(x, y) = x^T y$  in  $\mathbb{R}^n$  and  $\mathbb{R}^m$ .

Theorem 1 i) can be used to solve optimal estimation problems, and Theorem 1 ii) can be used to solve optimal control problems. Here we use them to define two Gramians.

### 4.2 The Reachability Gramian (Revisited)

Let us define the reachability operator  $\mathbf{R}_T : L_2[0, T] \rightarrow \mathbb{R}^n$  for the linear system  $\mathbf{G}$  by

$$x(T) = \mathbf{R}_T u : \quad x(T) = \int_0^T e^{A(T-t)} B u(t) dt.$$

If the realization is controllable, there exists control signals  $u \in L_2[0, T]$  that brings the state from  $x(0) = 0$  to any  $x(T) = x_T$ . That is, there are solutions to the equation,  $x_T = \mathbf{R}_T u$  for all  $x_T \in \mathbb{R}^n$ . The question is which control uses the least amount of energy, i.e., has the smallest possible  $\|u\|_2$ ? We can use Theorem 1 ii) to solve this problem.

To do that, we need the Hilbert adjoint  $\mathbf{R}_T^*$ , and  $\mathbf{R}_T \mathbf{R}_T^*$ . Simple calculations yield that

$$\begin{aligned}\mathbf{R}_T^* : \mathbb{R}^n &\rightarrow L_2[0, T]; & \mathbf{R}_T^* &= B^T e^{A^T(T-t)} \\ \mathbf{R}_T \mathbf{R}_T^* &\in \mathbb{R}^{n \times n}; & \mathbf{R}_T \mathbf{R}_T^* &= \int_0^T e^{A(T-t)} B B^T e^{A^T(T-t)} dt.\end{aligned}$$

If we make a change of variables in the last integral,  $\tau := T - t$ , we see that it coincides with the reachability Gramian (3.10),

$$\mathbf{R}_T \mathbf{R}_T^* = \int_0^T e^{A\tau} B B^T e^{A^T\tau} d\tau \equiv P(T).$$

The system is controllable if, and only if,  $P(T)$  is invertible for all  $T > 0$ . The optimal control and its energy is given by

$$u(t) = \mathbf{R}_T^* (\mathbf{R}_T \mathbf{R}_T^*)^{-1} x_T = B^T e^{A^T(T-t)} P(T)^{-1} x_T, \quad \|u\|_2 = \sqrt{x_T^T P(T)^{-1} x_T}.$$

It follows that for a fixed amount of input energy, say  $\|u\|_2 \leq 1$ , the reachable states  $x_T$  in the state space are contained in the ellipsoid  $\mathcal{R}_{\leq 1}$ ,

$$\mathcal{R}_{\leq 1} = \{x : x = U_P \Sigma_P z, \|z\| \leq 1\}, \quad P(T) = U_P \Sigma_P^2 U_P^T,$$

where  $U_P \Sigma_P^2 U_P^T$  is a singular value decomposition of  $P(T)$  ( $U_P = [u_1 \dots u_n]$  is unitary). Note that  $U_P$  and  $\Sigma_P$  are the component vectors and singular values of  $e^{At} B$ ,  $0 \leq t \leq T$ . The vector  $u_1$  points in the direction where it is easiest to control, and  $\sigma_1$  tells how far one can reach in that direction using energy  $\|u\|_2 \leq 1$ .

We sum up the results of this section in the following theorem.

**Theorem 2.** *A realization  $(A, B, C, D)$  of  $\mathbf{G}$  is controllable if, and only if,  $P(T)$  is invertible for all  $T > 0$ . The reachable subspace is spanned by the component vectors  $u_1, \dots, u_k$  in  $U_P$  that correspond to the  $k \leq n$  strictly positive singular values in  $\Sigma_P$ . The singular value  $\sigma_i$  quantifies how far one can reach in the direction  $u_i$  using the input energy  $\|u\|_2 \leq 1$ .*

### 4.3 The Observability Gramian

The observability Gramian is dual of the reachability Gramian. It can be derived in many ways, using PCA (see [13]) or by using Theorem 1 (see Exercise 4.2), for example. Here we use the perhaps simplest approach.

We want to study how much output energy there is in a particular initial state  $x(0) = x_0$ . That is, what is  $\|y\|_2^2 = \int_0^T y(t)^T y(t) dt$ , when  $x(0) = x_0$  and  $u(t) = 0$ ,  $0 \leq t \leq T$ ? Using that  $y(t) = C e^{At} x_0$ , straightforward calculations yield

$$\|y\|_2^2 = \int_0^T y(t)^T y(t) dt = \int_0^T x_0^T e^{A^T t} C^T C e^{At} x_0 dt = x_0 Q(T) x_0,$$

where we have defined the *observability Gramian* by

$$Q(T) = \int_0^T e^{A^T t} C^T C e^{At} dt.$$

In practice, it is often convenient to compute the reachability Gramian through the Lyapunov differential equation

$$\dot{Q} = A^T Q + Q A + C^T C, \quad Q(0) = 0. \quad (4.1)$$

When  $T \rightarrow \infty$  and  $A$  is Hurwitz,  $Q(= Q(\infty))$  can be computed from the algebraic Lyapunov equation

$$A^T Q + Q A + C^T C = 0. \quad (4.2)$$

The Gramian  $Q(T)$  contains a lot of information. For example, the system is observable if, and only if,  $Q(T)$  is nonsingular for all  $T > 0$ .

Just as for the reachability Gramian, it is useful to make a singular value decomposition of  $Q(T)$ ,  $Q(T) = U_Q \Sigma_Q^2 U_Q^T$ , where  $U_Q = [u_1 \ \dots \ u_n]$  is unitary. The vector  $u_1$  points in the direction of the state space that generates the largest amount of energy in the output. The initial states  $x_0$  that lie on the ellipsoid  $\mathcal{O}_{=1}$ ,

$$\mathcal{O}_{=1} = \{x : x = U_Q \Sigma_Q^{-1} z, \|z\| = 1\}, \quad Q(T) = U_Q \Sigma_Q^2 U_Q^T$$

all result in an output  $\|y\|_2 = 1$ . Hence, points on the ellipsoid very close to the origin are very observable in the output.

We sum up the results of this section in the following theorem.

**Theorem 3.** *A realization  $(A, B, C, D)$  of  $\mathbf{G}$  is observable if, and only if,  $Q(T)$  is invertible for all  $T > 0$ . The unobservable subspace is spanned by the component vectors  $u_{k+1}, \dots, u_n$  in  $U_Q$  that correspond to the (possible) zero singular values in  $\Sigma_Q$ . The singular value  $\sigma_i$  quantifies how much output signal energy  $\|y\|_2$  there is if  $x_0 = u_i$ .*

#### 4.4 Balanced Realizations

The goal of the model-reduction problem we set out to study, is to obtain a model  $\mathbf{G}_r$  from  $\mathbf{G}$  such that

$$\|G - G_r\|_\infty = \sup_{0 \neq u \in L_2[0, \infty)} \frac{\|y - y_r\|_2}{\|u\|_2}$$

is small, where  $y = \mathbf{G}u$  and  $y_r = \mathbf{G}_r u$ . We have argued in Chapter 1 that truncation and singular perturbation of certain state coordinates in  $\mathbf{G}$  can be a good way of obtaining  $\mathbf{G}_r$ . Since the norm  $\|G\|_\infty$  bounds how much signal energy is transferred from the input into the output, it seems like a good idea to remove states in  $\mathbf{G}$  that are not very involved in this energy transfer. This can be quantified using the analysis in Sections 4.2 and 4.3. We next try to find state coordinate directions that are very hard to excite with finite amounts of input energy, and also do not result in much energy in the output.

If we perform the coordinate transformation  $x = T\bar{x}$ , the reachability and observability Gramians transform as

$$\bar{P} = T^{-1} P T^{-T}, \quad \bar{Q} = T^T Q T.$$

One can note that the eigenvalues of the product  $PQ$  are invariant under coordinate transformations, and we define

$$\sigma_i := \sqrt{\lambda_i(PQ)} \quad (4.3)$$

as the *Hankel singular values* of  $\mathbf{G}$ . These turn out to be fundamental in model reduction, and are the *most important of all singular values in the course*.

We show next how the coordinates of  $\mathbf{G}$  can be balanced. The procedure has four steps:

1. Compute the Gramians  $P$  and  $Q$  for any stable and minimal realization  $(A, B, C, D)$  of  $\mathbf{G}$ .
2. Compute the Cholesky factor  $R$  of  $P$ , that is,  $P = R R^T$ .
3. Compute the singular value decomposition  $R^T Q R = U \Sigma^2 U^T$  of  $R^T Q R$ .
4. Use the coordinate transformation

$$x = T\bar{x}, \quad T = R U \Sigma^{-1/2}. \quad (4.4)$$

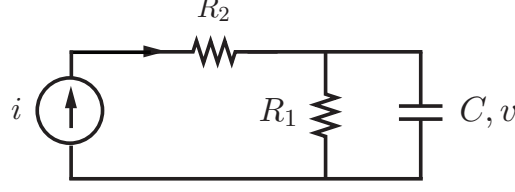


Figure 4.1: The circuit in Exercise 4.1.

Using these coordinates, it holds that

$$\Sigma = \text{diag}\{\sigma_1, \sigma_2, \dots, \sigma_n\}, \quad \sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n > 0$$

$$\bar{P} = \bar{Q} = \Sigma.$$

The reachability and observability Gramian become identical, and diagonal. This is called a *balanced realization*. It means that in the new coordinate vector,  $\bar{x} = (\bar{x}_1 \dots \bar{x}_n)$ ,  $\bar{x}_1$  is simultaneously the most controllable and observable coordinate, and  $\bar{x}_n$  is the least controllable and observable direction.

The balanced coordinates are very good for both truncation and singular perturbation. Among the highlights that we will discuss, we mention that:

- if  $A$  is Hurwitz and  $\sigma_{r+1} < \sigma_r$ , then  $\bar{A}_{11} \in \mathbb{R}^{r \times r}$  and  $\bar{A}_{11} - \bar{A}_{12}\bar{A}_{22}^{-1}\bar{A}_{21} \in \mathbb{R}^{r \times r}$  are Hurwitz; and
- $\sigma_{r+1} \leq \|G - G_r\|_\infty \leq 2 \sum_{i=r+1}^n \sigma_i$ .

The last bound can be compared to the bound in the Schmidt-Mirsky theorem.

## 4.5 Recommended Reading

Read the proof of Theorem 1 in [15, Sections 6.9–6.10]. Balanced realizations were introduced by Moore for model reduction in [13] using the PCA approach. Here a slightly different motivation for balanced realizations is used, but Moore's paper is still very readable. Most of the properties of balanced truncation are shown in [4, Sections 9.3–9.4].

## 4.6 Exercises

### EXERCISE 4.1 (Capacitor control)

Consider the electrical circuit in Figure 4.1, modelled by

$$C \frac{dv(t)}{dt} = -\frac{v(t)}{R_1} + i(t),$$

where  $v(t)$  is the voltage across the capacitor. We can control the current source  $i(t)$ , but there is an internal resistance  $R_2$  in the source, and an amount  $D(T) = \int_0^T R_2 i(t)^2 dt$  of energy is dissipated.

Characterize the voltages  $v(T)$  that are reachable from  $v(0) = 0$ , if the dissipated energy is bounded as

$$D(T) = \int_0^T R_2 i(t)^2 dt \leq 1.$$

**EXERCISE 4.2 (Optimal estimation)**

Suppose that we can measure the noisy output  $y_m(t)$  from a linear system  $\mathbf{G}$ ,

$$\begin{aligned}\dot{x} &= Ax, & x(0) &= x_0 \\ y_m &= Cx + e,\end{aligned}$$

where  $e(t)$  is measurement noise.

Define the observability operator for  $\mathbf{G}$  by  $\mathbf{O}_T : \mathbb{R}^n \rightarrow L_2[0, T]$ ,

$$y(t) = \mathbf{O}_T x_0 : \quad y(t) = Ce^{At}x_0.$$

Use Theorem 1 to find an optimal estimate  $\hat{x}_0$  of the initial state  $x_0$ , given a measurement  $y_m \in L_2[0, T]$ . That is, solve  $\min_x \|y_m - \mathbf{O}_T x\|_2$ .

What is the covariance of the estimation error  $\mathbf{E}(x_0 - \hat{x}_0)(x_0 - \hat{x}_0)^T$ ? You can assume that the measurement noise is approximately white, that is,  $\mathbf{E}e(t) = 0$  and  $\mathbf{E}e(t_1)e(t_2) = \delta(t_1 - t_2)$ .

**EXERCISE 4.3 (Balanced realizations)**

Find balanced realizations for the following systems  $\mathbf{G}$  and truncate them to find approximations  $\mathbf{G}_r$  which make  $\|G - G_r\|_\infty$  small. Compare the result to what you obtained in Exercise 1.2.

a)

$$A = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -3 \end{pmatrix}, \quad B = \begin{pmatrix} 1 \\ 0.1 \\ 0.1 \end{pmatrix}, \quad C = (1 \quad 1 \quad 0.1).$$

b)

$$A = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -0.1 & 0 \\ 0 & 0 & -0.101 \end{pmatrix}, \quad B = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}, \quad C = (1 \quad 1 \quad 1).$$

c)

$$A = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -0.1 & 0 \\ 0 & 0 & -0.101 \end{pmatrix}, \quad B = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad C = (1 \quad 1 \quad 1).$$

d)

$$G(s) = \frac{(s+2)(s+4)(s+6)(s+8)}{(s+1)(s+3)(s+5)(s+7)}.$$

How can you use the Hankel singular values to select approximation order  $r$ ?

**EXERCISE 4.4**

Let  $P$  and  $Q$  be reachability and observability Gramians for a linear system  $\mathbf{G}$ .

- a) Verify that (4.4) balances  $P$  and  $Q$ , i.e., that  $\bar{P} = \bar{Q} = \Sigma$ , and that  $\Sigma$  in Step 3 contains the Hankel singular values  $\sigma_i$  in (4.3).

b) Express

$$\max_x \frac{x^T Q x}{x^T P^{-1} x}, \quad (4.5)$$

using the Hankel singular values. Can you think of a system-theoretic interpretation of the ratio (4.5) for the system  $\mathbf{G}$ ?

c) Prove that a truncated balanced realization still is a balanced realization. You may assume that  $\bar{A}_{11}$  is Hurwitz.

## 5 Balanced Truncation and Balanced Singular Perturbation

In the previous lecture, we introduced the balanced realizations. It turns out that truncating or performing singular perturbation on the balanced realizations of  $\mathbf{G}$  yield good reduced models  $\mathbf{G}_r$  that make  $\|G - G_r\|_\infty$  small, as we see in this lecture. However, before moving to the properties of such reduced-order models, we will discuss some properties of the reachability and observability Gramians and their relation to Lyapunov equations.

### 5.1 Gramians, Reachability and Observability

As seen in the previous lecture, the reachability Gramian  $P(T)$  has a nice interpretation as it characterizes the least amount of (input) energy needed to reach (from the origin) a certain state  $x_T$ , leading to the definition of the reachability ellipsoid  $\mathcal{R}_{\leq 1}$ . From this interpretation, it can be concluded that a system  $\mathbf{G}$  is reachable if and only if all singular values of  $P(T)$  are strictly positive for all  $T > 0$ . Stated differently,  $\mathbf{G}$  is reachable if and only if  $P(T)$  is positive definite (denoted as  $P(T) > 0$ ) for all  $T > 0$ .

When  $\mathbf{G}$  is not reachable, the reachable subspace  $\mathcal{X}_{\text{reach}}$  characterizes the set of states that can be reached from the input ( $\mathcal{X}_{\text{reach}} = \mathbb{R}^n$  when  $\mathbf{G}$  is reachable). The reachable subspace can again be obtained from the reachability Gramian and the equality

$$\mathcal{X}_{\text{reach}} = \mathcal{R}(P(T)) = \mathcal{R}\begin{bmatrix} B & AB & \cdots & A^{n-1}B \end{bmatrix},$$

holds, where  $\mathcal{R}(M)$  denotes the range of a matrix  $M$ . Moreover,  $\begin{bmatrix} B & AB & \cdots & A^{n-1}B \end{bmatrix}$  is called the reachability matrix of  $\mathbf{G}$ .

Dual statements can be made for the observability Gramian  $Q(T)$ , which is positive definite ( $Q(T) > 0$ ) for all  $T > 0$  if and only if the system  $\mathbf{G}$  is observable. Next, the unobservable subspace  $\mathcal{X}_{\text{unobs}}$  is given by the null space of the observability Gramian as

$$\mathcal{X}_{\text{unobs}} = \mathcal{N}(Q(T)) = \mathcal{N}\left(\begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix}\right).$$

The matrix at the right-hand side is known as the observability matrix, and its null space corresponds to that of the observability Gramian.

In the previous lectures, it was seen that the Gramians can be found as the solutions of two Lyapunov *differential* equations. However, for asymptotically stable  $\mathbf{G}$  (i.e.,  $A$  Hurwitz) and infinite horizon ( $T \rightarrow \infty$ ), they satisfy the following *algebraic* Lyapunov equations

$$\begin{aligned} AP + PA^T + BB^T &= 0, \\ A^T Q + QA + C^T C &= 0. \end{aligned}$$

As a result, Gramians are easier to compute for an infinite horizon. Moreover, these Lyapunov equations provide the key to proving stability of reduced-order systems obtained by balanced truncation and balanced singular perturbation.

### 5.2 Interlude: Lyapunov Equations

As Lyapunov equations play an important role in systems and control theory in general (and in model reduction in particular), some of their properties are discussed here.



The Lyapunov equation

$$A^T X + X A + H = 0 \quad (5.1)$$

has a unique solution  $X \in \mathbb{R}^{n \times n}$  if, and only if,  $\lambda_i(A) + \bar{\lambda}_j(A) \neq 0$  for all  $i, j$ , where  $\bar{\lambda}_j$  denotes the complex conjugate of the eigenvalue  $\lambda_j$ . Using this condition, it immediately follows that (5.1) has a unique solution when  $A$  is Hurwitz. Namely, then  $\text{Re}(\lambda_i(A)) < 0$  for all  $i$ .

When  $A$  is Hurwitz, the following properties hold:

- i)  $X = \int_0^\infty e^{A^T t} H e^{A t} dt$ ;
- ii)  $X > 0$  if  $H > 0$ ;  $X \geq 0$  if  $H \geq 0$ ;
- iii) If  $H \geq 0$ , then  $(A, H)$  is observable if and only if  $X > 0$ .

The first property directly relates Lyapunov equations to the Gramians as discussed before, whereas the last property for  $H = C^T C$  retrieves our earlier result on the relation between observability and positive definiteness of the observability Gramian.

The Lyapunov equation (5.1) is directly related to Lyapunov stability of linear dynamical systems. To see this, the linear system  $\dot{x} = Ax$  is introduced as well as the Lyapunov function candidate  $V(x) = x^T X x$ . Then, differentiation of  $V$  along the trajectories of the linear system yields

$$\dot{V}(x) = \dot{x}^T X x + x^T X \dot{x} = (Ax)^T X x + x^T X (Ax) = x^T (A^T X + X A) x = -x^T H x, \quad (5.2)$$

where the latter equality follows from the Lyapunov equation. Using this perspective of the Lyapunov equation, the following properties can be shown:

- iv)  $\text{Re}(\lambda_i(A)) \leq 0$  if  $X > 0$  and  $H \geq 0$ ;
- v)  $A$  is Hurwitz if  $X > 0$  and  $H > 0$ ;
- vi)  $A$  is Hurwitz if  $X \geq 0$ ,  $H \geq 0$  and  $(A, H)$  is detectable.

### 5.3 Balanced Truncation and Singular Perturbation

When the realization of  $\mathbf{G}$  is balanced, the semi axes of the reachability and observability ellipsoids,  $\mathcal{R}_{\leq 1}$  and  $\mathcal{O}_{=1}$ , are lined up in order of importance. To truncate such a realization makes a lot of sense, from an intuitive point of view: The truncated states are not involved much in the energy transfer from input to output. Nevertheless, truncating such a realization is a heuristic, and to this day nobody knows if it is an optimal method in any sense. Even though balanced truncation is a heuristic, it has many good properties.

Before we state the error bounds for balanced truncation and singular perturbation, it is good to keep the following fundamental lower bound on the error in mind. It holds for all approximations  $\mathbf{G}_r$  of order at most  $r$  ( $\deg G_r \leq r$ ) that

$$\inf_{G_r \in H_\infty, \deg G_r \leq r} \|G - G_r\|_\infty \geq \sigma_{r+1}, \quad (5.3)$$

where  $\sigma_{r+1}$  is the  $(r + 1)$ -th largest Hankel singular value of  $\mathbf{G}$ . This can be proved using Hankel norm approximation, which is the topic of a later lecture in the course. Hence, no method can ever perform better than (5.3).

For the sake of convenience, assume that the realization  $(A, B, C, D)$  of  $\mathbf{G}$  is balanced as described in Section 4.4. The reachability and observability Gramians over the infinite time horizon then satisfy

$$\begin{aligned} A\Sigma + \Sigma A^T + BB^T &= 0 \\ A^T \Sigma + \Sigma A + C^T C &= 0 \end{aligned}$$

where  $\Sigma$  is diagonal and contains the Hankel singular values. It can be partitioned into

$$\Sigma = \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix},$$

where

$$\Sigma_1 = \begin{bmatrix} \sigma_1 I_{r_1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_l I_{r_l} \end{bmatrix}, \quad \Sigma_2 = \begin{bmatrix} \sigma_{l+1} I_{r_{l+1}} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_m I_{r_m} \end{bmatrix} \quad (5.4)$$

and  $n = r_1 + \dots + r_m$ ,  $r = r_1 + \dots + r_l$ , and  $\sigma_i \neq \sigma_j$ ,  $i \neq j$ . This notation is introduced to strengthen the results when singular values happen to have a multiplicity greater than one, i.e.,  $r_i > 1$  for some  $i$ . Conformably to  $\Sigma_1, \Sigma_2$ , the realization is partitioned into

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}, \quad C = [C_1 \quad C_2], \quad (5.5)$$

so that a truncated balanced realization is given by  $(A_{11}, B_1, C_1, D)$ .

### Properties of truncated balanced realizations

Truncated balanced realizations satisfy the following theorem.

**Theorem 4.** Suppose  $(A, B, C, D)$  is a balanced realization and that  $(A_{11}, B_1, C_1, D)$  is a balanced truncation. Then  $A_{11}$  is Hurwitz, and  $(A_{11}, B_1, C_1, D)$  is a minimal and balanced realization of  $\mathbf{G}_r$  with Gramian  $\Sigma_1$ . Furthermore,

$$\|G - G_r\|_\infty \leq 2 \sum_{i=l+1}^m \sigma_i.$$

When  $l = m - 1$  equality holds, and  $\|G(0) - G_r(0)\| = 2\sigma_m$  if  $r_m$  is odd.

Note that for  $A_{11}$  to be guaranteed Hurwitz it is important that  $\sigma_l \neq \sigma_{l+1}$ .

For truncation, we have an exact model match of the frequency response at infinite frequency,  $G(\infty) = G_r(\infty)$ . Note that in certain cases it holds that the maximum error is achieved at frequency zero, and one can generally expect that the error is largest for small frequencies.

### Properties of singularly perturbed balanced realizations

Singularly perturbed balanced realizations satisfy the following theorem.

**Theorem 5.** Suppose  $(A, B, C, D)$  is a balanced realization and that

$$(A_r, B_r, C_r, D_r) := (A_{11} - A_{12}A_{22}^{-1}A_{21}, B_1 - A_{12}A_{22}^{-1}B_2, C_1 - C_2A_{22}^{-1}A_{21}, D - C_2A_{22}^{-1}B_2)$$

is a singularly perturbed realization. Then  $A_r$  is Hurwitz, and  $(A_r, B_r, C_r, D_r)$  is a minimal and balanced realization of  $\mathbf{G}_r$  with Gramian  $\Sigma_1$ . Furthermore,

$$\|G - G_r\|_\infty \leq 2 \sum_{i=l+1}^m \sigma_i,$$

with equality if  $l = m - 1$ .

For singular perturbation we always have  $G_r(0) = G(0)$ .

Hence, the uniform error bound  $\|G - G_r\|_\infty$  holds in both cases, but the methods attain good model match at different frequencies.

## 5.4 Suggested Reading

The relation between Gramians and reachability and observability properties can be found in Sections 4.2–4.3 in [8], for example. The results on Lyapunov equations are taken from Section 7.1 of [16].

The stability properties and error bounds for regular balanced truncation are derived in Sections 9.2 and 9.4–9.5 of [4]. The fundamental lower bounds will be derived later in the course.

## 5.5 Exercises

### EXERCISE 5.1

Consider the model  $\mathbf{G}$  of a building that can be found in the file `building.mat`. It represents a model of vibrations in an eight-floor building, where the input  $u$  represents a force acting on the building and the output  $y$  gives the resulting velocity at the same floor.

- Load the realization  $(A, B, C, D)$  from the file `building.mat`. Is this model (asymptotically) stable, reachable, and observable?
- Compute and plot the Hankel singular values. Based on these Hankel singular values, what would be suitable reduction orders?
- Find a balanced realization and perform truncation and singular perturbation to obtain two reduced-order models of the same order  $r = 4$ . Compare their frequency-response functions to that of the high-order model. How are the approximations different?
- Compute the unit step responses (i.e., the response to an input  $u$  that satisfies  $u(t) = 1$  for all  $t \geq 0$  and  $x(0) = 0$ ) for the high-order system and the two reduced-order approximations. Compare the results.
- Verify that the reduced-order models are asymptotically stable, reachable and observable. Finally, compute the error bound and verify that it is satisfied.

### EXERCISE 5.2 [16]

Let

$$G(s) = \sum_{i=1}^5 \frac{\alpha^{2i}}{s + \alpha^{2i}}.$$

Find a balanced realization for each of the following  $\alpha$ :

$$\alpha = 2, 4, 20, 100.$$

Discuss the behavior of the Hankel singular values as  $\alpha \rightarrow \infty$ . What are the implications for the possibility of approximating  $G(s)$ ?

### EXERCISE 5.3

Let the Gramians of a system  $\mathbf{G}$  with realization  $(A, B, C, D)$  be  $P$  and  $Q$  (any coordinates), and let  $\mathbf{G}_r$  be a truncated balanced realization of  $\mathbf{G}$ . Let  $v_i$  and  $w_i$  satisfy

$$PQv_i = \sigma_i^2 v_i, \quad w_i^T PQ = \sigma_i^2 w_i^T,$$

and  $V = [v_1 \ \dots \ v_r]$  and  $W = [w_1 \ \dots \ w_r]$ . Furthermore, normalize  $v_i, w_i$  such that  $W^T V = I$ . Show that the system  $\mathbf{G}_p$  that is realized by  $(W^T A V, W^T B, C V, D)$  has the same input-output behavior as  $\mathbf{G}_r$ , i.e.,  $\|\mathbf{G}_r - \mathbf{G}_p\|_\infty = 0$ .



## 6 Frequency-Weighted Balanced Truncation and Controller Reduction

In this section, extensions of balanced truncation that aim at making the approximation criteria  $\|G^{-1}(G - G_r)\|_\infty$  and  $\|W_o(G - G_r)W_i\|_\infty$  small are presented. An important application of these extensions is found in the order reduction of feedback controllers, which is also discussed.

### 6.1 Frequency-Weighted Balanced Truncation

In control and filter design, and in many other applications, it is important to have good model-match at certain frequencies, not necessarily at  $s = 0$  or at  $s = \infty$ . Hence, we would like to have a method that is more flexible than simple truncation or singular perturbation.

A way this can be done is by introducing frequency weights (filters)  $W_o, W_i \in H_\infty$ , and to try to make the criterion

$$J := \|W_o(G - G_r)W_i\|_\infty$$

small. Hence, by choosing the weights to be large at the frequencies of interest, we can get a good match for those frequencies, provided we have a method to make  $J$  small. One such method is a frequency-weighted extension to balanced truncation. Note that when  $G$  is SISO (scalar), the weights  $W_o$  and  $W_i$  can be lumped into a single weight  $W$ .

The fundamental lower bound (5.3) can be generalized to the frequency-weighted case,

$$\inf_{G_r \in H_\infty, \deg G_r \leq r} \|W_o(G - G_r)W_i\|_\infty \geq \sigma_{r+1}([M_o G M_i]_+), \quad (6.1)$$

where  $\sigma_{r+1}([M_o G M_i]_+)$  is computed as follows: Compute the (unstable) spectral factors  $M_o$  and  $M_i$ ,

$$M_o^\sim M_o = W_o^\sim W_o, \quad M_i M_i^\sim = W_i W_i^\sim,$$

such that  $M_o, M_o^{-1}, M_i, M_i^{-1}$  are anti-stable and have all their poles in the open right half plane  $\mathbb{C}_+$ . Here  $M(s)^\sim := M(-s)^T$ , and  $P_+$  denotes the sum of the stable terms of the partial fraction expansion of  $P(s)$ . The system  $[M_o G M_i]_+$  is then stable and of order  $n$ , and it has  $n$  Hankel singular values. The  $(r + 1)$ -th largest singular value appears in (6.1). The bounds uses the Adamjan-Arov-Krein lemma, which will be discussed together with optimal Hankel norm approximation.

#### A Frequency-weighted Extension to Balanced Truncation

A realization  $(\tilde{A}, \tilde{B}, \tilde{C})$  of the weighted system  $W_o G W_i$  is given by

$$\tilde{A} = \begin{bmatrix} A & 0 & B C_i \\ B_o C & A_o & 0 \\ 0 & 0 & A_i \end{bmatrix}, \quad \tilde{B} = \begin{bmatrix} B D_i \\ 0 \\ B_i \end{bmatrix}, \quad \tilde{C} = [D_o C \quad C_o \quad 0].$$

We have assumed that  $G(\infty) = D = 0$  without loss of generality, since  $D$  can be copied into the  $G_r$  that is obtained by the following procedure.

Let us compute the reachability and observability Gramians  $\tilde{P}, \tilde{Q}$  for the weighted system with the above realization,

$$\begin{aligned} \tilde{A} \tilde{P} + \tilde{P} \tilde{A}^T + \tilde{B} \tilde{B}^T &= 0 \\ \tilde{A}^T \tilde{Q} + \tilde{Q} \tilde{A} + \tilde{C}^T \tilde{C} &= 0. \end{aligned}$$

The *weighted Gramians* for the system  $G$  are now defined by

$$P := [I_n \quad 0 \quad 0] \tilde{P} \begin{bmatrix} I_n \\ 0 \\ 0 \end{bmatrix}, \quad Q := [I_n \quad 0 \quad 0] \tilde{Q} \begin{bmatrix} I_n \\ 0 \\ 0 \end{bmatrix}. \quad (6.2)$$

These are the blocks of  $\tilde{P}$  and  $\tilde{Q}$  that correspond to the states in  $\mathbf{G}$ . The other states belong to  $\mathbf{W}_o$  and  $\mathbf{W}_i$ .

The weighted Gramians have an interpretation of how controllable and observable the states of  $\mathbf{G}$  are *as seen through* the weights  $\mathbf{W}_i$  and  $\mathbf{W}_o$ . For example, if the initial states of the filters are zero and the initial state of  $\mathbf{G}$  is  $x_0$ , the energy of the output signal of  $\mathbf{G}$  filtered through  $\mathbf{W}_o$  is given by  $\sqrt{x_0^T Q x_0}$  (if  $u = 0$ ). Conversely, if the initial state of all the systems are zero at  $t = -\infty$ , and we want to control the state of  $\mathbf{G}$  in  $\mathbf{W}_o \mathbf{G} \mathbf{W}_i$  at time  $t = 0$  to  $x_0$ , the least amount of energy in  $u$  that is needed is given by  $\sqrt{x_0^T P^{-1} x_0}$ . Here the states of  $\mathbf{W}_o$  and  $\mathbf{W}_i$  have been chosen so as to minimize the energy in  $u$  (they are free parameters).

Now the weighted Gramians (6.2) can be balanced just as the regular Gramians were in Section 4.4 and singular values can be computed. The corresponding realization can also be truncated to obtain the reduced model  $\mathbf{G}_r$ . (If  $G(\infty) = D \neq 0$ , then one should use the approximation  $\mathbf{G}_r + D$ .) This  $\mathbf{G}_r$  generally gives a small error  $\|W_o(G - G_r)W_i\|_\infty$ , if the order  $r$  is reasonably chosen. However, this method is a heuristic and it can fail. There are examples where  $\mathbf{G}_r$  even becomes unstable when  $\mathbf{G}$  is stable. The reason for this is that the weighted Gramians do not generally satisfy Lyapunov equations that ensure stability.

Despite this drawback, weighted balanced truncation is simple to apply and use, and it should be the first method of choice for weighted reduction. The method can also be modified in various ways, as we see next and in the provided references, so that stability is maintained and error bounds reminiscent of the ones in Section 5.3 are obtained.

### Balanced Stochastic Truncation

Many times it is important to make the relative error criterion

$$\|G^{-1}(G - G_r)\|_\infty \quad (6.3)$$

small. For example, if one is interested in matching the Bode plots of  $G$  and  $G_r$ . The criterion (6.3) is then suitable since the scales in Bode diagrams are logarithmic. In the SISO case, if we define  $\Delta(j\omega) = (G(j\omega) - G_r(j\omega))/G(j\omega)$ , it holds for small  $\Delta(j\omega)$  that

$$20 \log_{10} |G_r(j\omega)/G(j\omega)| \leq 8.69 |\Delta(j\omega)| \text{ dB}, \quad |\arg G(j\omega) - \arg G_r(j\omega)| \leq |\Delta(j\omega)| \text{ rad}.$$

We can directly apply frequency-weighted balanced truncation to the problem (6.3) under the additional assumption  $G, G^{-1} \in H_\infty$ . This means  $G$  should be minimum phase, and  $G(\infty) = D$  should be invertible. These are hard restrictions but the method can be extended to cope with them as mentioned below.

A realization of the system  $G^{-1}(G - D)$  is given by

$$\tilde{A} = \begin{bmatrix} A & 0 \\ -BD^{-1}C & A - BD^{-1}C \end{bmatrix}, \quad \tilde{B} = \begin{bmatrix} B \\ 0 \end{bmatrix}, \quad \tilde{C} = [D^{-1}C \quad D^{-1}C]. \quad (6.4)$$

The weighted Gramians can then be computed for this realization and be used for truncation. For this particular choice of weights, one can show the following theorem.

**Theorem 6.** Suppose  $G, G^{-1} \in H_\infty$ , and let  $\mathbf{G}_r$  be a truncated realization of  $\mathbf{G}$  that has been balanced with the weighted Gramians of (6.4). Then  $\mathbf{G}_r$  is stable and minimum phase ( $G_r, G_r^{-1} \in H_\infty$ ) and satisfies

$$\frac{\sigma_{l+1}}{\sqrt{1 + \sigma_{l+1}^2}} \leq \|G^{-1}(G - G_r)\|_\infty \leq \prod_{i=l+1}^m \left( 1 + 2\sigma_i(\sqrt{1 + \sigma_i^2} + \sigma_i) \right) - 1$$

$$\frac{\sigma_{l+1}}{\sqrt{1 + \sigma_{l+1}^2}} \leq \|G_r^{-1}(G - G_r)\|_\infty \leq \prod_{i=l+1}^m \left( 1 + 2\sigma_i(\sqrt{1 + \sigma_i^2} + \sigma_i) \right) - 1,$$

where the singular values  $\sigma_i$  are partitioned as in (5.4). The lower bounds are fundamental and hold for any approximation  $\mathbf{G}_r$ .

This method can be extended to the case when  $G(s)$  has zeros in the right complex half plane, i.e., when  $\mathbf{G}^{-1}$  is not stable. This more general method often goes under the name *balanced stochastic truncation*, see [17, Chapter 2], for instance.

## 6.2 Controller Reduction

In this section, the frequency weighted model reduction techniques of the previous section are applied to reduce the order of feedback controllers.

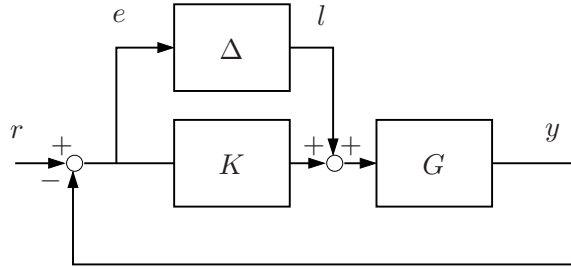


Figure 6.1: A feedback-control system. The plant  $\mathbf{G}$  is controlled by the reduced controller  $\mathbf{K}_r = \mathbf{K} + \Delta$ .

Consider the closed-loop system in Figure 6.1. A plant  $\mathbf{G}$  is controlled by a feedback controller  $\mathbf{K}_r$  (or  $\mathbf{K}$  if  $\Delta = 0$ ). If  $\mathbf{K}$  is designed using optimal control methods ( $H_2/H_\infty$ ), the order of the controller typically is at least as large as the plant  $\mathbf{G}$ . In order to simplify implementation of  $\mathbf{K}$ , we would like to obtain a low-order controller  $\mathbf{K}_r$ . To reduce the order of  $\mathbf{K}$ , it is generally not a good idea to solve the approximation problem using the criterion  $\|\Delta\|_\infty = \|\mathbf{K} - \mathbf{K}_r\|_\infty$ , however. The closed-loop behavior should be taken into account. Compare with Section 2.3.

### Stability Criterion

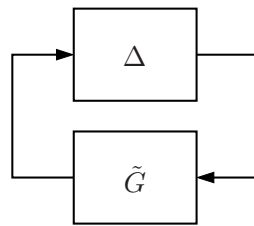


Figure 6.2: A feedback interconnection of  $\tilde{\mathbf{G}}$  and  $\Delta$ .

One of the many reasons to approximate models in the  $H_\infty$  norm is that robust stability can be proven by using the following sufficient condition.

**Theorem 7** (Small-gain theorem). *Suppose that  $\tilde{\mathbf{G}}, \Delta \in H_\infty$ . The feedback interconnection of  $\tilde{\mathbf{G}}$  and  $\Delta$  in Figure 6.2 is stable if  $\|\tilde{\mathbf{G}}\Delta\|_\infty \leq 1$  (or alternatively, if  $\|\tilde{\mathbf{G}}\|_\infty \|\Delta\|_\infty < 1$ ).*

If we look at the transfer function from  $l$  to  $e$  in Figure 6.1, it is given by  $e = -(I + GK)^{-1}Gl =: \tilde{\mathbf{G}}l$ . Hence, applying Theorem 7, the closed-loop system is stable if

$$\|(I + GK)^{-1}G\Delta\|_\infty < 1 \quad \text{or} \quad \|\Delta G(I + KG)^{-1}\|_\infty < 1. \quad (6.5)$$

Note that  $G(I + KG)^{-1} = (I + GK)^{-1}G$ .

Hence, to obtain a reduced-order controller  $\mathbf{K}_r$  that maintains stability, we can use the frequency-weighted reduction method from the previous lecture. For example, using the weights  $W_i = I$  and  $W_o = (I + GK)^{-1}G$ , or  $W_i = G(I + KG)^{-1}$  and  $W_o = I$ .

### Performance Criterion

A different approximation criterion is obtained if we try to match the complementary sensitivity functions of the system with the original and with the reduced order controller,  $\mathbf{GK}(\mathbf{I} + \mathbf{GK})^{-1}$  and  $\mathbf{GK}_r(\mathbf{I} + \mathbf{GK}_r)^{-1}$ . Using a Taylor expansion of the transfer function matrices, we obtain

$$GK_r(I + GK_r)^{-1} - GK(I + GK)^{-1} = (I + GK)^{-1}G(K_r - K)(I + GK)^{-1} + O(\|K_r - K\|_\infty^2). \quad (6.6)$$

Hence, the weighted reduction methods can be used with  $W_i = (I + GK)^{-1}$  and  $W_o = (I + GK)^{-1}G$ .

Of course, other closed-loop transfer functions can be matched in a similar way.

### Plant Reduction in the Closed Loop

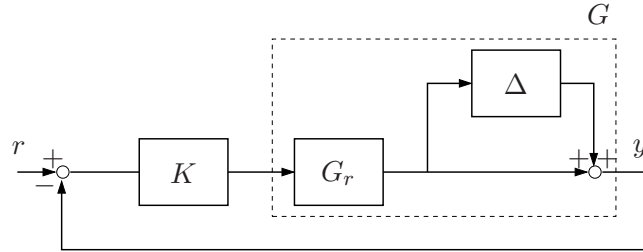


Figure 6.3: A feedback-controlled system. The perturbed reduced plant  $G = (I + \Delta)G_r$  is controlled by the controller  $K$ .

In some cases, one needs to reduce the model of the plant  $\mathbf{G}$ , before the controller  $\mathbf{K}$  can be designed. Then the following argument can be used.

Assume that we have a high-order plant model  $\mathbf{G}$ . We first reduce it to obtain  $\mathbf{G}_r$ , and design a controller  $\mathbf{K}$  for  $\mathbf{G}_r$ . Then we want to ensure that  $\mathbf{K}$  also works for the original plant model  $\mathbf{G}$ . Let us model the original model  $\mathbf{G}$  by a multiplicative perturbation to  $\mathbf{G}_r$ ,  $\mathbf{G} = (I + \Delta)\mathbf{G}_r$  where  $\Delta = (\mathbf{G} - \mathbf{G}_r)\mathbf{G}_r^{-1}$ , see Figure 6.3. This is a good way to parameterize uncertainty in the high-frequency dynamics.

Now, using the small-gain theorem, the closed-loop system in Figure 6.3 is stable if

$$\|\Delta G_r K (I + G_r K)^{-1}\|_\infty < 1.$$

We cannot directly use this relation to obtain  $\mathbf{G}_r$  since  $\mathbf{K}$  is not yet designed. However, assuming that the controller we will design is well working, it holds that  $G_r K (I + G_r K)^{-1} \approx I$  for frequencies up to the bandwidth. Hence, it makes a lot of sense to make the approximation criterion

$$\|\Delta\|_\infty = \|(G - G_r)G_r^{-1}\|_\infty = \|G_r^{-T}(G^T - G_r^T)\|_\infty \quad (6.7)$$

small. This can be done by employing balanced stochastic truncation, see Section 6.1, since using that method to make  $\|G^{-T}(G^T - G_r^T)\|_\infty$  small, also yields the bound

$$\|(G - G_r)G_r^{-1}\|_\infty \leq \prod_{i=l+1}^m \left( 1 + 2\sigma_i(\sqrt{1 + \sigma_i^2} + \sigma_i) \right) - 1.$$



Compare with Theorem 6.

Hence, if reduction of the plant model needs to be performed before the control design, a relative approximation criterion is better than the regular one. Mainly this is because a uniform bound on the phase error comes with the relative criterion.

### 6.3 Suggested Reading

Frequency-weighted model reduction is treated in great detail in [17, Chapter 3]. The design of weights is also covered. The special case of a relative or multiplicative error criterion is covered in [17, Chapter 2].

### 6.4 Exercises

#### EXERCISE 6.1

Let the model  $G$  be given by

$$G(s) = \frac{(s^2 + 0.004s + 0.04)(s^2 + 0.24s + 144)}{(s + 0.001)(s^2 + 0.002s + 0.01)(s^2 + 0.2s + 100)},$$

and let the frequency-dependent weight be

$$W(s) = \frac{s^2}{s^2 + 0.2s + 100}.$$

Perform model reduction to make  $\|G - G_r\|_\infty$  and  $\|W(G - G_r)\|_\infty$  small for some suitable  $r < 5$ . How are the approximations different?

#### EXERCISE 6.2 [18]

Let the model  $G$  be given by

$$G(s) = \frac{0.05(s^7 + 801s^6 + 1024s^5 + 599s^4 + 451s^3 + 119s^2 + 49s + 5.55)}{s^7 + 12.6s^6 + 53.48s^5 + 90.94s^4 + 71.83s^3 + 27.22s^2 + 4.75s + 0.3}.$$

Perform model reduction to make  $\|G - G_r\|_\infty$  and  $\|G^{-1}(G - G_r)\|_\infty$  small for some suitable  $r < 7$ . How are the approximations different?

#### EXERCISE 6.3

a) Show that the weighted reachability Gramian  $P$  satisfies

$$\begin{bmatrix} A & BC_i \\ 0 & A_i \end{bmatrix} \begin{bmatrix} P & P_{12} \\ P_{12}^T & P_{22} \end{bmatrix} + \begin{bmatrix} P & P_{12} \\ P_{12}^T & P_{22} \end{bmatrix} \begin{bmatrix} A & BC_i \\ 0 & A_i \end{bmatrix}^T + \begin{bmatrix} BD_i \\ B_i \end{bmatrix} \begin{bmatrix} BD_i \\ B_i \end{bmatrix}^T = 0.$$

Also derive a similar relation for the weighted observability Gramian  $Q$ . What are the advantages of solving these Lyapunov equations instead of the ones that  $\tilde{P}$  and  $\tilde{Q}$  satisfy?

b) Show that a realization of  $G^{-1}(G - D)$  is given by (6.4), when  $D$  is invertible.

c) Show that the observability Gramian  $\tilde{Q}$  for  $G^{-1}(G - D)$  has the form

$$\tilde{Q} = \begin{bmatrix} Q & Q \\ Q & Q \end{bmatrix},$$

where  $Q$  is the weighted observability Gramian of  $G$ . Also show that  $Q$  satisfies

$$Q(A - BD^{-1}C) + (A - BD^{-1}C)^T Q + C^T D^{-T} D^{-1} C = 0.$$

#### EXERCISE 6.4

Verify the Taylor expansion (6.6).

#### EXERCISE 6.5

A mechanical spring-mass system **G** can be modelled by

$$\begin{aligned} \dot{x} &= Ax + B(u + l) = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -174.7 & -1.362 & 174.7 & 0 \\ 0 & 0 & 0 & 1 \\ 195.7 & 0 & -195.7 & -1.825 \end{pmatrix} x + \begin{pmatrix} 0 \\ 1.293 \\ 0 \\ 0 \end{pmatrix} (u + l) \\ y &= Cx = (0 \quad 0 \quad 280 \quad 0) x, \end{aligned}$$

where  $u$  is the control input (a force),  $l$  is a load disturbance, and  $y$  the position of one mass. Using pole placement and an observer, the following controller  $-\mathbf{K}$  is obtained,

$$\begin{aligned} \dot{\hat{x}} &= (A - BL - K_f C)\hat{x} + K_f y = \begin{pmatrix} 0 & 1 & -194.5 & 0 \\ -294.6 & -33.38 & 478.5 & 14.54 \\ 0 & 0 & -80.01 & 1 \\ 195.7 & 0 & -2650 & -1.825 \end{pmatrix} \hat{x} + \begin{pmatrix} 0.6945 \\ -0.9197 \\ 0.2858 \\ 8.765 \end{pmatrix} y \\ u &= -L\hat{x} = -(92.78 \quad 24.77 \quad -35.85 \quad -11.25) \hat{x}. \end{aligned}$$

(Run the provided `model.m`, `design1.m` and `design2.m` to load the exact models into your MATLAB workspace.)

Perform controller reduction on  $\mathbf{K}$  to obtain a lower order  $\mathbf{K}_r$ , such that the performance of the closed-loop system is preserved as well as possible. For example, compare the load disturbance rejection using both  $\mathbf{K}$  and  $\mathbf{K}_r$ .

## 7 Nonlinear Model Order Reduction

Nonlinear model reduction is a difficult area, and there are not many rigorous results available. The ones we present here can be numerically implemented. However, there are no performance guarantees available.

### Principal Orthogonal Decomposition

An often used method to reduce the order of a nonlinear autonomous system

$$\dot{x} = f(x), \quad x(t) \in \mathbb{R}^n \quad (7.1)$$

is Principal Orthogonal Decomposition (POD). This is nothing but PCA (approximated with SVD) applied to some relevant trajectories of (7.1). For example, one can solve (7.1) once (or many times for different initial conditions) and construct the *snapshot matrix*  $X$ ,

$$X = [x(t_1) \quad x(t_2) \quad \dots \quad x(t_N)] \in \mathbb{R}^{n \times N},$$

out of samples of the representative states. Then one makes an SVD of  $X$  to find a subspace in  $\mathbb{R}^n$  that captures as much as possible of  $x(t)$  (in the  $\|\cdot\|_F$  norm),

$$X = U \Sigma V^T = [u_1 \quad u_2 \quad \dots \quad u_n] \Sigma V^T \approx U_r \Sigma_r V_r^T = [u_1 \quad u_2 \quad \dots \quad u_r] \Sigma_r V_r^T, \quad r < n.$$

One can then apply a linear coordinate transformation  $x = U \bar{x}$  followed by a truncation (Galerkin projection) to obtain a reduced model

$$\dot{z} = f_r(z) = U_r^T f(U_r z), \quad z(t) \in \mathbb{R}^r. \quad (7.2)$$

In general, nothing can be guaranteed about the closeness of the dynamics of (7.1) and (7.2).

### Empirical Gramians [19]

For nonlinear input-output models,

$$\begin{aligned} \dot{x} &= f(x, u), \quad x(t) \in \mathbb{R}^n, \quad u(t) \in \mathbb{R}^m \\ y &= h(x, u), \quad y(t) \in \mathbb{R}^p \end{aligned} \quad (7.3)$$

an extension to the balanced truncation method has been proposed. First, one needs to compute Gramians that quantify how controllable and observable the states are. One way to do this is to compute empirical Gramians. Let us assume that (7.3) is stable, and that  $x_0$  and  $u_0$  is an equilibrium point, i.e.,  $f(x_0, u_0) = 0$ .

The *empirical reachability Gramian* over the time interval  $[0, T]$  is defined by

$$P(T) = \sum_{i=1}^m \sum_{j=1}^r \sum_{k=1}^s \frac{1}{r s c_k^2} \int_0^T \Phi_{ijk}(t) dt,$$

where  $\Phi_{ijk}(t) \in \mathbb{R}^{n \times n}$  is given by  $\Phi_{ijk}(t) = (x_{ijk}(t) - x_{ijk,ss})(x_{ijk}(t) - x_{ijk,ss})^T$ . Here  $x_{ijk}(t)$  is the state of the system (7.3) corresponding to the input  $u(t) = c_k T_j e_i v(t) + u_0$ . Variables with the subscript  $ss$  denote constant steady-state values that should be subtracted. The constants  $c_k$  correspond to excitation sizes, the orthogonal matrices  $T_j \in \mathbb{R}^{m \times m}$  denote excitation directions, and  $e_i$  are the unit vectors in  $\mathbb{R}^m$ . Hence, the empirical Gramian is found by means of using a set of training inputs  $u(t)$  that are quantified by choosing  $c_k, T_j, e_i$ , and  $v(t)$ . We have the following theorem for linear systems.

**Theorem 8.** Assume (7.3) is a linear system and that  $v(t) = \delta(t)$  and  $x_{ijk,ss} = 0$ . Then the empirical reachability Gramian is equal to the regular reachability Gramian.

We can define the *empirical observability Gramian* over the time interval  $[0, T]$  in a similar way. We have that

$$Q(T) = \sum_{j=1}^r \sum_{k=1}^s \frac{1}{rsc_k^2} \int_0^T T_j \Psi_{jk}(t) T_j^T dt,$$

here  $\Psi_{jk}(t) \in \mathbb{R}^{n \times n}$ , and the entries  $\Psi_{jk}^{il}$ ,  $i, l = 1, \dots, n$ , are given by  $\Psi_{jk}^{il}(t) = (y_{ijk}(t) - y_{ijk,ss})^T (y_{ljk}(t) - y_{ljk,ss})$ . Here  $y_{ijk}(t)$  is the output of the system (7.3) when the initial state is  $x(0) = c_k T_j e_i + x_0$  and  $u(t) = u_0$ , where the constants  $c_k$  correspond to excitation sizes, the orthogonal matrices  $T_j \in \mathbb{R}^{n \times n}$  denote excitation directions, and  $e_i$  are the unit vectors in  $\mathbb{R}^n$ .

**Theorem 9.** Assume (7.3) is a linear system and  $y_{ljk,ss} = 0$ . Then the empirical observability Gramian is equal to the regular observability Gramian.

Just as for regular Gramians, we can find a linear coordinate transformation  $x = T\bar{x}$  that balances the empirical Gramians,

$$\bar{P}(T) = \bar{Q}(T) = \Sigma = \begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_n \end{pmatrix},$$

and the sizes of  $\sigma_i$  can be used to find initial guesses of approximation order  $r$ . The truncation of the nonlinear system is then done as follows. Let

$$\begin{aligned} T &= (t_1 \ t_2 \ \dots \ t_n) \in \mathbb{R}^{n \times n}, & V &= (t_1 \ t_2 \ \dots \ t_r) \in \mathbb{R}^{n \times r} \\ T^{-1} &= (s_1 \ s_2 \ \dots \ s_n)^T \in \mathbb{R}^{n \times n}, & W &= (s_1 \ s_2 \ \dots \ s_r) \in \mathbb{R}^{n \times r}. \end{aligned}$$

The reduced order system (using a Petrov-Galerkin projection) is given by

$$\begin{aligned} \dot{z} &= f_r(z, u) = W^T f(Vz, u), \quad z(t) \in \mathbb{R}^r, \ u(t) \in \mathbb{R}^m \\ y_r &= h_r(z, u) = h(Vz, u), \quad y(t) \in \mathbb{R}^p. \end{aligned} \tag{7.4}$$

In general, nothing can be guaranteed about the closeness of the dynamics of (7.3) and (7.4).

## 7.1 Suggested Reading

The paper [19] introduces the empirical Gramians, and motivates their construction. The paper [20] extends this concept to more general systems.

A separate theory for nonlinear Gramians and balancing is also available [21]. The theory is elegant, but it may be hard to compute the nonlinear Gramians and the balancing transformations for large systems. Hence, Taylor expansions of the nonlinear Gramians have also been considered [22].

## 8 Optimal Model Order Reduction in the Hankel Norm

In this lecture, we discuss optimal model reduction in the Hankel norm. There is no known computationally efficient solution to the optimal model reduction problem in the  $H_\infty$  norm. Since there is a connection between the Hankel norm and the  $H_\infty$  norm, the presented method is still interesting to us, and gives a lot of insight. The analysis also establishes a link to SVD.

### 8.1 The Hankel Norm and the Hankel Operator

The Hankel norm of a system  $y = \mathbf{G}u$  with transfer function  $G = (A, B, C, D) \in H_\infty$  is defined by

$$\|G\|_H^2 := \sup_{u \in L_2(\infty, 0]} \frac{\int_0^\infty y(t)^2 dt}{\int_{-\infty}^0 u(t)^2 dt}, \quad \text{where } y(t) = \int_{-\infty}^0 C e^{A(t-\tau)} B u(\tau) d\tau.$$

The Hankel norm tells how much signal energy can be transferred from past inputs into future outputs through the system  $G$ . In Exercise 4.4 b), we in fact showed that  $\|G\|_H = \sqrt{\lambda_{\max}(PQ)} = \sigma_1$  (the largest Hankel singular value). One defines the Hankel operator  $\Gamma_{\mathbf{G}}$  of the system  $\mathbf{G}$  by,

$$\Gamma_{\mathbf{G}} : L_2(-\infty, 0] \rightarrow L_2[0, \infty) : (\Gamma_{\mathbf{G}} u)(t) = \int_{-\infty}^0 C e^{A(t-\tau)} B u(\tau) d\tau, \quad t > 0.$$

The induced norm of  $\Gamma_{\mathbf{G}}$  is equal to the Hankel norm of  $\mathbf{G}$ ,  $\|\Gamma_{\mathbf{G}}\| = \|G\|_H$ .

A nice property of  $\Gamma_{\mathbf{G}}$  is that it has finite rank, with the rank being equal to the minimal number of states required to realize the input-output map  $\mathbf{G}$ ,

$$\text{Rank } \Gamma_{\mathbf{G}} = n.$$

Furthermore, one can make an SVD of  $\Gamma_{\mathbf{G}}$ , with the dyadic expansion

$$(\Gamma_{\mathbf{G}} u)(t) = \sum_{i=1}^n \sigma_i u_i(t) (v_i, u)_{L_2(-\infty, 0]}, \quad (8.1)$$

where  $\sigma_i$  are the Hankel singular values of  $\mathbf{G}$ , and the singular vectors are  $v_i \in L_2(-\infty, 0]$ ,  $u_i \in L_2[0, \infty)$ . Maybe one would think it is more natural to consider the operator  $\mathbf{G}$  directly,

$$\mathbf{G} : L_2(-\infty, \infty) \rightarrow L_2(-\infty, \infty) : (\mathbf{G}u)(t) = \int_{-\infty}^t C e^{A(t-\tau)} B u(\tau) d\tau + D u(t).$$

This operator is not of finite rank, however, and has no finite dyadic expansion such as (8.1). This makes further analysis difficult.

Since one can make an SVD of  $\Gamma_{\mathbf{G}}$ , one can essentially apply the Schmidt-Mirsky theorem to prove that

$$\|\Gamma_{\mathbf{G}} - \Gamma_{\mathbf{G}_r}\| = \|G - G_r\|_H \geq \sigma_{r+1}, \quad (8.2)$$

for all  $\mathbf{G}_r$  of order  $r$ . From the definition of the Hankel norm it should also be clear that  $\|G\|_H \leq \|G\|_\infty$  for all  $G \in H_\infty$ . Hence, the bound (8.2) also holds for the  $H_\infty$  norm, as has already been stated in earlier lectures.

### Relation between the Hankel norm and the $L_\infty$ norm

To proceed, we need to introduce more system spaces, namely  $L_\infty$ ,  $H_\infty^-$ , and  $H_\infty^-(r)$ . A transfer function  $G$  belongs to  $L_\infty$  if, and only if,

$$\|G\|_\infty := \sup_{\omega} \|G(j\omega)\|$$

is finite. Hence  $G(s)$  may have poles in both  $\mathbb{C}_+$  and  $\mathbb{C}_-$ , but not on the imaginary axis. A transfer function  $G(s)$  belongs to  $H_\infty^-$  if, and only if,  $G(-s) \in H_\infty$ . Hence, unstable systems with no stable poles belong to  $H_\infty^-$ . Such systems are called *anti-stable*. A transfer function belongs to  $H_\infty^-(r)$  if it belongs to  $L_\infty$  and it has at most  $r$  poles in the complex left half plane,  $\mathbb{C}_-$ .

We can now state the following fundamental theorems.

**Theorem 10** (Nehari). *Suppose that  $G \in H_\infty$ , and  $F \in H_\infty^-$ . Then  $G - F \in L_\infty$ , and*

$$\min_{F \in H_\infty^-} \|G - F\|_\infty = \|G\|_H (= \sigma_1).$$

Hence, one interpretation of  $\|G\|_H$  is that it is the minimum distance between  $G$  and an anti-stable system, measured in the  $L_\infty$  norm.

The following important extension to Nehari's theorem is available.

**Theorem 11** (Adamjan-Arov-Krein). *Suppose that  $G \in H_\infty$  and  $Q \in H_\infty^-(r)$ . Then  $G - Q \in L_\infty$ , and*

$$\min_{Q \in H_\infty^-(r)} \|G - Q\|_\infty = \sigma_{r+1}, \quad (8.3)$$

where  $\sigma_{r+1}$  is the  $(r + 1)$ -th largest Hankel singular value of  $G$ .

Denote an optimal  $Q \in H_\infty^-(r)$  by  $Q^*$ , i.e.,  $\|G - Q^*\|_\infty = \sigma_{r+1}$ . A stable/anti-stable decomposition of  $Q^*$  can be performed such that  $Q^* = G_r + F$ , where  $G_r \in H_\infty$ ,  $\deg G_r \leq r$ , and  $F \in H_\infty^-$ . It follows that

$$\|G - G_r\|_\infty = \|G - Q^* + F\|_\infty \leq \|G - Q^*\|_\infty + \|F\|_\infty = \sigma_{r+1} + \|F\|_\infty. \quad (8.4)$$

If we can show that the  $L_\infty$  norm of the unstable term  $F$  is small, then the stable part of  $Q^*$  can be a good candidate for  $H_\infty$  model reduction. A bound on  $\|F\|_\infty$  is derived in Section 8.3.

Note that the stable part  $G_r$  of  $Q^*$  solves the optimal Hankel norm approximation problem, since

$$\|G - G_r\|_H = \min_{F \in H_\infty^-} \|G - G_r - F\|_\infty = \min_{Q \in H_\infty^-(r)} \|G - Q\|_\infty = \sigma_{r+1},$$

and we know that a lower bound of the problem is  $\sigma_{r+1}$ , see (8.2). Note also that the direct term  $D_r$  of  $G_r$  is not determined by Hankel norm approximation since it does not appear in the Hankel operator.

## 8.2 State-Space Formulas for Construction of $Q^*$

In this section, we give state-space formulas for the computation of an optimal Hankel approximation  $Q^*$  whose existence is guaranteed by Theorem 11.

Suppose the system  $G$  has realization  $G = (A, B, C, D) \in H_\infty$  of order  $n$ , and that it is square ( $m = p$ ). The formulas for non-square systems are slightly more complicated, and are given in [4]. Assume the realization of  $G$  is such that the Gramians take the form

$$P = \begin{bmatrix} P_1 & 0 \\ 0 & \sigma_{r+1} I_l \end{bmatrix}, \quad Q = \begin{bmatrix} Q_1 & 0 \\ 0 & \sigma_{r+1} I_l \end{bmatrix},$$

where  $l$  is the multiplicity of the singular value  $\sigma_{r+1}$ . (For instance, permute a balanced realization.) One can choose a permuted balanced realization of  $G$ , for example. Conformably to  $P, Q$ , partition the state-space realization of  $G$ ,

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}, \quad C = [C_1 \quad C_2].$$

One can prove (Exercise 8.3) that under the given assumptions, there is a unitary matrix  $U \in \mathbb{R}^{m \times m}$  (we assumed  $p = m$ ),  $U^T U = I$ , such that

$$B_2 = -C_2^T U. \quad (8.5)$$

Also define

$$E_1 := Q_1 P_1 - \sigma_{r+1}^2 I.$$

A realization of the optimal approximation  $Q^*$  is now given by

$$\begin{aligned} \hat{A} &= E_1^{-1}(\sigma_{r+1}^2 A_{11}^T + Q_1 A_{11} P_1 - \sigma_{r+1} C_1^T U B_1^T) \\ \hat{B} &= E_1^{-1}(Q_1 B_1 + \sigma_{r+1} C_1^T U) \\ \hat{C} &= C_1 P_1 + \sigma_{r+1} U B_1^T \\ \hat{D} &= D - \sigma_{r+1} U. \end{aligned}$$

The transfer function  $Q^* = (\hat{A}, \hat{B}, \hat{C}, \hat{D})$  is of order  $n - l$  and belongs to  $H_\infty^-(r)$ . In fact, it has exactly  $r$  stable poles and  $n - r - l$  unstable poles. We verify below that  $E(s) := G(s) - Q^*(s)$  satisfies

$$E(s) \sim E(s) = \sigma_{r+1}^2 I,$$

where  $E(s) \sim = E(-s)^T$ , and thus  $\|E\|_\infty = \sigma_{r+1}$ . Note that  $\|E(j\omega)\| = \sigma_{r+1}$  for all  $\omega$ .

The given  $Q^*$  solves the optimization problem in Theorem 11. This is generally not the unique solution, however. All the solutions (also in the non-square  $\mathbf{G}$  case) are parameterized in [4] using a linear fractional transformation.

To verify that  $E(s) \sim E(s) = \sigma_{r+1}^2 I$ , we use the following lemma.

**Lemma 1.** *Let  $G = (A, B, C, D)$ . If there is a symmetric matrix  $P$  such that*

$$\begin{aligned} PA^T + AP + BB^T &= 0 \\ PC^T + BD^T &= 0 \\ D^T D &= \gamma^2 I, \end{aligned}$$

*then  $G(s) \sim G(s) = \gamma^2 I$  and  $G \in L_\infty$ . Conversely, if  $G = (A, B, C, D)$  is minimal and  $G(s) \sim G(s) = \gamma^2 I$ , then there is a symmetric matrix  $P$  that satisfies the above conditions.*

If we realize  $E(s) = G(s) - Q^*(s)$  with

$$A_e = \begin{bmatrix} A & 0 \\ 0 & \hat{A} \end{bmatrix}, \quad B_e = \begin{bmatrix} B \\ \hat{B} \end{bmatrix}, \quad C_e = [C \quad -\hat{C}], \quad D_e = D - \hat{D},$$

the assumptions of Lemma 1 are satisfied using

$$P_e = \begin{bmatrix} P_1 & 0 & I \\ 0 & \sigma_{r+1}^2 I_l & 0 \\ I & 0 & E_1^{-1} Q_1 \end{bmatrix}$$

as  $P$ , and  $\gamma = \sigma_{r+1}$ .

### 8.3 $H_\infty$ Error Bounds

In this section, we make the standing assumption that the Hankel singular values of  $G$  are distinct. This simplifies the notation. The results can be strengthened if some singular values are identical, just as in balanced truncation.

Using the error bound for balanced truncation, the  $H_\infty$  norm of  $G$  can be bounded using the Hankel singular values. If Theorem 4 is applied and all states are truncated, we obtain

$$\|G - G(\infty)\|_\infty \leq 2 \sum_{i=1}^n \sigma_i.$$

Optimal Hankel norm approximation can be used to improve this general bound.

If we construct the optimal Hankel approximation  $G_{n-1}$ , we obtain  $\|G - G_{n-1}\|_\infty = \sigma_n$ . The Gramians of the system  $G_{n-1}$  are given by  $P_1 E_1$  and  $E_1^{-1} Q_1$ , and hence the Hankel singular values of  $G_{n-1}$  are  $\sqrt{\lambda_i(P_1 E_1 E_1^{-1} Q_1)} = \sqrt{\lambda_i(P_1 Q_1)} = \sigma_i(G)$ ,  $i = 1, \dots, n-1$  (the remaining Hankel singular values). We can repeat this procedure on  $G_{n-1}$ , and remove Hankel singular values one by one. In the end, only a direct term  $\tilde{D} = G_0$  remains. Applying the triangle inequality, we have

$$\|G - \tilde{D}\|_\infty = \|(G - G_{n-1}) + (G_{n-1} - G_{n-2}) + \dots + (G_1 - \tilde{D})\|_\infty \leq \sum_{i=1}^n \sigma_i. \quad (8.6)$$

Hence, there always exists a constant  $\tilde{D}$  that will shift the Nyquist diagram of  $G$  so that it is contained in a circle of radius equal to the sum of the distinct Hankel singular values. If we use  $G(\infty)$  as the direct term, then we need to make the radius a factor two larger, in general.

Let us now return to the bound (8.4). We want an a priori bound on  $\|F\|_\infty$  since that yields a bound on  $\|G - G_r\|_\infty$ . We have that  $F \in H_\infty^-$  is of order  $n - r - 1$  (remember the multiplicity  $l = 1$ ). Furthermore,  $F^\sim \in H_\infty$  and  $\|F\|_\infty = \|F^\sim\|_\infty$ . The following lemma can be derived.

**Lemma 2.** Assume that  $\sigma_i(G)$  are distinct, and let the optimal  $L_\infty$  norm approximation  $Q \in H_\infty^-(r)$  of  $G$  be  $Q^* = G_r + F$ , such that  $G_r, F^\sim \in H_\infty$  and  $\deg G_r \leq r$ . Then

$$\sigma_i(F^\sim) \leq \sigma_{i+r+1}(G), \quad i = 1, \dots, n - r - 1.$$

Using (8.6) and Lemma 2 on  $F^\sim$ , there is a  $\tilde{D}$  such that  $\|F - \tilde{D}\|_\infty \leq \sum_{i=r+2}^n \sigma_i(G)$ . It follows that this  $\tilde{D}$  yields

$$\|G - G_r - \tilde{D}\|_\infty = \|G - G_r - F + F - \tilde{D}\|_\infty \leq \|G - Q^*\|_\infty + \|F - \tilde{D}\|_\infty = \sigma_{r+1} + \|F - \tilde{D}\|_\infty \leq \sum_{i=r+1}^n \sigma_i.$$

We sum the results up in the following theorem.

**Theorem 12.** Suppose  $G \in H_\infty$  is of order  $n$ , and has distinct Hankel singular values  $\sigma_i$ . Then for all approximations  $G_r \in H_\infty$  of order  $r$ , it holds that

$$\sigma_{r+1} \leq \|G - G_r\|_H \leq \|G - G_r\|_\infty.$$

Furthermore, there exists  $G_r \in H_\infty$  of order  $r$ , given by the stable part of the transfer function  $Q^* = (\hat{A}, \hat{B}, \hat{C}, \hat{D}) \in H_\infty^-(r)$ , and  $\tilde{D}$  such that

$$\begin{aligned} \|G - G_r\|_H &= \|\Gamma_G - \Gamma_{G_r}\| = \sigma_{r+1} \\ \|G - G_r - \tilde{D}\|_\infty &\leq \sum_{i=r+1}^n \sigma_i. \end{aligned}$$

In particular, when  $r = n - 1$ , a solution to the optimal  $H_\infty$  norm approximation problem is obtained.



## 8.4 Suggested Reading

[4, Chapter 10] gives a thorough and readable presentation of Hankel norm approximation. All the stated results above are proved there.

The Nehari and Adamjan-Arov-Krein theorems are proven in [23] (in discrete time).

## 8.5 Exercises

### EXERCISE 8.1 (Antoulas [2004])

Let the system  $\mathbf{G}$  have transfer function

$$G(s) = \frac{-s + 1}{s^6 + 3s^5 + 5s^4 + 7s^3 + 5s^2 + 3s + 1}.$$

Compute optimal Hankel norm approximations  $G_r$  of  $G$ , and the corresponding  $Q^*$  for  $r = 3$  and  $r = 5$ . Also plot the Bode diagrams of the approximation errors  $G - Q^*$  and  $G - G_r$ , and compute the  $L_\infty/H_\infty$  norm of the errors.

### EXERCISE 8.2

Derive a state-space algorithm that performs a stable/anti-stable decomposition of state-space realizations  $G = (A, B, C, D) \in L_\infty$ .

### EXERCISE 8.3

Prove that there is always a unitary  $U$  that satisfies (8.5). (Hint: Show first that  $B_2 B_2^T = C_2^T C_2$ .)

### EXERCISE 8.4

Prove that all (square) transfer functions  $G \in H_\infty$  of order  $n$  (with distinct Hankel singular values  $\sigma_i$ ) can be expanded as

$$G(s) = \tilde{D} + \sigma_1 E_1(s) + \sigma_2 E_2(s) + \dots + \sigma_n E_n(s),$$

where  $E_i \in H_\infty$  are all-pass, i.e.,  $E_i(s) \sim E_i(s) = I$ , and  $G_r(s) = \tilde{D} + \sigma_1 E_1(s) + \dots + \sigma_r E_r(s)$  is of order  $r$ .

### EXERCISE 8.5 (Extra)

Prove that the following lower bound on the weighted approximation criterion holds,

$$\min_{G_r \in H_\infty, \deg G_r \leq r} \|W_o(G - G_r)W_i\|_\infty \geq \sigma_{r+1}([M_o G M_i]_+),$$

where  $\sigma_{r+1}([M_o G M_i]_+)$  is computed as follows: Compute the (unstable) spectral factors  $M_o$  and  $M_i$ ,

$$M_o(s) \sim M_o(s) = W_o(s) \sim W_o(s), \quad M_i(s) M_i(s) \sim = W_i(s) W_i(s) \sim,$$

such that  $M_o(s), M_o(s)^{-1}, M_i(s), M_i(s)^{-1}$  have their poles in the open right half plane  $\mathbb{C}_+$ . Here  $[P]_+$  denotes the stable part of  $P$ .

# BIBLIOGRAPHY

- [1] H. Khalil, *Nonlinear Systems*, ser. Pearson Education. Prentice Hall, 2002.
- [2] R. V. Polyuga and A. J. van der Schaft, “Effort- and flow-constraint reduction methods for structure preserving model reduction of port-Hamiltonian systems,” *Systems & Control Letters*, vol. 61, no. 3, pp. 412 – 421, 2012.
- [3] M. Fiedler, “Algebraic connectivity of graphs,” *Czechoslovak Mathematical Journal*, vol. 23, no. 2, pp. 298–305, 1973.
- [4] M. Green and D. J. N. Limebeer, *Linear Robust Control*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1995.
- [5] A. van der Schaft, “Port-Hamiltonian systems: an introductory survey,” in *Proceedings of the International Congress of Mathematicians Vol. III*, M. Sanz-Sole, J. Soria, J. Varona, and J. Verdera, Eds., no. suppl 2. European Mathematical Society Publishing House (EMS Ph), 2006, pp. 1339–1365.
- [6] W. Ren, R. W. Beard, and E. M. Atkins, “A survey of consensus problems in multi-agent coordination,” in *Proceedings of the 2005 American Control Conference*, June 2005, pp. 1859–1864 vol. 3.
- [7] T. Ishizaki, K. Kashima, J. Imura, and K. Aihara, “Model reduction and clusterization of large-scale bidirectional networks,” *IEEE Transactions on Automatic Control*, vol. 59, no. 1, pp. 48–63, Jan 2014.
- [8] A. C. Antoulas, *Approximation of Large-Scale Dynamical Systems*. Philadelphia: SIAM, 2005.
- [9] G. Scarciootti and A. Astolfi, “Nonlinear model reduction by moment matching,” *Foundations and Trends in Systems and Control*, vol. 4, no. 3-4, pp. 224–409, 2017.
- [10] S. Skogestad, “Simple analytic rules for model reduction and PID controller tuning,” *Journal of Process Control*, vol. 13, pp. 291–309, 04 2004.
- [11] B. Bernhardsson and K. J. Åström, “Control system synthesis: Pole placement design,” Lecture Slides, Lund University, 2016.
- [12] J. C. Doyle, B. A. Francis, and A. R. Tannenbaum, *Feedback Control Theory*. Prentice Hall Professional Technical Reference, 1991.
- [13] B. Moore, “Principal component analysis in linear systems: Controllability, observability, and model reduction,” *IEEE Transactions on Automatic Control*, vol. 26, no. 1, pp. 17–32, February 1981.

- [14] R. A. Horn, *Topics in Matrix Analysis*. New York, NY, USA: Cambridge University Press, 1986.
- [15] D. G. Luenberger, *Optimization by Vector Space Methods*, 1st ed. New York, NY, USA: John Wiley & Sons, Inc., 1997.
- [16] K. Zhou and J. C. Doyle, *Essentials of Robust Control*. Prentice-Hall, 1998.
- [17] G. Obinata and B. D. O. Anderson, *Model reduction for control system design*. Springer New York, 2000.
- [18] M. G. Safonov and R. Y. Chiang, "Model reduction for robust control: A schur relative error method," *International Journal of Adaptive Control and Signal Processing*, vol. 2, no. 4, pp. 259–272, 1988.
- [19] S. Lall, J. E. Marsden, and S. Glavaski, "A subspace approach to balanced truncation for model reduction of nonlinear control systems," *International Journal of Robust and Nonlinear Control*, vol. 12, no. 6, pp. 519–535.
- [20] J. Hahn, T. F. Edgar, and W. Marquardt, "Controllability and observability covariance matrices for the analysis and order reduction of stable nonlinear systems," *Journal of Process Control*, vol. 13, no. 2, pp. 115 – 127, 2003.
- [21] J. Scherpen, "Balancing for nonlinear systems," *Systems & Control Letters*, vol. 21, no. 2, pp. 143 – 153, 1993.
- [22] A. J. Krener, *Reduced Order Modeling of Nonlinear Control Systems*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 41–62.
- [23] N. Young, *An Introduction to Hilbert Space*. Cambridge University Press, 1988.