4 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 yield good reduced models 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.

4.1 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 $\mathcal{R}$. From this interpretation, it can be concluded that a system G is reachable if and only if all singular values of $P(T)$ are strictly positive for some T. Stated differently, G is reachable if and only if $P(T)$ is positive definite (denoted as $P(T) > 0$) for some T.

When 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 G is reachable). The reachable subspace can again be obtained from the reachability Gramian and the equality

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

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

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

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

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 G (i.e., A Hurwitz) and infinite horizon ($T \to \infty$), they satisfy the following *algebraic* Lyapunov equations

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

As a result, Gramians are easier to compute for an infinite horizon. Moreover, the connection to these Lyapunov equations provides the key to proving many properties of reduced-order systems obtained by balanced truncation and balanced singular perturbation, as will be shown later.

4.2 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

$$A^TX + XA + H = 0 \tag{4.1}$$

has a unique solution X when $\lambda_i(A) + \bar{\lambda}_j(A) \neq 0$ for all i, j, where $\bar{\lambda}_j$ denotes the complex conjugate of the eigenvalue λ_j. Using this condition, it immediately follows that (4.1) 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) $X = \int_0^\infty e^{A^T t} H e^{At} dt$;

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 (4.1) 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

$$\dot{V}(x) = \dot{x}^T X x + x^T X x = (Ax)^T X x + x^T X (Ax) = x^T (A^T X + XA) x = -x^T H x, \tag{4.2}$$

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 $X > 0$, $H \geq 0$ and (A, H) is detectable.

4.3 Balanced Truncation and Singular Perturbation

When the realization of G is balanced, the semi axes of the reachability and observability ellipsoids, $\mathcal{R}$ and $\mathcal{O}$, 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 G_r that

$$\inf_{G_r \in H_\infty, \deg G_r \leq r} \|G - G_r\|_\infty \geq \sigma_{r+1}, \tag{4.3}$$

where σ_{r+1} is the $(r+1)$-th largest Hankel singular value of 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 (4.3).

For the sake of convenience, assume that the realization (A, B, C, D) of G is balanced as described in Section 3.4. The reachability and observability Gramians over the infinite time horizon then satisfy

$$A\Sigma + \Sigma A^T + BB^T = 0$$
$$A^T \Sigma + \Sigma A + C^T C = 0$$

where Σ 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

$$\Sigma_1 = \begin{bmatrix} \sigma_1 I_{r_1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_l I_{r_l} \end{bmatrix}, \quad \Sigma_2 = \begin{bmatrix} \sigma_{l+1} I_{r_{l+1}} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_m I_{r_m} \end{bmatrix} \tag{4.4}$$

and $n = r_1 + \ldots + r_m$, $r = r_1 + \ldots + r_l$, and $\sigma_i \neq \sigma_j$, $i \neq j$. This notation is introduced to exploit when singular values happen to have a multiplicity greater than one, $r_i > 1$. Conformably to Σ_1, Σ_2, the realization is partitioned into

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}, \quad C = \begin{bmatrix} C_1 & C_2 \end{bmatrix}, \tag{4.5}$$

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 6. *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 G_r with Gramian Σ_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 under certain cases it holds that the maximum error is achieved at frequency zero, 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 7. *Suppose (A, B, C, D) is a balanced realization and that*

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

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 G_r with Gramian Σ_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 error bound on $\|G - G_r\|_\infty$ holds in both cases, the question is whether one wants a good model match at low or high frequencies.

4.4 Suggested Reading

The relation between Gramians and reachability and observability properties can be found, for example, in *Approximation of Large-Scale Dynamical Systems* by Antoulas, whereas the discussion on Lyapunov equations is taken from Section 3.8 of *Robust and Optimal Control* by Zhou, Doyle, and Glover.

The stability properties and error bounds for regular balanced truncation are derived in Sections 9.2 and 9.4–9.5 of *Linear Robust Control*. The fundamental lower bounds will be derived later in the course.

4.5 Exercises

EXERCISE 4.1

Consider the model 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.

a) Load the matrices (A, B, C, D) from the file *building.mat*. Is this model (asymptotically) stable, reachable, and observable?

b) Compute and plot the Hankel singular values. Based on these Hankel singular values, what would be suitable reduction orders?

c) 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?

d) Compute the step response (i.e., the response to an input u that satisfies $u(t) = 1$ for all $t \geq 0$ and $u(t) = 0$ for all $t < 0$ and zero initial conditions) for the high-order system and the two reduced-order approximations. Compare the results.

e) Verify that the reduced-order models are asymptotically stable, reachable and observable. Finally, compute the error bound and verify that it is satisfied.

EXERCISE 4.2

Let the Gramians of $G = (A, B, C, D)$ be P and Q (any coordinates), and let G_r be a truncated balanced realization of G. Let v_i and w_i satisfy

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

and $V = \begin{bmatrix} v_1 & \dots & v_r \end{bmatrix}$ and $W = \begin{bmatrix} w_1 & \dots & w_r \end{bmatrix}$. Furthermore, normalize v_i, w_i such that $W^T V = I$. Show that the system G_p that is realized by $(W^T AV, W^T B, CV, D)$ has the same input-output behavior as G_r, i.e., $\|G_r - G_p\|_\infty = 0$.