4 Balanced Truncation and Frequency-Weighted Extensions

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. Many times other approximation criteria are important, however. We will show that variations of the balanced truncation method can be used to make the approximation criteria $\|G^{-1}(G - G_r)\|_\infty$ and $\|W_o(G - G_r)W_i\|_\infty$ small.

4.1 Balanced Truncation and Singular Perturbation

When the realization of $G$ is balanced, the semi axes of the reachability and observability ellipsoids, $R$ and $O$, 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

$$\inf_{G_r \in H_\infty, \deg G_r \leq r} \|G - G_r\|_\infty \geq \sigma_{r+1}, \quad (4.1)$$

where $\sigma_{r+1}$ 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 (4.1).

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 $\Sigma$ 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

$$\Sigma_1 = \begin{bmatrix} \sigma_1 I_{r_1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_l I_{r_l} \end{bmatrix}, \quad \Sigma_2 = \begin{bmatrix} \sigma_{l+1} I_{r_{l+1}} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_m I_{r_m} \end{bmatrix} \quad (4.2)$$

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 $\Sigma_1, \Sigma_2$, the realization is partitioned into

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}, \quad C = \begin{bmatrix} C_1 & C_2 \end{bmatrix}, \quad (4.3)$$

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 \(\Sigma_1\). 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 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

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

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 \(G_r\) 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 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.

4.2 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.

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 simple 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 (4.1) can be generalized to the frequency-weighted case,

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

(4.4)
The weighted Gramians to control the state of the reduced model and singular values can be computed. The corresponding realization can also be truncated to obtain the initial state of such that $\sqrt{W}$ the ones in Section 4.1 are obtained.

As we see next and in the provided references, so that stability is maintained and error bounds like the equations that ensure stability.

The weighted Gramians do not generally satisfy Lyapunov $G$ generally gives a small error that is obtained by the following procedure.

Let us compute the reachability and observability Gramians $P, Q$ for the system $G$ that 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 like the ones in Section 4.1 are obtained.
Balanced stochastic truncation

Many times it is important to make the relative error criterion
\[ \|G^{-1}(G - G_r)\|_\infty \] (4.6)
small. For example, if one is interested in matching the Bode plots of \( G \) and \( G_r \). The criterion (4.6) 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
\[
20 \log_{10} |G_r(j\omega)/G(j\omega)| \leq 8.69 |\Delta(j\omega)| \text{ dB, } \quad |\text{phase } G(j\omega) - \text{phase } G_r(j\omega)| \leq |\Delta(j\omega)| \text{ rad.}
\]

We can apply frequency-weighted balanced truncation to the problem (4.6) 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
\[
\tilde{A} = \begin{bmatrix} A & 0 \\ -BD^{-1}C & A - BD^{-1}C \end{bmatrix}, \quad \tilde{B} = \begin{bmatrix} B \\ 0 \end{bmatrix}, \quad \tilde{C} = [D^{-1}C \quad D^{-1}C].
\] (4.7)
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 8.** Suppose \( G, G^{-1} \in H_\infty \), and let \( G_r \) be a truncated realization of \( G \) that has been balanced with the weighted Gramians of (4.7). Then \( G_r \) is stable and minimum phase, \( G_r, G_r^{-1} \in H_\infty \), and satisfies
\[
\|G^{-1}(G - G_r)\|_\infty \leq m \prod_{i=l+1}^{m} \left( 1 + 2\sigma_i(\sqrt{1 + \sigma_i^2} + \sigma_i) \right) - 1,
\]
\[
\|G_r^{-1}(G - G_r)\|_\infty \leq m \prod_{i=l+1}^{m} \left( 1 + 2\sigma_i(\sqrt{1 + \sigma_i^2} + \sigma_i) \right) - 1,
\]
where the singular values \( \sigma_i \) are partitioned as in (4.2).

This method can be extended to the case when \( G \) has zeros in the right half plane, i.e., when \( G^{-1} \) is not stable. This more general method often goes under the name balanced stochastic truncation.

LMI-based frequency weighted balanced truncation∗

This section will be written and will show how weighted balanced truncation can be modified so that error bounds and stability hold for arbitrary weights. This requires theory for linear matrix inequalities (LMIs) which can be solved using convex optimization tools. The following paper introduces the method:


4.3 Suggested Reading

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.

Frequency-weighted reduction is treated in Section III in the survey paper *Controller Reduction: Concepts and Approaches*. Many references are also given in that paper. References to papers that discuss balanced stochastic truncation are given in *Linear Robust Control*. 

26
### 4.4 Exercises

**EXERCISE 4.1**

Let the model $G$ be given by

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

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 4.2** (Safonov and Chiang [1988])

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 4.3**

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

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

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 (4.7), 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 4.4** (Extra)

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

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

and $V = [v_1 \ldots v_r]$ and $W = [w_1 \ldots w_r]$. 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., $\|G_r - G_p\|_\infty = 0$.  

27