5 Frequency-Weighted Balanced Truncation and Controller Reduction

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

Finally, a selection of available methods for the reduction of nonlinear models is discussed.

5.1 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.

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

$$\inf_{G_r \in H_\infty,\, \deg G_r \le r} \|W_o(G-G_r)W_i\|_\infty \ge \sigma_{r+1}([M_o G M_i]_+), \tag{5.1}$$

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 $M_o, M_o^{-1}, M_i, M_i^{-1}$ have their poles in the 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 (5.1).

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

$$\tilde{A} = \begin{bmatrix} A & 0 & BC_i \\ B_o C & A_o & 0 \\ 0 & 0 & A_i \end{bmatrix}, \quad \tilde{B} = \begin{bmatrix} BD_i \\ 0 \\ B_i \end{bmatrix}, \quad \tilde{C} = \begin{bmatrix} D_o C & C_o & 0 \end{bmatrix}.$$

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

$$P := \begin{bmatrix} I_n & 0 \end{bmatrix} \tilde{P} \begin{bmatrix} I_n \\ 0 \end{bmatrix}, \quad Q := \begin{bmatrix} I_n & 0 \end{bmatrix} \tilde{Q} \begin{bmatrix} I_n \\ 0 \end{bmatrix}. \tag{5.2}$$

These are the blocks of $\tilde{P}$ and $\tilde{Q}$ that correspond to the states in G. The other states belong to W_o and W_i.

The weighted Gramians have an interpretation of how controllable and observable the states of G are *as seen through* the weights W_i and W_o. For example, if the initial states of the filters are zero and the initial state of G is x_0, the energy of the output signal of G filtered through 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 G in $W_o G W_i$ 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 W_o and W_i have been chosen so as to minimize the energy in u (they are free parameters).

Now the weighted Gramians (5.2) can be balanced just as the regular Gramians were in Section 3.4 and singular values can be computed. The corresponding realization can also be truncated to obtain the reduced model G_r. (If $G(\infty) = D \neq 0$, then one should use the approximation $G_r + D$.) This G_r generally gives a small error $\|W_o(G-G_r)W_i\|_\infty$, if the the order r is reasonably chosen. However, this method is truly a heuristic and it can fail. There are examples where G_r even becomes unstable when 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 like the ones in Section 4.3 are obtained.

Balanced stochastic truncation

Many times it is important to make the relative error criterion

$$\|G^{-1}(G - G_r)\|_\infty \tag{5.3}$$

small. For example, if one is interested in matching the Bode plots of G and G_r. The criterion (5.3) 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

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

We can apply frequency-weighted balanced truncation to the problem (5.3) 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

$$\tilde{A} = \begin{bmatrix} A & 0 \\ -BD^{-1}C & A - BD^{-1}C \end{bmatrix}, \quad \tilde{B} = \begin{bmatrix} B \\ 0 \end{bmatrix}, \quad \tilde{C} = \begin{bmatrix} D^{-1}C & D^{-1}C \end{bmatrix}. \tag{5.4}$$

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 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 (5.4). Then G_r is stable and minimum phase, $G_r, G_r^{-1} \in H_\infty$, and satisfies*

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

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

where the singular values σ_i are partitioned as in (4.4).

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*.

5.2 Controller Reduction

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

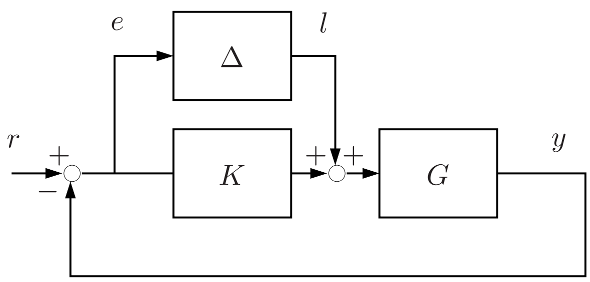

Figure 5.1: A feedback-control system. The plant G is controlled by the reduced controller $K_r = K + \Delta$.

Consider the closed-loop system in Figure 5.1. A plant G is controlled by a feedback controller K_r (or K if $\Delta = 0$). If K is designed using optimal control methods (H_2/H_∞), the order of the controller typically is at least as large as the plant G. In order to simplify implementation of K, we would like to obtain a low-order controller K_r. To reduce the order of K, it is generally not a good idea to solve the approximation problem using the criterion $\|\Delta\|_\infty = \|K - K_r\|_\infty$, however. The closed-loop behavior should be taken into account.

Robustness consideration

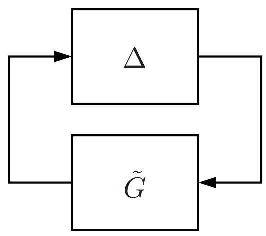

Figure 5.2: A feedback interconnection of $\tilde{G}$ and Δ.

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

Theorem 9 (Small-gain theorem). *Suppose that $\tilde{G}, \Delta \in H_\infty$. The feedback interconnection of $\tilde{G}$ and Δ in Figure 5.2 is stable if $\|\tilde{G}\Delta\|_\infty \le \|\tilde{G}\|_\infty \|\Delta\|_\infty < 1$.*

If we look at the transfer function from l to e in Figure 5.1, it is given by $e = -(I+GK)^{-1}Gl =: \tilde{G}l$. Hence, applying Theorem 9, the closed-loop system is stable if

$$\|(I + GK)^{-1}G\Delta\|_\infty < 1 \quad \text{or} \quad \|\Delta G(I + KG)^{-1}\|_\infty < 1. \tag{5.5}$$

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

Hence, to obtain a reduced-order controller 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 consideration

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, $GK(I + GK)^{-1}$ and $GK_r(I + GK_r)^{-1}$. Using a Taylor expansion, we obtain

$$GK_r(I + GK_r)^{-1} - GK(I + GK)^{-1} = (I + GK)^{-1}G(K_r - K)(I + GK)^{-1} + O(\|K_r - K\|_\infty^2). \quad (5.6)$$

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

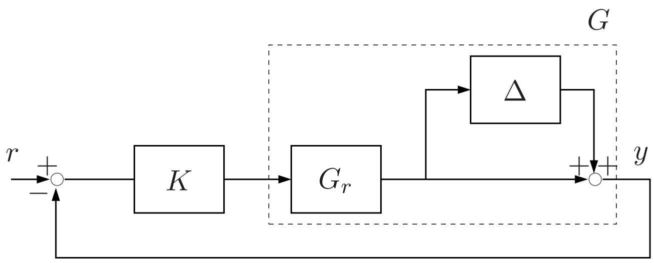

Figure 5.3: A feedback-controlled system. The perturbed reduced plant $G = (I + \Delta)G_r$ is controlled by the controller K.

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

Assume that we have a high-order plant model G. We first reduce it to obtain G_r, and design a controller K for G_r. Then we want to ensure that K also works for the original plant model G. Let us model the original model G by a multiplicative perturbation to G_r, $G = (I + \Delta)G_r$ where $\Delta = (G - G_r)G_r^{-1}$, see Figure 5.3. 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 5.3 is stable if

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

We cannot use this relation to obtain G_r since 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

$$\|(G - G_r)G_r^{-1}\|_\infty = \|G_r^{-T}(G^T - G_r^T)\|_\infty \quad (5.7)$$

small. This can be done by employing balanced stochastic truncation, see Section 5.1, 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 8.

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.

5.3 Nonlinear Model Order Reduction

Nonlinear model reduction is a difficult area, and there are not many rigorous results available. The ones we present here are some that 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

$$\dot{x} = f(x), \quad x(t) \in \mathbb{R}^n \tag{5.8}$$

is Principal Orthogonal Decomposition (POD). This is nothing but PCA (approximated with SVD) applied to some relevant trajectories of (5.8). For example, one can solve (5.8) once (or many times for different initial conditions) and construct the *snapshot matrix* X,

$$X = \begin{bmatrix} x(t_1) & x(t_2) & \dots & x(t_N) \end{bmatrix} \in \mathbb{R}^{n \times N},$$

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),

$$X = U\Sigma V^T = \begin{bmatrix} u_1 & u_2 & \dots & u_n \end{bmatrix} \Sigma V^T \approx U_r \Sigma_r V_r^T = \begin{bmatrix} u_1 & u_2 & \dots & u_r \end{bmatrix} \Sigma_r V_r^T, \quad r < n.$$

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

$$\dot{z} = f_r(z) = U_r^T f(U_r z), \quad z(t) \in \mathbb{R}^r. \tag{5.9}$$

In general, nothing can be guaranteed about the closeness of the dynamics of (5.8) and (5.9).

Empirical Gramians

For nonlinear input-output models,

$$\begin{aligned} \dot{x} &= f(x,u), \quad x(t) \in \mathbb{R}^n,\ u(t) \in \mathbb{R}^m \\ y &= h(x,u), \quad y(t) \in \mathbb{R}^p \end{aligned} \tag{5.10}$$

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 (5.10) 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

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

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 (5.10) 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 10. *Assume (5.10) 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 Ψ_{jk}^{il}, $i, l = 1, \ldots, n$, are given by $\Psi_{jk}^{il}(t) = (y_{ijk}(t) - y_{ijk,ss})^T(y_{ljk}(t) - y_{ljk,ss})^T$. Here $y_{ijk}(t)$ is the output of the system (5.10) 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 11. *Assume (5.10) 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 σ_i can be used to find initial guesses of approximation order r. The truncation of the nonlinear system is then done as follows. Let

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

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

$$\begin{aligned} \dot{z} &= f_r(z, u) = W^T f(Vz, u), \quad z(t) \in \mathbb{R}^r,\ u(t) \in \mathbb{R}^m \\ y_r &= h_r(z, u) = h(Vz, u), \qquad y(t) \in \mathbb{R}^p. \end{aligned} \tag{5.11}$$

In general, nothing can be guaranteed about the closeness of the dynamics of (5.10) and (5.11).

Nonlinear Gramians

A theory of nonlinear Gramians and balancing is available. It has been developed by Scherpen et al. The theory is elegant, but it may be hard to compute the nonlinear Gramians and the balancing transformations for large systems. See, for example:

J.M.A. Scherpen. *Balancing for nonlinear systems.* Systems & Control Letters, 21 (1993) 143-153.

Taylor expansions of the nonlinear Gramians are considered in:

A. J. Krener. *Reduced Order Modeling of Nonlinear Control Systems.* In Analysis and Design of Nonlinear Control Systems Analysis and Design of Nonlinear Control Systems, Springer, 2008.

5.4 Suggested Reading

The survey article *Controller Reduction: Concepts and Approaches* presents the most common methods for controller reduction. It also discusses frequency-weighted reduction in Section III. References to papers that discuss balanced stochastic truncation are given in *Linear Robust Control*.

The paper *A subspace approach to balanced truncation for model reduction of nonlinear control systems* introduced the empirical Gramians, and motivates their construction.

The paper *Controllabilty and observability covariance matrices for the analysis and order reduction of stable nonlinear systems* extends the concepts to more general systems.

5.5 Exercises

EXERCISE 5.1

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 5.2 (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 5.3

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 (5.4), 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 5.4

Verify the Taylor expansion (5.6).

EXERCISE 5.5

A mechanical spring-mass system G can be modelled by

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

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 $-K$ is obtained,

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

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

Perform controller reduction on K to obtain a lower order 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 K and K_r.