1 EL2520 Lecture notes 1: Introduction, Norms of Signals and Systems, The Small Gain Theorem

Feedback is a relatively simple yet extremely powerful principle for modifying the behavior of dynamical systems. Feedback is everywhere, both in natural/biological and engineered systems. The basic principle is that decisions are based on observations of the system behavior rather than on prior knowledge and information about external influences. The main reason for using feedback is that it is the most effective way to deal with uncertainty about a system and its surroundings. This course presents methods for analysing and designing feedback control algorithms for multivariable systems with respect to stability and performance in the presence of uncertainty.

The standard feedback control problem can be represented by the block diagram in Figure 1.1. Here G represents the system to be controlled, F_y is the feedback controller and F_r is a pre-filter on the setpoint r. The other signals (variables) are the plant input (manipulated variable) u, the plant output z, the measurement y, the disturbance on the output w and the measurement noise n. The aim of the control system is to make the output z follow the setpoint r in the presence of disturbances w and measurement noise n, or equivalently, keep the control error $e = r - z$ small[1]. It is usually of interest to also keep the control input u small in some sense.

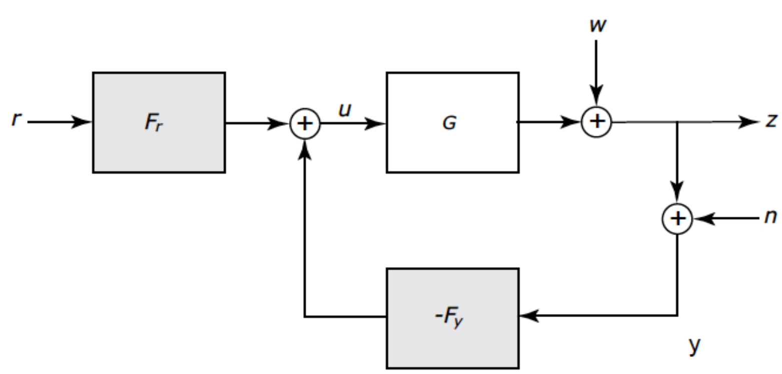


Figure 1.1: Two degree of freedom control system.

The feedback controller in Figure 1.1 is a so-called two degree of freedom controller since we have a prefilter F_r on the setpoint and a feedback controller F_y based on the measurement y

$$u = F_r r - F_y y$$

[1]The formulation of the control problem as a setpoint tracking problem may seem limiting, but note that many problems can be translated into this form. For instance, (local) optimization of a system can be achieved by letting the output be the derivative of the objective function and the setpoint $r = 0$.

The two degrees of freedom allow us to design the response to setpoints r more or less independently of the response to disturbances w and noise n. The standard one degree of freedom controller is obtained by letting $F_r = F_y = F$ which yields $u = F(r - y)$.

In this course we will consider multivariable systems, i.e., all signals are vectors and all transfer-functions are matrices. For instance, consider a 2×2 system with two inputs u_1, u_2 and two outputs y_1, y_2. We then have $u = [u_1\ u_2]^T$ and $y = [y_1\ y_2]^T$ and the transfer matrix $G(s)$ is a 2×2 matrix with the element $G_{ij}(s)$ being the transfer-function from input u_j to output y_i.

The focus will be on systems that can be described by a linear time invariant (LTI) model, which on state space form is

$$\begin{aligned} \dot{x} &= Ax(t) + Bu(t)\ ; x \in \mathbb{R}^n, u \in \mathbb{R}^m \\ y(t) &= Cx(t) + Du(t)\ ; y \in \mathbb{R}^p \end{aligned} \tag{1}$$

where x is the n-dimensional state vector. Taking the Laplace transform of (1) and eliminating the state x, we get the $p \times m$ transfer-matrix from input u to output y

$$Y(s) = G(s)U(s)\ ; \quad G(s) = C(sI - A)^{-1}B + D \tag{2}$$

We will also make extensive use of the frequency response of the system which is obtained by letting $s = i\omega$ in $G(s)$

$$Y(i\omega) = G(i\omega)U(i\omega) \tag{3}$$

where $Y(i\omega)$ and $U(i\omega)$ are the Fourier transforms of $y(t)$ and $u(t)$, respectively.

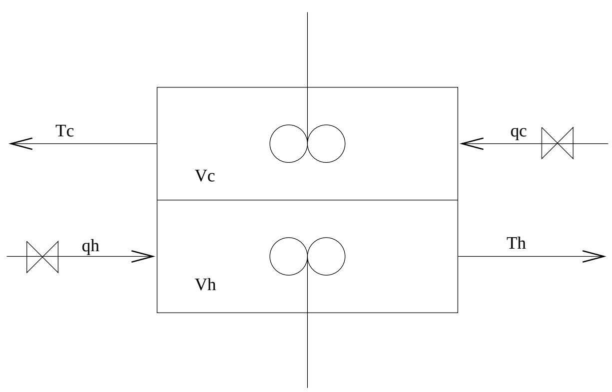


Figure 1.2: Simple heat exchanger with hot q_h and cold q_c flows.

As an example, consider the simple heat exchanger in Figure 1.2, used to transfer heat from a hot stream q_h to a cold stream q_c. Energy balances across the hot and cold side, respectively, yield the differential equations for the temperatures T_h and T_c.

$$\begin{aligned} V_h \frac{dT_h}{dt} &= \alpha_h(T_c - T_h) + q_h\beta_h(T_{hi} - T_h) \\ V_c \frac{dT_c}{dt} &= \alpha_c(T_h - T_c) + q_c\beta_c(T_{ci} - T_c) \end{aligned} \tag{4}$$

A linear time invariant model is obtained by linearizing this model about a desired steady-state. Assume $V_h = V_c = 1$, $\alpha_h = \alpha_c = 0.1$, $\beta_h = \beta_c = 0.5$, $T_{hi} = 100C, T_{ci} = 20C$,

$q_h = 0.1m^3/s$ and $q_c = 0.04m^3/s$. At steady-state the time-derivatives are zero, which yields the steady-state values $T_h^* = 80C$ and $T_c^* = 70C$. A Taylor series expansion of the right hand sides of (4) about this steady-state, keeping only the linear terms, yields

$$\begin{aligned} \dot{x} &= \begin{pmatrix} -0.15 & 0.1 \\ 0.1 & -0.12 \end{pmatrix} x(t) + \begin{pmatrix} 10 & 0 \\ 0 & -25 \end{pmatrix} u(t) + \begin{pmatrix} 0.05 & 0 \\ 0 & 0.02 \end{pmatrix} d(t) \\ y(t) &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} x(t) \end{aligned} \tag{5}$$

where the state vector $x = [\Delta T_h \ \Delta T_c]^T$, the input $u = [\Delta q_h \ \Delta q_c]^T$, the disturbance $d = [\Delta T_{hi} \ \Delta T_{ci}]$ and the output $y = [\Delta T_h \ \Delta T_c]^H$. Note that all variables in a linear model are deviation variables, e.g., $y_1 = \Delta T_h = T_h - T_h^*$, corresponding to the deviation from the steady-state value at which the linearized model was obtained. Taking the Laplace transform $G(s) = C(sI - A)^{-1}B$, we obtain

$$G(s) = \frac{1}{(29.5s+1)(4.24s+1)} \begin{pmatrix} 150(8.33s+1) & -312.5 \\ 125 & -469(6.67s+1) \end{pmatrix}$$

for the transfer-matrix from the input u to the output y. As an exercise you can derive the corresponding transfer-matrix $G_d(s)$ from the disturbance d to the output y. Note that, with reference to the block-diagram in Figure 1.1, the disturbance on the output is $w = G_d(s)d$.

Remark: Note that the controllers $F_y(s)$ and $F_r(s)$ we design in this course in general will be LTI models of the controllers, just like $G(s)$ and $G_d(s)$ are models of the physical systems we want to control. Thus, F_y and F_r must be *realized* in order to implement them on the real system. That is, we need to create a real system that has the dynamics corresponding to the models $F_y(s)$ or $F_r(s)$. The realization is in most cases solved by implementing the controller as an algorithm in a digital computer. Realization of the controller will not be covered in any detail in this course, but is covered e.g., in the course *EL2450 Hybrid and Embedded Control Systems.*

One may ask why we usually rely on feedback in control systems? We will in this course for the most part assume that a model is available for the plant G

$$z = G(s)u + w \tag{6}$$

Combining (6) with the control objective $z = r$, we get

$$u = G^{-1}(r - w) \quad \Rightarrow \quad z = r \tag{7}$$

which corresponds to feedforward control from the reference r and disturbance w. Thus, if we have a perfect model and we measure the disturbance w then there appears to be no need for feedback control. However, there are at least three reasons why we usually need feedback.

First, there will always be some mismatch between the true system and the model; we term this *model uncertainty.* Assume for instance that the true system $\tilde{G}$ is given by

$$\tilde{G} = G(I + \Delta_G)$$

where G is the model and Δ_G is the relative model error, i.e., the relative difference between the true plant and the model. Combining this with the feedforward control (7) we get

$$z = r + G\Delta_G G^{-1}(r - w)$$

Thus, for $\Delta_G \neq 0$ we do not get perfect following of the setpoint r. In fact, as we shall see later in the course, the second term representing the control error $e = r - z$ can become very large even for relatively small model errors Δ_G. Feedback is an efficient way of reducing the impact of model uncertainty. In Lecture 2 we will quantify the effect of feedback on uncertainty.

Second, there are usually many different disturbances acting on a system and we seldom measure all of these, if any at all. Unmeasured disturbances can be seen as uncertainty about the environment of the system we want to control and can only be dealt with using feedback control. Note that without knowledge of w and no feedback we have $z = w$, i.e., the disturbance goes directly through to the output. In Lecture 2 we will also quantify the effect of feedback on disturbance attenuation.

Finally, if the system $G(s)$ is unstable then the only way to make the system stable is by the use of feedback (or redesign of the system itself).

In summary, the main reasons for employing feedback, as opposed to feedforward, control are

- Model uncertainty (uncertain knowledge about system)
- Unmeasured disturbances (uncertain environment)
- Unstable systems

The main costs of feedback are partly that the controller F_y feeds measurement noise into the plant via the control input u, and partly that we potentially risk inducing instability in an otherwise stable system.

Note that we in general can combine feedforward and feedback control, and that for instance the pre-filter F_r in Figure 1.1 is a feedforward control from the setpoint r.

1.1 Controller Design

The classical control design and analysis methods, initially developed by pioneers such as Bode and Nyquist and typically dealt with in an introductory course in control, are valid for single-input-single-output (SISO) systems only. In this course we deal with multi-input-multi-output (MIMO) systems for which the classical methods are not directly applicable. However, it may be tempting to consider a MIMO system as a collection of SISO systems, and control each output with one input. We consider this case first to explain why it usually is not a viable approach.

As an example, consider the heat-exchanger above. We may consider controlling the hot temperature $y_1 = T_h$ using the hot flow $u_1 = q_h$ and the cold temperature $y_2 = T_c$ using the cold flow $u_2 = q_c$. If we for instance employ PI-controllers we then have

$$u_1 = \underbrace{K_{c1}\frac{T_{i1}s+1}{T_{i1}s}}_{C_1}(r_1 - y_1) \; ; \quad u_2 = \underbrace{K_{c2}\frac{T_{i2}s+1}{T_{i2}s}}_{C_2}(r_2 - y_2)$$

This control strategy is called *decentralized control* and is illustrated by the block diagram

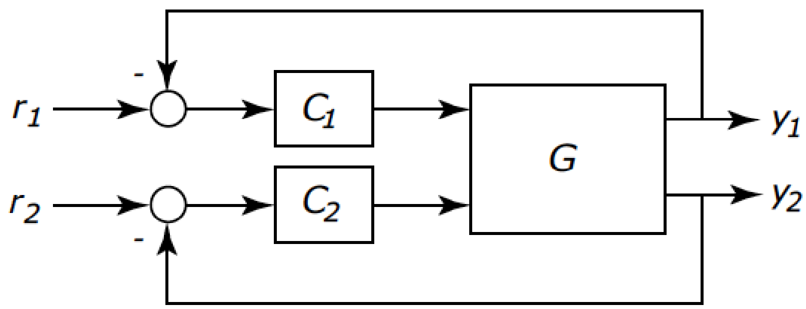


Figure 1.3: Decentralized control of a 2×2 system

in Figure 1.3. If we now consider each loop separately, we can tune these using classical control design methods. For the individual loops we then get the closed-loop transfer-functions

$$y_1 = \frac{G_{11}C_1}{\underbrace{G_{11}C_1 + 1}_{G_{c1}}} r_1 \; ; \quad y_2 = \frac{G_{22}C_2}{\underbrace{G_{22}C_2 + 1}_{G_{c2}}} r_2$$

However, since both inputs affect both outputs, the two loops will interact when we close both of them and the behavior will in general not be as expected from considering the individual loops (see also example 1.1 in course book or slides from Lecture 1).

If we consider the problem on the correct multivariable (matrix) form we get

$$y = \begin{pmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{pmatrix} \begin{pmatrix} C_1 & 0 \\ 0 & C_2 \end{pmatrix} (r - y) \quad \Rightarrow \quad y = \underbrace{(I + GC)^{-1}GC}_{G_c} r$$

where $y = [y_1 \; y_2]^T$ and $r = [r_1 \; r_2]^T$. Since G is a full matrix, so will also G_c be full and the diagonal elements of G_c will in general be very different from the individual elements G_{c1} and G_{c2}. In fact, G_c may be unstable even if the individual loops are stable. Furthermore, the off-diagonal elements of G_c will in general be non-zero, implying that changing the setpoint for one of the outputs will cause both outputs to change. Thus, in the general case it is not advisable to treat MIMO problems as a collection of SISO systems. Rather, one should approach them as multivariable systems as such, i.e., based on transfer-matrices and signal vectors. Since the classical design methods, like lead-lag design, are limited to SISO systems we will need to employ design and analysis methods that are specifically aimed at multivariable systems. One convenient approach to design controllers for multivariable systems is to formulate the control problem as an optimization problem, i.e., minimize the control error e and input usage u. For this we will need to quantify the size of signals and systems, and for this purpose we introduce below the concept of *signal and system norms.*

Note that there may be systems where a diagonal controller C, i.e., decentralized control, will work satisfactory. We will later in the course consider methods for determining when this will be the case.

1.2 Signal Norms and System Gain

In multivariable dynamic systems, all signals are vectors that furthermore are functions of time (or frequency). If we want to keep a signal, e.g., the control error e, small in some sense, we need to be able to quantify its size. A *vector norm* is a function that maps a vector into a scalar positive number, while a *signal norm* maps a function of time or frequency into a scalar positive number. Norms have to satisfy certain basic properties, but we will not go into any details here but rather define the specific norms that we will use in this course. For a more in-depth introduction to norms, we refer to a course on functional analysis.

There exists many different norms that can be used to quantify the size of a vector. In this course, we we will only use the most common vector norm, namely the Euclidian 2-norm. For a real vector $z \in \mathbb{R}^m$ the vector 2-norm is defined as

$$|z| = \sqrt{\Sigma_{i=1}^m z_i^2} = \sqrt{z^T z}$$

Note that this is the standard definition of the (Euclidian) length of a vector.

A signal norm measures the size of a time varying signal. If the signal is a vector, then a vector norm is used to convert it into a scalar time varying signal. The L_∞-norm, or peak-norm, is defined as

$$\|z\|_\infty = \sup_{t \geq 0} |z(t)|$$

where $|\cdot|$ denotes the Euclidian 2-norm (sup denotes the supremum, or least upper bound, which for most practical cases equals the maximum value). A signal is said to be bounded if the peak-norm is finite, i.e., $\|z\|_\infty < \infty$.

The L_2-norm, or energy-norm, is defined as

$$\|z\|_2 = \sqrt{\int_{-\infty}^{\infty} |z(t)|^2 dt}$$

A signal is said to be finite-energy if the energy-norm is finite, i.e., $\|z\|_2 < \infty$.

We will in this course only use the L_2-norm, or 2-norm for short.

A system $\mathcal{S}$ can be seen as a mapping of an input signal u to an output signal y. See Figure 1.4. Just like we use norms to obtain scalar measures of the size of signals, we

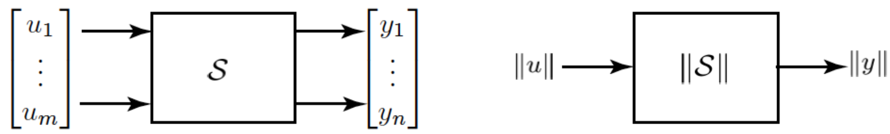


Figure 1.4: A system is a mapping of input signals to output signals. A norm $\|\cdot\|$ is function that assigns a positive real number, quantifying the size, to a time-varying or frequency dependent signal, or to the amplification of a system.

would like to have a scalar measure for the amplification of a system. For this purpose we employ system norms, also known as system gain. Consider first that we employ a specific input $u(t)$ which gives a specific output $y(t) = \mathcal{S}u(t)$. If we use the energy-norm, introduced above, to measure the size of the signals then the amplification exerted by the system is

$$\frac{\|y\|_2}{\|u\|_2} = \frac{\|\mathcal{S}u\|_2}{\|u\|_2}, \ \|u\|_2 \neq 0$$

The amplification will depend on the specific signal considered. To obtain a scalar measure of the system, we consider the maximum amplification over all possible signals

$$\|\mathcal{S}\| = \sup_{\|u\|_2 \neq 0} \frac{\|\mathcal{S}u\|_2}{\|u\|_2}$$

The maximum amplification, when both input and output are measured in the 2-norm, is called the energy-gain of the system, or simply the system gain.

The above definition of the energy-gain is valid for any system. Let us consider a stable linear time invariant SISO system with transfer-function $Y(s) = G(s)U(s)$, for which the frequency response is $Y(i\omega) = G(i\omega)U(i\omega)$ where $Y(i\omega)$ and $U(i\omega)$ are the Fourier transforms of $y(t)$ and $u(t)$, respectively. Assume that the peak amplitude of $|G(i\omega)|$ is K, i.e., $|G(i\omega)| \leq K \ \forall\omega$, and that $|G(i\omega^*)| = K$, i.e., the peak occurs at the frequency $\omega = \omega^*$. Then the energy-norm of $y(t)$ is, using Parseval's Theorem (the integral of the square of a function is equal to the integral of the square of its Fourier transform),

$$\|y\|_2^2 = \frac{1}{2\pi}\int_{-\infty}^{\infty} |Y(i\omega)|^2 d\omega = \frac{1}{2\pi}\int_{-\infty}^{\infty} |G(i\omega)|^2 |U(i\omega)|^2 d\omega \leq K^2 \|u\|_2^2$$

Note that equality holds if $u(t) = \sin(\omega^* t)$ and hence the system gain is

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

The energy-gain of a linear time invariant system G is called the H_∞-norm and is denoted $\|G\|_\infty$. Thus, for a stable SISO system it is simply the peak value of the amplitude $|G(i\omega)|$ in the Bode plot of G.

Consider next a simple static nonlinear nonlinear system

$$\mathcal{S} \ : \ y(t) = f(u(t)) \ ; \quad |f(x)| \leq K|x|$$

and $|f(x^*)| = K|x^*|$. Then the energy-norm of the output is

$$\|y\|_2^2 = \int_{-\infty}^{\infty} |f(u(t))|^2 dt \leq \int_{-\infty}^{\infty} K^2 |u(t)|^2 dt = K^2 \|u\|_2^2$$

and hence the energy-gain is

$$\|\mathcal{S}\| = \sup_u \frac{\|y\|_2}{\|u\|_2} = K$$

Since we in this course will focus on MIMO systems, we need to extend the definition of gain to transfer-matrices. We start by considering the gain for a static linear MIMO system $y = Au$. Then, the gain is

$$\|A\| = \sup_{u \neq 0} \frac{|y|}{|u|} = \sup_{u \neq 0} \frac{|Au|}{|u|}$$

Taking the square

$$\|A\|^2 = \sup_{u \neq 0} \frac{|Au|^2}{|u|^2} = \sup_{u \neq 0} \frac{u^T A^T A u}{u^T u} = \lambda_{max}(A^T A)$$

where λ_{max} is the maximum eigenvalue of $A^T A$. Hence, the gain of a static linear system (constant matrix) A is the square root of the maximum eigenvalue of $A^T A$. The square roots of the the eigenvalues of $A^T A$ are called the singular values of A and are denoted $\sigma_i(A)$. The largest singular value, which is the gain of A, is denoted $\bar{\sigma}(A)$, i.e.,

$$\|A\| = \bar{\sigma}(A)$$

We will return to the singular values later when we discuss properties of MIMO systems in more detail.

Using the results above we can show that the gain for MIMO LTI system G is

$$\|G\| = \sup_{\omega} \bar{\sigma}(G) = \|G\|_{\infty}$$

i.e., the peak value of the maximum singular value over all frequencies.

1.3 The Small Gain Theorem

The small-gain theorem is a general and very useful result on the stability of feedback systems that we will make extensive use of in this course. Essentially, it states that if the open-loop system is stable and the loop-gain is less than one, then also the closed-loop is stable. The definition of stability we use here is that of *input-output stability*. A system is said to be input-output stable if a bounded energy input gives a bounded energy output, or if the gain

$$\|\mathcal{S}\| < \infty$$

that is, the system has finite gain. We can then formally formulate the small gain theorem

The Small Gain Theorem: Consider the feedback interconnection of two systems $\mathcal{S}_1$ and $\mathcal{S}_2$ in Figure 1.5, where $\mathcal{S}_1$ and $\mathcal{S}_2$ both are input-output stable and the loop gain

$$\|\mathcal{S}_1\|\|\mathcal{S}_2\| < 1$$

Then, the closed-loop system is input-output stable from any input r_1, r_2 to any output e_1, e_2, y_1, y_2.

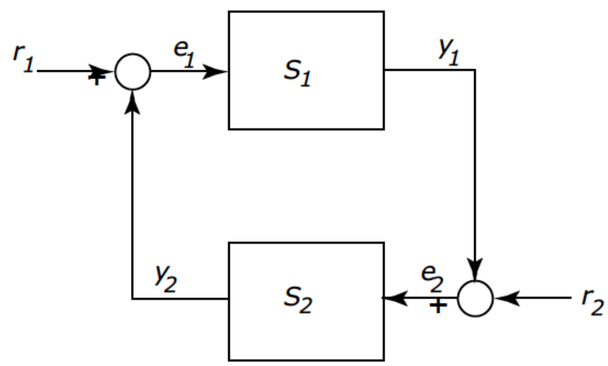


Figure 1.5: Feedback interconnection of two stable systems $\mathcal{S}_1$ and $\mathcal{S}_2$.

The proof in the linear case is based on the Generalized Nyquist Theorem[2] which we will introduce later in the course, and the proof is therefore left out for now. An informal sketch to a proof is given on the slides for Lecture 1.

Note that this is a sufficient condition for stability only, and by no means necessary. Consider for instance a linear negative SISO feedback loop with stable loop transfer-function $L(s)$. Then, from the Bode stability criterion the closed-loop is stable if the amplification $|L(i\omega)| < 1$ at the frequency where the phase lag is $\arg L(i\omega) = -\pi$. Thus, if we require $|L(i\omega)| < 1$ at all frequencies, as in the small gain theorem, then clearly we have closed-loop stability. But, this is then potentially highly conservative since we according to the Bode criterion can have $|L(i\omega)| >> 1$ at frequencies where the phase lag is not $-\pi$. Despite this conservativeness, we shall later see that the Small Gain Theorem is very useful, in particular when analyzing robustness, i.e., stability in the presence of uncertainty.

[2]and the proof in the general nonlinear time-varying case is based on what is essentially a generalization of the Nyquist theorem.

2 EL2520 Lecture notes 2: The Closed-Loop System

Consider the two-degree of freedom feedback control system in Figure 2.1. As discussed in Lecture 1, the aim of the control system is to make the output z follow the reference r, in the presence of disturbances on the input w_u, disturbances on the output w and measurement noise n (and model uncertainty Δ_G).

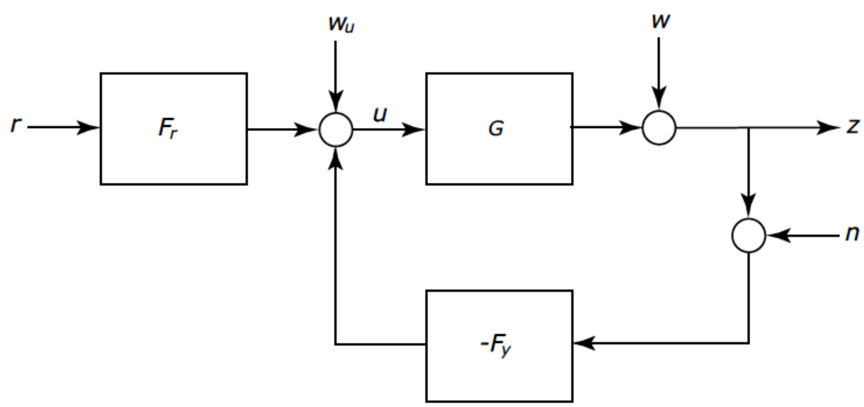


Figure 2.1: Feedback control system.

In classical control, dealing with SISO systems only, the control problem is typically solved in the frequency domain, i.e., based on the frequency response $G(i\omega)$. A standard approach to controller design is then to specify properties such as the closed-loop bandwidth and maximum peak of the closed-loop transfer-function G_c from r to z. To simplify the design, the specifications are then translated into specifications on the loop transfer-function $L = GF_y$, such as crossover frequency and phase margin, and these specifications are then met by *loop shaping* using the controller F_y (also known as lead-lag compensation). Robustness, i.e., stability in the presence of model uncertainty, is adressed through the phase- and amplitude margins in the classical framework.

In this course, we will introduce what is known as modern control[1]. This approach is also based on formulating and analyzing the control problem in the frequency domain. However, to accomodate for MIMO systems, specifications are instead based on norms of signals and/or systems. The use of norms will, as we shall see, also enable us to quantify limitations for what can achieved with feedback for a given system. For instance, if there is a time delay in the system, what is the maximum bandwidth that can be achieved even with the best possible controller? In addition, the use of norms in combination with the

[1]Sometimes modern control refers to the optimal control theory developed in the 1960s, using a state space approach to control, and then the control theory introduced here, developed mainly in the 1980s-90s, is referred to as postmodern. We will in this course also consider optimal control theory from the 60s, and then in the form of LQG optimal control in Lecture 8.

small gain theorem will enable us to quantify robust stability is a much more rigorous way than with the classical stability margins. Most importantly, we can extend the analysis of robust stability to MIMO systems for which robustness typically is a much more critical issue than for SISO systems.

We will start out by introducing the modern approach to control for SISO systems, and later extend the results to MIMO systems. Thus, in Lecture 2-4 we limit ourselves to systems with scalar inputs and outputs, i.e., SISO systems. In Lecture 5-13 we will deal with the more general MIMO problem.

The aim of control is, as stated above, to keep signals like the control error $e = r - z$ and the control input u small in some (norm) sense. This again implies making the transfer-functions from the external inputs w, w_u, r, n to the outputs e, u small in some (norm) sense. To derive the relevant closed-loop transfer-functions we use simple block-diagram algebra, i.e., we write up the relations in the block-diagram of Fig.1 and then solve for the variables of interest

$$z = w + G(w_u + F_r r - F_y(z + n))$$

Solving for z gives

$$z = \underbrace{\frac{1}{1 + GF_y}}_{S} w + \underbrace{\frac{G}{1 + GF_y}}_{SG} w_u + \underbrace{\frac{GF_r}{1 + GF_y}}_{G_c} r - \underbrace{\frac{GF_y}{1 + GF_y}}_{T} n \tag{1}$$

Here S is called the *Sensitivity function*, T is called the *Complementary Sensitivity function* and G_c is called the *Closed-Loop transfer function*. Note that

$$S + T = \frac{1}{1 + GF_y} + \frac{GF_y}{1 + GF_y} = 1$$

From (1) we see that we should make $|S(i\omega)|$ small for disturbance attenuation, $|T(i\omega)|$ small for noise attenuation and $|1 - G_c(i\omega)|$ small for setpoint following. Since $S + T = 1$, we see that we can not make both $|S|$ and $|T|$ small at the same frequency and hence we have to make a trade-off between attenuation of disturbances w, w_u on one hand and amplification of measurement noise n on the other hand.

The closed-loop transfer-functions for the control input u are derived in the same manner

$$u = w_u + F_r r - F_y(n + w + Gu) \quad \Rightarrow \quad u = \underbrace{\frac{1}{1 + GF_y}}_{S} w_u + \underbrace{\frac{F_r}{1 + GF_y}}_{SF_r} r - \underbrace{\frac{F_y}{1 + GF_y}}_{SF_y} (n + w)$$

Thus, to keep u small we need to make $|S|, |SF_r|$ and $|SF_y|$ small. We will below discuss how to define more precisely what we mean by ”small” transfer-functions. First, we will introduce the concept of *internal stability* since closed-loop stability is a pre-requisite for discussing anything related to control performance.

2.1 Internal Stability

A system is said to be *internally stable* if it is input-output stable from any external input to any output. For a SISO system, the outputs are u and z and the external inputs are r, w, w_u, n and we hence require the corresponding transfer-functions $S, SG, G_c, T, SF_r, SF_y, F_r$ to all be stable, i.e., have all poles strictly in the complex LHP. Note that we include also the feedforward controller F_r since it is outside the feedback loop and hence has to be stable. Thus, in principle we must check the poles of seven different transfer-functions to analyze internal stability. However, we can reduce the number to four transfer-functions since stability of certain transfer-functions implies stability of others. In particular,

- If S is stable, then so is T since $S + T = 1$.
- If S and F_r stable, then so is SF_r.
- If SG and F_r stable, then so is G_c.

Thus, it suffices to check stability of the transfer-functions S, SG, SF_y, F_r (also known as the Gang of Four).

Example: Consider the following control system

$$G = \frac{1}{s-1}\ , \quad F_y = F_r = \frac{s-1}{s}$$

which gives

$$S = \frac{s}{s+1}\ , \quad SG = \frac{s}{(s+1)(s-1)}\ , \quad SF_y = \frac{s-1}{s+1}\ , \quad F_r = \frac{s-1}{s}$$

and the closed-loop system is unstable since SG has a pole in the complex RHP.

Note that in the example above the controller has a zero at $s = 1$ which cancels the pole in G at $s = 1$. It is this cancellation that causes instability in this case. In fact, one should never cancel RHP poles and RHP zeros between the controller and plant since there always will be signals that can enter between the controller and plant which then implies instability from these signals. In the example above we have instability in SG, i.e., from w_u to z.

Internal stability as defined above is often also termed *nominal stability* since it only concerns stability of the closed-loop with the nominal model G. This in contrast to *robust stability* which concerns stability in the presence of model uncertainty, or model errors. Nominal stability, i.e., stability without uncertainty, is obviously a prerequisite for robust stability.

2.2 The Sensitivity Function

As seen from (1), the sensitivity function S is the transfer-function from disturbances on the output w to the output z. Without any feedback, i.e., $F_y = 0$, this transfer-function is simply 1.

– without feedback $z = w$

– with feedback $z = Sw$

Thus, feedback provides disturbance attenuation at frequencies where $|S(i\omega)| < 1$, while it amplifies disturbances at frequencies where $|S(i\omega)| > 1$. Ideally, considering only disturbance attenuation, we would like to achieve $|S(i\omega)| = 0\ \forall\omega$. However, there are several reasons why this is not possible. One simple reason, due to a result by Bode, is that any system for which the loop transfer-function $L = GF_y$ is stable and has a pole excess of a least 2 (at least two more poles than zeros, which is true for any real system) must have $|S(i\omega)| > 1$ at some frequency. This can easily be understood by considering the Nyquist plot of L shown in Figure 2.2. With a pole excess ≥ 2, L will have a phase-lag of $-\pi$ at some frequency which implies that the Nyquist curve must approach and possibly cross the negative real axis at some frequency and eventually end up in $|L| = 0$ at $\omega = \infty$. Furthermore, $L(i\omega)$ will have phase between $-\pi/2$ and $-\pi$ in some frequency range, implying that L must pass through the 3rd quadrant in the complex plane. According to the Nyquist criterion, L can not encircle -1 and hence L must pass through a unit circle centered at -1. Since $|1 + L|$ is the distance between L and the point -1 on the negative real axis, we must have $|1 + L| < 1$ at some frequency. Since $|S| = 1/|1 + L|$ it trivially follows that we must have $|S| > 1$ at some frequency. In fact, as we shall see in Lecture 4, any frequency for which $|S| < 1$ must be balanced by another frequency for which $|S| > 1$.

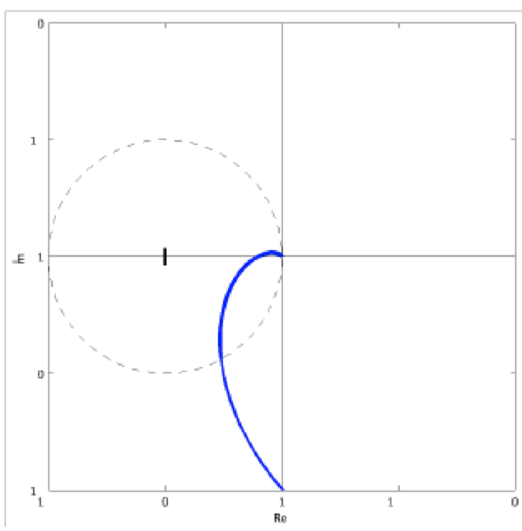


Figure 2.2: Nyquist plot, i.e., plot of loop gain $L(i\omega)$ as a function of frequency $\omega \in [0, \infty]$ in the complex plane. The dashed line is the unit circle with center in -1.

In general, to keep $|S(i\omega)| < M_S\ \ \forall\omega$ implies that the loop gain $L(i\omega)$ must be kept outside a circle centered at -1 and with radius M_S^{-1}. Typically, one allows maximum peak of the sensitivity $M_s = 2$ which implies avoiding that $L(i\omega)$ enters a circle centered at -1 and with a radius $M_S^{-1} = 0.5$

From the above, it is clear that we can not make the sensitivity small at all frequencies, and that we also need to have a peak in $|S|$ larger than one at some frequency. For control design, we then need to determine in what frequency ranges we want to make $|S|$ small and also what peak values we allow. A convenient way to do this is to define a frequency dependent bound[2] $|W_S^{-1}(i\omega)|$ that $|S|$ should stay below

$$|S(i\omega)| < |W_S^{-1}(i\omega)| \; \forall\omega \quad \Leftrightarrow \quad \|W_S S\|_\infty \leq 1$$

See also Figure 2.3. The second inequality follows from the fact the the $\|\cdot\|_\infty$-norm is

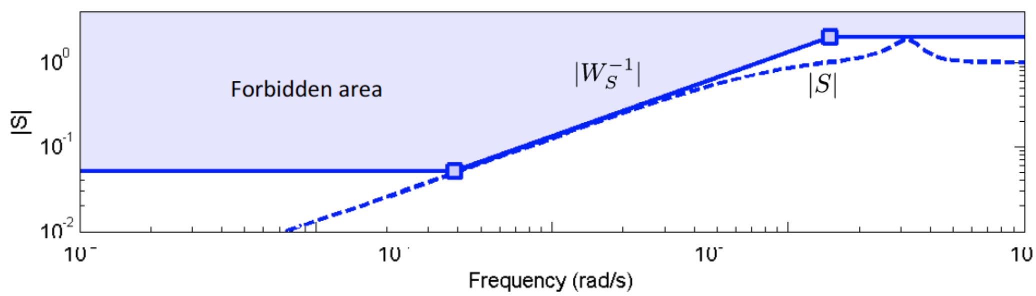


Figure 2.3: Plot of amplitude of sensitivity function S and bound W_S^{-1} as functions of frequency.

equal to the peak value of the amplitude evaluated over all frequencies (see Lecture 1). Thus, we have formulated our design objective for the sensitivity function in terms of the infinity-norm of the weighted sensitivity sensitivity $W_S S$ where W_S is the weight that we specify. Note that we now have the performance measure in the form of a single number $\|W_S S\|_\infty$ which makes it suitable for solving using optimization techniques. That is, we can determine the feedback controller F_y based on solving an optimization problem in which we minimize $\|W_S S\|_\infty$. We will return to such formulations later.

The origin of the name *sensitivity function* is in fact not related to the disturbance sensitivity discussed above. Rather, it is due to Bode and reflects the fact that the sensitivity function quantifies the relative sensitivity of the closed-loop transfer-function G_c to changes in the model G. Taking the derivative

$$\frac{dG_c}{dG} = \frac{F_r}{(1+GF_y)^2} = S\frac{F_r}{1+GF_y} = S\frac{G_c}{G}$$

and thus the relative sensitivity is

$$\frac{dG_c/G_c}{dG/G} = S$$

Considering the relative uncertainty description introduced in Lecture 1, we get

$$\tilde{G} = G(1+\Delta_G) \quad \Rightarrow \quad \tilde{G}_c = G_c(1+S\Delta_G)$$

where $\tilde{G}$ is the true system, G is the model and Δ_G is the relative model error. Thus, the feedback changes the relative error in the closed-loop relative to the open-loop by a factor equal to the sensitivity function S.

[2] We define the bound as the inverse of W_S since W_S serves as a weight on S in $\|W_S S\|_\infty$. We hence call W_S the performance weight, or weight on the sensitivity.

2.3 The Complementary Sensitivity Function

While the sensitivity function S reflects the ability of a feedback loop to attenuate disturbances and suppress the impact of model uncertainty on the closed-loop transfer-function G_c, the complementary sensitivity function T quantifies the amplification of measurement noise as well as the robust stability of a feedback loop.

From (1) we see that the complementary sensitivity function T is the closed-loop transfer-function from measurement noise n to the output z. Thus, to avoid amplifying noise in the loop we should make $|T(i\omega)|$ small at frequencies where noise is large. In fact, $T = 0$ without any feedback, so the fact that we feed measurement noise into the plant is a cost of employing feedback control. Note again that making $|T|$ small implies that we make $|S| \approx 1$, since $S + T = 1$, and hence a trade-off has to be made. More about this below.

Apart from reflecting the amplification of noise in a feedback loop, the complementary sensitivity function also has an important role in quantifying robust stability, that is, stability in the presence of model uncertainty. In particular, consider again the relative uncertainty description introduced in Lecture 1

$$\tilde{G}(s) = G(s)(1 + \Delta_G(s)) \tag{2}$$

where $\tilde{G}$ is the true system (unknown), G is the nominal model and $\Delta_G(s)$ is the relative model uncertainty. Given a controller which yields nominal stability, i.e., closed-loop stable with model G, the question of robust stability is then how large the uncertainty Δ_G can become without the closed-loop becoming unstable. In the basic control course (EL1000), this problem was solved using the Nyquist plot for the loop gain (assuming open-loop stability). Here we will instead approach the problem using the small gain theorem, mainly because we can then easily extend the results to MIMO systems and also to alternative uncertainty descriptions later (and we do not need to assume open-loop stability).

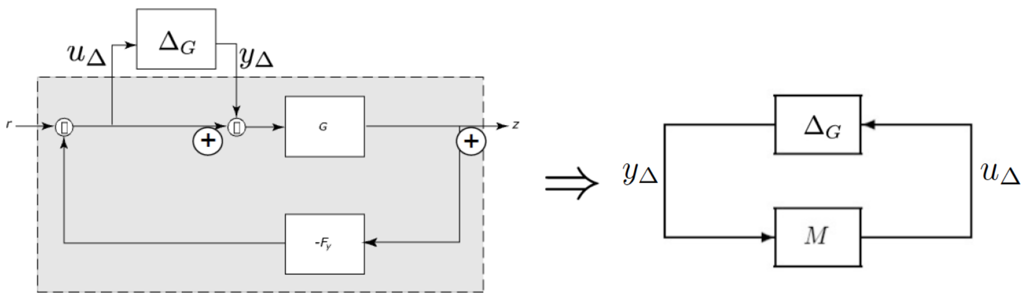
Consider the feedback loop in Figure 2.4, where we have included the uncertainty as given by (2). The question we would like to answer is what linear $\Delta_G(s)$ can be tolerated without the loop loosing stability? For this purpose we rewrite the block-diagram such that we isolate the uncertainty from the nominal system. See Figure 2.4. This then results in a feedback loop with two blocks; the uncertainty $\Delta_G(s)$ and the nominal transfer-function from the output y_Δ to the input u_Δ of Δ_G such that

$$u_\Delta = M y_\Delta$$

Using simple block-diagram algebra we identify

$$u_\Delta = -GF_y(y_\Delta + u_\Delta) \quad \Leftrightarrow \quad u_\Delta = \underbrace{\frac{GF_y}{1 + GF_y}}_{M} y_\Delta$$

That is, the block $M = T$, i.e., the complementary sensitivity function. Now, from the Small Gain Theorem we know that the M-Δ_G-loop in Figure 2.4 is stable if $M(s)$ and $\Delta_G(s)$ are both stable and the loop-gain $\|M\Delta_G\|_\infty < 1$. That $M(s)$ is stable follows from nominal closed-loop stability, i.e., stability without uncertainty, and the requirement of



$$u_\Delta = Ty_\Delta = My_\Delta$$

Figure 2.4: Block diagram with uncertainty. Corresponds to M-Δ_G-loop where M is nominal transfer-function (no uncertainty) and Δ_G is the relative model uncertainty.

$\Delta_G(s)$ stable means we can not allow unstable uncertainty $\Delta_G(s)$ when using the Small Gain Theorem for analysis. The robust stability condition is then $T(s)$ and $\Delta_G(s)$ stable and

$$\|T\Delta_G\|_\infty < 1 \quad \Leftrightarrow \quad |T(i\omega)| < \frac{1}{|\Delta_G(i\omega)|} \quad \forall\omega \tag{3}$$

Thus, we need to make the complementary sensitivity $|T|$ small for frequencies where the uncertainty $|\Delta_G(i\omega)|$ is large.

Example: Consider the nominal system

$$G(s) = \frac{1}{\tau s + 1} \ ; \quad \tau = 3$$

and the PI-controller

$$F_y(s) = 3\frac{3s+1}{3s}$$

which yields

$$T(s) = \frac{1}{s+1}$$

Assume now that the time-constant τ of the true system can change between 3 and 1, i.e., $\tau \in [1, 3]$. We model this uncertainty as

$$\tilde{G}(s) = \frac{1}{s+1} = \frac{1}{3s+1}(1 + \Delta_G(s)) \quad \Rightarrow \quad \Delta_G(s) = \frac{2s}{s+1}$$

Figure 2.5 shows the plot of $|T(i\omega)|$ and $1/|\Delta_G(i\omega)|$, and we see that condition (3) is just satisfied, thus we have robust stability for the changes in the plant time-constant between 3 and 1. Note that in this simple example we could simply have computed the closed-loop poles for the different time-constants in G instead and would then have found that it was robustly stable with some margin. In fact, if we allowed for variations in the plant time-constant between 0 and 3, the poles would still be all in the LHP, i.e., the system would be robustly stable, but the robust stability condition (3) would not be satisfied anymore. The conservativeness would in this case partly be due to sufficiency only of the small gain theorem, and partly by the fact that the uncertainty model we use would cover

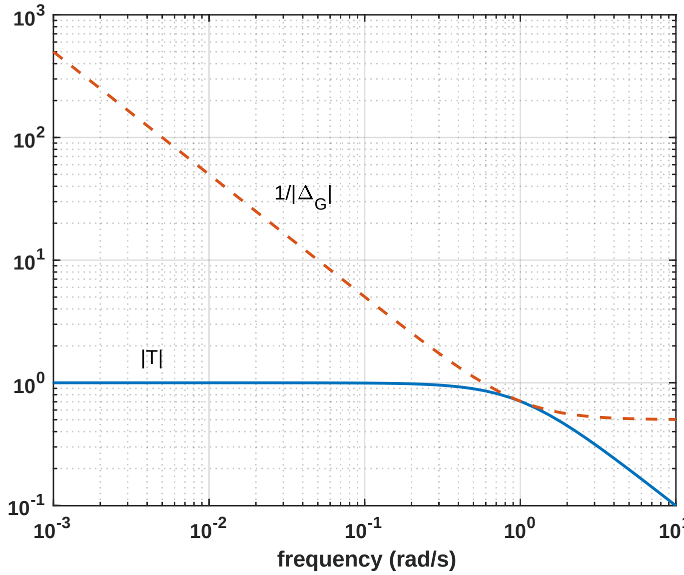


Figure 2.5: Plot of $|T(i\omega)|$ vs $1/|\Delta_G(i\omega)|$ for the example problem.

more plants than only those with time-constants between 0 and 3. More about this in Lecture 3.

Similar to what we did for the sensitivity function above, it is reasonable to define a frequency dependent bound $|W_T^{-1}(i\omega)|$ that $|T(i\omega)|$ should stay within. We then get

$$|T(i\omega)| \leq |W_T^{-1}(i\omega)| \; \forall \omega \quad \Leftrightarrow \quad \|W_T T\|_\infty \leq 1$$

Thus, by designing the weight $W_T(i\omega)$ we get a scalar measure $\|W_T T\|_\infty$ of the size of the complementary sensitivity T.

2.4 Sensitivity Shaping

Based on the above results we can see the control design problem as the problem of shaping the sensitivity functions S and T (and possibly other transfer-functions, such as those related to input usage) by designing the weights W_S and W_T and then e.g., leaving to an optimization algorithm to find a controller that minimizes the norms $\|W_S S\|_\infty$ and $\|W_T T\|_\infty$[3]. We will in the next few lectures focus on the design of the weights W_S and W_T, that is, formulating the control problem, and return to the problem of how to solve the resulting optimization problem later. Some simple observations are

- We should choose $|W_S|$ large in the frequency range where we want $|S|$ small. This is typically the frequency range where we need disturbance attenuation..

[3]Since we need a scalar objective function, we need to combine the two norms somehow.

- We should choose $|W_T|$ large in the frequency range where we want $|T|$ small. This is typically the frequency range where we have the most measurement noise and where the model uncertainty is the largest.

- Since $S + T = 1$, we can not choose $|W_S|$ and $|W_T|$ large in the same frequency range if we are to achieve the control objectives $\|W_S S\|_\infty \leq 1$, $\|W_T T\|_\infty \leq 1$.

See also Figure 2.6. In Lecture 4 we will discuss the trade-off discussed in the last point

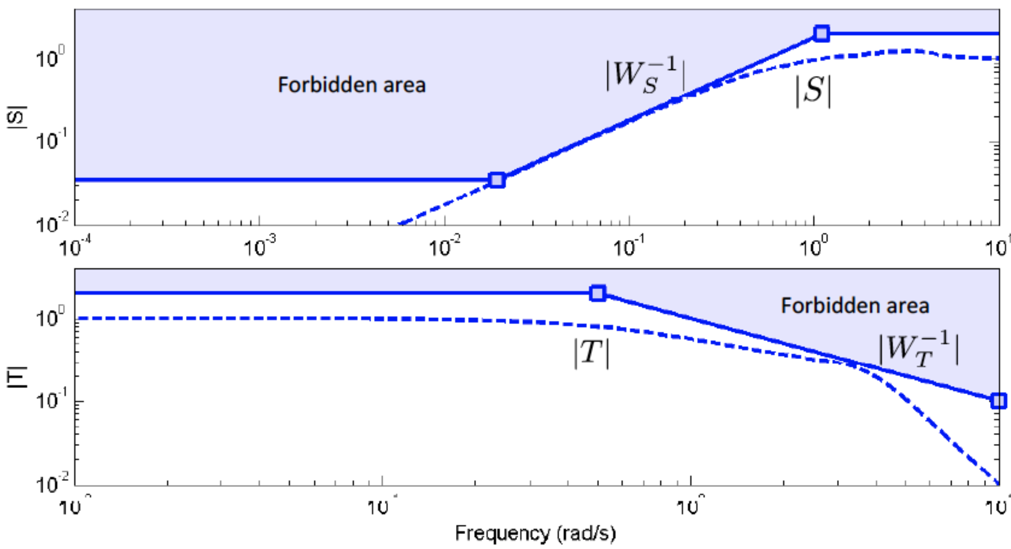


Figure 2.6: Shaping of sensitivity function S and complementary sensitivity function T.

above in more detail, and also other limitations for the design of the weights W_S and W_T. However, we will first discuss robustness in more detail in Lecture 3.

2 EL2520 Lecture notes 3: Robustness

In Lecture 2 we derived a robustness condition for the stability of a SISO feedback loop when the model of the plant is uncertain. In this lecture we will discuss modeling of uncertainty in more detail and derive robustness conditions for stability under alternative uncertainty descriptions We will also derive a condition for robust performance for SISO systems. As in the previous lectures, we focus for now on SISO systems but approach the problem with a framework that will allow us to more or less directly extend the results to the MIMO case later.

By the term *robust* we understand that a system is insensitive to uncertainty, in the sense that small changes in the system does not result in abrupt changes in the system behavior. Robustness can concern different system properties such as stability, in which case we talk about *robust stability*, or performance, for which we refer to *robust performance*. Robust stability is clearly a prerequisite for robust performance.

A system is either robust or not, and the quantitative measure of robustness is then the size of the uncertainty that the system can tolerate. To be able to quantify robustness we must hence define the uncertainty the system should be able to tolerate to be robust, i.e., what differences between the nominal model G and the true system $\tilde{G}$ should be considered. In general, there may be different sources of model uncertainty

- *Parametric uncertainty:* when certain parameters in a model, such as a time-constant or a time-delay, are uncertain. This is typically the case if we e.g., have identified a model from experimental data using methods from the field of *system identification*.
- *Unmodelled dynamics:* when our model is not sufficiently rich to capture the full dynamics of the system. For instance, it is often hard to describe the fast dynamics of many systems and these are then typically left out of the model. In many cases it is also desirable to work with a simple nominal model G and leave out dynamics that are considered less important for control.

Typically both sources of uncertainty are present in a given model, and it is then often reasonable to lump them together into what is termed a *lumped uncertainty description* and which should then be quite general. Below we will introduce such a lumped uncertainty description, based on model sets, which can accomodate for different sources of uncertainty.

2.1 Model Sets

One way of representing model uncertainty is to define a set of models rather than a single model. To generate such a model set, centered around a nominal model, we introduce a stable perturbation $\Delta_I(s)$ such that

$$\|\Delta_I\|_\infty \leq 1 \quad \Rightarrow \quad |\Delta_I(i\omega)| \leq 1 \ \forall\omega \tag{1}$$

Recall that the $\mathcal{H}_\infty$-norm $\|\cdot\|_\infty$ is only defined for stable systems. Thus, $\Delta_I(s)$ can be any stable transfer-function that has amplitude less than one (and any phase) at all frequencies. Examples of such transfer-functions are $\Delta_I(s) = e^{-\theta s}$ for any delay $\theta > 0$ and $\Delta_I(s) = \frac{1}{(\tau s+1)^n}$ for any $\tau > 0$ and $n > 0$. However, we will not consider any specific transfer-functions in the model set but rather allow any stable perturbation $\Delta_I(s)$ satisfying (1).

Given the above definition of a model perturbation $\Delta_I(s)$, we can define the uncertainty set

$$\Pi_I = \{G_p(s) = G(s)(1 + W_I(s)\Delta_I(s)), \quad \|\Delta_I\|_\infty < 1\} \tag{2}$$

where $W_I(s)$ is an *uncertainty weight* that defines the magnitude of the relative perturbation at each frequency. We will assume that the true system $\tilde{G}$ belongs to the set Π_I, such that if we ensure that the closed-loop system is stable for all models within the set Π_I then so will the true closed-loop system also be stable. Since the uncertainty set in (2) is on the form of relative model uncertainty introduced in Lecture 2, the robust stability condition derived using the small gain theorem applies to (2), i.e.,

$$\|W_I T\|_\infty \leq 1$$

See also Lecture notes 2.

Note that the model set (2) implies that, at each frequency ω, G_p is within a disc centered at $G(i\omega)$ and with radius $|W_I(i\omega)G(i\omega)|$ in the complex plane. See also Figure 2.1. The assumption is then that the frequency reponse of the true system at the given frequency is somewhere within the disc.

Example 1: uncertain gain. Consider a model $G_p(s) = k_p G(s)$ where the gain k is uncertain and can vary between $1 - \delta$ and $1 + \delta$ (nominal value 1). Thus,

$$\Pi_I = \{G_p(s) = k_p G(s), \quad k_p \in [1 - \delta, 1 + \delta]\}$$

which can be represented by (2) with a real $\Delta_I \in [-1, 1]$ and $W_I = \delta$.

Example 2: uncertain zero. Consider the uncertain set

$$\Pi_I = \{G_p = G_0(s)(s + z_p), \quad z_p \in [z_0 - \delta, z_0 + \delta]\}$$

which can be represented by (2) with $G(s) = G_0(s)(s + z_0)$, a real $\Delta_I \in [-1, 1]$ and the uncertainty weight $W_I(s) = \delta/(s + z_0)$.

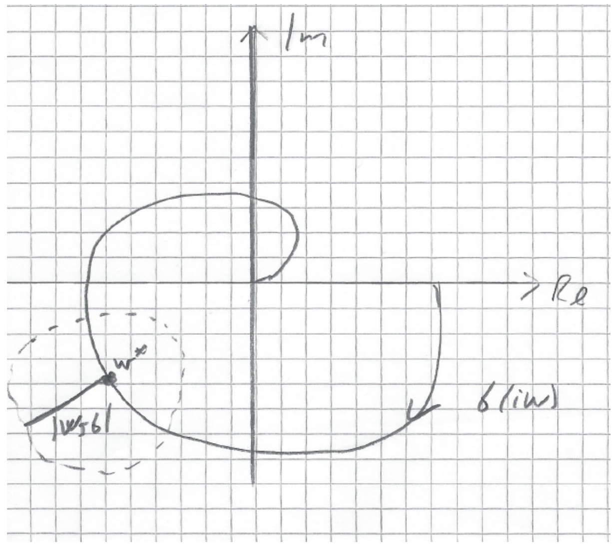


Figure 2.1: Nyquist plot of nominal model G. The dashed circle shows the uncertainty disc for G_p centered at $G(i\omega^*)|$ and with radius $|W_I(i\omega^*)G(i\omega^*)|$ at the frequency ω^*. The frequency response of the true system $\tilde{G}$ at ω^* is assumed to lie within the disc.

Note that in both examples above we should strictly use a real perturbation. However, since we will employ the small gain theorem for stability analysis of the set, we can only bound the magnitude of the uncertainty and hence we must allow a complex perturbation Δ_I, as defined by (1). Thereby we include models in the model set that are not part of the original model set and hence introduce some conservativeness in the robustness analysis.

Example 3: Fitting response data. According to the model set (2) we have

$$G_p(s) = G(1 + W_I\Delta_I) \; ; \quad \|\Delta_I\|_\infty \leq 1$$

From this we derive

$$\|W_I^{-1}G^{-1}(G_p - G)\|_\infty \leq 1$$

Hence, a model G_p is in the set Π_I if

$$|W_I(i\omega)| \geq \left|\frac{G_p(\omega) - G(i\omega)}{G(i\omega)}\right|, \quad \forall\omega \tag{3}$$

Thus, to obtain the uncertainty weight W_I one can simply plot the magnitude of $(G_p - G)/G$ as a function of frequency for all possible model candidates G_p and then fit a transfer-function such that $|W_I|$ is just larger than the maximum of these at all frequencies. Consider for instance the uncertain model

$$G_p = \frac{ke^{-\theta s}}{\tau s + 1}, \quad k, \theta, \tau \in [2, 3] \tag{4}$$

We first pick a nominal model, using the average values of k and τ while we choose the time-delay $\theta = 0$ to have a simple delay free nominal model

$$G(s) = \frac{2.5}{2.5s + 1}$$

We then evaluate G_p for values of k, τ and θ in the range $[2, 3]$ and plot the corresponding $|G_p(i\omega) - G(i\omega)|/|G(i\omega)|$ as a function of frequency. The result is shown in Figure 2.2. We first fit the weight

$$W_{I1}(s) = \frac{4s + 0.2}{1.6s + 1}$$

The magnitude of this weight is shown by the solid line in Figure 2.2, and as can be seen it satisfies (3) except for in a small frequency range around $\omega = 1$. To make sure the weight covers all model candidates, we add a small modification to the weight above to obtain

$$W_I(s) = W_{I1}(s)\frac{s^2 + 1.6s + 1}{s^2 + 1.4s + 1}$$

which magnitude is shown by the dashed line in Figure 2.2. We see that the weight now satisfies (3) at all frequencies. Hence, we have obtained a model set on the form (2) that covers the uncertainty in (4). Note again that the fitted lumped model set includes more models than the ones given by the original model set (4), hence introducing some conservativeness in the robustness analysis.

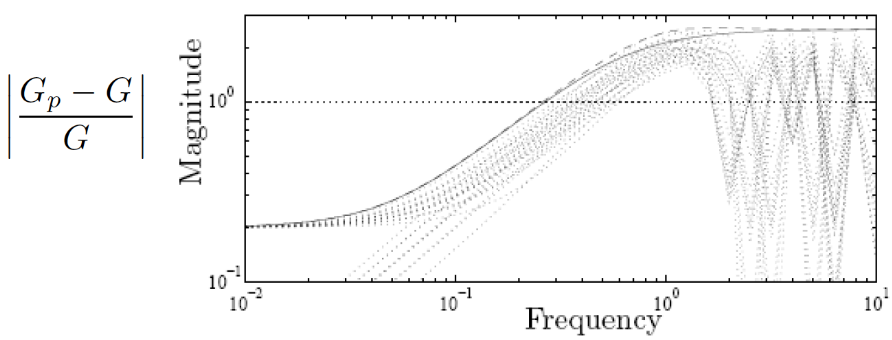


Figure 2.2: Example 3: amplitude of relative uncertainty $|G_p - G|/|G|$ for various parameter values of G_p. The solid and dashed lines shows fitted uncertainty weights W_I.

Let us now consider the controller

$$F_y(s) = K_c\frac{2.5s + 1}{2.5s}$$

According to the robust stability condition derived in Lecture 2 for model uncertainty on the form (2) we have robust stability if we have nominal stability, a stable perturbation $\Delta_G(s)$ and $\|W_I T\|_\infty < 1$. We first check nominal stability, i.e., closed-loop stability with the nominal model $G(s) = 2.5/(2.5s + 1)$. With the controller above we get

$$T(s) = \frac{F_y G}{1 + F_y G} = \frac{K_c}{s + K_c}$$

which is stable for any $K_c > 0$. For internal stability we need to check four transfer-functions (see Lec 2), but in general we have internal stability if one of them is stable and there are no cancellations of poles and zeros in the RHP between controller and plant. In this case, there are no such cancellations, and hence we have internal stability. The perturbation Δ_G is stable by definition (1). To check the robust stability condition $\|W_I T\|_\infty < 1$ we can either directly compute the norm (using e.g., the function `norm` in Matlab), or plot $|T|$ vs $|W_I^{-1}|$ in the frequency domain. With controller gain $K_c = 1$ we compute $\|W_I T\|_\infty = 1.75$ and hence we do not have robust stability. Reducing the controller gain to $K_c = 1/3$ we find $\|W_I T\|_\infty = 0.93$ and hence we have robust stability, i.e., we can guarantee closed-loop stability with any of the models in (4). The corresponding plots of $|T|$ vs $|W_I^{-1}|$ for the two values of K_c are shown in Figure 2.3 and as can be seen $|T|$ violates the bound for $K_c = 1$ but not for $K_c = 1/3$.

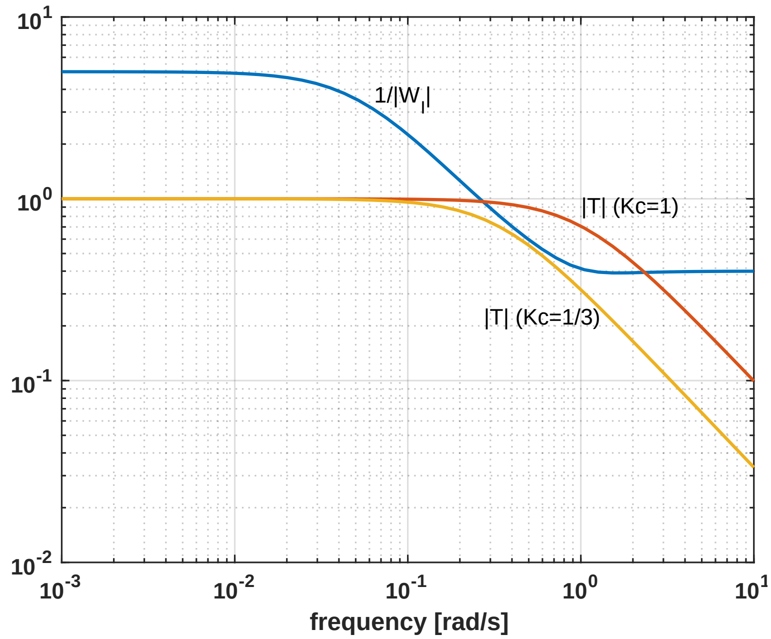


Figure 2.3: Example 3: plot of $|T|$ vs $|W_I^{-1}|$ for two different values of the controller gain K_c.

2.2 Uncertain Number of RHP Poles

One drawback with requiring the perturbation $\Delta_I(s)$ (1) to be stable is that we then can not allow poles to move between the LHP and RHP in the model set (2), i.e., all $G_p(s)$ (including $\tilde{G}$) must have the same number of RHP poles as the nominal model $G(s)$. Thus, if we for instance are uncertain about the open-loop stability of a system this can not be covered by the model set (2). However, it is possible to model an uncertain number of RHP poles by using a different model set. Define

$$\Pi_{iI} = \left\{ G_p(s) = G(s)(1 + W_{iI}(s)\Delta_I(s))^{-1}, \quad \|\Delta_I\|_\infty < 1 \right\} \tag{5}$$

The block diagram representation of (5) is shown in Figure 2.4. Note that we still only allow stable perturbations $\Delta_I(s)$, but by placing the perturbation within a feedback loop we can induce instabilty and thereby model poles crossing the imaginary axis. That is, $(1 + W_{iI}(s)\Delta_I(s))^{-1}$ can be unstable even if $\Delta_I(s)$ (and $W_{iI}(s)$) is stable.

Example 4: Uncertain stability. Assume we have a model with an uncertain pole

$$G_p(s) = \frac{1}{s - p}, \quad p \in [-3, 1]$$

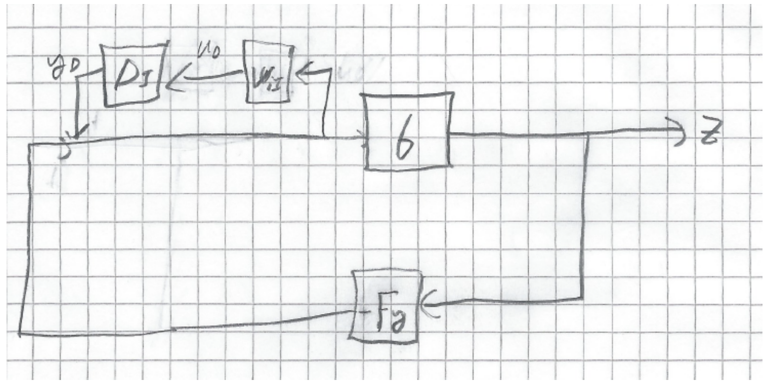


Figure 2.4: Block diagram with feedback around the uncertainty block.

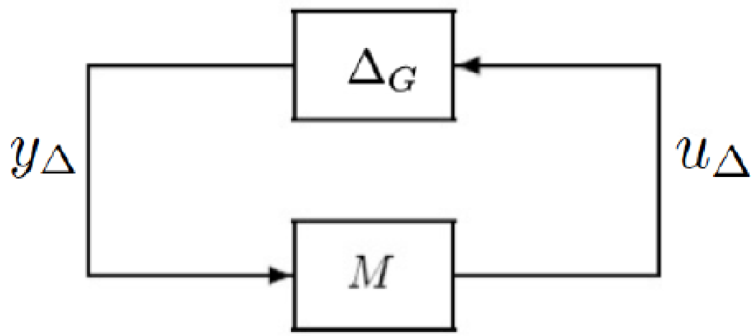


Figure 2.5: Block diagram in Figure 2.4 written on M-Δ-form.

We choose $p = -1$ in the nominal model, i.e., $G(s) = 1/(s+1)$, and the model set is then covered by (5) by choosing

$$W_{iI} = \frac{2}{s+1}$$

and $\Delta_I \in [-1, 1]$. Again we strictly only need a real perturbation Δ_I but must allow a complex perturbation with $|\Delta_I| < 1 \forall \omega$ to enable the use of the small gain theorem for robust stability analysis.

Let us now derive a robust stability condition for the model set (5). As before we rewrite the block-diagram in Figure 2.4 on M-Δ form (Figure 2.5), where M is a nominal transfer-function and Δ is a stable perturbation. Letting u_Δ and y_Δ be the input and outputs, respectively, of the Δ_I block in Figure 2.4, we identify

$$u_\Delta = W_{iI} y_\Delta - G F_y u_\Delta \;\Rightarrow\; u_\Delta = \underbrace{W_{iI} \frac{1}{1 + G F_y}}_{M} y_\Delta$$

Hence, $M = W_{iI} S$ and the robust stability condition is then W_{iI}, S stable, Δ_I stable and

$$\|W_{iI} S\|_\infty < 1$$

Thus, the robust condition now involves the sensitivity function S rather than the complementary sensitivity T, and we need to make $|S|$ small for frequencies where the uncertainty

$|W_{iI}|$ is large. Note that $|S|$ small means tight control and is as expected since uncertain open-loop instability implies that we need relatively high bandwidth to stabilize the system.

2.3 Robust Performance

Above we only considered closed-loop stability for all models in the respective model sets, i.e., robust stability. The assumption that the true system is within the considered model set then implies that we can guarantee closed-loop stability when the controller is applied to the real system. This is in some sense a least requirement on a controller; the system should not become unstable when we close the loop on the real system. However, it would obviously be an advantage if we also could guarantee that the performance specifications could be met for all models in the model set, and thereby the real system. This is called robust performance. It turns out that it is relatively easy to both analyze and achieve robust performance for SISO systems. We show this next.

Consider the performance specification

$$\|W_S S\|_\infty < 1 \quad \Rightarrow \quad |S(i\omega)| < |W_S^{-1}(i\omega)| \; \forall\omega$$

where S is the nominal sensitivity function, i.e., $S = 1/(1 + GF_y)$ with G being the nominal model. For robust performance with the model set (2) we get

$$|W_S S_p(i\omega)| < 1 \; \forall\omega; \quad S_p = \frac{1}{1 + G_p F_y} \quad \forall G_p \in \Pi_I$$

Inserting the relative model uncertainty of (2) we have

$$S_p = \frac{1}{1 + L + W_I \Delta_I L} \; ; \quad L = GF_y$$

At each frequency the worst case, i.e., the maximum $|S_p|$, is obtained when $1 + L$ and $w_I \Delta_I L$ point in opposite directions in the complex plane. Then

$$\sup_{G_p \in \Pi_I} |S_p(i\omega)| = \frac{1}{|1 + L(i\omega)| - |W_I(i\omega)L(i\omega)|}$$

and the requirement $|W_S S_p| < 1$ becomes

$$\frac{|W_S|}{|1 + L| - |W_I L|} < 1 \; \forall\omega$$

Dividing numerator and denominator by $|1 + L|$ we get

$$\frac{|W_S S|}{1 - |W_I T|} < 1 \forall\omega \quad \Leftrightarrow \quad |W_S S| + |W_I T| < 1 \; \forall\omega \tag{6}$$

Note that $|W_S S| < 1$ is the nominal performance requirement and $|W_I T| < 1$ is the robust stability criterion for the model set (2). Thus, we can conclude that if we just satisfy nominal performance and robust stability with some margin then we also get robust performance.

Note that the robust performance criterion (6) is valid for SISO systems only. In contrast to most other results we present in this course, this analysis can (unfortunately) not be easily extended to MIMO systems.

1 EL2520 Lecture notes 4: Fundamental Limitations and Conflicts

In many control courses the focus is mainly on control design, and analysis of the resulting closed-loop system. With this approach it is easy to get led to believe that the control performance relies only on the ability of the control engineer to design a sufficiently good controller. Also, one can get the impression that if one cannot meet the desired performance with one control design, then one should simply try with a more advanced controller.

The fact is that there always are hard limits to what can be achieved with feedback for a given system, and these limits are set by the system itself. Thus, the only way to overcome the limits, given that they prevent acceptable control performance, is a redesign of the system itself. This underlines that the traditional sequential approach to system and control design, i.e., first design the system then the controller to achieve desired dynamic behavior, is not optimal. Rather, the two activities should be integrated to ensure that the desired performance can be achieved.

In this lecture we derive quantitative limitations on the achievable control performance for a given system. We will mainly consider performance in terms of the sensitivity function S and the complementary sensitivity function T. As before, we specify the performance using weights W_S and W_T, respectively, so that the design requirements are

$$\|W_S S\|_\infty \leq 1 \quad \Leftrightarrow \quad |S(i\omega)| \leq |W_S^{-1}(i\omega)| \ \forall \omega$$

$$\|W_T T\|_\infty \leq 1 \quad \Leftrightarrow \quad |T(i\omega)| \leq |W_T^{-1}(i\omega)| \ \forall \omega$$

A standard choice of weight for the sensitivity function S is

$$W_S = \frac{1}{M_s} + \frac{\omega_{BS}}{s} \tag{1}$$

corresponding to infinite weight at $\omega = 0$ (enforcing $S(0) = 0$), weight $|W_S(i\omega_{BS})| \approx 1$ (enforcing minimum bandwidth for S approx. ω_{BS}) and, finally, maximum peak of $|S|$ equal to M_S at frequencies above ω_{BS}. Similarly, a standard weight for T is

$$W_T = \frac{1}{M_T} + \frac{s}{\omega_{BT}} \tag{2}$$

corresponding to allowing maximum peak M_T, minimum bandwidth ω_{BT} and enforcing $T = 0$ at infinite frequency. The inverse of these two weights, i.e., the corresponding bounds on $|S(i\omega)|$ and $|T(i\omega)|$, are shown in Figure 1.

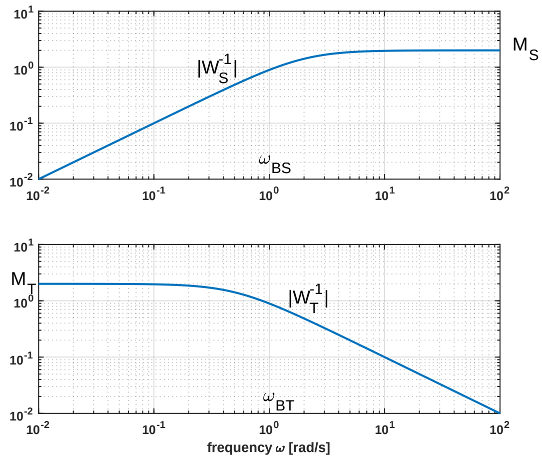


Figure 1.1: Bounds on $|S|$ and $|T|$ corresponding to the standard choice of weights used in this lecture.

1.1 Sensitivity Trade Off

With loop-gain $L = GF_y$ we have

$$S = \frac{1}{1+L} \; ; \quad T = \frac{L}{1+L}$$

and hence

$$S(i\omega) + T(i\omega) = 1 \quad \forall\omega$$

This implies that either $|S(i\omega)| > 0.5$ or $|T(i\omega)| > 0.5$ at any frequency, that is, we can not deal effectively with both disturbances and noise at the same frequency. In terms of the weights, a consequence is that we can not choose $|W_S| > 2$ and $|W_T| > 2$ at the same frequency.

Also note that since $|S(i\omega)+T(i\omega)| = 1$, the distance between $S(i\omega)$ and $-T(i\omega)$ is always 1 and hence

$$|S(i\omega)| >> 1 \quad \Leftrightarrow \quad |T(i\omega)| >> 1$$

at any given frequency ω. Thus, a large peak in $|S|$ implies a large peak in $|T|$, and vice versa.

1.2 The Bode Sensitivity Integral

In Lecture 2 we showed by simple arguments, based on the Nyquist plot of the loop gain, that the sensitivity function of any closed-loop system must exceed $|S| > 1$ at some frequency (provided the loop-gain has relative degree at least 2). A more powerful result is due to Bode:

Bode Integral Theorem: Suppose that the loop-gain $L(s) = GF_y(s)$ has relative degree (pole excess) ≥ 2, and that $L(s)$ has N_p poles in the RHP located at $s = p_i$. Then the sensitivity function must satisfy

$$\int_0^\infty \log|S(i\omega)|d\omega = \pi\Sigma_{i=1}^{N_p} Re(p_i)$$

The proof is based on Cauchy integral theorem (see course on complex analysis).

Thus, if we plot $\log|S|$ vs ω, then the area for which $|S| > 1$ ($\log|S| > 0$) must at least equal the area for which $|S| < 1$ ($\log|S| < 0$). If the system is open-loop stable ($N_p = 0$) then the two areas must exactly match. Thus, if we push down the sensivitiy in one frequency range we must pay for this by increasing the sensitivity in another frequency range. This is commonly known as the *waterbed effect* in control. Thus, we are enforced to make a trade-off between the performance in different frequency ranges.

1.3 Interpolation Constraints from RHP Poles and Zeros

Consider again the loop-gain

$$L(s) = G(s)F_y(s)$$

If $G(s)$ has a zero at $s = z$ in the complex RHP, then $G(z)$=0 and $L(z) = 0$. The last equality follows from the fact that we are not allowed to cancel RHP zeros in G by corresponding RHP poles in the controller $F_y(s)$ since then we do not get internal stability (see Lecture 2). The implications for the sensitivity functions are

$$S(z) = \frac{1}{1+L(z)} = 1\ ; \quad T(z) = \frac{L(z)}{1+L(z)} = 0 \tag{3}$$

Similarly, if $G(s)$ has a pole at $s = p$ in the complex RHP, then $G(p) = \infty$ and $L(p) = \infty$, and

$$S(p) = \frac{1}{1+L(p)} = 0\ ; \quad T(p) = \frac{L(p)}{1+L(p)} = 1 \tag{4}$$

These are interpolation constraints on $S(s)$ and $T(s)$ that any stabilizing controller must satisfy.

To see what the interpolation constraints above implies for the bounds on $|S(i\omega)|$ and $|T(i\omega)|$ we need the following result, known as the maximum modulus theorem in complex analysis

Maximum Modulus Theorem: Suppose that Ω is a region in the complex plane and that F is an analytic function on Ω and, furthermore, that F is not equal to a constant. Then $|F|$ attains its maximum value at the boundary of Ω.

For a proof, see a course book on complex analysis.

Since S and T are stable transfer-functions, they have no singularities (poles) in the RHP and are hence analytic in the complex RHP for which the boundary is the $i\omega$-axis. A trivial consequence of (3)-(4) and the maximum modulus theorem is then

$$\|S\|_\infty = \sup_\omega |S(i\omega)| > |S(z)| = 1\ ; \quad \|T\|_\infty = \sup_\omega |T(i\omega)| > |T(p)| = 1 \tag{5}$$

The bounds in (5) are not too useful as such since all they say is that we must have a peak in $|S|$ and $|T|$ exceeding 1, which we already knew applies to essentially any system, also those with no RHP poles and zeros (see Lecture 2). Too obtain more informative bounds, we include weights on S and T.

Consider first the weighted sensitivity for which the interpolation constraint and maximum modulus theorem implies

$$\|W_S S\|_\infty > |W_S(z)S(z)| = |W_S(z)|$$

Thus, in order to achieve $\|W_S S\|_\infty \leq 1$ the weight must fulfill the constraint

$$|W_S(z)| \leq 1$$

To get some insight into what this constraint implies, consider the weight (1)

$$W_S(s) = \frac{1}{M_S} + \frac{\omega_{BS}}{s}$$

where ω_{BS} is the required bandwidth, i.e., the frequency where $|S|$ becomes (approximately) larger than 1, and M_S the maximum allowed peak of $|S|$. Then

$$|W_S(z)| \leq 1 \quad \Rightarrow \quad \frac{1}{M_S} + \frac{\omega_{BS}}{z} \leq 1$$

and from this we derive

$$\omega_{BS} \leq (1 - M_S^{-1})z$$

Thus, if we allow a peak $M_S = \infty$ then the maximum bandwitdh $\omega_{BS} = z$. A more reasonable maximum peak is $M_S = 2$ for which we get the bandwidth limitation

$$\omega_{BS} \leq \frac{z}{2} \tag{6}$$

Thus, a RHP zero places a hard limitation on the achieveable bandwidth of a control system and the limitation is worse the closer the zero is to the imaginary axis.

Example 1: Consider the plant

$$G(s) = \frac{s-1}{(s+1)(s+2)}$$

Since the plant has a RHP zero at $z = 1$, we can not achieve a bandwidth for the sensitivity larger than 1 rad/s, or we if allow a maximum peak $M_S = 2$, not larger than 0.5 rad/s. Thus, we can not attenuate disturbances at higher frequencies.

A RHP zero is known as *non-minimum phase* since there exist transfer-functions with the same amplitude $|G(i\omega)|$ but with less negative phase. For instance, the plant $G = 1/(s+2)$ has the same amplitude curve as the plant in Example 1, but more positive phase. More negative phase implies that a system is harder to control. Another non-minimum phase phenomena is a time-delay. A time-delay θ has transfer-function $e^{-\theta s}$. By employing a 1st order Padé approximation of a delay we get

$$e^{-\theta s} \approx \frac{1 - \frac{\theta}{2}s}{1 + \frac{\theta}{2}s}$$

and hence a time-delay of θ can be seen as a RHP zero at $z = 2/\theta$. Then, the bounds derived above for a RHP zero yields, with $M_S = 2$,

$$\omega_{BS} \leq \frac{1}{\theta}$$

Consider next the interpolation constraint $T(p) = 1$ at a RHP pole p of $G(s)$. With the introduction of the weight W_T, the maximum modulus theorem yields

$$\|W_T T\|_\infty > |W_T T(p)| = |W_T(p)|$$

and hence we require

$$|W_T(p)| \leq 1$$

to achieve $\|W_T T\|_\infty \leq 1$. With the specific weight (2) we get

$$\frac{p}{\omega_{BT}} + \frac{1}{M_T} \leq 1$$

and with $M_T = 2$ we get the bandwidth limitation

$$\omega_{BT} \geq 2p \tag{7}$$

Note that this is a lower bound on the bandwidth of T and is as expected since we need a minimum bandwidth to stabilize an unstable system.

Given the fact that a RHP zero imposes an upper bound on the bandwidth of S and a RHP pole imposes a lower bound on the bandwidth of T, and these bandwidths must be close since $S + T = 1$, we would expect difficulties if a plant has both RHP zeros and RHP poles and these are close. To quantify this, recall that $S(p) = 0$ at a RHP pole p of $G(s)$. Factorizing the sensitivity function

$$S = S_{mp} \underbrace{\frac{s-p}{s+p}}_{S_{ap}}$$

where S_{mp} has no poles or zeros in the RHP and S_{ap} is all-pass, i.e., has amplitude 1 $\forall\omega$. If $G(s)$ also has a RHP zero at $s = z$, then from the constraint $S(z) = 1$ we get

$$S_{mp}(z) = S_{ap}^{-1}(z) = \frac{z+p}{z-p}$$

Hence

$$\|W_S S\|_\infty = \|W_S S_{mp}\|_\infty \geq |W_S(z) S_{mp}(z)| = |W_S(z)\frac{z+p}{z-p}|$$

For instance, with weight $W_S = 1$ we obtain

$$\|S\|_\infty \geq \frac{|z+p|}{|z-p|}$$

Thus, with poles and zeros close in the RHP in $G(s)$, large peaks in $|S|$ is unavoidable.

For the complementary sensitivity T we have $T(z) = 0$ and factorizing as for S above we get

$$\|W_T T\|_\infty \geq |W_T(p)\frac{p+z}{p-z}|$$

and with $W_T = 1$ we get

$$\|T\|_\infty \geq \frac{|z+p|}{|z-p|}$$

and large peaks in $|T|$ must also exist with z and p close, even with the best possible controller. This is as expected given the result for $\|S\|_\infty$ above and the fact that the difference between $\|T\|_\infty$ and $\|S\|_\infty$ is at most 1 since $S + T = 1$.

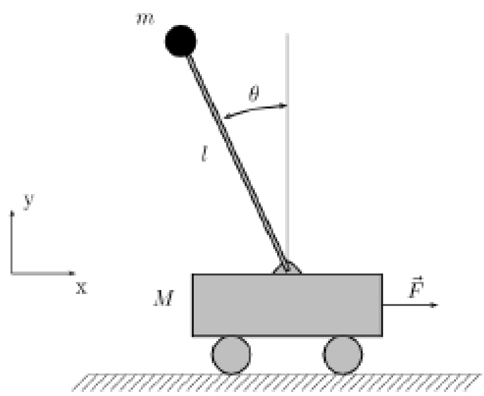


Figure 1.2: Inverted pendulum on a cart

Example 2: Consider the inverted pendulum placed on a cart in Figure 2. The aim is to stabilize the pendulum in an upright position using the movement of the wagon by the drag force F. Consider the position x of the wagon as the output, then the transfer-function from the input F is

$$X(s) = \frac{ls^2 - g}{s^2(Mls^2 - (M+m)g)}F(s)$$

The system has both a RHP zero and a RHP pole

$$z = \sqrt{\frac{g}{l}} \; ; \quad p = z\sqrt{1 + m/M}$$

For a pendulum with length $l = 1$ and mass of the pendulum equal to the mass of the wagon, i.e., $m = M$, we get $z = \sqrt{10}$ and $p = \sqrt{20}$ and

$$\|S\|_\infty \geq 5.8 \; ; \quad \|T\|_\infty \geq 5.8$$

If we reduce the mass of the pendulum to one tenth, i.e., $m = 0.1M$, we get $z = \sqrt{10}$ and $p = \sqrt{11}$ and

$$\|S\|_\infty \geq 42 \; ; \quad \|T\|_\infty \geq 42$$

Note that the rocket stabilization problem that puzzled scientists in the late 1950s and early 1960s corresponds to the cart pendulum problem since the air under the rocket becomes fluidized. From the result above one can understand why the problem was challenging. This is a good example of a system where it is crucial to consider integrated design of system and controller to ensure satisfactory dynamic performance.

In summary, we have shown that there exist several trade-offs and inherent limitations in control, and these are fundamental in the sense that no controller can overcome them. Thus, if a limitation is in conflict with the desired dynamic performance of a system, then the only viable solution is to redesign the system to reduce the limitations. Note that all results presented here are valid for SISO systems only. In Lecture 6 we will extend the results to the MIMO case.

5 EL2520 Lecture notes 5: Basics of Multivariable Control Systems, Decentralized Control and the RGA

So far the focus of this course has been on SISO systems. However, as stressed before, the main purpose has been to introduce a framework that enables us to more or less treat SISO and MIMO systems in the same way. Historically, the classical methods developed in the input-output, or frequency, domain by the pioneers of the likes of Nyquist, Bode and Black in the 1930s-1940s, were limited to SISO systems and not extendible to the MIMO case. As part of the space race in the 1950s-1960s one needed to solve challenging MIMO control problems and then moved to state space methods combined with optimization theory, resulting in the now famous state space optimal control theory. Some key names in this period were Bellman, Pontryagin and Kalman. Despite being an elegant solution to many control problems, one major drawback of the state space methods were that there was no obvious connection to the classial results in the frequency domain. In particular, while the latter addressed model uncertainty, e.g., in terms of phase and amplitude margins, as well as fundamental limitations, these issues were not easily addressed in the state-space domain. Indeed, in the 1980s George Zames at MIT noted that state space optimal methods for MIMO systems could be highly sensitive to model errors. To address this problem, he introduced an input-output framework suitable for MIMO systems, based on the H_∞-norm. On the basis of this, powerful results on robustness and performance limitations in MIMO control systems could be developed. Other key names from this period are John Doyle and Gunter Stein. It is the latter theory which is the focus of this course. However, as we shall see later in the course, there are also clear links to the state-space optimal control theory, not the least when it comes to solving the H_∞-optimal control problems formulated in the input-output (frequency) domain but typically solved in state-space.

Before extending the SISO results on robust stability, performance specifications and performance limitations to MIMO systems, we need to introduce some basic properties of MIMO systems, such as poles, zeros and gain. We will also briefly discuss more simplistic approaches to MIMO control, namely decentralized control and decoupling.

5.1 Transfer-Matrices, Poles and Zeros

Recall that a multi-input-multi-output (MIMO) system have inputs and outputs that are vector quantities

$$u = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_m \end{pmatrix} \quad ; \quad y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{pmatrix}$$

while the corresponding transfer-function $G(s)$ is a $p \times m$ matrix in which the ij-th element $G_{ij}(s)$ is the transfer-function from u_j to y_i. Recall also that if we have a linear time-invariant model on state-space form

$$\begin{aligned} \dot{x} &= Ax(t) + Bu(t) \; ; \quad x \in \mathbb{R}^n, u \in \mathbb{R}^m \\ y(t) &= Cx(t) + Du(t) \; ; \quad y \in \mathbb{R}^p \end{aligned}$$

then the transfer-matrix is

$$G(s) = C(sI - A)^{-1}B + D$$

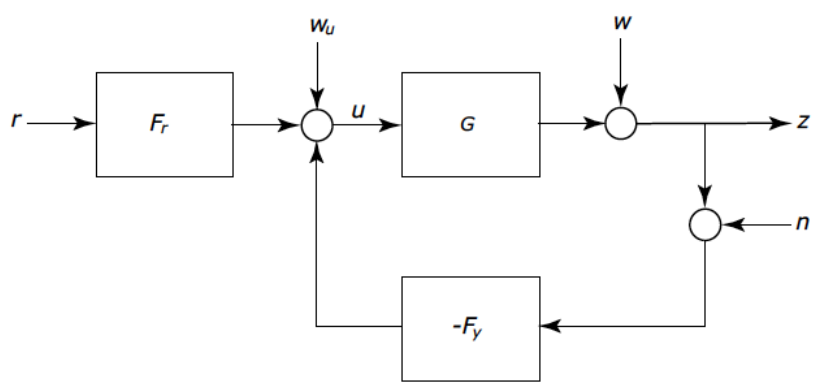


Figure 5.1: Block-diagram for two degree of freedom control system.

Consider the block-diagram in Figure 5.1. To derive closed-loop transfer-functions between inputs and outputs, we can as usual employ simple block-diagram algebra, i.e., write down relations from the block-diagram and solve for the inputs and outputs of interest. However, a simple rule that can be used instead is the following

1. Start from the output and move against the signal flow towards the input.
2. Write down the blocks, left to right, as you meet them
3. When you exit the loop, postmultiply by the term $(I+L)^{-1}$ where L is the loop-gain evaluated from the exit and against the signal flow
4. If there are parallell paths, simply add them toghether

Also useful is the so called "push through" rule (for matrices of appropriate dimensions)

$$A(I + BA)^{-1} = (I + AB)^{-1}A$$

For instance, consider the transfer-function from the noise n to the output z in Figure 5.1. Moving from z we get

$$z = -GF_y(I + GF_y)^{-1}n = -(I + GF_y)^{-1}GF_yn = -Tn$$

Also

$$z = (I + GF_y)^{-1}w = Sw \ ; \quad z = G(I + F_yG)^{-1}w_u = (I + GF_y)^{-1}Gw_u = SGw_u$$

and

$$u = (I + F_yG)^{-1}w_u = S_u w_u$$

Note that $(I + GF_y)^{-1} \neq (I + F_yG)^{-1}$ and hence the sensitivity at the input S_u differs from the sensitivity at the output S.

The pole polynomial $\lambda(s)$ of a linear system is the characteristic polynomial of the A matrix in the state space description

$$\lambda(s) = \det(sI - A)$$

The poles p_i are the zeros of the pole polynomial $\lambda(p_i) = 0$. The poles can be computed from the transfer-matrix using the following theorem

Theorem: The pole polynomial $\lambda(s)$ of a system with transfer-matrix $G(s)$ is the least common denominator of all minors of $G(s)$.

Recall that a minor of a matrix M is the determinant of any square matrix obtained by deleting one or more columns and rows of M. For a 2×2 matrix the minors are the matrix elements and the determininant of the matrix itself. Consider the example

$$G(s) = \begin{pmatrix} \frac{2}{s+1} & \frac{3}{s+2} \\ \frac{1}{s+1} & \frac{1}{s+1} \end{pmatrix}$$

which has minors $\frac{2}{s+1}, \frac{3}{s+2}, \frac{1}{s+1}$ and $\det G(s) = \frac{1-s}{(s+1)^2(s+2)}$. The characteristic polynomial is then $\lambda(s) = (s+1)^2(s+2)$ and hence there are two poles in $s = -1$ and one pole in $s = -2$.

The zeros of a SISO transfer-function $G(s)$ are those values of s where $G(s) = 0$. For a MIMO transfer-matrix, the most common corresponding definition of zeros (and the one we use in this course) are the values of s where $G(s)$ looses rank, i.e., has lower rank than for other values of s (except those of possible other zeros).

Theorem: The zero polynomial of $G(s)$ is the greatest common divisor of the maximal minors of $G(s)$, normed so that they have the pole polynomial of $G(s)$ as denominator. The zeros of $G(s)$ are the roots of the zero polynomial.

Note that the maximal minor of a square matrix is the determinant of the matrix. Consider again the 2×2 system above, for which the determinant is

$$\det G(s) = \frac{1-s}{(s+1)^2(s+2)}$$

which is normed so that the denominator is the pole polynomial. Thus, the zero polynomial is $(1 - s)$ and there is one zero at $s = 1$. You can easily check that $G(1)$ has rank 1.

When a matrix has a singularity or becomes rank deficient, these is always an associated subspace. For instance, if $G(z)$ is rank deficient then

$$G(z)u_z = 0 \cdot y_z$$

Here u_z and y_z are the input and output *zero directions*, respectively. Thus, u_z is in the nullspace of $G(z)$ while y_z^H is in the left nullspace of $G(z)$ where superscript H denotes conjugate (Hermetian) transpose. The latter follows from the fact that $y_z^H G(z) = 0 \cdot u_z^H$ where we have assumed that both directions are normalized to have length one, i.e., $y_z^H y_z = 1, u_z^H u_z = 1$. Likewise, at a pole p we have essentially

$$G(p)u_p = \infty \cdot y_p$$

where u_p and y_p are the input and output *pole directions*. We shall see in Lecture 6 that these directions are important when analyzing the limitations imposed on control performance by RHP poles and RHP zeros. Also note that a pole and a zero located at the same position in two transfer-matrices G_1 and G_2, respectively, does not necessarily cancel eachother when forming $G_1 G_2$. They only cancel if the corresponding pole and zero directions coincide.

5.2 Gain of a MIMO system

For a linear SISO system $G(s)$ we know that the amplification of the system depends on the frequency ω of the input, i.e.,

$$\frac{|Y(i\omega)|}{|U(i\omega)|} = |G(i\omega)|$$

Recall from Lecture 1 that the *gain* of a system was defined as the maximum amplification, and for a linear SISO system this is then

$$\sup_u \frac{\|y\|_2}{\|u\|_2} = \sup_\omega |G(i\omega)| = \|G\|_\infty$$

Thus, the gain a linear time invariant SISO system is the $\mathcal{H}_\infty$-norm of the transfer-function.

In a MIMO system, the amplification of the system will also depend on the direction of the input vector. To see this, consider first a static linear system

$$y = Ax$$

where A is real or complex constant matrix of dimension $m \times r$. As before, we consider amplification in terms of the 2-norm of the output over the 2-norm of the input, i.e., $|y|/|x|$ where $|\cdot|$ denotes the Euclidian 2-norm. We have

$$|y|^2 = |Ax|^2 = (Ax)^H Ax = x^H A^H Ax$$

Here $A^H A$ is a symmetric positive definite matrix and hence

$$|x|^2 \lambda_{min}(A^H A) \leq |y|^2 \leq |x|^2 \lambda_{max}(A^H A)$$

(follows from $A^H Ax = \lambda x \;\Rightarrow\; x^H A^H Ax = \lambda x^H x = \lambda |x|^2$). We then get the lower and upper bounds on the amplification

$$\underbrace{\sqrt{\lambda_{min}(A^H A)}}_{\underline{\sigma}(A)} \leq \frac{|y|}{|x|} \leq \underbrace{\sqrt{\lambda_{max}(A^H A)}}_{\bar{\sigma}(A)}$$

where $\underline{\sigma}(A)$ and $\bar{\sigma}(A)$ are the smallest and largest *singular values* of A, respectively. The corresponding input and output directions can be obtained from the Singular Value Decomposition (SVD) of A

$$A = U\Sigma V^H = [u_1 \ \cdots \ u_r]\mathtt{diag}(\sigma_i)[v_1 \ \cdots \ v_r]^H = \sum_{i=1}^{r} \sigma_i u_i v_i^H$$

Here U and V are orthonormal matrices (all columns orthogonal to each other and each of unit length) and Σ is a diagonal matrix with the singular values in descending order, i.e., $\sigma_1 = \bar{\sigma}$, $\sigma_r = \underline{\sigma}$. The input-output interpretation is that an input in the direction v_i gives an output in the direction u_i[1] and the amplification is σ_i. The maximum amplification is then $\bar{\sigma}(A)$. Thus, if we consider the frequency response of $G(s)$ at a given frequency

$$\underline{\sigma}(G(i\omega)) \leq \frac{|Y(i\omega)|}{|U(i\omega)|} \leq \bar{\sigma}(G(i\omega))$$

Considering the maximum amplification over both frequency and direction we get

$$\sup_u \frac{\|y\|_2}{\|u\|_2} = \sup_\omega \bar{\sigma}(G(i\omega)) = \|G\|_\infty$$

Thus, just like in the SISO case, the gain of a linear MIMO system is given by the $\mathcal{H}_\infty$-norm of the transfer-matrix $G(s)$.

5.3 Decentralized Control and the RGA

The rest of this course will focus on analyzing and designing multivariable controllers for multivariable plants, based on the framework introduced for SISO systems in Lecture 1-4. However, before moving on to discuss the extension of these results to the MIMO case, we shall briefly discuss one specific type of control structure which is relatively common in practical applications, namely *decentralized control*. In decentralized control one chooses to control each output y_i with only one input u_j, i.e,

$$u_j = F_{yi}(s)(r_i - y_i)$$

where the controller $F_{yi}(s)$ is a SISO transfer-function. Rearranging the inputs and outputs this implies that the full controller can be written on diagonal form

$$F_y(s) = \mathtt{diag}(F_{y1}(s) \ldots F_{y_p}(s))$$

[1]It may be somewhat confusing that we use U and u_j to denote output directions, since u normally is used for inputs in control, but we do this to follow the standard nomenclature for SVD.

Even though $F_y(s)$ is diagonal, the fact that the plant $G(s)$ in general is a full matrix implies that the closed-loop transfer-matrix

$$G_c(s) = G(s)F_y(s)(I + GF_y(s))^{-1}$$

also will be a full matrix. Thus, a setpoint change in any r_i will give a response in all outputs and all controllers F_{yi} will respond. This is usually termed *interactions* between the loops and can have a significant impact on both closed-loop performance and stability. Recall the examples discussed in Lecture 1. The strength of the interactions is in general dependent on how the inputs and outputs are paired, and it is advisable to choose a pairing that minimizes the interactions. The Relative Gain Array (RGA) is a model based tool that can be used to select such a pairing. We next derive the RGA and then discuss its use in selecing input-output pairings in decentralized control.

The idea behind the RGA is to quantify the impact of other control loops on the loop involving input u_j and output y_i. The Relative Gain for input u_j and output y_i is defined as the ratio

$$\lambda_{ij} = \frac{(y_i/u_j)_{\texttt{all loops open}}}{(y_i/u_j)_{\texttt{all other loops closed}}}$$

Here the numerator is simply the corresponding transfer-function element $G_{ij}(s)$. To determine the transfer-function (y_i/u_j) when all other loops are closed we must make some assumption on the controllers used. In the RGA, it is assumed that all other outputs are perfectly controlled, i.e., $y_{k,k\neq i} = 0$. Note that $u = G^{-1}y$ and hence the transfer-function we seek is

$$\left(\frac{y_i}{u_j}\right)_{y_{k,k\neq i}} = \frac{1}{(G^{-1})_{ji}}$$

Thus, the Relative Gain for the loop with output y_i and input u_j is

$$\lambda_{ij} = G_{ij}(G^{-1})_{ji} \tag{1}$$

and the Relative Gain Array (RGA) for all possible pairings become

$$\Lambda(G) = G(s) \times (G^{-1}(s))^T \tag{2}$$

where $\times$ denotes the Hadamard product, i.e., element-wise product. Since we prefer pairings with weak interactions with other loops we should choose pairings with a relative gain close to 1. In general, the RGA will be frequency dependent and then the most criticial is that interactions does not have a large impact around the crossover frequency ω_c of the loop considered. That is, we should prefer pairings with $\lambda_{ij}(i\omega_c) \approx 1$. Also note that if $\lambda_{ij}(0) < 0$ it implies that the steady-state gain of the loop changes sign as the other loops are closed, and negative feedback becomes positive feedback in the loop considered. Hence we should always avoid such pairings. This leads to the common rules for selecting pairings based on the RGA

1. Avoid pairings with $\lambda_{ij}(0) < 0$.
2. Prefer pairings with $\lambda_{ij}(i\omega_c)$ close to 1 (values between 0.5 and 3 are typically considered acceptable).

Note that if it is not possible to find any pairings that satisfy this rule, then it probably means it will be hard to obtain acceptable performance with decentralized control and one should instead consider true multivariable control, i.e., a full $F_y(s)$.

5.4 Decoupling

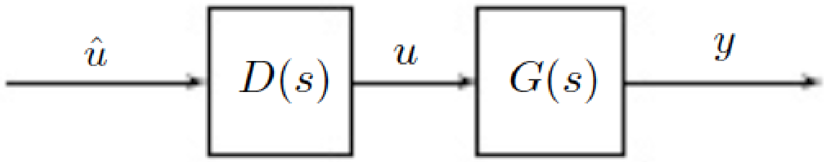


Figure 5.2: Pre-compensator for decoupling

When a plant has strong interactions, an apparantly simply way to resolve the problem is to remove the interactions by means of a compensator. Consider a plant $y = G(s)u$ and introduce the pre-compensator $u = D(s)\hat{u}(s)$. See Figure 5.2. Then

$$y = G(s)D(s)\hat{u}(s)$$

We can now choose the compensator $D(s)$ so that $G(s)D(s)$ has weak interactions. In particular, we can choose to have $G(s)D(s)$ diagonal, i.e., no interactions and $\Lambda(GD) = I$. A simple choice is then $D(s) = d(s)G^{-1}(s)$ where $d(s)$ are dynamics added to make $D(s)$ proper, i.e., have at least as many poles as zeros. The result is

$$G(s)D(s) = d(s)I$$

i.e., a diagonal system corresponding to a collection of r SISO systems.

Example 1: Consider the 2×2 system

$$G(s) = \frac{1}{s+1}\begin{pmatrix} 1 & -1 \\ 1.1 & -1 \end{pmatrix}$$

which has RGA

$$\Lambda = \begin{pmatrix} -10 & 11 \\ 11 & -10 \end{pmatrix}$$

at all frequencies. Thus, it is not recommended to use decentralized control. A decoupler is then

$$D(s) = \begin{pmatrix} 1 & -1 \\ 1.1 & -1 \end{pmatrix}^{-1} = \begin{pmatrix} -10 & 10 \\ -11 & 10 \end{pmatrix}$$

which yields

$$G(s)D(s) = \frac{1}{s+1}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

The decoupling above may at first sight seem to more or less solve all problems related to MIMO control; simply use a decoupler and you then have a number of SISO problems. However, it is not as simple as that. First, decoupling may require large input moves which can be costly. Second, the decoupling may not be a good solution when considering disturbance attenuation. Third, one can not employ the inverse of a model which has RHP poles or zeros, or time delays, since then internal stability is lost. However, the most important reason why one should be careful about using decoupling is uncertainty.

If one does not have a perfect model of the plant, then the result of the decoupler may be far from the expected. Consider the example above again, but now add 10% uncertainty to each of the elements of the model G when computing the decoupler

Example 1, cont'd: Assume our model of G is

$$\hat{G} = \frac{1}{s+1}\begin{pmatrix} 1.1 & -0.9 \\ 1.2 & -0.9 \end{pmatrix}$$

Then the decoupler becomes

$$D(s) = \hat{G}^{-1} = \begin{pmatrix} -10 & 10 \\ -13.3 & 12.2 \end{pmatrix}$$

and the compensated plant

$$G(s)D(s) = \frac{1}{s+1}\begin{pmatrix} 3.3 & -2.2 \\ 2.3 & -1.2 \end{pmatrix}$$

which is far from diagonal.

In fact, it can be shown that plants with large RGA numbers are highly sensitive to uncertainty in their inverse. One result that reflects this is the following

Theorem: A matrix G becomes singular if we make a relative change of $-1/\lambda_{ij}$ in element G_{ij}, where λ_{ij} is element i, j of the RGA of G. That is, if we change element G_{ij} to $G_{p,ij} = G_{ij}(1 - \frac{1}{\lambda_{ij}})$ then the matrix becomes singular.

For the example above we have $\lambda_{11} = 10$ and hence changing G_{11} by a relative factor 0.1 to $G_{11} = 1.1$ makes the model singular (with infinite RGA-elements).

In summary, one should be careful with using decoupling due to sensitivity to model uncertainty. In particular, this applies to plant with large RGA numbers (which somewhat ironically are plants for which decoupling is needed the most). Thus, it is better to solve the overall control problem, including some description of the model uncertainty, using optimization based approaches as we will discuss later in this course.

6 EL2520 Lecture notes 6: Performance Limitations in Multivariable Control

In this lecture we will extend the results on performance specifications and performance limitations in SISO systems, covered in Lecture 4, to the more general case of MIMO systems. We start by considering internal stability for MIMO systems since stability is a pre-requisite for any consideration of performance.

6.1 Internal Stability

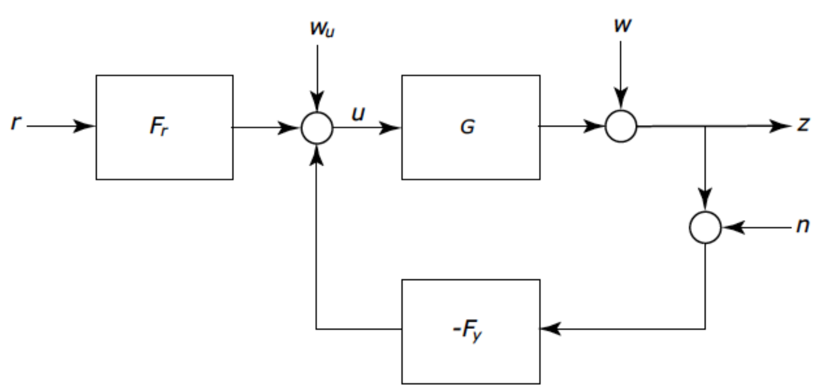


Figure 6.1: Two-degree of freedom control system.

Consider the closed-loop system in Figure 6.1. Recall that a system is internally stable if and only if it is input-output stable from any input to any output. It suffices to consider one input and one output on either side of the two blocks in the loop in Figure 6.1. Let us for instance consider the inputs w_u and w and the outputs u and z, for which we derive the transfer-functions

$$z = \underbrace{(I+GF_y)^{-1}}_{S} w + \underbrace{(I+GF_y)^{-1}G}_{SG} w_u \quad (1)$$

$$u = \underbrace{(I+F_yG)^{-1}F_y}_{S_uF_y} w + \underbrace{(I+F_yG)^{-1}}_{S_u} w_u \quad (2)$$

Thus, we have internal stability if S, SG, S_u, S_uF_y all stable. Note that we must also require the pre-filter $F_r(s)$ to be stable since it is outside the feedback loop. Note that cancellations of poles and zeros in the RHP between $G(s)$ and $F_y(s)$ always will cause instability of at least one of these transfer-functions and hence internal instability. On the other hand, if there are no cancellations in the RHP, then it suffices to check stability of one the four transfer-functions (in addition to $F_r(s)$).

Example: Consider the following plant and controller

$$G = \frac{1}{s+1}\begin{pmatrix} s+1 & 1 \\ 2 & 1 \end{pmatrix} \; ; \quad F_y = \frac{1}{s}G^{-1}(s)$$

The loop-gain is

$$L = G(s)F_y(s) = \frac{1}{s}I$$

and the sensitivity and complementary sensitivity functions are

$$S = \frac{s}{s+1}I \; ; \quad T = \frac{1}{s+1}I$$

which both are stable. However, if we consider the transfer-function from a disturbance on the output w to the control input u, we find

$$u = S_u F_y w = F_y S w = \frac{1}{s-1}\begin{pmatrix} 1 & -2 \\ -1 & s+1 \end{pmatrix} w$$

which is unstable and hence the closed-loop is unstable (note that $G(s)$ has a RHP zero at $s = 1$ which is canceled by an equivalent pole in $G^{-1}(s)$).

We finally remark that the fact that we are not allowed to cancel zeros or poles in the RHP implies that these always will appear as corresponding RHP zeros of $T(s)$ and $S(s)$, respectively, just like in the SISO case.

6.2 Performance Specifications

Consider again the closed-loop system in Figure 6.1. Let us for now limit our discussion to the problem of attenuating disturbances w and measurement noise n in the output z. From the block diagram we derive

$$z = (I + GF_y)^{-1}w - (I + GF_y)^{-1}GF_y n = Sw - Tn$$

Thus, we should make the sensitivity function S "small" for disturbance attenuation and the complementary sensitivity T "small" to avoid noise being amplified in the output. At each frequency ω we have have (see also Lec 5)

$$\underline{\sigma}(S(i\omega)) \leq \frac{|z|}{|w|} \leq \bar{\sigma}(S(i\omega)) \; ; \quad \underline{\sigma}(T(i\omega)) \leq \frac{|z|}{|n|} \leq \bar{\sigma}(T(i\omega))$$

depending on the direction of w and n. Thus, to bound the output $|z|$ in the presence of disturbances w and noise n in any direction we must bound $\bar{\sigma}(S)$ and $\bar{\sigma}(T)$. Similar to what we did for SISO systems, it is then natural to specify a frequency dependent bound on the sensitivity and complementary sensitivity, respectively

$$\bar{\sigma}(S(i\omega)) \leq |W_S^{-1}(i\omega)| \; \forall\omega \quad \Leftrightarrow \quad \|W_S S\|_\infty \leq 1$$

$$\bar{\sigma}(T(i\omega)) \leq |W_T^{-1}(i\omega)| \; \forall\omega \quad \Leftrightarrow \quad \|W_T T\|_\infty \leq 1$$

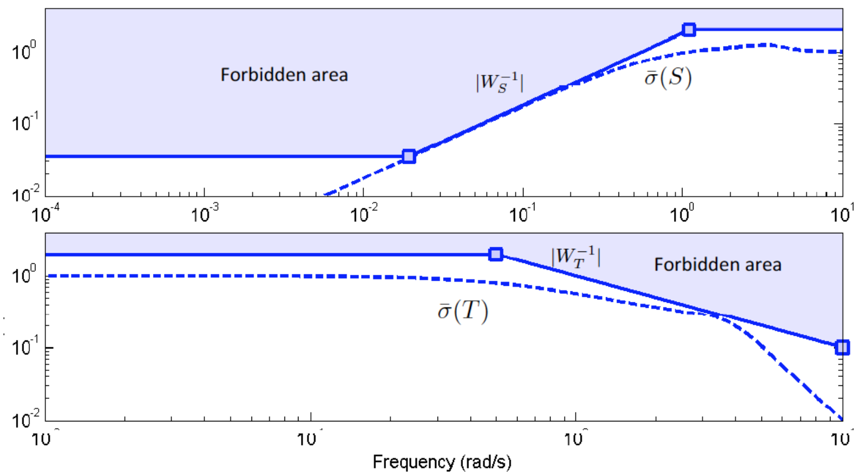


Figure 6.2: Performance bounds for maximum singular values of S and T, respectively. The bounds corresponds to the performance objective $\|W_S S\|_\infty <$ and $\|W_T T\|_\infty < 1$.

Thus, we have performance specifications in terms of the $\mathcal{H}_\infty$ norm of the weighted sensitivity functions, just like we derived for SISO systems earlier. The bounds are illustrated in Figure 6.2.

The control design problem then essentially consists of defining the weights W_S and W_T and then solving the resulting optimization problem[1]. Since the objective of the optimization is to make $\|W_S S\|_\infty \leq 1, \|W_T T\|_\infty \leq 1$, it is important to choose the weights W_S and W_T so that they not only reflect our control objectives, but also reflect any fundamental limitations and conflicts that must be satisfied. We therefore next discuss limitations and conflicts that must be respected when choosing the weights W_S amd W_T.

6.2.1 A note on scaling

In order to simplify the analysis of control requirements and limitations, and the design of weights, it is useful to scale all variables sucn that their expected/allowed magnitude is less than one. Consider for instance the unscaled problem

$$\hat{z} = \hat{G}(s)\hat{u} + \hat{G}_d(s)\hat{d}(s)$$

where $\hat{}$ is used to denote unscaled. Note that the disturbance on the output is $\hat{w} = \hat{G}_d(s)\hat{d}$. Assume now that acceptable control corresponds to keeping $|\hat{z}| \leq \hat{z}_{max}$, that the control input is constrained as $|\hat{u}| \leq \hat{u}_{max}$ and that the maximum expected disturbance is $\hat{d}_{max}$. Introduce the scaled variables

$$z = \frac{\hat{z}}{\hat{z}_{max}} \ ; \quad u = \frac{\hat{u}}{\hat{u}_{max}} \ ; \quad d = \frac{\hat{d}}{d_{max}}$$

[1]In many case, we may want to consider also other transfer-functions than S and T, but more about that in Lecture 7. For now we focus on S and T.

The scaled variables should then all vary in the interval $[-1, 1]$, provided we have acceptable performance. Introduce the scaling factors

$$D_d = \hat{d}_{max} \; ; \quad D_u = \hat{u}_{max} \; ; \quad D_z = \hat{z}_{max}$$

For MIMO systems the scaling factors will be different for different variables and then the scaling factors above will be diagonal matrices. The scaled model becomes

$$D_z z = \hat{G}(s)D_u u + \hat{G}_d(s)D_d d \quad \Rightarrow \quad z = D_z^{-1}\hat{G}(s)D_u u + D_z^{-1}\hat{G}_d(s)D_d d = G(s)u + G_d(s)d$$

With the scaling above, it is much easier to determine performance requirements for the closed-loop. In closed-loop we have for instance

$$z = SG_d d$$

and since the aim is to keep $|z| < 1$ for $|d| < 1$ we require

$$|SG_d| < 1 \;\forall\omega \quad \Rightarrow \quad \|SG_d\|_\infty < 1$$

For the SISO case, this implies that we need sensitivity reduction for frequencies where $|G_d(i\omega)| > 1$, and so this gives a lower bandwidth requirement on the sensitivity. Similar requirements for the MIMO case are derived below.

6.3 S+T=I

We have

$$S + T = (I + GF_y)^{-1} + (I + GF_y)^{-1}GF_y = I$$

To understand what limitations this impose on $\bar{\sigma}(S)$ and $\bar{\sigma}(T)$, and thereby our choice of W_S and W_T, we employ Fan's Theorem:

$$\sigma_i(A + B) \geq \sigma_i(A) - \bar{\sigma}(B)$$

Consider $\sigma_i = \bar{\sigma}$, i.e., the maximum singular value, and let $A = S, B = T$. Fan's Thm then yields

$$\bar{\sigma}(S) \leq 1 + \bar{\sigma}(T)$$

Similarly, let $A = I, B = -T$ or $A = -I, B = T$ which combined yields

$$\bar{\sigma}(S) \geq |1 - \bar{\sigma}(T)|$$

and thus

$$|1 - \bar{\sigma}(T)| \leq \bar{\sigma}(S) \leq 1 + \bar{\sigma}(T)$$

From this we conclude that we can not make both $\bar{\sigma}(S)$ and $\bar{\sigma}(T)$ small at the same frequency. Hence, we can not choose both $|W_S|$ and $|W_T|$ large at the same frequency. Also, if one is much larger than one at some frequency, then so is the other

$$\bar{\sigma}(S) >> 1 \quad \Leftrightarrow \quad \bar{\sigma}(T) >> 1$$

6.4 Extension of Bode Sensitivity Integral

Recall the Bode Sensitivity integral presented in Lecture 4. It can be shown that the determinant of a MIMO sensitivity function $\det S$ has the properties of a SISO sensitivity function. Application of Bode Sensitivity integral to $\det S$ then yields

$$\int_0^\infty \ln|\det S(i\omega)|d\omega = \sum_j \int_0^\infty \ln \sigma_j(S(i\omega))d\omega = \pi \sum_k \Re(p_i)$$

where we have used the fact that $|\det(S)| = \prod_j \sigma_j(S)$. The interpretation of the sensitivity integral is that we must trade-off sensitivity reduction at one frequency by a similar sensitivity increase at another frequency, but a similar trade-off can be made between different directions of the sensitivity function. That is, reducing sensitivity in one direction can be compensated for by an increased sensitivity in another direction.

The main conclusion to draw in terms of choice of the weight W_S is that it can not be chosen large at all frequencies and in all directions. Note that a scalar weight W_S often is preferred, which implies that all directions are given the same weight and the trade-off must then be made over frequencies in a similar way as for SISO systems.

6.5 RHP Zeros and Poles

We here show that the limitations from RHP zeros and poles, as derived in Lecture 4, carries over to the MIMO case more or less directly when considering the maximum singular values of S and T.

Theorem: Assume $G(s)$ has a RHP zero at $s = z > 0$. Then

$$\|W_S S\|_\infty \geq |W_S(z)|$$

Proof: By definition $G(z)$ is rank deficient, i.e., there exist a vector y_z such that

$$y_z^H G(z) = 0 \;\Rightarrow\; y_z^H T(z) = 0$$

Here y_z is the output zero direction as defined above. Now, $S + T = I$ and hence

$$y_z^H S(z) = y_z^H \;\Rightarrow\; S^H(z) y_z = y_z$$

which implies that

$$\bar{\sigma}(S(z)) \geq 1$$

where we have used the fact that $\bar{\sigma}(S^H) = \bar{\sigma}(S)$. Applying the Maximum Modulus Thm (see Lec 4) to the weighted sensitivity yields

$$\|W_S S\|_\infty \geq \bar{\sigma}(W_S(z)S(z)) \geq |W_S(z)|$$

where have assumed a scalar weight $W_S(s)$. Thus, we have the same restriction on W_S as derived for SISO systems in Lecture 4, i.e., the weight must satisfy the constraint

$$|W_S(z)| < 1$$

Consider next limitations imposed by RHP poles

Theorem: Assume $G(s)$ as a RHP pole at $s = p$. Then

$$\|W_T T\|_\infty \geq |W_T(p)|$$

Proof: As above, but with

$$S(p)y_p = 0 \;\Rightarrow\; T(p)y_p = y_p$$

Again, this is the same restriction as derived for SISO systems in Lecture 4 and we require that the weight on the complementary sensitivity is chosen so that

$$|W_T(p)| < 1$$

6.6 Requirements for Disturbance Attenuation

Consider a scalar disturbance d such that the disturbance on the output is

$$w = g_d(s)d \;\Rightarrow\; z = S(s)g_d(s)d \;, \quad |d| < 1 \; \forall w$$

Note that we assume the problem has been scaled such that the expected magnitude of the disturbance is less than 1. Assume we also have scaled the output z such that the performance requirement is $|z| < 1 \; \forall \omega$. This then gives the requirement

$$\bar{\sigma}(Sg_d) < 1 \; \forall \omega \;\Rightarrow\; \|Sg_d\|_\infty < 1$$

Define the disturbance direction as

$$y_d(i\omega) = \frac{g_d(i\omega)}{|g_d(i\omega)|}$$

The requirement then becomes

$$\bar{\sigma}(Sy_d) < \frac{1}{|g_d|} \quad \forall \omega$$

Thus, the requirement on the sensitivity S is only in the direction y_d. Perform a singular value decomposition (SVD) of S at a given frequency and consider in particular the high-gain and low-gain directions of S

$$S\bar{v} = \bar{\sigma}(S)\bar{u} \;, \quad S\underline{v} = \underline{\sigma}(S)\underline{u}$$

Now, if the disturbance direction y_d is completely aligned with $\bar{v}$ we get the requirement

$$\bar{\sigma}(S) < \frac{1}{|g_d|}$$

However, if y_d is aligned with the weak direction $\underline{v}$ we get the requirement

$$\underline{\sigma}(S) < \frac{1}{|g_d|}$$

Thus, the requirement imposed by a disturbance on the sensitivity function is highly dependent on which direction the disturbance acts in.

6.7 Disturbances and RHP Zeros

From the above we have seen that RHP zeros in the open-loop plant $G(s)$ impose limitations on the achievable maximum singular value of the sensitivity function $\bar{\sigma}(S)$. However, we also noted that a given disturbance may not be aligned with the worst direction of the sensitivity function and that we therefore in principle may achieve acceptable disturbance attenuation even if $\bar{\sigma}(S) > 1$ at a given frequency. The directions of the sensitivity S depends on the specific controller used. We here show that it is possible to determine if a RHP zero makes it fundamentally infeasible to get acceptable attenuation of a given disturbance with any controller.

Assume $G(s)$ has a RHP zero at $s = z$, then as discussed above

$$y_z^H S(z) = y_z^H$$

From the Maximum Modulus Thm we then get

$$\|Sg_d\|_\infty \geq \|y_z^H S g_d\|_\infty \geq |y_z^H g_d(z)|$$

Thus, for acceptable disturbance attenuation to be possible we must require

$$|y_z^H g_d(z)| < 1$$

Otherwise, no controller exists that will provide $\|Sg_d\|_\infty < 1$, i.e., acceptable disturbance attenuation.

Note that the size of the inner product $y_z^H g_d(z)$ depends on how the two vectors y_z and $g_d(z)$ are aligned. Two extreme cases are

- $y_z \perp g_d(z) \;\Rightarrow\; y_z^H g_d(z) = 0$, i.e., no apparant limitation from the RHP zero.
- $y_z \parallel g_d(z) \;\Rightarrow\; |y_z^H g_d(z)| = |g_d(z)|$, i.e., can in principle not atteunate disturbances for frequencies above $\omega = z$.

Example: Consider the 2×2 system

$$G(s) = \begin{pmatrix} \frac{2}{s+1} & \frac{1}{s+2} \\ \frac{s+3}{s+1} & \frac{2}{s+2} \end{pmatrix}$$

which has a zero at $s = 1$. To determine y_z, consider

$$G(1) = \begin{pmatrix} 1 & 1/3 \\ 2 & 2/3 \end{pmatrix} \quad \Rightarrow \quad y_z^H = \frac{1}{\sqrt{5}} \begin{pmatrix} -2 & 1 \end{pmatrix}$$

There are two disturbances affecting the output z. The first disturbance d_1 has the transfer-function

$$z = g_{d1}(s) d_1 \; ; \quad g_{d1} = \frac{2}{s+1} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

From this we compute $|y_z^H g_{d1}(1)| = 1/\sqrt{5} < 1$, and hence the RHP zero does not prevent acceptable disturbance attenuation. For the second disturbance d_2 we have

$$z = g_{d2}(s) d_2 \ ; \quad g_{d2} = \frac{2}{s+1} \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

and we get $|y_z^H g_{d2}(1)| = 3/\sqrt{5} > 1$. Thus, it is not possible to achieve acceptable disturbance attenuation of this disturbance with any controller due to the existence of the RHP zero at $s = 1$.

7 EL2520 Lecture notes 7: Robust Stability in MIMO systems, Loop Shaping

We start this lecture by deriving a robust stability condition for MIMO systems. We then move on to discuss how to design controllers that satisfy robustness and performance specifications in terms of norm bounds on weighted closed-loop transfer-functions, e.g., $\|W_S S\|_\infty < 1$ and $\|W_T T\|_\infty < 1$. As we show, this can either be done using classical shaping of the open-loop $L = GF_y$, or shaping the closed-loop transfer-functions directly using optimization.

7.1 Robust Stability

We consider modeling uncertainty in the same way as we did for SISO systems in Lecture 3. That is, we consider a set of models G_p obtained by perturbing the nominal model $G(s)$ by a stable perturbation $\Delta_o(s)$

$$\Pi_o = \{G_p(s) = (I + W_o(s)\Delta_o(s))G(s), \quad \|\Delta_o\|_\infty < 1\} \tag{1}$$

Note that this corresponds to modeling the uncertainty at the output side. See also Figure 7.1. The stable perturbation $\Delta_o(s)$ is in general a full matrix of the same dimension as $G(s)$ and such that $\bar{\sigma}(\Delta_o) < 1\ \forall\omega \Leftrightarrow \|\Delta_o\|_\infty < 1$. The uncertainty weight $W_o(s)$ can be a matrix of compatible dimension, but is often chosen to be a scalar weight. The assumption is, as before, that the true plant is within the model set Π_o (1), such that if we ensure closed-loop stability for all plants in the set Π_o then also the true system will be closed-loop stable.

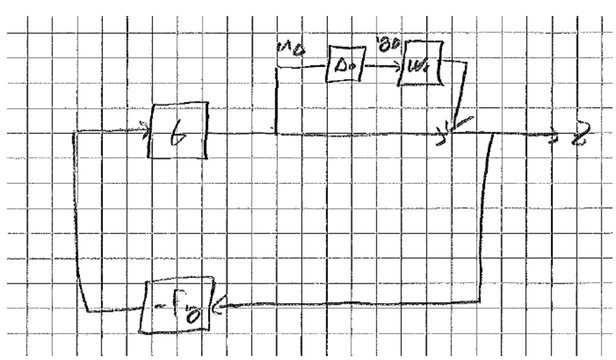


Figure 7.1: Closed-loop system with output uncertainty.

To analyze if the loop in Figure 7.1 is stable for all possible perturbations $\Delta_o(s)$, such that $\|\Delta_o\|_\infty < 1$, we employ the Small Gain Theorem. First rewrite the block-diagram in Figure 7.1 on M-Δ-form as shown in Figure 7.2. The block M is identified by considering the transfer-function from y_Δ to u_Δ in Figure 7.1.

$$u_\Delta = -GF_y(I + GF_y)^{-1}W_o y_\Delta = -TW_o y_\Delta$$

Thus, $M = -TW_o$ in Figure 7.2, where T is the complementary sensitivity function and W_o is the uncertainty weight. Applying the small gain theorem to the M-Δ loop then gives that the closed-loop is robustly stable if $T(s)$ is stable (nominal stability), $\Delta_o(s)$ stable (by assumption) and

$$\|TW_o\|_\infty < 1$$

Note that we must also choose the weight W_o to be stable, but this is no limitation since it is only the magnitude of the weight that matters in the uncertainty model (1). Thus, we arrive at a similar robust stability condition for MIMO systems as derived for SISO systems before.

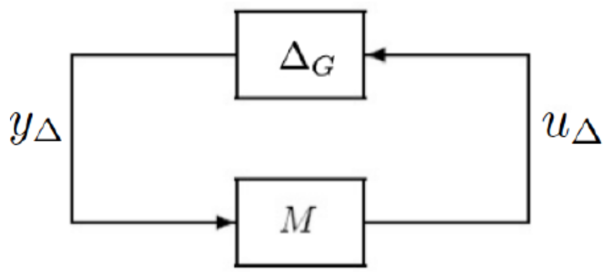


Figure 7.2: M-Δ-block for analyzing robust stability of system in Figure 7.1

For SISO systems it does not make any difference if we model the uncertainty on the input or output side of $G(s)$ since we can just move the uncertainty through $G(s)$, i.e.,

$$(1 + W(s)\Delta(s))G(s) = G(s)(1 + W(s)\Delta(s))$$

However, this is not true for MIMO systems and it makes a difference if we model the uncertainty on the input or output side. Let us consider modeling the uncertainty on the input side

$$\Pi_i = \{G_p(s) = G(s)(I + W_i(s)\Delta_i(s)), \quad \|\Delta_i\|_\infty < 1\} \tag{2}$$

Repeating the analysis done for output uncertainty above, we now obtain for M in the M-Δ_i-loop in Figure 7.2

$$u_\Delta = -F_yG(I + F_yG)^{-1}W_i y_\Delta = -T_u W_i y_\Delta$$

and hence the small gain theorem yields the robust stability criterion

$$\|T_u W_i\|_\infty < 1$$

where T_u is the complementary sensitivity function at the input side. Recall that in general $T_u \neq T$, and hence we get a different robust stability condition when we model the uncertainty on the input side (2) as compared to when we model it on the output side (1).

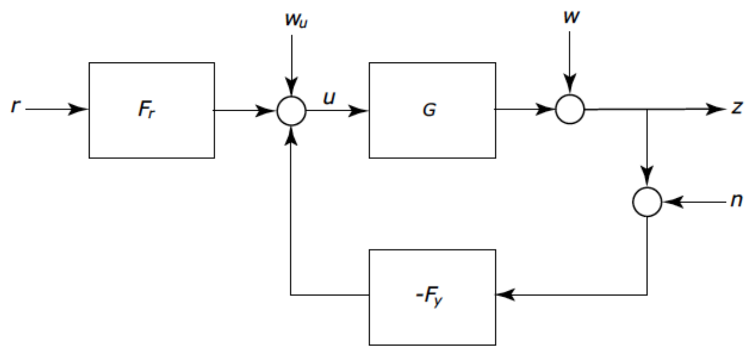


Figure 7.3: Two-degree of freedom control system.

7.2 Controller Design

Consider the closed-loop system in Figure 7.3. Recall that the aim of the control system is to make the output z follow the reference r despite disturbances on the input w_u, disturbances on the output w and measurement noise n. In addition, it is usually preferable to limit the use of the input u. Finally, we want to make sure the closed-loop is stable even if our model is uncertain, i.e., we want to make the closed-loop robustly stable. These performance and robustness specifications can all be formulated as requirements on the relevant closed-loop transfer-functions. For instance, to attenuate disturbances w in the output z we should make S small and to make the system robustly stable wrt output uncertainty and to attenuate measurement noise we should make T small. As discussed previously, we can not make a given transfer-function small att all frequencies, e.g., recall the Bode sensitivity integral, and we must also make a trade off between different transfer-functions, e.g., $S+T=I$. Finally, we should make sure that all specifications are feasible, that is, satisfy all limitations as discussed in Lecture 6. The aim of the control design is then to shape the relevant closed-loop transfer-functions, e.g., S and T, to fulfill the specifications using the controllers F_y and F_r. The shaping is done in two steps:

1. Define boundaries that a given closed-loop transfer-function should stay within, e.g.,

$$\bar{\sigma}(S) < |W_S^{-1}| \;\forall\omega \;\Rightarrow\; \|W_S S\|_\infty < 1 \tag{3}$$

$$\bar{\sigma}(T) < |W_T^{-1}| \;\forall\omega \;\Rightarrow\; \|W_T T\|_\infty < 1 \tag{4}$$

   The boundaries, or weights, should reflect the control objectives while respecting all fundamental trade-offs and limitations such that a controller satisfying all bounds exists.

2. Determine controller $F_y(s)$ (and $F_r(s)$) such that all specified bounds are satisfied.

The second step above can be approached either using classical loop shaping, i.e., shaping the open-loop $L = GF_y$, or using an optimization based approach, i.e., $\mathcal{H}_\infty$-synthesis, to directly shape the closed-loop transfer-functions. We discuss the two different approaches below.

7.2.1 Classic Loop Shaping

In classical loop shaping, the idea is to translate specifications on the closed-loop transfer-functions into specifications on the open-loop transfer-function (loop gain) $L = GF_y$ and then use F_y to shape L. The main motivation behind this approach is that the relation between the controller $F_y(s)$ and the loop gain $L = GF_y$ is much more transparant than the relationship between the controller and closed-loop transfer-functions like the sensitivity S and the complementary sensitivty T.

For MIMO plants, the specifications on the closed-loop (3) and (4) are in terms of bounds on the maximum singular values of S and T. Thus, we must translate the specifications on these singular values into specifications on the singular values of the loop gain $L = GF_y$. Starting with the sensitivity function S we have

$$S = (I + L)^{-1}$$

For a square matrix A we have $\bar{\sigma}(A^{-1}) = 1/\underline{\sigma}(A)$ and we get

$$\bar{\sigma}(S) = \frac{1}{\underline{\sigma}(I + L)}$$

Now, from Fan's Theorem (see Lec 6)

$$|\sigma_i(L) - 1| \leq \sigma_i(I + L) \leq \sigma_i(L) + 1 \tag{5}$$

If $\underline{\sigma}(L) >> 1$ then $\underline{\sigma}(I + L) \approx \underline{\sigma}(L)$ and

$$\bar{\sigma}(S) \approx \frac{1}{\underline{\sigma}(L)}, \quad \underline{\sigma}(L) >> 1$$

Thus, in the frequency range where we want to make the sensitivity function small, corresponding to $|W_S|$ large, we get

$$\bar{\sigma}(S) \leq |W_S^{-1}| \quad \Rightarrow \quad \underline{\sigma}(L) \geq |W_S| \ , \quad |W_S| >> 1 \tag{6}$$

Similarly, for T we use the fact that $\bar{\sigma}(A)\underline{\sigma}(B) \leq \bar{\sigma}(AB) \leq \bar{\sigma}(A)\bar{\sigma}(B)$ with $A = L$ and $B = (I + L)^{-1}$ together with (5) to derive

$$\bar{\sigma}(T) \approx \bar{\sigma}(L), \quad \bar{\sigma}(L) << 1$$

Hence, for frequencies where $|W_T| >> 1$ we get

$$\bar{\sigma}(T) \leq |W_T^{-1}| \quad \Rightarrow \quad \bar{\sigma}(L) \leq |W_T^{-1}| \ , \quad |W_T| >> 1 \tag{7}$$

Typically, we want to make the sensitivity small at low frequencies and the complementary sensitivity small at high frequencies. Thus, at low frequencies the loop gain should satisfy the bound (6) while at high frequencies it should satisfy the bound (7). This is illustrated in Figure 7.4.

Note that it is difficult to adress stability in MIMO loop shaping; there is no phase lag defined for MIMO systems, and hence no gain and phase margins. Thus, there is a need to adress stability after having shaped the loop to satisfy the given bounds. We will return to this problem in Lecture 10 when discussing robust loopshaping.

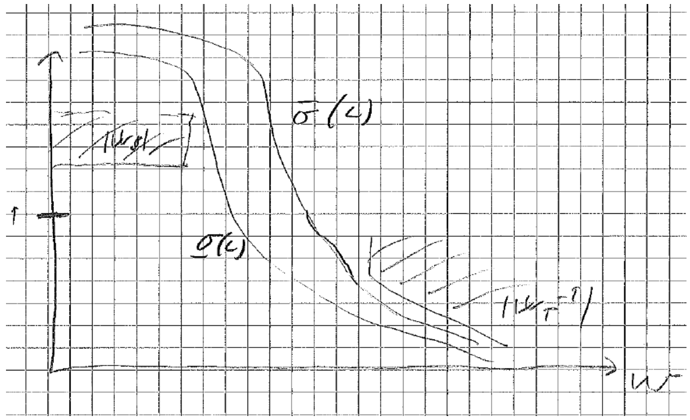


Figure 7.4: Illustration of shaping singular values of loop gain $L = GF_y$ to shape bounds $|W_S|$ and $|W_T^{-1}|$.

7.2.2 Shaping the closed-loop - $\mathcal{H}_\infty$-Synthesis

An attractive alternative to "manually" shaping the open-loop, is to employ an optimization algorithm to directly shape the closed-loop according to the specifications. With such an approach we can also solve more general problems, i.e., with specifications on closed-loop transfer-functions other than S and T. As an example, assume we want to attenuate disturbances on the output w and measurement noise n in z while keeping the input usage u small. The corresponding transfer-functions are

$$z = Sw + Tn \ ; \quad u = -F_y(I + GF_y)(w + n) = G_{wu}(w + n)$$

As before, we introduce weights to make a suitable trade-off between different frequencies and different transfer-functions, and to respect any fundamental limitations, and thereby get the performance objectives

$$\|W_S S\|_\infty \leq 1 \ , \quad \|W_T T\|_\infty \leq 1 \ , \quad \|W_u G_{wu}\|_\infty \leq 1 \tag{8}$$

Note that we can also include robust stability for output uncertainty, as discussed above, by including the uncertainty weight in W_T. In this course we will, unless stated otherwise, assume that all weights are scalar transfer-functions.

For optimization purposes we need a single objective function and this can be obtained by stacking all objectives in (8) into one big matrix and minimizing the norm of the stacked matrix

$$F_y = \arg\min_{F_y} \left\| \begin{pmatrix} W_S S \\ W_T T \\ W_u G_{wu} \end{pmatrix} \right\|_\infty \tag{9}$$

Note that if the $\mathcal{H}_\infty$-norm of the stacked matrix is less than some positive number γ, then so is the $\mathcal{H}_\infty$ norm of each of the elements $W_S S, W_T T$ and $W_u G_{wu}$ less than γ (with some margin).

Ideally, we could give the problem in (9) to an optimization algorithm to determine the feedback controller $F_y(s)$ that minimizes the objective function. Unfortunately, this is usually not possible for several reasons. Rather, we need to solve an associated signal minimization problem that reflects the objective function in (9). The signal minimization problem we can solve is, with reference to the system shown in Figure 7.5,

$$\min_{F_y} \sup_{w} \frac{\|z_e\|_2}{\|w\|_2} \tag{10}$$

where $z_e = G_{ec}w$. Note that

$$\sup_{w} \frac{\|z_e\|_2}{\|w\|_2} = \|G_{ec}\|_\infty \tag{11}$$

and hence solving (10) corresponds to minimizing the $\mathcal{H}_\infty$-norm of the closed-loop G_{ec} in Figure 7.5. The system in Figure 7.5 is called *the extended system* for reasons that will come obvious below. An important aspect of the extended system is that we can separate the controller $F_y(s)$, to be determined through optimization, from the open-loop system $G_0(s)$.

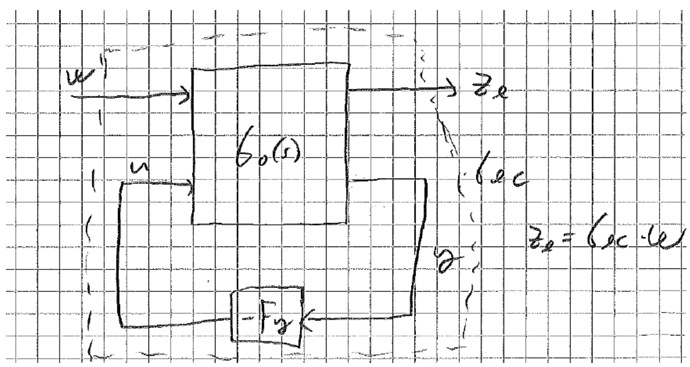


Figure 7.5: Extended system for $\mathcal{H}_\infty$-optimization.

The fact that the signal minimization problem in (10) corresponds to

$$\min_{F_y} \|G_{ec}\|_\infty \tag{12}$$

implies that we can solve the original problem (9) if we choose the signals z_e and w of the extended system such that

$$G_{ec} = \begin{pmatrix} W_S S \\ W_T T \\ W_u G_{wu} \end{pmatrix} \tag{13}$$

That is, we need to choose the output z_e of the extended system G_{ec} such that

$$z_e = \begin{pmatrix} W_S S \\ W_T T \\ W_u G_{wu} \end{pmatrix} w \tag{14}$$

where w is the disturbance on the output in the original block-diagram in Figure 7.3 and also the input to the extended system in Figure 7.5[1]. For the optimization, we need to provide the corresponding open-loop system $G_0(s)$ in Fig. 7.5.

[1]Depending on what problem we want to solve, we may choose different inputs to the exended system. For the problem considered here, use of w as the input is sufficient.

Let us derive the outputs $z_{ei}, i = 1, 3$, and corresponding open-loop transfer-functions G_{0i}, needed to reflect each of the three transfer-functions in (14).

- The first output we seek should reflect the first transfer-function in (14), i.e,

$$z_{e1} = W_S S w$$

To relate this to the orignal block-diagram in Figure 7.3, we have that $z = Sw$. Hence we choose the output

$$z_{e1} = W_S z$$

In open-loop we have $z = Gu + w$, and hence for G_0 we get

$$z_{e1} = W_S(Gu + w)$$

- The second output z_{e2} should reflect $W_T T$, i.e.,

$$z_{e2} = W_T T w$$

In the original system we have $z = Sw = (I - T)w$ and hence $z - w = -Tw$ which leads to

$$z_{e2} = -W_T(z - w)$$

In open-loop we have $z - w = Gu$ and hence for G_0 we get

$$z_{e2} = -W_T G u$$

- The final output z_{e3} should reflect $W_u G_{wu}$, i.e.,

$$z_{e3} = W_u G_{wu} u$$

In the original system we have $u = G_{wu} w$ and hence

$$z_{e3} = W_u u$$

In open-loop we have $u = u$ and hence for G_0 we get

$$z_{e3} = W_u u$$

In summary, we choose the output

$$z_e = \begin{pmatrix} W_S z \\ -W_T(z - w) \\ W_u u \end{pmatrix}$$

which corresponds to the open-loop transfer-matrix

$$\begin{pmatrix} z_e \\ y \end{pmatrix} = G_0(s) \begin{pmatrix} w \\ u \end{pmatrix} \quad ; \quad G_0(s) = \begin{pmatrix} W_S & W_S G \\ 0 & -W_T G \\ 0 & W_u \\ I & G \end{pmatrix} \tag{15}$$

and the corresponding closed-loop is the desired (13). Note that the open-loop extended system contains several instances of the model $G(s)$ as well as the weights W_S, W_T and W_u.

The solution of the signal minimization problem (10) is based on the state-space realization of the open-loop model $G_0(s)$ (15)

$$\begin{aligned} \dot{x} &= Ax(t) + Bu(t) + Nw(t) \\ z_e(t) &= Mx(t) + Du(t) \\ y(t) &= Cx(t) + w(t) \end{aligned} \tag{16}$$

where the realization is normalized so that $D^T M = 0$ and $D^T D = I$ (for reasons that become obvious when proving the main result below). Solving the optimal control problem in (10) is hard, and it is therefore usually preferred to solve the sub-optimal control problem. The sub-optimal problem is, given a positive number γ, determine a stabilizing controller $F_y(s)$ which provides

$$\sup_w \frac{\|z_e\|_2}{\|w\|_2} = \|G_{ec}\|_\infty = \gamma$$

If no such controller exist, then the minimum achievable value of $\|G_{ec}\|_\infty$ is larger than γ. To determine if a controller $F_y(s)$ exists that gives $\|G_{ec}\|_\infty = \gamma$ one can solve the algebraic Riccati equation

$$A^T P + PA + M^T M + P(\gamma^{-2} N N^T - BB^T)P = 0 \tag{17}$$

where A, B, M, N are all from the state-space description of $G_0(s)$ (16). If a positive definite solution $P > 0$ to (17) exists and if $A - BB^T P$ is stable, then the controller

$$\begin{aligned} \dot{\hat{x}} &= A\hat{x}(t) + Bu(t) + N(y(t) - C\hat{x}(t)) \; ; \quad \hat{x}(0) = 0 \\ u(t) &= -L_\infty \hat{x}(t), \quad L_\infty = B^T P \end{aligned} \tag{18}$$

will give $\|G_{ec}\|_\infty = \gamma$. The Laplace transform of (18) is $u = -F_y(s)y$. Note that the feedback controller (18) is on the form of an observer combined with state feedback. As we shall see in the next lecture, this control structure also results when solving the classical optimal control problems based on objective functions defined in the time domain. Also note that the controller will have the same number as states as the open-loop model $G_0(s)$, which implies that the number of states in the controller will be at least the number of states in $G(s)$ plus the number of states in the weights W_S, W_T and W_u.

To arrive at the $\mathcal{H}_\infty$-optimal controller, i.e., the controller that solves (10) and (12), we can solve the sub-optimal problem iteratively, i.e., by iterating on γ untill $\gamma = \gamma_{min}$. The basis for this is that for every $\gamma > \gamma_{min}$, there is a solution $P > 0$ to the Riccati equation (17) which stabilizes the loop with $\|G_{ec}\|_\infty = \gamma$, while there exists no stabilizing solution $P > 0$ if $\gamma < \gamma_{min}$.

To prove that the controller (18) solves the sub-optimal control problem $\|G_{ec}\|_\infty = \gamma$, consider the function

$$V(t) = x^T(t)Px(t) + \int_0^t (z_e^T(\tau) z_e(\tau) - \gamma^2 w^T(\tau) w(\tau) d\tau$$

Now, if P is a positive definite matrix, i.e., $x^T P x > 0$, and we can show that $V(t) \leq 0 \; \forall t$, then the integral term must be negative for all t and hence it follows that $\|z_e\|_2^2 < \gamma^2 \|w\|_2^2$ for any disturbance w. Note that $V(0) = 0$ and hence it is sufficient to show that $\dot{V}(t) \leq 0 \; \forall t$. Taking the time derivative of $V(t)$ we get

$$\begin{aligned}
\dot{V} &= \dot{x}^T P x + x^T P \dot{x} + z_e^T z_e - \gamma^2 w^T w \\
&= x^T A^T P x + u^t B^T P x + w^T N^T P X + x^T P A X + x^T P B u + x^T P N w + x^T M M x + u^T u - \gamma^2 w^T w \\
&= x^T (A^T P + P A + M^T M) x + u^T B^T P + x^T P B u + u^T u + w^T N^T P x + x^T P N w - \gamma^2 w^T w \qquad (19)
\end{aligned}$$

We complete the square for $u^T B^T P + x^T P B u + u^T u$ by adding $x^T P B B^T P x$ and complete the square for $w^T N^T P x + x^T P N w - \gamma^2 w^T w$ by adding $-\gamma^{-2} x^T P N N^T P x$ (and substracting the same terms from the overall expression for $\dot{V}$) to obtain

$$\begin{aligned}
\dot{V} &= x^T (A^T P + P A + M^T M - P(B B^T - \gamma^{-2} N N^T) P) x + (u + B^T P x)^T (u + B^T P x) \\
&\quad -\gamma^{-2} (w - \gamma^{-2} N^T P x)^T (w - \gamma^{-2} N^T P x) \qquad (20)
\end{aligned}$$

Now it is easily seen that if P is a positive definite solution to (17) and we choose $u = -B^T P x$, then $\dot{V} \leq 0 \; \forall t$ and the proof is complete.

8 EL2520 Lecture notes 8: LQG - Linear Quadratic Optimal Control

In lecture 7, we showed that we could solve the problem of minimizing the $\mathcal{H}_\infty$-norm of weighted transfer-functions, e.g., $\min_{F_y} \|W_S S\|_\infty$, by solving a corresponding signal minimization problem

$$\min_{Fy} \sup_{w} \frac{\|z_e\|_2}{\|w\|_2}$$

in state space, where z_e and w are the output and input, respectively, of the system we want to minimize the $\mathcal{H}_\infty$-norm of. The resulting controller was on the form of state feedback combined with an observer. The classical optimal control theory from the 1960's is also based on solving a signal minimization problem in state space, but then with an objective function formulated in the time-domain and using a stochastic framework for the input signals. Below we present the main results related to this theory, the LQG-controller which also can be written as state feedback combined with an observer (the Kalman filter).

8.1 The LQG control problem

We consider a system on state space form

$$\begin{aligned} \dot{x} &= Ax(t) + Bu(t) + Nv_1(t) \\ y(t) &= Cx(t) + v_2(t) \\ z(t) &= Mx(t) \end{aligned} \tag{1}$$

The aim of the control is, similar to before, to keep $z(t)$ small in the presence of the disturbance v_1 and measurement noise v_2. Recall that in H_∞-optimal control we considered the worst-case disturbances. In the LQG framework, we rather consider stochastic disturbances and noise. Assume that v_1 and v_2 are Gaussian (normally distributed) white noise with covariance matrices[1]

$$E\{v_1 v_1^T\} = R_1 \ , \quad E\{v_2 v_2^T\} = R_2$$

The optimal control problem we aim to solve, given these disturbances[2], is

$$\min_u E\{\lim_{T\to\infty} \frac{1}{T} \int_0^T z^T Q_1 z + u^T Q_2 u dt\} \tag{2}$$

The name LQG comes from <u>L</u>inear system, <u>Q</u>uadratic cost and <u>G</u>aussian disturbances.

[1] For a discussion on stochastic descriptions and modeling of disturbances, see course book and lecture slides for Lecture 8.

[2] A more general problem formulation allows for cross correlation between v_1 and v_2, but that is not considered here.

An important property of the solution to (2) is the *Separation Principle* which states that, provided (A, B) in (1) is stabilizable[3], (A, C) is detectable[4], and (A, R_1) stabilizable and $(A, M^T Q_1 M)$ detectable[5], the optimal control problem can be split into two subproblems

1. Optimal state feedback (LQ)
2. Optimal observer (Kalman filter)

and that these two sub-problems can be solved independently. The combined solutions is then the optimal solution to the overall LQG problem.

8.2 The LQ Problem

The LQ-problem is a deterministic control problem with no stochastic disturbances and assuming all states of the system are known (measured without noise)

$$\begin{aligned} \dot{x} &= Ax(t) + Bu(t) \ ; \quad x(0) = x_0 \\ z(t) &= Mx(t) \end{aligned} \tag{3}$$

The control problem we aim to solve is

$$\min_u \lim_{T \to \infty} \frac{1}{T} \int_0^T z^T Q_1 z + u^T Q_2 z dt$$

Thus, minimizing the square objective function for any initial condition x_0. The solution, which we state without proof here, is

$$u(t) = -Lx(t) \ , \quad L = Q_2^{-1} B^T P$$

and where $P \geq 0$ solves the algebraic Riccati equation

$$A^T P + PA + M^T Q_1 M - PBQ_2^{-1} B^T P = 0$$

Note that the design variables are the weights Q_1 and Q_2 that can be used to trade-off between outputs and inputs.

Example: Consider a simple first-order system

$$\dot{x} = ax(t) + u(t) \ ; \quad z(t) = x(t)$$

and performance objective

$$J = \int_0^\infty x^2 + \rho u^2 dt$$

The Riccati equation is

$$2ap + 1 - \frac{1}{\rho} p^2 = 0 \quad \Rightarrow \quad p = a\rho \pm \sqrt{(\rho a)^2 + \rho}$$

[3]All unstable states controllable

[4]All unstable states observable

[5]The last two conditions ensure a unique positive definite solution to the Riccati equations below

Thus, the optimal feedback is

$$u(t) = -\frac{p}{\rho}x(t) = -(a + \sqrt{a^2 + \frac{1}{\rho}})x(t)$$

and the closed-loop becomes

$$\dot{x} = -\sqrt{a^2 + \frac{1}{\rho}}x(t)$$

Note that if the weight $\rho \to \infty$, i.e., all weight is put on the control input u, then $u(t) = 0 \cdot x(t)$ if the system is open-loop stable ($a < 0$) while $u(t) = -2ax(t)$ if the system is open-loop unstable ($a > 0$), in which case the pole of the system is mirrored into the LHP by the feedback.

8.3 The Optimal Observer

Given that we in general do not measure all states, and furthemore, that the measurements are contaminated by measurement noise, we need to estimate the states with an observer. The system considered is

$$\begin{aligned} \dot{x} &= Ax(t) + Bu(t) + Nv_1(t) \ , \quad E\{v_1 v_1^T\} = R_1 \\ y(t) &= Cx(t) + v_2(t) \ , \quad E\{v_2 v_2^T\} = R_2 \end{aligned} \tag{4}$$

To estimate the state $x(t)$ we construct an observer

$$\dot{\hat{x}} = A\hat{x}(t) + Bu(t) + K_f(y(t) - C\hat{x}(t)) \tag{5}$$

where $\hat{x}$ denotes the estimated state. The design parameter is the observer gain K_f. The optimal observer is the observer which minimizes the variance of the estimation error

$$\min_{K_f} E\{(x - \hat{x})^T(x - \hat{x})\}$$

The optimal observer gain is then given by

$$K_f = PC^T R_2^{-1}$$

where $P \geq 0$ solves the Riccati equation

$$PA^T + AP - PC^T R_2^{-1} CP + NR_1N^T = 0$$

The optimal observer is called *the Kalman filter.*

Example: Consider the system

$$\dot{x} = ax(t) + u(t) + v_1(t) \ , \quad y(t) = x(t) + v_2(t)$$

where the disturbance and measurement noise variances are given by

$$E\{v_1^2\} = R_1 \ ; \quad E\{v_2^2\} = R_2$$

The Riccati equation and corresponding gain for the optimal observer is

$$2ap - p^2/R_2 + R_1 = 0 \quad \Rightarrow \quad k_f = a + \sqrt{a^2 + R_1/R_2}$$

Thus, if $a < 0$ and $R_2 >> R_1$ then $k_f = 0$, i.e., disregard measurement if there is large variance in the noise v_2 relative to the variance of the process disturbance v_1. The equation for the estimation error $\tilde{x}(t) = x(t) - \hat{x}(t)$ is

$$\dot{\tilde{x}} = -\sqrt{a^2 + R_1/R_2}\tilde{x}(t)$$

Thus, we get faster convergence with increasing R_1/R_2.

8.4 The Optimal LQG Controller

As stated above, the solution to the LQG problem fulfills the separation principle which means that the overall solution is a combination of the two separate problems of optimal control without any disturbance or noise (LQ) and the optimal observer without any control (Kalman). Thus, the LQG controller is given by

$$\begin{aligned} u(t) &= -L\hat{x}(t) \\ \dot{\hat{x}} &= A\hat{x}(t) + Bu(t) + K_f(y - C\hat{x}(t)) \end{aligned} \tag{6}$$

where the state controller feedback gain L is the LQ-optimal gain and the observer gain K_f is the Kalman gain. The structure of the LQG controller is illustrated in Figure 8.1. The tuning parameters of the LQG controller are the control weights Q_1, Q_2 and the

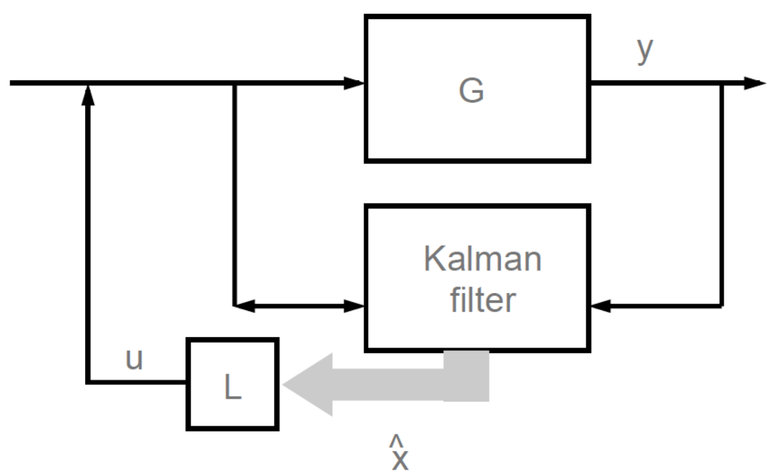


Figure 8.1: Structure of the LQG controller.

disturbance covariances R_1, R_2.

Advantages with LQG is partly that it is directly applicable to MIMO systems (which was one of the main reasons for its development in the 1950s-1960s), partly that it is relatively easy to make a trade-off between keeping the control error small on one hand and keeping input usage small on the other hand by choosing appropriate relative values of the weights Q_1 and Q_2. However, it may in general be difficult to see a direct connection between the choice of the tuning parameters Q_1, Q_2, R_1, R_2 and the real control objectives. In practice, this implies that one usually has to iterate untill finding an acceptable solution. A systematic method known as loop transfer recovery exists, but is not covered in this course. We will rather, in Lecture 9, show that LQG can be used as a machinery to shape closed-loop transfer-functions using H_2-optimal control, in which case Q_1, Q_2, R_1, R_2 are given by the objective function formulated in the input-output space.

One main disadvantage with LQG is that the only uncertainty that can be modelled is signal uncertainty, in the form of v_1 and v_2, and signals do not affect system properties like stability. Hence, it is not possible to explicitly adress robust stability with LQG. In fact, it has been shown that the LQG controller typically will have poor robustness margins and this was one of the main motivations for the introduction of robust control, or $\mathcal{H}_\infty$-optimal control, in the 1980s and 90s. We will compare the different control design methods in Lecture 11.

9 EL2520 Lecture notes 9: $\mathcal{H}_2$-Optimal Control

In lecture 7 we showed how the problem of minimizing the $\mathcal{H}_\infty$-norm of weighted closed-loop transfer-functions could be solved by translating the system norm minimization problem, in input-output space, into an equivalent signal norm minimization problem in state space. In LQG, introduced in lecture 8, the starting point was a signal minimization problem in state space that could be solved explicitly using the separation principle. In this lecture we will show that the signal minimization problem treated in LQG under certain conditions is equivalent to the problem of minimizing the $\mathcal{H}_2$-norm of weighted closed-loop transfer-functions. Thus, LQG can in principle be used as a machinery to shape closed-loop transfer-functions, but then based on the $\mathcal{H}_2$-norm. At the end of the lecture we will discuss some important differences between $\mathcal{H}_2$- and $\mathcal{H}_\infty$-optimal control.

We start by formulating the $\mathcal{H}_2$-optimal control problem and then show that it can be formulated as a signal minimization problem equivalent to the one treated in LQG.

The $\mathcal{H}_2$-optimal control problem is[1]

$$F_y = \arg\min_{F_y} \|P\|_2 \tag{1}$$

where P is the transfer-function we want to minimize and $\|\cdot\|_2$ denotes the $\mathcal{H}_2$-norm of P. Similar to what we considered for $\mathcal{H}_\infty$-optimal control in lecture 7, the matrix P is typically a stacked matrix of weighted transfer-functions, e.g.,

$$P = \begin{pmatrix} W_S S \\ W_T T \\ W_u G_{wu} \end{pmatrix} \tag{2}$$

Recall that the $\mathcal{H}_2$-norm of a scalar (SISO) transfer-function $G(s)$ is defined as

$$\|G\|_2^2 = \frac{1}{2\pi}\int_{-\infty}^{\infty} |G(i\omega)|^2 d\omega$$

Note that the $\mathcal{H}_2$-norm is defined for strictly proper[2] and stable $G(s)$ only. Thus, if we for instance consider minimizing the weighted sensitivity $P = W_S S$, then we must choose the weight W_S strictly proper since S is always semi-proper only. Compared to the $\mathcal{H}_\infty$-norm, which is the peak value of $|G(i\omega)|$, we see that the $\mathcal{H}_2$-norm essentially is a measure of the area under $|G(i\omega)|$ when plotted against frequency. For a multivariable (MIMO) system, the $\mathcal{H}_2$-norm is defined as

$$\|G\|_2^2 = \sum_{ij} \|G_{ij}\|_2^2 = \frac{1}{2\pi}\sum_{ij}\int_{-\infty}^{\infty} |G_{ij}(i\omega)|^2 d\omega = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{tr}(G(i\omega)^H G(i\omega)) d\omega$$

[1] We also include the prefilter F_r in the design when P depends on it

[2] Recall that a proper transfer-function has at least as many poles as zeros, while strictly proper implies more poles than zeros. A system with as many poles as zeros is called semi-proper.

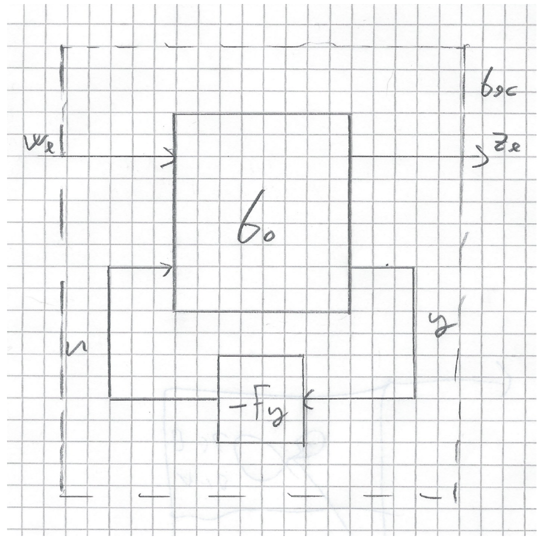


Figure 9.1: Extended system with closed-loop transfer-function $z_e = G_{ec}(s)w_e$ and corresponding open-loop $(z_e, y) = G_0(s)(w_e, u)$.

To solve (1), we formulate an extended system which has P as the closed-loop transfer-matrix from the input w_e to the output z_e. See Figure 9.1. In the same fashion as when solving the $\mathcal{H}_\infty$-optimal control problem in lecture 7, we pick the input w_e and output z_e from the signals of the original feedback system such that the closed-loop transfer-function G_{ec} from w_e to z_e equals P. As an example, assume that we want to minimize the norm of the weighted sensitivity function, i.e., $P = W_S S$. Since $z = Sw$ in the original feedback problem, we pick $w_e = w$ and $z_e = W_S z$ in the extended system, such that $z_e = W_S S w_e$.

The solution of the $\mathcal{H}_2$-optimal control problem is based on a state-space realization of the corresponding open-loop of the extended system, i.e., $G_0(s)$

$$\begin{aligned} \dot{x} &= Ax(t) + Bu(t) + Nw_e(t) \\ z_e(t) &= Mx(t) + Du(t) \\ y(t) &= Cx(t) + w_e(t) \end{aligned} \tag{3}$$

where the realization is chosen such that $D^T M = 0$ and $D^T D = I$ (for reasons that will become obvious below). Note that the formulation of the extended system to represent P, and the state-space realization of the corresponding open-loop system $G_0(s)$, is identical to the one used in $\mathcal{H}_\infty$-optimal control in lecture 7. However, while we in the latter case considered worst-case disturbances w_e, to reflect $\|P\|_\infty$, we now consider the case with w_e being white noise with covariance

$$E\{w_e w_e^T\} = I \quad \Rightarrow \quad \Phi_{w_e} = I$$

Note that this corresponds to each scalar signal w_{ei} in the vector w_e having a flat frequency spectrum $\Phi_{w_{ei}} = 1$, i.e., the noise has the same energy (amplitude) at all frequencies. Now, using Parseval's theorem we get for the output

$$\|z_e(t)\|_2^2 = \frac{1}{2\pi}\int_{-\infty}^{\infty} |Z_e(i\omega)|^2 d\omega = \frac{1}{2\pi}\int_{-\infty}^{\infty} |G_{ec}(i\omega)W_e(i\omega)|^2 d\omega \tag{4}$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{tr}(G_{ec}^H(i\omega)G_{ec}(i\omega))d\omega = \|G_{ec}\|_2^2 = \|P\|_2^2 \tag{5}$$

Thus, minimizing the 2-norm of the output $\|z_e\|_2$ when the input w_e is white noise with spectrum $\Phi_{w_e} = I$ corresponds to minimizing the 2-norm of the closed-loop transfer-function from w_e to z_e, i.e., $\|G_{ec}\|_2$ ($\|P\|_2$).

With $D^T M = 0$ and $D^T D = I$ we get

$$\|z_e(t)\|_2^2 = ||Mx(t) + Du(t)\|_2^2 = \|Mx(t)\|_2^2 + \|u(t)\|_2^2 = \int_0^\infty x^T M^T M x + u^T u dt$$

Note that this is equivalent to the objective function in the LQG-problem, discussed in lecture 8, with weights $Q_1 = Q_2 = I$. Also note that the model (3) corresponds to the model used in LQG with disturbance and noise covariances $R_1 = R_2 = I$. Thus, we conclude that solution of (1) corresponds to the solution of an LQG problem based on the extended system (3) with weights $Q_1 = Q_2 = I$ and covariances $R_1 = R_2 = I$. Note that the weighting of the problem is included in P and hence in the model (3).

As shown in lecture 8, the solution to the LQG problem is an LQ-controller combined with a Kalman filter. Since (3) is on innovation form, i.e., we essentially measure the process disturbance w_e, the Kalman filter is

$$\dot{\hat{x}} = A\hat{x}(t) + Bu(t) + N(y(t) - C\hat{x}(t))$$

and the feedback law is

$$u = -L\hat{x}(t)$$

where $L = B^T S$ and $S \geq 0$ solves the Riccati equation

$$A^T S + SA + M^T M - SBB^T S = 0$$

And the controller $F_y(s)$ is obtained by Laplace transformation to yield

$$F_y(s) = L(sI - A + BB^T S + NC)^{-1} N$$

Note that we require the observer above to be stable, i.e., $A - NC$ stable, otherwise the Kalman filter must be computed in a more involved manner not treated here. However, the most common situation is that $A - NC$ has one or more zero eigenvalues due to pure integrators in the weights used in P and this can be avoided by using weights with poles close to zero but slightly inside the LHP. For instance, rather than using a weight $W_S = \frac{s+1}{s(0.001s+1)}$ for the sensitivity, we can use a weight $W_S = \frac{s+1}{(s+\epsilon)(0.001s+1)}$, $\epsilon > 0$ with ϵ being a small positive constant.

9.1 $\mathcal{H}_2$- vs $\mathcal{H}_\infty$-optimal control

Recall that the $\mathcal{H}_\infty$-norm of a transfer-function G_{ec} is equal to the peak value of the maximum singular value over all frequencies, i.e.,

$$\|G_{ec}\|_\infty = \sup_\omega \bar{\sigma}(G_e c)$$

Thus, by minimizing the $\mathcal{H}_\infty$-norm we focus on the worst frequency and worst direction of the disturbance input w_e.

In terms of the singular values, the $\mathcal{H}_2$-norm is

$$\|G_{ec}\|_2^2 = \frac{1}{2\pi}\int_{-\infty}^{\infty}\sum_i \sigma_i^2(G_{ec}(i\omega)d\omega$$

Thus, when minimizing the $\mathcal{H}_2$-norm of G_{ec} we consider all frequencies and all directions of the disturbance input w_e, and we determine the optimal trade-off between them.

From a performance point of view it may seem more reasonable to optimize with respect to some form of average over all frequencies and all directions, as is done in $\mathcal{H}_2$-optimization and LQG. Putting all effort into improving the worst case disturbance attenuation, as is done in $\mathcal{H}_\infty$-optimization, may seem somewhat irrational. However, since we include weights in the problem, i.e., the weights W_S on the sensitivity S and W_T on the complementary sensitivity T, it makes sense to try to satisfy the corresponding bounds $|W_S^{-1}|$ and $|W_T^{-1}|$ for all frequencies and directions. Furthermore, the main motivation for introducing the $\mathcal{H}_\infty$-norm in the first place was that we thereby could include robustness as a criterion in the control design. It is quite obvious that stability can not be dealt with on an average basis; if the closed-loop is not stable for the worst case model within the uncertainty set then it is not robustly stable. However, it is of course possible to design controllers using $\mathcal{H}_2$-optimal control or LQG, and then analyze the robustness of the resulting closed-loop using $\mathcal{H}_\infty$.

We will return to a comparison between $\mathcal{H}_\infty$-optimal control, $\mathcal{H}_2$-optimal control and LQG in lecture 11 where we will consider control design for an example process.

10 EL2520 Lecture notes 10: Robust Loop Shaping

In this lecture we will consider a special $\mathcal{H}_\infty$-optimal control problem, namely one in which the objective is to maximize robust stability for a certain, quite general, form of model uncertainty known as coprime uncertainty. The main reason why this problem is of interest is that it can be used to ”robustify” a controller that has been designed to meet certain performance specifications, but which has not addressed robust stability. For instance, in lecture 7 we considered controller design using loop shaping ideas, i.e., shaping the singular values of the loop gain $L = GF_y$, but noted that it was diffcult to address robust stability and stability margins when loopshaping for multivariable systems. In this case it is therefore relevant to first design for performance only and then use $\mathcal{H}_\infty$-optimization in a second step to make the system robustly stable. This approach is also known as Glover-MacFarlane loop shaping.

10.1 Robust Stabilization

We start by describing coprime uncertainty and then derive the corresponding robust stability condition using the Small Gain Theorem. To maximize robust stability, i.e., maximize the size of the uncertainty set that can be tolerated without loosing stability, we formulate an $\mathcal{H}_\infty$-optimal control problem which we then show can be solved explicitly. We finally illustrate the results with a simple example.

A left coprime factorization of a model $G(s)$ is

$$G(s) = M^{-1}(s)N(s) \tag{1}$$

where $M(s)$ and $N(s)$ are stable coprime transfer-functions. Two transfer-functions are coprime if they satisfy the Bezout identity

$$N(s)U(s) + M(s)V(s) = I$$

for some stable $U(s)$ and $V(s)$. Two scalar stable transfer-functions are coprime if they have no common RHP zeros. How to compute a (normalized) left coprime factorization of a transfer-matrix $G(s)$ is given in Appendix of these lecture notes.

Since $M(s)$ and $N(s)$ are stable, it implies that all RHP poles of $G(s)$ will appear as RHP zeros of $M(s)$ and all RHP zeros of $G(s)$ will appear as RHP zeros of $N(s)$. A coprime factorization is normalized if it satisfies

$$M(s)M^T(-s) + N(s)N^T(-s) = I$$

We next introduce uncertainty, in the form of stable perturbations $\Delta_M(s)$ and $\Delta_N(s)$ to $M(s)$ and $N(s)$, respectively, so that the set of plants is

$$G_p(s) = (M(s) + \Delta_M(s))^{-1}(N(s) + \Delta_N(s)) \; ; \quad \|\Delta_N \; \Delta_M\|_\infty < \epsilon \tag{2}$$

Note that this uncertainty description allows both zeros and poles to cross between the LHP and RHP and is as such quite general.

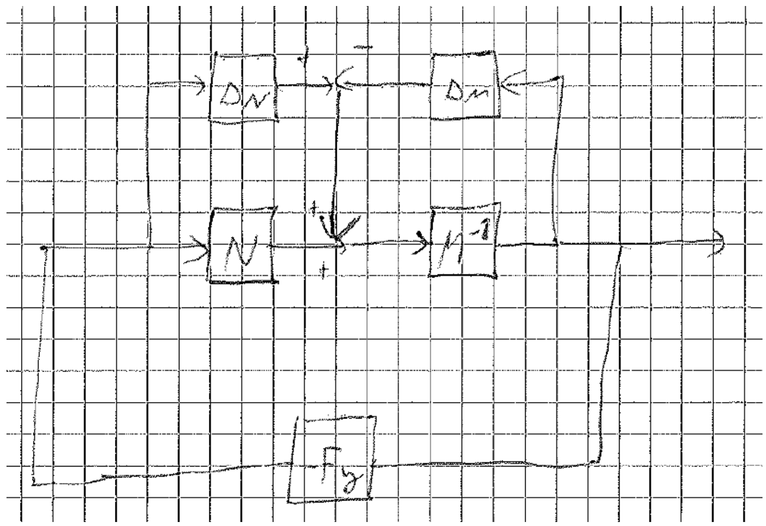


Figure 10.1: Closed-loop with left coprime factor uncertainty.

The aim of robust stabilization is to determine a controller that maximizes the robust stability region, given the uncertainty set (2), in the sense that it allows the largest possible ϵ while providing robust stability. For this purpose, we next derive a robust stability condition for a closed-loop system with the uncertain plant given by (2). The loop is illustrated in Figure 10.1. As usual, we rewrite the loop on the form of a P-Δ_G-loop, where P is a nominal transfer-function and Δ_G is the uncertainty, and apply the Small Gain Theorem. See also Lecture 7. Let $\Delta_G = [\Delta_N \; \Delta_M]$, then we identify P from the block diagram in Figure 1 as

$$P = -\begin{pmatrix} F_y \\ I \end{pmatrix} (I + GF_y)^{-1} M^{-1}$$

Applying the Small Gain Theorem, we can conclude that the closed-loop system is robustly stable if Δ_G is stable (Δ_N and Δ_M both stable), P is stable (nominal stability) and

$$\|\Delta_G\|_\infty \|P\|_\infty < 1$$

Since the aim of robust stabilization is to maximize the allowed $\|\Delta_G\|_\infty$, this can be formulated as minimizing $\|P\|_\infty$, i.e.,

$$\min_{F_y} \|P\|_\infty = \min_{F_y} \| \begin{pmatrix} F_y \\ I \end{pmatrix} (I + GF_y)^{-1} M^{-1} \|_\infty = \gamma_{min}$$

With the minimum achievable $\|P\|_\infty = \gamma_{min}$ we can allow $\|\Delta_G\|_\infty < 1/\gamma_{min}$.

The minimum γ_{min}, and the corresponding robustly stabilizing controller $F_y(s)$, can be computed directly from a state-space realization of $G(s)$

$$\begin{aligned} \dot{x} &= Ax(t) + Bu(t) \\ y(t) &= Cx(t) \end{aligned} \tag{3}$$

First, solve the two algebraic Riccati equations for postitive definite matrices $X > 0, Z > 0$

$$AZ + ZA^T - ZC^TCZ + BB^T = 0$$

$$A^TX + XA - XB^TBX + C^TC = 0$$

Now, let $\lambda_m = \max_i \lambda_i(XZ)$. Then

$$\gamma = \alpha(1 + \lambda_m)^{1/2}, \ \ \alpha \geq 1$$

where α is a tuning parameter. Note that $\gamma = \gamma_{min}$ for the choice $\alpha = 1$. The main reason for introducing $\alpha \geq 1$ is that in many cases the maximally robustifying controller can influence too strongly on the performance, and this can be avoided to some extent by choosing α somewhat larger than 1. A typical rule of thumb is $\alpha = 1.1$. Introduce

$$R = I - \frac{1}{\gamma^2}(I + ZX)$$

Then, the optimal state feedback gain and state observer gains are, respectively,

$$L = B^TX \ ; \quad K = R^{-1}ZC^T$$

and the controller providing $\|P\|_\infty = \gamma$ is

$$\dot{\hat{x}} = A\hat{x}(t) + Bu(t) + K(y(t) - C\hat{x}(t)) \ ; \quad u(t) = -L\hat{x}(t)$$

10.2 Robust Loopshaping

Robust stabilization is usually of little interest on its own; we also need to address performance. In Glover-MacFarlane loop shaping, one first shapes the open-loop to satisfy certain performance specifications

$$\hat{G}(s) = W_2(s)G(s)W_1(s)$$

where W_1 and W_2 are post- and pre-compensators, respectively. For a discussion on loop shaping for performance, see Lecture 7. In a second step, one solves the robust stabilization problem for the shaped plant $\hat{G}(s)$. This give a robustly stabilizing controller $\hat{F}_y(s)$, and the overall controller is then

$$F_y(s) = W_1(s)\hat{F}_y(s)W_2(s)$$

In general, provided γ_{min} for the robust stabilization is small ($\gamma_{min} \approx< 4$) the controller $\hat{F}_y(s)$ will usually have a small influence on the performance. If γ_{min} is large ($\approx> 4$) then this indicates that there is a conflict between the nominal performance objectives and robust stability.

Example: Consider the system

$$z = \underbrace{\frac{200}{10s+1}\frac{1}{(0.05s+1)^2}}_{G(s)}u + \underbrace{\frac{100}{10s+1}}_{G_d(s)}d$$

The model is scaled such that expected disturbances have magnitude $|d| < 1\ \forall\omega$ and acceptable performance corresponds to keeping $|z| < 1\ \forall\omega$. Thus, since $z = SG_d d$ acceptable performance corresponds to

$$|SG_d(i\omega)| < 1\ \forall\omega \quad \Leftrightarrow \quad |S(i\omega)| < |G_d^{-1}(i\omega)|\ \forall\omega$$

Since $S = 1/(1 + L)$ where L is the loop-gain, we have that $|S| \approx 1/|L|$ when $|S| << 1$. Thus, we should shape the loop such that

$$|L(i\omega)| > |G_d(i\omega)| \ \ \omega < \omega_d$$

where $|G_d(i\omega)| > 1$ for $\omega < \omega_d$. A controller that satsifies this is

$$W_1 = \frac{s+2}{s}$$

The corresponding loop-gain $L = GW_1$ is shown by the dashed line in Figure 2. Also shown is a simulation of the closed-loop response to a unit step in the disturbance d. As can be seen, the controller maintains $|z| < 1$ but the response is quite oscillatory, indicating poor robust stability. We therefore next robustify the shaped loop, according to the results above, using the Matlab robust control toolbox command `ncfsyn`. The maximal robustness corresponds to $\gamma_{min} = 2.34$ which should be OK. The robustified loop gain is shown by the solid line in Figure 2, and as can be seen the robustification mainly lowers the slope of the loop gain around the crossover frequency, thereby increasing the phase margin. As can be seen from the response to the step disturbance, the performance, in terms of maximum peak and settling time, is about the same but with the oscillatory behavior removed.

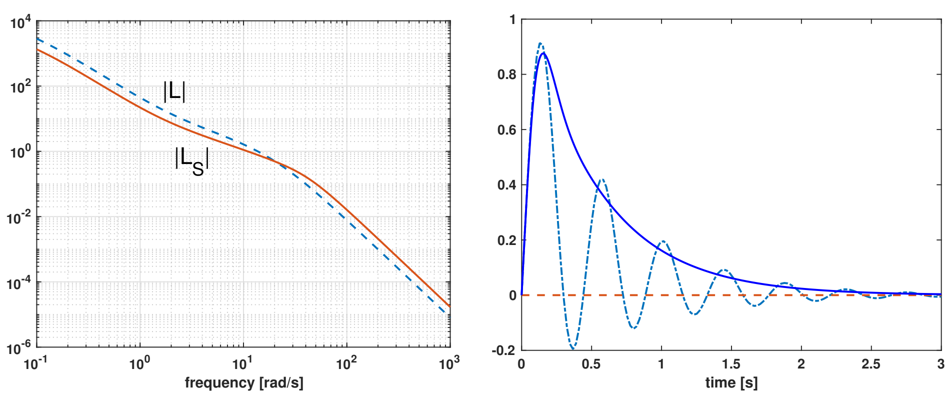


Figure 10.2: Loop gains and simulations for Example.

11 Appendix. Computing normalized left coprime factorizations

This appendix is for the interested only (not exam relevant).

Consider a state-space realization of a square transfer-matrix $G(s)$

$$\begin{aligned} \dot{x} &= Ax(t) + Bu(t) \\ y(t) &= Cx(t) + Du(t) \end{aligned} \tag{4}$$

Then, the left normalized coprime factors $M(s)$ and $N(s)$, in $G(s) = M^{-1}(s)N(s)$, are given by

$$N(s) = R^{-1/2}C(sI - A - HC)^{-1}(B + HD) + R^{-1/2}D$$

$$M(s) = R^{-1/2}C(sI - A - HC)^{-1}H + R^{-1/2}$$

where $H = -(BD^T + ZC^T)R^{-1}$ and $R = I + DD^T$. Here $Z > 0$ is the positive definite solution to the algebraic Riccati equation

$$(A - BS^{-1}D^TC)Z + Z(A - BS^{-1}D^TC)^T - ZC^TR^{-1}CZ + BS^{-1}B^T = 0$$

with $S = I + D^TD$.

The result is due to Vidyasagar, *Control Systems Synthesis: A Factorization Aproach*, MIT Press (1985)

EL2520
Control Theory and Practice

Lecture 11:
Classical and Modern Optimal Control revisited & Case Study

Elling W. Jacobsen
School of Electrical Engineering and Computer Science
KTH, Stockholm, Sweden

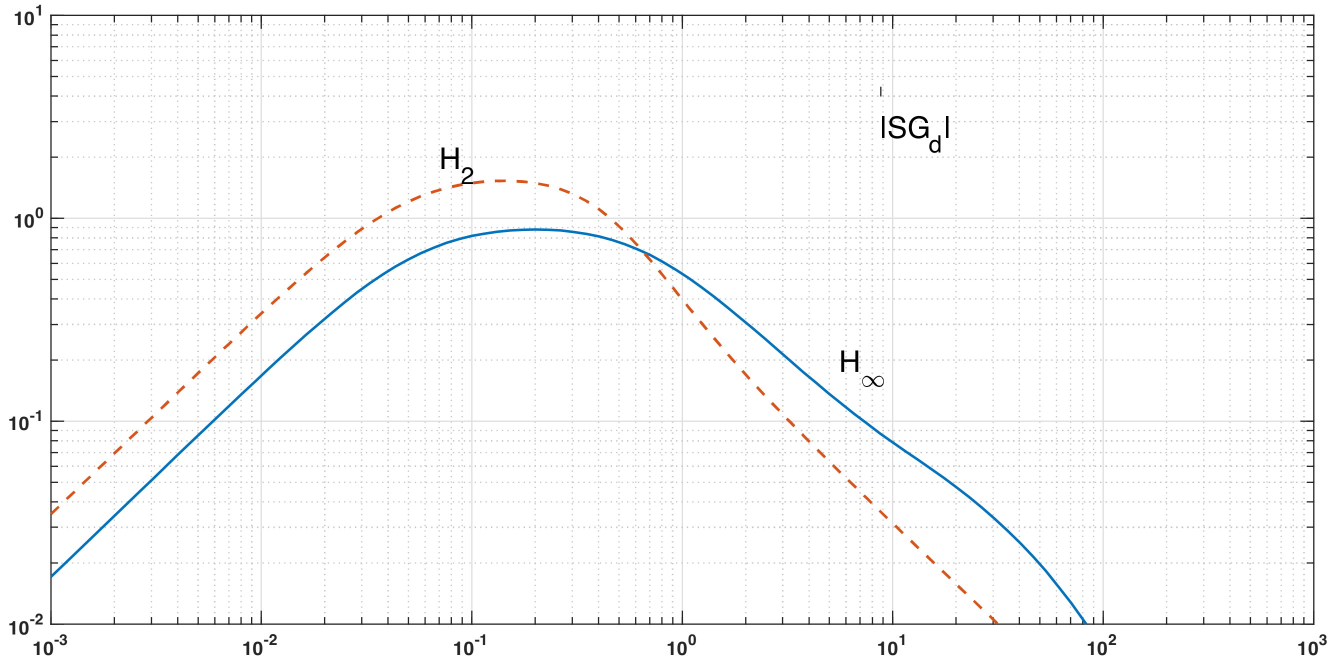
Today's lecture

- Brief review of design methods covered so far, including interpretations and comparisons
- Case study: disturbance attenuation in a chemical reactor

Summary

- All synthesis methods considered in this course can be used for signal minimization, or for shaping closed-loop transfer-functions (equivalence exist in all cases)
- Although possible, LQG not well suited for shaping closed-loop transfer-functions. Then, H2-optimal control more direct.
- H_∞-optimal control can be used for explicit design for robust stability, and can provide performance guarantees

Disturbance sensitivity



Optimal Control

- Classical optimal control LQG
  - originally motivated by the need for methods that could deal with MIMO systems
  - formulated as optimization problem in state-space (time)

- Modern optimal control H_∞ and H_2
  - originally motivated by need to explicitly address robustness
  - formulated as optimization problem in input-output space (frequency domain)
  - solved in state-space

- All formulations result in controllers on the form observer + state feedback

Robust Loop shaping

- Combines classical loop shaping with H_∞-optimal control where the optimization step only addresses robust stability

Design III: H_2 signal based control

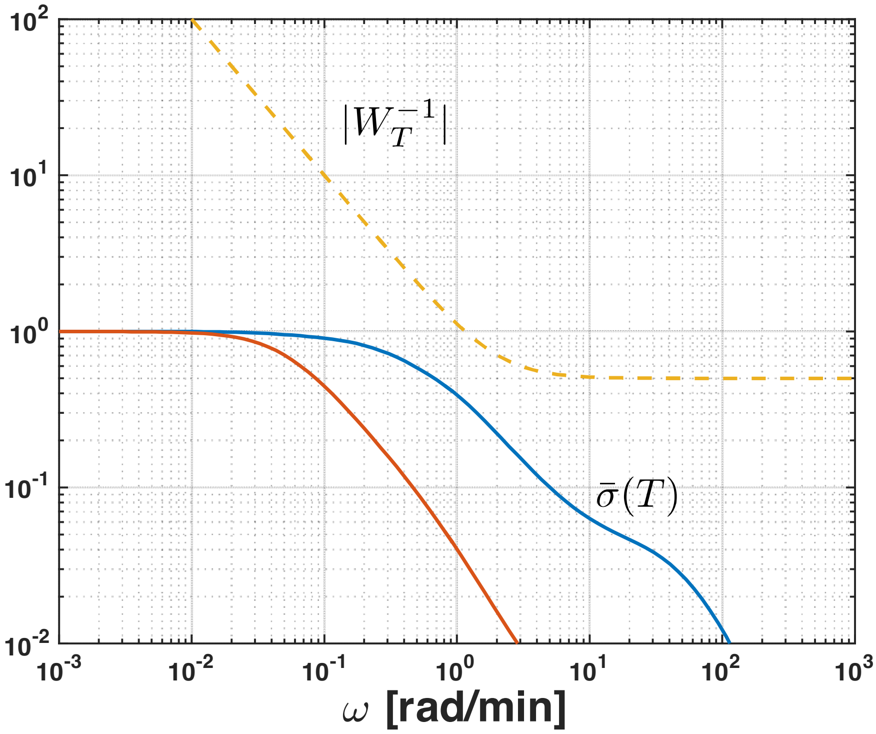
```
WS=((s+.05)/(s+1.d-4)); % Adds integral action
WU=1;

d=icsignal(1);
u=icsignal(2);
y=icsignal(2);
Wy=icsignal(2);
Wu=icsignal(2);
P=iconnect;
P.Input=[d;u];
P.Output=[Wy;Wu;-y];
P.equation{1}=equate(y,G*u+Gd*d);
P.equation{2}=equate(Wy,WS*y);
P.equation{3}=equate(Wu,WU*u);

[C,CL,gam]=h2syn(P.System,2,2);

>> gam
   gam =
       0.720
```

Robustness



Linear Quadratic Gaussian control

Model: linear system with white noise

$$\frac{d}{dt}x(t) = Ax(t) + Bu(t) + Nv_1(t)$$
$$y(t) = Cx(t) + v_2(t)$$
$$z(t) = Mx(t)$$

where v_1, v_2 are white noise with

$$\text{cov}([v_1,\ v_2]) = \begin{bmatrix} R_1 & R_{12} \\ R_{12}^T & R_2 \end{bmatrix}$$

Objective: minimize effect of v on z, punish control cost

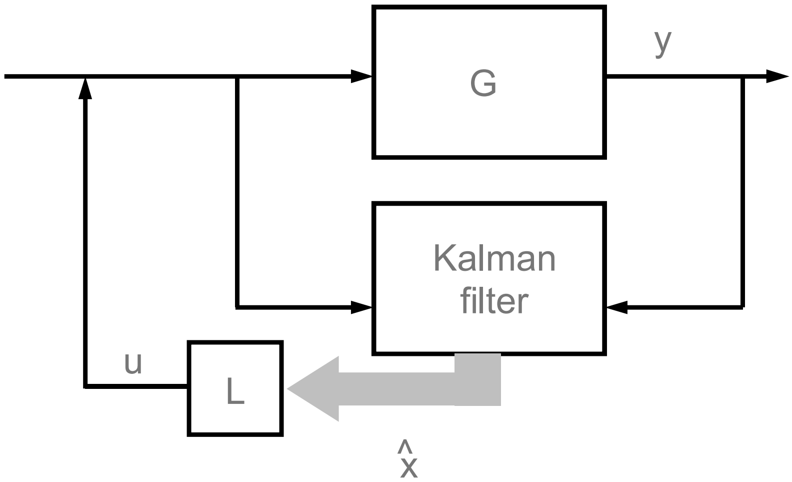
$$J = \mathbf{E}\left\{\lim_{T\to\infty} \frac{1}{T}\int_0^T [z^T Q_1 z + u^T Q_2 u]\ dt\right\}$$

LQG: Linear system, Quadratic cost, Gaussian noise

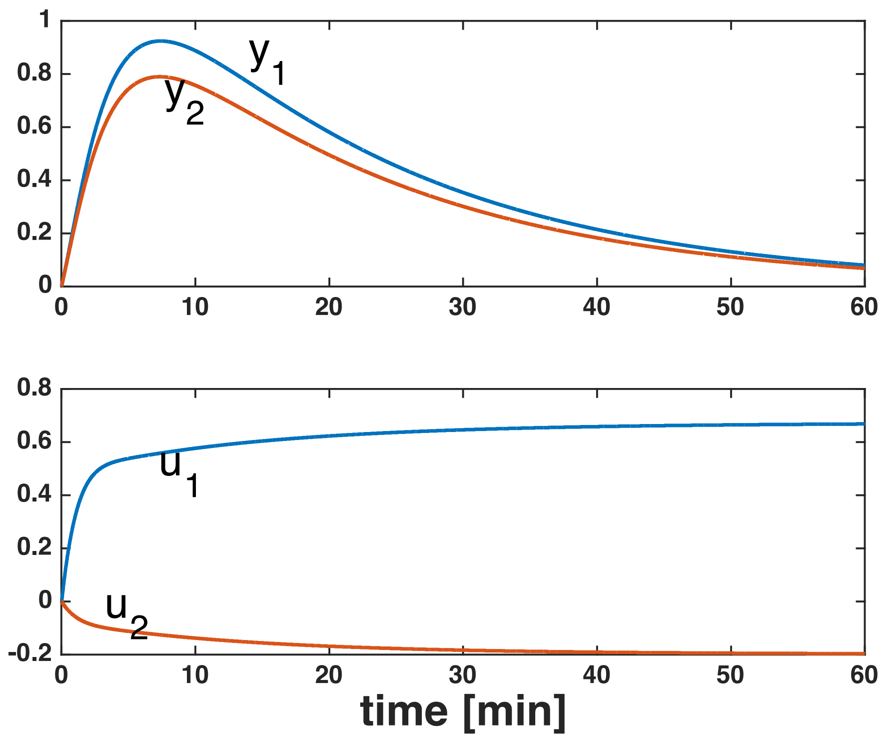
Solution structure

Optimal solution satisfies *separation principle*, composed of

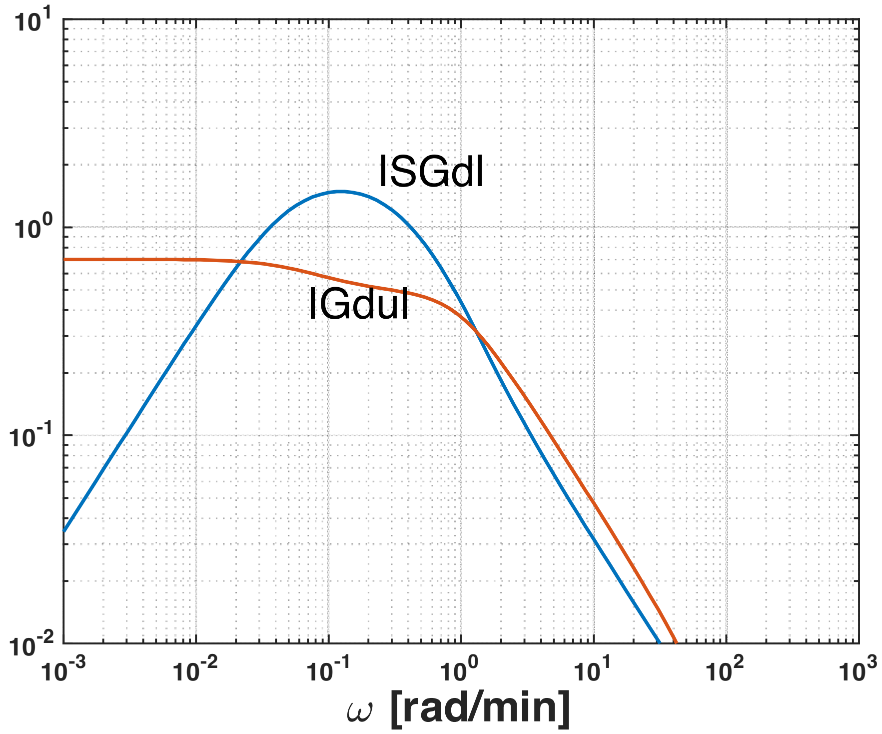
- Optimal linear state feedback (LQ-problem, no noise)
- Optimal observer (Kalman filter, no control)



Simulation



Performance



Optimal solution

State feedback (LQ problem)

$$u(t) = -L\hat{x}(t) = -Q_2^{-1}B^T S\hat{x}(t)$$

where S>0 is the solution to the algebraic Riccati equation

$$A^T S + SA + M^T Q_1 M - SBQ_2^{-1}B^T S = 0$$

Kalman filter

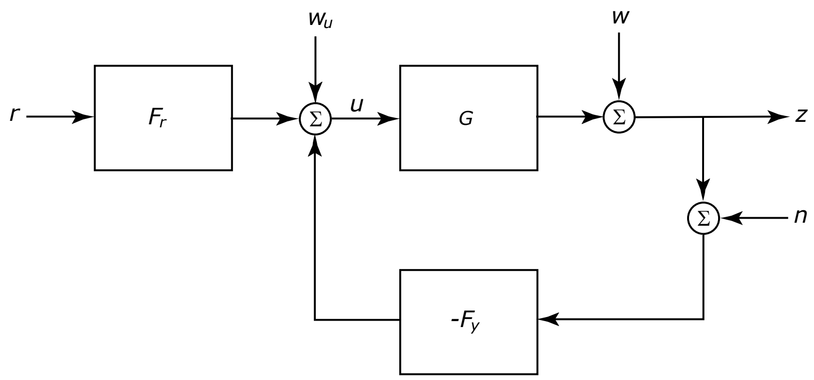
$$\dot{\hat{x}} = A\hat{x}(t) + Bu(t) + K(y(t) - C\hat{x}(t)); \quad K = (PC^T + NR_{12})R_2^{-1}$$

where P>0 is the solution to algebraic Riccati equation

$$AP + PA^T + NR_1N^T - (PC^T + NR_{12})R_2^{-1}(PC^T + NR_{12})^T = 0$$

- Combination of optimal state feedback, with no noise, and optimal observer, with no control, is the optimal controller to the combined problem
- Tuning parameters: Q_1, Q_2, R_1, R_2 (usually assume $R_{12} = R_{21} = 0$)

H_∞–optimal control



Aim: shape closed loop transfer-functions, e.g., S, T, G_{wu}, to achieve desired system properties

How: introduce weights W_S, W_T, W_u and determine F_y, F_r such that

$$\|W_S S\|_\infty < 1 \quad \|W_T T\|_\infty < 1 \quad \|W_u G_{wu}\|_\infty < 1$$

$$\Downarrow$$

$$\bar{\sigma}(S(i\omega)) < |w_S^{-1}(i\omega)| \quad \bar{\sigma}(T(i\omega)) < |w_T^{-1}(i\omega)| \quad \bar{\sigma}(G_{uw}(i\omega)) < |w_u^{-1}(i\omega)| \quad \forall\omega$$

where we assume $W_S = w_S I$ etc.

Design II: add robustness criterion

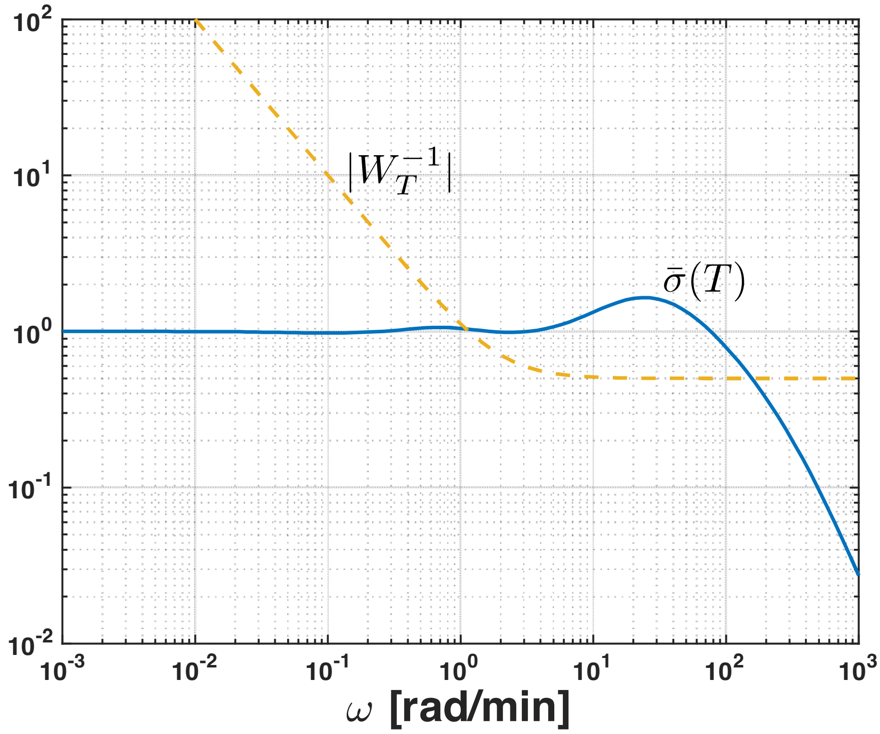
```
WS=((s+.05)/(s+1.d-4)); % Adds integral action
WU=(s+0.5)/(s+1); % to put relatively smaller weight on low frequency inputs
WT=2*s/(s+2); % For robust stability in presence of output uncertainty

d=icsignal(1);
r=icsignal(2);
u=icsignal(2);
y=icsignal(2);
Wy=icsignal(2);
Wu=icsignal(2);
Wt=icsignal(2);
P=iconnect;
P.Input=[d;r;u];
P.Output=[Wy;Wt;Wu;r-y];
P.equation{1}=equate(y,G*u+Gd*d);
P.equation{2}=equate(Wy,WS*(r-y));
P.equation{3}=equate(Wu,WU*u);
P.equation{4}=equate(Wt,WT*y);

[C,CL,gam]=hinfsyn(P.System,2,2);

>>gam
   gam = 1.796
```

Robustness



32

Selecting Weights

Weights W_S, W_T, W_u should

- reflect our requirements on performance and robustness, e.g., W_S large for frequencies where we need disturbance attenuation, W_T large where we want noise attenuation and where model uncertainty is large.
- take into account trade-offs and limitations, e.g., S+T=I, RHP poles, RHP zeros and time delays, such that $\|\cdot\|_\infty < 1$ is feasible, i.e., there exists stabilizing controller that meets specifications.

Usually a good idea to scale all signals, such that their expected / allowed magnitude is less than 1, prior to designing weights.

Controller Design – H_∞

How determine $F_y(s)$ to achieve $\|W_S S\|_\infty < 1 \quad \|W_T T\|_\infty < 1 \quad \|W_u G_{wu}\|_\infty < 1$?

- *Loop shaping*, i.e., shape loop gain $L = GF_y$

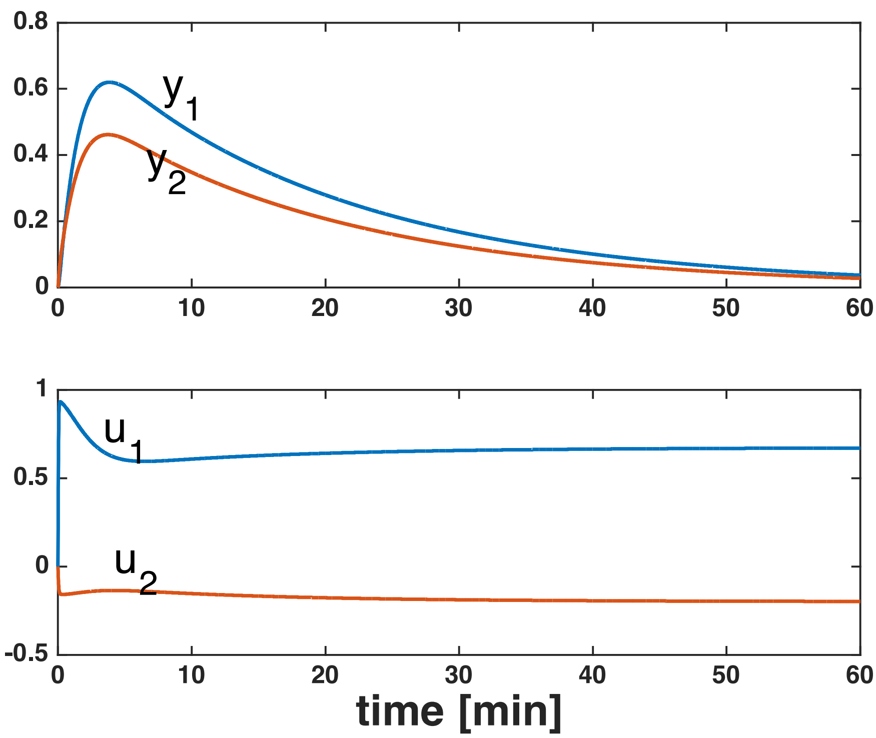
  or

- *Synthesis,* i.e., solve optimization problem

$$F_{y,opt}(s) = \arg\min_{F_y} \left\| \begin{matrix} W_S S \\ W_T T \\ W_u G_{wu} \end{matrix} \right\|_\infty \qquad (*)$$

Note that the if the "stacked" objective above is less than 1, then we have achieved the three individual objectives

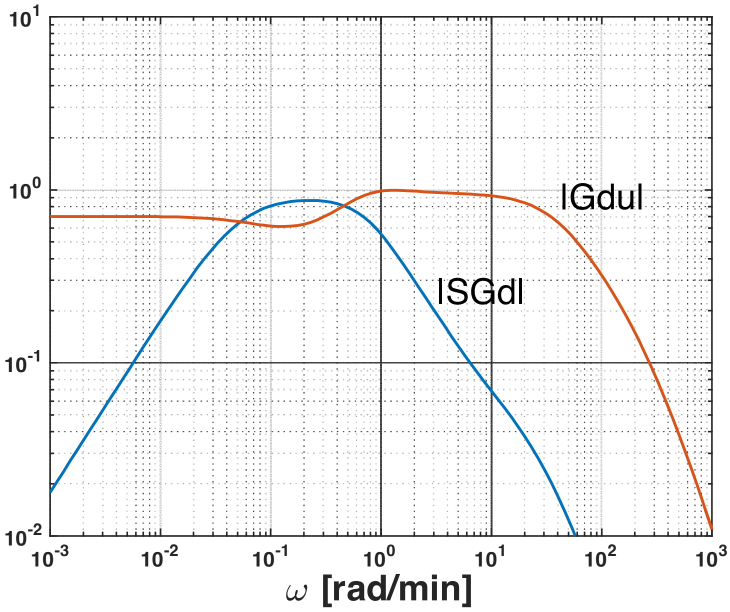
Simulation



Add weight to input

- Try to push down input usage at high frequencies by introducing

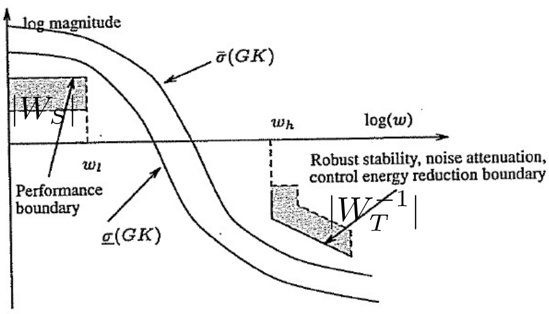
$$W_U(s) = \frac{s + 0.5}{s + 1}$$



Loop Shaping

Need to translate bounds on $\bar{\sigma}(S)$ and $\bar{\sigma}(T)$ into bounds on $\sigma_i(L),\ L = GF_y$

- $\underline{\sigma}(L) >> 1 \ \Rightarrow \ \bar{\sigma}(S) \approx 1/\underline{\sigma}(L)$ and hence $\underline{\sigma}(L) > |w_S|$ where $|w_S| >> 1$
- $\bar{\sigma}(L) << 1 \ \Rightarrow \ \bar{\sigma}(T) \approx \bar{\sigma}(L)$ and hence $\bar{\sigma}(L) < |w_T^{-1}|$ where $|w_T| >> 1$



- Robust loop shaping; robustify the shaped plant by maximizing robustness margin for coprime uncertainty

H_∞ Synthesis

- Given a state space system G_0 on the form

$$\begin{aligned} \dot{x} &= Ax + Bu + Nw_e \\ z_e &= Mx + Du \qquad\qquad (**) \\ y &= Cx + w_e \end{aligned}$$

- Determine if a controller $u = -F_y(s)y$ exists such that for the resulting closed-loop system G_{ec} and given γ

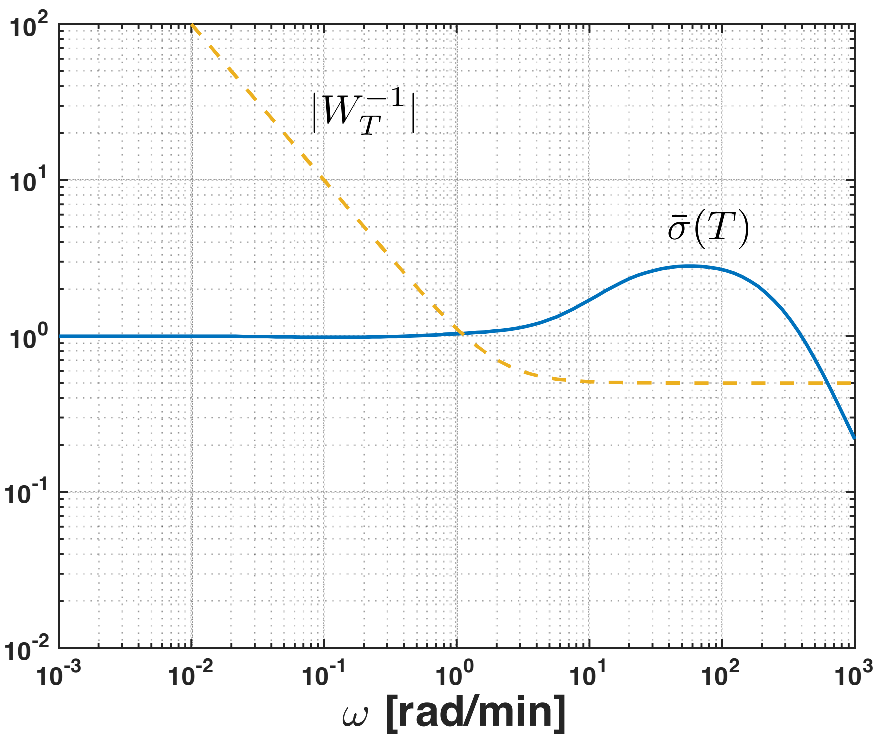
$$\sup_{w_e \neq 0} \frac{\|z_e\|_2}{\|w_e\|_2} < \gamma \qquad (1)$$

- Let P>0 be a solution to the algebraic Riccati equation

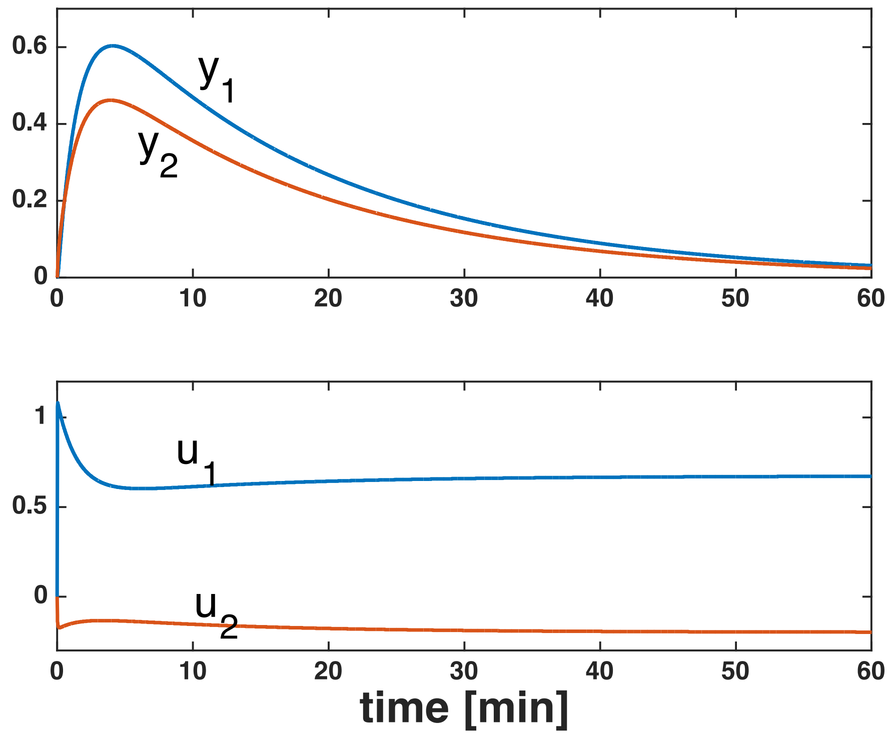
$$A^T P + PA + M^T M + P(\gamma^{-2} N N^T - BB^T)P = 0$$

if $A - BB^T P$ is stable then the controller exists, otherwise not

Robust Stability



Simulation unit step disturbance



The H_∞–optimal controller

- A controller satisfying requirement (1) is then given by

$$\begin{aligned} \dot{\hat{x}} &= A\hat{x} + Bu + N(y - C\hat{x}) \\ u &= -L_\infty \hat{x} \; ; \quad L_\infty = B^T P \end{aligned}$$

i.e., state observer combined with state feedback

- The optimal controller can be found by iterating on γ untill $\gamma \approx \gamma_{min}$

- To solve the original problem $(*)$ we note that with $z_e = G_{ec} w_e$

$$\sup_{w_e \neq 0} \frac{\|z_e\|_2}{\|w_e\|_2} < \gamma \;\Leftrightarrow\; \|G_{ec}\|_\infty < \gamma$$

- Thus, select the output z_e and input w_e such that

$$z_e = G_{ec} w_e = \begin{bmatrix} W_S S \\ W_T T \\ W_u G_{wu} \end{bmatrix} w_e$$

and determine corresponding open-loop system G_0

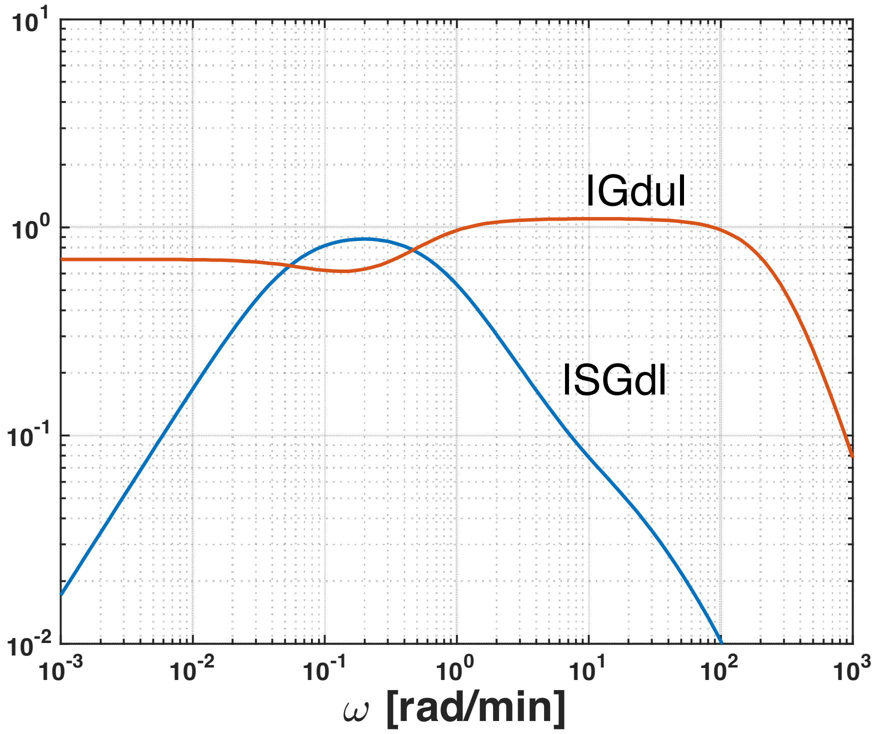
H2-optimal control

- Similar to in H$_\infty$-synthesis, define the extended system G_0

$$
\begin{aligned}
\dot{x} &= Ax + Bu + Nw_e \\
z_e &= Mx + Du \\
y &= Cx + w_e
\end{aligned}
$$

  such that that the closed-loop transfer-matrix, with $u = -F_y y$, is the one we want to minimize
- If w_e is white noise with intensity Φ_{w_e}= I, then minimizing $\|z_e\|_2$ corresponds to minimizing $\|G_{ec}\|_2$ where G_{ec} is closed-loop transfer-matrix from w_e to z_e
- Corresponds to LQG problem with $Q_1 = Q_2 = R_1 = R_2 = I$

Performance



Design I: H_∞ signal based, no robustness

```
WS=((s+.05)/(s+1.d-4)); % Adds integral action
WU=1;

d=icsignal(1);
u=icsignal(2);
y=icsignal(2);
Wy=icsignal(2);
Wu=icsignal(2);
P=iconnect;
P.Input=[d;u];
P.Output=[Wy;Wu;-y];
P.equation{1}=equate(y,G*u+Gd*d);
P.equation{2}=equate(Wy,WS*y);
P.equation{3}=equate(Wu,WU*u);

[C,CL,gam]=hinfsyn(P.System,2,2);

>> gam
   gam =
       1.104
```


H_2 and H_∞ optimal control

H_2-optimal control

$$\min_{F_y} \|G_{ec}\|_2^2 = \min_{F_y} \int_{-\infty}^{\infty} \sum_i \sigma_i^2(G_{ec}(i\omega))\, d\omega$$

(reduce all singular values at all frequencies)

H_∞-optimal control

$$\min_{F_y} \|G_{ec}\|_\infty = \min_{F_y} \sup_\omega \overline{\sigma}(G_{ec}(i\omega))$$

(reduce maximum singular value at worst frequency)

Design example

DC servo from Lecture 5:

$$G(s) = \frac{1}{s(s+1)}$$

Same performance requirements as in Lec 9

Two key points:

– H_∞ optimal design allows to work directly with constraints
– The relation between H_2 and H_∞ optimal controllers

Controllability Analysis (see lec 6)

- Is control needed?

  Yes, since $\|G_d\|_\infty > 1$ and requirement is $\|SG_d\|_\infty < 1$

- What is required bandwidth?

  $$\|G_d(i\omega)\|_2 > 1, \;\; \omega < 0.35$$

  Thus, need to attenuate disturbance up to $\omega \approx 0.35$

- Any fundamental limitations?

  Yes, there is a RHP zero $z = 0.62$ which gives bandwidth limitation for S $\omega_{BS} \lesssim 0.31$

- Conflict? Need to attenuate disturbances only in disturbance direction
  – zero direction $y_z = [0.816 \; 0.578]^T$
  – with $G_d(z) = [0.313 \; 0.340]^T$ we get $|y_z^H G_d(z)| = 0.45 < 1$ OK!

Chemical Reactor Model

- Scaled model (so that all signals should have magnitude <1 at all frequencies)

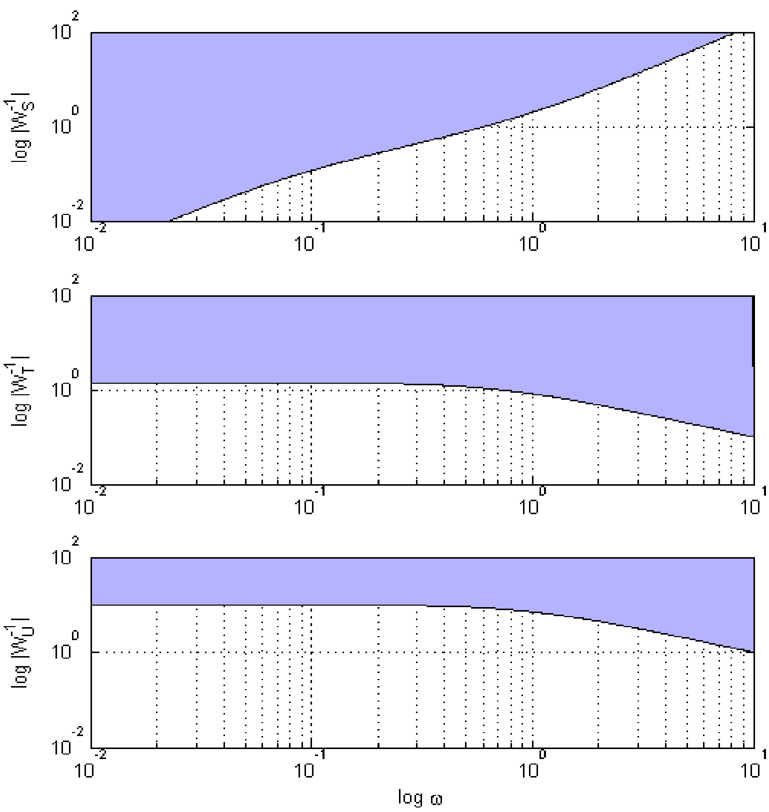
$$G(s) = \frac{1}{(15.2s+1)(3.1s+1)} \begin{pmatrix} 22.4(3.1s+1) & 59.4(8.3s+1) \\ -12.6(10.2s+1) & -60.6(12.1s+1) \end{pmatrix}$$

$$G_d(s) = \frac{1}{(15.2s+1)} \begin{pmatrix} 3.28 \\ 3.56 \end{pmatrix}$$

  - Aim: keep output deviations less than 1 in presence of disturbances with magnitude up to 1.
  - uncertainty: measurement uncertainty exceeds 100% for frequencies above 1 rad/min. Use uncertainty weight

$$W_T(s) = \frac{2s}{s+2} I$$

Weights

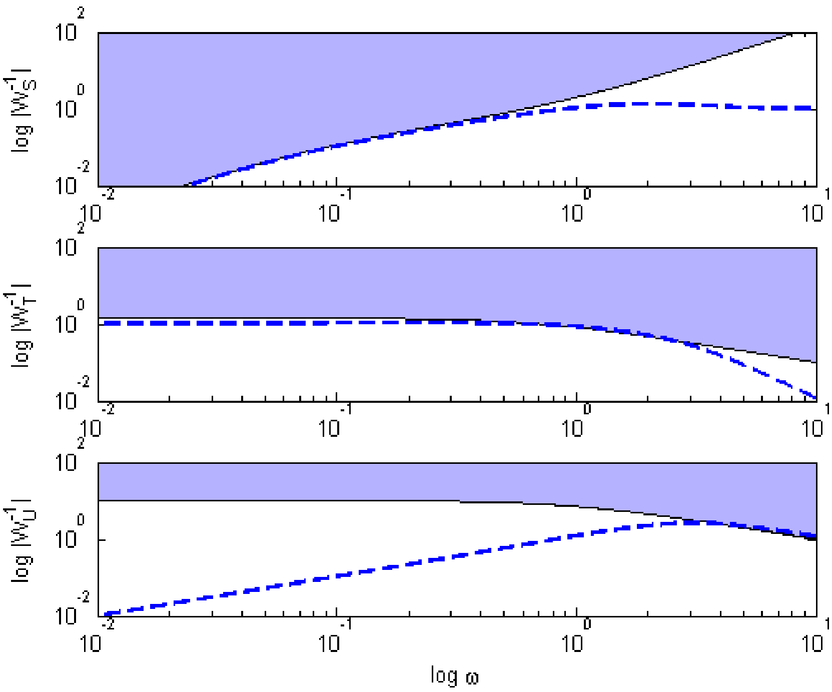


$$W_S(s) = \frac{0.71s + 0.05}{s^2(s+1)}$$

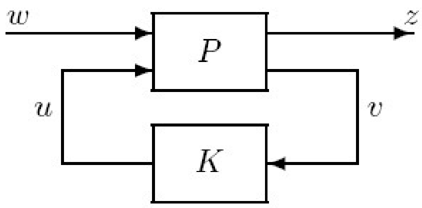
$$W_T(s) = s + 0.71$$

$$W_U(s) = \frac{10s + 10}{s + 100}$$

H_∞ optimal control



H_2/H_∞ optimal control for signal minimization

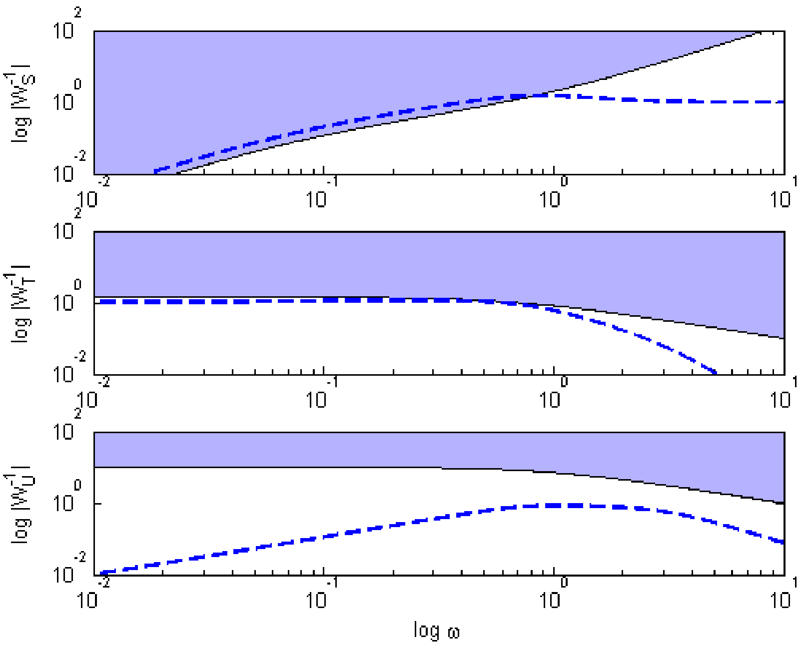


- So far: select input w and output z to reflect system transfer-function G_{ec} that we want to shape
- But: in many cases specifications may be directly on signals, e.g., keep an output less than some constraint in the presence of a bounded disturbance. Then w and z given directly by problem formulation.
- Note that weightings still will be in the frequency domain
- Case study: controller design for chemical reactor using signal minimization and sensitivity shaping, respectively

LQG for System Optimization

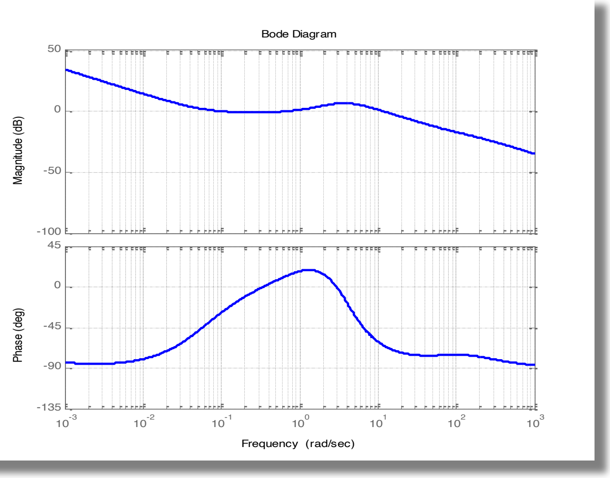
- LQG may be used to obtain desired system properties, e.g., sensitivity and complementary sensitivity
- Systematic method, called Loop Transfer Recovery (LTR)
  - Main drawback: based on cancellations, and will even cancel RHP zeros and hence loss of internal stability
  - Not treated in this course
- Better to use H_2 - optimization which is a direct method, also based on using the LQG machinery

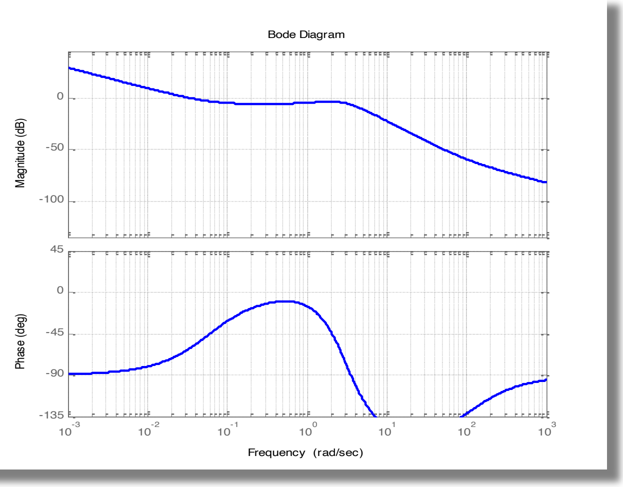
H_2-optimal controller



Quiz: why doesn't the H_2-optimal controller "meet the specs"?

Comparing the controllers





Signal vs System Optimization

- LQG is formulated as signal minimization problem, i.e., minimize weighted control error and weighted control input in the presence of (filtered) white noise disturbances
- H_2– and H_∞-optimal control usually considered as system norm minimization, e.g., minimize weighted sensitivity and weighted complementary sensitivity
- But, equivalence exists between system properties and signal properties, e.g.,

$$\sup_{w_e \neq 0} \frac{\|z_e\|_2}{\|w_e\|_2} = \|G_{ec}\|_\infty$$

- Also, equivalence exist between LQG and H_2–optimal control; minimizing 2-norm of output in presence of white noise equals minimizing 2-norm of corresponding transfer-function ver 1.0, 200429 Elling W. Jacobsen, KTH

12 EL2520 Lecture notes 12: Model Predictive Control

Model Predictive Control (MPC) has gained an increasing popularity in almost all branches of industry in the last decades and is becoming something of a standard controller on par with PID, albeit significantly more advanced. The basic idea in MPC is to use the available dynamic model to predict the future, and use the control input to optimize the predicted future. The objective is similar to what is used in LQG, namely minimizing a quadratic objective function in the future process outputs and control inputs. However, while LQG was introduced in continuous time and with an infinite time horizon, MPC is based on discrete time computations and considers a finite time horizon. Although the considered horizon is finite, the optimization is repeated at every sample instance, and with a constant horizon this implies that the horizon is constantly moving forward in time. For this reason, MPC is also often called *receding horizon control*. One of the main advantages of MPC is that constraints on the control inputs and outputs can be included in the optimization.

Below we give a brief introduction to MPC. We start by introducing sampling and discrete time models, and then formulate and solve the discrete time finite horizon LQR[1] problem. A key is that the LQR solution is implemented in a receding fashion and we show that we then can include constraints in the problem.

12.1 Sampling and Discrete Time Models

When a controller is implemented on a digital computer, computations can only be done at discrete time instances. See Figure 12.1. The control input is computed at discrete time points and typically kept constant between these time instants (zero-order-hold). The continuous output from the process G is sampled every h seconds, the sampling time, and fed into the computer. The time of sampling the output and computing the control input are usually assumed to coincide.

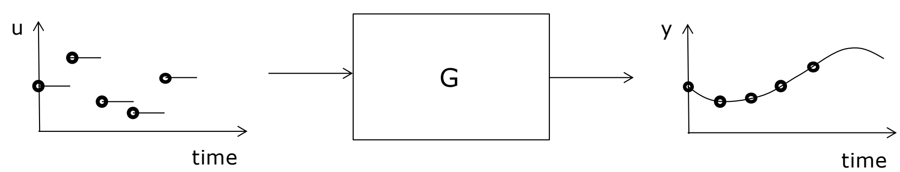


Figure 12.1: Zero-order hold input and output sampling.

To describe the process dynamics at the sampling instances, consider the system evolution

[1] Linear Quadratic Regulator, corresponding to the determinstic part of the LQG problem.

between two sample instances. The solution of the linear differential equations

$$\dot{x} = Ax(t) + Bu(t)$$

is

$$x(t+h) = e^{Ah}x(t) + \int_{s=0}^{h} e^{As}Bu(t+s)ds \tag{1}$$

If u is held constant between samples so that $u(t) = u_t, t \in [t, t+h]$, then

$$\begin{aligned} x(t+h) &= A_D x(t) + B_D u_t \ ; \quad A_D = e^{Ah} \ ; \quad B_D = \int_{s=0}^{h} e^{As}Bds & (2) \\ y(t) &= Cx(t) + Du_t & (3) \end{aligned}$$

This a discrete-time linear dynamic system. For notational convenience we drop the reference to physical time and write

$$\begin{aligned} x_{k+1} &= Ax_k + Bu_k & (4) \\ y_k &= Cx_k + Du_k & (5) \end{aligned}$$

where $u_0, u_1, \ldots$ is the input sequence, $y_0, y_1, \ldots$ is the output sequence and $x_0, x_1, \ldots$ is the state evolution. Note that the discrete time system is stable if all eigenvalues of A are inside the unit circle in the complex plane.

12.2 Finite Horizon LQR

The Linear Quadratic Regulator (LQR) over a finite horizon N determines the input sequence $U = u_0, \ldots, u_{N-1}$ that minimizes a square function of the system states and control inputs over the considered horizon

$$\min_U J(U) = \min_U \sum_{k=0}^{N-1} (x_k^T Q_1 x_k + u_k^T Q_2 u_k) + x_N^T Q_f x_N \tag{6}$$

Here Q_1, Q_2, Q_f are positive definite weight matrices. Note that the final state cost, with weight Q_f, mainly is used to ensure stability.

To solve (6), introduce the notation $X = (x_0, x_1, \ldots, x_N), U = (u_0, u_1, \ldots, u_{N-1})$. Then, from model (5) we get

$$\begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ B & 0 & \cdots & 0 \\ AB & B & 0 & \cdots \\ \vdots & \vdots & & \\ A^{N-1}B & A^{N-2}B & \cdots & B \end{bmatrix} \begin{bmatrix} u_0 \\ u_1 \\ \vdots \\ u_{N-1} \end{bmatrix} + \begin{bmatrix} I \\ A \\ A^2 \\ \vdots \\ A^N \end{bmatrix} x_0$$

which can be expressed as

$$X = GU + Hx_0$$

where $G \in \mathbb{R}^{Nn \times Nm}, \ H \in \mathbb{R}^{Nn \times n}$.

The objective function $J(U)$ can be written on matrix form

$$J(U) = X^T \text{diag}(Q_1 \quad Q_1 \quad \cdots \quad Q_f) X + U^T \text{diag}(Q_2 \quad Q_2 \quad \cdots \quad Q_2) U = X^T \bar{Q}_1 X + U^T \bar{Q}_2 U$$

Now, inserting the model $X = GU + Hx_0$ we get

$$J(U) = (U^T G^T + x_0^T H^T) \bar{Q}_1 (GU + Hx_0) + U^T \bar{Q}_2 U = U^T \underbrace{(G^T \bar{Q}_1 G + \bar{Q}_2)}_{P_{LQ}} U + 2 \underbrace{x_0^T H^T \bar{Q}_1 G}_{q_{LQ}} U + \underbrace{x_0^T H^T \bar{Q}_1 H x_0}_{r_{LQ}}$$

Thus, we can write the objective function as

$$J(U) = U^T P_{LQ} U + 2 q_{LQ} U + r_{LQ}$$

Taking the derivative with respect to the input sequence U we get

$$\frac{dJ(U)}{dU} = 2 P_{LQ} U + 2 q_{LQ}$$

and setting this to zero[2] we get the optimal control

$$U^* = -P_{LQ}^{-1} q_{LQ}$$

12.3 Constrained Receding Horizon Control

The solution to the discrete time LQR problem above is an open-loop control with no feedback. In order to incorporate feedback, to deal with model uncertainty and disturbances, one can implement it in a receding horizon fashion in which the optimization is repeated at every sample and the system state is updated at every sample using an observer, e.g., a Kalman filter. By keeping a fixed length of the horizon in every sample, the horizon will then move forward in time for each sample, hence the name receding horizon control. The use of an observer to update the state estimate at every sample ensures that there is feedback from the output measurements. The algorithm is then, at each sample,

1. Solve LQR over horizon N
2. Implement u_0
3. Let system evolve one sample and update state estimate using an observer, and repeat from 1.

Model Predictive Control (MPC) is based on this receding horizon control principle, but with the addition of constraints in the LQR problem. The constrained LQR problem is

$$\begin{array}{ll} \text{minimize} & J(U) = \sum_{k=0}^{N-1} (x_k^T Q_1 x_k + u_k^T Q_2 u_k) + x_N^T Q_f x_N \\ \text{subject to} & u_{\min} \leq u_k \quad \leq u_{\max}, \;\; k = 0, \ldots, N-1 \\ & y_{\min} \leq C x_k \leq y_{\max}, \;\; k = 1, \ldots, N \\ & x_{k+1} = A x_k + B u_k \end{array} \tag{7}$$

[2]The second order derivative is positive

The addition of constraints is important since the control inputs are essentially always constrained in real applications and MPC can deal with this in an optimal fashion. The ability to set constraints on the outputs is also important in many applications, e.g., set a maximum temperature in a chemical reactor or set the minimum speed of an airplane.

The constrained LQR problem above can be written on the form of a Quadratic Programming (QP) problem

$$\begin{array}{ll} \text{minimize} & u^T P u + 2q^T u + r \\ \text{subject to} & Au \leq b \end{array} \tag{8}$$

To write the original problem (7) on the QP form (8), note first that we as above can write the objective function on the form

$$J(U) = U^T P_{LQ} U + 2q_{LQ} U + r_{LQ}$$

The objective function $J(U)$ is convex provided $Q_2 > 0$ (positive definite). To write the constraints on the form in (8), consider first the constraint $u_{min} \leq u_k \leq u_{max}$ which can be written

$$\begin{bmatrix} 1 \\ -1 \end{bmatrix} u_k \leq \begin{bmatrix} u_{max} \\ -u_{min} \end{bmatrix}$$

For the constraints on the output y we introduce $Y = \begin{pmatrix} y_0 & y_1 & \cdots & y_N \end{pmatrix}^T$ and the constraint

$$y_{\min}\, \mathbf{1} \leq Y \leq y_{\max}\, \mathbf{1}$$

where $\mathbf{1}$ is a vector with all elements 1. Let

$$\bar{C} = \begin{bmatrix} C & 0 & \dots & 0 \\ 0 & C & 0 & 0 \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & C \end{bmatrix}$$

Then $Y = \bar{C}X$ and

$$Y \geq y_{min}\, \mathbf{1} \;\Leftrightarrow\; \bar{C}(GU + Hx_0) \geq y_{min}\, \mathbf{1} \;\Leftrightarrow\; \underbrace{\bar{C}G}_{A_{\underline{y}}} U \geq \underbrace{y_{min}\, \mathbf{1} - \bar{C}Hx_0}_{b_{\underline{y}}}$$

Similarly,

$$\underbrace{\bar{C}G}_{A_{\bar{y}}} U \leq \underbrace{y_{max}\, \mathbf{1} - \bar{C}Hx_0}_{b_{\bar{y}}}$$

The overall QP problem is then

$$\min_U U^T P_{LQ} U + 2q_{LQ} U + r_{LQ}$$

subject to

$$\begin{bmatrix} A_{\bar{y}} \\ -A_{\underline{y}} \\ I \\ -I \end{bmatrix} U \leq \begin{bmatrix} b_{\bar{y}} \\ -b_{\underline{y}} \\ u_{max}\mathbf{1} \\ -u_{min}\mathbf{1} \end{bmatrix}$$

In summary, at each sample the MPC controller receives a state estimate for x_0 from an observer, e.g., a Kalman filter, and then solves the above QP problem. The first input

u_0 in the determined input sequence is then implemented on the system and the system evolves on sample time before obtaining a new state estimate, solving the QP problem and so on.

For an example, MPC control of the DC servo, see the lecture slides for Lecture 12.

Some remarks on the MPC

- The main tuning parameters are the weights Q_1, Q_2, Q_f and the length of the prediction horizon N. Often different horizons are used for the prediction and the control input.
- In some cases, the QP problem may become infeasible, that is, there exists no solution that meets all constraints. To avoid this it is common to soften the constraints on the outputs by introducing so called *slack* variables. In essence, one then allows these constraints to be violated but penalize the violation in the objective function. Infeasibility may still happen, and one therefore often have a backup solution, e.g., in terms of a standard LQ controller.

The main purpose of this lecture has been to introduce the concept of MPC. To learn more in-depth about the theoretical basis of MPC and its application we refer to the course EL2700 Model Predictive Control given in period 1.

13 EL2520 Lecture notes 13: Dealing with Hard Constraints

The focus of this course is on control of systems that can be described by linear time-invariant (LTI) models. This is the case for many, if not most, applications. The reason is that the aim of a control system typically is to keep a system close to some desired steady-state, and then the behavior can usually be well described by an LTI model[1]. However, there is one non-linearity present in essentially any control system that can not be described by a linear model, and that is hard constraints on the control inputs. Almost any control input, be it e.g., a valve or the power of an engine, has hard lower and upper limits. A valve can for instance at most be fully open or fully closed. In the previous lecture we saw how such constraints can be included in the control optimization in MPC, providing optimal control under input (and possibly output) constraints. In this lecture we will consider an approach based on adding a modification to a controller that has been designed based on an LTI model, e.g., an LQG-controller, an $\mathcal{H}_\infty$-optimal controller or a simple PID controller. This approach is usually termed *Anti Reset Windup*, referring to the fact that the integral (reset) part of a controller "winds up", i.e., just grows and grows, when there is a non-zero control error and the input is at a constraint.

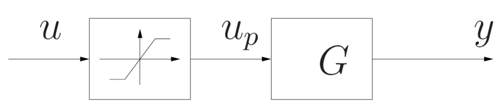


Figure 13.1: Hard constraint on input.

Consider the open-loop system in Figure 13.1. Here u is the input computed by the controller, i.e., the controller output, while u_p is the input actually implemented on the system. A hard constraint implies that $u_p = u_{min}$ when $u < u_{min}$, $u_p = u$ when $u_{min} < u < u_{max}$ and $u_p = u_{max}$ when $u > u_{max}$. Thus, when the computed input is outside the hard constraints, the implemented input is constant implying that the feedback loop is broken. See also Figure 13.2. With the feedback loop broken it implies that the poles of the system equals the open-loop poles, i.e., the union of the poles of $F_y(s)$ and $G(s)$. Thus, if either $F_y(s)$ or $G(s)$ is unstable the system becomes unstable as the input u_p saturates (is at a constraint). If $G(s)$ is unstable, it is not much we can do other than ensure we do not hit a constraint in the input. However, if the controller $F_y(s)$ is unstable then we can in principle modify the controller so that it becomes stable when the input is saturated. Note that most controllers are input-output unstable since they typically contain integral action, i.e., poles at $s = 0$.

We shall develop a modification of the controller as discussed above, starting with a

[1] In fact, feedback control usually tends to make a system much more linear than it is without control.

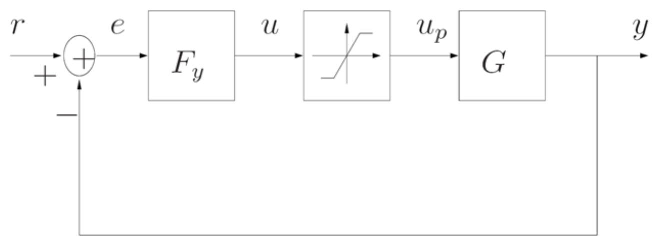


Figure 13.2: Feedback loop with constraint on input.

controller consisting of an observer with feedback of observed states

$$\begin{aligned} \dot{\hat{x}} &= A\hat{x}(t) + Bu(t) + K(y(t) - C\hat{x}(t)) && (1) \\ u(t) &= -L\hat{x}(t) && (2) \end{aligned}$$

On input-output form, the controller transfer-function is

$$U(s) = -F_y(s)Y(s) = -L(sI - A + BL + KC)^{-1}KY(s) \tag{3}$$

Note that we use the computed input u in the observer. A simple modification, provided we know the input constraints or we measure the true input u_p, is to replace u with the true input u_p in the observer. In the former case we effectively get a non-linear observer since we include a model of the saturation in the observer. The observer can now be written

$$\dot{\hat{x}} = A\hat{x}(t) + Bu_p(t) + K(y(t) - C\hat{x}(t)) = (A - KC)\hat{x}(t) + Bu_p(t) + Ky(t) \tag{4}$$

and on input-output form the controller becomes

$$U(s) = -L(sI - A + KC)^{-1}Y(s) + L(sI - A + KC)^{-1}BU_p(s) \tag{5}$$

When the input is not saturated, $u_p = u$ and the controller is the same as in (3) with poles given by the eigenvalues of $A - BL - KC$. However, when the input is saturated, u_p is a constant and the controller poles are given by the eigenvalues of $A - KC$ as can be seen from (5). Note that the controller in the latter case has poles equal to the poles of the observer which always should be stable. Thus, when the input saturates, the controller becomes stable. The modification correponds to what is known as anti reset windup.

The modification of the observer, as discussed above, is only applicable to controllers which can be written on the form of an observer with state feedback. However, we shall see that the resulting modification of the controller can be seen as input tracking, i.e., attempting to force u to be equal to u_p, and this idea can then be used also for other type of controllers. To this end, write (4) as

$$\dot{\hat{x}} = (A - KC)\hat{x}(t) + B(u_p(t) + u(t) - u(t)) + Ky(t) \tag{6}$$

With $u = -L\hat{x}$ we get

$$\dot{\hat{x}} = (A - KC - BL)\hat{x} + B(u_p(t) - u(t)) + Ky(t) \tag{7}$$

The controller can then be written

$$
\begin{aligned}
U(s) &= -L(sI - A + BL + KC)^{-1}KY(s) - L(sI - A + BL + KC)^{-1}B(U_p(s) - U(s)) \\
&= -F_y(s)Y(s) + W(s)(U_p(s) - U(s))
\end{aligned}
$$

Thus, the controller is partly trying to keep Y small (or, more generally, equal to the setpoint) and partly trying to keep the computed control input $u(t)$ equal to the implemented control input u_p. The control structure is illustrated in Figure 13.3. A typical choice for $W(s)$ is

$$
W(s) = \frac{1}{T_t s}
$$

that is, a pure integrator. The smaller the integral time T_t is chosen, the more emphasis is put on keeping $u(t) = u_p(t)$ and hence avoiding integral windup in the feedback controller $F_y(s)$ when the input $u_p(t)$ is at a constraint. Such input tracking is known as *anti reset windup*.

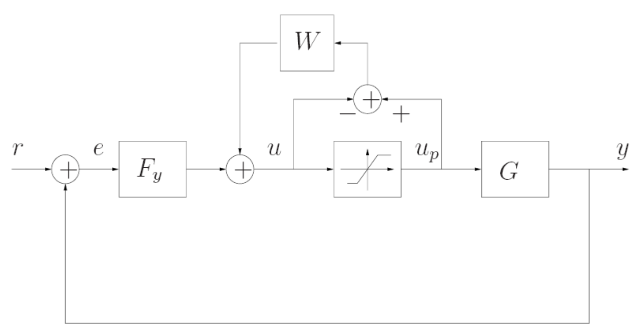


Figure 13.3: Feedback loop with constraint on input and input tracking.

EL2520
Control Theory and Practice

Lecture 14: Summary

Elling W. Jacobsen
School of Electrical Engineering and Computer Science
KTH, Stockholm, Sweden

Outline

- Course objectives
- Checklist
- The exam and beyond
- Brief review of lectures

Preparations for Exam

- Old exams with answers (usually not complete solutions) in Canvas
- An old exam will be covered in the final exercise (tomorrow Friday)
- A discussion group will be opened in Canvas to answer questions that may come up before the exam. I will do my best to answer all questions, but all of you feel free to post answers!
- Remember to motivate all your answers on the exam!

Course Objectives

The course aims to provide the participants basic knowledge of approaches and methods in advanced control, in particular linear multivariable feedback systems.

Key ingredients:

- A modern view of control
  - enabling systematic trade-off between various performance goals, respecting fundamental limitations and ensuring robust stability
- Multivariable control
  - multivariable systems (poles, zeros, gains and directions), decentralized control (RGA) and decoupling, several design methods based on controller synthesis through optimization, robust loop shaping
- Dealing with hard constraints
  - anti-windup and model predictive control

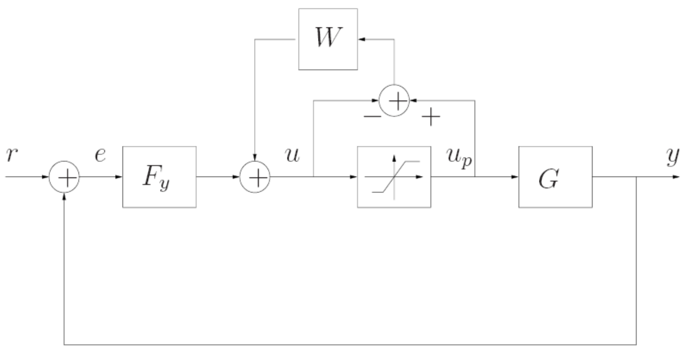
Checklist

The basics
- Matrix manipulations, eigenvalue computations, singular values
- Complex numbers
- Differential equations, state-space models, transfer functions and frequency response

A modern perspective on SISO control:
- Signal norms, system gain and the Small Gain Theorem
- The closed loop system and the central transfer functions
- Internal stability
- Fundamental limitations due to RHP poles/zeros, time delays
- Reasonable design goals and mapping to loop gain specifications
- Uncertainty sets and robust stability / performance

Lecture 13 – Dealing with hard constraints

Classical approach to deal with input constraints: input tracking, i.e., use feedback from difference between computed control and saturated control (Anti Reset Windup)



Lecture 12 – Model Predictive Control

Structured way of dealing with control and state constraints

All based on discrete time models!

1. Predict how state evolves (as function of future controls)

$$x_k = A^k x_0 + \sum_{j=0}^{k-1} A^{k-1-j} B u_j$$

2. Determine optimal control by minimizing LQ criterion w constraints

$$\begin{array}{ll} \text{minimize} & \sum_{i=0}^{N_p-1} [(x_i - x_i^{\text{ref}})^T Q_1 (x_i - x_i^{\text{ref}}) + (u_i - u_i^{\text{ref}})^T Q_2 (u_i - u_i^{\text{ref}})] \\ & +(x_{N_p} - x_{N_p}^{\text{ref}})^T S (x_{N_p} - x_{N_p}^{\text{ref}}) \\ \text{subject to} & u_{\text{min}} \leq u_i \leq u_{\text{max}} \\ & y_{\text{min}} \leq C x_i \leq y_{\text{max}} \end{array}$$

3. Implement first control, return to 1 at next sampling instant

Can be solved via efficient optimization (quadratic programming)

Checklist

Multivariable linear systems

- Transfer matrices and block diagram algebra
- Multivariable poles and zeros, directions
- Amplification, gain and directions
- Extending SISO results to MIMO

Multivariable control design techniques:

- H_2 and H_∞-optimal control: weighting functions and extended system, controller structure
- LQG: design, optimal control structure and disturbance models
- Loop shaping and Glover-McFarlane robustification
- The relative gain array for decentralized control structure design

Checklist

Dealing with hard constraints:

- Sampling of linear systems (continuous→discrete time)
- Model Predictive Control – MPC
  - finite horizon LQR w constraints –> Quadratic Program
- Anti-windup (tracking) to deal with actuator saturation

Lecture 11 – Comparing Design Methods, Case Study

Lecture 10 – Robust loop shaping

Glover Mc-Farlane: first perform classic loop shaping and then add robustifying controller in a second step

Solves the problem: find controller that stabilizes

$$G(s) = (M(s) + \Delta_M(s))^{-1}(N(s) + \Delta_N(s))$$

for all uncertainties

$$\|\Delta_M(s)\ \Delta_N(s)\|_\infty \leq \epsilon$$

Dynamic controller of high order (plant+nominal controller): model reduction based on balanced realization

About the exam

When? Tue May 31 at 08.00-13.00

How? Written on-campus exam (remember to register)

What? 5 problems, 5 hours. Grading criteria on homepage

Allowed aids?
Course book (Glad&Ljung *or* Skogestad&Postlethwaite)
Basic control book (Glad&Ljung)
Lecture notes and slides from this years course (printed!)
Mathematical handbook
Pocket calculator (not symbolic)
Note: all aids must be physical, i.e., not allowed to use any aids on computer or mobile phone

What is on the exam

- Signal and system norms, Small Gain Theorem
- Poles, zeros, directionality, singular values
- Closed-loop transfer functions
- Internal stability
- Performance limitations
- Input-output controllability analysis (requirements vs limitations)
- Robust stability
- RGA, decentralized control and decoupling
- LQG, H2- and Hinf-optimal control
- Glover-McFarlane robust loop shaping
- Discrete time systems, finite horizon LQR and MPC
- Anti reset windup

Lecture 8 – Linear Quadratic Control

Objective: minimize the cost

$$J = \mathbf{E}\left\{ \lim_{T\to\infty} \int_0^T [z^T Q_1 z + u^T Q_2 u]\, dt \right\}$$

Optimal solution:

- State feedback and linear observer (separation principle)
- Feedback and observer gains by solving Riccati equations
- LQG controller can have very poor robustness margins
- Some equivalence to $\mathcal{H}_2$- optimal control

Question

- How do we choose suitable weights in $\mathcal{H}_2$-optimal design?
  - in the same way as in $\mathcal{H}_\infty$-optimal design?
  - is aim still to achieve e.g., $\|W_T T\|_2 \leq 1$?

The main difference between $\mathcal{H}_2$ and $\mathcal{H}_\infty$ is that, in the latter case, the aim is to push down the peak value of $\bar{\sigma}(W_T T(i\omega))$ while in the former case the aim is to minimize the area under $\bar{\sigma}(W_T T(i\omega))$

Thus, $\|W_T T\|_2 \leq 1$ does not have a simple interpretation. Rather
- choose weight large at frequencies where you want T small
- if resulting T is too large in some frequency range, increase weight in that range and redo design

Beyond the exam

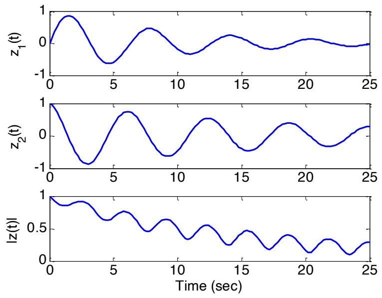
Learn more in our other advanced courses
- EL2820 Modeling of Dynamical Systems, per 1
- EL2700 Model Predictive Control, per 1
- EL2805 Reinforcement Learning, per 2
- EL2620 Nonlinear Control, per 2
- EL2450 Hybrid and Embedded Systems, per 3
- EL2810 Machine Learning Theory, per 3
- EL2425 Automatic Control Project Course, per 1-2

Put your skills to the test: do your master thesis at Decision and Control Systems
- see the web or come talk to us!

Contribute to frontline research: enroll in our PhD program!
- an exciting career – come and talk to us!

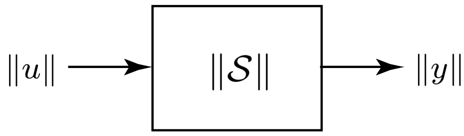
Lecture 1

- *Signal norms*: measure signal size across space (channels) and time



$\Rightarrow \|z\|$

- *System gain*: bounds signal amplification



Question

How choose weight W_u in objective $\|W_u G_{wu}\|_\infty < 1$

- We have

$$u = G_{wu}(s)w$$

- If system properly scaled so that $|w| < 1, \; |u| < 1 \; \forall\omega$, then we require

$$\bar{\sigma}(G_{wu}(i\omega)) < 1 \; \forall\omega \quad \Rightarrow \quad \|G_{wu}\|_\infty < 1$$

  – i.e., weight is $W_u = I$
- More generally
  – increase $|W_u(i\omega)|$ at frequencies where you want to use less input
  – e.g., if initial u(t) too large, increase the weight at higher frequencies
  – to limit derivative of u(t), use $W_u' = sW_u$

Lecture 7/9 – H_∞ / H_2 -optimal control

H_2-optimal

$$\min_{F_y} \|G_{ec}\|_2^2 = \min_{F_y} \int_{-\infty}^{\infty} \sum_i \sigma_j^2(G_{ec}(i\omega))\, d\omega$$

H_{inf}-optimal

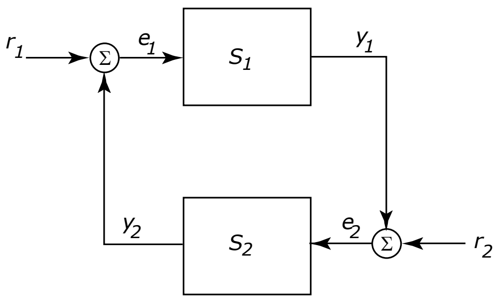
$$\min_{F_y} \|G_{ec}\|_\infty = \min_{F_y} \sup_\omega \overline{\sigma}(G_{ec}(i\omega))$$

Computed from state-space description of G_{ec}

- Solution is observer + static feedback from observed states

Small Gain Theorem

Theorem. Consider the interconnection



If $\mathcal{S}_1$ and $\mathcal{S}_2$ are input-output stable and

$$\|\mathcal{S}_1\| \cdot \|\mathcal{S}_2\| < 1$$

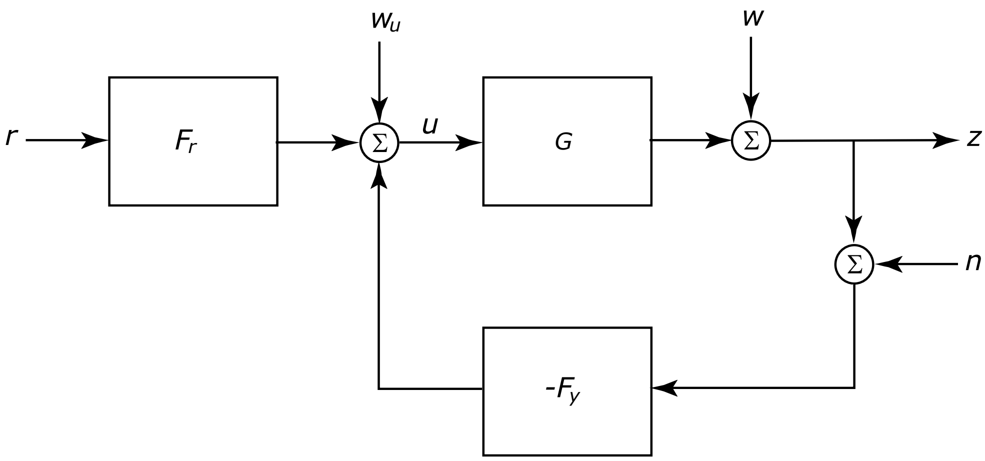
Then, the closed-loop system with r_1, r_2 as inputs and e_1, e_2, y_1, y_2 outputs is input-output stable.

Note: can use any norm that satisfies multiplicative property

$$\|AB\| \leq \|A\| \|B\|$$

e.g., inf-norm but not 2-norm

Lecture 2 - The closed-loop system



Controller: feedback F_y and feedforward F_r
Disturbances: w, w_u drive system from desired state
Measurement noise: corrupts information about z

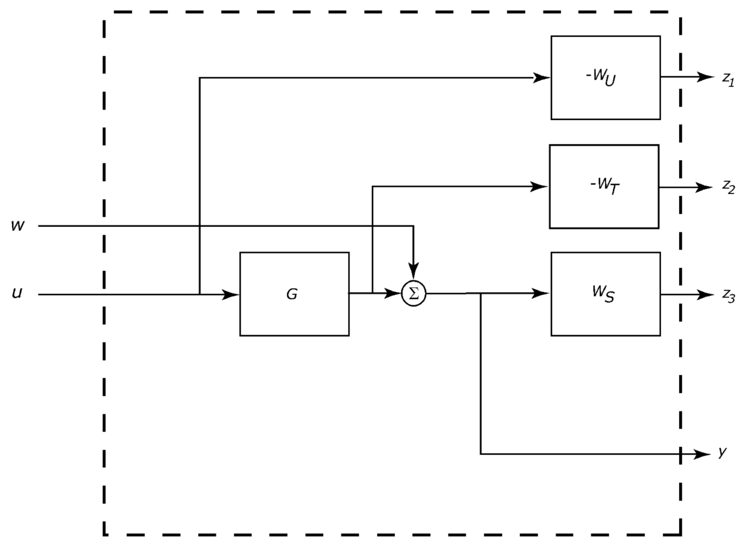
Aim: find controller such that z follows r.

Lecture 7/9 – H_∞ / H_2 -optimal control

Specification: optimize over all stabilizing controllers to achieve

$$\left\| \begin{pmatrix} W_S S \\ W_T T \\ W_U S F_y \end{pmatrix} \right\| \leq 1$$

Based on 'extended system'



Optimal controller:

- observer+linear feedback from estimated states

Question

- How to get the MIMO limitations in Lecture 6 from math?
  - Essentially combine interpolation constraints, e.g.,
    $$y_z^H S(z) = y_z^H$$
    with the Maximum Modulus Thm

  See also notes and slides from Lecture 6

Transfer functions and observations

$$S = \frac{1}{1+GF_y} \qquad (w \to z,\ w_u \to u) \quad \text{sensitivity function}$$

$$T = \frac{GF_y}{1+GF_y} \qquad (n \to z) \quad \text{complementary sensitivity}$$

$$G_c = \frac{GF_r}{1+GF_y} \qquad (r \to z) \quad \text{closed loop system}$$

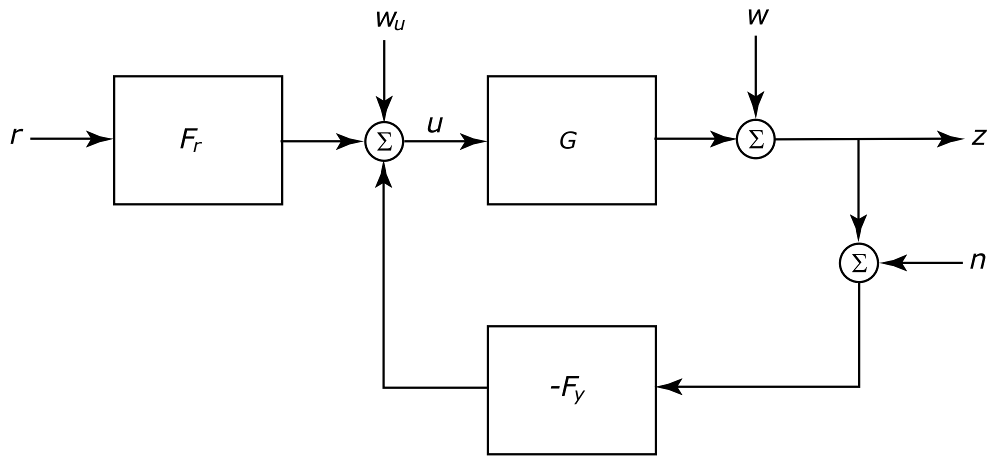
$$SG = \frac{G}{1+GF_y} \qquad (w_u \to z)$$

$$SF_y = \frac{F_y}{1+GF_y} \qquad (n \to u)$$

$$SF_r = \frac{F_r}{1+GF_y} \qquad (r \to u)$$

Observations: need to look at all! Many tradeoffs (e.g. S+T=1)

Internal Stability



Definition. The closed loop system above is *internally stable* iff it is input-output stable from all inputs r, w_u, w, n to all outputs u, z, y

Theorem. The closed-loop system is stable if and only if

$$S,\ SG,\ SF_y,\ F_r$$

are stable

Lecture 6 – MIMO Limitations

- Extend SISO limitations to MIMO case
- Essentially, we get interpolation constraints in certain directions (zero and pole directions)
- Otherwise derivations and limitations as for SISO case

Question

- How does decoupling affect robustness when the inputs are uncertain?
  - if we have input uncertainty then $u_p = (I+\Delta)u$ and we get for the compensated plant

$$G(s)(I+\Delta)d(s)G^{-1}(s) = d(s)(I+G(s)\Delta G^{-1}(s))$$

  - the term $G(s)\Delta G^{-1}(s)$ can become very large for ill-conditioned G(s)
  - Example with 10% input uncertainty

$$G = \frac{1}{s+1}\begin{pmatrix} 1 & -1 \\ 1.1 & -1 \end{pmatrix} \qquad \Delta = \begin{pmatrix} 0.1 & 0 \\ 0 & -0.1 \end{pmatrix}$$

$$\Rightarrow \quad G(s)\Delta G^{-1}(s) = \begin{pmatrix} -2.1 & 2 \\ -2.2 & 2.1 \end{pmatrix}$$

Example

Example Let

$$G(s) = \frac{1}{s}\left(\frac{-s+2}{s+2}\right)$$

Determine a controller U(s)=F(s)[R(s)-Y(s)] that achieves

$$G_c(s) = \frac{1}{1+sT}\left(\frac{-s+2}{s+2}\right)$$

and show that the closed-loop system is internally stable

The Sensitivity Functions

The sensitivity function (S):

- Quantifies disturbance attenuation due to feedback

The complementary sensitivity function (T)

- Equals the closed-loop system G_c with 1-DOF control
- Quantifies the amplification of noise at the output
- Determines robust stability properties

A first trade-off: S+T=1

Decoupling

- Full decoupling

$$D(s) = d(s)G^{-1}(s) \qquad \Rightarrow \qquad G(s)D(s) = d(s)I$$

- Static decoupling

$$D = G^{-1}(0)$$

The Relative Gain Array

Definition. the *relative gain array, RGA,* of a square system is defined as

$$\Lambda(G) = G \times (G^{-1})^T$$

or, in Matlab notation, RGA(G)= G.*inv(G).'

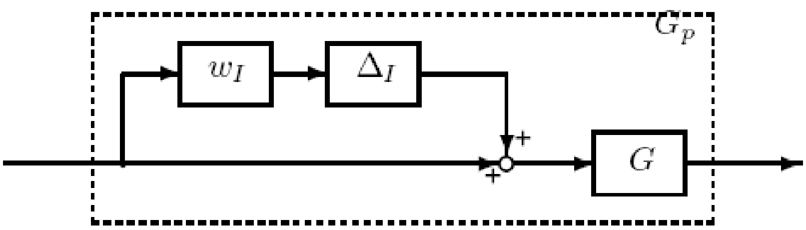
Rules of thumb:
1. Avoid pairings with $\lambda_{ij}(0) < 0$ (why?)
2. Prefer pairings with $|\lambda_{ij}(i\omega_c)| \approx 1$

Lecture 3 - Robustness

Robustness = property maintained under uncertainty

Idea: specify uncertainty set and guarantee stability and performance for all possible models within set

We have focused on multiplicative (relative) uncertainty



SISO Robustness

Robust stability via small-gain theorem

$$|T(i\omega)| \leq |w_I^{-1}(i\omega)| \qquad \forall\omega$$

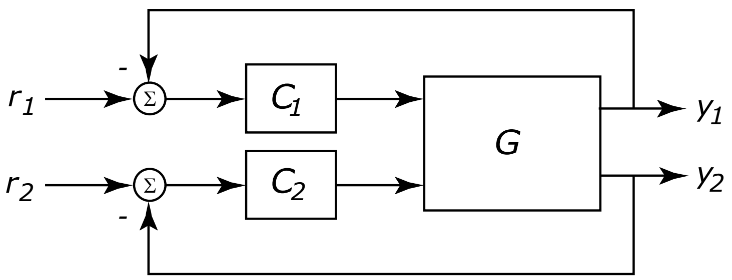
where

$$|w_I(i\omega)| \geq \left|\frac{G_p(i\omega) - G(i\omega)}{G(i\omega)}\right| \qquad \forall G_p \in \Pi_i$$

Robust performance puts simultaneous bound on S and T.

Decentralized Control

Decentralized control:



Interactions: each input affects multiple outputs

Qualitatively: the more interactions, the harder to control

- The relative gain array tries to quantify the degree of interactions

Question

Please repeat MIMO zeros and zero directions

- A zero z is a value of s where G(s) has less rank than normal.
  - Example:

$$G(s) = \frac{1}{s+1}\begin{pmatrix} s+1 & 1 \\ 3 & 1 \end{pmatrix} \quad \Rightarrow \quad G(2) = \frac{1}{3}\begin{pmatrix} 3 & 1 \\ 3 & 1 \end{pmatrix}$$

- When a matrix is rank deficient it has a left and right null space; the output zero direction y_z is the left nullspace and the input zero direction u_z is the right nullspace for the zero

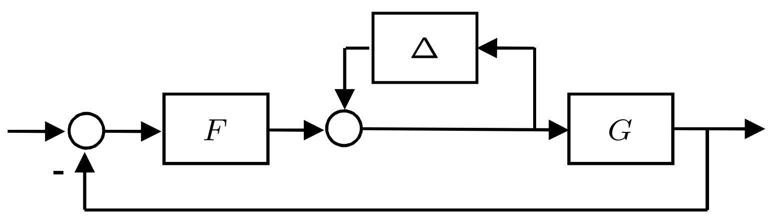
$$G(z)u_z = 0 \cdot y_z$$

  - for the example:

$$u_z^H = \begin{pmatrix} \frac{1}{3} & -1 \end{pmatrix} \quad y_z^H = \begin{pmatrix} -1 & 1 \end{pmatrix}$$

SISO Robustness

Example: Consider the uncertain system



Use the small-gain theorem to derive a robustness criterion

$$\|P\Delta\|_\infty \leq 1$$

for some transfer function P independent of Δ

Question

- Why is robustness important, because we can not operate at nominal operating where model is obtained?
  - partly, this usually gives rise to a model error
  - but, we always have uncertainty in actuators (inputs) and measurements, and model uncertainty even at nominal operating point
- How is uncertainty modeled in practice
  - from system identification, e.g., parametric uncertainty. See e.g., example in Lecture notes 3
  - from knowledge of uncertainty in actuators and sensors
  - by adding some generic uncertainty to achieve a reasonable robustness (similar to AM and PM in SISO control), e.g., robust loopshaping

Lecture 5 – Multivariable linear systems

Poles, zeros and gains.

Theorem. The *pole polynomial* of a system with transfer matrix G(s) is the least common denominator of all minors of G(s). The poles of G(s) are the roots of the pole polynomial.

Theorem. The *zero polynomial* of G(s) is the greatest common divisor of the maximal minors of G(s), normed so that they have the pole polynomial of G(s) as denominator. The *zeros* of G(s) are the roots of its zero polynomial.

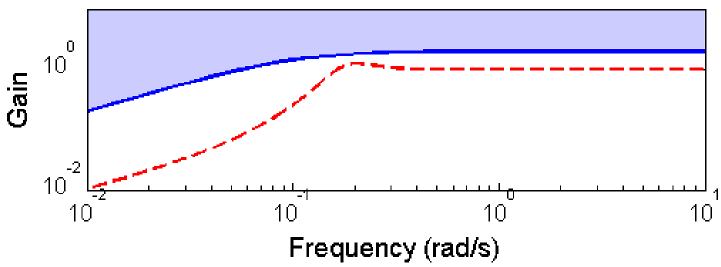
Theorem. The gain of a linear system G(s) is given by

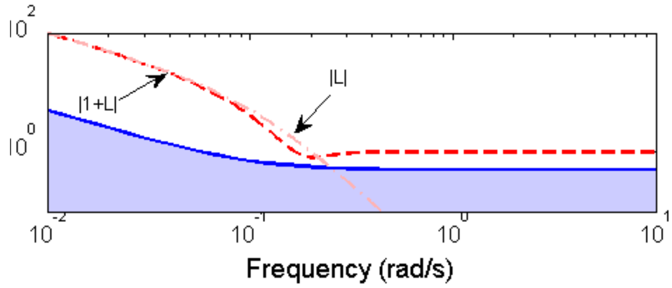
$$\|G\|_\infty = \sup_\omega |G(i\omega)| = \sup_\omega \overline{\sigma}(G(i\omega))$$

Classical Loop Shaping

In certain frequency ranges, there is a reasonable approximate mapping between constraints on S and T into requirements on the loop gain L.

$$|S(i\omega)| \leq |W_S^{-1}(i\omega)| \Leftrightarrow |1 + L(i\omega)| \geq |W_S(i\omega)|$$

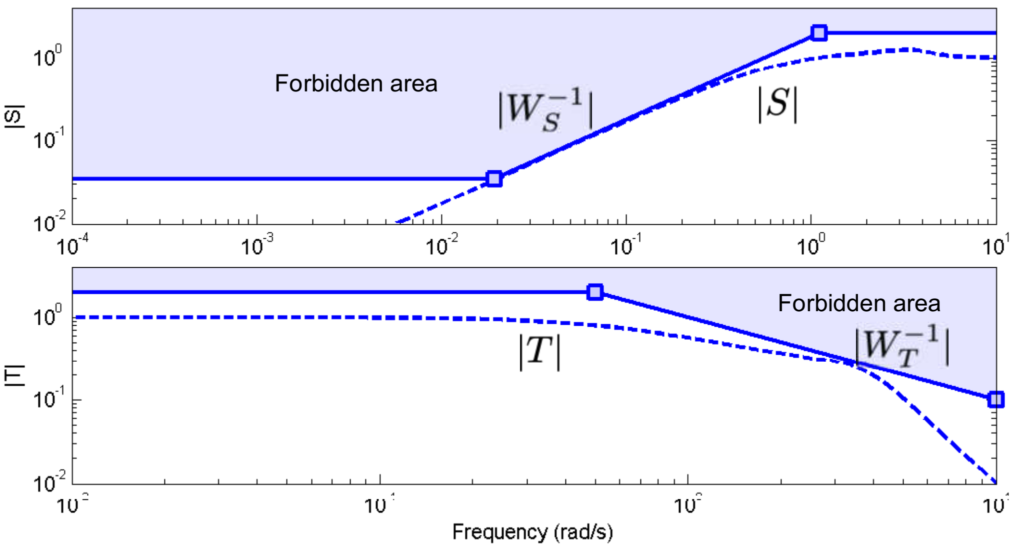




Problematic area is around cross-over frequency

- Put requirements on phase and amplitude margin

Lecture 4– Limitations and Conflicts



$$|S(i\omega)| \leq |W_S^{-1}(i\omega)| \ \forall \omega \quad \Leftrightarrow \quad \|W_S S\|_\infty < 1$$

$$|T(i\omega)| \leq |W_T^{-1}(i\omega)| \ \forall \omega \quad \Leftrightarrow \quad \|W_T T\|_\infty < 1$$

Can we choose weights w_S, w_T ("forbidden areas") freely?

– No, there are many constraints and limitations!

Limitations and conflicts

- Fundamental trade-offs in control systems design
  - S+T=1 (both cannot be small at the same frequency)
  - Cannot attenuate disturbances at all frequencies (Bode Sensitivity Integral)

- Fundamental limitations:
  - Unstable poles
  - Non-minimum phase zeros
  - Time delays

- Practical limitation:
  - Control input constraints

Rules of thumb

RHP zeros limit bandwidth (of S)

$$\boxed{\omega_{BS} \leq \frac{z}{2}}$$

Time-delays impose a similar bound

$$\boxed{\omega_{BS} \leq \frac{1}{T}}$$

RHP poles require a minimum bandwidth (for T)

$$\boxed{\omega_{BT} \geq 2p}$$