سایت اختصاصی
مهندسی کنترل

Design Methods for Control Systems

Okko H. Bosgra
Delft University of Technology
Delft, The Netherlands

Huibert Kwakernaak
University of Twente
Enschede, The Netherlands

Notes for a course of the
Dutch Institute of Systems and Control
Winter term 2000–2001

DISC

Preface

Part of these notes were developed for a course of the Dutch Network on Systems and Control with the title "Robust control and H_∞ optimization," which was taught in the Spring of 1991. These first notes were adapted and much expanded for a course with the title "Design Methods for Control Systems," first taught in the Spring of 1994. They were thoroughly revised for the Winter 1995–1996 course. For the Winter 1996–1997 course Chapter 4 was extensively revised and expanded, and a number of corrections and small additions were made to the other chapters. In the Winter 1997–1998 edition some material was added to Chapter 4 but otherwise there were minor changes only. The changes in the 1999–2000 version were limited to a number of minor corrections. In the 2000-2001 version an index and an appendix are added and Chapter 4 is revised.

The aim of the course is to present a mature overview of several important design techniques for linear control systems, varying from classical to "post-modern." The emphasis is on ideas, methodology, results, and strong and weak points, not on proof techniques.

All the numerical examples were prepared using MATLAB. For many examples and exercises the Control Toolbox is needed. For Chapter 6 the Robust Control Toolbox or the μ-Tools toolbox is indispensable.

Contents

1

Introduction to Feedback Control Theory

Overview *– Feedback is an essential element of automatic control systems. The primary requirements for feedback control systems are stability, performance and robustness.*

The design targets for linear time-invariant feedback systems may be phrased in terms of frequency response design goals and loop shaping. The design targets need to be consistent with the limits of performance imposed by physical realizability.

Extra degrees of freedom in the feedback system configuration introduce more flexibility.

1.1 Introduction

Designing a control system is a creative process involving a number of choices and decisions. These choices depend on the properties of the system that is to be controlled and on the requirements that are to be satisfied by the controlled system. The decisions imply compromises between conflicting requirements. The design of a control system involves the following steps:

1. Characterize the system boundary, that is, specify the scope of the control problem and of the system to be controlled.
2. Establish the type and the placement of actuators in the system, and thus specify the inputs that control the system.
3. Formulate a model for the dynamic behavior of the system, possibly including a description of its uncertainty.
4. Decide on the type and the placement of sensors in the system, and thus specify the variables that are available for feedforward or feedback.
5. Formulate a model for the disturbances and noise signals that affect the system.

6. Specify or choose the class of command signals that are to be followed by certain outputs.

7. Decide upon the functional structure and the character of the controller, also in dependence on its technical implementation.

8. Specify the desirable or required properties and qualities of the control system.

In several of these steps it is crucial to derive useful mathematical models of systems, signals and performance requirements. For the success of a control system design the depth of understanding of the dynamical properties of the system and the signals often is more important than the *a priori* qualifications of the particular design method.

The models of systems we consider are in general linear and time-invariant. Sometimes they are the result of physical modelling obtained by application of first principles and basic laws. On other occasions they follow from experimental or empirical modelling involving experimentation on a real plant or process, data gathering, and fitting models using methods for system identification.

Some of the steps may need to be performed repeatedly. The reason is that they involve design decisions whose consequences only become clear at later steps. It may then be necessary or useful to revise an earlier decision. Design thus is a process of gaining experience and developing understanding and expertise that leads to a proper balance between conflicting targets and requirements.

The functional specifications for control systems depend on the application. We distinguish different types of control systems:

Regulator systems. The primary function of a regulator system is to keep a designated output within tolerances at a predetermined value despite the effects of load changes and other disturbances.

Servo or ***positioning systems.*** In a servo system or positioning control system the system is designed to change the value of an output as commanded by a reference input signal, and in addition is required to act as a regulator system.

Tracking systems. In this case the reference signal is not predetermined but presents itself as a measured or observed signal to be tracked by an output.

Feedback is an essential element of automatic control. This is why § 1.2 presents an elementary survey of a number of basic issues in feedback control theory. These include *robustness*, *linearity* and *bandwidth improvement*, and *disturbance reduction*.

Stability is a primary requirement for automatic control systems. After recalling in § 1.3 various definitions of stability we review several well known ways of determining stability, including the Nyquist criterion.

In view of the importance of stability we elaborate in § 1.4 on the notion of stability robustness. First we recall several classical and more recent notions of stability margin. More refined results follow by using the Nyquist criterion to establish conditions for robust stability with respect to loop gain perturbations and inverse loop gain perturbations.

For single-input single-output feedback systems realizing the most important design targets may be viewed as a process of loop shaping of a one-degree-of-freedom feedback loop. The targets include

targets {
- closed-loop stability,
- disturbance attenuation,
- stability robustness,

within the limitations set by

limitations
- plant capacity,
- corruption by measurement noise.

Further design targets, which may require a two-degree-of-freedom configuration, are

further targets
- satisfactory closed-loop response,
- robustness of the closed-loop response.

Loop shaping and prefilter design are discussed in § 1.5. This section introduces various important closed-loop system functions such as the sensitivity function, the complementary sensitivity function, and the input sensitivity function.

Certain properties of the plant, in particular its pole-zero pattern, impose inherent restrictions on the closed-loop performance. In § 1.7 the limitations that right-half plane poles and zeros imply are reviewed. Ignoring these limitations may well lead to unrealistic design specifications. These results deserve more attention than they generally receive.

$1\frac{1}{2}$ and 2-degree-of-freedom feedback systems, designed for positioning and tracking, are discussed in Section 1.8.

1.2 Basic feedback theory

1.2.1 Introduction

In this section feedback theory is introduced at a low conceptual level[1]. It is shown how the simple idea of feedback has far-reaching technical implications.

Example 1.2.1 (Cruise control system). Figure 1.1 shows a block diagram of an automobile cruise control system, which is used to maintain the speed of a vehicle automatically at a constant level. The speed v of the car depends on the throttle opening u. The throttle opening is controlled by the cruise controller in such a way that the throttle opening is *increased* if the difference $v_r - v$ between the reference speed v_r and the actual speed is positive, and *decreased* if the difference is negative.

This feedback mechanism is meant to correct automatically any deviations of the actual vehicle speed from the desired cruise speed.

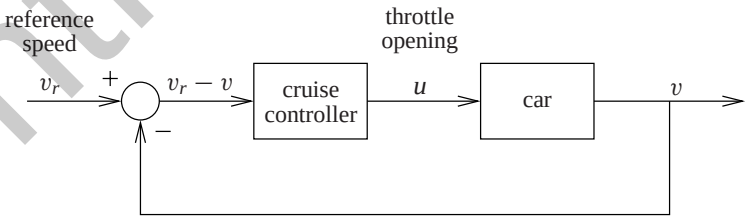


Figure 1.1: Block diagram of the cruise control system

For later use we set up a simple model of the cruising vehicle that accounts for the major physical effects. By Newton's law

$$m\dot{v}(t) = F_{\text{total}}(t), \qquad t \geq 0, \tag{1.1}$$

[1]This section has been adapted from Section 11.2 of Kwakernaak and Sivan (1991).

where m is the mass of the car, the derivative $\dot{v}$ of the speed v its acceleration, and F_{total} the total force exerted on the car in forward direction. The total force may be expressed as

$$F_{\text{total}}(t) = cu(t) - \rho v^2(t). \tag{1.2}$$

The first term $cu(t)$ represents the propulsion force of the engine, and is proportional to the throttle opening $u(t)$, with proportionality constant c. The throttle opening varies between 0 (shut) and 1 (fully open). The second term $\rho v^2(t)$ is caused by air resistance. The friction force is proportional to the square of the speed of the car, with ρ the friction coefficient. Substitution of F_{total} into Newton's law results in

$$m\dot{v}(t) = cu(t) - \rho v^2(t), \qquad t \geq 0. \tag{1.3}$$

If $u(t) = 1$, $t \geq 0$, then the speed has a corresponding steady-state value $v_{\max}$, which satisfies $0 = c - \rho v_{\max}^2$. Hence, $v_{\max} = \sqrt{c/\rho}$. Defining

$$w = \frac{v}{v_{\max}} \tag{1.4}$$

as the speed expressed as a fraction of the top speed, the differential equation reduces to

$$T\dot{w}(t) = u(t) - w^2(t), \qquad t \geq 0, \tag{1.5}$$

where $T = m/\sqrt{\rho c}$. A typical practical value for T is $T = 10$ [s].

We linearize the differential equation (1.5). To a constant throttle setting u_0 corresponds a steady-state cruise speed w_0 such that $0 = u_0 - w_0^2$. Let $u = u_0 + \tilde{u}$ and $w = w_0 + \tilde{w}$, with $|\tilde{w}| \ll w_0$. Substitution into (1.5) while neglecting second-order terms yields

$$T\dot{\tilde{w}}(t) = \tilde{u}(t) - 2w_0\tilde{w}(t). \tag{1.6}$$

Omitting the circumflexes we thus have the first-order linear differential equation

$$\dot{w} = -\frac{1}{\theta}w + \frac{1}{T}u, \qquad t \geq 0, \tag{1.7}$$

with

$$\theta = \frac{T}{2w_0}. \tag{1.8}$$

The time constant θ strongly depends on the operating conditions. If the cruise speed increases from 25% to 75% of the top speed then θ decreases from 20 [s] to 6.7 [s]. □

Exercise 1.2.2 (Acceleration curve). Show that the solution of the scaled differential equation (1.5) for a constant maximal throttle position

$$u(t) = 1, \qquad t \geq 0, \tag{1.9}$$

and initial condition $w(0) = 0$ is given by

$$w(t) = \tanh\left(\frac{t}{T}\right), \qquad t \geq 0. \tag{1.10}$$

Plot the scaled speed w as a function of t for $T = 10$ [s]. Is this a powerful car? □

1.2.2 Feedback configurations

To understand and analyze feedback we first consider the configuration of Fig. 1.2(a). The signal r is an external control input. The "plant" is a given system, whose output is to be controlled. Often the function of this part of the feedback system is to provide power, and its dynamical properties are not always favorable. The output y of the plant is fed back via the *return compensator* and subtracted from the external input r. The difference e is called the *error signal* and is fed to the plant via the *forward compensator.*

The system of Fig. 1.2(b), in which the return compensator is a unit gain, is said to have *unit feedback.*

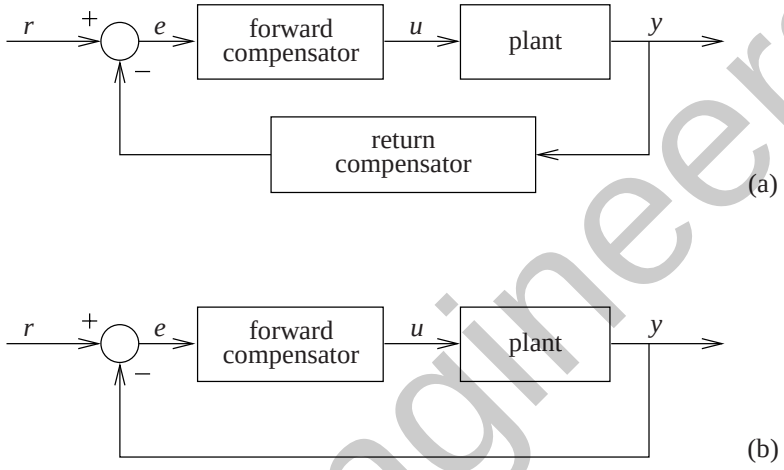


Figure 1.2: Feedback configurations: (a) General. (b) Unit feedback.

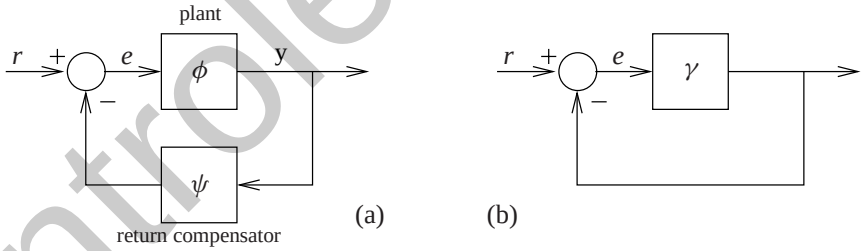


Figure 1.3: (a) Feedback configuration with input-output maps. (b) Equivalent unit feedback configuration.

Example 1.2.3 (Unit feedback system). The cruise control system of Fig. 1.1 is a unit feedback system. □

For the purposes of this subsection we reduce the configuration of Fig. 1.2(a) to that of Fig. 1.3(a), where the forward compensator has been absorbed into the plant. The plant is represented as an input-output-mapping system with input-output (IO) map ϕ, while the return compensator has the IO map ψ. The control input r, the error signal e and the output signal y usually all are time signals. Correspondingly, ϕ and ψ are IO maps of dynamical systems, mapping time signals to time signals.

The feedback system is represented by the equations

$$y = \phi(e), \qquad e = r - \psi(y). \tag{1.11}$$

These equations may or may not have a solution e and y for any given control input r. If a solution exists, the error signal e satisfies the equation $e = r - \psi(\phi(e))$, or

$$e + \gamma(e) = r. \tag{1.12}$$

Here $\gamma = \psi \circ \phi$, with $\circ$ denoting map composition, is the IO map of the series connection of the plant followed by the return compensator, and is called the *loop IO map*. Equation (1.12) reduces the feedback system to a unit feedback system as in Fig. 1.3(b). Note that because γ maps time functions into time functions, (1.12) is a *functional* equation for the time signal e. We refer to it as the *feedback equation*.

1.2.3 High-gain feedback

Feedback is most effective if the loop IO map γ has "large gain." We shall see that one of the important consequences of this is that the map from the external input r to the output y is approximately the inverse ψ^{-1} of the IO map ψ of the return compensator. Hence, the IO map from the control input r to the control system output y is almost independent of the plant IO map.

Suppose that for a given class of external input signals r the feedback equation

$$e + \gamma(e) = r \tag{1.13}$$

has a solution e. Suppose also that for this class of signals the "gain" of the map γ is large, that is,

$$\|\gamma(e)\| \gg \|e\|, \tag{1.14}$$

with $\|\cdot\|$ some norm on the signal space in which e is defined. This class of signals generally consists of signals that are limited in bandwidth and in amplitude. Then in (1.13) we may neglect the first term on the left, so that

$$\gamma(e) \approx r. \tag{1.15}$$

Since by assumption $\|e\| \ll \|\gamma(e)\|$ this implies that

$$\|e\| \ll \|r\|. \tag{1.16}$$

In words: If the gain is large then the error e is small compared with the control input r. Going back to the configuration of Fig. 1.3(a), we see that this implies that $\psi(y) \approx r$, or

$$y \approx \psi^{-1}(r), \tag{1.17}$$

where ψ^{-1} is the *inverse* of the map ψ (assuming that it exists).

Note that it is assumed that the feedback equation has a bounded solution[2] e for every bounded r. This is not necessarily always the case. If e is bounded for every bounded r then the closed-loop system by definition is BIBO stable[3]. Hence, the existence of solutions to the feedback equation is equivalent to the (BIBO) stability of the closed-loop system.

Note also that generally the gain may only be expected to be large for a *class* of error signals, denoted E. The class usually consists of band- and amplitude-limited signals, and depends on the "capacity" of the plant.

[2] A signal is bounded if its norm is finite. Norms of signals are discussed in § 5.4.4. See also § 1.3.

[3] A system is BIBO (bounded-input bounded-output) stable if every bounded input results in a bounded output (see § 1.3).

Example 1.2.4 (Proportional control of the cruise control system). A simple form of feedback that works reasonably well but not more than that for the cruise control system of Example 1.2.1 is *proportional feedback*. This means that the throttle opening is controlled according to

$$u(t) - u_0 = g[r(t) - w(t)], \tag{1.18}$$

with the *gain* g a constant and u_0 a nominal throttle setting. Denote w_0 as the steady-state cruising speed corresponding to the nominal throttle setting u_0, and write $w(t) = w_0 + \tilde{w}(t)$ as in Example 1.2.1. Setting $\tilde{r}(t) = r(t) - w_0$ we have

$$\tilde{u}(t) = g[\tilde{r}(t) - \tilde{w}(t)]. \tag{1.19}$$

Substituting this into the linearized equation (1.7) (once again omitting the circumflexes) we have

$$\dot{w} = -\frac{1}{\theta}w + \frac{g}{T}(r - w), \tag{1.20}$$

that is,

$$\dot{w} = -\left(\frac{1}{\theta} + \frac{g}{T}\right)w + \frac{g}{T}r. \tag{1.21}$$

Stability is ensured as long as

$$\frac{1}{\theta} + \frac{g}{T} > 0. \tag{1.22}$$

After Laplace transformation of (1.21) and solving for the Laplace transform of w we identify the closed-loop transfer function H_{cl} from

$$w = \underbrace{\frac{\frac{g}{T}}{s + \frac{1}{\theta} + \frac{g}{T}}}_{H_{\text{cl}}(s)} r. \tag{1.23}$$

We follow the custom of operational calculus not to distinguish between a time signal and its Laplace transform.

Figure 1.4 gives Bode magnitude plots of the closed-loop frequency response $H_{\text{cl}}(\text{j}\omega)$, $\omega \in \mathbb{R}$, for different values of the gain g. If the gain g is large then $H_{\text{cl}}(\text{j}\omega) \approx 1$ for low frequencies. The larger the gain, the larger the frequency region is over which this holds. □

1.2.4 Robustness of feedback systems

The approximate identity $y \approx \psi^{-1}(r)$ (1.17) *remains* valid as long as the feedback equation has a bounded solution e for every r and the gain is large. The IO map ψ of the return compensator may often be implemented with good accuracy. This results in a matching accuracy for the IO map of the feedback system as long as the gain is large, even if the IO map of the plant is poorly defined or has unfavorable properties. The fact that

$$y \approx \psi^{-1}(r) \tag{1.24}$$

in spite of uncertainty about the plant dynamics is called *robustness* of the feedback system with respect to plant uncertainty.

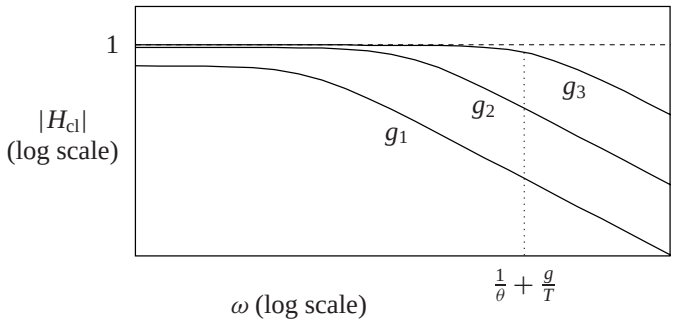


Figure 1.4: Magnitude plots of the closed-loop frequency response function for three values of the gain with $g_1 < g_2 < g_3$.

Example 1.2.5 (Cruise control system). The proportional cruise feedback control system of Example 1.2.4 is a first-order system, like the open-loop system. The closed-loop time constant θ_{cl} follows by inspection of (1.23) as

$$\frac{1}{\theta_{cl}} = \frac{1}{\theta} + \frac{g}{T}. \tag{1.25}$$

As long as $g \gg \frac{T}{\theta}$ the closed-loop time constant θ_{cl} approximately equals $\frac{T}{g}$. Hence, θ_{cl} does not depend much on the open-loop time constant θ, which is quite variable with the speed of the vehicle. For $g \gg \frac{T}{\theta}$ we have

$$H_{cl}(j\omega) \approx \frac{\frac{g}{T}}{j\omega + \frac{g}{T}} \approx 1 \qquad \text{for } |\omega| \leq \frac{g}{T}. \tag{1.26}$$

Hence, up to the frequency $\frac{g}{T}$ the closed-loop frequency response is very nearly equal to the unit gain. The frequency response of the open-loop system is

$$H(j\omega) = \frac{\frac{1}{T}}{j\omega + \frac{1}{\theta}} \approx \frac{\theta}{T} \quad \text{for } |\omega| < \frac{1}{\theta}. \tag{1.27}$$

The open-loop frequency response function obviously is much more sensitive to variations in the time constant θ than the closed-loop frequency response. □

1.2.5 Linearity and bandwidth improvement by feedback

Besides robustness, several other favorable effects may be achieved by feedback. They include linearity improvement, bandwidth improvement, and disturbance reduction.

Linearity improvement is a consequence of the fact that if the loop gain is large enough, the IO map of the feedback system approximately equals the inverse ψ^{-1} of the IO map of the return compensator. If this IO map is linear, so is the IO map of the feedback system, with good approximation, no matter how nonlinear the plant IO map ϕ is.

Also *bandwidth* improvement is a result of the high gain property. If the return compensator is a unit gain, the IO map of the feedback system is close to unity over those frequencies for which the feedback gain is large. This increases the bandwidth.

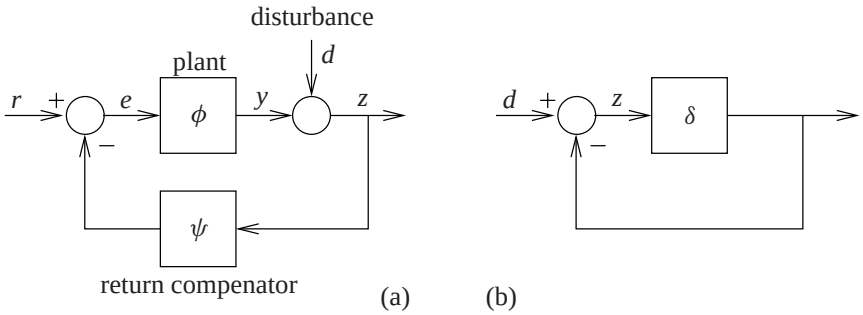


Figure 1.5: (a) Feedback system with disturbance. (b) Equivalent unit feedback configuration in the absence of the control input r.

Example 1.2.6 (Bandwidth improvement of the cruise control system). In Example 1.2.5 the time constant of the closed-loop proportional cruise control system is

$$\theta_{\text{cl}} = \frac{\theta}{1 + \frac{g\theta}{T}}. \tag{1.28}$$

For positive gain g the closed-loop time constant is smaller than the open-loop time constant θ and, hence, the closed-loop bandwidth is greater than the open-loop bandwidth. □

Exercise 1.2.7 (Steady-state linearity improvement of the proportional cruise control system). The dynamics of the vehicle are given by

$$T\dot{w} = u - w^2. \tag{1.29}$$

For a given steady-state solution (u_0, w_0), with $w_0 = \sqrt{u_0}$, consider the proportional feedback scheme

$$u - u_0 = g(r - w). \tag{1.30}$$

Calculate the steady-state dependence of $w - w_0$ on $r - w_0$ (assuming that r is constant). Plot this dependence for $w_0 = 0.5$ and $g = 10$.

To assess the linearity improvement by feedback compare this plot with a plot of $w - w_0$ versus $u - u_0$ for the open-loop system. Comment on the two plots. □

1.2.6 Disturbance reduction

A further useful property of feedback is that the effect of (external) *disturbances* is reduced. It frequently happens that in the configuration of Fig. 1.2(a) external disturbances affect the output y. These disturbances are usually caused by environmental effects.

The effect of disturbances may often be modeled by adding a *disturbance signal* d at the output of the plant as in Fig. 1.5(a). For simplicity we study the effect of the disturbance in the absence of any external control input, that is, we assume $r = 0$. The feedback system then is described by the equations $z = d + y$, $y = \phi(e)$, and $e = -\psi(z)$. Eliminating the output y and the error signal e we have $z = d + \phi(e) = d + \phi(-\psi(z))$, or

$$z = d - \delta(z), \tag{1.31}$$

where $\delta = (-\phi) \circ (-\psi)$. The map δ is also called a *loop IO map*, but it is obtained by "breaking the loop" at a different point compared with when constructing the loop IO map $\gamma = \psi \circ \phi$.

The equation (1.31) is a feedback equation for the configuration of Fig. 1.5(b). By analogy with the configuration of Fig. 1.3(b) it follows that if the gain is *large* in the sense that $\|\delta(z)\| \gg \|z\|$ then we have

$$\|z\| \ll \|d\|. \tag{1.32}$$

This means that the output z of the feedback system is small compared with the disturbance d, so that the effect of the disturbance is much reduced. All this holds provided the feedback equation (1.31) has at all a bounded solution z for any bounded d, that is, provided the closed-loop system is BIBO stable.

Example 1.2.8 (Disturbance reduction in the proportional cruise control system). The progress of the cruising vehicle of Example 1.2.1 may be affected by head or tail winds and up- or downhill grades. These effects may be represented by modifying the dynamical equation (1.3) to $m\dot{v} = cu - \rho v^2 + d$, with d the disturbing force. After scaling and linearization as in Example 1.2.1 this leads to the modification

$$\dot{w} = -\frac{1}{\theta} w - \frac{1}{T} u + d \tag{1.33}$$

of (1.7). Under the effect of the proportional feedback scheme (1.18) this results in the modification

$$\dot{w} = -\left(\frac{1}{\theta} + \frac{g}{T}\right) w + \frac{g}{T} r + d \tag{1.34}$$

of (1.21). Laplace transformation and solution for w (while setting $r = 0$) shows that the effect of the disturbance on the closed-loop system is represented by

$$w_{\mathrm{cl}} = \frac{1}{s + \frac{1}{\theta_{\mathrm{cl}}}} d. \tag{1.35}$$

From (1.33) we see that in the open-loop system the effect of the disturbance on the output is

$$w_{\mathrm{ol}} = \frac{1}{s + \frac{1}{\theta}} d. \tag{1.36}$$

This signal w_{ol} actually is the "equivalent disturbance at the output" of Fig. 1.6(a). Comparison of (1.34) and (1.35) shows that

$$w_{\mathrm{cl}} = \underbrace{\frac{s + \frac{1}{\theta}}{s + \frac{1}{\theta_{\mathrm{cl}}}}}_{S(s)} w_{\mathrm{ol}}. \tag{1.37}$$

S is known as the *sensitivity function* of the closed-loop system. Figure 1.6 shows the Bode magnitude plot of the frequency response function $S(\mathrm{j}\omega)$. The plot shows that the open-loop disturbances are attenuated by a factor

$$\frac{\theta_{\mathrm{cl}}}{\theta} = \frac{1}{1 + \frac{g\theta}{T}} \tag{1.38}$$

until the angular frequency $1/\theta$. After a gradual rise of the magnitude there is no attenuation or amplification of the disturbances for frequencies over $1/\theta_{\rm cl}$.

The disturbance attenuation is not satisfactory at very low frequencies. In particular, constant disturbances (that is, zero-frequency disturbances) are not completely eliminated because $S(0) \neq 0$. This means that a steady head wind or a long uphill grade slow the car down. In § 2.3 it is explained how this effect may be overcome by applying *integral control.* □

1.2.7 Pitfalls of feedback

As we have shown in this section, feedback may achieve very useful effects. It also has pitfalls:

1. Naïvely making the gain of the system large may easily result in an *unstable* feedback system. If the feedback system is unstable then the feedback equation has no bounded solutions and the beneficial effects of feedback are nonexistent.

2. Even if the feedback system is stable then high gain may result in overly large inputs to the plant, which the plant cannot absorb. The result is reduction of the gain and an associated loss of performance.

3. Feedback implies measuring the output by means of an output sensor. The associated *measurement errors* and *measurement noise* may cause loss of accuracy.

We return to these points in § 1.5.

1.3 Closed-loop stability

1.3.1 Introduction

In the remainder of this chapter we elaborate some of the ideas of Section 1.2 for linear time-invariant feedback systems. Most of the results are stated for single-input single-output (SISO) systems but from time to time also multi-input multi-output (MIMO) results are discussed.

We consider the *two-degree-of-freedom* configuration of Fig. 1.7. A MIMO or SISO plant with transfer matrix P is connected in feedback with a forward compensator with transfer matrix C. The function of the feedback loop is to provide stability, robustness, and disturbance attenuation. The feedback loop is connected in series with a prefilter with transfer matrix F. The function of the prefilter is to improve the closed-loop response to command inputs.

The configuration of Fig. 1.7 is said to have two degrees of freedom because both the compensator C and the prefilter F are free to be chosen by the designer. When the prefilter is replaced

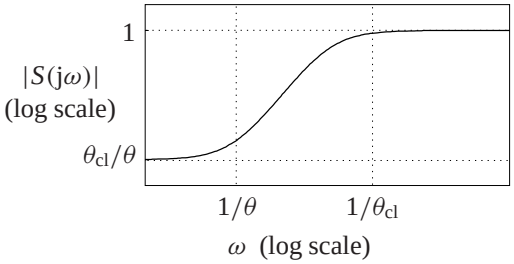


Figure 1.6: Magnitude plot of the sensitivity function of the proportional cruise control system

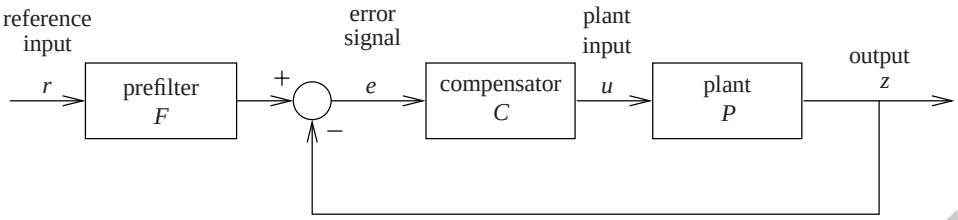


Figure 1.7: Two-degree-of-freedom feedback system configuration

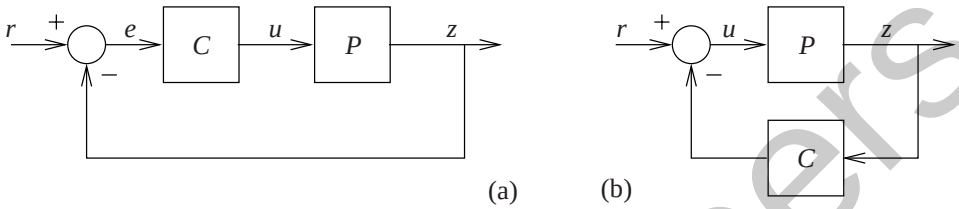


Figure 1.8: Single-degree-of-freedom feedback system configurations

with the unit system as in Fig. 1.8(a) the system is called a *single-degree-of-freedom* feedback system. Also the configuration of Fig. 1.8(b), with a return compensator instead of a forward compensator, is called a single-degree-of-freedom system.

In § 1.8 we consider alternative two-degree-of-freedom configurations, and study which is the most effective configuration.

1.3.2 Stability

In the rest of this section we discuss the stability of the closed-loop control system of Fig. 1.7. We assume that the overall system, including the plant, the compensator and the prefilter, has a state representation

$$\dot{x}(t) = Ax(t) + Br(t), \tag{1.39}$$

$$\begin{bmatrix} z(t) \\ u(t) \\ e(t) \end{bmatrix} = Cx(t) + Dr(t). \tag{1.40}$$

The command signal r is the external input to the overall system, while the control system output z, the plant input u and the error signal e jointly form the output. The signal x is the state of the overall system. A, B, C, and D are constant matrices of appropriate dimensions.

The state representation of the overall system is formed by combining the state space representations of the component systems. We assume that these state space representations include all the important dynamic aspects of the systems. They may be uncontrollable or unobservable. Besides the reference input r the external input to the overall system may include other exogenous signals such as disturbances and measurement noise.

Definition 1.3.1 (Stability of a closed-loop system). The feedback system of Fig. 1.7 (or any other control system) is *stable* if the state representation (1.39–1.40) is *asymptotically stable*, that is, if for a zero input and any initial state the state of the system asymptotically approaches the zero state as time increases. □

Given the state representation (1.39–1.40), the overall system is asymptotically stable if and only if all the eigenvalues of the matrix A have strictly negative real part.

There is another important form of stability.

Definition 1.3.2 (BIBO stability). The system of Fig. 1.7 is called *BIBO stable* (bounded-input-bounded-output stable) if every bounded input r results in bounded outputs z, u, and e for any initial condition on the state. □

To know what "bounded" means we need a norm for the input and output signals. A signal is said to be bounded if its norm is finite. We discuss the notion of the norm of a signal at some length in § 4.3.5. For the time being we say that a (vector-valued) signal $v(t)$ is bounded if there exists a constant M such that $|v_i(t)| \leq M$ for all t and for each component v_i of v.

Exercise 1.3.3 (Stability and BIBO stability).

1. Prove that if the closed-loop system is stable in the sense of Definition 1.3.1 then it is also BIBO stable.

2. Conversely, prove that if the system is BIBO stable and has no unstable unobservable modes[4] then it is stable in the sense of Definition 1.3.1.

3. Often BIBO stability is defined so that bounded input signals are required to result in bounded output signals for *zero* initial conditions of the state. With this definition, Part (1) of this exercise obviously still holds. Conversely, prove that if the system is BIBO stable in this sense and has no unstable unobservable and uncontrollable modes[5] then it is stable in the sense of Definition 1.3.1.

□

We introduce a further form of stability. It deals with the stability of interconnected systems, of which the various one- and two-degree-of-freedom feedback systems we encountered are examples. Stability in the sense of Definition 1.3.1 is independent of the presence or absence of inputs and outputs. BIBO stability, on the other hand, is strongly related to the presence and choice of input and output signals. *Internal* stability is BIBO stability but decoupled from a particular choice of inputs and outputs. We define the notion of internal stability of an interconnected systems in the following manner.

Definition 1.3.4 (Internal stability of an interconnected system). In each "exposed interconnection" of the interconnected system, inject an "internal" input signal v_i (with i an index), and define an additional "internal" output signal w_i just after the injection point. Then the system is said to be *internally stable* if the system whose input consists of the joint (external and internal) inputs and whose output is formed by the joint (external and internal) outputs is BIBO stable. □

To illustrate the definition of internal stability we consider the two-degree-of-freedom feedback configuration of Fig. 1.9. The system has the external input r, and the external output z. Identifying five exposed interconnections, we include five internal input-output signal pairs as shown in Fig. 1.10. The system is internally stable if the system with input (r, v_1, v_2, v_3, v_4, v_5) and output (z, w_1, w_2, w_3, w_4, w_5) is BIBO stable.

Exercise 1.3.5 (Stability and internal stability).

[4] A state system $\dot{x} = Ax + Bu$, $y = Cx + Du$ has an unobservable mode if the homogeneous equation $\dot{x} = Ax$ has a nontrivial solution x such that $Cx = 0$. The mode is unstable if this solution $x(t)$ does not approach 0 as $t \to \infty$.

[5] The state system $\dot{x} = Ax + Bu$, $y = Cx + Du$ has an uncontrollable mode if the state differential equation $\dot{x} = Ax + Bu$ has a solution x that is independent of u.

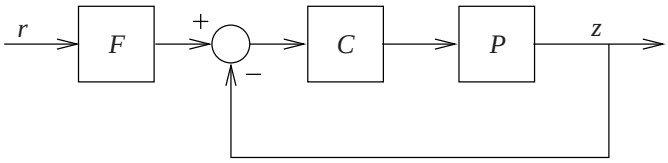


Figure 1.9: Two-degree-of-freedom feedback system

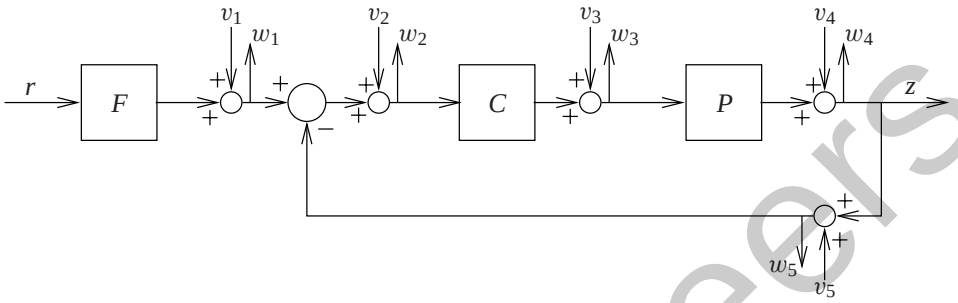


Figure 1.10: Two-degree-of-freedom system with internal inputs and outputs added

1. Prove that if the system of Fig. 1.9 is stable then it is internally stable.
2. Conversely, prove that if the system is internally stable and none of the component systems has any unstable unobservable modes then the system is stable in the sense of Definition 1.3.1. *Hint:* This follows from Exercise 1.3.3(b).

□

When using input-output descriptions, such as transfer functions, then internal stability is usually easier to check than stability in the sense of Definition 1.3.1. If no unstable unobservable and uncontrollable modes are present then internal stability is equivalent to stability in the sense of Definition 1.3.1.

1.3.3 Closed-loop characteristic polynomial

For later use we discuss the relation between the state and transfer functions representations of the closed-loop configuration of Fig. 1.9.

The characteristic polynomial χ of a system with state space representation

$$\dot{x}(t) = Ax(t) + Bu(t), \tag{1.41}$$

$$y(t) = Cx(t) + Du(t), \tag{1.42}$$

is the characteristic polynomial of its *system matrix* A,

$$\chi(s) = \det(sI - A). \tag{1.43}$$

The roots of the characteristic polynomial χ are the eigenvalues of the system. The system is stable if and only if the eigenvalues all have strictly negative real parts, that is, all lie in the open left-half complex plane.

The configuration of Fig. 1.9 consists of the series connection of the prefilter F with the feedback loop of Fig. 1.11(a).

Exercise 1.3.6 (Stability of a series connection). Consider two systems, one with state x and one with state z and let χ_1 and χ_2 denote their respective characteristic polynomials. Prove that the characteristic polynomial of the series connection of the two systems with state $\left[\begin{smallmatrix} x \\ z \end{smallmatrix}\right]$ is $\chi_1\chi_2$. From this it follows that the eigenvalues of the series connection consist of the eigenvalues of the first system together with the eigenvalues of the second system. □

We conclude that the configuration of Fig. 1.9 is stable if and only if both the prefilter and the feedback loop are stable. To study the stability of the MIMO or SISO feedback loop of

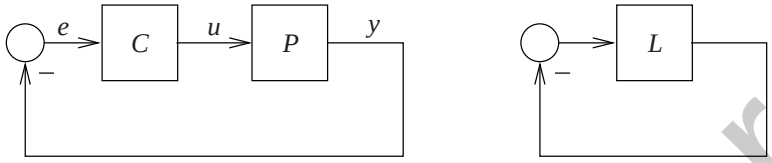


Figure 1.11: MIMO or SISO feedback systems

Fig. 1.11(a) represent it as in Fig. 1.11(b). $L = PC$ is the transfer matrix of the series connection of the compensator and the plant. L is called the *loop transfer matrix*. We assume that L is proper, that is, $L(\infty)$ is finite.

Suppose that the series connection L has the characteristic polynomial χ. We call χ the *open-loop characteristic polynomial*. It is proved in § 1.10 that the characteristic polynomial of the closed-loop system of Fig. 1.11 is

$$\chi_{\rm cl}(s) = \chi(s)\frac{\det[I + L(s)]}{\det[I + L(\infty)]}. \tag{1.44}$$

We call $\chi_{\rm cl}$ the *closed-loop characteristic polynomial*.

In the SISO case we may write the loop transfer function as

$$L(s) = \frac{R(s)}{Q(s)}, \tag{1.45}$$

with R and Q polynomials, where $Q = \chi$ is the open-loop characteristic polynomial. Note that we allow *no* cancellation between the numerator polynomial and the characteristic polynomial in the denominator. It follows from (1.44) that within the constant factor $1 + L(\infty)$ the closed-loop characteristic polynomial is

$$\chi(s)(1 + L(s)) = Q(s)[1 + \frac{R(s)}{Q(s)}] = Q(s) + R(s). \tag{1.46}$$

We return to the configuration of Fig. 1.11(a) and write the transfer functions of the plant and the compensator as

$$P(s) = \frac{N(s)}{D(s)}, \qquad C(s) = \frac{Y(s)}{X(s)}. \tag{1.47}$$

D and X are the open-loop characteristic polynomials of the plant and the compensator, respectively. N and Y are their numerator polynomials. Again we allow no cancellation between the numerator polynomials and the characteristic polynomials in the denominators. Since $R(s) = N(s)Y(s)$ and $Q(s) = D(s)X(s)$ we obtain from (1.46) the well-known result that within a constant factor the closed-loop characteristic polynomial of the configuration of Fig. 1.11(a) is

$$D(s)X(s) + N(s)Y(s). \tag{1.48}$$

With a slight abuse of terminology this polynomial is often referred to as the closed-loop characteristic polynomial. The actual characteristic polynomial is obtained by dividing (1.48) by its leading coefficient[6].

Exercise 1.3.7 (Hidden modes). Suppose that the polynomials N and D have a common polynomial factor. This factor corresponds to one or several unobservable or uncontrollable modes of the plant. Show that the closed-loop characteristic polynomial also contains this factor. Hence, the eigenvalues corresponding to unobservable or uncontrollable modes cannot be changed by feedback. In particular, any unstable uncontrollable or unobservable modes cannot be stabilized.

The same observation holds for any unobservable and uncontrollable poles of the compensator. □

The stability of a feedback system may be tested by calculating the roots of its characteristic polynomial. The system is stable if and only if each root has strictly negative real part. The *Routh-Hurwitz stability criterion*, which is reviewed in Section 5.2, allows to test for stability without explicitly computing the roots. A *necessary* but not sufficient condition for stability is that all the coefficients of the characteristic polynomial have the same sign. This condition is known as *Descartes' rule of signs.*

1.3.4 Pole assignment

The relation

$$\chi = DX + NY \tag{1.49}$$

for the characteristic polynomial (possibly within a constant) may be used for what is known as *pole assignment* or *pole placement*. If the plant numerator and denominator polynomials N and D are known, and χ is specified, then (1.49) may be considered as an equation in the unknown polynomials X and Y. This equation is known as the *Bézout* equation. If the polynomials N and D have a common nontrivial polynomial factor that is not a factor of χ then obviously no solution exists. Otherwise, a solution always exists.

The Bézout equation (1.49) may be solved by expanding the various polynomials as powers of the undeterminate variable and equate coefficients of like powers. This leads to a set of linear equations in the coefficients of the unknown polynomials X and Y, which may easily be solved. The equations are known as the *Sylvester* equations (Kailath 1980a).

To set up the Sylvester equations we need to know the degrees of the polynomials X and Y. Suppose that $P = N/D$ is *strictly proper*[7], with $\deg D = n$ and $\deg N < n$ given. We try to find a strictly proper compensator $C = Y/X$ with degrees $\deg X = m$ and $\deg Y = m - 1$ to be determined. The degree of $\chi = DX + NY$ is $n + m$, so that by equating coefficients of like powers we obtain $n + m + 1$ equations. Setting this number equal to the number $2m + 1$ of unknown coefficients of the polynomials Y and X it follows that $m = n$. Thus, we expect to solve the pole assignment problem with a compensator of the same order as the plant.

Example 1.3.8 (Pole assignment). Consider a second-order plant with transfer function

$$P(s) = \frac{1}{s^2}. \tag{1.50}$$

[6] That is, the coefficient of the highest-order term.

[7] A rational function or matrix P is *strictly proper* if $\lim_{|s|\to\infty} P(s) = 0$. A rational function P is strictly proper if and only if the degree of its numerator is less than the degree of its denominator.

Because the compensator is expected to have order two we need to assign four closed-loop poles. We aim at a dominant pole pair at $\frac{1}{2}\sqrt{2}(-1 \pm \mathrm{j})$ to obtain a closed-loop bandwidth of 1 [rad/s], and place a non-dominant pair at $5\sqrt{2}(-1 \pm \mathrm{j})$. Hence,

$$\begin{aligned} \chi(s) &= (s^2 + s\sqrt{2} + 1)(s^2 + 10\sqrt{2}s + 100) \\ &= s^4 + 11\sqrt{2}s^3 + 121s^2 + 110\sqrt{2}s + 100. \end{aligned} \tag{1.51}$$

Write $X(s) = x_2 s^2 + x_1 s + x_0$ and $Y(s) = y_1 s + y_0$. Then

$$\begin{aligned} D(s)X(s) + N(s)Y(s) &= s^2(x_2 s^2 + x_1 s + x_0) + (y_1 s + y_0) \\ &= x_2 s^4 + x_1 s^3 + x_0 s^2 + y_1 s + y_0. \end{aligned} \tag{1.52}$$

Comparing (1.51) and (1.52) the unknown coefficients follow by inspection, and we see that

$$X(s) = s^2 + 11\sqrt{2}s + 121, \tag{1.53}$$

$$Y(s) = 110\sqrt{2}s + 100. \tag{1.54}$$

□

Exercise 1.3.9 (Sylvester equations). More generally, suppose that

$$P(s) = \frac{b_{n-1}s^{n-1} + b_{n-2}s^{n-2} + \cdots + b_0}{a_n s^n + a_{n-1}s^{n-1} + \cdots + a_0}, \tag{1.55}$$

$$C(s) = \frac{y_{n-1}s^{n-1} + y_{n-2}s^{n-2} + \cdots + y_0}{x_n s^n + x_{n-1}s^{n-1} + \cdots + x_0}, \tag{1.56}$$

$$\chi(s) = \chi_{2n}s^{2n} + \chi_{2n-1}s^{2n-1} + \cdots + \chi_0. \tag{1.57}$$

Show that the equation $\chi = DX + NY$ may be arranged as

$$\underbrace{\begin{bmatrix} a_n & 0 & \cdots & \cdots & 0 \\ a_{n-1} & a_n & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_0 & a_1 & \cdots & \cdots & a_n \\ 0 & a_0 & a_1 & \cdots & a_{n-1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & \cdots & 0 & a_0 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x_n \\ x_{n-1} \\ \cdots \\ \cdots \\ \cdots \\ \cdots \\ x_0 \end{bmatrix}}_{\mathbf{x}} + \underbrace{\begin{bmatrix} 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \cdots & \cdots & \cdots & 0 \\ b_{n-1} & 0 & 0 & \cdots & 0 \\ b_{n-2} & b_{n-1} & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ b_0 & b_1 & \cdots & \cdots & b_{n-1} \\ 0 & b_0 & b_1 & \cdots & b_{n-2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & \cdots & 0 & b_0 \end{bmatrix}}_{B} \underbrace{\begin{bmatrix} y_{n-1} \\ y_{n-2} \\ \cdots \\ \cdots \\ \cdots \\ \cdots \\ y_0 \end{bmatrix}}_{\mathbf{y}} = \underbrace{\begin{bmatrix} \chi_{2n} \\ \chi_{2n-1} \\ \cdots \\ \cdots \\ \cdots \\ \cdots \\ \chi_0 \end{bmatrix}}_{\mathbf{c}} \tag{1.58}$$

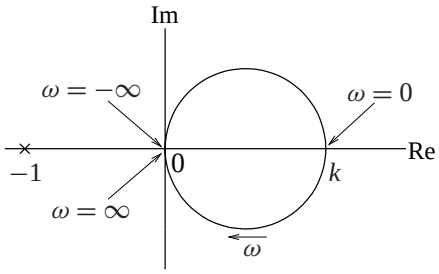


Figure 1.12: Nyquist plot of the loop gain transfer function $L(s) = k/(1 + s\theta)$

This in turn may be represented as

$$\begin{bmatrix} A & B \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} = \mathbf{c} \tag{1.59}$$

and solved for $\mathbf{x}$ and $\mathbf{y}$. If the polynomials D and N are coprime[8] then the square matrix $\begin{bmatrix} A & B \end{bmatrix}$ is nonsingular. □

1.3.5 Nyquist criterion

In classical control theory closed-loop stability is often studied with the help of the *Nyquist stability criterion*, which is a well-known graphical test. Consider the simple MIMO feedback loop of Fig. 1.11. The block marked "L" is the series connection of the compensator K and the plant P. The transfer matrix $L = PK$ is called the *loop gain matrix* — or *loop gain*, for short — of the feedback loop.

For a SISO system, L is a scalar function. Define the *Nyquist plot*[9] of the scalar loop gain L as the curve traced in the complex plane by

$$L(\mathrm{j}\omega), \qquad \omega \in \mathbb{R}. \tag{1.60}$$

Because for finite-dimensional systems L is a rational function with real coefficients, the Nyquist plot is symmetric with respect to the real axis. Associated with increasing ω we may define a positive direction along the locus. If L is *proper*[10] and has no poles on the imaginary axis then the locus is a closed curve. By way of example, Fig. 1.12 shows the Nyquist plot of the loop gain transfer function

$$L(s) = \frac{k}{1 + s\theta}, \tag{1.61}$$

with k and θ positive constants. This is the loop gain of the cruise control system of Example 1.2.5 with $k = g\theta/T$.

We first state the best known version of the Nyquist criterion.

Summary 1.3.10 (Nyquist stability criterion for SISO open-loop stable systems). Assume that in the feedback configuration of Fig. 1.11 the SISO system L is open-loop stable. Then the closed-loop system is stable if and only if the Nyquist plot of L does not encircle the point -1. □

[8] That is, they have no nontrivial common factors.

[9] The Nyquist plot is discussed at more length in § 2.4.3.

[10] A rational matrix function L is *proper* if $\lim_{|s|\to\infty} L(s)$ exists. For a rational function L this means that the degree of its numerator is not greater than that of its denominator.

It follows immediately from the Nyquist criterion and Fig. 1.12 that if $L(s) = k/(1 + s\theta)$ and the block "L" is stable then the closed-loop system is stable for all positive k and θ.

Exercise 1.3.11 (Nyquist plot). Verify the Nyquist plot of Fig. 1.12. □

Exercise 1.3.12 (Stability of compensated feedback system). Consider a SISO single-degree-of-freedom system as in Fig. 1.8(a) or (b), and define the loop gain $L = PK$. Prove that if both the compensator and the plant are stable and L satisfies the Nyquist criterion then the feedback system is stable. □

The result of Summary 1.3.10 is a special case of the *generalized Nyquist criterion*. The generalized Nyquist principle applies to a MIMO unit feedback system of the form of Fig. 1.11, and may be phrased as follows:

Summary 1.3.13 (Generalized Nyquist criterion). Suppose that the loop gain transfer function L of the MIMO feedback system of Fig. 1.11 is proper such that $I + L(j\infty)$ is nonsingular (this guarantees the feedback system to be well-defined) and has no poles on the imaginary axis. Assume also that the Nyquist plot of $\det(I + L)$ does not pass through the origin. Then

$$\begin{gathered}\text{the number of unstable closed-loop poles}\\ =\\ \text{the number of times the Nyquist plot of } \det(I+L) \text{ encircles the origin clockwise}\\ +\\ \text{the number of unstable open-loop poles.}\end{gathered}$$

It follows that the closed-loop system is stable if and only if the number of encirclements of $\det(I + L)$ equals the negative of the number of unstable open-loop poles. □

Similarly, the "unstable open-loop poles" are the right-half plane eigenvalues of the system matrix of the state space representation of the open-loop system. This includes any uncontrollable or unobservable eigenvalues. The "unstable closed-loop poles" similarly are the right-half plane eigenvalues of the system matrix of the closed-loop system.

In particular, it follows from the generalized Nyquist criterion that if the open-loop system is stable then the closed-loop system is stable if and only if the number of encirclements is zero (i.e., the Nyquist plot of $\det(I + L)$ does *not* encircle the origin).

For SISO systems the loop gain L is scalar, so that the number of times the Nyquist plot of $\det(I + L) = 1 + L$ encircles the origin equals the number of times the Nyquist plot of L encircles the point -1.

The condition that $\det(I + L)$ has no poles on the imaginary axis and does not pass through the origin may be relaxed, at the expense of making the analysis more complicated (see for instance Dorf (1992)).

The proof of the Nyquist criterion is given in § 1.10. More about Nyquist plots may be found in § 2.4.3.

1.3.6 Existence of a stable stabilizing compensator

A compensator that stabilizes the closed-loop system but by itself is unstable is difficult to handle in start-up, open-loop, input saturating or testing situations. There are unstable plants for which a stable stabilizing controller does not exist. The following result was formulated and proved by Youla, Bongiorno, and Lu (1974); see also Anderson and Jury (1976) and Blondel (1994).

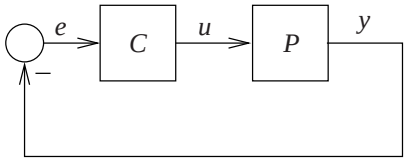


Figure 1.13: Feedback system configuration

Summary 1.3.14 (Existence of stable stabilizing controller). Consider the unit feedback system of Fig. 1.11(a) with plant P and compensator C.

The plant possesses the *parity interlacing property* if it has an even number of poles (counted according to multiplicity) between each pair of zeros on the positive real axis (including zeros at infinity.)

There exists a stable compensator C that makes the closed-loop stable if and only if the plant P has the parity interlacing property. □

If the denominator of the plant transfer function P has degree n and its numerator degree m then the plant has n poles and m (finite) zeros. If $m < n$ then the plant is said to have $n - m$ zeros at infinity.

Exercise 1.3.15 (Parity interlacing property). Check that the plant

$$P(s) = \frac{s}{(s-1)^2} \tag{1.62}$$

possesses the parity interlacing property while

$$P(s) = \frac{(s-1)(s-3)}{s(s-2)} \tag{1.63}$$

does not. Find a stabilizing compensator for each of these two plants (which for the first plant is itself stable.) □

1.4 Stability robustness

1.4.1 Introduction

In this section we consider SISO feedback systems with the configuration of Fig. 1.13. We discuss their *stability robustness*, that is, the property that the closed-loop system remains stable under changes of the plant and the compensator. This discussion focusses on the *loop gain* $L = PC$, with P the plant transfer function, and C the compensator transfer function. For simplicity we assume that the system is *open-loop stable*, that is, both P and C represent the transfer function of a stable system.

We also assume the existence of a *nominal* feedback loop with loop gain L_0, which is the loop gain that is supposed to be valid under nominal circumstances.

1.4.2 Stability margins

The closed-loop system of Fig. 1.13 remains stable under perturbations of the loop gain L as long as the Nyquist plot of the perturbed loop gain does not encircle the point -1. Intuitively, this

may be accomplished by "keeping the Nyquist plot of the nominal feedback system away from the point -1."

The classic *gain margin* and *phase margin* are well-known indicators for how closely the Nyquist plot approaches the point -1.

Gain margin The gain margin is the smallest positive number k_m by which the Nyquist plot must be multiplied so that it passes through the point -1. We have

$$k_m = \frac{1}{|L(\mathrm{j}\omega_r)|}, \tag{1.64}$$

where ω_r is the angular frequency for which the Nyquist plot intersects the negative real axis furthest from the origin (see Fig. 1.14).

Phase margin The phase margin is the extra phase ϕ_m that must be added to make the Nyquist plot pass through the point -1. The phase margin ϕ_m is the angle between the negative real axis and $L(\mathrm{j}\omega_m)$, where ω_m is the angular frequency where the Nyquist plot intersects the unit circle closest to the point -1 (see again Fig. 1.14).

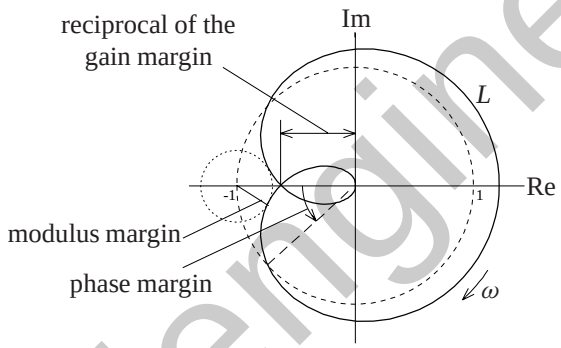


Figure 1.14: Robustness margins

In classical feedback system design, robustness is often specified by establishing minimum values for the gain and phase margin. Practical requirements are $k_m > 2$ for the gain margin and $30^\circ < \phi_m < 60^\circ$ for the phase margin.

The gain and phase margin do not necessarily adequately characterize the robustness. Figure 1.15 shows an example of a Nyquist plot with excellent gain and phase margins but where a relatively small *joint* perturbation of gain and phase suffices to destabilize the system. For this reason Landau, Rolland, Cyrot, and Voda (1993) introduced two more margins.

Modulus margin [11] The modulus margin s_m is the radius of the smallest circle with center -1 that is tangent to the Nyquist plot. Figure 1.14 illustrates this. The modulus margin very directly expresses how far the Nyquist plot stays away from -1.

Delay margin [12] The delay margin τ_m is the smallest extra delay that may be introduced in the loop that destabilizes the system. The delay margin is linked to the phase margin ϕ_m by the relation

$$\tau_m = \frac{\phi_m}{\omega_m}, \tag{1.65}$$

[11] French: *marge de module.*
[12] French: *marge de retard.*

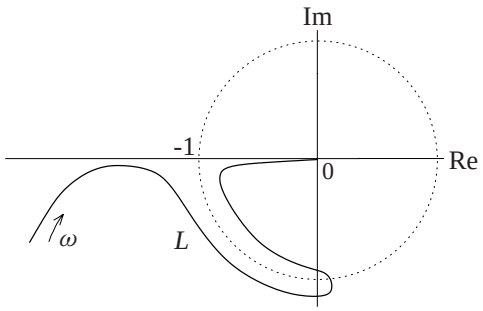


Figure 1.15: This Nyquist plot has good gain and phase margins but a small simultaneous perturbation of gain and phase destabilizes the system

with ω_m the frequency at which the Nyquist plot intersects the unit circle.

A practical specification for the modulus margin is $s_m > 0.5$. The delay margin should be at least of the order of $\frac{1}{2B}$, where B is the bandwidth (in terms of angular frequency) of the closed-loop system.

Adequate margins of these types are not only needed for robustness, but also to achieve a satisfactory time response of the closed-loop system. If the margins are small, the Nyquist plot approaches the point -1 closely. This means that the stability boundary is approached closely, manifesting itself by closed-loop poles that are very near to the imaginary axis. These closed-loop poles may cause an oscillatory response (called "ringing" if the resonance frequency is high and the damping small.)

Exercise 1.4.1 (Relation between robustness margins). Prove that the gain margin k_m and the phase margin ϕ_m are related to the modulus margin s_m by the inequalities

$$k_m \geq \frac{1}{1 - s_m}, \qquad \phi_m \geq 2 \arcsin \frac{s_m}{2}. \tag{1.66}$$

This means that if $s_m \geq \frac{1}{2}$ then $k_m \geq 2$ and $\phi_m \geq 2 \arcsin \frac{1}{4} \approx 28.96^\circ$ (Landau, Rolland, Cyrot, and Voda 1993). The converse is not true in general. □

1.4.3 Robustness for loop gain perturbations

The robustness specifications discussed so far are all rather qualitative. They break down when the system is not open-loop stable, and, even more spectacularly, for MIMO systems. We introduce a more refined measure of stability robustness by considering the effect of plant perturbations on the Nyquist plot more in detail. For the time being the assumptions that the feedback system is SISO and open-loop stable are upheld. Both are relaxed later.

Naturally, we suppose the nominal feedback system to be well-designed so that it is closed-loop stable. We investigate whether the feedback system *remains* stable when the loop gain is perturbed from the nominal loop gain L_0 to the actual loop gain L.

By the Nyquist criterion, the Nyquist plot of the nominal loop gain L_0 does not encircle the point -1, as shown in Fig. 1.16. The actual closed-loop system is stable if also the Nyquist plot of the actual loop gain L does not encircle -1.

It is easy to see by inspection of Fig. 1.16 that the Nyquist plot of L definitely does not encircle the point -1 if for all $\omega \in \mathbb{R}$ the distance $|L(\mathrm{j}\omega) - L_0(\mathrm{j}\omega)|$ between any point $L(\mathrm{j}\omega)$ and

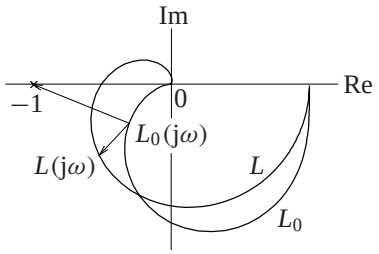


Figure 1.16: Nominal and perturbed Nyquist plots

the corresponding point $L_0(j\omega)$ is *less* than the distance $|L_0(j\omega)+1|$ of the point $L_0(j\omega)$ and the point -1, that is, if

$$|L(j\omega) - L_0(j\omega)| < |L_0(j\omega)+1| \quad \text{for all } \omega \in \mathbb{R}. \tag{1.67}$$

This is equivalent to

$$\frac{|L(j\omega) - L_0(j\omega)|}{|L_0(j\omega)|} \cdot \frac{|L_0(j\omega)|}{|L_0(j\omega)+1|} < 1 \quad \text{for all } \omega \in \mathbb{R}. \tag{1.68}$$

Define the *complementary sensitivity function* T_0 of the nominal closed-loop system as

$$T_0 = \frac{L_0}{1+L_0}. \tag{1.69}$$

T_0 bears its name because its complement

$$1 - T_0 = \frac{1}{1+L_0} = S_0 \tag{1.70}$$

is the *sensitivity function*. The sensitivity function plays an important role in assessing the effect of disturbances on the feedback system, and is discussed in Section 1.5.

Given T_0, it follows from (1.68) that if

$$\frac{|L(j\omega) - L_0(j\omega)|}{|L_0(j\omega)|} \cdot |T_0(j\omega)| < 1 \quad \text{for all } \omega \in \mathbb{R} \tag{1.71}$$

then the perturbed closed-loop system is stable.

The factor $|L(j\omega) - L_0(j\omega)|/|L_0(j\omega)|$ in this expression is the *relative* size of the perturbation of the loop gain L from its nominal value L_0. The relation (1.71) shows that the closed-loop system is guaranteed to be stable as long as the relative perturbations satisfy

$$\frac{|L(j\omega) - L_0(j\omega)|}{|L_0(j\omega)|} < \frac{1}{|T_0(j\omega)|} \quad \text{for all } \omega \in \mathbb{R}. \tag{1.72}$$

The larger the magnitude of the complementary sensitivity function is, the smaller is the allowable perturbation.

This result is discussed more extensively in Section 5.6, where also its MIMO version is described. It originates from Doyle (1979). The stability robustness condition has been obtained under the assumption that the open-loop system is stable. In fact, it also holds for open-loop unstable systems, *provided* the number of right-half plane poles remains invariant under perturbation.

Summary 1.4.2 (Doyle's stability robustness criterion). Suppose that the closed-loop system of Fig. 1.13 is nominally stable. Then it remains stable under perturbations that do not affect the number of open-loop unstable poles if

$$\frac{|L(j\omega) - L_0(j\omega)|}{|L_0(j\omega)|} < \frac{1}{|T_0(j\omega)|} \quad \text{for all } \omega \in \mathbb{R}, \tag{1.73}$$

with T_0 the nominal complementary sensitivity function of the closed-loop system. □

Exercise 1.4.3 (No poles may cross the imaginary axis). Use the general form of the Nyquist stability criterion of Summary 1.3.13 to prove the result of Summary 1.4.2. □

Doyle's stability robustness condition is a *sufficient* condition. This means that there may well exist perturbations that do not satisfy (1.73) but nevertheless do not destabilize the closed-loop system. This limits the applicability of the result. With a suitable modification the condition is also necessary, however. Suppose that the relative perturbations are known to bounded in the form

$$\frac{|L(j\omega) - L_0(j\omega)|}{|L_0(j\omega)|} \leq |W(j\omega)| \quad \text{for all } \omega \in \mathbb{R}, \tag{1.74}$$

with W a given function. Then the condition (1.73) is implied by the inequality

$$|T_0(j\omega)| < \frac{1}{|W(j\omega)|} \quad \text{for all } \omega \in \mathbb{R}. \tag{1.75}$$

Thus, if the latter condition holds, robust stability is guaranteed for all perturbations satisfying (1.74). Moreover, (1.75) is not only sufficient but also *necessary* to guarantee stability for *all* perturbations satisfying (1.74) (Vidysagar 1985). Such perturbations are said to "fill the uncertainty envelope."

Summary 1.4.4 (Stability robustness). Suppose that the closed-loop system of Fig. 1.13 is nominally stable. It remains stable under all perturbations that do not affect the number of open-loop unstable poles satisfying the bound

$$\frac{|L(j\omega) - L_0(j\omega)|}{|L_0(j\omega)|} \leq |W(j\omega)| \quad \text{for all } \omega \in \mathbb{R}, \tag{1.76}$$

with W a given function, if and only if

$$|T_0(j\omega)| < \frac{1}{|W(j\omega)|} \quad \text{for all } \omega \in \mathbb{R}. \tag{1.77}$$

□

Again, the MIMO version is presented in Section 5.6. The result is further discussed in § 1.5.

1.4.4 Inverse loop gain perturbations

According to the Nyquist criterion, the closed-loop system remains stable under perturbation as long as under perturbation the Nyquist plot of the loop gain does not cross the point -1. Equivalently, the closed-loop system remains stable under perturbation as long as the *inverse* $1/L$ of the loop gain does not cross the point -1. Thus, the sufficient condition (1.67) may be replaced with the sufficient condition

$$\left|\frac{1}{L(j\omega)} - \frac{1}{L_0(j\omega)}\right| < \left|\frac{1}{L_0(j\omega)} + 1\right| \quad \text{for all } \omega \in \mathbb{R}. \tag{1.78}$$

Dividing by the inverse $1/L_0$ of the nominal loop gain we find that a sufficient condition for robust stability is that

$$\left|\frac{\frac{1}{L(j\omega)}-\frac{1}{L_0(j\omega)}}{\frac{1}{L_0(j\omega)}}\right| < \left|\frac{\frac{1}{L_0(j\omega)}+1}{\frac{1}{L_0(j\omega)}}\right| = |1+L_0(j\omega)| = \frac{1}{|S_0(j\omega)|} \tag{1.79}$$

for all $\omega \in \mathbb{R}$. This in turn leads to the following conclusions.

Summary 1.4.5 (Inverse loop gain stability robustness criterion). Suppose that the closed-loop system of Fig. 1.13 is nominally stable. It remains stable under perturbations that do not affect the number of open-loop right-half plane zeros of the loop gain if

$$\left|\frac{\frac{1}{L(j\omega)}-\frac{1}{L_0(j\omega)}}{\frac{1}{L_0(j\omega)}}\right| < \frac{1}{|S_0(j\omega)|} \quad \text{for all } \omega \in \mathbb{R}, \tag{1.80}$$

with S_0 the nominal sensitivity function of the closed-loop system. □

Exercise 1.4.6 (Reversal of the role of the right half plane poles and the right-half plane zeros). Note that the role of the right-half plane poles has been taken by the right-half plane zeros. Explain this by deriving a stability condition based on the *inverse* Nyquist plot, that is, the polar plot of $1/L$. □

Again the result may be generalized to a sufficient and necessary condition.

Summary 1.4.7 (Stability robustness under inverse perturbation). Suppose that the closed-loop system of Fig. 1.13 is nominally stable. It remains stable under all perturbations that do not affect the number of right-half plane zeros satisfying the bound

$$\left|\frac{\frac{1}{L(j\omega)}-\frac{1}{L_0(j\omega)}}{\frac{1}{L_0(j\omega)}}\right| \le |W(j\omega)| \quad \text{for all } \omega \in \mathbb{R}, \tag{1.81}$$

with W a given function, if and only if

$$|S_0(j\omega)| < \frac{1}{|W(j\omega)|} \quad \text{for all } \omega \in \mathbb{R}. \tag{1.82}$$

□

Thus, for robustness *both* the sensitivity function S and its complement T are important. Later it is seen that for practical feedback design the complementary functions S and T need to be made small in complementary frequency regions (for low frequencies and for high frequencies, respectively).

We illustrate these results by an example.

Example 1.4.8 (Frequency dependent robustness bounds). Consider a SISO feedback loop with loop gain

$$L(s) = \frac{k}{1+s\theta}, \tag{1.83}$$

with k and θ positive constants. The nominal sensitivity and complementary sensitivity functions S_0 and T_0 are

$$S_0(s) = \frac{1}{1+L_0(s)} = \frac{1+s\theta_0}{1+k_0+s\theta_0}, \tag{1.84}$$

$$T_0(s) = \frac{L_0(s)}{1 + L_0(s)} = \frac{k_0}{1 + k_0 + s\theta_0}. \tag{1.85}$$

with k_0 and θ_0 the nominal values of k and θ, respectively. Figure 1.17 displays doubly logarithmic magnitude plots of $1/S_0$ and $1/T_0$.

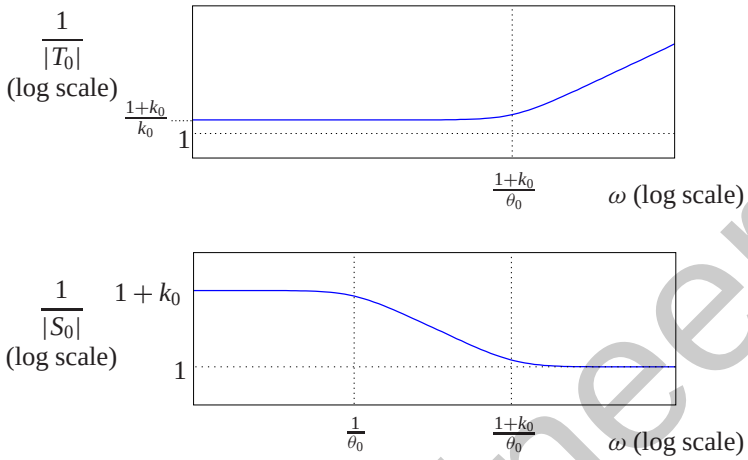


Figure 1.17: Magnitude plots of $1/T_0$ and $1/S_0$.

By the result of Summary 1.4.4 we conclude from the magnitude plot of $1/T_0$ that for low frequencies (up to the bandwidth $(k_0 + 1)/\theta_0$) relative perturbations of the loop gain L of relative size up to 1 and slightly larger are permitted while for higher frequencies increasingly larger perturbations are allowed without danger of destabilization of the closed-loop system.

By the result of Summary 1.4.7, on the other hand, we conclude from the magnitude plot of $1/S_0$, that for low frequencies (up to the frequency $1/\theta_0$) relative perturbations of the loop gain up to $1 + k_0$ are permitted. For high frequencies (greater than the bandwidth) the allowable relative size of the perturbations drops to the value 1. □

Exercise 1.4.9 (Landau's modulus margin and the sensitivity function).

1. In Subsection 1.4.2 the modulus margin s_m is defined as the distance from the point -1 to the Nyquist plot of the loop gain L:

$$s_m = \inf_{\omega \in \mathbb{R}} |1 + L(\mathrm{j}\omega)|. \tag{1.86}$$

   Prove that $1/s_m$ is the peak value of the magnitude of the sensitivity function S.

2. If the Nyquist plot of the loop gain L approaches the point -1 closely then so does that of the inverse loop gain $1/L$. Therefore, the number

$$r_m = \inf_{\omega \in \mathbb{R}} \left| 1 + \frac{1}{L(\mathrm{j}\omega)} \right| \tag{1.87}$$

   may also be viewed as a robustness margin. Prove that $1/r_m$ is the peak value of the magnitude of the complementary sensitivity function T.

□

1.5 Frequency response design goals

1.5.1 Introduction

In this section we translate the design targets for a linear time-invariant two-degree-of-freedom feedback system as in Fig. 1.18 into requirements on various closed-loop frequency response functions. The design goals are

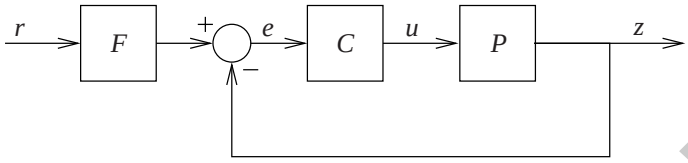


Figure 1.18: Two-degree-of-freedom feedback system

- closed-loop stability,
- disturbance attenuation,
- satisfactory closed-loop command response,
- stability robustness, and
- robustness of the closed-loop response,

within the limitations set by

- plant capacity, and
- corruption by measurement noise.

We discuss these aspects one by one for single-input-single-output feedback systems.

1.5.2 Closed-loop stability

Suppose that the feedback system of Fig. 1.13 is open-loop stable. By the Nyquist stability criterion, for closed-loop stability the loop gain should be shaped such that the Nyquist plot of L does not encircle the point -1.

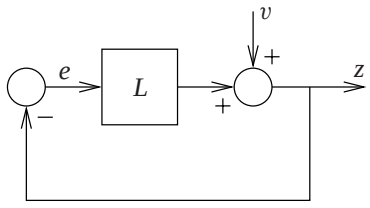


Figure 1.19: Feedback loop with disturbance

1.5.3 Disturbance attenuation and bandwidth — the sensitivity function

To study disturbance attenuation, consider the block diagram of Fig. 1.19, where v represents the equivalent disturbance at the output of the plant. In terms of Laplace transforms, the signal balance equation may be written as $z = v - Lz$. Solution for z results in

$$z = \underbrace{\frac{1}{1+L}}_{S} v = Sv, \tag{1.88}$$

where S is the sensitivity function of the closed-loop system. The smaller $|S(\mathrm{j}\omega)|$ is, with $\omega \in \mathbb{R}$, the more the disturbances are attenuated at the angular frequency ω. $|S|$ is small if the magnitude of the loop gain L is large. Hence, for disturbance attenuation it is necessary to shape the loop gain such that it is large over those frequencies where disturbance attenuation is needed.

Making the loop gain L large over a large frequency band easily results in error signals e and resulting plant inputs u that are larger than the plant can absorb. Therefore, L can only be made large over a limited frequency band. This is usually a low-pass band, that is, a band that ranges from frequency zero up to a maximal frequency B. The number B is called the *bandwidth* of the feedback loop. Effective disturbance attenuation is only achieved up to the frequency B.

The larger the "capacity" of the plant is, that is, the larger the inputs are the plant can handle before it saturates or otherwise fails, the larger the maximally achievable bandwidth usually is. For plants whose transfer functions have zeros with nonnegative real parts, however, the maximally achievable bandwidth is limited by the location of the right-half plane zero closest to the origin. This is discussed in Section 1.5.

Figure 1.20(a) shows an "ideal" shape of the magnitude of the sensitivity function. It is small for low frequencies and approaches the value 1 at high frequencies. Values greater than 1 and peaking are to be avoided. Peaking easily happens near the point where the curve crosses over the level 1 (the 0 dB line).

The desired shape for the sensitivity function S implies a matching shape for the magnitude of the complementary sensitivity function $T = 1 - S$. Figure 1.20(b) shows a possible shape[13] for the complementary sensitivity T corresponding to the sensivity function of Fig. 1.20(a). When S is as shown in Fig. 1.20(a) then T is close to 1 at low frequencies and decreases to 0 at high frequencies.

It may be necessary to impose further requirements on the shape of the sensitivity function if the disturbances have a distinct frequency profile. Consider for instance the situation that the actual disturbances enter the plant internally, or even at the plant input, and that the plant is highly oscillatory. Then the equivalent disturbance at the output is also oscillatory. To attenuate these disturbances effectively the sensitivity function should be small at and near the resonance frequency. For this reason it sometimes is useful to replace the requirement that S be small with the requirement that

$$|S(\mathrm{j}\omega)V(\mathrm{j}\omega)| \tag{1.89}$$

be small over a suitable low frequency range. The shape of the weighting function V reflects the frequency contents of the disturbances. If the actual disturbances enter the system at the plant input then a possible choice is to let $V = P$.

[13] Note that S and T are complementary, not $|S|$ and $|T|$.

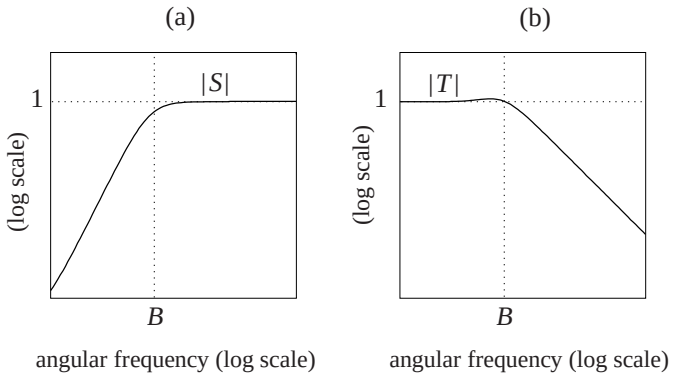


Figure 1.20: (a) "Ideal" sensitivity function. (b) A corresponding complementary sensitivity function.

Exercise 1.5.1 (Plant input disturbance for oscillatory plant). Consider an oscillatory second-order plant with transfer function

$$P(s) = \frac{\omega_0^2}{s^2 + 2\zeta_0\omega_0 s + \omega_0^2}. \tag{1.90}$$

Choose the compensator

$$C(s) = \frac{k}{\omega_0^2}\frac{s^2 + 2\zeta_0\omega_0 s + \omega_0^2}{s(s+\alpha)}. \tag{1.91}$$

Show that the sensitivity function S of the closed-loop system is independent of the resonance frequency ω_0 and the relative damping ζ_0. Select k and α such that a well-behaved high-pass sensitivity function is obtained.

Next, select the resonance frequency ω_0 well within the closed-loop bandwidth and take the relative damping ζ_0 small so that the plant is quite oscillatory. Demonstrate by simulation that the closed-loop response to disturbances at the plant input reflects this oscillatory nature even though the closed-loop sensitivity function is quite well behaved. Show that this oscillatory behavior also appears in the response of the closed-loop system to a nonzero initial condition of the plant. □

1.5.4 Command response — the complementary sensitivity function

The response of the two-degree-of-freedom configuration of Fig. 1.21 to the command signal r follows from the signal balance equation $z = PC(-z + Fr)$. Solution for z results in

$$z = \underbrace{\frac{PC}{1+PC}}_{H} F\ r. \tag{1.92}$$

The *closed-loop transfer function* H may be expressed as

$$H = \underbrace{\frac{L}{1+L}}_{T} F = TF, \tag{1.93}$$

with $L = PC$ the loop gain and T the complementary sensitivity function.

Adequate loop shaping ideally results in a complementary sensitivity function T that is close to 1 up to the bandwidth, and transits smoothly to zero above this frequency. Thus, without a prefilter F (that is, with $F = 1$), the closed-loop transfer function H ideally is low-pass with the same bandwidth as the frequency band for disturbance attenuation.

Like for the sensitivity function, the plant dynamics impose limitations on the shape that T may assume. In particular, right-half plane plant poles constrain the frequency above which T may be made to roll off. This is discussed in Section 1.5.

If the shape and bandwidth of T are satisfactory then no prefilter F is needed. If the closed-loop bandwidth is *greater* than necessary for adequate command signal response then the prefilter F may be used to reduce the bandwidth of the closed-loop transfer function H to prevent overly large plant inputs. If the bandwidth is *less* than required for good command signal response the prefilter may be used to compensate for this. A better solution may be to increase the closed-loop bandwidth. If this is not possible then probably the plant capacity is too small.

1.5.5 Plant capacity — the input sensitivity function

Any physical plant has limited "capacity," that is, can absorb inputs of limited magnitude only. The configuration of Fig. 1.21 includes both the disturbances v and the measurement noise m. In terms of Laplace transforms we have the signal balance $u = C(Fr - m - v - Pu)$. This may be solved for u as

$$u = \underbrace{\frac{C}{I + CP}}_{M} (Fr - m - v). \tag{1.94}$$

The function M determines the sensitivity of the plant input to disturbances and the command signal. It is sometimes known as the *input sensitivity function*.

If the loop gain $L = CP$ is large then the input sensitivity M approximately equals the *inverse* $1/P$ of the plant transfer function. If the open-loop plant has zeros in the right-half complex plane then $1/P$ is unstable. For this reason the right-half plane open-loop plant zeros limit the closed-loop bandwidth. The input sensitivity function M may only be made equal to $1/P$ up to the frequency which equals the magnitude of the right-half plane plant zero with the smallest magnitude.

The input sensitivity function M is connected to the complementary sensitivity function T by the relation

$$T = MP. \tag{1.95}$$

By this connection, for a fixed plant transfer function P design requirements on the input sensitivity function M may be translated into corresponding requirements on the complementary sensitivity T, and vice-versa.

To prevent overly large inputs, generally M should not be too large. At low frequencies a high loop gain and correspondingly large values of M are prerequisites for low sensitivity. If these large values are not acceptable, the plant capacity is inadequate, and either the plant needs to be replaced with a more powerful one, or the specifications need to be relaxed. At high frequencies — that is, at frequencies above the bandwidth — M should decrease as fast as possible. This is consistent with the robustness requirement that T decrease fast.

Except by Horowitz (Horowitz 1963) the term "plant capacity" does not appear to be used widely in the control literature. Nevertheless it is an important notion. The maximum bandwidth

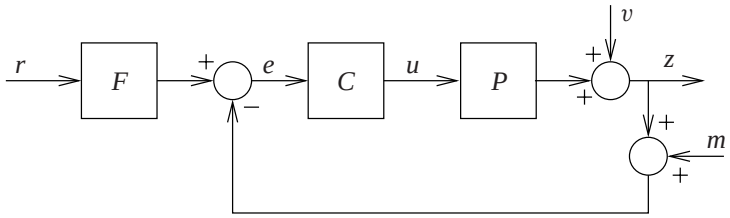


Figure 1.21: Two-degree-of-freedom system with disturbances and measurement noise

determined by the plant capacity may roughly be estimated as follows. Consider the response of the plant to the step input $a\,\mathbb{1}(t)$, $t \in \mathbb{R}$, with a the largest amplitude the plant can handle before it saturates or otherwise fails, and $\mathbb{1}$ the unit step function. Let θ be half the time needed until the output either reaches

1. 85% of its steady-state value, or
2. 85% of the largest feasible output amplitude,

whichever is less. Then the angular frequency $1/\theta$ may be taken as an indication of the largest possible bandwidth of the system.

This rule of thumb is based on the observation that the first-order step response $1 - \mathrm{e}^{-t/\theta}$, $t \geq 0$, reaches the value 0.865 at time 2θ.

Exercise 1.5.2 (Input sensitivity function for oscillatory plant). Consider a stabilizing feedback compensator of the form (1.91) for the plant (1.90), with k and α selected as suggested in Exercise 1.5.1. Compute and plot the resulting input sensitivity function M and discuss its behavior. □

1.5.6 Measurement noise

To study the effect of measurement noise on the closed-loop output we again consider the configuration of Fig. 1.21. By solving the signal balance $z = v + PC(Fr - m - z)$ for the output z we find

$$z = \underbrace{\frac{1}{1+PC}}_{S} v + \underbrace{\frac{PC}{1+PC}}_{T} Fr - \underbrace{\frac{PC}{1+PC}}_{T} m. \tag{1.96}$$

This shows that the influence of the measurement noise m on the control system output is determined by the complementary sensitivity function T. For low frequencies, where by the other design requirements T is close to 1, the measurement noise fully affects the output. This emphasizes the need for good, low-noise sensors.

1.5.7 Stability robustness

In Section 1.4 it is seen that for stability robustness it is necessary to keep the Nyquist plot "away from the point -1." The target is to achieve satisfactory gain, phase, and modulus margins.

Alternatively, as also seen in Section 1.4, robustness for loop gain perturbations requires the complementary sensitivity function T to be small. For robustness for inverse loop gain perturbations, on the other hand, the sensitivity function S needs to be small.

By complementarity T and S cannot be simultaneously small, at least not *very* small. The solution is to make T and S small in different frequency ranges. It is consistent with the other design targets to have the sensitivity S small in the *low frequency* range, and T small in the complementary high frequency range.

The faster T decreases with frequency — this is called *roll-off* — the more protection the closed-loop system has against high-frequency loop perturbations. This is important because owing to neglected dynamics — also known as *parasitic effects* — high frequency uncertainty is ever-present.

Small values of the sensitivity function for low frequencies, which are required for adequate disturbance attenuation, ensure protection against perturbations of the inverse loop gain at low frequencies. Such perturbations are often caused by load variations and environmental changes.

In the *crossover region* neither S nor T can be small. The crossover region is the frequency region where the loop gain L crosses the value 1 (the zero dB line.) It is the region that is most critical for robustness. Peaking of S and T in this frequency region is to be avoided. Good gain, phase and modulus margins help to ensure this.

1.5.8 Performance robustness

Feedback system performance is determined by the sensitivity function S, the complementary sensitivity function T, the input sensitivity function M, and the closed-loop transfer function H, successively given by

$$S = \frac{1}{1+L}, \quad T = \frac{L}{1+L}, \tag{1.97}$$

$$M = \frac{C}{1+L} = SC, \quad H = \frac{L}{1+L}F = TF. \tag{1.98}$$

We consider the extent to which each of these functions is affected by plant variations. For simplicity we suppose that the system environment is sufficiently controlled so that the compensator transfer function C and the prefilter transfer function F are not subject to perturbation. Inspection of (1.97–1.98) shows that under this assumption we only need to study the effect of perturbations on S and T: The variations in M are proportional to those in S, and the variations in H are proportional to those in T.

Denote by L_0 the *nominal* loop gain, that is, the loop gain that is believed to be representative and is used in the design calculations. Correspondingly, S_0 and T_0 are the nominal sensitivity function and complementary sensitivity function.

It is not difficult to establish that when the loop gain changes from its nominal value L_0 to its actual value L the corresponding *relative change* of the reciprocal of the sensitivity function S may be expressed as

$$\frac{\frac{1}{S} - \frac{1}{S_0}}{\frac{1}{S_0}} = \frac{S_0 - S}{S} = T_0 \frac{L - L_0}{L_0}. \tag{1.99}$$

Similarly, the relative change of the reciprocal of the complementary sensitivity function may be written as

$$\frac{\frac{1}{T} - \frac{1}{T_0}}{\frac{1}{T_0}} = \frac{T_0 - T}{T} = S_0 \frac{L_0 - L}{L} = S_0 \frac{\frac{1}{L} - \frac{1}{L_0}}{\frac{1}{L_0}}. \tag{1.100}$$

These relations show that for the sensitivity function S to be robust with respect to changes in the loop gain we desire the nominal complementary sensitivity function T_0 to be small. On the other

hand, for the complementary sensitivity function T to be robust we wish the nominal sensitivity function S_0 to be small. These requirements are conflicting, because S_0 and T_0 add up to 1 and therefore cannot simultaneously be small.

The solution is again to have each small in a different frequency range. As seen before, normal control system design specifications require S_0 to be small at low frequencies (below the bandwidth). This causes T to be robust at low frequencies, which is precisely the region where its values are significant. Complementarily, T_0 is required to be small at high frequencies, causing S to be robust in the high frequency range.

Exercise 1.5.3 (Formulas for relative changes). Prove (1.99–1.100). □

Exercise 1.5.4 (MIMO systems). Suppose that the configuration of Fig. 1.21 represents a MIMO rather than a SISO feedback system. Show that the various closed-loop system functions encountered in this section generalize to the following matrix system functions:

- the *sensitivity matrix* $S = (I + L)^{-1}$, with $L = PC$ the *loop gain matrix* and I an identity matrix of suitable dimensions;
- the *complementary sensitivity matrix* $T = I - S = L(I + L)^{-1} = (I + L)^{-1}L$;
- the *input sensitivity matrix* $M = (I + CP)^{-1}C = C(I + PC)^{-1}$;
- the *closed-loop transfer matrix* $H = (I + L)^{-1}LF = TF$.

□

1.5.9 Review of the design requirements

We summarize the conclusions of this section as follows:

- The sensitivity S should be small at low frequencies to achieve
  - disturbance attenuation,
  - good command response, and
  - robustness at low frequencies.
- The complementary sensitivity T should be small at high frequencies to prevent
  - exceeding the plant capacity,
  - adverse effects of measurement noise, and
  - loss of robustness at high frequencies.
- In the intermediate (crossover) frequency region peaking of both S and T should be avoided to prevent
  - overly large sensitivity to disturbances,
  - excessive influence of the measurement noise, and
  - loss of robustness.

1.6 Loop shaping

1.6.1 Introduction

The design of a feedback control system may be viewed as a process of *loop shaping*. The problem is to determine the feedback compensator C as in Fig. 1.18 such that the loop gain frequency response function $L(\mathrm{j}\omega)$, $\omega \in \mathbb{R}$, has a suitable shape. The design goals of the previous section may be summarized as follows.

Low frequencies. At low frequencies we need S small and T close to 1. Inspection of

$$S = \frac{1}{1+L}, \qquad T = \frac{L}{1+L} \tag{1.101}$$

shows that these targets may be achieved simultaneously by making the loop gain L large, that is, by making $|L(\mathrm{j}\omega)| \gg 1$ in the low-frequency region.

High frequencies. At high frequencies we need T small and S close to 1. This may be accomplished by making the loop gain L small, that is, by making $|L(\mathrm{j}\omega)| \ll 1$ in the high-frequency region.

Figure 1.22 shows how specifications on the magnitude of S in the low-frequency region and that of T in the high-frequency region result in bounds on the loop gain L.

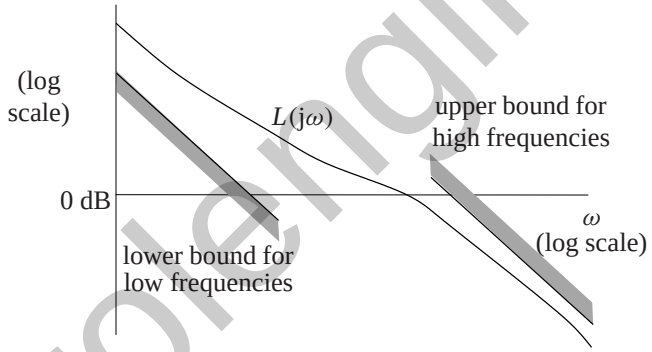


Figure 1.22: Robustness bounds on L in the Bode magnitude plot

Crossover region. In the crossover region we have $|L(\mathrm{j}\omega)| \approx 1$. In this frequency region it is not sufficient to consider the behavior of the magnitude of L alone. The behavior of the magnitude and phase of L in this frequency region together determine how closely the Nyquist plot approaches the critical point -1.

The more closely the Nyquist plot of L approaches the point -1 the more

$$S = \frac{1}{1+L} \tag{1.102}$$

peaks. If the Nyquist plot of L comes very near the point -1, so does the inverse Nyquist plot, that is, the Nyquist plot of $1/L$. Hence, the more closely the Nyquist plot of L approaches -1 the more

$$T = \frac{L}{1+L} = \frac{1}{1+\frac{1}{L}} \tag{1.103}$$

peaks.

Gain and phase of the loop gain are not independent. This is clarified in the next subsection.

1.6.2 Relations between phase and gain

A well-known and classical result of Bode (1940) is that the magnitude $|L(\mathrm{j}\omega)|$, $\omega \in \mathbb{R}$, and the phase $\arg L(\mathrm{j}\omega)$, $\omega \in \mathbb{R}$, of a minimum phase[14] linear time-invariant system with real-rational[15] transfer function L are uniquely related. If on a log-log scale the plot of the magnitude $|L|$ versus ω has an approximately constant slope n [decade/decade] then

$$\arg L(\mathrm{j}\omega) \approx n \times \frac{\pi}{2}. \tag{1.104}$$

Thus, if $|L(\mathrm{j}\omega)|$ behaves like $1/\omega$ then we have a phase of approximately $-\pi/2$ [rad] $= -90^\circ$, while if $|L(\mathrm{j}\omega)|$ behaves like $1/\omega^2$ then the phase is approximately $-\pi$ [rad] $= -180^\circ$.

Exercise 1.6.1 (Bode's gain-phase relationship). Why (1.104) holds may be understood from the way asymptotic Bode magnitude and phase plots are constructed (see § 2.4.2). Make it plausible that between any two break points of the Bode plot the loop gain behaves as

$$L(\mathrm{j}\omega) \approx c(\mathrm{j}\omega)^n, \tag{1.105}$$

with c a real constant. Show that n follows from the numbers of poles and zeros of L whose magnitude is less than the frequency corresponding to the lower break point. What is c? □

More precisely Bode's gain-phase relationship may be phrased as follows (Bode 1945).

Summary 1.6.2 (Bode's gain-phase relationship). Let L be a minimum phase real-rational proper transfer function. Then magnitude and phase of the corresponding frequency response function are related by

$$\arg L(\mathrm{j}\omega_0) = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{d \log |L_{\omega_0}(\mathrm{j}u)|}{du}\, W(u)\, du, \qquad \omega_0 \in \mathbb{R}, \tag{1.106}$$

with log denoting the natural logarithm. The intermediate variable u is defined by

$$u = \log \frac{\omega}{\omega_0}. \tag{1.107}$$

W is the function

$$W(u) = \log \coth \frac{|u|}{2} = \log \left| \frac{\frac{\omega}{\omega_0} + 1}{\frac{\omega}{\omega_0} - 1} \right|. \tag{1.108}$$

L_{ω_0}, finally, is given by

$$L_{\omega_0}(\mathrm{j}u) = L(\mathrm{j}\omega), \qquad u = \log \frac{\omega}{\omega_0}. \tag{1.109}$$

L_{ω_0} is the frequency response function L defined on the logaritmically transformed and scaled frequency axis $u = \log \omega/\omega_0$. □

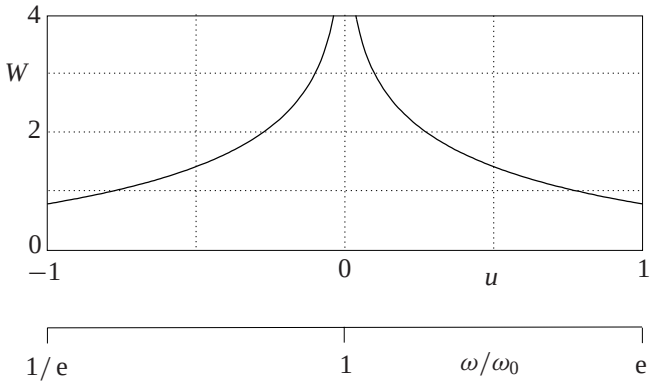


Figure 1.23: Weighting function W in Bode's gain-phase relationship

In (1.106) the first factor under the integral sign is the slope of the magnitude Bode plot (in [decades/decade]) as previously expressed by the variable n. W is a weighting function of the form shown in Fig. 1.23. W has most of it weight near 0, so that $\arg L(j\omega_0)$ is determined by the behavior of L_{ω_0} near 0, that is, by the behavior of L near ω_0. If $\log|L_{\omega_0}|$ would have a constant slope then (1.104) would be recovered exactly, and $\arg L(j\omega_0)$ would be determined by n alone. If the slope of $\log|L|$ varies then the behavior of $|L|$ in neighboring frequency regions also affects $\arg L(j\omega_0)$.

The Bode gain-phase relationship leads to the following observation. Suppose that the general behavior of the Nyquist plot of the loop gain L is as in Fig. 1.14, that is, the loop gain is greater than 1 for low frequencies, and enters the unit disk once (at the frequency $\pm\omega_m$) without leaving it again. The frequency ω_m at which the Nyquist plot of the loop gain L crosses the unit circle is at the center of the crossover region. For stability the phase of the loop gain L at this point should be between -180° and 180°. To achieve a phase margin of at least 60° the phase should be between -120° and 120°. Figure 1.24 (Freudenberg and Looze 1988) illustrates the bounds on the phase in the crossover region. Since $|L|$ decreases at the point of intersection $\arg L$ generally may be expected to be negative in the crossover region.

If at crossover the phase of L is, say, -90° then by Bode's gain-phase relationship $|L|$ decreases at a rate of about 1 [decade/decade]. To avoid instability the rate of decrease cannot be greater than 2 [decades/decade]. Hence, for robust stability in the crossover region the magnitude $|L|$ of the loop gain cannot decreases faster than at a rate somewhere between 1 and 2 [decade/decade]. This bound on the rate of decrease of $|L|$ in turn implies that the crossover region cannot be arbitrarily narrow.

Bode's gain-phase relationship holds for minimum phase systems. For non-minimum phase systems the trade-off between gain attenuation and phase lag is even more troublesome. Let L be non-minimum phase but stable[16]. Then $L = L_m \cdot L_z$, where L_m is minimum phase and L_z is an all-pass function[17] such that $|L_z(j\omega)| = 1$ for all ω. It follows that $|L(j\omega)| = |L_m(j\omega)|$ and

$$\arg L(j\omega) = \arg L_m(j\omega) + \arg L_z(j\omega) \leq \arg L_m(j\omega). \tag{1.110}$$

As L_z only introduces phase lag, the trade-off between gain attenuation and limited phase lag

[14] A rational transfer function is minimum phase if all its poles and zeros have strictly negative real parts (see Exercise 1.6.3).

[15] That is, $L(s)$ is a rational function in s with real coefficients.

[16] That is, L has right-half plane zeros but no right-half plane poles.

[17] That is, $|L_z(j\omega)|$ is constant for all $\omega \in \mathbb{R}$.

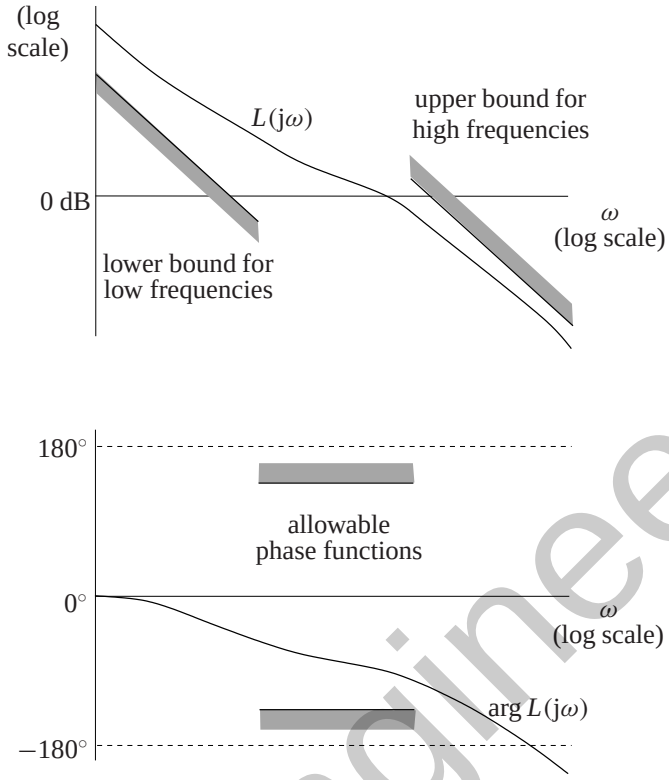


Figure 1.24: Allowable regions for gain and phase of the loop gain L

is further handicapped. Non-minimum phase behavior generally leads to reduction of the open-loop gains — compared with the corresponding minimum phase system with loop gain L_m — or reduced crossover frequencies. The effect of right-half plane zeros — and also that of right-half plane poles — is discussed at greater length in Section 1.7.

Exercise 1.6.3 (Minimum phase). Let

$$L(s) = k\frac{(s-z_1)(s-z_2)\cdots(s-z_m)}{(s-p_1)(s-p_2)\cdots(s-p_n)} \tag{1.111}$$

be a rational transfer function with all its poles $p_1, p_2, \cdots, p_n$ in the left-half complex plane. Then there exists a well-defined corresponding frequency response function $L(j\omega)$, $\omega \in \mathbb{R}$.

Changing the sign of the real part of the ith zero z_i of L leaves the behavior of the magnitude $|L(j\omega)|$, $\omega \in \mathbb{R}$, of L unaffected, but modifies the behavior of the phase $\arg L(j\omega)$, $\omega \in \mathbb{R}$. Similarly, changing the sign of the gain k does not affect the magnitude. Prove that under such changes the phase $\arg L(j\omega)$, $\omega \in \mathbb{R}$, is *minimal* for all frequencies if all zeros z_i lie in the left-half complex plane and k is a positive gain.

This is why transfer functions whose poles and zeros are all in the left-half plane and have positive gain are called *minimum phase* transfer functions. □

1.6.3 Bode's sensitivity integral

Another well-known result of Bode's pioneering work is known as *Bode's sensitivity integral*

Summary 1.6.4 (Bode's sensitivity integral). Suppose that the loop gain L has no poles in the open right-half plane[18]. Then if L has at least two more poles than zeros the sensitivity function

$$S = \frac{1}{1+L} \tag{1.112}$$

satisfies

$$\int_0^\infty \log|S(j\omega)|\, d\omega = 0. \tag{1.113}$$

□

The assumption that L is rational may be relaxed (see for instance Engell (1988)). The statement that L should have at least two more poles than zeros is sometimes phrased as "the pole-zero excess of L is at least two[19]." If the pole-zero excess is one, the integral on the left-hand side of (1.113) is finite but has a nonzero value. If L has right-half plane poles then (1.113) needs to be modified to

$$\int_0^\infty \log|S(j\omega)|\, d\omega = \pi \sum_i \mathrm{Re}\ \ p_i. \tag{1.114}$$

The right-hand side is formed from the open right-half plane poles p_i of L, included according to their multiplicity. The proof of (1.114) may be found in § 1.10.

We discuss some of the implications of Bode's sensitivity integral. Suppose for the time being that L has no open right-half plane poles, so that the sensitivity integral vanishes. Then the integral over all frequencies of $\log|S|$ is zero. This means that $\log|S|$ both assumes negative and positive values, or, equivalently, that $|S|$ both assumes values less than 1 and values greater than 1.

For the feedback system to be useful, $|S|$ needs to be less than 1 over an effective low-frequency band. Bode's sensitivity integral implies that *if*, this can be achieved at all then it is at the cost of disturbance *amplification* (rather than attenuation) at high frequencies. Figure 1.25 illustrates this. If the open-loop system has right-half plane poles this statement still holds. If

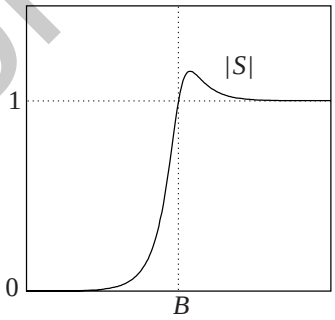


Figure 1.25: Low frequency disturbance attenuation may only be achieved at the cost of high frequency disturbance amplification

the pole-zero excess of the plant is zero or one then disturbance attenuation is possible over all frequencies.

[18] That is, no poles with strictly positive real part.

[19] In adaptive control the expression is "L has relative degree greater than or equal to 2."

Exercise 1.6.5 (Bode integral).

(a) Suppose that the loop gain is $L(s) = k/(1+s\theta)$, with k and θ positive constants. Calculate the sensitivity function S and plot its magnitude. Does Bode's theorem apply?

(b) Repeat this for the loop gain $L(s) = k/(1+s\theta)^2$. First check that the closed-loop system is stable for all positive k and θ.

□

Freudenberg and Looze (1988) use Bode's sensitivity integral to derive a lower bound on the peak value of the sensitivity function in the presence of constraints on the low-frequency behavior of S and the high-frequency behavior of L. Suppose that L is real-rational, minimum phase and has a pole-zero excess of two or more. Let ω_L and ω_H be two frequencies such that $0 < \omega_L < \omega_H$. Assume that S and L are bounded by

$$|S(\mathrm{j}\omega)| \leq \alpha < 1, \qquad 0 \leq \omega \leq \omega_L, \tag{1.115}$$

and

$$|L(\mathrm{j}\omega)| \leq \varepsilon \left(\frac{\omega_H}{\omega}\right)^{k+1}, \qquad \omega \geq \omega_H, \tag{1.116}$$

with $0 < \varepsilon < 0.5$ and $k \geq 0$. Then the peak value of the sensitivity function is bounded by

$$\sup_{\omega_l \leq \omega \leq \omega_H} |S(\mathrm{j}\omega)| \geq \frac{1}{\omega_H - \omega_L}\left(\omega_L \log\frac{1}{\alpha} - \frac{3\varepsilon\omega_H}{2k}\right). \tag{1.117}$$

The proof is given in § 1.10.

This result is an example of the more general phenomenon that bounds on the loop gain in different frequency regions interact with each other. Control system design therefore involves trade-offs between phenomena that occur in different frequency regions. The interaction becomes more severe as the bounds become tighter. If α or ε decrease or if ω_L or k increase then the lower bound for the peak value of the sensitivity function increases. Also if the frequencies ω_L and ω_H are required to be closely together, that is, the crossover region is required to be small, then this is paid for by a high peak value of S in the crossover region.

The bounds (1.115) and (1.116) are examples of the bounds indicated in Fig. 1.24. The inequality (1.117) demonstrates that stiff bounds in the low- and high-frequency regions may cause serious stability robustness problems in the crossover frequency range.

The natural limitations on stability, performance and robustness as discussed in this section are aggravated by the presence of right-half plane plant poles and zeros. This is the subject of Section 1.7.

Exercise 1.6.6 (Bode's integral for the complementary sensitivity). Let T be the complementary sensitivity function of a stable feedback system that has integrating action of at least order two[20]. Prove that

$$\int_0^\infty \log\left|T\left(\frac{1}{\mathrm{j}\omega}\right)\right| d\omega = \pi \prod_i \mathrm{Re}\,\frac{1}{z_i}, \tag{1.118}$$

with the z_i the right-half plane zeros of the loop gain L (Middleton 1991; Kwakernaak 1995). What does this equality imply for the behavior of T? □

[20] This means that $1/L(s)$ behaves as $O(s^2)$ for $s \to 0$, (see § 2.3).

1.7 Limits of performance

1.7.1 Introduction

In this section[21] we present a brief review of several inherent limitations of the behavior of the sensitivity function S and its complement T which result from the pole-zero pattern of the plant.

In particular, right-half plane zeros and poles play an important role. The reason is that if the the plant has right-half plane poles, it is unstable, which imposes extra requirements on the loop gain. If the plant has right-half plane zeros, its inverse is unstable, which imposes limitations on the way the dynamics of the plant may be compensated.

More extensive discussions on limits of performance may be found in Engell (1988) for the SISO case and Freudenberg and Looze (1988) for both the SISO and the MIMO case.

1.7.2 Freudenberg-Looze equality

A central result is the *Freudenberg-Looze equality*, which is based on the Poisson integral formula from complex function theory. The result that follows was originally obtained by Freudenberg and Looze (1985) and Freudenberg and Looze (1988).

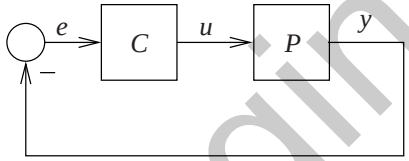


Figure 1.26: SISO feedback system

Summary 1.7.1 (Freudenberg-Looze equality). Suppose that the closed-loop system of Fig. 1.26 is stable, and that the loop gain has a right-half plane zero $z = x + jy$ with $x > 0$. Then the sensitivity function $S = 1/(1 + L)$ must satisfy

$$\int_{-\infty}^{\infty} \log(|S(j\omega)|) \frac{x}{x^2 + (y - \omega)^2}\, d\omega = \pi \log |B_{\text{poles}}^{-1}(z)|. \tag{1.119}$$

B_{poles} is the *Blaschke product*

$$B_{\text{poles}}(s) = \prod_i \frac{p_i - s}{\bar{p}_i + s}, \tag{1.120}$$

formed from the open right-half plane poles p_i of the loop gain $L = PC$. The overbar denotes the complex conjugate. □

The proof is given in § 1.10. It relies on the Poisson integral formula from complex function theory.

1.7.3 Trade-offs for the sensitivity function

We discuss the consequences of the Freudenberg-Looze relation (1.119), which holds at any right-half plane zero $z = x + jy$ of the loop gain, and, hence, at any right-half plane zero of

[21] The title has been taken from Engell (1988) and Boyd and Barratt (1991).

the plant. The presentation follows Freudenberg and Looze (1988) and Engell (1988). We first rewrite the equality in the form

$$\int_0^\infty \log(|S(\mathrm{j}\omega)|)\; w_z(\omega)\; d\omega = \log|B_{\text{poles}}^{-1}(z)|, \tag{1.121}$$

with w_z the function

$$w_z(\omega) = \frac{1}{\pi}\left(\frac{x}{x^2+(\omega-y)^2} + \frac{x}{x^2+(\omega+y)^2}\right). \tag{1.122}$$

We arrange (1.121) in turn as

$$\int_0^\infty \log(|S(\mathrm{j}\omega)|)\; dW_z(\omega) = \log|B_{\text{poles}}^{-1}(z)|, \tag{1.123}$$

with W_z the function

$$W_z(\omega) = \int_0^\omega w_z(\eta)\; d\eta = \frac{1}{\pi}\arctan\frac{\omega-y}{x} + \frac{1}{\pi}\arctan\frac{\omega+y}{x}. \tag{1.124}$$

The function W_z is plotted in Fig. 1.27 for different values of the ratio y/x of the imaginary part to the real part. The plot shows that W_z increases monotonically from 0 to 1. Its steepest increase is about the frequency $\omega = |z|$.

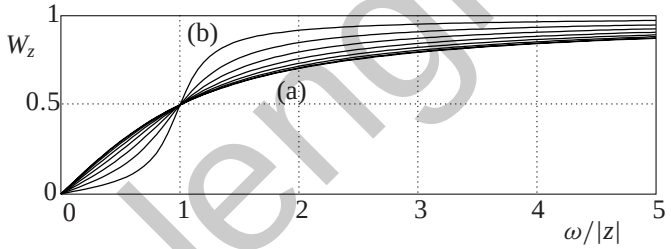


Figure 1.27: The function W_z for values of $\arg z$ increasing from (a) 0 to (b) almost $\pi/2$

Exercise 1.7.2 (Weight function). Show that the weight function W_z represents the extra phase lag contributed to the phase of the plant by the fact that the zero z is in the right-half plane, that is, $W_z(\omega)$ is the phase of

$$\frac{z+\mathrm{j}\omega}{z-\mathrm{j}\omega}. \tag{1.125}$$

□

Exercise 1.7.3 (Positivity of right-hand side of the Freudenberg-Looze equality). Prove that $\log|B_{\text{poles}}^{-1}(z)|$ is positive for any z whose real part is positive. □

The comments that follow hold for plants that have at least one right-half plane zero.

The function w_z is positive and also $\log|B_{\text{poles}}^{-1}(z)|$ is positive. Hence, (1.121) implies that if $\log|S|$ is negative over some frequency range so that, equivalently, $|S|$ is less than 1, then necessarily $|S|$ is greater than 1 over a complementary frequency range. This we already concluded from Bode's sensitivity integral.

The Freudenberg-Looze equality strengthens the Bode integral because of the weighting function w_z included in the integrand. The quantity $dW_z(\mathrm{j}\omega) = w_z(\mathrm{j}\omega)d\omega$ may be viewed as a weighted length of the frequency interval. The weighted length equals the extra phase added by the right-half plane zero z over the frequency interval. The larger the weighted length is, the more the interval contributes to the right-hand side of (1.123). The weighting function determines to what extent small values of $|S|$ at low frequencies need to be compensated by large values at high frequencies. We argue that if $|S|$ is required to be small in a certain frequency band—in particu-

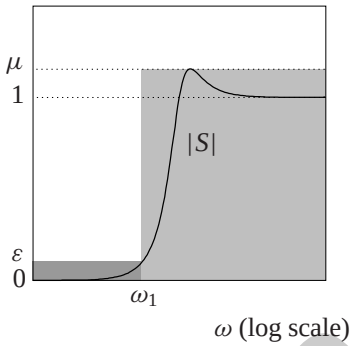


Figure 1.28: Bounds on $|S|$

lar, a low-frequency band—it necessarily peaks in another band. Suppose that we wish $|S(\mathrm{j}\omega)|$ to be less than a given small number ε in the frequency band $[0,\ \omega_1]$, with ω_1 given. We should like to know something about the peak value μ of $|S|$ in the complementary frequency range. Figure 1.28 shows the numbers ε and μ and the desired behavior of $|S|$. Define the bounding function

$$b(\omega) = \begin{cases} \varepsilon & \text{for } |\omega| \leq \omega_1, \\ \mu & \text{for } |\omega| > \omega_1. \end{cases} \tag{1.126}$$

Then $|S(\mathrm{j}\omega)| \leq b(\omega)$ for $\omega \in \mathbb{R}$ and the Freudenberg-Looze equality together imply that b needs to satisfy

$$\int_0^\infty \log(b(\omega))\, dW_z(\omega) \geq \log|B_{\text{poles}}^{-1}(z)|. \tag{1.127}$$

Evaluation of the left-hand side leads to

$$W_z(\omega_1)\log\varepsilon + (1 - W_z(\omega_1))\log\mu \geq \log|B_{\text{poles}}^{-1}(z)|. \tag{1.128}$$

Resolution of (1.128) results in the inequality

$$\mu \geq \left(\frac{1}{\varepsilon}\right)^{\frac{W_z(\omega_1)}{1-W_z(\omega_1)}} \cdot \left|B_{\text{poles}}^{-1}(z)\right|^{\frac{1}{1-W_z(\omega_1)}}. \tag{1.129}$$

We note this:

- For a fixed zero $z = x + \mathrm{j}y$ and fixed $\omega_1 > 0$ the exponents in this expression are positive. By 1.7.3 we have $|B_{\text{poles}}^{-1}(z)| \geq 1$. Hence, for $\varepsilon < 1$ the peak value μ is greater than 1. Moreover, the smaller ε is, the larger is the peak value. Thus, small sensitivity at low frequencies is paid for by a large peak value at high frequencies.

- For fixed ε, the two exponents increase monotonically with ω_1, and approach ∞ as ω_1 goes to ∞. The first exponent (that of $1/\varepsilon$) crosses the value 1 at $\omega = \sqrt{x^2+y^2} = |z|$. Hence, if the width of the band over which sensitivity is required to be small is greater than the magnitude of the right-half plane zero z, the peak value assumes excessive values.

The Freudenberg-Looze equality holds for *any* right-half plane zero, in particular the one with smallest magnitude. Therefore, if excessive peak values are to be avoided, the width of the band over which sensitivity may be made small cannot be extended beyond the magnitude of the *smallest* right-half plane zero.

The number

$$|B^{-1}_{\text{poles}}(z)| = \prod_i \left| \frac{\bar{p}_i + z}{p_i - z} \right| \tag{1.130}$$

is replaced with 1 if there are no right-half plane poles. Otherwise, it is greater than 1. Hence, right-half plane poles make the plant more difficult to control.

The number (1.130) is large if the right-half plane zero z is close to any of the right-half plane poles p_i. If this situation occurs, the peak values of $|S|$ are correspondingly large, and, as a result, the plant is difficult to control. The Freudenberg-Looze equality holds for any right-half plane zero. Therefore, plants with a right-half plane zero close to a right-half plane pole are difficult to control. The situation is worst when a right-half plane zero coincides with a right-half plane pole—then the plant has either an uncontrollable or an unobservable unstable mode.

We summarize the qualitative effects of right-half plane zeros of the plant on the shape of the sensitivity function S (Engell 1988).

Summary 1.7.4 (Effect of right-half plane open-loop zeros).

(a) The upper limit of the band over which effective disturbance attenuation is possible is constrained from above by the magnitude of the smallest right-half plane zero.

(b) If the plant has unstable poles, the achievable disturbance attenuation is further impaired. This effect is especially pronounced when one or several right-half plane pole-zero pairs are close.

(c) If the plant has no right-half plane zeros the maximally achievable bandwidth is solely constrained by the plant capacity. As seen in the next subsection the right-half plane pole with largest magnitude constrains the *smallest* bandwidth that is required.

The trade-off between disturbance attenuation and amplification is subject to Bode's sensitivity integral.

□

We consider an example.

Example 1.7.5 (Double inverted pendulum). To illustrate these results we consider the double inverted pendulum system of Fig. 1.29. Two pendulums are mounted on top of each other. The input to the system is the horizontal position u of the pivot of the lower pendulum. The measured output is the angle ϕ that the lower pendulum makes with the vertical. The pendulums have equal lengths L, equal masses m and equal moments of inertia J (taken with respect to the center of gravity).

The transfer function P from the input u to the angle ϕ may be found (Kwakernaak and Westdijk 1985) to be given by

$$P(s) = \frac{1}{L} \frac{s^2(-(3K+1)s^2+3)}{(K^2+6K+1)s^4 - 4(K+2)s^2 + 3}, \tag{1.131}$$

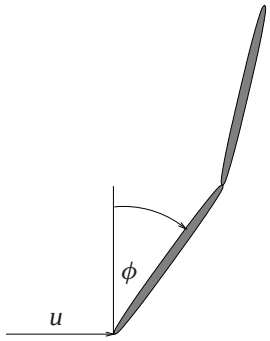


Figure 1.29: Double inverted pendulum system

with K the ratio $K = J/(mL^2)$. For a pendulum whose mass is homogeneously distributed along its length K is $\frac{1}{3}$. If we furthermore let $L = 1$ then

$$P(s) = \frac{s^2(-2s^2+3)}{\frac{28}{9}s^4 - \frac{28}{3}s^2 + 3}. \tag{1.132}$$

This plant transfer function has zeros at 0, 0, and ± 1.22474, and poles at ± 0.60507 and ± 1.62293. The plant has two right-half plane poles and one right-half plane zero.

By techniques that are explained in Chapter 6 we may calculate the transfer function C of the compensator that makes the transfer matrices S and T stable and at the same time minimizes the *peak value* of $\sup_{\omega\in\mathbb{R}} |S(\mathrm{j}\omega)|$ of the sensitivity function S of the closed-loop system. This compensator transfer function is

$$C(s) = 1.6292\,\frac{(s+1.6229)(s+0.6051)(s-0.8018)}{(s+1.2247)s^2}. \tag{1.133}$$

Figure 1.30 shows the magnitude plot (a) of the corresponding sensitivity function S. This magnitude does not depend on the frequency ω and has the constant value 21.1178. Note that the compensator has a double pole at 0, which in the closed loop cancels against the double zero at 0 of the plant. The corresponding double closed-loop pole at 0 actually makes the closed-loop system unstable. It causes the plant input u to drift.

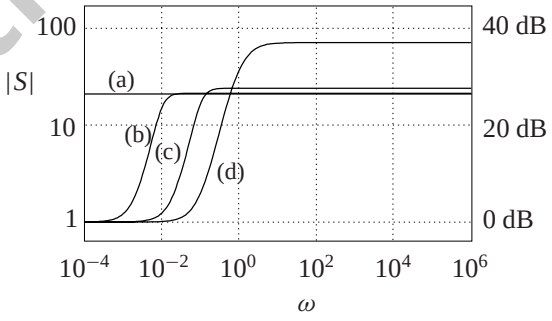


Figure 1.30: Sensitivity functions for the double pendulum system

The peak value 21.1178 for the magnitude of the sensitivity function is quite large (as compared with 1). No compensator exists with a smaller peak value. To reduce the sensitivity of

the feedback system to low frequency disturbances it is desirable to make $|S|$ smaller at low frequencies. This is achieved, for instance, by the compensator with transfer function

$$C(s) = 1.6202\,\frac{(s+1.6229)(s+0.6051)(s-0.8273)}{(s+1.2247)(s+0.017390)(s-0.025715)}. \tag{1.134}$$

Figure 1.30 shows the magnitude (b) of the corresponding sensitivity function. This compensator no longer cancels the double zero at 0. As a result, the sensitivity S at zero frequency is 1, which means that constant disturbances are not attenuated. The reason is structural: Because the plant transfer function P equals zero at frequency 0, the plant is unable to compensate for constant disturbances. Actually, the zeros at 0 play the role of right-half plane zeros, except that they bound the frequencies where S may be made small from *below* rather than from above.

The plot of Fig. 1.30 also shows that compared with the compensator (1.133) the compensator (1.134) reduces the sensitivity to disturbances up to a frequency of about 0.01. By further modifying the compensator this frequency may be pushed up, at the price of an increase in the peak value of the magnitude of S. Figure 1.30 shows two more magnitude plots (c) and (d). The closer the magnitude 1.2247 of the right-half plane plant zero is approached the more the peak value increases.

There exist compensators that achieve magnitudes less than 1 for the sensitivity in the frequency range, say, between 0.01 and 0.1. The cost is a further increase of the peak value.

The compensators considered so far all result in sensitivity functions that do not approach the value 1 as frequency increases to ∞. The reason is that the loop gain does not approach 0. This undesirable phenomenon, which results in large plant input amplitudes and high sensitivity to measurement noise, is removed in Example 1.7.9. □

Exercise 1.7.6 (Interlacing property). Check that the double inverted pendulum does not have the parity interlacing property of § 1.3.6 (p. 19). Hence, no stabilizing compensator exists that by itself is stable. □

Exercise 1.7.7 (Lower bound for the peak value of S).

(a) Define $\|S\|_\infty$ as the peak value of $|S|$, that is, $\|S\|_\infty = \sup_{\omega\in\mathbb{R}} |S(\mathrm{j}\omega)|$. Use (1.129) to prove that if the closed-loop system is stable then

$$\|S\|_\infty \geq \left|B_{\text{poles}}^{-1}(z)\right|, \tag{1.135}$$

where B_{poles} is the Blaschke product formed from the right-half plane poles of the plant, and z any right-half plane zero of the plant.

(b) Check that the compensator (1.133) actually achieves this lower bound.

□

1.7.4 Trade-offs for the complementary sensitivity function

Symmetrically to the results for the sensitivity function well-defined trade-offs hold for the complementary sensitivity function. The role of the right-half plane zeros is now taken by the right-half plant open-loop *poles*, and vice-versa. This is seen by writing the complementary sensitivity function as

$$T = \frac{L}{1+L} = \frac{1}{1+\frac{1}{L}}. \tag{1.136}$$

Comparison with Freudenberg-Looze equality of 1.7.1 leads to the conclusion that for any right-half plane open-loop pole $p = x + jy$ we have (Freudenberg and Looze 1988)

$$\int_{-\infty}^{\infty} \log(|T(j\omega)|) \frac{x}{x^2 + (y-\omega)^2}\, d\omega = \pi \log|B_{\text{zeros}}^{-1}(p)|, \tag{1.137}$$

with B_{zeros} the Blaschke product

$$B_{\text{zeros}}(s) = \prod_i \frac{z_i - s}{\bar{z}_i + s} \tag{1.138}$$

formed from the open right-half plane *zeros* z_i of L.

We consider the implications of the equality (1.137) on the shape of T. Whereas the sensitivity S is required to be small at *low* frequencies, T needs to be small at *high* frequencies. By an argument that is almost dual to that for the sensitivity function it follows that if excessive peaking of the complementary sensitivity function at low and intermediate frequencies is to be avoided, $|T|$ may only be made small at frequencies that *exceed* the magnitude of the open-loop right-half plane *pole* with *largest* magnitude. Again, close right-half plane pole-zero pairs make things worse.

We summarize the qualitative effects of right-half plane zeros of the plant on the shape achievable for the complementary sensitivity function T (Engell 1988).

Summary 1.7.8 (Effect of right-half plane open-loop poles).

(a) The lower limit of the band over which the complementary sensitivity function may be made small is constrained from below by the magnitude of the largest right-half plane open-loop pole. Practically, the achievable bandwidth is always greater than this magnitude.

(b) If the plant has right-half plane zeros, the achievable reduction of T is further impaired. This effect is especially pronounced when one or several right-half plane pole-zero pairs are very close.

□

Figure 1.31 summarizes the difficulties caused by right-half plane zeros and poles of the plant transfer function P. S can only be small up to the magnitude of the smallest right-half plane zero. T can only start to roll off to zero at frequencies greater than the magnitude of the largest right-half plane pole. The crossover region, where S and T assume their peak values, extends over the intermediate frequency range.

Example 1.7.9 (Double inverted pendulum). We return to the double inverted pendulum of Example 1.7.5. For robustness to plant uncertainty and reduction of the susceptibility to measurement noise it is necessary that the loop gain decreases to zero at high frequencies. Correspondingly, the complementary sensitivity function also decreases to zero while the sensitivity function approaches 1. The compensator

$$C(s) = -\frac{1.5136(s+1.6229)(s+0.60507)(s-0.82453)}{(s+1.2247)(s+0.017226)(s-0.025394)(1+0.00061682s)}, \tag{1.139}$$

whose transfer function is strictly proper, accomplishes this. Figure 1.32 shows that for low frequencies the magnitude plot (e) of the corresponding sensitivity function closely follows the magnitude plot (c) of Fig. 1.30, which is repeated in Fig. 1.32. At high frequencies the magnitude of S drops off to 1, however, starting at a frequency of about 100.

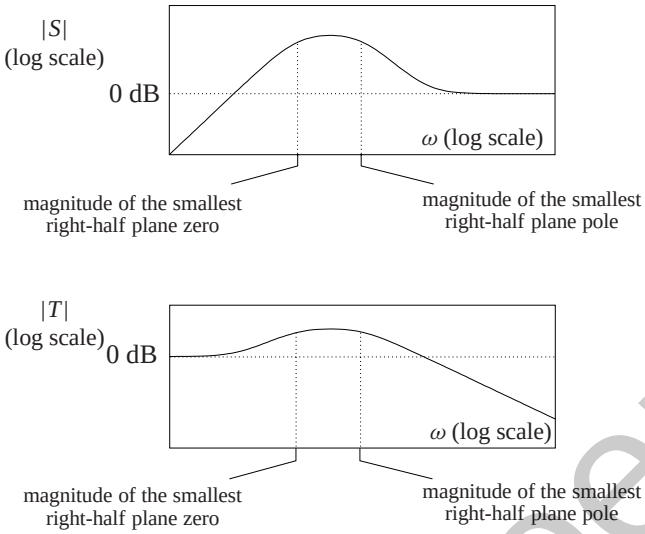


Figure 1.31: Right-half plane zeros and poles constrain S and T

The lowest frequency at which $|S|$ may start to drop off to 1 coincides with the lowest frequency at which the complementary sensitivity may be made to start decreasing to zero. This, in turn, is determined by the magnitude 1.6229 of the right-half plane plant pole with largest magnitude. The magnitude plot (f) in Fig. 1.32 shows that making $|S|$ drop off at a lower frequency than in (e) causes the peak value to increase. □

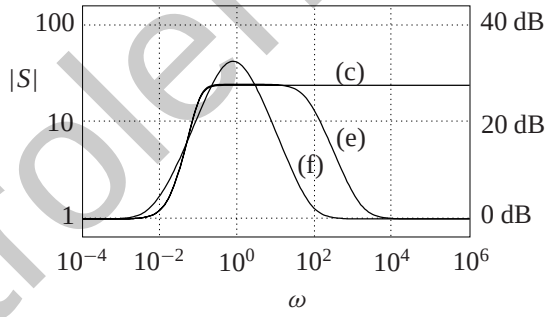


Figure 1.32: More sensitivity functions for the double pendulum system

1.8 Two-degrees-of-freedom feedback systems

In § 1.3 we introduced the two-degrees-of-freedom configuration of Fig. 1.33. The function of the precompensator F is to improve the closed-loop response to command inputs r.

Figure 1.34 shows two other two-degrees-of-freedom configurations. In this section we study whether it makes a difference which configuration is chosen. We restrict the discussion to SISO systems.

In the configuration of Fig. 1.33 we write the plant and compensator transfer functions in the

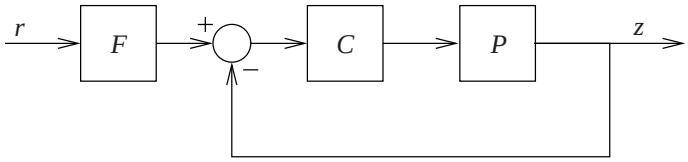


Figure 1.33: Two-degrees-of-freedom feedback system configuration

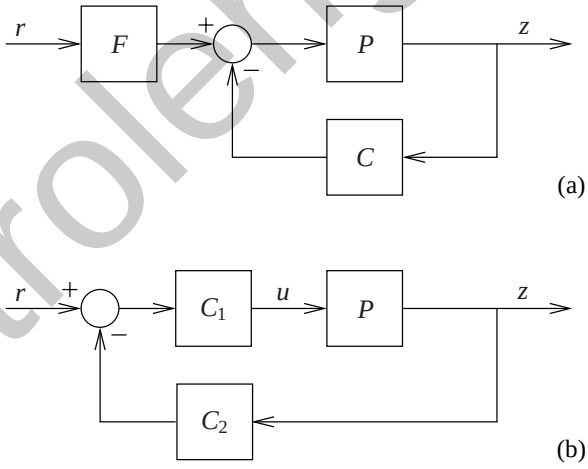


Figure 1.34: Further two-degrees-of-freedom feedback system configurations

polynomial fraction form

$$P = \frac{N}{D}, \qquad C = \frac{Y}{X}. \tag{1.140}$$

The feedback loop is stable if and only if the roots of the closed-loop characteristic polynomial $D_{\text{cl}} = DX + NY$ are all in the open left-half complex plane.

In this same configuration, the closed-loop transfer function H from the command signal r to the control system output z is

$$H = \frac{PC}{1+PC}F = \frac{NY}{D_{\text{cl}}}F. \tag{1.141}$$

The prefilter transfer function F is available to compensate for any deficiencies of the uncompensated closed-loop transfer function

$$H_0 = \frac{NY}{D_{\text{cl}}}. \tag{1.142}$$

Right-half plane zeros of this uncompensated transfer function are a handicap for the compensation. Right-half plane roots of N (that is, open-loop right-half plane zeros) and right-half plane roots of Y may well be present. Such zeros cannot be canceled by corresponding poles of F because this would make the precompensator, and, hence, the whole control system, unstable.

Next consider the two-degrees-of-freedom configuration of Fig. 1.34(a). We now have for the closed-loop transfer function

$$H = \frac{P}{1+PC}F = \underbrace{\frac{NX}{D_{\text{cl}}}}_{H_0}F. \tag{1.143}$$

Inspection shows that the open-loop plant zeros re-occur in the uncompensated closed-loop transfer function H_0 but that instead of the roots of Y (the compensator zeros) now the roots of X (the compensator poles) appear as zeros. Hence, the precompensator design problem for this configuration is different from that for the configuration of Fig. 1.33. In fact, if the compensator has right-half plane poles or zeros, or both, it is impossible to achieve identical overall closed-loop transfer functions for the two configurations.

Comparison of (1.142) and (1.143) suggests that there may exist a configuration such that the numerator of the uncompensated closed-loop transfer function is independent of the compensator. To investigate this, consider the configuration of Fig. 1.34(b). C_1 and C_2 have the polynomial fractional representations

$$C_1 = \frac{Y_1}{X_1}, \quad C_2 = \frac{Y_2}{X_2}. \tag{1.144}$$

To match the closed-loop characteristics of the configurations of Figs. 1.33 and 1.34(a) we need $C = C_1C_2$. This implies that $X_1X_2 = X$ and $Y_1Y_2 = Y$. The closed-loop transfer function now is

$$H = \frac{PC_1}{1+PC} = \frac{NX_2Y_1}{D_{\text{cl}}}. \tag{1.145}$$

Inspection shows that the numerator of H is independent of the compensator if we let $X_2Y_1 = 1$, so that

$$C_1 = \frac{Y_1}{X_1} = \frac{1}{X_1X_2} = \frac{1}{X}, \quad C_2 = \frac{Y_2}{X_2} = Y_1Y_2 = Y. \tag{1.146}$$

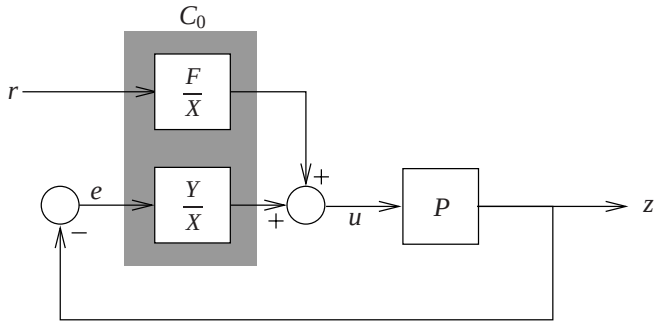


Figure 1.35: $1\frac{1}{2}$-degrees-of-freedom control system

The closed-loop transfer function now is

$$H = \frac{N}{D_{\text{cl}}}. \tag{1.147}$$

The corresponding configuration of Fig. 1.34(b) has two disadvantages:

1. The configuration appears to require the implementation of a block with the purely polynomial transfer function $C_2(s) = Y(s)$, which is physically impossible (unless Y is of degree zero).

2. The configuration actually has only one degree of freedom. The reason is that one degree of freedom has been used to make the numerator of the closed-loop transfer function independent of the compensator.

The first difficulty may be remedied by noticing that from the block diagram we have

$$u = C_1 r - C_1 C_2 z = \frac{1}{X} r + \frac{Y}{X} e. \tag{1.148}$$

This implies

$$Xu = r + Ye. \tag{1.149}$$

This input-output relation — with r and e as inputs and u as output — may be implemented by a state realization of order equal to the degree of the polynomial X.

The second disadvantange may be overcome by modifying (1.149) to

$$Xu = Fr + Ye, \tag{1.150}$$

with F a polynomial of degree less than or equal to that of X. This still allows implementation by a state realization of order equal to the degree of X (see Exercise 1.8.1). The compensator is represented in the block diagram of Fig. 1.35. The combined block C_0 is jointly realized as a single input-output-state system. The closed-loop transfer function is

$$H = \frac{NF}{D_{\text{cl}}}. \tag{1.151}$$

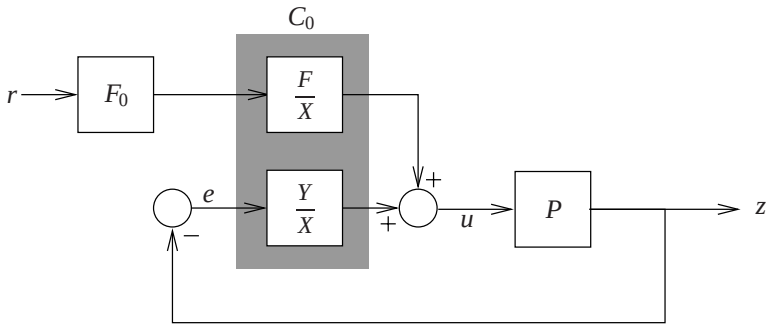


Figure 1.36: $2\frac{1}{2}$-degrees-of-freedom configuration

The design of the prefilter amounts to choosing the polynomial F. This might be called a $1\frac{1}{2}$-*degrees-of-freedom* control system[22]. By application of a further prefilter F_0 as in Fig. 1.36 the closed-loop transfer function becomes

$$H = \frac{NF}{D_{\rm cl}} F_0. \tag{1.152}$$

This results in a $2\frac{1}{2}$-degrees-of-freedom control system.

An application is described in Example 2.9.5 in § 2.9.5.

Exercise 1.8.1 (Realization of the $1\frac{1}{2}$-degrees-of-freedom compensator). Represent the polynomials X, F and Y as

$$X(s) = s^n + a_{n-1}s^{n-1} + a_{n-2}s^{n-2} + \cdots + a_0, \tag{1.153}$$
$$F(s) = b_n s^n + b_{n-1}s^{n-1} + b_{n-2}s^{n-2} + \cdots + b_0, \tag{1.154}$$
$$Y(s) = c_n s^n + c_{n-1}s^{n-1} + c_{n-2}s^{n-2} + \cdots + c_0. \tag{1.155}$$

1. Show that the $1\frac{1}{2}$-degrees-of-freedom compensator $Xu = Fr + Ye$ may be realized as in Fig. 1.37.
2. Find a state representation for the compensator.
3. Prove that the feedback system of Fig. 1.35, with the dashed block realized as in Fig. 1.37, has the closed-loop characteristic polynomial $DX + NY$.

□

1.9 Conclusions

It is interesting to observe a number of "symmetries" or "dualities" in the results reviewed in this chapter (Kwakernaak 1995). For good performance and robustness the loop gain L of a well-designed linear feedback system should be

- large at low frequencies and
- small at high frequencies.

[22] Half degrees of freedom were introduced in control engineering terminology by Grimble (1994), though with a different connotation than the one used here.

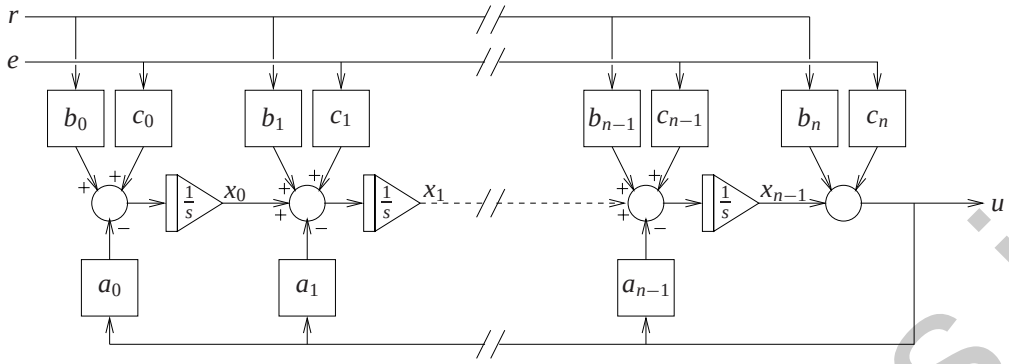


Figure 1.37: Realization of the $1\frac{1}{2}$-degrees-of-freedom compensator $Xu = Fr + Ye$

As the result, the sensitivity function S is

- small at low frequencies and
- approximately equal to 1 at high frequencies.

The complementary sensitivity function T is

- approximately equal to 1 at low frequencies and
- small at high frequencies.

Such well-designed feedback systems are

- robust with respect to perturbations of the inverse loop gain at low frequencies, and
- robust with respect to perturbations of the loop gain at high frequencies.

Furthermore,

- right-half plane open-loop zeros limit the frequency up to which S may be made small at low frequencies, and
- right-half plane open-loop poles limit the frequency from which T may be made small at high frequencies.

Note that to a large extent performance and robustness go hand in hand, that is, the requirements for good performance imply good robustness, and vice-versa. This is also true for the critical crossover region, where peaking of both S and T is to be avoided, both for performance and robustness.

1.10 Appendix: Proofs

In this section we collect a number of proofs for Chapter 1.

1.10.1 Closed-loop characteristic polynomial

We first prove (1.44) in Subsection 1.3.3.

Proof 1.10.1 (Closed-loop characteristic polynomial). Let

$$\dot{x} = Ax + Bu, \quad y = Cx + Du, \tag{1.156}$$

be a state realization of the block L in the closed-loop system of Fig. 1.11. It follows that $L(s) = C(sI - A)^{-1} + D$. From $u = -y$ we obtain with the output equation that $u = -Cx - Du$, so that $u = -(I + D)^{-1}Cx$. Since by assumption $I + D = I + L(j\infty)$ is nonsingular the closed-loop system is well-defined. Substitution of u into the state differential equation shows that the closed-loop system is described by the state differential equation

$$\dot{x} = [A - B(I + D)^{-1}C]x. \tag{1.157}$$

The characteristic polynomial χ_{cl} of the closed-loop system hence is given by

$$\begin{aligned} \chi_{cl}(s) &= \det[sI - A + B(I + D)^{-1}C] \\ &= \det(sI - A) \cdot \det[I + (sI - A)^{-1}B(I + D)^{-1}C]. \end{aligned} \tag{1.158}$$

Using the well-known determinant equality $\det(I + MN) = \det(I + NM)$ it follows that

$$\begin{aligned} \chi_{cl}(s) &= \det(sI - A) \cdot \det[I + (I + D)^{-1}C(sI - A)^{-1}B] \\ &= \det(sI - A) \cdot \det[(I + D)^{-1}] \cdot \det[I + D + C(sI - A)^{-1}B] \\ &= \det(sI - A) \cdot \det[(I + D)^{-1}] \cdot \det[I + L(s)]. \end{aligned} \tag{1.159}$$

Denoting the open-loop characteristic polynomial as $\det(sI - A) = \chi(s)$ we thus have

$$\frac{\chi_{cl}(s)}{\chi(s)} = \frac{\det[I + L(s)]}{\det[I + L(j\infty)]}. \tag{1.160}$$

■

1.10.2 The Nyquist criterion

The proof of the generalized Nyquist criterion of Summary 1.3.13 in Subsection 1.3.5 relies on the *principle of the argument* of complex function theory[23].

Summary 1.10.2. Principle of the argument Let R be a rational function, and C a closed contour in the complex plane as in Fig. 1.38. As the complex number s traverses the contour C in clockwise direction, its image $R(s)$ under R traverses a closed contour that is denoted as $R(C)$, also shown in Fig. 1.38. Then as s traverses the contour C exactly once in clockwise direction,

(the number of times $R(s)$ encircles the origin in clockwise direction as s traverses C)
=
(the number of zeros of R inside C) − (the number of poles of R inside C).

We prove the generalized Nyquist criterion of Summary 1.3.13.

Proof of the generalized Nyquist criterion. We apply the principle of the argument to (1.160), where we choose the contour C to be the so-called *Nyquist contour* or *D-contour* indicated in Fig. 1.39. The radius ρ of the semicircle is chosen so large that the contour encloses all the right-half plane roots of both χ_{cl} and χ_{ol}. Then by the principle of the argument the number of times that the image of $\det(I + L)$ encircles the origin equals the number of right-half plane roots of χ_{cl} (i.e., the number of unstable closed-loop poles) minus the number of right-half plane roots of χ_{ol} (i.e., the number of unstable open-loop poles). The Nyquist criterion follows by letting the radius ρ of the semicircle approach ∞. Note that as ρ approaches ∞ the image of the semicircle under $\det(I + L)$ shrinks to the single point $\det(I + L(j\infty))$. ■

[23] See Henrici (1974). The generalized form in which we state the principle may be found in Postlethwaite and MacFarlane (1979a).

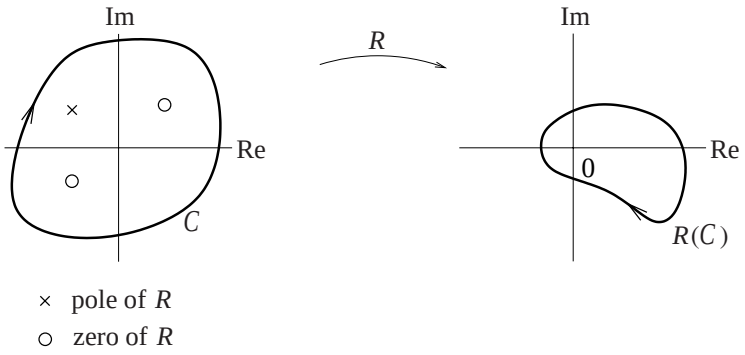


Figure 1.38: Principle of the argument. Left: a closed contour C in the complex plane. Right: the image $R(C)$ of C under a rational function R.

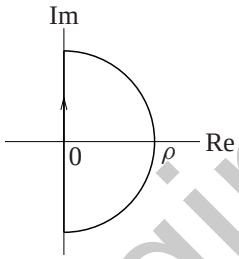


Figure 1.39: Nyquist contour

1.10.3 Bode's sensitivity integral

The proof of Bode's sensitivity integral is postponed until the next subsection. Accepting it as true we use it to derive the inequality (1.117) of Subsection 1.6.3.

Proof 1.10.3 (Lower bound for peak value of sensitivity). If the open-loop system is stable then we have according to Bode's sensitivity integral

$$\int_0^\infty \log|S(\mathrm{j}\omega)|\,d\omega = 0. \tag{1.161}$$

From the assumption that $|S(\mathrm{j}\omega)| \le \alpha < 1$ for $0 \le \omega \le \omega_L$ it follows that if $0 < \omega_L < \omega_H < \infty$ then

$$\begin{aligned} 0 &= \int_0^\infty \log|S(\mathrm{j}\omega)|\,d\omega \\ &= \int_0^{\omega_L} \log|S(\mathrm{j}\omega)|\,d\omega + \int_{\omega_L}^{\omega_H} \log|S(\mathrm{j}\omega)|\,d\omega + \int_{\omega_H}^\infty \log|S(\mathrm{j}\omega)|\,d\omega \\ &\le \omega_L \log\alpha + (\omega_H - \omega_L) \sup_{\omega_L \le \omega \le \omega_H} \log|S(\mathrm{j}\omega)| + \int_{\omega_H}^\infty \log|S(\mathrm{j}\omega)|\,d\omega. \end{aligned} \tag{1.162}$$

As a result,

$$(\omega_H - \omega_L) \sup_{\omega_L \le \omega \le \omega_H} \log|S(\mathrm{j}\omega)| \ge \omega_L \log\frac{1}{\alpha} - \int_{\omega_H}^\infty \log|S(\mathrm{j}\omega)|\,d\omega. \tag{1.163}$$

Next consider the following sequence of (in)equalities on the tail part of the sensitivity integral

$$\begin{aligned}\left|\int_{\omega_H}^{\infty} \log|S(\mathrm{j}\omega)|\, d\omega\right| &\leq \int_{\omega_H}^{\infty} |\log|S(\mathrm{j}\omega)||\, d\omega \\ &\leq \int_{\omega_H}^{\infty} |\log S(\mathrm{j}\omega)|\, d\omega = \int_{\omega_H}^{\infty} |\log[1+L(\mathrm{j}\omega)]|\, d\omega. \end{aligned} \tag{1.164}$$

From the inequalities (4.1.38) and (4.1.35) of Abramowitz and Stegun (1965) we have for any complex number z such that $0 \leq |z| \leq 0.5828$

$$|\log(1+z)| \leq -\log(1-|z|) \leq |\log(1-|z|)| \leq \frac{3|z|}{2}. \tag{1.165}$$

The assumption that

$$|L(\mathrm{j}\omega)| \leq \varepsilon \left(\frac{\omega_H}{\omega}\right)^{k+1} \quad \text{for } \omega > \omega_H \tag{1.166}$$

with $0 < \varepsilon < 0.5$, implies that $|L(\mathrm{j}\omega)| \leq \varepsilon < 0.5$ for $\omega > \omega_H$. With this it follows from (1.164) and (1.165) that

$$\left|\int_{\omega_H}^{\infty} \log|S(\mathrm{j}\omega)|d\omega\right| \leq \int_{\omega_H}^{\infty} \frac{3}{2}|L(\mathrm{j}\omega)|\, d\omega \leq \int_{\omega_H}^{\infty} \frac{3\varepsilon}{2}\left(\frac{\omega_H}{\omega}\right)^{k+1} d\omega = \frac{3\varepsilon}{2}\frac{\omega_H}{k}. \tag{1.167}$$

The final step of of the proof is to conclude from (1.163) that

$$\sup_{\omega_L \leq \omega \leq \omega_H} \log|S(\mathrm{j}\omega)| \geq \frac{1}{\omega_H - \omega_L}\left(\omega_L \log\frac{1}{\alpha} - \frac{3\varepsilon\omega_H}{2k}\right). \tag{1.168}$$

This is what we set out to prove. ∎

1.10.4 Limits of performance

The proof of the Freudenberg-Looze equality of Summary 1.7.1 relies on Poisson's integral formula from complex function theory.

Summary 1.10.4 (Poisson's integral formula). Let F be a function $\mathbb{C} \to \mathbb{C}$ that is analytic in the closed right-half plane[24] and is such that

$$\lim_{R\to\infty} \frac{|F(R\,\mathrm{e}^{\mathrm{j}\theta})|}{R} = 0 \tag{1.169}$$

for all $\theta \in [-\frac{\pi}{2}, -\frac{\pi}{2}]$. Then the value of $F(s)$ at any point $s = x + \mathrm{j}y$ in the open right-half plane[25] can be recovered from the values of $F(\mathrm{j}\omega)$, $\omega \in \mathbb{R}$, by the integral relation

$$F(s) = \frac{1}{\pi}\int_{-\infty}^{\infty} F(\mathrm{j}\omega)\, \frac{x}{x^2 + (y-\omega)^2}\, d\omega. \tag{1.170}$$

A sketch of the proof of Poisson's integral formula follows.

Proof of Poisson's integral formula. We present a brief proof of Poisson's integral formula based on elementary properties of the Laplace and Fourier transforms (see for instance Kwakernaak and Sivan (1991)). Since by assumption the function F is analytic in the closed right-half plane, its inverse Laplace transform f is zero on $(-\infty, 0)$. Hence, for $s = x + \mathrm{j}y$ we may write

$$F(s) = \int_{-\infty}^{\infty} f(t)\,\mathrm{e}^{-(x+\mathrm{j}y)t}\, dt = \int_{-\infty}^{\infty} f(t)\,\mathrm{e}^{-x|t|}\, e^{-\mathrm{j}yt}\, dt. \tag{1.171}$$

[24] For rational F this means that F has no poles in $\mathrm{Re}\,(s) \geq 0$.

[25] That is, for $\mathrm{Re}\,(s) > 0$.

For $x > 0$ the function $e^{-x|t|}$, $t \in \mathbb{R}$, is the inverse Fourier transform of the frequency function

$$\frac{1}{x - j\omega} + \frac{1}{x + j\omega} = \frac{2x}{x^2 + \omega^2}, \quad \omega \in \mathbb{R}. \tag{1.172}$$

Thus, we have

$$\begin{aligned} F(s) &= \int_{-\infty}^{\infty} f(t)\, e^{-jyt} \int_{-\infty}^{\infty} \frac{2x}{x^2 + \omega^2}\, e^{j\omega t}\, \frac{d\omega}{2\pi}\, dt \\ &= \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{x}{x^2 + \omega^2} \underbrace{\int_{-\infty}^{\infty} f(t)\, e^{-j(y-\omega)t}\, dt}_{F(j(y-\omega))}\, d\omega \end{aligned} \tag{1.173}$$

By replacing the integration variable ω with $y - \omega$ we obtain the desired result

$$F(s) = \frac{1}{\pi} \int_{-\infty}^{\infty} F(j\omega) \frac{x}{x^2 + (y - \omega)^2}\, d\omega. \tag{1.174}$$

■

We next consider the Freudenberg-Looze equality of Summary 1.7.1.

Proof 1.10.5 (Freudenberg-Looze equality). The proof of the Freudenberg-Looze equality of Summary 1.7.1 follows that of Freudenberg and Looze (1988). We first write L as $L = N/D$, with N and D coprime polynomials[26]. Then

$$S = \frac{D}{D + N}. \tag{1.175}$$

Since by assumption the closed-loop system is stable, the denominator $D + N$ has all its roots in the open left-half plane. Hence, S is analytic in the closed right-half plane. Moreover, any right-half plane pole z of L is a root of D and, hence, a zero of S.

We should like to apply Poisson's formula to the logarithm of the sensitivity function. Because of the right-half plane roots p_i of D, however, $\log S$ is not analytic in the right-half plane, and Poisson's formula cannot be used. To remedy this we *cancel* the right-half plane zeros of S by considering

$$\tilde{S} = B_{\text{poles}}^{-1} S. \tag{1.176}$$

Application of Poisson's formula to $\log \tilde{S}$ yields

$$\log \tilde{S}(s) = \frac{1}{\pi} \int_{-\infty}^{\infty} \log(\tilde{S}(j\omega))\, \frac{x}{x^2 + (y - \omega)^2}\, d\omega \tag{1.177}$$

for any open right-half plane point $s = x + jy$. Taking the real parts of the left- and right-hand sides we have

$$\log |\tilde{S}(s)| = \frac{1}{\pi} \int_{-\infty}^{\infty} \log(|\tilde{S}(j\omega)|)\, \frac{x}{x^2 + (y - \omega)^2}\, d\omega \tag{1.178}$$

Now replace s with a right-half plane zero $z = x + jy$ of L, that is, a right-half plane zero of N. Then

$$S(z) = \frac{1}{1 + L(z)} = 1, \tag{1.179}$$

so that $\tilde{S}(z) = B_{\text{poles}}^{-1}(z)$. Furthermore, $|B_{\text{poles}}(j\omega)| = 1$ for $\omega \in \mathbb{R}$, so that $|\tilde{S}(j\omega)| = |S(j\omega)|$ for $\omega \in \mathbb{R}$. Thus, after setting $s = z$ we may reduce (1.178) to

$$\log |B_{\text{poles}}^{-1}(z)| = \frac{1}{\pi} \int_{-\infty}^{\infty} \log(|S(j\omega)|)\, \frac{x}{x^2 + (y - \omega)^2}\, d\omega, \tag{1.180}$$

which is what we set out to prove.

[26] That is, N and D have no common factors.

Bode's sensitivity integral (1.114) follows from Proof 1.10.5.

Proof 1.10.6 (Bode's sensitivity integral). The starting point for the proof of Bode's sensitivity integral (1.114) is (1.178). Setting $y = 0$, replacing $\tilde{S}$ with $B_{\text{poles}}^{-1}S$, and multiplying on the left and the right by πx we obtain (exploiting the fact that $|B_{\text{poles}}| = 1$ on the imaginary axis)

$$\int_{-\infty}^{\infty} \log(|S(\mathrm{j}\omega)|)\, \frac{x^2}{x^2+\omega^2}\, d\omega = \pi x \log|S(x)| + \pi x \log|B_{\text{poles}}^{-1}(x)|. \tag{1.181}$$

Letting x approach ∞, the left-hand side of this expression approaches the Bode integral, while under the assumption that L has pole excess two the quantity $x \log|S(x)|$ approaches 0. Finally,

$$\begin{aligned} \lim_{x\to\infty} x \log|B_{\text{poles}}^{-1}(x)| &= \lim_{x\to\infty} x \log \prod_i \left|\frac{\bar{p}_i + x}{p_i - x}\right| = \lim_{x\to\infty} \operatorname{Re} \sum_i x \log \frac{1+\frac{\bar{p}_i}{x}}{1-\frac{p_i}{x}} \\ &= 2\sum_i \operatorname{Re}\ p_i. \end{aligned} \tag{1.182}$$

This completes the proof.

2

Classical Control System Design

Overview *– Classical criteria for the performance of feedback control systems are the error constants and notions such as bandwidth and peaking of the closed-loop frequency response, and rise time, settling time and overshoot of the step response.*

The graphical tools Bode, Nyquist, Nichols plots, and M-, N- and root loci belong to the basic techniques of classical and modern control.

Important classical control system design methods consist of loop shaping by lag compensation (including integral control), lead compensation and lag-lead compensation. Quantitative feedback design (QFT) allows to satisfy quantitative bounds on the performance robustness.

2.1 Introduction

In this chapter we review methods for the design of control systems that are known under the name of *classical control theory*. The main results in classical control theory emerged in the period 1930–1950, the initial period of development of the field of feedback and control engineering. The methods obtained a degree of maturity during the fifties and continue to be of great importance for the practical design of control systems, especially for the case of single-input, single-output linear control systems. Much of what now is called modern robust control theory has its roots in these classical results.

The historical development of the "classical" field started with H. Nyquist's stability criterion (Nyquist 1932), H. S. Black's analysis of the feedback amplifier (Black 1934), H. W. Bode's frequency domain analysis (Bode 1940), and W. R. Evans' root locus method (Evans 1948). To an extent these methods are of a *heuristic* nature, which both accounts for their success and for their limitations. With these techniques a designer attempts to synthesize a compensation network or controller that makes the closed-loop system perform as required. The terminology in use in that era (with expressions such as "synthesize," "compensation," and "network") is from the field of amplifier circuit design (Boyd and Barratt 1991).

In this chapter an impression is given of some of the classical highlights of control. The presentation is far from comprehensive. More extensive introductions may be found in classical and modern textbooks, together with an abundance of additional material. Well-known sources are for instance Bode (1945), James, Nichols, and Philips (1947), Evans (1954), Truxal (1955), Savant (1958), Horowitz (1963), Ogata (1970), Thaler (1973), Franklin, Powell, and Emami-Naeini (1986), D'Azzo and Houpis (1988), Van de Vegte (1990), Franklin, Powell, and Emami-Naeini (1991), and Dorf (1992).

In § 2.2 (p. 60) we discuss the steady-state error properties of feedback control systems. This naturally leads to review of the notion of integral control in § 2.3 (p. 64).

The main emphasis in classical control theory is on frequency domain methods. In § 2.4 (p. 69) we review various important classical graphic representations of frequency responses: Bode, Nyquist and Nichols plots.

The design goals and criteria of classical control theory are considered in § 2.5 (p. 79). In § 2.6 (p. 82) the basic classical techniques of lead, lag and lag-lead compensation are discussed. A brief survey of root locus theory as a method for parameter selection of compensators is presented in § 2.7 (p. 88). The historically interesting Guillemin-Truxal design procedure is considered in § 2.8 (p. 89). In the 1970s *quantitative feedback theory* (QFT) was initiated by Horowitz (1982). This powerful extension of the classical frequency domain feedback design methodology is explained in § 2.9 (p. 93).

All the design methods are *model-based*. They rely on an underlying and explicit model of the plant that needs to be controlled. The experimental Ziegler-Nichols rules for tuning a PID-controller mentioned in § 2.3 (p. 64) form an exception.

2.2 Steady state error behavior

2.2.1 Steady-state behavior

One of the fundamental reasons for adding feedback control to a system is that *steady-state errors*are reduced by the action of the control system. Consider the typical single-loop control system of Fig. 2.1. We analyze the steady-state behavior of this system, that is, the asymptotic behavior in the time domain for $t \to \infty$ when the reference input r is a polynomial time function of degree n. Thus,

$$r(t) = \frac{t^n}{n!}\mathbb{1}(t), \qquad t \geq 0, \tag{2.1}$$

where $\mathbb{1}$ is the *unit step* function, $\mathbb{1}(t) = 1$ for $t \geq 0$ and $\mathbb{1}(t) = 0$ for $t < 0$. For $n = 0$ we have a step of unit amplitude, for $n = 1$ the input is a ramp with unit slope and for $n = 2$ it is a parabola with unit second derivative.

The Laplace transform of the reference input is $\hat{r}(s) = 1/s^{n+1}$. The control error is the signal

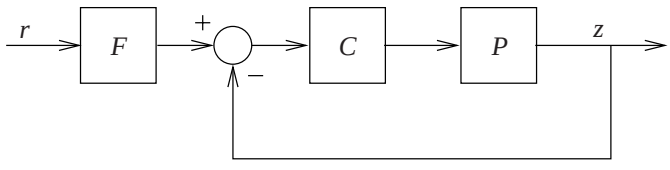


Figure 2.1: Single-loop control system configuration

ε defined by

$$\varepsilon(t) = r(t) - y(t), \qquad t \geq 0. \tag{2.2}$$

The steady-state error, *if it exists*, is

$$\varepsilon_\infty = \lim_{t\to\infty} \varepsilon(t). \tag{2.3}$$

The Laplace transform of the control error is

$$\hat{\varepsilon}(s) = \hat{r}(s) - \hat{y}(s) = \frac{1}{s^{n+1}} - \frac{H(s)}{s^{n+1}} = \frac{1 - H(s)}{s^{n+1}}. \tag{2.4}$$

The function

$$H = \frac{PCF}{1 + PC} = \frac{L}{1 + L}F \tag{2.5}$$

is the closed-loop transfer function. $L = PC$ is the loop gain.

Assuming that the closed-loop system is stable, so that all the poles of H are in the left-half plane, we may apply the final value theorem of Laplace transform theory. It follows that the steady-state error, *if it exists*, is given by

$$\varepsilon_\infty^{(n)} = \lim_{s\downarrow 0} s\hat{\varepsilon}(s) = \lim_{s\downarrow 0} \frac{1 - H(s)}{s^n}, \tag{2.6}$$

with n the order of the polynomial reference input. Assume for the moment that no prefilter is installed, that is, if $F(s) = 1$. Then

$$1 - H(s) = \frac{1}{1 + L(s)}, \tag{2.7}$$

and the steady-state error is

$$\varepsilon_\infty^{(n)} = \lim_{s\downarrow 0} \frac{1}{s^n[1 + L(s)]}. \tag{2.8}$$

This equation allows to compute the steady-state error of the response of the closed-loop system to the polynomial reference input (2.1).

2.2.2 Type k systems

A closed-loop system with loop gain L is *of type* k if for some integer k the limit $\lim_{s\downarrow 0} s^k L(s)$ exists and is nonzero. The system is of type k if and only if the loop gain L has exactly k poles in the origin. If the system is of type k then

$$\lim_{s\downarrow 0} s^n L(s) \begin{cases} = \infty & \text{for } 0 \leq n < k, \\ \neq 0 & \text{for } n = k, \\ = 0 & \text{for } n > k. \end{cases} \tag{2.9}$$

Consequently, from (2.8) we have for a type k system without prefilter (that is, if $F(s) = 1$)

$$\lim_{s\downarrow 0} \varepsilon_\infty^{(n)} \begin{cases} = 0 & \text{for } 0 \leq n < k, \\ \neq 0 & \text{for } n = k, \\ = \infty & \text{for } n > k. \end{cases} \tag{2.10}$$

Hence, if the system is of type k and stable then it has a zero steady-state error for polynomial reference inputs of order less than k, a nonzero finite steady-state error for an input of order k, and an infinite steady-state error for inputs of order greater than k.

A type 0 system has a nonzero but finite steady-state error for a step reference input, and an infinite steady-state error for ramp and higher-order inputs.

A type 1 system has zero steady-state error for a step input, a finite steady-state error for a ramp input, and infinite steady-state error for inputs of order two or higher.

A type 2 system has zero steady-state error for step and ramp inputs, a finite steady-state error for a second-order input, and infinite steady-state error for inputs of order three or higher.

Figure 2.2 illustrates the relation between system type and steady-state errors.

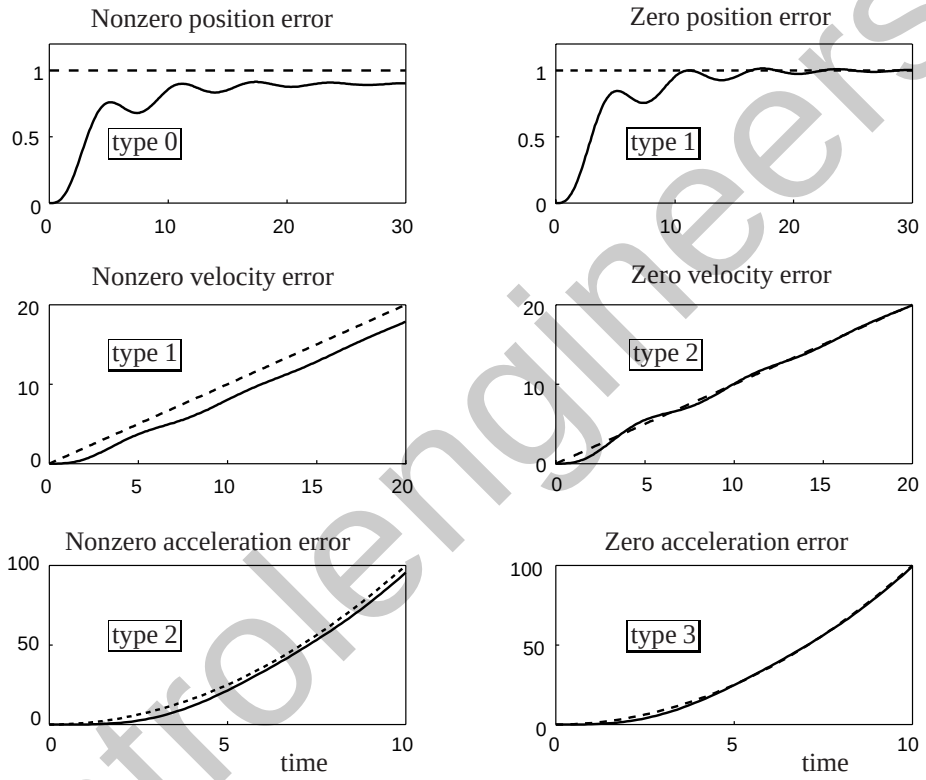


Figure 2.2: System type and steady-state errors

Exercise 2.2.1 (Type k systems with prefilter). The results of this subsection have been proved if no prefilter is used, that is, $F(s) = 1$. What condition needs to be imposed on F in case a prefilter is used? □

2.2.3 Error constants

If the system is of type 0 then the steady-state error for step inputs is

$$\varepsilon_\infty^{(0)} = \frac{1}{1+L(0)} = \frac{1}{1+K_p}. \tag{2.11}$$

The number $K_p = L(0)$ is the *position error constant.*

Table 2.1: Steady-state errors

System	Input		
	step	ramp	parabola
type 0	$\frac{1}{1+K_p}$	∞	∞
type 1	0	$\frac{1}{K_v}$	∞
type 2	0	0	$\frac{1}{K_a}$

For a type 1 system the steady-state error for a ramp input is

$$\varepsilon_\infty^{(1)} = \lim_{s\downarrow 0} \frac{1}{s[1+L(s)]} = \lim_{s\downarrow 0} \frac{1}{sL(s)} = \frac{1}{K_v}. \tag{2.12}$$

The number $K_v = \lim_{s\downarrow 0}\ sL(s)$ is the *velocity constant.*

The steady-state error of a type 2 system to a second-order input is

$$\varepsilon_\infty^{(2)} = \lim_{s\downarrow 0} \frac{1}{s^2 L(s)} = \frac{1}{K_a}. \tag{2.13}$$

The number $K_a = \lim_{s\downarrow 0}\ s^2 L(s)$ is the *acceleration constant.*

Table 2.1 summarizes the various steady-state errors. In each case, the larger the error constant is the smaller is the corresponding steady-state error.

The position, velocity and acceleration error provide basic requirements that must be satisfied by servomechanism systems depending upon their functional role.

The steady-state error results are robust in the sense that if the coefficients of the transfer function of the system vary then the error constants also vary but the zero steady-state error properties are preserved — as long as the system does not change its type. Thus, for a type 1 system the velocity error constant may vary due to parametric uncertainty in the system. As long as the type 1 property is not destroyed the system preserves its zero steady-state position error.

Exercise 2.2.2 (Steady-state response to polynomial disturbance inputs). Consider the feedback loop with disturbance input of Fig. 2.3. Suppose that the closed-loop system is stable, and that the disturbance is polynomial of the form

$$v(t) = \frac{t^n}{n!}\mathbb{1}(t), \qquad t \geq 0. \tag{2.14}$$

Show that the steady-state response of the output is given by

$$z_\infty^n = \lim_{t\to\infty} z(t) = \lim_{s\downarrow 0} \frac{1}{s^n[1+L(s)]}. \tag{2.15}$$

This formula is identical to the expression (2.8) for the steady-state error of the response to a polynomial reference input.

It follows that a type k system has a zero steady-state response to polynomial disturbances of order less than k, a nonzero finite steady-state response for polynomial disturbances of order k, and an infinite steady-state response for polynomial disturbances of order greater than k. □

Exercise 2.2.3 (Steady-state errors and closed-loop poles and zeros). Suppose that the closed-loop transfer function G_{cl} of (2.5) is expanded in factored form as

$$G_{\mathrm{cl}}(s) = k\frac{(s+z_1)(s+z_2)\cdots(s+z_m)}{(s+p_1)(s+p_2)\cdots(s+p_n)}. \tag{2.16}$$

1. Prove that the position error constant may be expressed as

$$K_p = \frac{k\prod_{j=1}^{m} z_j}{\prod_{j=1}^{n} p_j - k\prod_{j=1}^{m} z_j}. \tag{2.17}$$

2. Suppose that $K_p = \infty$, that is, the system has zero position error. Prove that the velocity and acceleration error constants may be expressed as

$$\frac{1}{K_v} = \sum_{j=1}^{n}\frac{1}{p_j} - \sum_{j=1}^{m}\frac{1}{z_j} \tag{2.18}$$

and

$$\frac{1}{K_a} = \frac{1}{2}\left(\sum_{j=1}^{m}\frac{1}{z_j^2} - \sum_{j=1}^{n}\frac{1}{p_j^2} - \frac{1}{K_v^2}\right), \tag{2.19}$$

respectively. *Hint:* Prove that $1/K_v = -G'_{\mathrm{cl}}(0)$ and $1/K_a = -G''_{\mathrm{cl}}(0)/2$, with the prime denoting the derivative. Next differentiate $\ln G_{\mathrm{cl}}(s)$ twice with respect to s at $s = 0$, with G_{cl} given by (2.16).

These results originate from Truxal (1955).

The relations (2.18) and (2.19) represent the connection between the error constants and the system response characteristics. We observe that the further the closed-loop poles are from the origin, the larger the velocity constant K_v is. The velocity constant may also be increased by having closed-loop zeros close to the origin. □

2.3 Integral control

Integral control is a remarkably effective classical technique to achieve low-frequency disturbance attenuation. It moreover has a useful robustness property.

Disturbance attenuation is achieved by making the loop gain large. The loop gain may be made large at low frequencies, and indeed infinite at zero frequency, by including a factor $1/s$ in the loop gain $L(s) = P(s)C(s)$. If $P(s)$ has no "natural" factor $1/s$ then this is accomplished by including the factor in the compensator transfer function C by choosing

$$C(s) = \frac{C_0(s)}{s}. \tag{2.20}$$

The rational function C_0 remains to be selected. The compensator $C(s)$ may be considered as the series connection of a system with transfer function $C_0(s)$ and another with transfer function $1/s$. Because a system with transfer function $1/s$ is an integrator, a compensator of this type is said to have *integrating action*.

If the loop gain $L(s)$ has a factor $1/s$ then in the terminology of § 2.2 (p. 60) the system is of type 1. Its response to a step reference input has a zero steady-state error.

Obviously, if $L_0(s)$ contains no factor s then the loop gain

$$L(s) = \frac{L_0(s)}{s} \tag{2.21}$$

is infinite at zero frequency and very large at low frequencies. As a result, the sensitivity function S, which is given by

$$S(s) = \frac{1}{1+L(s)} = \frac{1}{1+\frac{L_0(s)}{s}} = \frac{s}{s+L_0(s)}, \tag{2.22}$$

is zero at zero frequency and, by continuity, small at low frequencies. The fact that S is zero at zero frequency implies that zero frequency disturbances, that is, constant disturbances, are completely eliminated. This is called *constant disturbance rejection*.

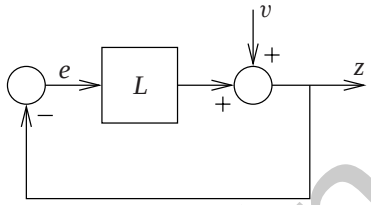


Figure 2.3: Feedback loop

Exercise 2.3.1 (Rejection of constant disturbances). Make the last statement clearer by proving that if the closed-loop system of Fig. 2.3 has integrating action and is stable, then its response z (from any initial condition) to a constant disturbance $v(t) = 1$, $t \geq 0$, has the property

$$\lim_{t\to\infty} z(t) = 0. \tag{2.23}$$

Hint: Use Laplace transforms and the final value property. □

The constant disturbance rejection property is *robust* with respect to plant and compensator perturbations as long as the perturbations are such that

- the loop gain still contains the factor $1/s$, and
- the closed-loop system remains stable.

In a feedback system with integrating action, the transfer function of the series connection of the plant and compensator contains a factor $1/s$. A system with transfer function $1/s$ is capable of generating a constant output with zero input. Hence, the series connection may be thought of as containing a *model* of the mechanism that generates constant disturbances, which are precisely the disturbances that the feedback system rejects. This notion has been generalized (Wonham 1979a) to what is known as the *internal model principle*. This principle states that if a feedback system is to reject certain disturbances then it should contain a model of the mechanism that generates the disturbances.

Exercise 2.3.2 (Type k control). The loop gain of a type k system contains a factor $1/s^k$, with k a positive integer. Prove that if a type k closed-loop system as in Fig. 2.3 is stable then it rejects disturbances of the form

$$v(t) = \frac{t^n}{n!}, \quad t \geq 0, \tag{2.24}$$

with n any nonnegative integer such that $n \leq k-1$. "Rejects" means that for such disturbances $\lim_{t\to\infty} z(t) = 0$. □

Compensators with integrating action are easy to build. Their effectiveness in achieving low frequency disturbance attenuation and the robustness of this property make "integral control" a popular tool in practical control system design. The following variants are used:

- *Pure integral control*, with compensator transfer function

$$C(s) = \frac{1}{sT_i}. \tag{2.25}$$

  The single design parameter T_i (called the *reset time*) does not always allow achieving closed-loop stability or sufficient bandwidth.

- *Proportional and integral control*, also known as *PI control*, with compensator transfer function

$$C(s) = g\left(1 + \frac{1}{sT_i}\right), \tag{2.26}$$

  gives considerably more flexibility.

- *PID* (proportional-integral-derivative) *control*, finally, is based on a compensator transfer function of the type

$$C(s) = g\left(sT_d + 1 + \frac{1}{sT_i}\right). \tag{2.27}$$

  T_d is the *derivative time*. The derivative action may help to speed up response but tends to make the closed-loop system less robust for high frequency perturbations.

  Derivative action technically cannot be realized. In any case it would be undesirable because it greatly amplifies noise at high frequencies. Therefore the derivative term sT_d in (2.27) in practice is replaced with a "tame" differentiator

$$\frac{sT_d}{1+sT}, \tag{2.28}$$

  with T a small time constant.

Standard PID controllers are commercially widely available. In practice they are often tuned experimentally with the help of the rules developed by Ziegler and Nichols (see for instance Franklin, Powell, and Emami-Naeini (1991)). The Ziegler-Nichols rules (Ziegler and Nichols 1942) were developed under the assumption that the plant transfer function is of a well-damped low-pass type. When tuning a P-, PI- or PID-controller according to the Ziegler-Nichols rules first a P-controller is connected to the plant. The controller gain g is increased until undamped oscillations occur. The corresponding gain is denoted as as g_0 and the period of oscillation as T_0. Then the parameters of the PID-controller are given by

$$\begin{array}{rlll} \text{P-controller:} & g = 0.5g_0, & T_i = \infty, & T_d = 0, \\ \text{PI-controller:} & g = 0.45g_0, & T_i = 0.85T_0, & T_d = 0, \\ \text{PID-controller:} & g = 0.6g_0, & T_i = 0.5T_0, & T_d = 0.125T_0. \end{array}$$

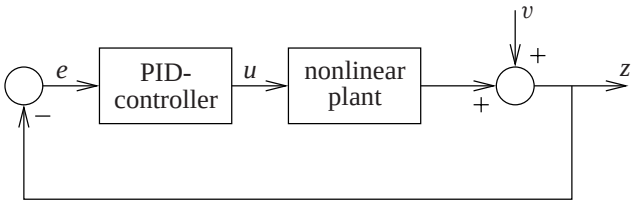


Figure 2.4: Nonlinear plant with integral control

The corresponding closed-loop system has a relative damping of about 0.25, and its closed-loop step response to disturbances has a pak value of about 0.4. Normally experimental fine tuning is needed to obtain the best results.

Integral control also works for nonlinear plants. Assume that the plant in the block diagram of Fig. 2.4 has the reasonable property that for every constant input u_0 there is a unique constant steady-state output w_0, and that the plant may maintain any constant output w_0. The "integral controller" (of type I, PI or PID) has the property that it maintains a constant output u_0 if and only if its input e is zero. Hence, if the disturbance is a constant signal v_0 then the closed-loop system is in steady-state if and only if the error signal e is zero. Therefore, if the closed-loop system is stable then it rejects constant disturbances.

Example 2.3.3 (Integral control of the cruise control system). The linearized cruise control system of Example 1.2.1 (p. 3) has the linearized plant transfer function

$$P(s) = \frac{\frac{1}{T}}{s + \frac{1}{\theta}}. \tag{2.29}$$

If the system is controlled with pure integral control

$$C(s) = \frac{1}{sT_i} \tag{2.30}$$

then the loop gain and sensitivity functions are

$$L(s) = P(s)C(s) = \frac{\frac{1}{TT_i}}{s(s + \frac{1}{\theta})}, \qquad S(s) = \frac{1}{1 + L(s)} = \frac{s(s + \frac{1}{\theta})}{s^2 + \frac{1}{\theta}s + \frac{1}{TT_i}}. \tag{2.31}$$

The roots of the denominator polynomial

$$s^2 + \frac{1}{\theta}s + \frac{1}{TT_i} \tag{2.32}$$

are the closed-loop poles. Since θ, T and T_i are all positive these roots have negative real parts, so that the closed-loop system is stable. Figure 2.5 shows the loci of the roots as T_i varies from ∞ to 0 (see also § 2.7 (p. 88)). Write the closed-loop denominator polynomial (2.32) as $s^2 + 2\zeta_0\omega_0 s + \omega_0^2$, with ω_0 the resonance frequency and ζ_0 the relative damping. It easily follows that

$$\omega_0 = \frac{1}{\sqrt{TT_i}}, \qquad \zeta_0 = \frac{1/\theta}{2\omega_0} = \frac{\sqrt{TT_i}}{2\theta}. \tag{2.33}$$

The best time response is obtained for $\zeta_0 = \frac{1}{2}\sqrt{2}$, or

$$T_i = \frac{2\theta^2}{T}. \tag{2.34}$$

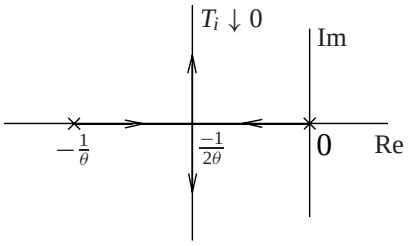


Figure 2.5: Root loci for the cruise control system. The $\times$s mark the open-loop poles

If $T = \theta = 10$ [s] (corresponding to a cruising speed of 50% of the top speed) then $T_i = 20$ [s]. It follows that $\omega_0 = 1/\sqrt{200} \approx 0.07$ [rad/s]. Figure 2.6 shows the Bode magnitude plot of the resulting sensitivity function. The plot indicates that constant disturbance rejection is obtained as well as low-frequency disturbance attenuation, but that the closed-loop bandwidth is not greater than the bandwidth of about 0.1 [rad/s] of the open-loop system.

Increasing T_i decreases the bandwidth. Decreasing T_i beyond $2\theta^2/T$ does not increase the bandwidth but makes the sensitivity function S peak. This adversely affects robustness. Bandwidth improvement without peaking may be achieved by introducing proportional gain. (See Exercise 2.6.2, p. 85.) □

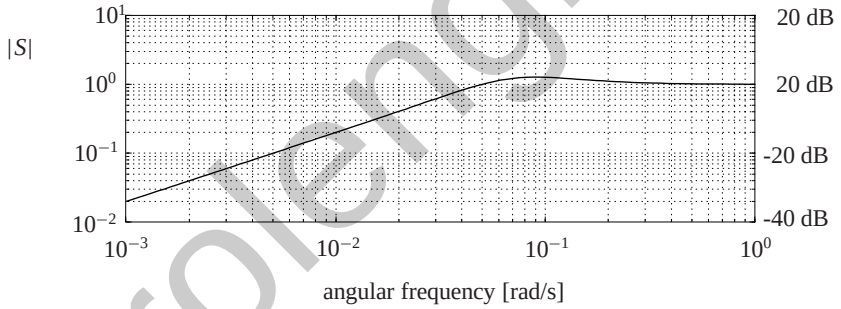


Figure 2.6: Magnitude plot of the sensitivity function of the cruise control system with integrating action

Exercise 2.3.4 (PI control of the cruise control system). Show that by PI control constant disturbance rejection and low-frequency disturbance attenuation may be achieved for the cruise control system with satisfactory gain and phase margins for any closed-loop bandwidth allowed by the plant capacity. □

Exercise 2.3.5 (Integral control of a MIMO plant). Suppose that Fig. 2.3 (p. 65) represents a stable MIMO feedback loop with rational loop gain matrix L such that also L^{-1} is a well-defined rational matrix function. Prove that the feedback loop rejects constant disturbances if and only if $L^{-1}(0) = 0$. □

Exercise 2.3.6 (Integral control and constant input disturbances). Not infrequently disturbances enter the plant at the input, as schematically indicated in Fig. 2.7. In this case the transfer

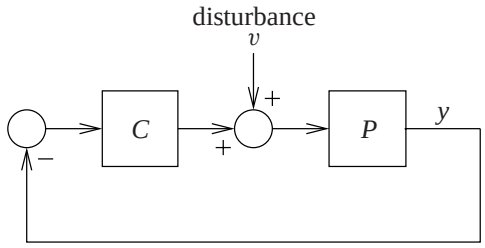


Figure 2.7: Plant with input disturbance

function from the disturbances v to the control system output z is

$$R = \frac{P}{1+PC}. \tag{2.35}$$

Suppose that the system has integrating action, that is, the loop gain $L = PC$ has a pole at 0. Prove that constant disturbances at the input are rejected if and only the integrating action is localized in the compensator. □

2.4 Frequency response plots

2.4.1 Introduction

In classical as in modern control engineering the graphical representation of frequency responses is an important tool in understanding and studying the dynamics of control systems. In this section we review three well-known ways of presenting frequency responses: Bode plots, Nyquist plots, and Nichols plots.

Consider a stable linear time-invariant system with input u, output y and transfer function L. A sinusoidal input $u(t) = \hat{u}\sin(\omega t)$, $t \geq 0$, results in the steady-state sinusoidal output $y(t) = \hat{y}\sin(\omega t + \phi)$, $t \geq 0$. The amplitude $\hat{y}$ and phase ϕ of the output are given by

$$\hat{y} = |L(\mathrm{j}\omega)|\,\hat{u}, \qquad \phi = \arg L(\mathrm{j}\omega). \tag{2.36}$$

The magnitude $|L(\mathrm{j}\omega)|$ of the frequency response function $L(\mathrm{j}\omega)$ is the *gain* at the frequency ω. Its argument $\arg L(\mathrm{j}\omega)$ is the *phase shift*.

Write

$$L(s) = k\frac{(s-z_1)(s-z_2)\cdots(s-z_m)}{(s-p_1)(s-p_2)\cdots(z-p_n)}, \tag{2.37}$$

with k a constant, $z_1, z_2, \cdots, z_m$ the zeros, and $p_1, p_2, \cdots, p_n$ the poles of the system. Then for any $s = \mathrm{j}\omega$ the magnitude and phase of $L(\mathrm{j}\omega)$ may be determined by measuring the vector lengths and angles from the pole-zero pattern as in Fig. 2.8. The magnitude of L follows by appropriately multiplying and dividing the lengths. The phase of L follows by adding and subtracting the angles.

A pole p_i that is close to the imaginary axis leads to a short vector length $s - p_i$ at values of ω in the neighborhood of Im p_i. This results in large magnitude $|L(\mathrm{j}\omega)|$ at this frequency, and explains why a pole that is close to the imaginary axis leads to a resonance peak in the frequency response.

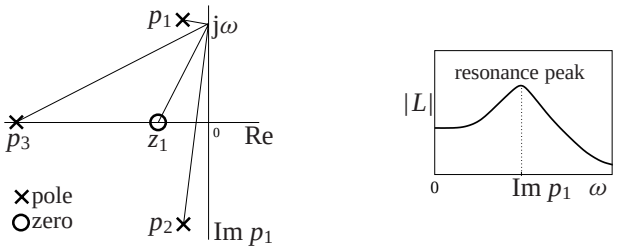


Figure 2.8: Frequency response determined from the pole-zero pattern

2.4.2 Bode plots

A frequency response function $L(\mathrm{j}\omega)$, $\omega \in \mathbb{R}$, may be plotted in two separate graphs, magnitude as a function of frequency, and phase as a function of frequency. When the frequency and magnitude scales are logarithmic the combined set of the two graphs is called the *Bode diagram* of L. Individually, the graphs are referred to as the *Bode magnitude plot* and the *Bode phase plot*. Figure 2.9 shows the Bode diagrams of a second-order frequency response function for different values of the relative damping.

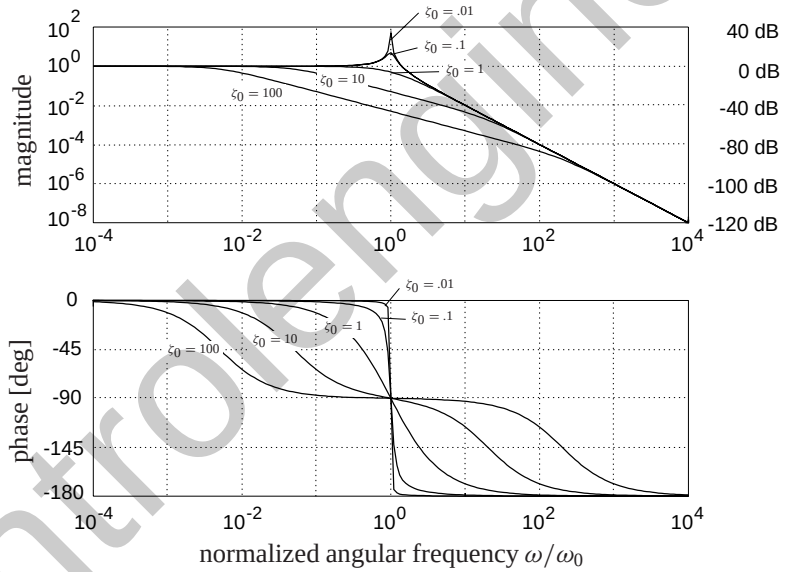


Figure 2.9: Bode diagram of the transfer function $\omega_0^2/(s^2 + 2\zeta_0\omega_0 + \omega_0^2)$ for various values of ζ_0

The construction of Bode diagrams for rational transfer functions follows simple steps. Write the frequency response function in the factored form

$$L(\mathrm{j}\omega) = k\frac{(\mathrm{j}\omega - z_1)(\mathrm{j}\omega - z_2)\cdots(\mathrm{j}\omega - z_m)}{(\mathrm{j}\omega - p_1)(\mathrm{j}\omega - p_2)\cdots(\mathrm{j}\omega - p_n)}. \tag{2.38}$$

It follows for the logarithm of the amplitude and for the phase

$$\log|L(\mathrm{j}\omega)| = \log|k| + \sum_{i=1}^{m}\log|\mathrm{j}\omega - z_i| - \sum_{i=1}^{n}\log|\mathrm{j}\omega - p_i|, \tag{2.39}$$

$$\arg L(\mathrm{j}\omega) = \arg k + \sum_{i=1}^{m} \arg(\mathrm{j}\omega - z_i) - \sum_{i=1}^{n} \arg(\mathrm{j}\omega - p_i). \tag{2.40}$$

The asymptotes of the individual terms of (2.39) and (2.40) may be drawn without computation. For a first-order factor $s + \omega_0$ we have

$$\log|\mathrm{j}\omega + \omega_0| \approx \begin{cases} \log|\omega_0| & \text{for } 0 \le \omega \ll |\omega_0|, \\ \log\omega & \text{for } \omega \gg |\omega_0|, \end{cases} \tag{2.41}$$

$$\arg(\mathrm{j}\omega + \omega_0) \approx \begin{cases} \arg(\omega_0) & \text{for } 0 \le \omega \ll |\omega_0|, \\ 90^\circ & \text{for } \omega \gg |\omega_0|. \end{cases} \tag{2.42}$$

The asymptotes for the doubly logarithmic amplitude plot are straight lines. The low frequency asymptote is a constant. The high frequency asymptote has slope 1 decade/decade. If the amplitude is plotted in decibels then the slope is 20 dB/decade. The amplitude asymptotes intersect at the frequency $|\omega_0|$. The phase moves from $\arg(\omega_0)$ (0° if ω_0 is positive) at low frequencies to 90° ($\pi/2$ rad) at high frequencies. Figure 2.10 shows the amplitude and phase curves for the first order factor and their asymptotes (for ω_0 positive).

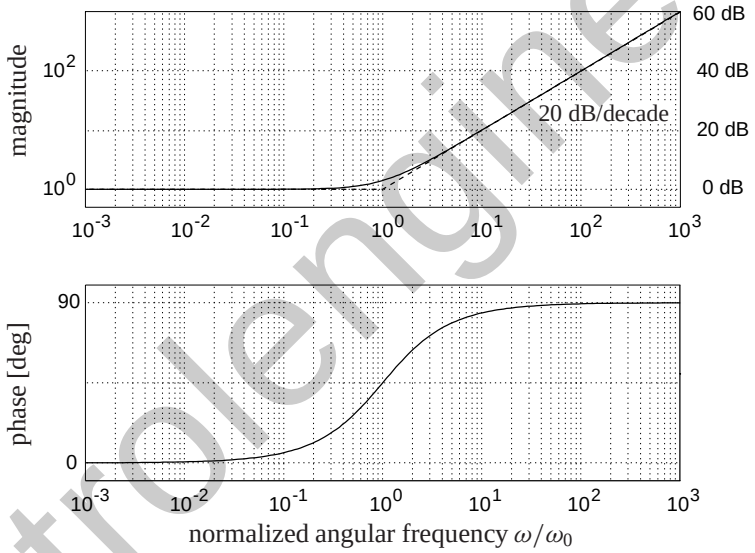


Figure 2.10: Bode amplitude and phase plots for the factor $s + \omega_0$. Dashed: low- and high-frequency asymptotes.

Factors corresponding to complex conjugate pairs of poles or zeros are best combined to a second-order factor of the form

$$s^2 + 2\zeta_0\omega_0 s + \omega_0^2. \tag{2.43}$$

Asymptotically,

$$\log|(\mathrm{j}\omega)^2 + 2\zeta_0\omega_0(\mathrm{j}\omega) + \omega_0^2| \approx \begin{cases} 2\log|\omega_0| & \text{for } 0 \le \omega \ll |\omega_0|, \\ 2\log\omega & \text{for } \omega \gg |\omega_0|, \end{cases} \tag{2.44}$$

$$\arg((\mathrm{j}\omega)^2 + 2\zeta_0\omega_0(\mathrm{j}\omega) + \omega_0^2) \approx \begin{cases} 0 & \text{for } 0 \le \omega \ll |\omega_0|, \\ 180^\circ & \text{for } \omega \gg |\omega_0|. \end{cases} \tag{2.45}$$

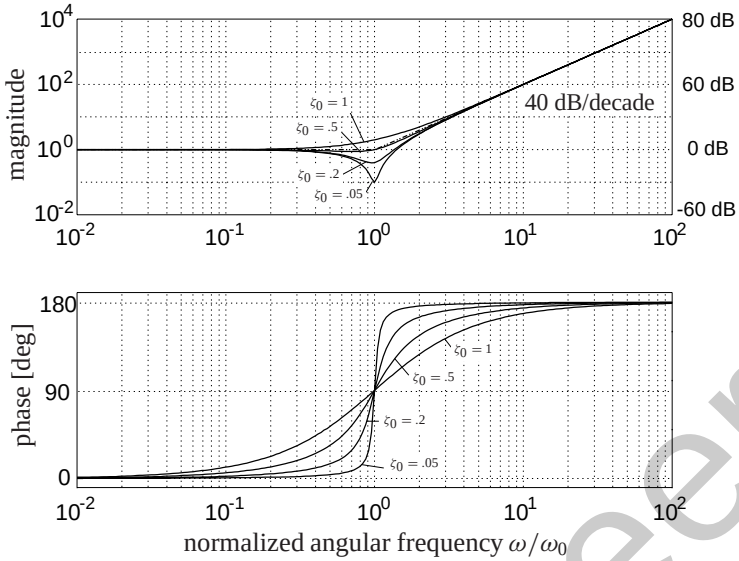


Figure 2.11: Bode plots for a second-order factor $s^2 + 2\zeta_0\omega_0 s + \omega_0^2$ for different values of the relative damping ζ_0. Dashed: low- and high-frequency asymptotes

The low frequency amplitude asymptote is again a constant. In the Bode magnitude plot the high frequency amplitude asymptote is a straight line with slope 2 decades/decade (40 dB/decade). The asymptotes intersect at the frequency $|\omega_0|$. The phase goes from 0° at low frequencies to 180° at high frequencies. Figure 2.11 shows amplitude and phase plots for different values of the relative damping ζ_0 (with ω_0 positive).

Bode plots of high-order transfer functions, in particular asymptotic Bode plots, are obtained by adding log magnitude and phase contributions from first- and second-order factors.

The "asymptotic Bode plots" of first- and second-order factors follow by replacing the low frequency values of magnitude and phase by the low frequency asymptotes at frequencies below the break frequency, and similarly using the high-frequency asymptotes above the break frequency. High-order asymptotic Bode plots follow by adding and subtracting the asymptotic plots of the first- and second-order factors that make up the transfer function.

As shown in Fig. 2.12 the gain and phase margins of a stable feedback loop may easily be identified from the Bode diagram of the loop gain frequency response function.

Exercise 2.4.1 (Complex conjugate pole pair). Consider the factor $s^2 + 2\zeta_0\omega_0 s + \omega_0^2$. The positive number ω_0 is the characteristic frequency and ζ_0 the relative damping.

1. Prove that for $|\zeta_0| < 1$ the roots of the factor are the complex conjugate pair

$$\omega_0 \left(-\zeta_0 \pm \mathrm{j}\sqrt{1 - \zeta_0^2} \right). \tag{2.46}$$

   For $|\zeta_0| \geq 1$ the roots are real.

2. Prove that in the diagram of Fig. 2.13 the distance of the complex conjugate pole pair to the origin is ω_0, and that the angle ϕ equals $\arccos \zeta_0$.

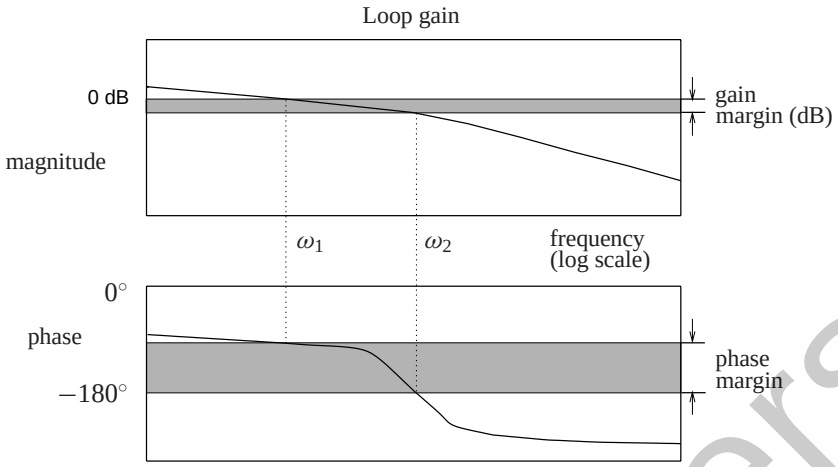


Figure 2.12: Gain and phase margins from the Bode plot of the loop gain

3. Assume that $0 < \zeta_0 < 1$. At which frequency has the amplitude plot of the factor $(\mathrm{j}\omega)^2 + 2\zeta_0\omega_0(\mathrm{j}\omega) + \omega_0^2$, $\omega \in \mathbb{R}$, its minimum (and, hence, has the amplitude plot of its reciprocal its maximum)? Note that this frequency is not precisely ω_0.

□

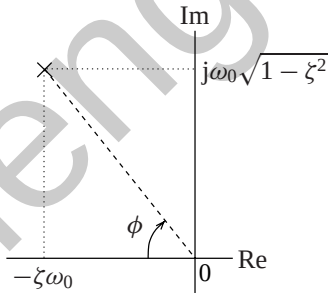


Figure 2.13: Complex conjugate root pair of the factor $s^2 + 2\zeta_0\omega_0 + \omega_0^2$

Exercise 2.4.2 (Transfer function and Bode plot). Consider the loop gain L whose Bode diagram is given in Fig. 2.14.

1. Use Bode's gain-phase relationship (§ 1.6, p. 34) to conclude from the Bode plot that the loop gain is (probably) minimum-phase, and, hence, stable. Next argue that the corresponding closed-loop system is stable.

2. Fit the Bode diagram as best as possible by a rational pole-zero representation.

The conclusion of (a) is correct as long as outside the frequency range considered the frequency response behaves in agreement with the low- and high-frequency asymptotes inferred from the plot. This is especially important for the low-frequency behavior. Slow dynamics that change the number of encirclements of the Nyquist plot invalidate the result. □

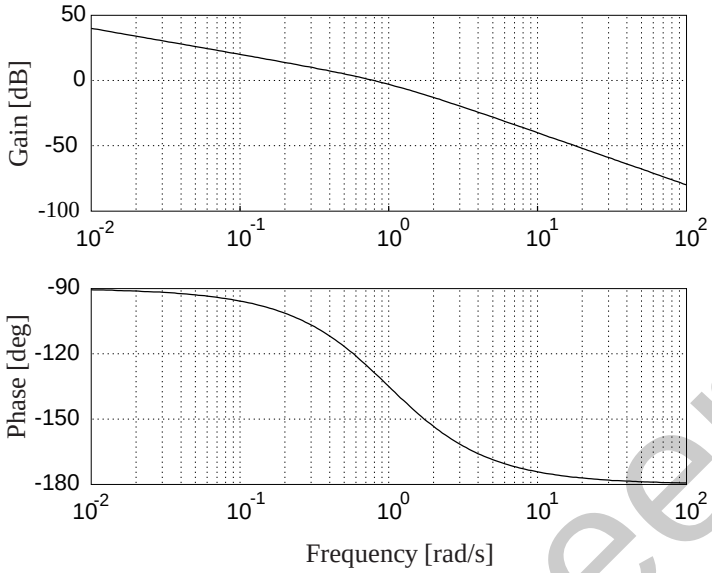


Figure 2.14: Bode diagram of a loop gain L

2.4.3 Nyquist plots

In § 1.3 (p. 11) we already encountered the *Nyquist plot*, which is a polar plot of the frequency response function with the frequency ω as parameter. If frequency is not plotted along the locus — a service that some packages fail to provide — then the Nyquist plot is less informative than the Bode diagram. Figure 2.15 shows the Nyquist plots of the second-order frequency response functions of Fig. 2.9.

Normally the Nyquist plot is only sketched for $0 \leq \omega < \infty$. The plot for negative frequencies follows by mirroring with respect to the real axis.

If L is strictly proper then for $\omega \to \infty$ the Nyquist plot approaches the origin. Write L in terms of its poles and zeros as in (2.37). Then asymptotically

$$L(\mathrm{j}\omega) \approx \frac{k}{(\mathrm{j}\omega)^{n-m}} \qquad \text{for} \qquad \omega \to \infty. \tag{2.47}$$

If k is positive then the Nyquist plot approaches the origin at an angle $-(n-m) \times 90^\circ$. The number $n-m$ is called the *pole-zero excess* or *relative degree* of the system.

In control systems of type k the loop gain L has a pole of order k at the origin. Hence, at low frequencies the loop frequency response asymptotically behaves as

$$L(\mathrm{j}\omega) \approx \frac{c}{(\mathrm{j}\omega)^k} \qquad \text{for} \qquad \omega \downarrow 0, \tag{2.48}$$

with c a real constant. If $k = 0$ then for $\omega \downarrow 0$ the Nyquist plot of L approaches a point on the real axis. If $k > 0$ and c is positive then the Nyquist plot goes to infinity at an angle $-k \times 90^\circ$. Figure 2.16 illustrates this.

Exercise 2.4.3 (Nyquist plots). Prove the following observations.

1. The shape of the Nyquist plot of $L(s) = 1/(1+sT)$ is a circle whose center and radius are independent of T.

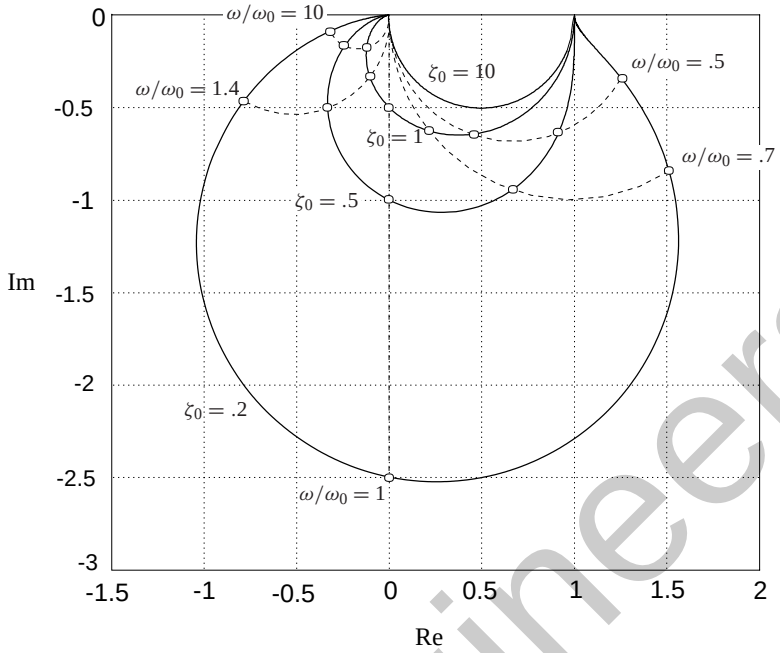
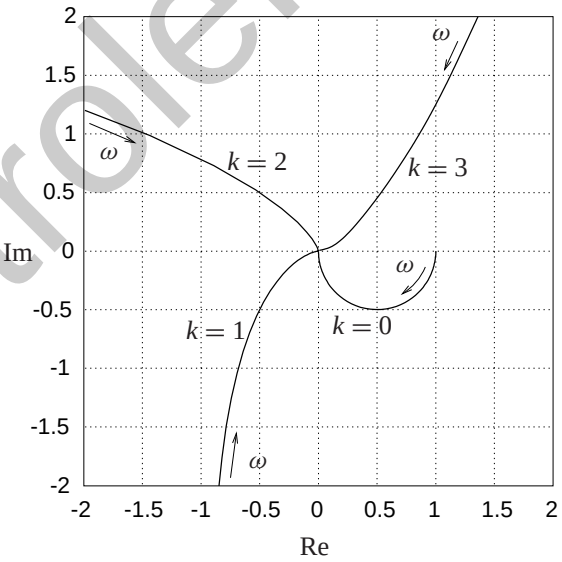


Figure 2.15: Nyquist plots of the transfer function $\omega_0^2/(s^2 + 2\zeta_0\omega_0 s + \omega_0^2)$ for different values of the relative damping ζ_0



$$L(s) = \frac{1}{s^k(1+s)}$$

Figure 2.16: Nyquist plots of the loop gain for different values of system type

2. The shape of the Nyquist plot of $L(s) = 1/(1+sT_1)(1+sT_2)$ only depends on the ratio T_1/T_2. The shape is the same for $T_1/T_2 = \alpha$ and $T_2/T_1 = \alpha$.

3. The shape of the Nyquist plot of $L(s) = \omega_0^2/(\omega_0^2 + 2\zeta\omega_0 s + s^2)$ is independent of ω_0.

□

2.4.4 M- and N-circles

Consider a simple unit feedback loop with loop gain L as in Fig. 2.17. The closed-loop transfer function of the system equals the complementary sensitivity function

$$H = T = \frac{L}{1+L}. \tag{2.49}$$

M- and N-circles are a graphical tool — typical for the classical control era — to determine the closed-loop frequency response function from the Nyquist plot of the loop gain L.

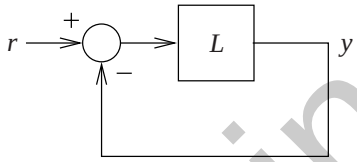


Figure 2.17: Unit feedback loop

An M-circle is the locus of points z in the complex plane where the magnitude of the complex number

$$\frac{z}{1+z} \tag{2.50}$$

is constant and equal to M. An M-circle has center and radius

$$\text{center} \quad \left(\frac{M^2}{1-M^2}, 0\right), \qquad \text{radius} \quad \left|\frac{M}{1-M^2}\right|. \tag{2.51}$$

An N-circle is the locus of points in the complex plane where the argument of the number (2.50) is constant and equal to $\arctan N$. An N-circle has center and radius

$$\text{center} \quad \left(-\frac{1}{2}, \frac{1}{2N}\right), \qquad \text{radius} \quad \frac{1}{2}\sqrt{1+\frac{1}{N^2}}. \tag{2.52}$$

Figure 2.18 shows the arrangement of M- and N-circles in the complex plane.

The magnitude of the closed-loop frequency response and complementary sensitivity function T may be found from the points of intersection of the Nyquist plot of the loop gain L with the M-circles. Likewise, the phase of T follows from the intersections with the N-circles.

Figure 2.18 includes a typical Nyquist plot of the loop gain L. These are some of the features of the closed-loop response that are obtained by inspection of the intersections with the M-circles:

- The height of the resonance peak is the maximum value of M encountered along the Nyquist plot.
- The resonance frequency ω_m is the frequency where this maximum occurs.

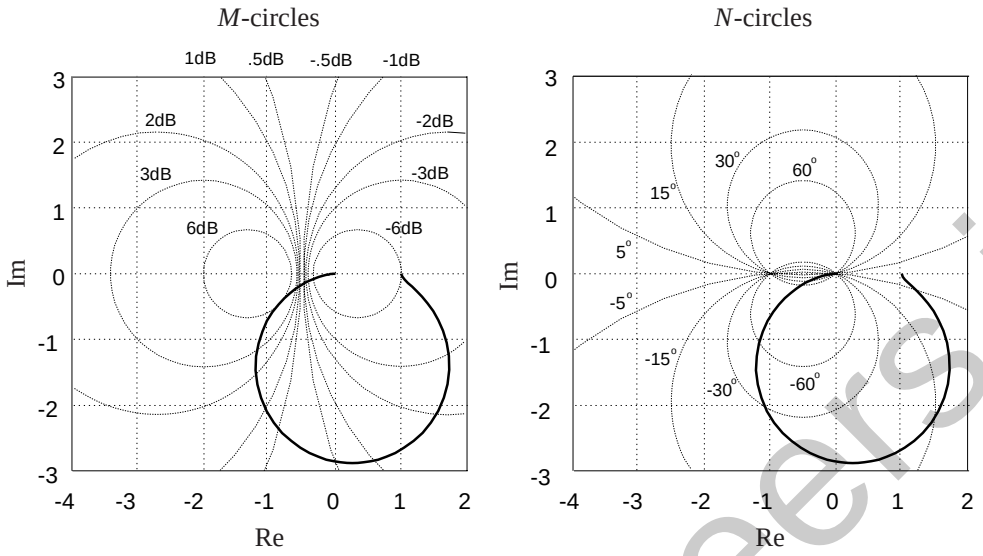


Figure 2.18: M- circles (left) and N-circles (right)

- The bandwidth is the frequency at which the Nyquist plot intersects the 0.707 circle (the -3 dB circle).

These and other observations provide useful indications how to modify and shape the loop frequency response to improve the closed-loop properties. The M- and N-loci are more often included in Nichols plots (see the next subsection) than in Nyquist plots.

Exercise 2.4.4 (M- and N-circles). Verify the formulas (2.51) and (2.52) for the centers and radii of the M- and N-circles. □

2.4.5 Nichols plots

The linear scales of the Nyquist plot sometimes obscure the large range of values over which the magnitude of the loop gain varies. Also, the effect of changing the compensator frequency response function C or the plant frequency response function P on the Nyquist plot of the loop gain $L = PC$ cannot always easily be predicted.

Both difficulties may be overcome by plotting the loop gain in the form of a *Nichols plot* (James, Nichols, and Philips 1947). A Nichols plot is obtained by plotting the log magnitude of the loop gain frequency response function versus its phase. In these coordinates, the M- en N-circles transform to M- and N- loci. The phase–log magnitude plane together with a set of M- and N-loci is called a *Nichols chart*. In Fig. 2.19 Nichols plots are given of the second-order frequency response functions whose Bode diagrams and Nyquist plots are shown in Figs. 2.9 (p. 70) and 2.15 (p. 75), respectively.

In a Nichols diagram, gain change corresponds to a vertical shift and phase change to a horizontal shift. This makes it easy to assess the effect of changes of the compensator frequency response function C or the plant frequency response function P on the loop gain $L = PC$.

Exercise 2.4.5 (Gain and phase margins in the Nichols plot). Explain how the gain margin and phase margin of a stable feedback loop may be identified from the Nichols plot. □

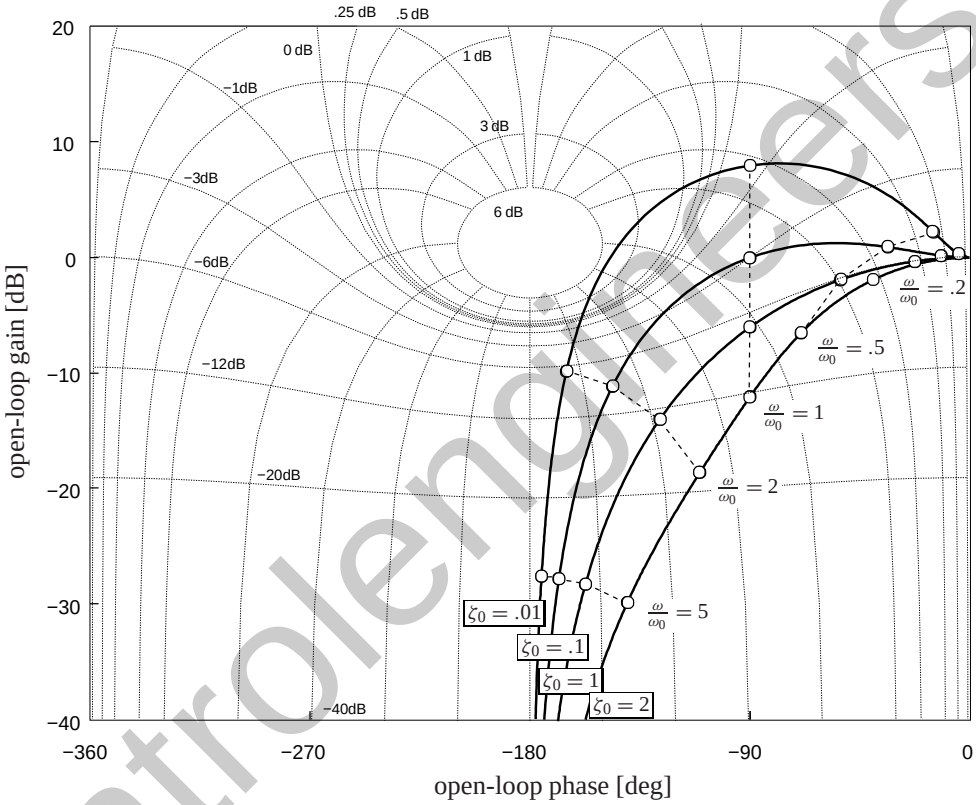


Figure 2.19: Nichols plots of the transfer function $\omega_0^2/(s^2 + 2\zeta_0\omega_0 s + \omega_0^2)$ for different values of ζ_0

Exercise 2.4.6 (Lag-lead compensator). Consider a compensator with the second-order transfer function

$$C(s) = \frac{(1+sT_1)(1+sT_2)}{(1+sT_1)(1+sT_2)+sT_{12}}. \tag{2.53}$$

T_1, T_2 and T_{12} are time constants. The corresponding frequency response function is

$$C(\mathrm{j}\omega) = \frac{(1-\omega^2 T_1 T_2)+\mathrm{j}\omega(T_1+T_2)}{(1-\omega^2 T_1 T_2)+\mathrm{j}\omega(T_1+T_2+T_{12})}, \qquad \omega \in \mathbb{R}. \tag{2.54}$$

By a proper choice of the time constants the network acts as a lag network (that is, subtracts phase) in the lower frequency range and as a lead network (that is, adds phase) in the higher frequency range.

Inspection of the frequency response function (2.54) shows that numerator and denominator simultaneously become purely imaginary at the frequency $\omega = 1/\sqrt{T_1 T_2}$. At this frequency the frequency response function is real. This frequency is the point where the character of the network changes from lag to lead, and where the magnitude of the frequency response is minimal.

1. Plot the Bode, Nyquist, and Nichols diagrams of this frequency response function.
2. Prove that the Nyquist plot of L has the shape of a circle in the right half of the complex plane with its center on the real axis. Since $C(0) = C(\mathrm{j}\infty) = 1$ the plot begins and ends in the point 1.

□

2.5 Classical control system design

2.5.1 Design goals and criteria

For SISO systems we have the following partial list of typical classical performance specifications. Consider the feedback loop of Fig. 2.20. These are the basic requirements for a well-

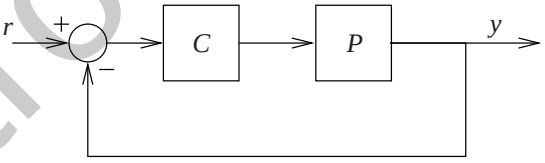


Figure 2.20: Basic feedback system

designed control system:

1. The transient response is sufficiently fast.
2. The transient response shows satisfactory damping.
3. The transient response satisfies accuracy requirements, often expressed in terms of the error constants of § 2.2 (p. 60).
4. The system is sufficiently insensitive to external disturbances and variations of internal parameters.

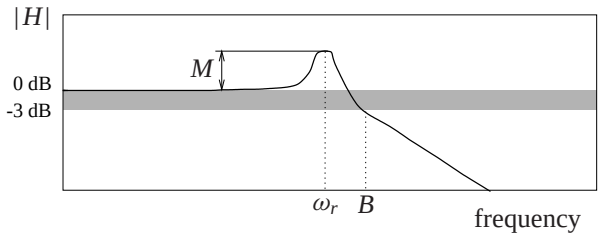


Figure 2.21: Frequency-domain performance quantities

These basic requirements may be further specified in terms of both a number of *frequency-domain specifications* and certain *time-domain specifications*.

Figures 2.12 (p. 73) and 2.21 illustrate several important frequency-domain quantities:

Gain margin. The gain margin — see § 1.4 (p. 20) — measures relative stability. It is defined as the reciprocal of the magnitude of the loop frequency response L, evaluated at the frequency ω_π at which the phase angle is -180 degrees. The frequency ω_π is called the *phase crossover frequency*.

Phase margin. The phase margin — again see § 1.4 — also measures relative stability. It is defined as 180° plus the phase angle ϕ_1 of the loop frequency response L at the frequency ω_1 where the gain is unity. The frequency ω_1 is called the *gain crossover frequency*.

Bandwidth. The bandwidth B measures the speed of response in frequency-domain terms. It is defined as the range of frequencies over which the closed-loop frequency response H has a magnitude that is at least within a factor $\frac{1}{2}\sqrt{2} = 0.707$ (3 dB) of its value at zero frequency.

Resonance peak. Relative stability may also be measured in terms of the peak value M of the magnitude of the closed-loop frequency response H (in dB), occurring at the *resonance frequency* ω_r.

Figure 2.22 shows five important time-domain quantities that may be used for performance specifications for the response of the control system output to a step in the reference input:

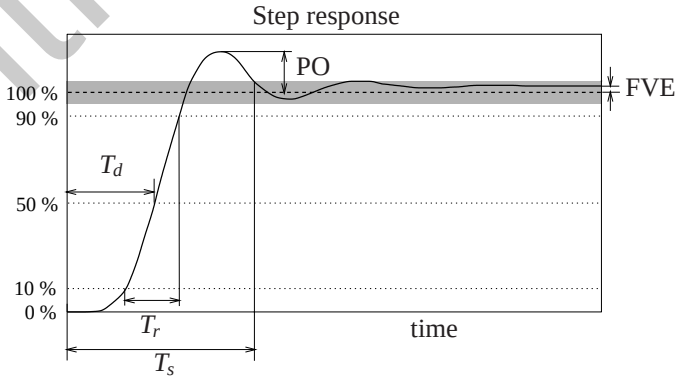


Figure 2.22: Time-domain quantities

Delay time T_d. The delay time measures the total average delay between reference and output. It may for instance be defined as time where the response is at 50% of the step amplitude.

Rise time T_r. The rise time expresses the "sharpness" of the leading edge of the response. Various definitions exist. One defines T_R as the time needed to rise from 10% to 90% of the final value.

Percentage overshoot PO. This quantity expresses the maximum difference (in % of the steady-state value) between the transient and the steady-state response to a step input.

Settling time T_s. The settling time is often defined as time required for the response to a step input to reach and remain within a specified percentage (typically 2 or 5%) of its final value.

Final value of error FVE. The FVE is the steady-state position error.

This list is not exhaustive. It includes no specifications of the *disturbance attenuating properties*. These specifications can not be easily expressed in general terms. They should be considered individually for each application.

Horowitz (1963, pp. 190–194) lists a number of quasi-empirical relations between the time domain parameters T_d, T_r, T_s and the overshoot on the one hand and the frequency domain parameters B, M and the phase at the frequency B on the other. The author advises to use them with caution.

Exercise 2.5.1. Cruise control system Evaluate the various time and frequency performance indicators for the integral cruise control system design of Example 2.3.3 (p. 67). □

2.5.2 Compensator design

In the classical control engineering era the design of feedback compensation to a great extent relied on trial-and-error procedures. Experience and engineering sense were as important as a thorough theoretical understanding of the tools that were employed.

In this section we consider the basic goals that may be pursued from a classical point of view. In the classical view the following series of steps leads to a successful control system design:

- Determine the plant transfer function P based on a (linearized) model of the plant.
- Investigate the shape of the frequency response $P(\mathrm{j}\omega)$, $\omega \in \mathbb{R}$, to understand the properties of the system fully.
- Consider the desired steady-state error properties of the system (see § 2.2, p. 60). Choose a compensator structure — for instance by introducing integrating action or lag compensation — that provides the required steady-state error characteristics of the compensated system.
- Plot the Bode, Nyquist or Nichols diagram of the loop frequency response of the compensated system. Adjust the gain to obtain a desired degree of stability of the system. M- and N-circles are useful tools. The gain and phase margins are measures for the success of the design.
- If the specifications are not met then determine the adjustment of the loop gain frequency response function that is required. Use lag, lead, lag-lead or other compensation to realize the necessary modification of the loop frequency response function. The Bode gain-phase relation sets the limits.

The graphic tools essential to go through these steps that were developed in former time now are integrated in computer aided design environments.

The design sequence summarizes the main ideas of classical control theory developed in the period 1940–1960. It is presented in terms of *shaping loop transfer functions* for single-input, single-output systems.

In § 2.6 (p. 82) we consider techniques for loop shaping using simple controller structures — lead, lag, and lead-lag compensators. In § 2.8 (p. 89) we discuss the Guillemin-Truxal design procedure. Section 2.9 (p. 93) is devoted to Horowitz's Quantitative Feedback Theory (Horowitz and Sidi 1972), which allows to impose and satisfy quantitative bounds on the robustness of the feedback system.

2.6 Lead, lag, and lag-lead compensation

2.6.1 Introduction

In this section we discuss the classical techniques of lead, lag, and lag-lead compensation. An extensive account of these techniques is given by Dorf (1992).

2.6.2 Lead compensation

Making the loop gain L large at low frequencies — by introducing integrating action or making the static gain large — may result in a Nyquist plot that shows unstable behavior. Even if the closed-loop system is stable the gain and phase margins may be unacceptably small, resulting in nearly unstable, oscillatory behavior.

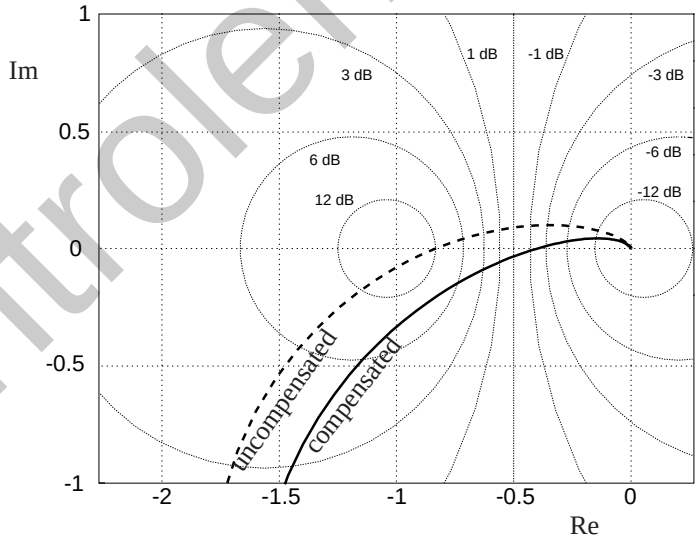


Figure 2.23: Nyquist plot of uncompensated and compensated plant

Figure 2.23 shows an instance of this. To obtain satisfactory stability we may reshape the loop gain in such a way that its Nyquist plot remains outside an M-circle that guarantees sufficient closed-loop damping. A minimal value of $M = 1.4$ (3 dB) might be a useful choice.

The required phase advance in the resonance frequency region may be obtained by utilizing a phase-advance network in series with the plant. The network may be of first order with frequency response function

$$C(\mathrm{j}\omega) = \alpha \frac{1 + \mathrm{j}\omega T}{1 + \mathrm{j}\omega\alpha T}. \tag{2.55}$$

For $0 < \alpha < 1$ we obtain a lead compensator and for $\alpha > 1$ a lag compensator. In the first case the compensator creates phase advance, in the second it creates extra phase lag. Figure 2.24 shows the Bode diagrams.

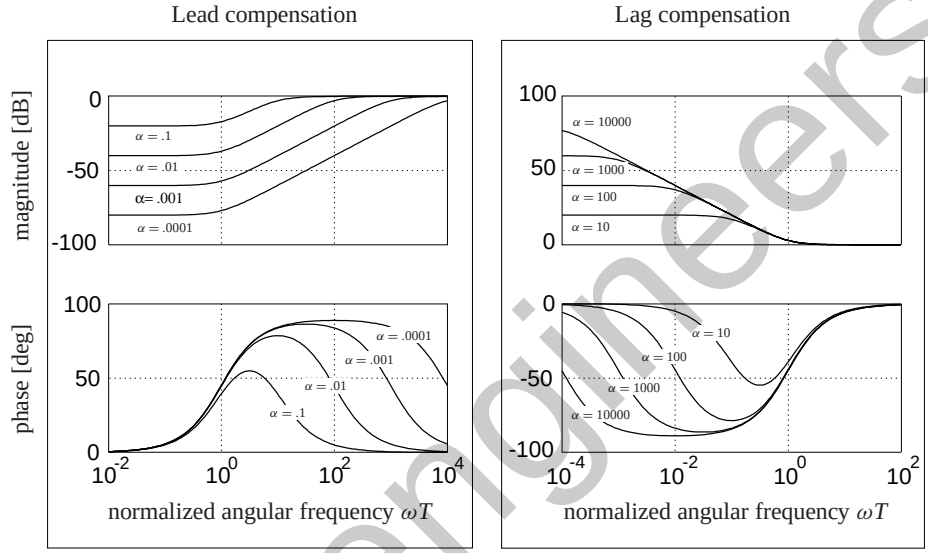


Figure 2.24: Log magnitude and phase of lead and lag compensators $C(s) = \alpha \frac{1+sT}{1+s\alpha T}$

Over the frequency interval $(1/T, 1/\alpha T)$ the phase advance compensator has the character of a differentiating network. By making α sufficiently small the compensator may be given the character of a differentiator over a large enough frequency range.

Phase lead compensation, also used in PD control, increases the bandwidth and, hence, makes the closed-loop system faster. Keeping the Nyquist plot away from the critical point -1 has the effect of improving the transient response.

Phase lead compensation results in an increase of the resonance frequency. If very small values of α are used then the danger of undesired amplification of measurement noise in the loop exists. The bandwidth increase associated with making α small may aggravate the effect of high frequency parasitic dynamics in the loop.

The characteristics of phase-lead compensation are reviewed in Table 2.2. An application of lead compensation is described in Example 2.6.3 (p. 86).

Exercise 2.6.1 (Specifics of the first-order lead or lag compensator). Inspection of Fig. 2.24 shows that the maximum amount of phase lead or lag that may be obtained with the compensator (2.55) is determined by α. Also the width of the frequency window over which significant phase lead or lag is achieved depends on α. Finally, the low frequency gain loss (for lead compensation) or gain boost (for lag compensation) depend on α.

1. Prove that the peak phase lead or lag occurs at the normalized frequency

$$\omega_{\text{peak}} T = 1/\sqrt{\alpha}, \tag{2.56}$$

and that the peak phase lead or lag equals

$$\phi_{\max} = \arctan \frac{1}{2} \left| \frac{1}{\sqrt{\alpha}} - \sqrt{\alpha} \right|. \tag{2.57}$$

2. Show that the width of the window over which phase lead or lag is effected is roughly $\left|\log_{10} \alpha\right|$ decades.

3. Show that the low frequency gain loss or boost is $20 \left|\log_{10} \alpha\right|$ dB.

Figure 2.25 shows plots of the peak phase lead or lag, the window width, and the low-frequency gain loss or boost. □

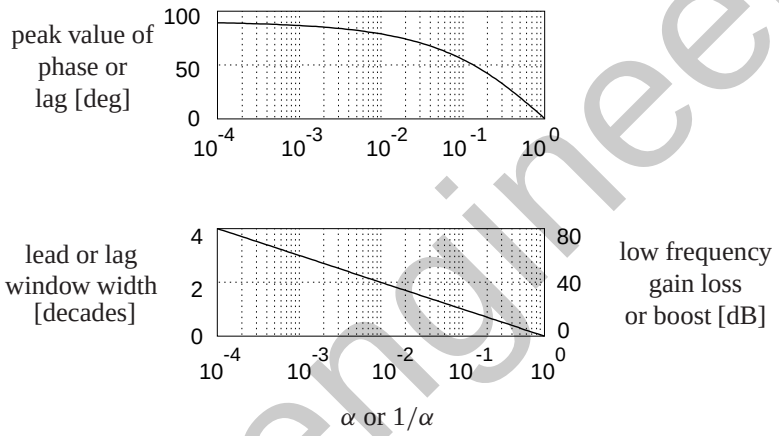


Figure 2.25: Peak phase lead or lag

First-order phase advance compensation is not effective against resonant modes in the plant corresponding to second order dynamics with low damping. The rapid change of phase from 0 to -180 degrees caused by lightly damped second-order dynamics cannot adequately be countered. This requires compensation by a second order filter (called a *notch filter*) with zeros near the lightly damped poles and stable poles on the real line at a considerable distance from the imaginary axis.

2.6.3 Lag compensation

The loop gain may be increased at low frequencies by a lag compensator. If the time constant T in

$$C(\mathrm{j}\omega) = \alpha \frac{1 + \mathrm{j}\omega T}{1 + \mathrm{j}\omega\alpha T} \tag{2.58}$$

is chosen such that $1/T$ is much greater than the resonance frequency ω_R of the loop gain then there is hardly any additional phase lag in the crossover region. In the limit $\alpha \to \infty$ the compensator frequency response function becomes

$$C(\mathrm{j}\omega) = 1 + \frac{1}{\mathrm{j}\omega T}. \tag{2.59}$$

Table 2.2: Characteristics of lead and lag compensation design

Compensation	Phase-lead	Phase-lag
Method	Addition of phase-lead angle near the crossover frequency	Increase the gain at low frequencies
Purpose	Improve the phase margin and the transient response	Increase the error constants while maintaining the phase margin and transient response properties
Applications	When a fast transient response is desired	When error constants are specified
Results	Increases the system bandwidth	Decreases the system bandwidth
Advantages	Yields desired response	Suppresses high frequency noise
	Speeds dynamic response	Reduces the steady-state error
Disadvantages	Increases the bandwidth and thus the susceptibility to measurement noise	Slows down the transient response
Not applicable	If the phase decreases rapidly near the crossover frequency	If no low frequency range exists where the phase is equal to the desired phase margin

This is a compensator with proportional and integral action.

Increasing the low frequency gain by lag compensation reduces the steady-state errors. It also has the effect of decreasing the bandwidth, and, hence, making the closed-loop system slower. On the other hand, the effect of high frequency measurement noise is reduced. Table 2.2 reviews and summarizes the characteristics of lead and lag compensation.

Lag compensation is fully compatible with phase-lead compensation as the two compensations affect frequency regions that are widely apart.

Exercise 2.6.2 (Phase lag compensation). An example of phase lag compensation is the integral compensation scheme for the cruise control system of Example 2.3.3 (p. 67). The first-order plant requires a large gain boost at low frequencies for good steady-state accuracy. This gain is provided by integral control. As we also saw in Example 2.3.3 (p. 67) pure integral control limits the bandwidth. To speed up the response additional phase lead compensation is needed.

To accomplish this modify the pure integral compensation scheme to the PI compensator

$$C(s) = k\frac{1+sT_i}{sT_i}. \tag{2.60}$$

This provides integrating action up to the frequency $1/T_i$. At higher frequencies the associated 90° phase lag vanishes. A suitable choice for the frequency $1/T_i$ is, say, half a decade below the desired bandwidth.

Suppose that the desired bandwidth is 0.3 [rad/s]. Select T_i as recommended, and choose the gain k such that the loop gain crossover frequency is 0.3 [rad/s]. Check whether the resulting design is satisfactory. □

2.6.4 Lag-lead compensation

We illustrate the design of a lag-lead compensator by an example. Note the successive design steps.

Example 2.6.3 (Lag-lead compensator). Consider the simple second-order plant with transfer function

$$P(s) = \frac{\omega_0^2}{s^2 + 2\zeta_0\omega_0 s + \omega_0^2}, \tag{2.61}$$

with $\omega_0 = 0.1$ [rad/s] and $\zeta_0 = 0.2$. The system is poorly damped. The design specifications are

- Constant disturbance rejection by integral action.
- A closed-loop bandwidth of 1 [rad/s].
- Satisfactory gain and phase margins.

Step 1: Lag compensation. To achieve integral control we introduce lag compensation of the form

$$C_0(s) = k_0 \frac{1 + sT_i}{sT_i}. \tag{2.62}$$

The phase lag compensation may be extended to 1 decade below the desired bandwidth by choosing $1/T_i = 0.1$ [rad/s], that is, $T_i = 10$ [s]. Letting $k_0 = 98.6$ makes sure that the crossover frequency of the loop gain is 1 [rad/s]. Figure 2.26 shows the Bode diagram of the resulting loop gain. Inspection reveals a negative phase margin, so that the closed-loop

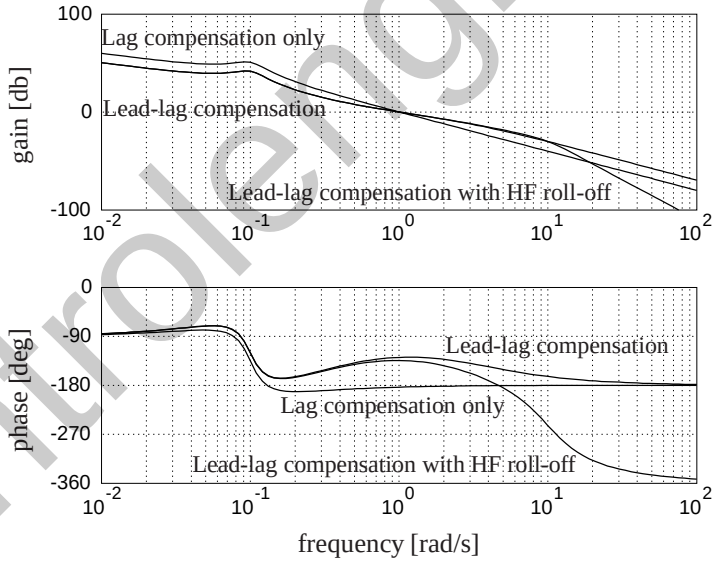


Figure 2.26: Bode diagrams of the loop gain

system is unstable.

Step 2. Phase lead compensation. We stabilize the closed loop by lead compensation of the form

$$C_1(s) = k_1\alpha \frac{1 + sT}{1 + s\alpha T}. \tag{2.63}$$

Phase advance is needed in the frequency region between, say, 0.1 and 10 [rad/s]. Inspection of Fig. 2.24 or 2.25 and some experimenting leads to the choice $\alpha = 0.1$ and $T = 3$ [rad/s]. Setting $k_1 = 3.3$ makes the crossover frequency equal to 1 [rad/s]. The resulting Bode diagram of the loop gain is included in Fig. 2.26. The closed-loop system is stable with infinite gain margin (because the phase never goes below -180°) and a phase margin of more than 50°.

Figure 2.27 shows the Bode magnitude plot of the closed-loop frequency response function and of the closed-loop step response. They are quite adequate.

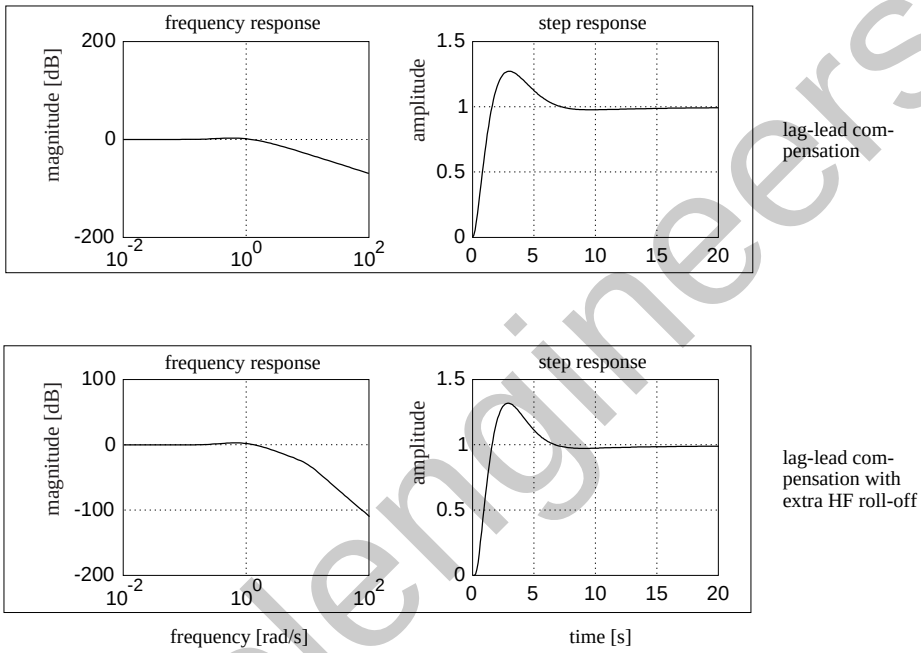


Figure 2.27: Closed-loop frequency and step responses

Step 3. High-frequency roll-off. For measurement noise reduction and high-frequency robustness we provide high-frequency roll-off of the compensator by including additional lag compensation of the form

$$C_2(s) = \frac{\omega_1^2}{s^2 + 2\zeta_1\omega_1 s + \omega_1^2}. \tag{2.64}$$

Setting $\omega_1 = 10$ [rad/s] and $\zeta_1 = 0.5$ makes the roll-off set in at 10 [rad/s] without unnecessary peaking and without appreciable effect in the crossover region. The corresponding loop gain is shown in Fig. 2.26. The gain margin is now about 17 dB and the phase margin about 45°. Figure 2.27 shows the extra roll-off of the closed-loop frequency response. Enhancing high-frequency roll-off slightly increases the overshoot of the closed-loop step response.

□

2.7 The root locus approach to parameter selection

2.7.1 Introduction

The root locus technique was conceived by Evans (1950) — see also Evans (1954). It consists of plotting the loci of the roots of the characteristic equation of the closed-loop system as a function of a proportional gain factor in the loop transfer function. This graphical approach yields a clear picture of the stability properties of the system as a function of the gain. It leads to a design decision about the value of the gain.

The root locus method is not a complete design procedure. First the controller structure, including its pole and zero locations, should be chosen. The root locus method then allows to adjust the gain. Inspection of the loci often provides useful indications how to revise the choice of the compensator poles and zeros.

2.7.2 Root loci rules

We review the basic construction rules for root loci. Let the loop transfer function of a feedback system be given in the form

$$L(s) = k\frac{(s-z_1)(s-z_2)\cdots(s-z_m)}{(s-p_1)(s-p_2)\cdots(s-p_n)}. \tag{2.65}$$

For physically realizable systems the loop transfer function L is proper, that is, $m \le n$. The roots $z_1, z_2, \cdots, z_m$ of the numerator polynomial are the *open-loop zeros* of the system. The roots $p_1, p_2, \cdots, p_n$ of the denominator polynomial are the *open-loop poles*. The constant k is the *gain*.

The closed-loop poles are those values of s for which $1 + L(s) = 0$, or, equivalently,

$$(s-p_1)(s-p_2)\cdots(s-p_n) + k(s-z_1)(s-z_2)\cdots(s-z_m) = 0. \tag{2.66}$$

Under the assumption that $m \le n$ there are precisely n closed-loop poles. The *root loci* are the loci of the closed-loop poles as k varies from 0 to $+\infty$.

Computer calculations based on subroutines for the calculation of the roots of a polynomial are commonly used to provide accurate plots of the root loci. The graphical rules that follow provide useful insight into the general properties of root loci.

Summary 2.7.1 (Basic construction rules for root loci).

1. For $k = 0$ the closed-loop poles coincide with the open-loop poles $p_1, p_2, \cdots, p_n$.
2. If $k \to \infty$ then m of the closed-loop poles approach the (finite) open-loop zeros $z_1, z_2, \cdots, z_m$. The remaining $n - m$ closed-loop poles tend to infinity.
3. There are as many locus branches as there are open-loop poles. Each branch starts for $k = 0$ at an open-loop pole location and ends for $k = \infty$ at an open-loop zero (which thus may be at infinity).
4. If $m < n$ then $n - m$ branches approach infinity along straight line asymptotes. The directions of the asymptotes are given by the angles

$$\alpha_i = \frac{2i+1}{n-m}\pi \quad \text{[rad]} \qquad i = 0, 1, \cdots, n-m-1. \tag{2.67}$$

Thus, for $n - m = 1$ we have $\alpha = \pi$, for $n - m = 2$ we have $\alpha = \pm\pi/2$, and so on. The angles are evenly distributed over $[0, 2\pi]$.

5. All asymptotes intersect the real axis at a single point at a distance s_0 from the origin, with

$$s_0 = \frac{(\text{sum of open-loop poles}) - (\text{sum of open-loop zeros})}{n - m}. \tag{2.68}$$

6. As we consider real-rational functions only the loci are symmetrical about the real axis.

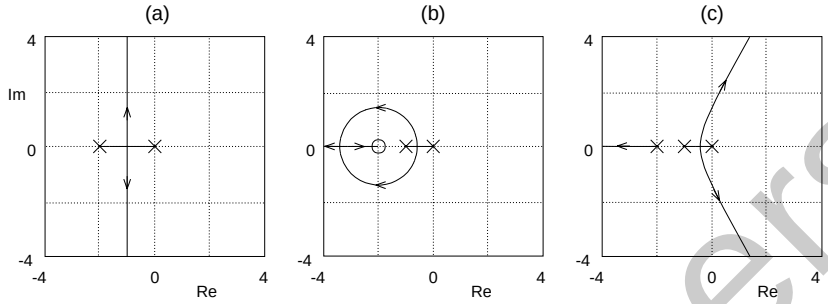


Figure 2.28: Examples of root loci:
(a) $L(s) = \frac{k}{s(s+2)}$, (b) $L(s) = \frac{k(s+2)}{s(s+1)}$, (c) $L(s) = \frac{k}{s(s+1)(s+2)}$

7. Those sections of the real axis located to the left of an odd total number of open-loop poles and zeros on this axis belong to a locus.

8. There may exist points where a locus breaks away from the real axis and points where a locus arrives on the real axis. Breakaway points occur only if the part of the real axis located between two open-loop poles belongs to a locus. Arrival points occur only if the part of the real axis located between two open-loop zeros belongs to a locus.

□

Figure 2.28 illustrates several typical root loci plots.

The root locus method has received much attention in the literature subsequent to Evans' pioneering work. Its theoretical background has been studied by Föllinger (1958), Berman and Stanton (1963), Krall (1961), Krall (1963), Krall (1970), and Krall and Fornaro (1967). The application of the root locus method in control design is described in almost any basic control engineering book — see for instance Dorf (1992), Franklin, Powell, and Emami-Naeini (1986), Franklin, Powell, and Emami-Naeini (1991), and Van de Vegte (1990).

Exercise 2.7.2 (Root loci). Check for each of the root locus diagrams of Fig. 2.28 which of the rules (a)–(h) of Summary 2.7.1 applies. □

2.8 The Guillemin-Truxal design procedure

2.8.1 Introduction

A network-theory oriented approach to the synthesis of feedback control systems was proposed by Truxal (1955). The idea is simple. Instead of designing a compensator on the basis of an analysis of the open-loop transfer function the closed-loop transfer function H is directly chosen such that it satisfies a number of favorable properties. Next, the compensator that realizes this behavior is computed. Generally an approximation is necessary to arrive at a practical compensator of low order.

2.8.2 Procedure

Let H be the chosen closed-loop transfer function. For unit feedback systems as in Fig. 2.29 with plant and compensator transfer functions P and C, respectively, we have

$$H = \frac{PC}{1+PC}. \tag{2.69}$$

Solving for the compensator transfer function C we obtain

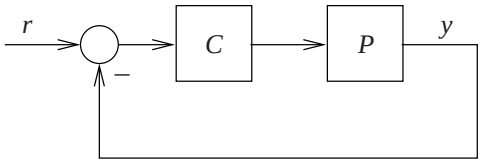


Figure 2.29: Unit feedback system

$$C = \frac{1}{P}\frac{H}{1-H}. \tag{2.70}$$

The determination of the desired H is not simple. Sometimes H may be selected on the basis of a preliminary analysis of the behavior of a closed-loop system with a low-order compensator.

A classical approach is to consider the steady-state errors in selecting the closed-loop system. Suppose that the closed-loop system is desired to be of type k (see § 2.2, p. 60). Then the loop gain L needs to be of the form

$$L(s) = \frac{N(s)}{s^k D(s)}, \tag{2.71}$$

with N and D polynomials that have no roots at 0. It follows that

$$H(s) = \frac{L(s)}{1+L(s)} = \frac{N(s)}{s^k D(s)+N(s)} = \frac{b_{k-1}s^{k-1}+b_{k-2}s^{k-2}+\cdots+b_0}{s^n+b_{n-1}s^{n-1}+\cdots+b_0}. \tag{2.72}$$

Conversely, choosing the numerator and denominator coefficients this way ensures the system to be of type k.

This still leaves considerable freedom to achieve other goals. Suppose that we select the closed-loop transfer function as

$$H(s) = \frac{b_0}{s^n+b_{n-1}s^{n-1}+\cdots+b_0}, \tag{2.73}$$

which implies a zero steady-state error for step inputs.

Exercise 2.8.1 (Zero steady-state error). Prove this. □

One way to choose the coefficients $b_0, b_1, \cdots, b_{n-1}$ is to place the closed-loop poles evenly distributed on the left half of a circle with center at the origin and radius ω_0. This yields closed-loop responses with a desired degree of damping. The resulting polynomials are known as *Butterworth polynomials*. For the normalized case $\omega_0 = 1$ the reference step responses are given in Fig. 2.30(a). Table 2.3 shows the coefficients for increasing orders. For general ω_0 the polynomials follow by substituting $s := s/\omega_0$.

Table 2.3: Normalized denominator polynomials for Butterworth pole patterns

Order	Denominator
1	$s+1$
2	$s^2+1.4s+1$
3	$s^3+2.0s^2+2.0s+1$
4	$s^4+2.6s^3+3.4s^2+2.6s+1$
5	$s^5+3.24s^4+5.24s^3+5.24s^2+3.24s+1$
6	$s^6+3.86s^5+7.46s^4+9.14s^3+7.46s^2+3.86s+1$
7	$s^7+4.49s^6+10.1s^5+14.6s^4+14.6s^3+10.1s^2+4.49s+1$
8	$s^8+5.13s^7+13.14s^6+21.85s^5+25.69s^4+21.85s^3+13.14s^2+5.13s+1$

Another popular choice is to choose the coefficients such that the integral of the time multiplied absolute error

$$\int_0^\infty t|e(t)|\,dt \tag{2.74}$$

is minimal, with e the error for a step input (Graham and Lathrop 1953). The resulting step responses and the corresponding so-called *ITAE standard forms* are shown in Fig. 2.30(b) and Table 2.4, respectively. The ITAE step responses have a shorter rise time and less overshoot than the corresponding Butterworth responses.

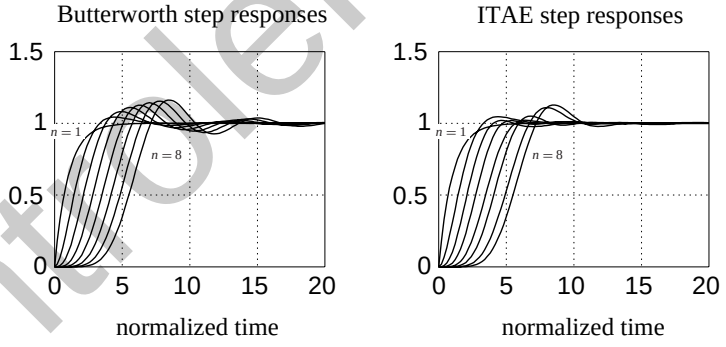


Figure 2.30: Step responses for Butterworth (left) and ITAE (right) denominator polynomials

2.8.3 Example

We consider the Guillemin-Truxal design procedure for the cruise control system of Example 2.3.3 (p. 67). The plant has the first-order transfer function

$$P(s)=\frac{\frac{1}{T}}{s+\frac{1}{\theta}}, \tag{2.75}$$

Table 2.4: Normalized denominator polynomials for ITAE criterion

Order	Denominator
1	$s+1$
2	$s^2+1.4s+1$
3	$s^3+1.75s^2+2.15s+1$
4	$s^4+2.1s^3+3.4s^2+2.7s+1$
5	$s^5+2.8s^4+5.0s^3+5.5s^2+3.4s+1$
6	$s^6+3.25s^5+6.60s^4+8.60s^3+7.45s^2+3.95s+1$
7	$s^7+4.475s^6+10.42s^5+15.08s^4+15.54s^3+10.64s^2+4.58s+1$
8	$s^8+5.20s^7+12.80s^6+21.60s^5+25.75s^4+22.20s^3+13.30s^2+5.15s+1$

with $T=\theta=10$ [s]. We specify the desired closed-loop transfer function

$$H(s)=\frac{\omega_0^2}{s^2+1.4\omega_0+\omega_0^2}. \tag{2.76}$$

The denominator is a second-order ITAE polynomial. The numerator has been chosen for a zero position error, that is, type 1 control. It is easy to find that the required compensator transfer function is

$$C(s)=\frac{1}{P(s)}\frac{H(s)}{1-H(s)}=\frac{\omega_0^2T(s+\frac{1}{\theta})}{s(s+1.4s)}. \tag{2.77}$$

The integrating action is patent. As seen in Example 2.3.3 (p. 67) the largest obtainable bandwidth with pure integral control is about $1/\sqrt{200}\approx 0.07$ [rad/s]. For the Guillemin-Truxal design we aim for a closed-loop bandwidth $\omega_0=1/\sqrt{2}\approx 0.7$ [rad/s].

Figure 2.31 shows the resulting sensitivity function and the closed-loop step response. It confirms that the desired bandwidth has been obtained. Inspection of (2.75) and (2.77) shows

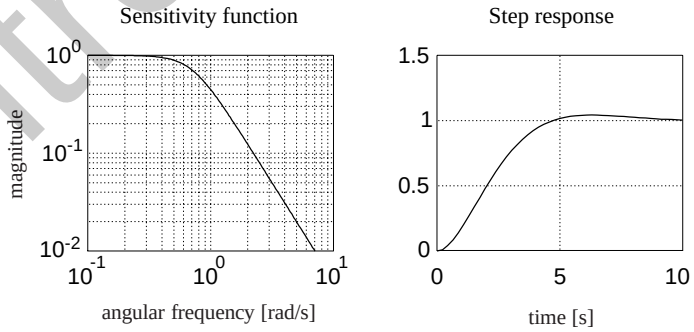


Figure 2.31: Sensitivity function and closed-loop step response of a Guillemin-Truxal design for the cruise control system

that in the closed loop the plant pole at $-1/\theta$ is canceled by a compensator zero at the same location. This does not bode well for the design, even though the sensitivity function and the

closed-loop step response of Fig. 2.31 look quite attractive. The canceling pole at $-1/\theta$ is also a closed-loop pole. It causes a slow response (with the open-loop time constant θ) to nonzero initial conditions of the plant and slow transients (with the same time constant) in the plant input.

This cancelation phenomenon is typical for naïve applications of the Guillemin-Truxal method. Inspection of (2.70) shows that cancelation may be avoided by letting the closed-loop transfer function H have a zero at the location of the offending pole. This constrains the choice of H, and illustrates what is meant by the comment that the selection of the closed-loop transfer function is not simple.

2.9 Quantitative feedback theory (QFT)

2.9.1 Introduction

Quantitative feedback theory (QFT) is a term coined by Horowitz (1982) (see also Horowitz and Sidi (1972)). A useful account is given by Lunze (1989). The method is deeply rooted in classical control. It aims at satisfying quantatitive bounds that are imposed on the variations in the closed-loop transfer function as a result of specified variations of the loop gain. The design method relies on the graphical representation of the loop gain in the Nichols chart.

2.9.2 Effect of parameter variations

Nichols plots may be used to study the effect of parameter variations and other uncertainties in the plant transfer function on a closed-loop system.

In particular, it may be checked whether the closed-loop system remains stable. By the Nyquist criterion, closed-loop stability is retained as long as the loop gain does not cross the point -1 under perturbation. In the Nichols chart, the critical point that is to be avoided is the point (-180°, 0 dB), located at the heart of the chart.

The effect of the perturbations on the closed-loop transfer function may be assessed by studying the width of the track that is swept out by the perturbations among the M-loci.

Example 2.9.1 (Uncertain second-order system). As an example we consider the plant with transfer function

$$P(s) = \frac{g}{s^2(1+s\theta)}. \tag{2.78}$$

Nominally $g = 1$ and $\theta = 0$. Under perturbation the gain g varies between 0.5 and 2. The parasitic time constant may independently vary from 0 to 0.2 [s]. We assume that a preliminary study has led to a tentative design in the form of a lead compensator with transfer function

$$C(s) = \frac{k + T_d s}{1 + T_0 s}, \tag{2.79}$$

with $k = 1$, $T_d = \sqrt{2}$ [s] and $T_0 = 0.1$ [s]. The nominal system has closed-loop poles $-0.7652 \pm$ j0.7715 and -8.4697. The closed-loop bandwidth is 1 [rad/s]. Figure 2.32 shows the nominal and perturbed complementary sensitivity function and closed-loop step response. Figure 2.33 shows the Nichols plot of the nominal loop gain $L_0 = P_0 C$, with $P_0(s) = 1/s^2$. The figure also shows with the uncertainty regions caused by the parameter variations at a number of fixed frequencies. These diagrams are constructed by calculating the loop gain $L(\mathrm{j}\omega)$ with ω fixed as a function of the uncertain parameters g and θ along the edges of the uncertainty regions. The corners of the uncertainty regions, as marked in the plot for $\omega = 10$, correspond to extreme values of the parameters as follows:

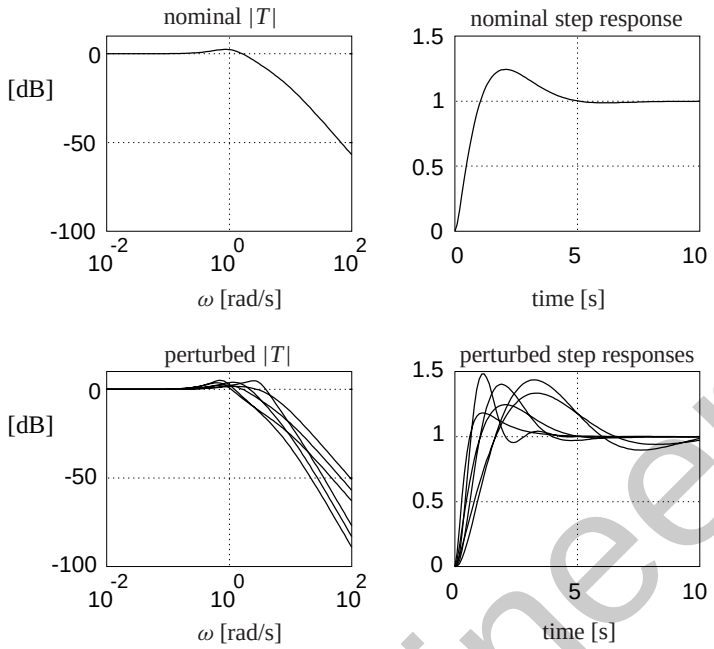


Figure 2.32: Nominal and perturbed complementary sensitivity functions and step responses for the nominal design

A: $\theta = 0, \quad g = 0.5$
B: $\theta = 0.2, \quad g = 0.5,$
C: $\theta = 0.2, \quad g = 2,$
D: $\theta = 0, \quad g = 2.$

Inspection shows that no perturbation makes the Nichols plot cross over the center of the chart. This means that the closed-loop system remains stable under all perturbations. □

2.9.3 Stability and performance robustness

Robust stability of the closed-loop system is guaranteed if perturbations do not cause the Nichols plot of the loop gain to cross over the center of the chart.

In the QFT approach, destabilization caused by *unmodeled perturbations* is prevented by specifying a *forbidden region* about the origin for the loop gain as in Fig. 2.33. The forbidden region is a region enclosed by an M-locus, for instance the 6 dB locus. If the Nichols plot of L never enters the forbidden region, not even under perturbation, then the modulus margin is always greater than 6 dB. Besides providing stability robustness, the guaranteed distance of L from the critical point prevents ringing.

In the QFT approach, in the simplest situation *performance robustness* is specified in the form of bounds on the variation of the magnitude of the closed-loop frequency response function H. Typically, for each frequency ω the maximally allowable variation $\Delta(\omega)$ of $|H(\mathrm{j}\omega)|$, called the *tolerance band*, is specified. Since $H = TF$, with T the complementary sensitivity function and F the prefilter transfer function, it follows after taking logarithms that

$$\log|H| = \log|T| + \log|F|. \tag{2.80}$$

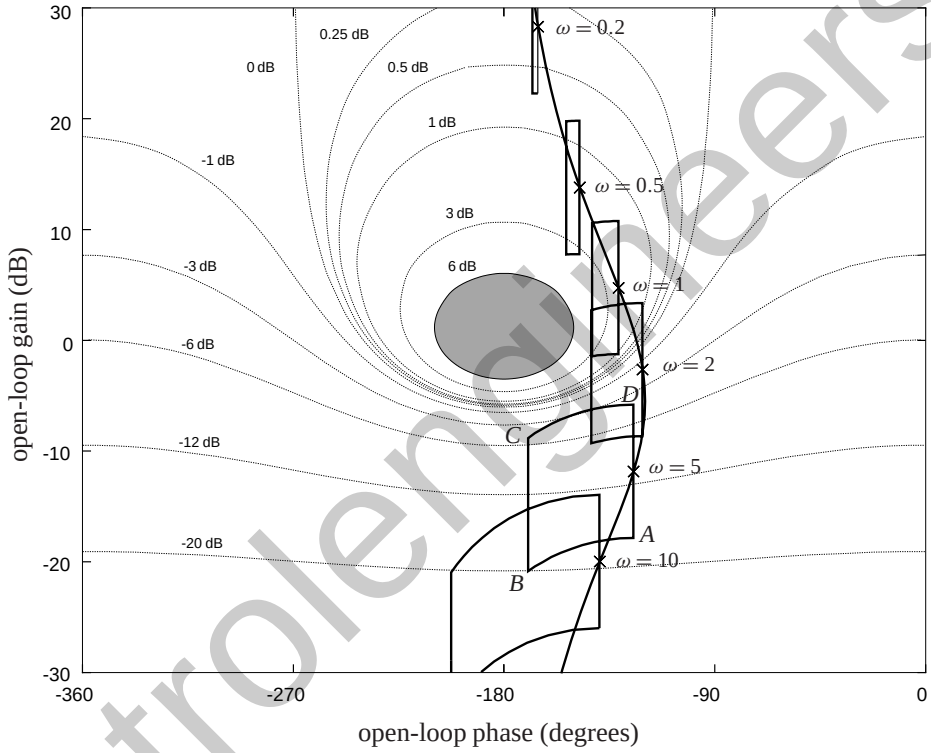


Figure 2.33: Nominal Nichols plot and uncertainty regions. (Shaded: forbidden region)

For simplicity we suppress the angular frequency ω. Inspection of (2.80) shows that if F is not subject to uncertainty then robust performance is obtained if and only if for each frequency $\log|T|$ varies by at most Δ on the uncertainty region. Whether this condition is satisfied may be verified graphically by checking in the Nichols chart whether the uncertainty region fits between two M-loci whose values differ by less than Δ.

In the next subsection we discuss how to design the feedback loop such that T satisfies the stability and performance robustness conditions.

Example 2.9.2 (Performance robustness of the design example). Inspection of the plots of Fig. 2.33 reveals that the perturbations sweep out a very narrow band of variations of $|T|$ at frequencies less than 0.2, a band with a width of about 5 dB at frequency 1, a band with a width of about 10 dB between the frequencies 2 and 10, while the width of the band further increases for higher frequencies. This is borne out by Fig. 2.32. □

2.9.4 QFT design of robust feedback systems

A feedback system design may easily fail to satisfy the performance robustness specifications. This often may be remedied by re-shaping the loop gain L.

Changing the compensator frequency response $C(j\omega)$ for some frequency ω amounts to *shifting* the loop gain $L(j\omega)$ at that same frequency in the Nichols plot. By visual inspection the shape of the Nichols plot of L may be adjusted by suitable shifts at the various frequencies so that the plot fits the tolerance bounds on T.

Part of the technique is to prepare *templates* of the uncertainty regions at a number of frequencies (usually not many more than five), and shifting these around in the Nichols chart. The translations needed to shift the Nichols plot to make it fit the tolerance requirements are achieved by a frequency dependent correction of the compensator frequency response C. Note that changing the loop gain by changing the compensator frequency response function does not affect the shapes of the templates.

The procedure is best explained by an example.

Example 2.9.3 (QFT design). We continue the second-order design problem of the previous examples, and begin by specifying the tolerance band Δ for a number of critical frequencies as in Table 2.5. The desired bandwidth is 1 rad/s.

Table 2.5: Tolerance band specifications.

frequency	tolerance band Δ
0.2	0.5 dB
1	2 dB
2	5 dB
5	10 dB
10	18 dB

Determination of the performance boundaries. The first step of the procedure is to trace for each selected critical frequency the locus of the *nominal* points such that the tolerance band is satisfied with the tightest fit. This locus is called the *performance boundary.* Points on the performance boundary may for instance be obtained by fixing the nominal point at a certain phase, and shifting the template up or down until the *lowest* position is found where the tolerance band condition is satisfied.

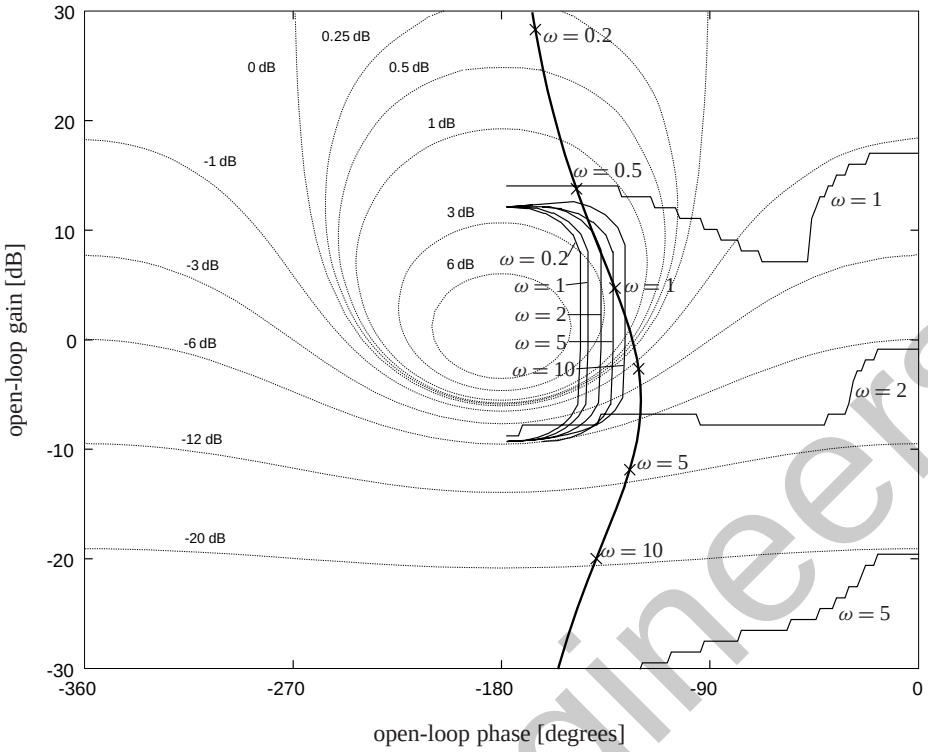


Figure 2.34: Performance and robustness boundaries

Determination of the robustness boundaries. Next, by shifting the template around the forbidden region so that it touches it but does not enter it the *robustness boundary* is traced for each critical frequency.

A feedback design satisfies the performance bounds and robustness bounds if for each critical frequency the corresponding value of the loop gain is on or above the performance boundary and to the right of or on the robustness boundary. If it is on the boundaries then the bounds are satisfied tightly. Figure 2.34 shows the performance boundaries thus obtained for the critical frequencies 1, 2 and 5 rad/s to the right in the Nichols chart. The performance boundary for the frequency .1 rad/s is above the portion that is shown and that for 10 rad/s below it. The robustness boundaries are shown for all five critical frequencies to the right of the center of the chart.

Inspection shows that the nominal design satisfies the specifications for the critical frequencies ω = 2, 5 and 10 rad/s, but not for $\omega = 1$ rad/s, and also for $\omega = 0.2$ it may be shown that the specifications are not satisfied.

Loop gain shaping. The crucial step in the design is to shape the loop gain such that

1. at each critical frequency ω the corresponding loop gain $L(\mathrm{j}\omega)$ is on or above the corresponding performance boundary;
2. at each critical frequency ω the corresponding loop gain $L(\mathrm{j}\omega)$ is to the right of the corresponding stability robustness boundary.

This target should be achieved with a compensator transfer function of least complexity and without overdesign (that is, the loop gain should be *on* the boundaries rather than

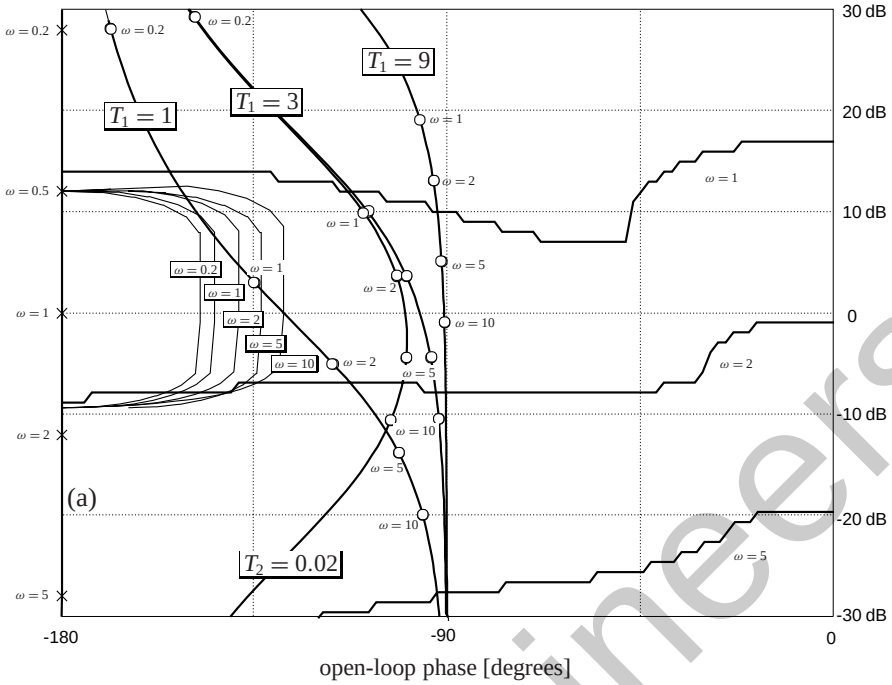


Figure 2.35: Redesign of the loop gain

above or to the right). This stage of the design requires experience and intuition, and is the least satisfactory from a methodical point of view.

In the problem at hand a design may be found in the following straightforward manner. The vertical line (a) in Fig. 2.35 is the Nichols plot of the nominal plant transfer function $P(s) = 1/s^2$. Obviously phase lead is needed. This is provided with a compensator with transfer function

$$C(s) = 1 + sT_1. \tag{2.81}$$

The curves marked $T_1 = 1$, $T_1 = 3$ and $T_1 = 9$ represent the corresponding loop gains $L = PC$. The loop gains for $T_1 = 3$ and $T_1 = 9$ satisfy the requirements; the latter with wide margins. We choose $T_1 = 3$.

To reduce the high-frequency compensator gain we modify its transfer function to

$$C(s) = \frac{1 + sT_1}{1 + sT_2}. \tag{2.82}$$

The resulting loop gain for $T_2 = 0.02$ is also included in Fig. 2.35. It very nearly satisfies the requirements[1]. Figure 2.36 gives plots of the resulting nominal and perturbed step responses and complementary sensitivity functions. The robustness improvement is evident. □

Exercise 2.9.4 (Performance and robustness boundaries). Traditional QFT relies on shifting paper templates around on Nichols charts to determine the performance and robustness bounds such as in Fig. 2.34. Think of ways to do this using routines from the MATLAB Control Toolbox. Practice these ideas by re-creating Fig. 2.35. □

[1] The requirements may be completely satisfied by adding a few dB to the loop gain.

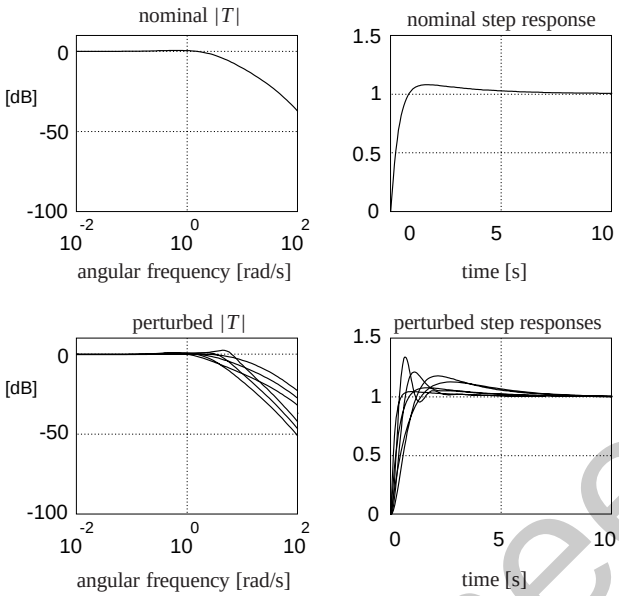


Figure 2.36: Nominal and perturbed complementary sensitivity functions and step responses of the revised design

2.9.5 Prefilter design

Once the feedback compensaor has been selected the QFT design needs to be completed with the design of the prefilter.

Example 2.9.5 (Prefilter design). We continue the design example, and complete it as a $2\frac{1}{2}$-degree-of-freedom design as proposed in § 1.8 (p. 47). Figure 2.37 shows the block diagram. The choice of the numerator polynomial F provides half a degree of freedom and the rational

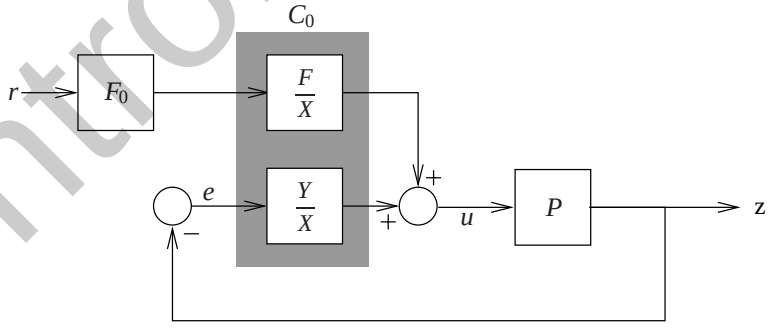


Figure 2.37: $2\frac{1}{2}$-degree-of-freedom feedback system

transfer function F_0 of the rational prefilter constitutes another degree of freedom. The closed-loop transfer function (from the reference input r to the controlled output z) is

$$H = \frac{NF}{D_{\mathrm{cl}}} F_0. \tag{2.83}$$

$P = N/D$ is the plant transfer function and $D_{\rm cl} = DX + NY$ the closed-loop characteristic polynomial.

In the problem at hand $N(s) = 1$ and $D(s) = s^2$. The compensator $Y(s) = 3s + 1$, $X(s) = 0.02s + 1$ constructed in Example 2.9.3 (p. 96) results in the closed-loop characteristic polynomial

$$D_{\rm cl}(s) = 0.02s^3 + s^2 + 3s + 1. \tag{2.84}$$

Its roots are -0.3815, -2.7995, and -46.8190. Completing the design amounts to choosing the correct polynomial F and transfer function F_0 to provide the necessary compensation in

$$H(s) = \frac{F(s)}{0.02(s+0.3815)(s+2.7995)(s+46.8190)} F_0(s). \tag{2.85}$$

The pole at -0.3815 slows the response down, so we cancel it by selecting the polynomial F — whose degree can be at most 1 — as $F(s) = s/0.3815 + 1$. To reduce the bandwidth to the desired 1 rad/s and to obtain a critically damped closed-loop step response we let

$$F_0(s) = \frac{\omega_0^2}{s^2 + 2\zeta_0\omega_0 s + \omega_0^2}, \tag{2.86}$$

with $\omega_0 = 1$ rad/s and $\zeta = \frac{1}{2}\sqrt{2}$.

Figure 2.38 displays the ensuing nominal and perturbed step responses and closed-loop transfer functions. Comparison with Fig. 2.33 makes it clear that the robustness has been drastically improved. □

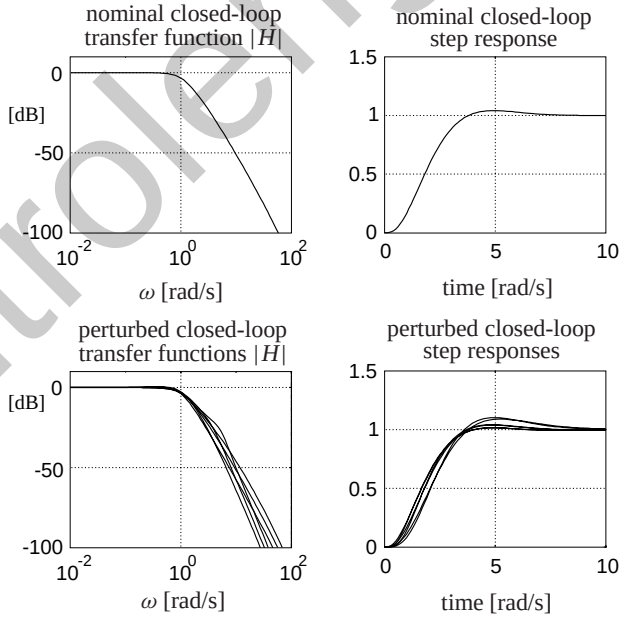


Figure 2.38: Nominal and perturbed step responses and closed-loop transfer functions of the final QFT design

2.9.6 Concluding remark

The QFT method has been extended to open-loop unstable plants and non-minimum phase plants, and also to MIMO and nonlinear plants. Horowitz (1982) provides references and a review. Recently a MATLAB toolbox for QFT design has appeared[2].

2.10 Concluding remarks

This chapter deals with approaches to classical compensator design. The focus is on compensation by shaping the open-loop frequency response.

The design goals in terms of shaping the loop gain are extensively considered in the classical control literature. The classical design techniques cope with them in an *ad hoc* and qualitative manner. It requires profound experience to handle the classical techniques, but if this experience is available then for single-input single-output systems it is not easy to obtain the quality of classical design results by the analytical control design methods that form the subject of the later chapters of this book.

If the design problem has a much more complex structure, for instance with multi-input multi-output plants or when complicated uncertainty descriptions and performance requirements apply, then the analytical techniques are the only reliable tools. Even in this case a designer needs considerable expertise and experience with classical techniques to appreciate and understand the design issues involved.

[2] Quantitative Feedback Theory Toolbox, The MathWorks Inc., Natick, MA, USA, 1995 release.

3

LQ, LQG and H_2 Control System Design

Overview *– LQ and LQG design methods convert control system design problems to an optimization problem with quadratic time-domain performance criteria. Disturbances and measurement noise are modeled as stochastic processes.*

The H_2 formulation of the same method eliminates the stochastic element. It permits a frequency domain view and allows the introduction of frequency dependent weighting functions.

MIMO problems can be handled almost as easily as SISO problems.

3.1 Introduction

The application of optimal control theory to the practical design of multivariable control systems attracted much attention during the period 1960–1980. This theory considers linear finite-dimensional systems represented in state space form, with quadratic performance criteria. The system may be affected by disturbances and measurement noise represented as stochastic processes, in particular, by Gaussian white noise. The theoretical results obtained for this class of design methods are known under the general name of *LQG theory*. Standard references are Anderson and Moore (1971), Kwakernaak and Sivan (1972) and Anderson and Moore (1990). The deterministic part is called *LQ theory.*

In the period since 1980 the theory has been further refined under the name of H_2 theory (Doyle, Glover, Khargonekar, and Francis 1989), in the wake of the attention for the so-called H_∞ control theory.

In the present chapter we present a short overview of a number of results of LQG and H_2 theory with an eye to using them for control system design. LQ theory is basic to the whole chapter, and is dealt with at some length in Section 3.2 (p. 104). Besides presenting the solution to the LQ problem we discuss its properties, the choice of the weighting matrices, and how to obtain systems with a prescribed degree of stability. Using the notion of return difference and the associated return difference equality we discuss the asymptotic properties and the guaranteed

gain and phase margins associated with the LQ solution. The section concludes with a subsection on the numerical solution of Riccati equations.

Section 3.3 (p. 115) deals with the LQG problem. The LQ paradigm leads to state feedback. By using optimal observers — Kalman filters — compensators based on output feedback may be designed. For well-behaved plants — specifically, plants that have no right-half plane zeros — the favorable properties of the LQ solution may asymptotically be recovered by assuming very small measurement noise. This is known as *loop transfer recovery.*

In Section 3.4 (p. 124) it is demonstrated that LQG optimization amounts to minimization of the H_2-norm of the closed-loop system. This interpretation removes the stochastic ingredient from the LQG framework, and reduces the role of the intensity matrices that describe the white noise processes in the LQG problem to that of design parameters. The H_2 interpretation naturally leads to H_2 optimization with frequency dependent weighting functions. These permit a great deal of extra flexibility and make H_2 theory a tool for shaping closed-loop system functions. A useful application is the design of feedback systems with integral control.

Multi-input multi-output systems are handled almost (but not quite) effortlessly in the LQG and H_2 framework. In Section 3.6 (p. 133) we present both a SISO and a MIMO design example.

In Section 3.7 (p. 139) a number of proofs for this chapter are collected.

3.2 LQ theory

3.2.1 Introduction

In this section we describe the LQ paradigm. The acronym refers to Linear systems with Quadratic performance criteria. Consider a linear time-invariant system represented in state space form as

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t), \\ z(t) &= Dx(t), \end{aligned} \qquad t \geq 0. \tag{3.1}$$

For each $t \geq 0$ the state $x(t)$ is an n-dimensional vector, the input $u(t)$ a k-dimensional vector, and the output $z(t)$ an m-dimensional vector.

We wish to control the system from any initial state $x(0)$ such that the output z is reduced to a very small value as quickly as possible without making the input u unduly large. To this end we introduce the performance index

$$J = \int_0^\infty [z^{\mathrm{T}}(t)Qz(t) + u^{\mathrm{T}}(t)Ru(t)]\, dt. \tag{3.2}$$

Q and R are symmetric weighting matrices, that is, $Q = Q^{\mathrm{T}}$ and $R = R^{\mathrm{T}}$. Often it is adequate to let the two matrices simply be diagonal.

The two terms $z^{\mathrm{T}}(t)Qz(t)$ and $u^{\mathrm{T}}(t)Ru(t)$ are quadratic forms in the components of the output z and the input u, respectively. The first term in the integral criterion (3.2) measures the accumulated deviation of the output from zero. The second term measures the accumulated amplitude of the control input. It is most sensible to choose the weighting matrices Q and R such that the two terms are nonnegative, that is, to take Q and R nonnegative-definite[1]. If the matrices are diagonal then this means that their diagonal entries should be nonnegative.

The problem of controlling the system such that the performance index (3.2) is minimal along all possible trajectories of the system is the *optimal linear regulator problem.*

[1] An $n \times n$ symmetric matrix R is nonnegative-definite if $x^{\mathrm{T}}Rx \geq 0$ for every n-dimensional vector x. R is positive-definite if $x^{\mathrm{T}}Rx > 0$ for all nonzero x.

3.2.2 Solution of the LQ problem

There is a wealth of literature on the linear regulator problem. The reason why it attracted so much attention is that its solution may be represented in *feedback* form. An optimal trajectory is generated by choosing the input for $t \geq 0$ as

$$u(t) = -Fx(t). \tag{3.3}$$

This solution requires that the state $x(t)$ be fully accessible for measurement at all times. We return to this unreasonable assumption in § 3.3 (p. 115). The $k \times n$ *state feedback gain matrix* F is given by

$$F = R^{-1}B^{\mathrm{T}}X. \tag{3.4}$$

The symmetric $n \times n$ matrix X is the nonnegative-definite solution of the *algebraic matrix Riccati equation* (ARE)

$$A^{\mathrm{T}}X + XA + D^{\mathrm{T}}QD - XBR^{-1}B^{\mathrm{T}}X = 0. \tag{3.5}$$

The proof is sketched in § 3.7 (p. 139), the appendix to this chapter. The solution of the algebraic Riccati equation is discussed in § 3.2.9 (p. 114).

We summarize a number of well-known important facts about the solution of the LQ problem. An outline of the proof is given in § 3.7.1 (p. 142).

Summary 3.2.1 (Properties of the solution of the optimal linear regulator problem).
Assumptions:

- The system (3.1) is stabilizable[2] and detectable[3]. Sufficient for stabilizability is that the system is controllable. Sufficient for detectability is that it is observable.
- The weighting matrices Q and R are positive-definite.

The following facts are well documented (see for instance Kwakernaak and Sivan (1972) and Anderson and Moore (1990)).

1. The algebraic Riccati equation (ARE)

   $$A^{\mathrm{T}}X + XA + D^{\mathrm{T}}QD - XBR^{-1}B^{\mathrm{T}}X = 0 \tag{3.6}$$

   has a unique nonnegative-definite symmetric solution X. If the system $\dot{x}(t) = Ax(t)$, $z(t) = Dx(t)$ is observable then X is positive-definite. There are finitely many other solutions of the ARE.

2. The minimal value of the performance index (3.2) is $J_{\min} = x^{\mathrm{T}}(0)Xx(0)$.

3. The minimal value of the performance index is achieved by the feedback control law

   $$u(t) = -Fx(t), \qquad t \geq 0, \tag{3.7}$$

   with $F = R^{-1}B^{\mathrm{T}}X$.

[2] That is, there exists a state feedback $u(t) = -Fx(t)$ such that the closed-loop system $\dot{x}(t) = (A - BF)x(t)$ is stable.
[3] That is, there exists a matrix K such that the system $\dot{e}(t) = (A - KD)e(t)$ is stable.

4. The closed-loop system

$$\dot{x}(t) = (A - BF)x(t), \qquad t \geq 0, \tag{3.8}$$

is stable, that is, all the eigenvalues of the matrix $A - BF$ have strictly negative real parts.

□

The reasons for the assumptions may be explained as follows. If the system is not stabilizable then obviously it cannot be stabilized. If it is not detectable then there exist state feedback controllers that do not stabilize the system but hide the instability from the output—hence, stability of the optimal solution is not guaranteed. R needs to be positive-definite to prevent infinite input amplitudes. If Q is not positive-definite then there may be unstable closed-loop modes that have no effect on the performance index.

Exercise 3.2.2 (Cruise control system). The linearized dynamics of the vehicle of Example 1.2.1 (p. 3) may be described by the equation $\dot{x}(t) = -ax(t) + au(t)$, where $a = 1/\theta$ is a positive constant. Without loss of generality we may take $a = 1$.

Consider finding the linear state feedback that minimizes the criterion

$$\int_0^\infty [x^2(t) + \rho u^2(t)]\, dt \tag{3.9}$$

with ρ a positive constant. Determine the ARE, find its positive solution, and compute the optimal state feedback gain.

Compute the closed-loop pole of the resulting optimal feedback system and check that the closed-loop system is always stable. How does the closed-loop pole vary with ρ? Explain.

Plot the closed-loop response of the state $x(t)$ and input $u(t)$ to the initial state $x(0) = 1$ in dependence on ρ. Explain the way the closed-loop responses vary with ρ. □

3.2.3 Choice of the weighting matrices

The choice of the weighting matrices Q and R is a trade-off between control performance (Q large) and low input energy (R large). Increasing both Q and R by the same factor leaves the optimal solution invariant. Thus, only relative values are relevant. The Q and R parameters generally need to be tuned until satisfactory behavior is obtained, or until the designer is satisfied with the result.

An initial guess is to choose both Q and R diagonal

$$Q = \begin{bmatrix} Q_1 & 0 & 0 & \cdots & 0 \\ 0 & Q_2 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & \cdots & 0 & Q_m \end{bmatrix}, \qquad R = \begin{bmatrix} R_1 & 0 & 0 & \cdots & 0 \\ 0 & R_2 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & \cdots & 0 & R_k \end{bmatrix}, \tag{3.10}$$

where Q and R have positive diagonal entries such that

$$\sqrt{Q_i} = \frac{1}{z_i^{\max}}, \quad i = 1, 2, \cdots, m, \qquad \sqrt{R_i} = \frac{1}{u_i^{\max}}, \quad i = 1, 2, \cdots, k. \tag{3.11}$$

The number $z_i^{\max}$ denotes the maximally acceptable deviation value for the ith component of the output z. The other quantity $u_i^{\max}$ has a similar meaning for the ith component of the input u.

Starting with this initial guess the values of the diagonal entries of Q and R may be adjusted by systematic trial and error.

3.2.4 Prescribed degree of stability

By including a time-dependent weighting function in the performance index that grows exponentially with time we may force the optimal solutions to decay faster than the corresponding exponential rate. The modified performance index is

$$J_\alpha = \int_0^\infty \mathrm{e}^{2\alpha t}[z^{\mathrm{T}}(t)Qz(t) + u^{\mathrm{T}}(t)Ru(t)]\,dt, \tag{3.12}$$

with α a real number. Define

$$x_\alpha(t) = x(t)\,\mathrm{e}^{\alpha t}, \qquad u_\alpha(t) = u(t)\,\mathrm{e}^{\alpha t}, \qquad t \geq 0. \tag{3.13}$$

These signals satisfy

$$\dot{x}_\alpha(t) = (A + \alpha I)x_\alpha(t) + Bu_\alpha(t) \tag{3.14}$$

and

$$J_\alpha = \int_0^\infty [z_\alpha^{\mathrm{T}}(t)Qz_\alpha(t) + u_\alpha^{\mathrm{T}}(t)Ru_\alpha(t)]\,dt. \tag{3.15}$$

Consequently, the minimizing u_α is

$$u_\alpha(t) = -R^{-1}B^{\mathrm{T}}X_\alpha x_\alpha(t), \tag{3.16}$$

or

$$u(t) = -F_\alpha x(t), \tag{3.17}$$

with $F_\alpha = R^{-1}B^{\mathrm{T}}X_\alpha$. X_α is the positive-definite solution of the modified algebraic Riccati equation

$$(A^{\mathrm{T}} + \alpha I)X_\alpha + X_\alpha(A + \alpha I) + D^{\mathrm{T}}QD - X_\alpha BR^{-1}B^{\mathrm{T}}X_\alpha = 0. \tag{3.18}$$

The stabilizing property of the optimal solution of the modified problem implies that

$$\text{Re } \lambda_i(A + \alpha I - BF_\alpha) < 0, \qquad i = 1, 2, \cdots, n, \tag{3.19}$$

with $\lambda_i(A + \alpha I - BF_\alpha)$ denoting the ith eigenvalue. Application of the control law (3.17) to the system (3.1) creates a closed-loop system matrix $\bar{A}_\alpha = A - BF_\alpha$. It follows from (3.19) that its eigenvalues satisfy

$$\text{Re } \lambda_i(\bar{A}_\alpha) < -\alpha. \tag{3.20}$$

Thus, choosing α positive results in an optimal closed-loop system with a *prescribed degree of stability*.

Exercise 3.2.3 (Cruise control system). Modify the criterion (3.9) to

$$\int_0^\infty \mathrm{e}^{2\alpha t}[x^2(t) + \rho u^2(t)]\,dt, \tag{3.21}$$

with α a positive constant. Rework Exercise 3.2.2 (p. 106) while explaining and interpreting the effect of α. □

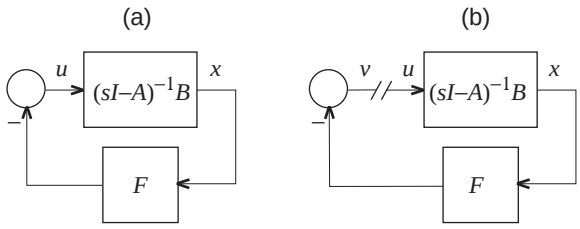


Figure 3.1: State feedback

3.2.5 Return difference equality and inequality

Figure 3.1(a) shows the feedback connection of the system $\dot{x} = Ax + Bu$ with the state feedback controller $u = -Fx$. If the loop is broken as in Fig. 3.1(b) then the loop gain is

$$L(s) = F(sI - A)^{-1}B. \tag{3.22}$$

The quantity

$$J(s) = I + L(s) \tag{3.23}$$

is known as the *return difference* $J(s)u$ is the difference between the signal u in Fig. 3.1(b) and the "returned" signal $v = -L(s)u$.

Several properties of the closed-loop system may be related to the return difference. Consider

$$\det J(s) = \det[I + L(s)] = \det[I + F(sI - A)^{-1}B]. \tag{3.24}$$

Using the well-known matrix equality $\det(I + MN) = \det(I + NM)$ we obtain

$$\begin{aligned} \det J(s) &= \det[I + (sI - A)^{-1}BF] \\ &= \det\,(sI - A)^{-1}\det(sI - A + BF) \\ &= \frac{\det(sI - A + BF)}{\det(sI - A)} = \frac{\chi_{\rm cl}(s)}{\chi_{\rm ol}(s)}. \end{aligned} \tag{3.25}$$

The quantities $\chi_{\rm ol}(s) = \det(sI - A)$ and $\chi_{\rm cl}(s) = \det(sI - A + BF)$ are the open- and the closed-loop characteristic polynomial, respectively. We found the same result in § 1.3 (p. 11) in the form $\chi_{\rm cl}(s) = \chi_{\rm ol}(s)\det[I + L(s)]$.

Suppose that the gain matrix F is optimal as in Summary 3.2.1. It is proved in § 3.7.2 (p. 143) by manipulation of the algebraic Riccati equation (3.6) that the corresponding return difference satisfies the equality

$$J^{\rm T}(-s)RJ(s) = R + G^{\rm T}(-s)QG(s). \tag{3.26}$$

$G(s) = D(sI - A)^{-1}B$ is the *open-loop transfer matrix* of the system (3.1).

The relation (3.26) is known as the *return difference equality* or as the *Kalman-Jakubovič-Popov (KJP)* equality, after its discoverers. In Subsection 3.2.6 (p. 109) we use the return difference equality to study the root loci of the optimal closed-loop poles.

By setting $s = j\omega$, with $\omega \in \mathbb{R}$, we obtain the *return difference inequality*[4]

$$J^{\rm T}(-{\rm j}\omega)RJ({\rm j}\omega) \geq R \qquad \text{for all } \omega \in \mathbb{R}. \tag{3.27}$$

In Subsection 3.2.7 (p. 112) we apply the return difference inequality to establish a well-known robustness property of the optimal state feedback system.

[4]If P and Q are symmetric matrices of the same dimensions then $P \geq Q$ means that $x^{\rm T}Px \geq x^{\rm T}Qx$ for every real n-dimensional vector x.

3.2.6 Asymptotic performance weighting

For simplicity we first consider the case that (3.1) is a SISO system. To reflect this in the notation we rewrite the system equations (3.1) in the form

$$\begin{aligned} \dot{x}(t) &= Ax(t) + bu(t), \\ z(t) &= dx(t), \end{aligned} \tag{3.28}$$

with b a column vector and d a row vector. Similarly, we represent the optimal state feedback controller as

$$u(t) = -fx(t), \tag{3.29}$$

with f a row vector. The open-loop transfer function $G(s) = d(sI - A)^{-1}b$, the loop gain $L(s) = f(sI - A)^{-1}b$ and the return difference $J(s) = 1 + L(s)$ now all are scalar functions. Without loss of generality we consider the performance index

$$J = \int_0^\infty [z^2(t) + \rho u^2(t)]\, dt, \tag{3.30}$$

with ρ a positive number. This amounts to setting $Q = 1$ and $R = \rho$.

Under these assumptions the return difference equality (3.26) reduces to

$$J(-s)J(s) = 1 + \frac{1}{\rho}G(s)G(-s). \tag{3.31}$$

From (3.27) we have

$$J(s) = \frac{\chi_{\rm cl}(s)}{\chi_{\rm ol}(s)}, \tag{3.32}$$

with $\chi_{\rm cl}$ the closed-loop characteristic polynomial and $\chi_{\rm ol}$ the open-loop characteristic polynomial. We furthermore write

$$G(s) = \frac{k\psi(s)}{\chi_{\rm ol}(s)}. \tag{3.33}$$

The constant k is chosen such that the polynomial ψ is *monic*[5]. From (3.31–3.33) we now obtain

$$\chi_{\rm cl}(-s)\chi_{\rm cl}(s) = \chi_{\rm ol}(-s)\chi_{\rm ol}(s) + \frac{k^2}{\rho}\psi(-s)\psi(s). \tag{3.34}$$

The left-hand side $\chi_{\rm cl}(-s)\chi_{\rm cl}(s)$ of this relation defines a polynomial whose roots consists of the closed-loop poles (the roots of $\chi_{\rm cl}(s)$) together with their *mirror images* with respect to the imaginary axis (the roots of $\chi_{\rm cl}(-s)$). It is easy to separate the two sets of poles, because by stability the closed-loop poles are always in the left-half complex plane.

From the right-hand side of (3.34) we may determine the following facts about the loci of the closed-loop poles as the weight ρ on the input varies.

Infinite weight on the input term. If $\rho \to \infty$ then the closed-loop poles and their mirror images approach the roots of $\chi_{\rm ol}(s)\chi_{\rm ol}(-s)$. This means that the closed-loop poles approach

[5] That is, the coefficient of the term of highest degree is 1.

- those open-loop poles that lie in the left-half complex plane (the "stable" open-loop poles), and
- the mirror images of those open-loop poles that lie in the right-half complex plane (the "unstable" open-loop poles).

If the open-loop system is stable to begin with then the closed-loop poles approach the open-loop poles as the input is more and more heavily penalized. In fact, in the limit $\rho \to \infty$ all entries of the gain F become zero—optimal control in this case amounts to no control at all.

If the open-loop system is unstable then in the limit $\rho \to \infty$ the least control effort is used that is needed to stabilize the system but no effort is spent on regulating the output.

Vanishing weight on the input term. As $\rho \downarrow 0$ the closed-loop poles the open-loop zeros (the roots of ψ) come into play. Suppose that q_- open-loop zeros lie in the left-half complex plane or on the imaginary axis, and q_+ zeros in the right-half plane.

- If $\rho \downarrow 0$ then q_- closed-loop poles approach the q_- left-half plane open-loop zeros.
- A further q_+ closed-loop poles approach the mirror images in the left-half plane of the q_+ right-half plane open-loop zeros.
- The remaining $n - q_- - q_+$ closed-loop poles approach infinity according to a Butterworth pattern of order $n - q_- - q_+$ (see § 2.7, p. 88).

Just how small or large ρ should be chosen depends on the desired or achievable bandwidth. We first estimate the radius of the Butterworth pole configuration that is created as ρ decreases. Taking leading terms only, the right-hand side of (3.34) reduces to

$$(-s)^n s^n + \frac{k^2}{\rho}(-s)^q s^q, \tag{3.35}$$

with q the degree of ψ, and, hence, the number of open-loop zeros. From this the $2(n-q)$ roots that go to infinity as $\rho \downarrow 0$ may be estimated as the roots of

$$s^{2(n-q)} + (-1)^{n-q}\frac{k^2}{\rho} = 0. \tag{3.36}$$

The $n-q$ left-half plane roots are approximations of the closed-loop poles. They form a Butterworth pattern of order $n-q$ and radius

$$\omega_c = \left(\frac{k^2}{\rho}\right)^{\frac{1}{2(n-q)}}. \tag{3.37}$$

If the plant has no right-half plane zeros then this radius is an estimate of the closed-loop bandwidth. The smaller ρ is the more accurate the estimate is. The bandwidth we refer to is the bandwidth of the closed-loop system with z as output.

If the plant has right-half plane open-loop zeros then the bandwidth is limited to the magnitude of the right-half plane zero that is closest to the origin. This agrees with the limits of performance established in § 1.7 (p. 40).

Exercise 3.2.4 (Closed-loop frequency response characteristics). Verify these statements by considering the SISO configuration of Fig. 3.2. Let $L(s) = f(sI - A)^{-1}b$ and $G(s) = d(sI - A)^{-1}b$.

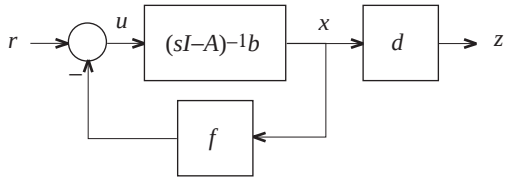


Figure 3.2: State feedback with reference input r and output z

1. Prove that

$$u = \frac{1}{1+L(s)}r, \qquad z = \underbrace{\frac{G(s)}{1+L(s)}}_{H(s)}r. \tag{3.38}$$

It follows that the closed-loop transfer function H is

$$H(s) = \frac{k\psi(s)}{\chi_{\rm cl}(s)}. \tag{3.39}$$

2. Assume that the open-loop transfer function G has no right-half plane zeros. Prove that as $\rho \downarrow 0$ the closed-loop transfer function behaves as

$$H(s) \approx \frac{k}{B_{n-q}(s/\omega_c)}. \tag{3.40}$$

B_k denotes a Butterworth polynomial of order k (see Table 2.3, p. 91). Hence, the closed-loop bandwidth is ω_c.

3. Next assume that the open-loop transfer function has a single (real) right-half plane zero ζ. Prove that asymptotically the closed-loop transfer function behaves as

$$H(s) \approx \frac{s-\zeta}{s+\zeta}\,\frac{k}{B_{n-q}(s/\omega_c)}. \tag{3.41}$$

Argue that this means that asymptotically the bandwidth is ζ (that is, the frequency response function $H(\mathrm{j}\omega)/H(0) - 1$, $\omega \in \mathbb{R}$, is small over the frequency range $[0,\ \zeta]$).

□

For the MIMO case the situation is more complex. The results may be summarized as follows. Let $R = \rho R_0$, with ρ a positive number, and R_0 a fixed positive-definite symmetric matrix. We study the root loci of the closed-loop poles as a function of ρ.

- As $\rho \to \infty$ the closed-loop poles approach those open-loop poles that lie in the left-half plane and the mirror images in the left-half plane of the right-half plane open-loop poles.
- If $\rho \downarrow 0$ then those closed-loop poles that remain finite approach the left-half plane zeros of $\det G^{\mathrm{T}}(-s)\,QG(s)$.

  If the open-loop transfer matrix $G(s) = D(sI - A)^{-1}B$ is square then we define the zeros of $\det G(s)$ as the open-loop zeros. In this case the closed-loop poles approach the left-half plane open-loop zeros and the left-half plane mirror images of the right-half plane open-loop zeros.

The closed-loop poles that do not remain finite as $\rho \downarrow 0$ go to infinity according to several Butterworth patterns of different orders and different radii. The number of patterns and their radii depend on the open-loop plant (Kwakernaak 1976).

Understanding the asymptotic behavior of the closed-loop poles provides insight into the properties of the closed-loop systems. We note some further facts:

- As $\rho \downarrow 0$ the gain matrix F approaches ∞, that is, some or all of its entries go to infinity.
- Assume that the open-loop transfer matrix $D(sI - A)^{-1}B$ is square, and that all its zeros are in the left-half plane. Then as we saw the closed-loop bandwidth inreases without bound as $\rho \downarrow 0$. Correspondingly, the solution X of the Riccati equation approaches the zero matrix.

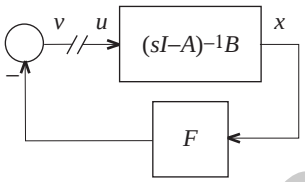


Figure 3.3: Loop gain for state feedback

3.2.7 Guaranteed gain and phase margins

If the state feedback loop is opened at the plant input as in Fig. 3.3 then the loop gain is $L(s) = F(sI - A)^{-1}B$. For the single-input case discussed in Subsection 3.2.6 (see Eqn. (3.31)) the return difference inequality (3.27) takes the form

$$|1 + L(j\omega)| \geq 1, \qquad \omega \in \mathbb{R}. \tag{3.42}$$

This inequality implies that the Nyquist plot of the loop gain stays outside the circle with center at -1 and radius 1. Figure 3.4 shows two possible behaviors of the Nyquist plot.

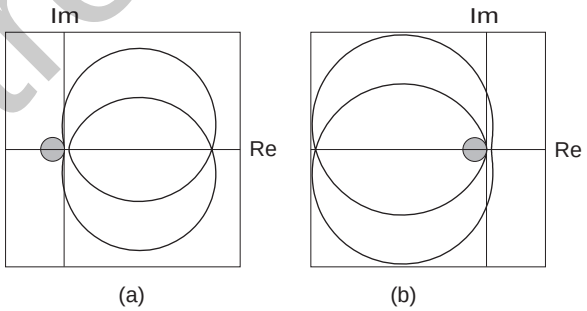


Figure 3.4: Examples of Nyquist plots of optimal loop gains.
(a): Open-loop stable plant. (b): Open-loop unstable plant with one right-half plane pole.
● = unit circle

Inspection shows that the modulus margin of the closed-loop system is 1. The gain margin is infinite and the phase margin is at least 60°. More precisely, for open-loop stable systems the gain

may vary between 0 and ∞ without destabilizing the system. For open-loop unstable systems it may vary between $\frac{1}{2}$ and ∞.

Exercise 3.2.5 (Phase margin). Prove that the phase margin is at least 60°. □

The guaranteed stability margins are very favorable. Some caution in interpreting these results is needed, however. The margins only apply to perturbations at the point where the loop is broken, that is, at the plant input. The closed-loop system may well be very sensitive to perturbations at any other point.

The SISO results may be generalized to the multi-input case. Suppose that the loop gain satisfies the return difference inequality (3.27). Assume that the loop gain $L(s)$ is perturbed to $W(s)L(s)$, with W a stable transfer matrix. It is proved in Subsection 3.7.3 (p. 144) of § 3.7, the appendix to this chapter, that the closed-loop *remains* stable provided

$$RW(\mathrm{j}\omega) + W^T(-\mathrm{j}\omega)R > R, \qquad \omega \in \mathbb{R}. \tag{3.43}$$

If both R and W are diagonal then this reduces to

$$W(\mathrm{j}\omega) + W^T(-\mathrm{j}\omega) > I, \qquad \omega \in \mathbb{R}. \tag{3.44}$$

This shows that if the ith diagonal entry W_i of W is real then it may have any value in the interval $(\frac{1}{2}, \infty)$ without destabilizing the closed-loop system. If the ith diagonal entry is $W_i(\mathrm{j}\omega) = \mathrm{e}^{\mathrm{j}\phi}$ then the closed-loop system remains stable as long as the angle ϕ is less than $\frac{\pi}{3}$, that is, 60°. Thus, the SISO results applies to each input channel separately.

3.2.8 Cross term in the performance index

In the optimal regulator problem for the stabilizable and detectable system

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t), \\ z(t) &= Dx(t), \end{aligned} \qquad t \geq 0, \tag{3.45}$$

we may consider the generalized quadratic performance index

$$J = \int_0^\infty \begin{bmatrix} z^{\mathrm{T}}(t) & u^{\mathrm{T}}(t) \end{bmatrix} \begin{bmatrix} Q & S \\ S^{\mathrm{T}} & R \end{bmatrix} \begin{bmatrix} z(t) \\ u(t) \end{bmatrix} dt. \tag{3.46}$$

We assume that

$$\begin{bmatrix} Q & S \\ S^{\mathrm{T}} & R \end{bmatrix} \tag{3.47}$$

is positive-definite. Define $v(t) = u(t) + R^{-1}S^{\mathrm{T}}z(t)$. Then minimization of J is equivalent to minimizing

$$J = \int_0^\infty [z^{\mathrm{T}}(t)(Q - SR^{-1}S^{\mathrm{T}})z(t) + v^{\mathrm{T}}(t)Rv(t)]\, dt \tag{3.48}$$

for the system

$$\dot{x}(t) = (A - BR^{-1}S^{\mathrm{T}}D)x(t) + Bv(t). \tag{3.49}$$

The condition that (3.47) be positive-definite is equivalent to the condition that both R and $Q - SR^{-1}S^{\mathrm{T}}$ be positive-definite (see Exercise 3.2.7, p. 114). Thus we satisfy the conditions of Summary 3.2.1. The Riccati equation now is

$$A^{\mathrm{T}}X + XA + D^{\mathrm{T}}QD - (XB + D^{\mathrm{T}}S)R^{-1}(B^{\mathrm{T}}X + S^{\mathrm{T}}D) = 0. \tag{3.50}$$

The optimal input for the system (3.45) is $u(t) = -Fx(t)$, with

$$F = R^{-1}(B^{\mathrm{T}}X + S^{\mathrm{T}}D). \tag{3.51}$$

Exercise 3.2.6 (LQ problem for system with direct feedthrough). The system

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t), \\ z(t) &= Dx(t) + Eu(t), \end{aligned} \qquad t \geq 0, \tag{3.52}$$

has what is called "direct feedthrough" because of the term with u in the output z. Show that the problem of minimizing

$$J = \int_0^\infty [z^{\mathrm{T}}(t)Qz(t) + u^{\mathrm{T}}(t)Ru(t)]\,dt \tag{3.53}$$

for this system may be converted into the cross term problem of this subsection. □

Exercise 3.2.7 (Positive-definiteness and Schur complement). Prove that the condition that the matrix (3.47) be positive-definite is equivalent to either of the following two conditions:

1. Both R and $Q - SR^{-1}S^{\mathrm{T}}$ are positive-definite. $Q - SR^{-1}S^{\mathrm{T}}$ is called the *Schur complement* of R. *Hint:*
$$\begin{bmatrix} Q & S \\ S^{\mathrm{T}} & R \end{bmatrix} = \begin{bmatrix} I & SR^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} Q - SR^{-1}S^{\mathrm{T}} & 0 \\ 0 & R \end{bmatrix} \begin{bmatrix} I & 0 \\ R^{-1}S^{\mathrm{T}} & I \end{bmatrix}. \tag{3.54}$$
2. Both Q and $R - S^{\mathrm{T}}Q^{-1}S$ are positive-definite. $R - S^{\mathrm{T}}Q^{-1}S$ is the Schur complement of Q.

□

3.2.9 Solution of the ARE

There are several algorithms for the solution of the algebraic Riccati equation (3.6) or (3.50). For all but the simplest problem recourse needs to be taken to numerical computer calculation. Equation (3.50) is the most general form of the Riccati equation. By redefining $D^{\mathrm{T}}QD$ as Q and $D^{\mathrm{T}}S$ as S the ARE (3.50) reduces to

$$A^{\mathrm{T}}X + XA + Q - (XB + S)R^{-1}(B^{\mathrm{T}}X + S^{\mathrm{T}}) = 0. \tag{3.55}$$

The most dependable solution method relies on the *Hamiltonian matrix*

$$H = \begin{bmatrix} A - BR^{-1}S^{\mathrm{T}} & -BR^{-1}B^{\mathrm{T}} \\ -Q + SR^{-1}S^{\mathrm{T}} & -(A - BR^{-1}S^{\mathrm{T}})^{\mathrm{T}} \end{bmatrix} \tag{3.56}$$

associated with the LQ problem. Under the assumptions of Summary 3.2.1 (p. 105) or § 3.2.7 (p. 112) the Hamiltonian matrix H has no eigenvalues on the imaginary axis. If λ is an eigenvalue of the $2n \times 2n$ matrix H then $-\lambda$ is also an eigenvalue. Hence, H has exactly n eigenvalues

with negative real part. Let the columns of the real $2n \times n$ matrix E form a basis for the n-dimensional space spanned by the eigenvectors and generalized eigenvectors of H corresponding to the eigenvalues with strictly negative real parts. Partition

$$E = \begin{bmatrix} E_1 \\ E_2 \end{bmatrix}, \tag{3.57}$$

with E_1 and E_2 both square. It is proved in § 3.7.4 (p. 145) that

$$X = E_2 E_1^{-1} \tag{3.58}$$

is the desired solution of the algebraic Riccati equation.

E may efficiently be computed by Schur decomposition (Golub and Van Loan 1983)

$$H = UTU^{\mathrm{H}} \tag{3.59}$$

of the Hamiltonian matrix H. U is unitary, that is, $UU^{\mathrm{H}} = U^{\mathrm{H}}U = I$. I is a unit matrix and the superscript H denotes the complex-conjugate transpose. T is upper triangular, that is, all entries below the main diagonal of T are zero. The diagonal entries of T are the eigenvalues of H. The diagonal entries of T may be arranged in any order. In particular, they may be ordered such that the eigenvalues with negative real part precede those with positive real parts. Partition $U = [U_1 \;\; U_2]$, where U_1 and U_2 both have n columns. Then the columns of U_1 span the same subspace as the eigenvectors and generalized eigenvectors corresponding to the eigenvalues of H with negative real parts. Hence, we may take $E = U_1$.

This is the algorithm implemented in most numerical routines for the solution of algebraic Riccati equations. An up-to-date account of the numerical aspects of the solution of the ARE may be found in Sima (1996).

Exercise 3.2.8 ("Cruise control system"). Use the method of this subsection to solve the ARE that arises in Exercise 3.2.2 (p. 106). □

3.2.10 Concluding remarks

The LQ paradigm would appear to be useless as a design methodology because full state feedback is almost never feasible. Normally it simply is too costly to install the instrumentation needed to measure all the state variables. Sometimes it is actually impossible to measure some of the state variables.

In Section 3.3 (p. 115) we see how instead of using state feedback control systems may be designed based on feedback of selected output variables only. The idea is to reconstruct the state as accurately as possible using an observer or Kalman filter. By basing feedback on *estimates* of the state several of the favorable properties of state feedback may be retained or closely recovered.

3.3 LQG Theory

3.3.1 Introduction

In this section we review what is known as LQG theory. LQG stands for Linear Quadratic Guassian. By including Gaussian white noise in the LQ paradigm linear optimal feedback systems based on *output feedback* rather than state feedback may be found.

We consider the system

$$\left.\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t) + Gv(t) \\ y(t) &= Cx(t) + w(t) \\ z(t) &= Dx(t) \end{aligned}\right\} \qquad t \in \mathbb{R}. \tag{3.60}$$

The *measured output* y is available for feedback. As in § 2.3 (p. 64) the output z is the *controlled output.* The noise signal v models the *plant disturbances* and w the *measurement noise.*

The signals v and w are vector-valued Gaussian white noise processes with

$$\left.\begin{aligned} Ev(t)v^{\mathrm{T}}(s) &= V\delta(t-s) \\ Ev(t)w^{\mathrm{T}}(s) &= 0 \\ Ew(t)w^{\mathrm{T}}(s) &= W\delta(t-s) \end{aligned}\right\} \qquad t, s \in \mathbb{R}. \tag{3.61}$$

V and W are nonnegative-definite symmetric constant matrices, called the *intensity matrices* of the two white noise processes. We do not go into the theory of stochastic processes in general and that of white noise in particular, but refer to texts such as Wong (1983) and Bagchi (1993). The initial state $x(0)$ is assumed to be a random vector.

The various assumptions define the state $x(t)$, $t \in \mathbb{R}$, and the controlled output $z(t)$, $t \in \mathbb{R}$, as random processes. As a result, also the quadratic error expression

$$z^{\mathrm{T}}(t)Qz(t) + u^{\mathrm{T}}(t)Ru(t), \qquad t \geq 0, \tag{3.62}$$

is a random process. The problem of controlling the system such that the integrated expected value

$$\int_0^T E[z^{\mathrm{T}}(t)Qz(t) + u^{\mathrm{T}}(t)Ru(t)]\,dt \tag{3.63}$$

is minimal is the *stochastic linear regulator problem.* The time interval $[0, T]$ at this point is taken to be finite but eventually we consider the case that $T \to \infty$. At any time t the entire past measurement signal $y(s)$, $s \leq t$, is assumed to be available for feedback. Figure 3.5 clarifies the situation.

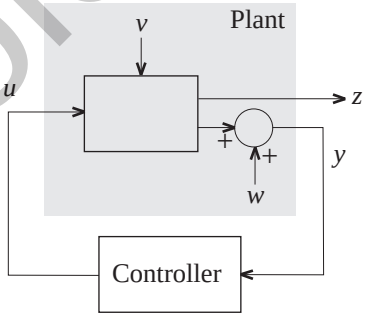


Figure 3.5: LQG feedback

3.3.2 Observers

Consider the observed system

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t), \\ y(t) &= Cx(t), \end{aligned} \qquad t \in \mathbb{R}. \tag{3.64}$$

This is the system (3.60) but without the state noise v and the measurement noise w. The state x of the system (3.64) is not directly accessible because only the output y is measured. We may reconstruct the state with arbitrary precision by connecting an *observer* of the form

$$\dot{\hat{x}}(t) = A\hat{x}(t) + Bu(t) + K[y(t) - C\hat{x}], \qquad t \in \mathbb{R}. \tag{3.65}$$

The signal $\hat{x}$ is meant to be an estimate of the state $x(t)$. It satisfies the state differential equation of the system (3.64) with an additional input term $K[y(t) - C\hat{x}(t)]$ on the right-hand side. K is the *observer gain matrix*. It needs to be suitably chosen. The *observation error* $y(t) - C\hat{x}(t)$ is the difference between the actual measured output $y(t)$ and the output $\hat{y}(t) = C\hat{x}(t)$ as reconstructed from the estimated state $\hat{x}(t)$. The extra input term $K[y(t) - C\hat{x}(t)]$ on the right-hand side of (3.65) provides a correction that is active as soon as the observation error is nonzero. Figure 3.6

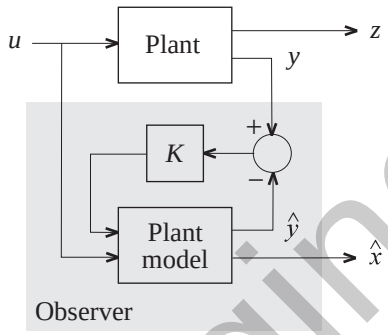


Figure 3.6: Structure of an observer

shows the structure of the observer. Define

$$e(t) = \hat{x}(t) - x(t) \tag{3.66}$$

as the *state estimation error*. Differentiation of e yields after substitution of (3.65) and (3.64) that the error satisfies the differential equation

$$\dot{e}(t) = (A - KC)e(t), \qquad t \in \mathbb{R}. \tag{3.67}$$

If the system (3.64) is detectable then there always exists a gain matrix K such that the error system (3.67) is stable. If the error system is stable then $e(t) \to 0$ as $t \to \infty$ for any initial error $e(0)$. Hence,

$$\hat{x}(t) \overset{t\to\infty}{\longrightarrow} x(t), \tag{3.68}$$

so that the estimated state converges to the actual state.

3.3.3 The Kalman filter

Suppose that we connect the observer

$$\dot{\hat{x}}(t) = A\hat{x}(t) + Bu(t) + K[y(t) - C\hat{x}(t)], \qquad t \in \mathbb{R}. \tag{3.69}$$

to the noisy system

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t) + Gv(t), \\ y(t) &= Cx(t) + w(t), \end{aligned} \qquad t \in \mathbb{R}. \tag{3.70}$$

Differentiation of $e(t) = \hat{x}(t) - x(t)$ leads to the error differential equation

$$\dot{e}(t) = (A - KC)e(t) - Gv(t) + Kw(t), \qquad t \in \mathbb{R}. \tag{3.71}$$

Owing to the two noise terms on the right-hand side the error now no longer converges to zero, even if the error system is stable. Suppose that the error system is stable. It is proved in § 3.7.7 (p. 149) that as $t \to \infty$ the *error covariance matrix*

$$Ee(t)e^{\mathrm{T}}(t) \tag{3.72}$$

converges to a constant steady-state value Y that satisfies the linear matrix equation

$$(A - KC)Y + Y(A - KC)^{\mathrm{T}} + GVG^{\mathrm{T}} + KWK^{\mathrm{T}} = 0. \tag{3.73}$$

This type of matrix equation is known as a *Lyapunov equation*. It is made plausible in Subsection 3.7.5 that as a function of the gain matrix K the steady-state error covariance matrix Y is *minimal* if K is chosen as

$$K = YC^{\mathrm{T}}W^{-1}. \tag{3.74}$$

"Minimal" means here that if $\bar{Y}$ is the steady-state error covariance matrix corresponding to any other observer gain $\bar{K}$ then $\bar{Y} \geq Y$. This inequality is to be taken in the sense that $\bar{Y} - Y$ is nonnegative-definite.

A consequence of this result is that the gain (3.74) minimizes the steady-state mean square state reconstruction error $\lim_{t\to\infty} Ee^{\mathrm{T}}(t)e(t)$. As a matter of fact, the gain minimizes the weighted mean square construction error $\lim_{t\to\infty} Ee^{\mathrm{T}}(t)W_e e(t)$ for any nonnegative-definite weighting matrix W_e.

Substitution of the optimal gain matrix (3.74) into the Lyapunov equation (3.73) yields

$$AY + YA^{\mathrm{T}} + GVG^{\mathrm{T}} - YC^{\mathrm{T}}W^{-1}CY = 0. \tag{3.75}$$

This is another matrix Riccati equation. The observer

$$\dot{\hat{x}}(t) = A\hat{x}(t) + Bu(t) + K[y(t) - C\hat{x}], \qquad t \in \mathbb{R}, \tag{3.76}$$

with the gain chosen as in (3.74) and the covariance matrix Y the nonnegative-definite solution of the Riccati equation (3.75) is the famous *Kalman filter* (Kalman and Bucy 1961).

We review several properties of the Kalman filter. They are the duals of the properties listed in Summary 3.2.1 (p. 105) for the Riccati equation associated with the regulator problem[6] .

Summary 3.3.1 (Properties of the Kalman filter).
Assumptions:

- The system

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Gv(t), \\ y(t) &= Cx(t), \end{aligned} \qquad t \in \mathbb{R}, \tag{3.77}$$

  is stabilizable and detectable.

[6]The optimal regulator and the Kalman filter are dual in the following sense. Given the regulator problem of § 3.2 (p. 104), replace A with A^{T}, B with C^{T}, D with G^{T}, Q with V, and R with W. Then the regulator Riccati equation (3.6) becomes the observer Riccati equation (3.75), its solution X becomes Y, the state feedback gain F is the transpose of the observer gain K, and the closed-loop system matrix $A - BF$ is the transpose of the error system matrix $A - KC$. By matching substitutions the observer problem may be transposed to a regulator problem.

- The noise intensity matrices V and W are positive-definite.

The following facts follow from Summary 3.2.1 (p. 105) by duality:

1. The algebraic Riccati equation

$$AY + YA^{\mathrm{T}} + GVG^{\mathrm{T}} - YC^{\mathrm{T}}W^{-1}CY = 0 \tag{3.78}$$

has a unique nonnegative-definite symmetric solution Y. If the system (3.77) is controllable rather than just stabilizable then Y is positive-definite.

2. The minimal value of the steady-state weighted mean square state reconstruction error $\lim_{t\to\infty} Ee^{\mathrm{T}}(t)W_e e(t)$ is[7] tr YW_e.

3. The minimal value of the mean square reconstruction error is achieved by the observer gain matrix $K = YC^{\mathrm{T}}W^{-1}$.

4. The error system

$$\dot{e}(t) = (A - KC)e(t), \qquad t \in \mathbb{R}, \tag{3.79}$$

is stable, that is, all the eigenvalues of the matrix $A - KC$ have strictly negative real parts. As a consequence also the observer

$$\dot{\hat{x}}(t) = A\hat{x}(t) + Bu(t) + K[y(t) - C\hat{x}(t)], \qquad t \in \mathbb{R}, \tag{3.80}$$

is stable.

□

The reasons for the assumptions may be explained as follows. If the system (3.77) is not detectable then no observer with a stable error system exists. If the system is not stabilizable (with v as input) then there exist observers that are not stable but are immune to the state noise v. Hence, stability of the error system is not guaranteed. W needs to be positive-definite to prevent the Kalman filter from having infinite gain. If V is not positive-definite then there may be unstable modes that are not excited by the state noise and, hence, are not stabilized in the error system.

Exercise 3.3.2 (Asymptotic results). The asymptotic results for the regulator problem may be "dualized" to the Kalman filter.

1. Define the "observer return difference"

$$J_f(s) = I + C(sI - A)^{-1}K. \tag{3.81}$$

Prove that

$$\det J_f(s) = \frac{\chi_f(s)}{\chi_{\mathrm{ol}}(s)}, \tag{3.82}$$

where $\chi_{\mathrm{ol}}(s) = \det(sI - A)$ is the system characteristic polynomial and $\chi_f(s) = \det(sI - A + KC)$ the observer characteristic polynomial.

[7] The quantity tr $M = \sum_{i=1}^{m} M_{ii}$ is called the *trace* of the $m \times m$ matrix M with entries M_{ij}, $i, j = 1, 2, \cdots, m$.

2. Prove that the return difference J_f of the Kalman filter satisfies

$$J_f(s)WJ_f^{\mathrm{T}}(-s) = W + M(s)VM^{\mathrm{T}}(-s), \tag{3.83}$$

where M is the open-loop transfer matrix $M(s) = C(sI - A)^{-1}G$.

3. Consider the SISO case with $V = 1$ and $W = \sigma$. J_f and M are now scalar functions. Prove that

$$\chi_f(s)\chi_f(-s) = \chi_{\mathrm{ol}}(s)\chi_{\mathrm{ol}}(-s)[1 + \frac{1}{\sigma}M(s)M(-s)]. \tag{3.84}$$

4. Write

$$M(s) = g\frac{\zeta(s)}{\chi_{\mathrm{ol}}(s)}, \tag{3.85}$$

with ζ a monic polynomial and g a constant.

- Prove that as $\sigma \to \infty$ the optimal observer poles approach the open-loop poles that lie in the left-half plane and the mirror images of the open-loop poles that lie in the right-half plane.
- Prove that as $\sigma \downarrow 0$ the optimal observer poles that do not go to ∞ approach the open-loop zeros that lie in the left-half plane and the mirror images of the open-loop zeros that lie in the right-half plane.
- Establish the asymptotic pattern of the optimal observer poles that approach ∞ as $\sigma \downarrow 0$.

□

3.3.4 Kalman filter with cross correlated noises

A useful generalization of the Kalman filter follows by assuming cross correlation of the white noise processes v and w. Suppose that

$$E\begin{bmatrix} v(t) \\ w(t) \end{bmatrix}\begin{bmatrix} v^{\mathrm{T}}(s) & w^{\mathrm{T}}(s) \end{bmatrix} = \begin{bmatrix} V & U \\ U^{\mathrm{T}} & W \end{bmatrix}\delta(t-s), \qquad t, s \in \mathbb{R}. \tag{3.86}$$

Assume that

$$\begin{bmatrix} V & U \\ U^{\mathrm{T}} & W \end{bmatrix} \tag{3.87}$$

is positive-definite, and, as before, that the system $\dot{x}(t) = Ax(t) + Gv(t)$, $y(t) = Cx(t)$ is stabilizable and detectable. Then the optimal observer gain is

$$K = (YC^{\mathrm{T}} + GU)W^{-1}, \tag{3.88}$$

where the steady-state error covariance matrix Y is the positive-definite solution of the Riccati equation

$$AY + YA^{\mathrm{T}} + GVG^{\mathrm{T}} - (YC^{\mathrm{T}} + GU)W^{-1}(CY + U^{\mathrm{T}}G^{\mathrm{T}}) = 0. \tag{3.89}$$

Exercise 3.3.3 (Cross correlated noises). Prove this result. □

3.3.5 Solution of the stochastic linear regulator problem

The stochastic linear regulator problem consists of minimizing

$$\int_0^T E[z^{\mathrm{T}}(t)Qz(t) + u^{\mathrm{T}}(t)Ru(t)]\,dt \tag{3.90}$$

for the system

$$\left.\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t) + Gv(t) \\ y(t) &= Cx(t) + w(t) \\ z(t) &= Dx(t) \end{aligned}\right\} \qquad t \in \mathbb{R}. \tag{3.91}$$

We successively consider the situation of no state noise, state feedback, and output feedback.

No state noise. From § 3.2 (p. 139) we know that if the disturbance v is absent and the state $x(t)$ may be directly and accurately accessed for measurement, then for $T \to \infty$ the performance index is minimized by the state feedback law

$$u(t) = -Fx(t), \tag{3.92}$$

with the feedback gain F as in Summary 3.2.1 (p. 105).

State feedback. If the white noise disturbance v is present then the state and input cannot be driven to 0, and the integrated generalized square error (3.90) does not converge to a finite number as $T \to \infty$. It is proved in Subsection 3.7.6 (p. 148) that the state feedback law (3.92) minimizes the *rate* at which (3.90) approaches ∞, that is, it minimizes

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T E[z^{\mathrm{T}}(t)Qz(t) + u^{\mathrm{T}}(t)Ru(t)]\,dt. \tag{3.93}$$

This limit equals the *steady-state mean square error* indexsteady-state mean square error

$$\lim_{t\to\infty} E[z^{\mathrm{T}}(t)Qz(t) + u^{\mathrm{T}}(t)Ru(t)]. \tag{3.94}$$

Hence, the state feedback law minimizes the steady-state mean square error.

Output feedback. We next consider the situation that the state *cannot* be accessed for measurement. The state may be optimally estimated, however, with the help of the Kalman filter. Then the solution of the stochastic linear regulator problem with *output feedback* (rather than state feedback) is to replace the state $x(t)$ in the state feedback law (3.92) with the *estimated state* $\hat{x}(t)$. Thus, the optimal controler is given by

$$\begin{aligned} \dot{\hat{x}}(t) &= A\hat{x}(t) + Bu(t) + K[y(t) - C\hat{x}(t)], \\ u(t) &= -F\hat{x}(t), \end{aligned} \qquad t \in \mathbb{R}. \tag{3.95}$$

The controller minimizes the steady-state mean square error (3.94) under output feedback. The feedback gain F and the observer gain K follow from Summaries 3.2.1 (p. 105) and 3.3.1 (p. 118), respectively. Figure 3.7 shows the arrangement of the closed-loop system.

Using the estimated state as if it were the actual state is known as *certainty equivalence*. It divorces state estimation and control input selection. This idea is often referred to as the *separation principle*.

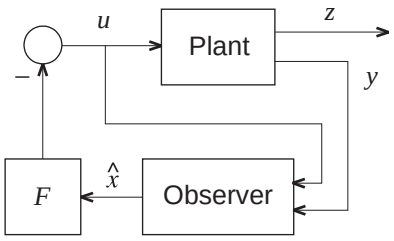


Figure 3.7: Observer based feedback control

The closed-loop system that results from interconnecting the plant (3.91) with the compensator (3.95) is stable — under the assumptions of Summaries 3.2.1 and 3.3.1, of course. This is most easily recognized as follows. Substitution of $u(t) = -F\hat{x}(t)$ into $\dot{x}(t) = Ax(t) + Bu(t) + Gv(t)$ yields with the further substitution $\hat{x}(t) = x(t) + e(t)$

$$\dot{x}(t) = (A - BF)x(t) - BFe(t) + Gv(t). \tag{3.96}$$

Together with (3.71) we thus have

$$\begin{bmatrix} \dot{x}(t) \\ \dot{e}(t) \end{bmatrix} = \begin{bmatrix} A - BF & -BF \\ 0 & A - KC \end{bmatrix} \begin{bmatrix} x(t) \\ e(t) \end{bmatrix} + \begin{bmatrix} Gv(t) \\ -Gv(t) + Kw(t) \end{bmatrix}. \tag{3.97}$$

The eigenvalues of this system are the eigenvalues of the closed-loop system. Inspection shows that these eigenvalues consist of the eigenvalues of $A - BF$ (the *regulator poles*) together with the eigenvalues of $A - KC$ (the *observer poles*). If the plant (3.91) has order n then the compensator also has order n. Hence, there are $2n$ closed-loop poles.

Exercise 3.3.4 (Closed-loop eigenvalues).

1. Prove that the eigenvalues of (3.97) are the eigenvalues of the closed-loop system.
2. Show (most easily by a counterexample) that the fact that the observer and the closed-loop system are stable does not mean that the compensator (3.95) by itself is stable.

□

Exercise 3.3.5 (Compensator transfer function). The configuration of Fig. 3.7 may be rearranged as in Fig. 3.8. Show that the equivalent compensator C_e has the transfer matrix $C_e(s) = F(sI - A + BF + KC)^{-1}K$. □

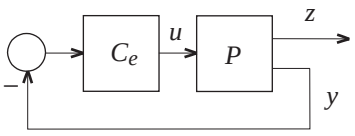


Figure 3.8: Equivalent unit feedback configuration

3.3.6 Asymptotic analysis and loop transfer recovery

In this subsection we study the effect of decreasing the intensity W of the measurement noise. Suppose that $W = \sigma W_0$, with W_0 a fixed symmetric positive-definite weighting matrix and σ a

positive number. We investigate the asymptotic behavior of the closed-loop system as $\sigma \downarrow 0$. Before doing this we need to introduce two assumptions:

- The disturbance v is additive to the plant input u, that is, $G = B$. This allows the tightest control of the disturbances.
- The open-loop plant transfer matrix $C(sI - A)^{-1}B$ is square, and its zeros all have negative real parts.

Breaking the loop at the plant input as in Fig. 3.9 we obtain the loop gain

$$L_\sigma(s) = C_e(s)P(s) = F(sI - A + BF + K_\sigma C)^{-1}K_\sigma C(sI - A)^{-1}B. \tag{3.98}$$

(Compare Exercise 3.3.5, p. 122.) To emphasize the dependence on σ the observer gain and the

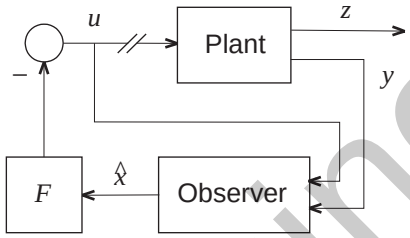


Figure 3.9: Breaking the loop

loop gain are each provided with a subscript. As $\sigma \downarrow 0$ the gain K_σ approaches ∞. At the same time the error covariance matrix Y_σ approaches the zero matrix. This is the dual of the conclusion of Subsection 3.2.6 (p. 109) that the state feedback gain F goes to ∞ and the solution X of the Riccati equation approaches the zero matrix as the weight on the input decreases.

The fact that $Y_\sigma \downarrow 0$ indicates that in the limit the observer reconstructs the state with complete accuracy. It is proved in § 3.7.7 (p. 149) that as $\sigma \downarrow 0$ the loop gain L_σ approaches the expression

$$L_0(s) = F(sI - A)^{-1}B. \tag{3.99}$$

The asymptotic loop gain L_0 is precisely the loop gain for full state feedback. Accordingly, the guaranteed gain and phase margins of Subsection § 3.2.7 (p. 112) are recouped. This is called *loop transfer recovery* (LTR).

The term loop transfer recovery appears to have been coined by Doyle and Stein (1981). Extensive treatments may be found in Anderson and Moore (1990) and Saberi, Chen, and Sannuti (1993). We use the method in the design examples of § 3.6 (p. 127).

Exercise 3.3.6 (Dual loop transfer recovery). Dual loop recovery provides an alternative approach to loop recovery. Dual loop recovery results when the loop is broken at the plant *output* y rather than at the input (Kwakernaak and Sivan 1972, § 5.6). In this case it is necessary to assume that $D = C$, that is, the controlled output is measured. Again we need $C(sI - A)^{-1}B$ to be square with left-half plane zeros only. We let $R = \rho R_0$ and consider the asymptotic behavior for $\rho \downarrow 0$.

1. Make it plausible that the loop gain approaches

$$L_0(s) = C(sI - A)^{-1}K. \tag{3.100}$$

2. Show that the corresponding return difference $J_0(s) = I + L_0(s)$ satisfies the return difference inequality

$$J_0(\mathrm{j}\omega) W J_0^{\mathrm{T}}(-\mathrm{j}\omega) \geq W, \qquad \omega \in \mathbb{R}. \tag{3.101}$$

3. Show that gain and phase margins apply that agree with those found in Subsection 3.2.7 (p. 112).

□

3.4 H_2 optimization

3.4.1 Introduction

In this section we define the LQG problem as a special case of a larger class of problems, which lately has become known as H_2 optimization. Most importantly, this approach allows to remove the stochastic ingredient of the LQG formulation. In many applications it is difficult to establish the precise stochastic properties of disturbances and noise signals. Very often in the application of the LQG problem to control system design the noise intensities V and W play the role of *design parameters* rather than that they model reality.

The stochastic element is eliminated by recognizing that the performance index for the LQG problem may be represented as a *system norm*—the H_2-norm. To introduce this point of view, consider the stable system

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bv(t), \\ y(t) &= Cx(t), \end{aligned} \qquad t \in \mathbb{R}. \tag{3.102}$$

The system has the transfer matrix $H(s) = C(sI - A)^{-1}B$. Suppose that the signal v is white noise with covariance function $Ev(t)v^{\mathrm{T}}(s) = V\delta(t-s)$. Then the output y of the system is a stationary stochastic process with spectral density matrix

$$S_y(f) = H(\mathrm{j}2\pi f) V H^{\mathrm{T}}(-\mathrm{j}2\pi f), \qquad f \in \mathbb{R}. \tag{3.103}$$

As a result, the mean square output is

$$Ey^{\mathrm{T}}(t)y(t) = \operatorname{tr} \int_{-\infty}^{\infty} S_y(f)\, df = \operatorname{tr} \int_{-\infty}^{\infty} H(\mathrm{j}2\pi f) V H^{\sim}(\mathrm{j}2\pi f)\, df. \tag{3.104}$$

Here we introduce the notation $H^{\sim}(s) = H^{\mathrm{T}}(-s)$. The quantity

$$\|H\|_2 = \sqrt{\operatorname{tr} \int_{-\infty}^{\infty} H(\mathrm{j}2\pi f) H^{\sim}(\mathrm{j}2\pi f)\, df} \tag{3.105}$$

is called the H_2-*norm* of the system. If the white noise v has intensity $V = I$ then the mean square output $Ey^{\mathrm{T}}(t)y(t)$ equals precisely the square of the H_2-norm of the system.

H_2 refers to the space of square integrable functions on the imaginary axis whose inverse Fourier transform is zero for negative time.

Exercise 3.4.1 (Lyapunov equation). Prove that the H_2-norm of the stable system (3.102) is given by $\|H\|_2^2 = \operatorname{tr} CYC^{\mathrm{T}}$, where the matrix Y is the unique symmetric solution of the Lyapunov equation $AY + YA^{\mathrm{T}} + BB^{\mathrm{T}} = 0$. □

3.4.2 H_2 optimization

In this subsection we rewrite the time domain LQG problem into an equivalent frequency domain H_2 optimization problem. While the LQG problem requires state space realizations, the H_2-optimization problem is in terms of transfer matrices. To simplify the expressions to come we assume that $Q = I$ and $R = I$, that is, the LQG performance index is

$$\lim_{t\to\infty} E[z^{\mathrm{T}}(t)z(t) + u^{\mathrm{T}}(t)u(t)]. \tag{3.106}$$

This assumption causes no loss of generality because by scaling and transforming the variables z and u the performance index may always be brought into this form.

For the open-loop system

$$\dot{x} = Ax + Bu + Gv, \tag{3.107}$$
$$z = Dx, \tag{3.108}$$
$$y = Cx + w \tag{3.109}$$

we have in terms of transfer matrices

$$z = \underbrace{D(sI-A)^{-1}G}_{P_{11}(s)}\, v + \underbrace{D(sI-A)^{-1}B}_{P_{12}(s)}\, u, \tag{3.110}$$
$$y = \underbrace{C(sI-A)^{-1}G}_{P_{21}(s)}\, v + \underbrace{C(sI-A)^{-1}B}_{P_{22}(s)}\, u + w. \tag{3.111}$$

Interconnecting the system as in Fig. 3.10 with a compensator C_e we have the signal balance

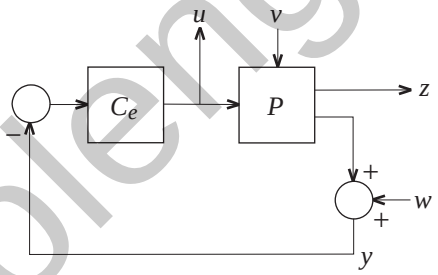


Figure 3.10: Feedback system with stochastic inputs and outputs

$u = -C_e y = -C_e(P_{21}v + P_{22}u + w)$, so that

$$u = \underbrace{-(I + C_eP_{22})^{-1}C_eP_{21}}_{H_{21}(s)}\, v \underbrace{-(I + C_eP_{22})^{-1}C_e}_{H_{22}(s)}\, w. \tag{3.112}$$

From $z = P_{11}v + P_{12}u$ we obtain

$$z = \underbrace{P_{11} - P_{12}(I + C_eP_{22})^{-1}C_eP_{21}}_{H_{11}(s)}\, v \underbrace{-P_{12}(I + C_eP)^{-1}C_e}_{H_{12}(s)}\, w. \tag{3.113}$$

A more compact notation is

$$\begin{bmatrix} z \\ u \end{bmatrix} = \underbrace{\begin{bmatrix} H_{11}(s) & H_{12}(s) \\ H_{21}(s) & H_{22}(s) \end{bmatrix}}_{H(s)} \begin{bmatrix} v \\ w \end{bmatrix}. \tag{3.114}$$

From this we find for the steady-state mean square error

$$\lim_{t\to\infty} E\big(z^{\mathrm{T}}(t)z(t)+u^{\mathrm{T}}(t)u(t)\big) = \lim_{t\to\infty} E\Big(\begin{bmatrix}z(t)\\u(t)\end{bmatrix}^{\mathrm{T}}\begin{bmatrix}z(t)\\u(t)\end{bmatrix}\Big) \tag{3.115}$$

$$= \operatorname{tr}\int_{-\infty}^{\infty} H(\mathrm{j}2\pi f)H^{\sim}(\mathrm{j}2\pi f)\,df \tag{3.116}$$

$$= \|H\|_2^2. \tag{3.117}$$

Hence, solving the LQG problem amounts to minimizing the H_2 norm of the closed-loop system of Fig. 3.10 with (v, w) as input and (z, u) as output.

The configuration of Fig. 3.10 is a special case of the configuration of Fig. 3.11. In the latter diagram w is the *external input* (v and w in Fig. 3.10). The signal z is the *error signal*, which ideally should be zero (z and u in Fig. 3.10). Furthermore, u is the *control input*, and y the *observed output*. The block G is the *generalized plant*, and K the compensator. Note that the sign reversion at the compensator input in Fig. 3.10 has been absorbed into the compensator K.

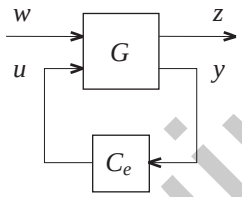


Figure 3.11: The standard H_2 problem

Exercise 3.4.2 (Generalized plant for the LQG problem). Show that for the LQG problem the generalized plant in state space form may be represented as

$$\dot{x}(t) = Ax(t)+\begin{bmatrix}G & 0\end{bmatrix}\begin{bmatrix}v(t)\\w(t)\end{bmatrix}+Bu(t), \tag{3.118}$$

$$\begin{bmatrix}z(t)\\u(t)\\y(t)\end{bmatrix} = \begin{bmatrix}D\\0\\C\end{bmatrix}x(t)+\begin{bmatrix}0&0\\0&0\\0&I\end{bmatrix}\begin{bmatrix}v(t)\\w(t)\end{bmatrix}+\begin{bmatrix}0\\I\\0\end{bmatrix}u(t). \tag{3.119}$$

□

3.4.3 The standard H_2 problem and its solution

The *standard* H_2 optimization problem is the problem of choosing the compensator K in the block diagram of Fig. 3.11 such that it

1. stabilizes the closed-loop system, and
2. minimizes the H_2-norm of the closed-loop system (with w as input and z as output).

We represent the generalized plant G of Fig. 3.11 in state space form as

$$\dot{x}(t) = Ax(t)+B_1w(t)+B_2u(t), \tag{3.120}$$

$$z(t) = C_1x(t)+D_{11}w(t)+D_{12}u(t), \tag{3.121}$$

$$y(t) = C_2x(t)+D_{21}w(t)+D_{22}u(t). \tag{3.122}$$

The H_2 problem may be solved by reducing it to an LQG problem. This is done in § 3.7.8 (p. 149). The derivation necessitates the introduction of some assumptions, which are listed in the summary that follows. They are natural assumptions for LQG problems.

Summary 3.4.3 (Solution of the H_2 problem). Consider the standard H_2 optimization problem for the generalized plant

$$\begin{aligned} \dot{x}(t) &= Ax(t) + B_1 w(t) + B_2 u(t), && (3.123)\\ z(t) &= C_1 x(t) + D_{12} u(t), && (3.124)\\ y(t) &= C_2 x(t) + D_{21} w(t) + D_{22} u(t). && (3.125) \end{aligned}$$

Assumptions:

- The system $\dot{x}(t) = Ax(t) + B_2 u(t)$, $z(t) = C_1 x(t)$ is stabilizable and detectable.
- The system $\dot{x}(t) = Ax(t) + B_1 w(t)$, $y(t) = C_2 x(t)$ is stabilizable and detectable.
- The matrix $\left[\begin{smallmatrix} A-sI & B_1 \\ C_2 & D_{21} \end{smallmatrix}\right]$ has full row rank for every $s = \mathrm{j}\omega$, and D_{21} has full row rank.
- The matrix $\left[\begin{smallmatrix} A-sI & B_2 \\ C_1 & D_{12} \end{smallmatrix}\right]$ has full column rank for every $s = \mathrm{j}\omega$, and D_{12} has full column rank.

Under these assumptions the optimal output feedback controller is

$$\begin{aligned} \dot{\hat{x}}(t) &= A\hat{x}(t) + B_2 u(t) + K[y(t) - C_2\hat{x}(t) - D_{22}u(t)] && (3.126)\\ u(t) &= -F\hat{x}(t). && (3.127) \end{aligned}$$

The observer and state feedback gain matrices are

$$F = (D_{12}^{\mathrm{T}} D_{12})^{-1}(B_2^{\mathrm{T}} X + D_{12}^{\mathrm{T}} C_1), \quad K = (YC_2^{\mathrm{T}} + B_1 D_{21}^{\mathrm{T}})(D_{21} D_{21}^{\mathrm{T}})^{-1}. \tag{3.128}$$

The symmetric matrices X and Y are the unique positive-definite solutions of the algebraic Riccati equations

$$\begin{aligned} A^{\mathrm{T}}X + XA + C_1^{\mathrm{T}}C_1 - (XB_2 + C_1^{\mathrm{T}}D_{12})(D_{12}^{\mathrm{T}}D_{12})^{-1}(B_2^{\mathrm{T}}X + D_{12}^{\mathrm{T}}C_1) &= 0,\\ AY + AY^{\mathrm{T}} + B_1 B_1^{\mathrm{T}} - (YC_2^{\mathrm{T}} + B_1 D_{21}^{\mathrm{T}})(D_{21}D_{21}^{\mathrm{T}})^{-1}(C_2 Y + D_{21}B_1^{\mathrm{T}}) &= 0. \end{aligned} \tag{3.129}$$

□

The condition that D_{12} has full column rank means that there is "direct feedthrough" from the input u to the error signal z. Likewise, the condition that D_{21} has full row rank means that the noise w is directly fed through to the observed output y.

The H_2 optimization problem and its solution are discussed at length in Saberi, Sannuti, and Chen (1995). In §§ 3.5 (p. 127) and 3.6 (p. 133) we discuss the application of H_2 optimization to control system design.

3.5 Feedback system design by H_2 optimization

3.5.1 Introduction

In this section we review how LQG and H_2 optimization may be used to design SISO and MIMO linear beedback systems.

3.5.2 Parameter selection for the LQG problem

We discuss how to select the design parameters for the LQG problem without a cross term in the performance index and without cross correlation between the noises. The LQG problem consists of the minimization of

$$\lim_{t\to\infty} E[z^{\mathrm{T}}(t)Qz(t) + u^{\mathrm{T}}(t)Ru(t)] \tag{3.130}$$

for the system

$$\left.\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t) + Gv(t) \\ z(t) &= Dx(t) \\ y(t) &= Cx(t) + w(t) \end{aligned}\right\} \qquad t \in \mathbb{R}. \tag{3.131}$$

Important design parameters are the weighting matrices Q and R and the intensities V and W. In the absence of specific information about the nature of the disturbances also the noise input matrix G may be viewed as a design parameter. Finally there usually is some freedom in the selection of the control output z; this means that also the matrix D may be considered a design parameter.

We discuss some rules of thumb for selecting the design parameters. They are based on the assumption that we operate in the asymptotic domain where the weighting matrices R (the weight on the input) and W (the measurement noise intensity) are small.

1. First the parameters D, Q and R for the *regulator part* are selected. These quantities determine the dominant characteristics of the closed-loop system.

   (a) D determines the controlled output z. Often the controlled output z is also the measured output y. The case where z is not y is called *inferential control*. There may be compelling engineering reasons for selecting z different from y.

   (b) In the SISO case Q may be chosen equal to 1.

   In the MIMO case Q is best chosen to be diagonal according to the rules of § 3.2.3 (p. 106).

   (c) In the SISO case R is a scalar design parameter. It is adjusted by trial and error until the desired bandwidth is achieved (see also § 3.2.6, p. 109)).

   In the MIMO case one may let $R = \rho R_0$, where the fixed matrix R_0 is selected according to the rules of § 3.2.3 (p. 106) and ρ is selected to achieve the desired bandwidth.

2. Next, the design parameters for the *observer part* are determined. They are chosen to achieve loop transfer recovery, as described in § 3.3.6 (p. 122).

   (a) To take advantage of loop transfer recovery we need to take $G = B$. LTR is only effective if the open-loop transfer function $P(s) = C(sI - A)^{-1}B$ has no right-half plane zeros, or only has right-half plane zeros whose magnitudes are sufficiently much greater than the desired bandwidth.

   (b) In the SISO case we let $V = 1$.

   In the MIMO case we may select V to be diagonal by the "dual" of the rules of § 3.2.3 (p. 106). This amounts to choosing each diagonal entry of V proportional to the inverse of the square root of the amplitude of the largest disturbance that may occur at the corresponding input channel.

(c) In the SISO case W is a scalar parameter that is chosen small enough to achieve loop transfer recovery. Asymptotically the finite observer poles are the zeros of $\det C(sI-A)^{-1}B$. These closed-loop poles correspond to canceling pole-zero pairs between the plant and the compensator. The far-away observer poles determine the bandwidth of the compensator and the high-frequency roll-off frequency for the complementary sensitivity. The magnitude of the dominant observer poles should be perhaps a decade larger than the magnitude of the dominant regulator poles.

In the MIMO case we let $W = \sigma W_0$. W_0 is chosen diagonally with each diagonal entry proportional to the inverse of the square root of the largest expected measurement error at the corresponding output channel. The scalar σ is chosen small enough to achieve LTR.

3.5.3 H_2 Interpretation

In this subsection we discuss the interpretation of the LQG problem as an H_2 optimization problem.

If we choose $G = B$ to take advantage of loop transfer recovery then the open-loop equations (3.110–3.111) may be rewritten as

$$z = R(s)(u+v), \tag{3.132}$$
$$y = P(s)(u+v)+w, \tag{3.133}$$

where

$$P(s) = C(sI-A)^{-1}B, \qquad R(s) = D(sI-A)^{-1}B. \tag{3.134}$$

The corresponding block diagram is represented in Fig. 3.12. If P is invertible then by block

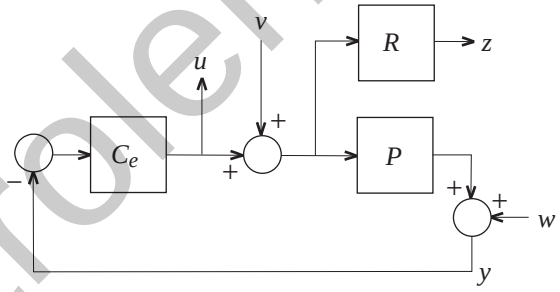


Figure 3.12: Block diagram for loop recovery design

diagram substitution Fig. 3.12 may be redrawn as in Fig. 3.13, where $W_0 = RP^{-1}$. In the case of non-inferential control $W_0 = I$.

We consider the frequency domain interpretation of the arrangement of Fig. 3.13. By setting up two appropriate signal balances it is easy to find that

$$z = W_0 SPv - W_0 Tw, \tag{3.135}$$
$$u = -T'v - Uw. \tag{3.136}$$

Here

$$S = (I+PC_e)^{-1}, \qquad T = (I+PC_e)^{-1}PC_e, \tag{3.137}$$
$$T' = (I+C_eP)^{-1}C_eP, \qquad U = C_e(I+PC_e)^{-1}. \tag{3.138}$$

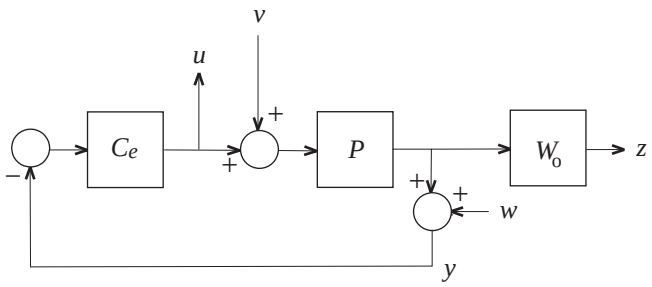


Figure 3.13: Equivalent block diagram

S is the sensitivity matrix, T the complementary sensitivity matrix, and U the input sensitivity matrix. T' is the complementary sensitivity function if the loop is broken at the plant input rather than at the plant output. In the SISO case $T' = T$.

Exercise 3.5.1 (Transfer matrices). Derive (3.135–3.138). □

From (3.135–3.136) we find for the performance index

$$\|H\|_2^2 = \mathrm{tr} \int_{-\infty}^{\infty} (W_0 S P P^{\sim} S^{\sim} W_0^{\sim} + W_0 T T^{\sim} W_0^{\sim} + T' T'^{\sim} + U U^{\sim})\, df. \tag{3.139}$$

The argument $\mathrm{j}2\pi f$ is omitted from each term in the integrand. Inspection reveals that the performance index involves a trade-off of the sensitivity S, the complementary sensitivities T and T', and the input sensitivity U. The importance given to each of the system functions depends on W_0 and P, which act as frequency dependent weighting functions.

The weighting functions in (3.139) arise from the LQG problem and are not very flexible. For more freedom we generalize the block diagram of Fig. 3.13 to that of Fig. 3.14. V_1 and V_2 are shaping filters and W_1 and W_2 weighting filters that may be used to modify the design. It is not difficult to establish that

$$z_1 = W_1 S P V_1 v - W_1 T V_2 w, \tag{3.140}$$
$$z_2 = -W_2 T' V_1 v - W_2 U V_2 w. \tag{3.141}$$

As a result, the performance index now takes the form

$$\|H\|_2^2 = \mathrm{tr} \int_{-\infty}^{\infty} (W_1 S P V_1 V_1^{\sim} P^{\sim} S^{\sim} W_1^{\sim} + W_1 T V_2 V_2^{\sim} T^{\sim} W_1^{\sim} + W_2 T' V_1 V_1'^{\sim} T'^{\sim} W_2^{\sim} + W_2 U V_2 V_2^{\sim} U^{\sim} W_2^{\sim})\, df. \tag{3.142}$$

In the next subsections some applications of this generalized problem are discussed.

3.5.4 Design for integral control

There are various ways to obtain integrating action in the LQG framework. We discuss a solution that follows logically from the frequency domain interpretation (3.142). For simplicity we only consider the SISO case. For the MIMO case the idea may be carried through similarly by introducing integrating action in each input channel as for the SISO case.

Integral control aims at suppressing constant disturbances, which requires making $S(0) = 0$. If the system has no natural integrating action then integrating action needs to be introduced in

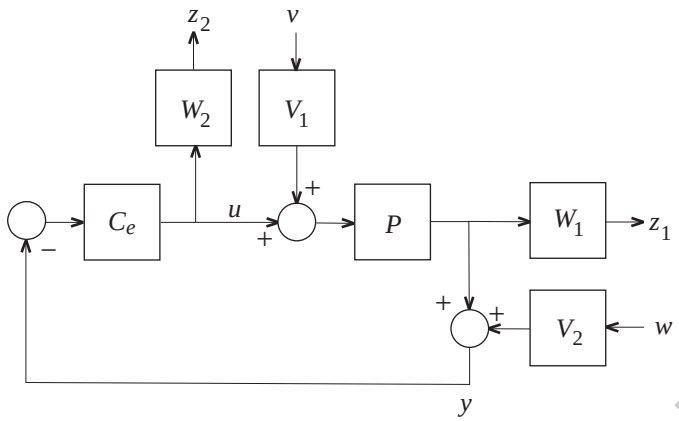


Figure 3.14: Generalized configuration

the compensator. Inspection of (3.142) shows that taking $V_1(0) = \infty$ forces $S(0)$ to be zero — otherwise the integral cannot be finite. Hence, we take V_1 as a rational function with a pole at 0. In particular, we let

$$V_1(s) = \frac{s+\alpha}{s}. \tag{3.143}$$

V_1 may be seen as a filter that shapes the frequency distribution of the disturbance. The positive constant α models the width of the band over which the low-frequency disturbance extends.

Further inspection of (3.142) reveals that the function V_1 also enters the third term of the integrand. In the SISO case the factor T' in this term reduces to T. If $S(0) = 0$ then by complementarity $T(0) = 1$. This means that this third term is infinite at frequency 0, *unless* W_2 has a factor s that cancels the corresponding factor s in the numerator of V_1. This has a clear interpretation: If the closed-loop system is to suppress constant disturbances then we need to allow constant inputs — hence we need $W_2(0) = 0$.

More in particular we could take

$$W_2(s) = \frac{s}{s+\alpha} W_{2o}(s), \tag{3.144}$$

where W_{2o} remains to be chosen but usually is taken constant. This choice of W_2 reduces the weight on the input over the frequency band where the disturbances are large. This allows the gain to be large in this frequency band.

A practical disadvantage of chooosing V_1 as in (3.143) is that it makes the open-loop system unstabilizable, because of the integrator outside the loop. This violates one of the assumptions of § 3.4.3 (p. 126) required for the solution of the H_2 problem. The difficulty may be circumvented by a suitable partitioning of the state space and the algebraic Riccati equations (Kwakernaak and Sivan 1972). We prefer to eliminate the problem by the block diagram substitutions (a) → (b) → (c) of Fig. 3.15. The end result is that an extra factor $\frac{s+\alpha}{s}$ is included in both the plant transfer function and the weighting function for the input. The extra factor in the weighting function on the input cancels against the factor $\frac{s}{s+\alpha}$ that we include according to (3.144). W_{2o} remains.

Additionally an extra factor $\frac{s}{s+\alpha}$ is included in the compensator. If the modified problem leads to an optimal compensator C_0 then the optimal compensator for the original problem is

$$C(s) = \frac{s+\alpha}{s} C_0(s). \tag{3.145}$$

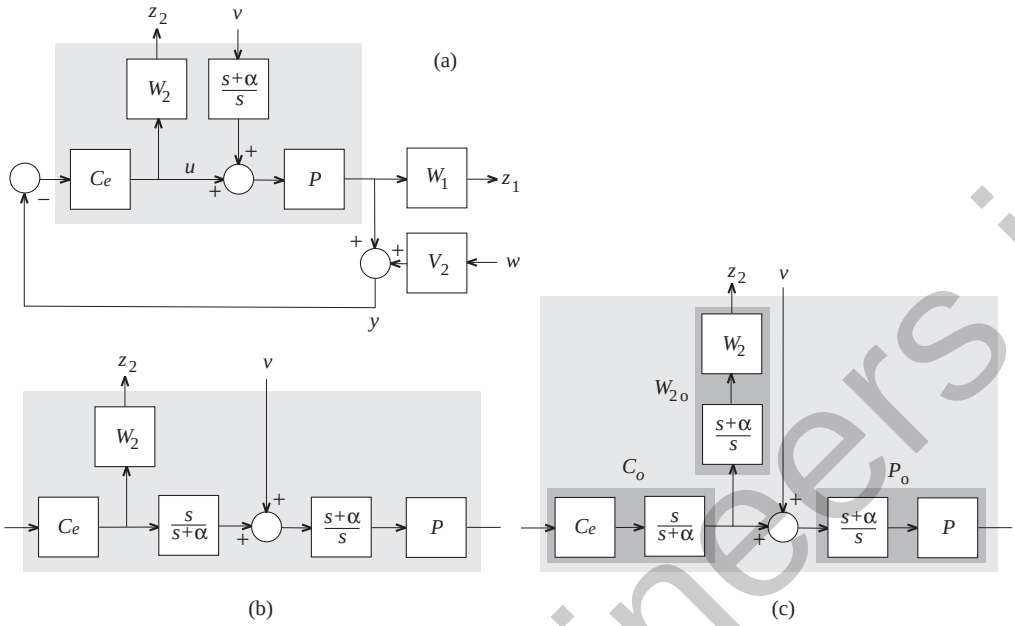


Figure 3.15: Block diagram substitutions for integral control (a)→(b)→(c)

This compensator explicitly includes integrating action.

Note that by the block diagram substitutions this method of obtaining integral control comes down to including an integrator in the plant. After doing the design for the modified plant the extra factor is moved over to the compensator. This way of ensuring integrating action is called the *integrator in the loop* method. We apply it in the example of § 3.6.3 (p. 135). The method is explained in greater generality in § 6.7 (p. 285) in the context of H_∞ optimization.

3.5.5 High-frequency roll-off

Solving the LQG problems leads to compensators with a strictly proper transfer matrix. This means that the high-frequency roll-off of the compensator and of the input sensitivity is 1 decade/decade (20 dB/decade). Correspondingly the high-frequency roll-off of the complementary sensitivity is at least 1 decade/decade.

Exercise 3.5.2 (High-frequency roll-off). Prove that LQG optimal compensators are strictly proper. Check for the SISO case what the resulting roll-off is for the input sensitivity function and the complementary sensitivity function, dependent on the high-frequency roll-off of the plant.

For some applications it may be desirable to have a steeper high-frequency roll-off. Inspection of (3.142) shows that extra roll-off may be imposed by letting the weighting function W_2 *increase* with frequency. Consider the SISO case and suppose that $V_2(s) = 1$. Let

$$W_2(s) = \rho(1 + rs), \tag{3.146}$$

with r a positive constant. Then by inspecting the fourth term in the integrand of (3.142) we conclude that the integral can only converge if at high frequencies the input sensitivity U, and, hence, also the compensator transfer function C_e, rolls off at at least 2 decades/decade.

The difficulty that there is no state space realization for the block W_2 with the transfer function (3.146) may be avoided by the block diagram substitution of Fig. 3.16. If the modified problem

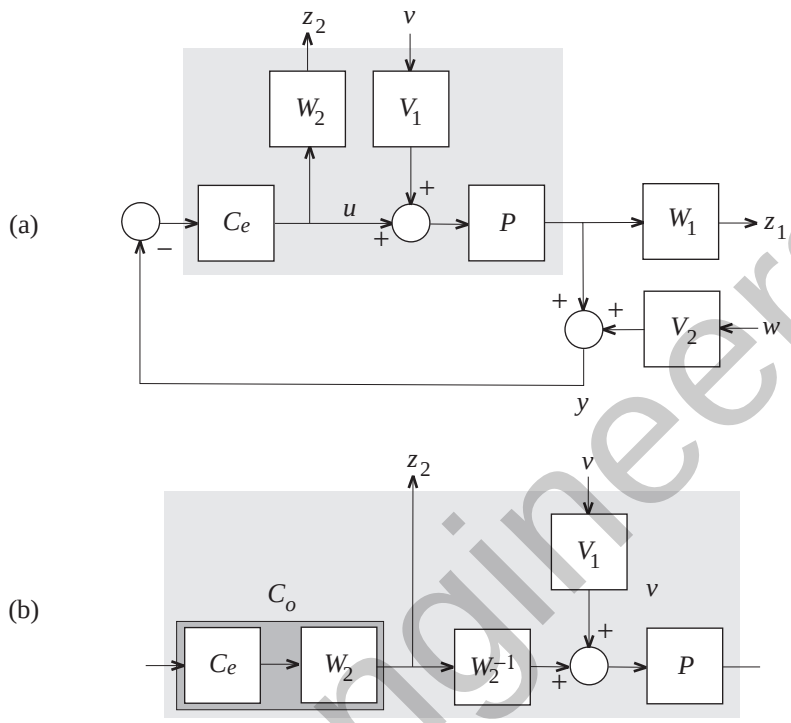


Figure 3.16: Block diagram substitution for high-frequency roll-off

is solved by the compensator C_0 then the optimal compensator for the original problem is

$$C_e(s) = \frac{C_0(s)}{W_2(s)} = \frac{C_0(s)}{\rho(1+rs)}. \tag{3.147}$$

The extra roll-off is apparent. Even more roll-off may be obtained by letting $W_2(s) = O(s^m)$ as $|s| \to \infty$, with $m \geq 2$.

For a more general exposition of the block diagram substitution method see § 6.7 (p. 285).

3.6 Examples and applications

3.6.1 Introduction

In this section we present two design applications of H_2 theory: A simple SISO system and a not very complicated MIMO system.

3.6.2 LQG design of a double integrator plant

We consider the double integrator plant

$$P(s) = \frac{1}{s^2}. \tag{3.148}$$

The design target is a closed-loop system with a bandwidth of 1 rad/s. Because the plant has a natural double integration there is no need to design for integral control and we expect excellent low-frequency characteristics. The plant has no right-half zeros that impair the achievable performance. Neither do the poles.

In state space form the system may be represented as

$$\dot{x} = \underbrace{\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}}_{A} x + \underbrace{\begin{bmatrix} 0 \\ 1 \end{bmatrix}}_{B} u, \qquad y = \underbrace{\begin{bmatrix} 1 & 0 \end{bmatrix}}_{C} x. \tag{3.149}$$

Completing the system equations we have

$$\dot{x} = Ax + Bu + Gv, \tag{3.150}$$

$$y = Cx + w, \tag{3.151}$$

$$z = Dx. \tag{3.152}$$

We choose the controlled variable z equal to the measured variable y. so that $D = C$. To profit from loop transfer recovery we let $G = B$. In the SISO case we may choose $Q = V = 1$ without loss of generality. Finally we write $R = \rho$ and $W = \sigma$, with the constants ρ and σ to be determined.

We first consider the regulator design. In the notation of § 3.2.6 (p. 109) we have $k = 1$, $\psi(s) = 1$ and $\chi_{\text{ol}}(s) = s^2$. It follows from (3.34) that the closed-loop characteristic polynomial χ_{cl} for state feedback satisfies

$$\chi_{\text{cl}}(-s)\chi_{\text{cl}}(s) = \chi_{\text{ol}}(-s)\chi_{\text{ol}}(s) + \frac{k^2}{\rho}\psi(-s)\psi(s) = s^4 + \frac{1}{\rho}. \tag{3.153}$$

The roots of the polynomial on the right-hand side are $\frac{1}{2}\sqrt{2}(\pm 1 \pm \mathrm{j})/\rho^{\frac{1}{4}}$. To determine the closed-loop poles we select those two roots that have negative real parts. They are given by

$$\frac{1}{2}\sqrt{2}(-1 \pm \mathrm{j})/\rho^{\frac{1}{4}}. \tag{3.154}$$

Figure 3.17 shows the loci of the closed-loop poles as ρ varies. As the magnitude of the closed-

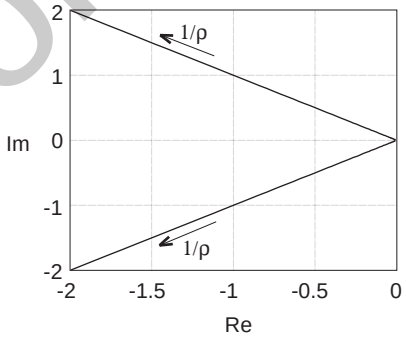


Figure 3.17: Loci of the closed-loop poles

loop pole pair is $1/\rho^{\frac{1}{4}}$ the desired bandwidth of 1 rad/s is achieved for $\rho = 1$.

We next turn to the observer design. By Exercise 3.3.2(c) (p. 119) the observer characteristic polynomial χ_f satisfies

$$\chi_f(-s)\chi_f(s) = \chi_{\text{ol}}(s)\chi_{\text{ol}}(-s)[1 + \frac{1}{\sigma}M(s)M(-s)] = s^4 + \frac{1}{\sigma^4}, \tag{3.155}$$

where $M(s) = C(sI - A)^{-1}G = 1/s^2$ and $\chi_{ol}(s) = s^2$. This expression is completely similar to that for the regulator characteristic polynomial, and we conclude that the observer characteristic values are

$$\frac{1}{2}\sqrt{2}(-1 \pm \mathrm{j})/\sigma^{\frac{1}{4}}. \tag{3.156}$$

By the rule of thumb of § 3.5.2 (p. 128) we choose the magnitude $1/\sigma^{\frac{1}{4}}$ of the observer pole pair 10 times greater than the bandwidth, that is, 10 rad/s. It follows that $\sigma = 0.0001$.

By numerical computation[8] it is found that the optimal compensator transfer function is

$$C_e(s) = \frac{155.6(s + 0.6428)}{s^2 + 15.56s + 121.0}. \tag{3.157}$$

This is a lead compensator.

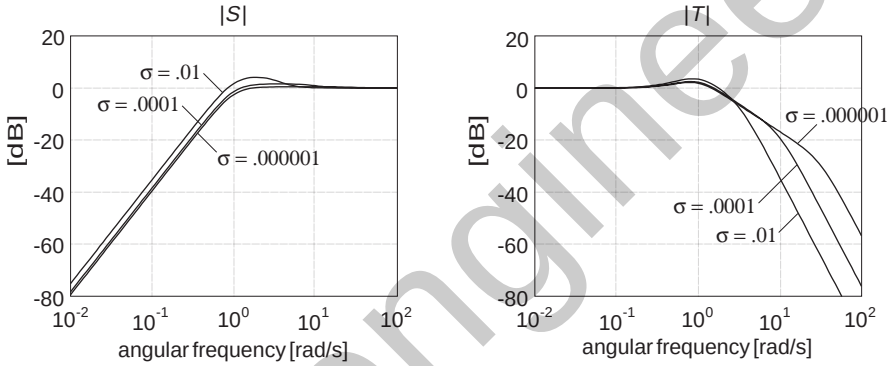


Figure 3.18: Sensitivity function and complementary sensitivity function for the H_2 design

Figure 3.18 shows the magnitude plots of the sensitivity and complementary sensitivity functions for $\sigma = .01$, $\sigma = .0001$ and $\sigma = .000001$. The smaller σ is the better loop transfer is recovered. Since the high-frequency roll-off of the complementary sensitivity of 40 dB/decade sets in at the angular frequency $1/\sigma^{\frac{1}{4}}$ it is advantageous not to choose σ too small. Taking σ large, though, results in extra peaking at crossover. The value $\sigma = .0001$ seems a reasonable compromise. Figure 3.19 gives the closed-loop step response. The value $\sigma = .000001$ gives the best response but that for $\sigma = .0001$ is very close.

3.6.3 A MIMO system

As a second example we consider the two-input two-output plant with transfer matrix (Kwakernaak 1986)

$$P(s) = \begin{bmatrix} \frac{1}{s^2} & \frac{1}{s+2} \\ 0 & \frac{1}{s+2} \end{bmatrix}. \tag{3.158}$$

[8] This may be done very conveniently with MATLAB using the Control Toolbox (Control Toolbox 1990).

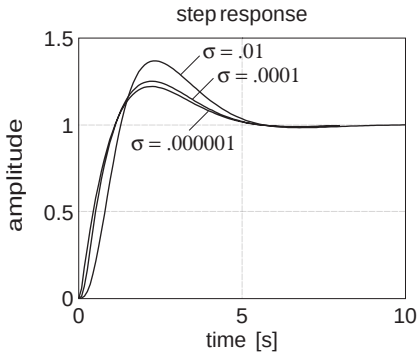


Figure 3.19: Closed-loop step response for the H_2 design

Figure 3.20 shows the block diagram. The plant is triangularly coupled. It is easy to see that it may be represented in state space form as

$$\dot{x} = \underbrace{\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -2 \end{bmatrix}}_{A} x + \underbrace{\begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}}_{B} u, \qquad y = \underbrace{\begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}}_{C} x. \tag{3.159}$$

The first two components of the state represent the block $1/s^2$ in Fig. 3.20, and the third the block $1/(s+2)$.

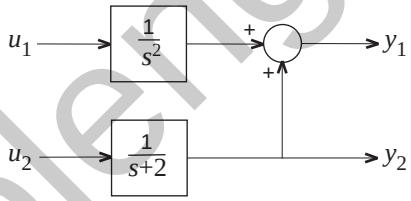


Figure 3.20: MIMO system

The plant has no right-half plane poles or zeros, so that there are no fundamental limitations to its performance.

Exercise 3.6.1 (Poles and zeros). What are the open-loop poles and zeros? *Hint:* For the definition of the zeros see § 3.2.6 (p. 109). □

We aim at a closed-loop bandwidth of 1 rad/s on both channels, with good low- and high-frequency characteristics.

We complete the system description to

$$\dot{x} = Ax + Bu + Gv, \tag{3.160}$$

$$y = Cx + w, \tag{3.161}$$

$$z = Dx. \tag{3.162}$$

To take advantage of loop transfer recovery we let $G = B$. As the controlled output z is available for feedback we have $D = C$. Assuming that the inputs and outputs are properly scaled we choose

$$Q = I, \quad R = \rho I, \quad V = I, \quad W = \sigma I, \tag{3.163}$$

with the positive scalars ρ and σ to be determined. First we consider the design of the regulator

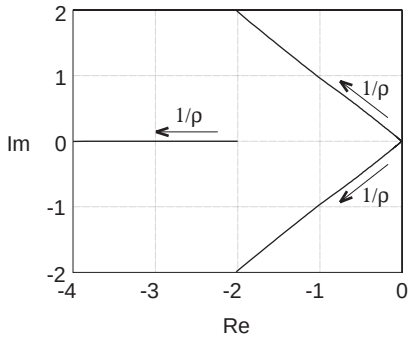


Figure 3.21: Loci of the regulator poles

part, which determines the dominant characteristics. By numerical solution of the appropriate algebraic Riccati equation for a range of values of ρ we obtain the loci of closed-loop poles of Fig. 3.21. As ρ decreases the closed-loop poles move away from the double open-loop pole at 0 and the open-loop pole at -2. For $\rho = 0.8$ the closed-loop poles are -2.5424 and $-.7162 \pm$ j .7034. The latter pole pair is dominant with magnitude 1.0038, which is the correct value for a closed-loop bandwidth of 1 rad/s.

Next we consider the loci of the optimal observer poles as a function of σ. Like in the double integrator example, they are identical to those of the regulator poles.

Exercise 3.6.2 (Identical loci). Check that this is a consequence of choosing $Q = V = I$, $R = \rho I$, $W = \sigma I$, $D = C$ and $G = B$. □

Again following the rule that the dominant observer poles have magnitude 10 times that of the dominant regulator poles we let $\sigma = 5 \times 10^{-5}$. This results in the optimal observer poles -200.01 and $-7.076 \pm$ j 7.067. The latter pole pair has magnitude 10. Using standard software the H_2 solution may now be found. Figure 3.22 shows the magnitudes of the four entries of the resulting 2×2 sensitivity matrix S.

The attenuation of disturbances that enter the system at the first output correponds to the entries S_{11} and S_{21} and is quite adequate, thanks to the double integrator in the corresponding input channel. The attenuation of disturbances that affect the second output (represented by S_{12} and S_{22}) is disappointing, however. The reason is that the low-frequency disturbances generated by the double integrator completely dominate the disturbances generated in in the other channel.

We improve the performance of the second channel by introducing integrating action. Application of the integrator-in-the-loop method of § 3.5.4 (p. 130) amounts to including an extra block

$$\frac{s+\alpha}{s} \tag{3.164}$$

in the second channel, as indicated in Fig. 3.23. After completion of the design the extra block is absorbed into the compensator. Representing the extra block by the state space realization

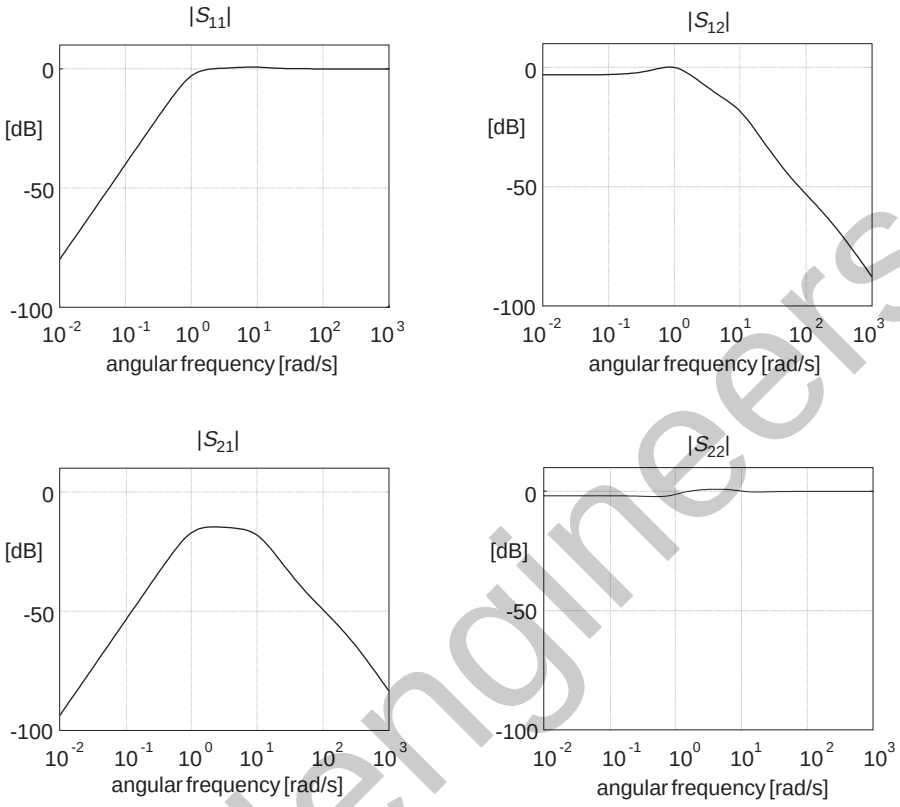


Figure 3.22: Magnitudes of the entries of the sensitivity matrix S

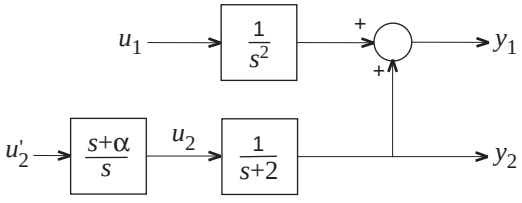


Figure 3.23: Expanded plant for integrating action in the second channel

$\dot{x}_4 = u_2'$, $u_2 = \alpha x_4 + u_2'$ we obtain the modified plant

$$\dot{x} = \underbrace{\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -2 & \alpha \\ 0 & 0 & 0 & 0 \end{bmatrix}}_{A} x + \underbrace{\begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}}_{B} u, \qquad y = \underbrace{\begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}}_{C} x. \tag{3.165}$$

The input u now has the components u_1 and u_2', and x has the components x_1, x_2, x_3, and x_4.

Again we let $D = C$, $G = B$, $Q = V = I$, $R = \rho I$ and $W = \sigma I$, with ρ and σ to be determined. We choose $\alpha = 1$ so that the low-frequency disturbance at the second input channel extends up to the desired bandwidth. Figure 3.24 shows the loci of the regulator poles with ρ as parameter.

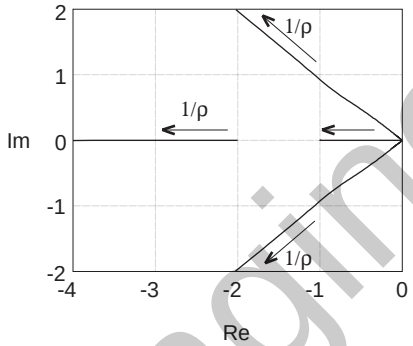


Figure 3.24: Loci of the regulator poles of the extended system

Three of the loci move out to ∞ from the open-loop poles 0, 0, and -2. The fourth moves out from the open-loop pole at 0 to the open-loop zero at $-\alpha = -1$. For $\rho = 0.5$ the regulator poles are -2.7200, $-0.8141 \pm \mathrm{j}\,0.7394$ and -0.6079. The latter pole turns out to be nondominant. The pole pair $-0.8141 \pm \mathrm{j}\,0.7394$, which has magnitude 1.0998, determines the bandwidth.

The loci of the optimal observer poles are again identical to those for the regulator poles. For $\sigma = 5 \times 10^{-5}$ the observer poles are -200.01, $-7.076 \pm \mathrm{j}\,7.067$ and -1. The latter pole is nondominant and the pole pair $-7.076 \pm \mathrm{j}\,7.067$ has magnitude 10.

Figure 3.25 shows the magnitudes of the four entries of the sensitivity and complementary sensitivity matrices S and T that follow for $\rho = .5$ and $\sigma = 5 \times 10^{-5}$. The results are now much more acceptable.

Note that the off-diagonal entries of S and T are small (though less so in the crossover region). This means that the feedback compensator to an extent achieves decoupling. This is a consequence of the high feedback gain at low frequencies. Infinite gain at all frequencies with unit feedback would make the closed-loop transfer matrix equal to the unit matrix, and, hence, completely decouple the system. The decoupling effect is also visible in Fig. 3.26, which shows the entries s_{ij}, i, $j = 1, 2$ of the closed-loop response to unit steps on the two inputs.

3.7 Appendix: Proofs

In this appendix we provide sketches of several of the proofs for this chapter.

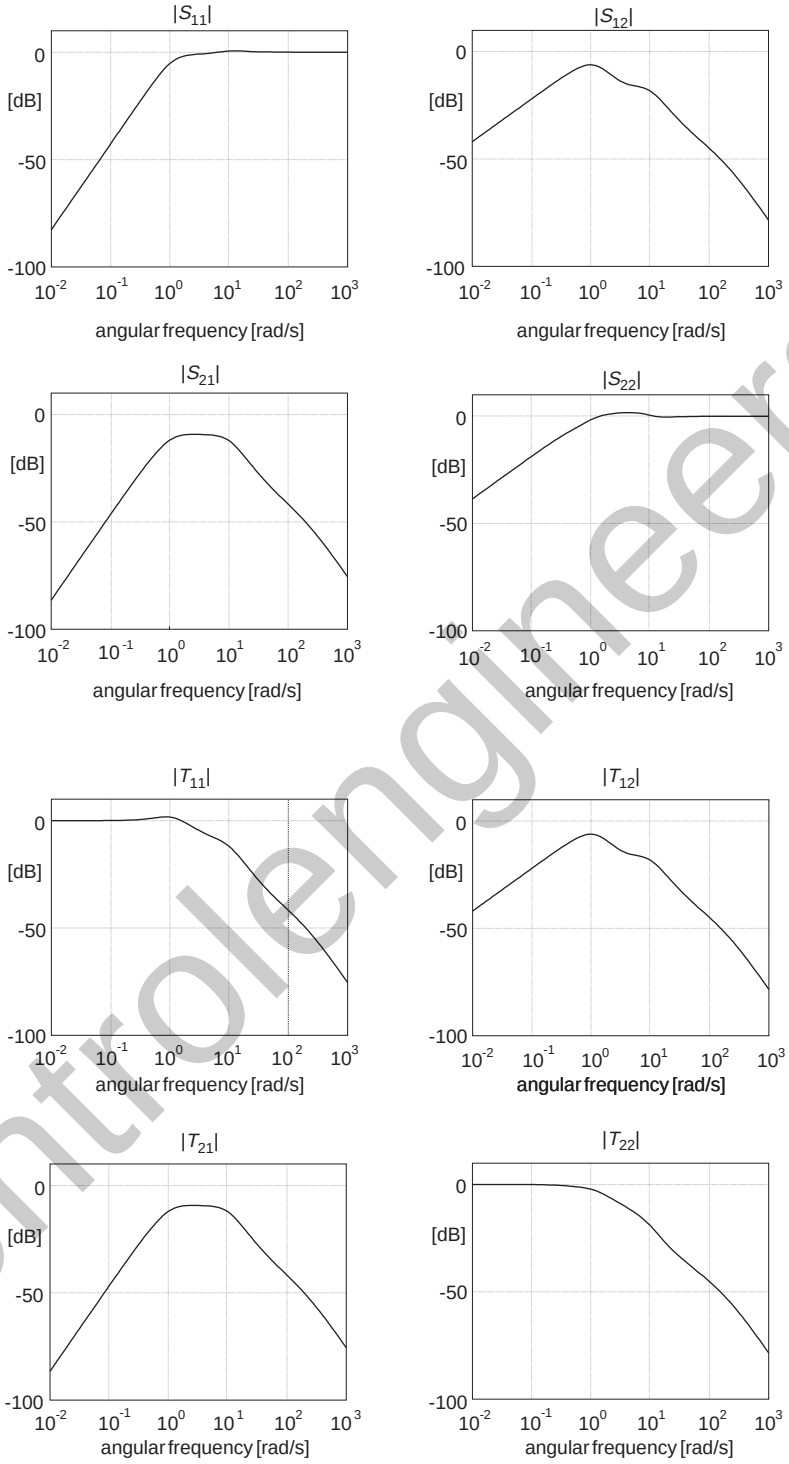


Figure 3.25: Magnitudes of the entries of the sensitivity matrix S and the complementary sensitivity matrix T

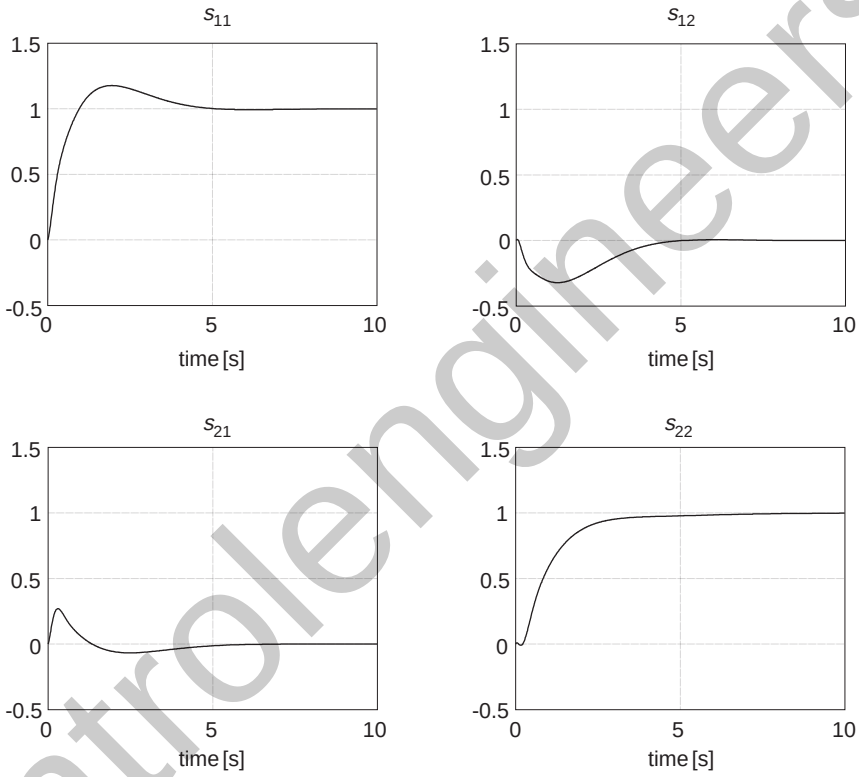


Figure 3.26: Entries of the closed-loop unit step response matrix

3.7.1 Outline of the solution of the regulator problem

We consider the problem of minimizing

$$\int_0^\infty [z^{\mathrm{T}}(t)Qz(t) + u^{\mathrm{T}}(t)Ru(t)]\,dt \tag{3.166}$$

for the stabilizable and detectable system

$$\dot{x}(t) = Ax(t) + Bu(t), \qquad z(t) = Dx(t). \tag{3.167}$$

Lyapunov equation. We first study how to compute the quantity

$$\int_0^{t_1} x^{\mathrm{T}}(s)Q(s)x(s)\,dt + x^{\mathrm{T}}(t_1)X_1x(t_1) \tag{3.168}$$

for solutions of the time-varying state differential equation $\dot{x}(t) = A(t)x(t)$. $Q(t)$ is a time-dependent nonnegative-definite symmetric matrix and X_1 is a nonnegative-definite symmetric matrix. Substitution of $x(s) = \Phi(s,0)x(0)$, with Φ the state transition matrix of the system, yields

$$\begin{aligned}\int_0^{t_1} x^{\mathrm{T}}(s)Q(s)x(s)\,ds + x^{\mathrm{T}}(t_1)X_1x(t_1)\\ = x^{\mathrm{T}}(0)\left(\int_0^{t_1} \Phi^{\mathrm{T}}(s,0)Q(s)\Phi(s,0)\,ds + \Phi^{\mathrm{T}}(t_1,0)X_1\Phi(t_1,0)\right)x(0).\end{aligned} \tag{3.169}$$

Define the time-dependent matrix

$$X(t) = \int_t^{t_1} \Phi^{\mathrm{T}}(s,t)Q(s)\Phi(s,t)\,ds + \Phi^{\mathrm{T}}(t_1,t)X_1\Phi(t_1,t). \tag{3.170}$$

Then

$$\int_0^{t_1} x^{\mathrm{T}}(s)Q(s)x(s)\,ds + x^{\mathrm{T}}(t_1)X_1x(t_1) = x^{\mathrm{T}}(0)X(0)x(0). \tag{3.171}$$

Differentiation of $X(t)$ with respect to t using $\frac{\partial}{\partial t}\Phi(s,t) = \Phi(s,t)A(t)$ shows that $X(t)$ satisfies the matrix differential equation and terminal condition

$$-\dot{X}(t) = A^{\mathrm{T}}(t)X(t) + X(t)A(t) + Q(t), \qquad X(t_1) = X_1. \tag{3.172}$$

This equation is a *Lyapunov matrix differential equation*. If A is constant and stable and R is constant then as $t_1 \to \infty$ the matrix $X(t)$ approaches a constant nonnegative-definite matrix $\bar{X}$ that is independent of X_1 and is the unique solution of the *algebraic Lyapunov equation*

$$0 = A^{\mathrm{T}}\bar{X} + \bar{X}A + Q. \tag{3.173}$$

Solution of the regulator problem. We consider how to determine the time-dependent gain $F(t)$ such that the state feedback $u(t) = -F(t)x(t)$, $0 \le t \le t_1$, minimizes the performance criterion

$$\begin{aligned}\int_0^{t_1} [z^{\mathrm{T}}(t)Qz(t) + u^{\mathrm{T}}(t)Ru(t)]\,dt + x^{\mathrm{T}}(t_1)X_1x(t_1)\\ = \int_0^{t_1} x^{\mathrm{T}}(t)\left(D^{\mathrm{T}}QD + F^{\mathrm{T}}(t)RF(t)\right)x(t)\,dt + x^{\mathrm{T}}(t_1)X_1x(t_1)\end{aligned} \tag{3.174}$$

for the system $\dot{x}(t) = Ax(t) + Bu(t) = [A - BF(t)]x(t)$. X_1 is a nonnegative-definite symmetric matrix. It follows that

$$\int_0^\infty [z^{\mathrm{T}}(t)Qz(t) + u^{\mathrm{T}}(t)Ru(t)]\,dt + x^{\mathrm{T}}(t_1)X_1x(t_1) = x^{\mathrm{T}}(0)X(0)x(0), \tag{3.175}$$

where $X(t)$ is the solution of

$$-\dot{X}(t) = [A - BF(t)]^{\mathrm{T}} X(t) + X(t)[A - BF(t)] + D^{\mathrm{T}} QD + F^{\mathrm{T}}(t) RF(t), \tag{3.176}$$

$X(t_1) = X_1$. Completion of the square on the right-hand side (with respect to $F(t)$ results in

$$\begin{aligned} -\dot{X}(t) \quad &= \quad [F(t) - R^{-1}B^{\mathrm{T}}X(t)]^{\mathrm{T}} R[F(t) - R^{-1}B^{\mathrm{T}}X(t)] \\ &\quad + A^{\mathrm{T}}X(t) + X(t)A + D^{\mathrm{T}}QD - X(t)BR^{-1}B^{\mathrm{T}}X(t), \qquad X(t_1) = X_1. \end{aligned} \tag{3.177}$$

Tracing the solution $X(t)$ backwards from t_1 to 0 we see that at each time t the least increase of $X(t)$ results if the gain $F(t)$ is selected as

$$F(t) = -R^{-1}B^{\mathrm{T}}X(t). \tag{3.178}$$

Correspondingly, (3.177) reduces to the *matrix differential Riccati equation*

$$-\dot{X}(t) = A^{\mathrm{T}}X(t) + X(t)A + D^{\mathrm{T}}QD - X(t)BR^{-1}B^{\mathrm{T}}X(t), \qquad X(t_1) = X_1. \tag{3.179}$$

If the system is stabilizable then the minimum of the performance criterion on the left-hand side of (3.174) is a nonincreasing function of the terminal time t_1. Because it is bounded from below (by 0) the criterion has a well-defined limit as $t_1 \to \infty$. Hence, also $X(t)$ has a well-defined limit as $t_1 \to \infty$. Because of the time independence of the system this limit $\bar{X}$ is independent of t. The limit is obviously nonnegative-definite, and satisfies the algebraic Riccati equation

$$0 = A^{\mathrm{T}}\bar{X} + \bar{X}A + D^{\mathrm{T}}QD - \bar{X}BR^{-1}B^{\mathrm{T}}\bar{X}. \tag{3.180}$$

Correspondingly the optimal feedback gain is time-invariant and equal to

$$F = -R^{-1}B^{\mathrm{T}}\bar{X}. \tag{3.181}$$

The left-hand side of (3.174) can only converge to a finite limit if $z(t) = Dx(t) \to 0$ as $t \to \infty$. By detectability this implies that $x(t) \to 0$, that is, the closed-loop system is stable. If the system is observable then $\bar{X}$ is positive-definite; otherwise there exist nonzero initial states such that $z(t) = Dx(t) = 0$ for $t \geq 0$.

3.7.2 Kalman-Jakubovič-Popov equality

The Kalman-Jakubovič-Popov equality is a generalization of the return difference equality that we use in § 3.2.5 (p. 108). The KJP equality establishes the connection between factorizations and algebraic Riccati equations.

Summary 3.7.1 (Kalman-Jakubovič-Popov equality). Consider the linear time-invariant system $\dot{x}(t) = Ax(t) + Bu(t)$, $y(t) = Cx(t) + Du(t)$, with transfer matrix $G(s) = C(sI - A)^{-1}B + D$, and let Q and R be given symmetric constant matrices. Suppose that the algebraic matrix Riccati equation

$$0 = A^{\mathrm{T}}X + XA + C^{\mathrm{T}}QC - (XB + C^{\mathrm{T}}QD)(D^{\mathrm{T}}QD + R)^{-1}(B^{\mathrm{T}}X + D^{\mathrm{T}}QC) \tag{3.182}$$

has a symmetric solution X. Then

$$R + G^{\sim}(s)QG(s) = J^{\sim}(s)LJ(s). \tag{3.183}$$

The constant symmetric matrix L and the rational matrix function J are given by

$$L = R + D^{\mathrm{T}}QD, \qquad J(s) = I + F(sI - A)^{-1}B, \tag{3.184}$$

with $F = L^{-1}(B^{\mathrm{T}}X + D^{\mathrm{T}}QC)$. The zeros of the numerator of $\det J$ are the eigenvalues of the matrix $A - BF$. □

We use the notation $G^{\sim}(s) = G^{\mathrm{T}}(-s)$.

The KJP equality arises in the study of the regulator problem for the system $\dot{x}(t) = Ax(t) + Bu(t)$, $y(t) = Cx(t) + Du(t)$, with the criterion

$$\int_0^\infty [y^{\mathrm{T}}(t)Qy(t) + u^{\mathrm{T}}(t)Ru(t)]\,dt. \tag{3.185}$$

The equation (3.182) is the algebraic Riccati equation associated with this problem, and $u(t) = -Fx(t)$ is the corresponding optimal state feedback law.

The KJP equality is best known for the case $D = 0$ (see for instance Kwakernaak and Sivan (1972)). It then reduces to the *return difference equality*

$$J^{\sim}(s)RJ(s) = R + G^{\sim}(s)QG(s). \tag{3.186}$$

Kalman-Jakubovič-Popov equality. The proof starts from the algebraic Riccati equation

$$0 = A^{\mathrm{T}}X + XA + C^{\mathrm{T}}QC - (XB + C^{\mathrm{T}}QD)L^{-1}(B^{\mathrm{T}}X + D^{\mathrm{T}}QC), \tag{3.187}$$

with $L = R + D^{\mathrm{T}}QD$. From the relation $F = L^{-1}(B^{\mathrm{T}}X + D^{\mathrm{T}}QC)$ we have $B^{\mathrm{T}}X + D^{\mathrm{T}}QC = LF$, so that the Riccati equation may be written as

$$0 = A^{\mathrm{T}}X + XA + C^{\mathrm{T}}QC - F^{\mathrm{T}}LF. \tag{3.188}$$

This in turn we rewrite as

$$0 = -(-sI - A^{\mathrm{T}})X - X(sI - A) + C^{\mathrm{T}}QC - F^{\mathrm{T}}LF. \tag{3.189}$$

Premultiplication by $B^{\mathrm{T}}(-sI - A^{\mathrm{T}})^{-1}$ and postmultiplication by $(sI - A)^{-1}B$ results in

$$\begin{aligned} 0 = &- B^{\mathrm{T}}X(sI - A)^{-1}B - B^{\mathrm{T}}(-sI - A^{\mathrm{T}})^{-1}XB \\ &+ B^{\mathrm{T}}(-sI - A^{\mathrm{T}})^{-1}(C^{\mathrm{T}}QC - F^{\mathrm{T}}LF)(sI - A)^{-1}B. \end{aligned} \tag{3.190}$$

Substituting $B^{\mathrm{T}}X = LF - D^{\mathrm{T}}QC$ we find

$$\begin{aligned} 0 = (D^{\mathrm{T}}QC - LF)(sI - A)^{-1}B + B^{\mathrm{T}}(-sI - A^{\mathrm{T}})^{-1}(C^{\mathrm{T}}QD - F^{\mathrm{T}}L) \\ + B^{\mathrm{T}}(-sI - A^{\mathrm{T}})^{-1}(C^{\mathrm{T}}QC - F^{\mathrm{T}}LF)(sI - A)^{-1}B. \end{aligned} \tag{3.191}$$

Expansion of this expression, substitution of $C(sI - A)^{-1}B = G(s) - D$ and $F(sI - A)^{-1}B = J(s) - I$ and simplification lead to the desired result

$$R + G^{\sim}(s)QG(s) = J^{\sim}(s)LJ(s). \tag{3.192}$$

■

3.7.3 Robustness under state feedback

We consider an open-loop stable system with loop gain matrix L that satisfies the return difference inequality

$$(I + L)^{\sim}R(I + L) \geq R \qquad \text{on the imaginary axis.} \tag{3.193}$$

We prove that if the loop gain is perturbed to WL, with W stable rational, then the closed-loop system remains stable as long as

$$RW + W^{\sim}R > R \qquad \text{on the imaginary axis.} \tag{3.194}$$

The proof follows Anderson and Moore (1990). First consider the case that $R = I$. It follows from the return difference inequality that $L^{\sim} + L + L^{\sim}L \geq 0$ on the imaginary axis, or

$$L^{-1} + (L^{-1})^{\sim} + I \geq 0 \qquad \text{on the imaginary axis.} \tag{3.195}$$

The perturbed system is stable if $I + WL$ has no zeros in the right-half complex plane. Equivalently, the perturbed system is stable if for $0 \leq \varepsilon \leq 1$ no zeros of

$$I + [(1-\varepsilon)I + \varepsilon W]L \tag{3.196}$$

cross the imaginary axis. Hence, the perturbed system is stable if only if

$$L^{-1} + (1-\varepsilon)I + \varepsilon W = M_\varepsilon \tag{3.197}$$

is nonsingular on the imaginary axis for all $0 \leq \varepsilon \leq 1$. Substitution of $L^{-1} = M_\varepsilon - (1-\varepsilon)I - \varepsilon W$ into (3.195) yields

$$M_\varepsilon + M_\varepsilon^\sim \geq 2(1-\varepsilon)I + \varepsilon(W + W^\sim) = (2-\varepsilon)I + \varepsilon(W + W^\sim - I). \tag{3.198}$$

Inspection shows that if

$$W + W^\sim > I \tag{3.199}$$

on the imaginary axis then $M_\varepsilon + M_\varepsilon^\sim > 0$ on the imaginary axis for all $0 \leq \varepsilon \leq 1$, which means that M_ε is nonsingular on the imaginary axis. Hence, if $W + W^\sim > I$ on the imaginary axis then the perturbed system is stable.

Exercise 3.7.2 (The case $R \neq I$). Show that the case that R is not necessarily the unit matrix may be reduced to the previous case by factoring R — for instance by Choleski factorization — as $R = R_0^{\mathrm{T}} R_0$. Work this out to prove that the closed-loop system remains stable under perturbation satisfying $RW + W^\sim R > R$ on the imaginary axis. □

3.7.4 Riccati equations and the Hamiltonian matrix

We consider the algebraic Riccati equation

$$A^{\mathrm{T}}X + XA + Q - (XB + S)R^{-1}(B^{\mathrm{T}}X + S^{\mathrm{T}}) = 0 \tag{3.200}$$

and the associated Hamiltonian matrix

$$H = \begin{bmatrix} A - BR^{-1}S^{\mathrm{T}} & -BR^{-1}B^{\mathrm{T}} \\ -Q + SR^{-1}S^{\mathrm{T}} & -(A - BR^{-1}S^{\mathrm{T}})^{\mathrm{T}} \end{bmatrix}. \tag{3.201}$$

Summary 3.7.3 (Riccati equation and the Hamiltonian matrix).

1. If λ is an eigenvalue of H then also $-\lambda$ is an eigenvalue of H.
2. Given a solution X of the Riccati equation, define $F = R^{-1}(B^{\mathrm{T}}X + S^{\mathrm{T}})$. Then

$$H \begin{bmatrix} I \\ X \end{bmatrix} = \begin{bmatrix} I \\ X \end{bmatrix} (A - BF). \tag{3.202}$$

3. If λ is an eigenvalue of $A - BF$ corresponding to the eigenvector x then λ is also an eigenvalue of H, corresponding to the eigenvector

$$\begin{bmatrix} I \\ X \end{bmatrix} x. \tag{3.203}$$

Hence, if the $n \times n$ matrix $A - BF$ has n eigenvalues with negative real parts — such as in the solution of the LQ problem of Summary 3.2.1 (p. 105) — then the eigenvalues of H consist of these n eigenvalues of $A - BF$ and their negatives.

4. Assume that H has no eigenvalues with zero real part. Then there is a similarity transformation U that brings H into upper triangular form T such that

$$H = UTU^{-1} = U\begin{bmatrix} T_{11} & T_{12} \\ 0 & T_{22} \end{bmatrix} U, \tag{3.204}$$

where the eigenvalues of the $n \times n$ diagonal block T_{11} all have negative real parts and those of T_{22} have positive real parts. Write

$$U = \begin{bmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{bmatrix}, \tag{3.205}$$

where each of the subblocks has dimensions $n \times n$. Then there is a solution X of the Riccati equation such that the eigenvalues of $A - BF$ all have strictly negative real part, if and only if U_{11} is nonsingular. In that case

$$X = U_{21}U_{11}^{-1} \tag{3.206}$$

is the unique solution of the Riccati equation such that the eigenvalues of $A - BF$ all have strictly negative real part. For the LQ problem of Summary 3.2.1 (p. 105) U_{11} is nonsingular.

□

For the transformation under 4 there are several possibilities. One is to bring H into Jordan normal form. For numerical computation it is to great advantage to use the Schur transformation.

Riccati equation and the Hamiltonian matrix (sketch).

1. The Hamiltonian matrix is of the form

$$H = \begin{bmatrix} A & Q \\ R & -A^{\mathrm{T}} \end{bmatrix}, \tag{3.207}$$

with all blocks square, and Q and R symmetric. We have for the characteristic polynomial of H

$$\begin{aligned} \det(\lambda I - H) &= \det\begin{bmatrix} \lambda I - A & -Q \\ -R & \lambda I + A^{\mathrm{T}} \end{bmatrix} \overset{(1)}{=} \det\begin{bmatrix} -\lambda I + A & Q \\ R & -\lambda I - A^{\mathrm{T}} \end{bmatrix} \\ &\overset{(2)}{=} (-1)^n \det\begin{bmatrix} R & -\lambda I - A^{\mathrm{T}} \\ -\lambda I + A & Q \end{bmatrix} \overset{(3)}{=} \det\begin{bmatrix} -\lambda I - A^{\mathrm{T}} & R \\ Q & -\lambda I + A \end{bmatrix} \\ &\overset{(4)}{=} \det\begin{bmatrix} -\lambda I + A & Q \\ R & -\lambda I - A^{\mathrm{T}} \end{bmatrix} \overset{(5)}{=} \det\begin{bmatrix} -\lambda I + A & -Q \\ -R & -\lambda I - A^{\mathrm{T}} \end{bmatrix} \end{aligned} \tag{3.208}$$

In step (1) we multiply the matrix by -1. In step (2) we interchange the first and second rows of blocks and in step (3) the firt and second columns of blocks. In step (4) we transpose the matrix. In step (5) we multiply the second row and the second column of blocks by -1.

Inspection shows that the characteristic polynomial does not change if λ is replaced with λ. Hence, if λ is an eigenvalue, so is $-\lambda$.

2. Using the Riccati equation we obtain from (3.201)

$$H\begin{bmatrix} I \\ X \end{bmatrix} = \begin{bmatrix} A - BF \\ -Q + SR^{-1}S^{\mathrm{T}} - (A - BR^{-1}S^{\mathrm{T}})^{\mathrm{T}}X \end{bmatrix} \tag{3.209}$$

$$= \begin{bmatrix} A - BF \\ XA - XBF \end{bmatrix} = \begin{bmatrix} I \\ X \end{bmatrix}(A - BF). \tag{3.210}$$

3. If $(A - BF)x = \lambda x$ then

$$H\begin{bmatrix} I \\ X \end{bmatrix} x = \begin{bmatrix} I \\ X \end{bmatrix}(A - BF)x = \lambda \begin{bmatrix} I \\ X \end{bmatrix} x. \tag{3.211}$$

4. From $H\,U = UT$ we obtain

$$H \begin{bmatrix} U_{11} \\ U_{21} \end{bmatrix} = \begin{bmatrix} U_{11} \\ U_{21} \end{bmatrix} T_{11}. \tag{3.212}$$

After multiplying on the right by U_{11} it follows that

$$H \begin{bmatrix} I \\ U_{21}U_{11}^{-1} \end{bmatrix} = \begin{bmatrix} I \\ U_{21}U_{11}^{-1} \end{bmatrix} U_{11}T_{11}U_{11}^{-1}. \tag{3.213}$$

We identify $X = U_{21}U_{11}^{-1}$ and $A - BF = U_{11}T_{11}U_{11}^{-1}$. For the LQ problem the nonsingularity of U_{11} follows by the existence of X such that $A - BF$ is stable. ■

3.7.5 The Kalman filter

In this subsection we outline the derivation of the Kalman filter.

Linear system driven by white noise Consider the stable linear system $\dot{x}(t) = Ax(t) + v(t)$, driven by white noise with intensity V, that is, $Ev(t)v^{\mathrm{T}}(s) = V\delta(t-s)$. The state of the system

$$x(t) = \int_{-\infty}^{t} \mathrm{e}^{A(t-s)}\, v(s)\, ds, \qquad t \in \mathbb{R}, \tag{3.214}$$

is a stationary stochastic process with covariance matrix

$$\begin{aligned} Y &= Ex(t)x^{\mathrm{T}}(t) = \int_{-\infty}^{t}\int_{-\infty}^{t} \mathrm{e}^{A(t-s_1)}\big(Ev(s_1)v^{\mathrm{T}}(s_2)\big)\,\mathrm{e}^{A^{\mathrm{T}}(t-s_2)}\, ds_1 ds_2 \\ &= \int_{-\infty}^{t} \mathrm{e}^{A(t-s)}\, V\,\mathrm{e}^{A^{\mathrm{T}}(t-s)}\, ds = \int_{0}^{\infty} \mathrm{e}^{A\tau}\, V\,\mathrm{e}^{A^{\mathrm{T}}\tau}\, d\tau. \end{aligned} \tag{3.215}$$

It follows that

$$\begin{aligned} AY + YA^{\mathrm{T}} &= \int_0^\infty \left(A\,\mathrm{e}^{A\tau}\, V\,\mathrm{e}^{A^{\mathrm{T}}\tau} + \mathrm{e}^{A\tau}\, V\,\mathrm{e}^{A^{\mathrm{T}}\tau}\, A^{\mathrm{T}}\right) d\tau \\ &= \int_0^\infty \frac{d}{d\tau}\left(\mathrm{e}^{A\tau}\, V\,\mathrm{e}^{A^{\mathrm{T}}\tau}\right) d\tau = \mathrm{e}^{A\tau}\, V\,\mathrm{e}^{A^{\mathrm{T}}\tau}\Big|_0^\infty = -V. \end{aligned} \tag{3.216}$$

Hence, the covariance matrix Y is the unique solution of the Lyapunov equation

$$AY + YA^{\mathrm{T}} + V = 0. \tag{3.217}$$

Observer error covariance matrix. We consider the system $\dot{x}(t) = Ax(t) + Bu(t) + v(t)$, $y(t) = Cx(t) + w(t)$, where v is white noise with intensity V and w white noise with intensity W. The estimation error $e(t) = \hat{x}(t) - x(t)$ of the observer

$$\dot{\hat{x}}(t) = A\hat{x}(t) + Bu(t) + K[y(t) - C\hat{x}(t)] \tag{3.218}$$

satisfies the differential equation

$$\dot{e}(t) = (A - KC)e(t) - Gv(t) + Kw(t). \tag{3.219}$$

The noise process $-Gv(t) + Kw(t)$ is white noise with intensity $G^{\mathrm{T}}VG + K^{\mathrm{T}}WK$. Hence, if the error system is stable then the error covariance matrix $Y = Ee(t)e^{\mathrm{T}}(t)$ is the unique solution of the Lyapunov equation

$$(A - KC)Y + Y(A - KC)^{\mathrm{T}} + G^{\mathrm{T}}VG + K^{\mathrm{T}}WK = 0. \tag{3.220}$$

The Kalman filter. We discuss how to choose the observer gain K to minimize the error covariance matrix Y. To this end we complete the square (in K) and rewrite the Lyapunov equation (3.220) as

$$(K - YC^{\mathrm{T}}W^{-1})W(K - YC^{\mathrm{T}}W^{-1})^{\mathrm{T}} + AY + YA^{\mathrm{T}} + GVG^{\mathrm{T}} - YC^{\mathrm{T}}W^{-1}CY = 0. \tag{3.221}$$

Suppose that there exists a gain K that stabilizes the error system and minimizes the error variance matrix Y. Then changing the gain to $K + \varepsilon\tilde{K}$, with ε a small scalar and $\tilde{K}$ an arbitrary matrix of the same dimensions as K, should only affect Y quadratically in ε. Inspection of (3.221) shows that this implies

$$K = YC^{\mathrm{T}}W^{-1}. \tag{3.222}$$

With this gain the observer reduces to the Kalman filter. The minimal error variance matrix Y satisfies the Riccati equation

$$AY + YA^{\mathrm{T}} + GVG^{\mathrm{T}} - YC^{\mathrm{T}}W^{-1}CY = 0. \tag{3.223}$$

3.7.6 Minimization of the steady-state mean square error under state feedback

We consider the problem of choosing the gain F of the state feedback law $u(t) = -Fx(t)$ to minimize the steady state mean square error

$$E\left(z^{\mathrm{T}}(t)Qz(t) + u^{\mathrm{T}}(t)Ru(t)\right) \tag{3.224}$$

for the system $\dot{x}(t) = Ax(t) + Bu(t) + v(t)$. The white noise v has intensity V.

If the feedback law stabilizes the system $\dot{x}(t) = (A - BF)x(t) + v(t)$ then the steady-state covariance matrix Y of the state is given by

$$Y = Ex(t)x^{\mathrm{T}}(t) = \int_0^\infty \mathrm{e}^{(A-BF)s}\, V\, \mathrm{e}^{(A-BF)^{\mathrm{T}}s}\, ds. \tag{3.225}$$

Hence we have for the steady-state mean square error

$$\begin{aligned} E\left(z^{\mathrm{T}}(t)Qz(t) + u^{\mathrm{T}}(t)Ru(t)\right) &= E\left(x^{\mathrm{T}}(t)D^{\mathrm{T}}QDx(t) + x^{\mathrm{T}}(t)F^{\mathrm{T}}RFx(t)\right) \\ &= E\ \mathrm{tr}\left(x(t)x^{\mathrm{T}}(t)D^{\mathrm{T}}QD + x(t)x^{\mathrm{T}}(t)F^{\mathrm{T}}RF\right) = \mathrm{tr}\ Y\left(D^{\mathrm{T}}QD + F^{\mathrm{T}}RF\right). \end{aligned} \tag{3.226}$$

We rewrite this in the form

$$\begin{aligned} E\left(z^{\mathrm{T}}(t)Qz(t) + u^{\mathrm{T}}(t)Ru(t)\right) &= \mathrm{tr}\ Y\left(D^{\mathrm{T}}QD + F^{\mathrm{T}}RF\right) \\ &= \mathrm{tr} \int_0^\infty \mathrm{e}^{(A-BF)s}\, V\, \mathrm{e}^{(A-BF)^{\mathrm{T}}s}\, ds \left(D^{\mathrm{T}}QD + F^{\mathrm{T}}RF\right) \\ &= \mathrm{tr}\ V \underbrace{\int_0^\infty \mathrm{e}^{(A-BF)^{\mathrm{T}}s}\left(D^{\mathrm{T}}QD + F^{\mathrm{T}}RF\right)\mathrm{e}^{(A-BF)s}\, ds}_{X} = \mathrm{tr}\ VX. \end{aligned} \tag{3.227}$$

X is the solution of the Lyapunov equation

$$(A - BF)^{\mathrm{T}}X + X(A - BF) + D^{\mathrm{T}}QD + F^{\mathrm{T}}RF = 0. \tag{3.228}$$

X and, hence, $\mathrm{tr}\ VX$, is minimized by choosing $F = R^{-1}B^{\mathrm{T}}X$, with X the solution of the Riccati equation $A^{\mathrm{T}}X + XA + D^{\mathrm{T}}QD - XBR^{-1}B^{\mathrm{T}}X = 0$.

3.7.7 Loop transfer recovery

We study an LQG optimal system with measurement noise intensity $W = \sigma W_0$ as $\sigma \downarrow 0$ under the assumptions that $G = B$ and that the plant transfer matrix $G(s) = C(sI - A)^{-1}B$ is square with stable inverse $G^{-1}(s)$.

Under the assumption $G = B$ we have in the absence of any measurement noise w

$$y = C(sI - A)^{-1}B(u + v) = G(s)(u + v). \tag{3.229}$$

Because by assumption G^{-1} is stable the input noise v may be recovered with arbitrary precision by approximating the inverse relation

$$v = G^{-1}(s)y - u \tag{3.230}$$

with sufficient accuracy. From the noise v and the known input u the state x may in turn be reconstructed with arbitrary precision. Hence, we expect that as σ decreases to 0 the covariance Y_σ of the estimation error decreases to the zero matrix.

Under the assumption $G = B$ the Riccati equation for the optimal observer is

$$AY_\sigma + Y_\sigma A^{\mathrm{T}} + BVB^{\mathrm{T}} - Y_\sigma C^{\mathrm{T}} W^{-1} C Y_\sigma = 0. \tag{3.231}$$

We rewrite this as

$$AY_\sigma + Y_\sigma A^{\mathrm{T}} + BVB^{\mathrm{T}} - \sigma K_\sigma W_0 K_\sigma^{\mathrm{T}} = 0, \tag{3.232}$$

with $K_\sigma = Y_\sigma C^{\mathrm{T}} W^{-1}$ the gain. Inspection shows that if $Y_\sigma \downarrow 0$ then necessarily $K_\sigma \to \infty$. In fact we may write

$$K_\sigma \approx \frac{1}{\sqrt{\sigma}} BU_\sigma \quad \text{as } \sigma \downarrow 0, \tag{3.233}$$

where U_σ is a square nonsingular matrix (which may depend on σ) such that $U_\sigma W_0 U_\sigma^{\mathrm{T}} = V$.

We study the asymptotic behavior of the loop gain

$$\begin{aligned}
L_\sigma(s) &= F(sI - A + BF + K_\sigma C)^{-1} K_\sigma C(sI - A)^{-1}B \\
&\approx F(sI - A + BF + \frac{1}{\sqrt{\sigma}} BU_\sigma C)^{-1} \frac{1}{\sqrt{\sigma}} BU_\sigma C(sI - A)^{-1}B \\
&\approx F(sI - A + \frac{1}{\sqrt{\sigma}} BU_\sigma C)^{-1} \frac{1}{\sqrt{\sigma}} BU_\sigma C(sI - A)^{-1}B \\
&= F(sI - A)^{-1}\left(I + \frac{1}{\sqrt{\sigma}} BU_\sigma C(sI - A)^{-1}\right)^{-1} \frac{1}{\sqrt{\sigma}} BU_\sigma C(sI - A)^{-1}B \\
&\overset{(1)}{=} F(sI - A)^{-1}B\frac{1}{\sqrt{\sigma}} U_\sigma \left(I + \frac{1}{\sqrt{\sigma}} C(sI - A)^{-1}BU_\sigma\right)^{-1} C(sI - A)^{-1}B \\
&= F(sI - A)^{-1}BU_\sigma \left(I\sqrt{\sigma} + C(sI - A)^{-1}BU_\sigma\right)^{-1} C(sI - A)^{-1}B.
\end{aligned} \tag{3.234}$$

In step (1) we use the well-known matrix identity $(I + AB)^{-1}A = A(I + BA)^{-1}$. Inspection of the final equality shows that

$$L_\sigma(s) \xrightarrow{\sigma \downarrow 0} F(sI - A)^{-1}B, \tag{3.235}$$

which is the loop gain under full state feedback.

3.7.8 Solution of the H_2 optimization problem

We solve the standard H_2 optimization problem of § 3.4.3 (p. 126) as if it is an LQG problem, that is, we set out to minimize the steady-state value of

$$Ez^{\mathrm{T}}(t)z(t) \tag{3.236}$$

under the assumption that w is a white noise input with intensity matrix I.

State feedback. We first consider the solution with state feedback. For this it is enough to study the equations

$$\dot{x}(t) = Ax(t) + B_1 w(t) + B_2 u(t), \tag{3.237}$$
$$z(t) = C_1 x(t) + D_{11} w(t) + D_{12} u(t). \tag{3.238}$$

If $D_{11} \neq 0$ then the output z has a white noise component that may well make the mean square output (3.236) infinite. We therefore assume that $D_{11} = 0$. Under this assumption we have

$$z(t) = C_1 x(t) + D_{12} u(t) = \begin{bmatrix} I & D_{12} \end{bmatrix} \begin{bmatrix} C_1 x(t) \\ u(t) \end{bmatrix} = \begin{bmatrix} I & D_{12} \end{bmatrix} \begin{bmatrix} z_0(t) \\ u(t) \end{bmatrix}, \tag{3.239}$$

where $z_0(t) = C_1 x(t)$. As a result,

$$\begin{aligned} Ez^{\mathrm{T}}(t)z(t) &= E \begin{bmatrix} z_0^{\mathrm{T}}(t) & u^{\mathrm{T}}(t) \end{bmatrix} \begin{bmatrix} I \\ D_{12}^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} I & D_{12} \end{bmatrix} \begin{bmatrix} z_0 \\ u(t) \end{bmatrix} \\ &= E \begin{bmatrix} z_0^{\mathrm{T}}(t) & u^{\mathrm{T}}(t) \end{bmatrix} \begin{bmatrix} I & D_{12} \\ D_{12}^{\mathrm{T}} & D_{12}^{\mathrm{T}} D_{12} \end{bmatrix} \begin{bmatrix} z_0 \\ u(t) \end{bmatrix}. \end{aligned} \tag{3.240}$$

This defines a linear regulator problem with a cross term in the output and input. It has a solution if the system $\dot{x}(t) = Ax(t) + B_2 u(t)$, $z_0(t) = C_1 x(t)$ is stabilizable and detectable, and the weighting matrix

$$\begin{bmatrix} I & D_{12} \\ D_{12}^{\mathrm{T}} & D_{12}^{\mathrm{T}} D_{12} \end{bmatrix} \tag{3.241}$$

is positive-definite. A necessary and sufficient condition for the latter is that $D_{12}^{\mathrm{T}} D_{12}$ be nonsingular. The solution to the regulator problem is a state feedback law of the form

$$u(t) = -Fx(t). \tag{3.242}$$

The gain matrix F may easily be found from the results of § 3.2.8 (p. 113) and is given in Summary 3.4.3 (p. 118).

Output feedback. If the state is not available for feedback then it needs to be estimated with a Kalman filter. To this end we consider the equations

$$\dot{x}(t) = Ax(t) + B_1 w(t) + B_2 u(t), \tag{3.243}$$
$$y(t) = C_2 x(t) + D_{21} w(t) + D_{22} u(t). \tag{3.244}$$

The second equation may be put into the standard form for the Kalman filter if we consider $y(t) - D_{22}u(t)$ as the observed variable rather than $y(t)$. If we denote the observation noise as $v(t) = D_{21} w(t)$ then

$$\dot{x}(t) = Ax(t) + B_1 w(t) + B_2 u(t), \tag{3.245}$$
$$y(t) - D_{22} u(t) = C_2 x(t) + v(t) \tag{3.246}$$

defines a stochastic system with cross correlated noise terms. We have

$$E \begin{bmatrix} w(t) \\ v(t) \end{bmatrix} \begin{bmatrix} w^{\mathrm{T}}(s) & v^{\mathrm{T}}(s) \end{bmatrix} = E \begin{bmatrix} I \\ D_{21} \end{bmatrix} w(t) w^{\mathrm{T}}(s) \begin{bmatrix} I & D_{21}^{\mathrm{T}} \end{bmatrix} \tag{3.247}$$

$$= \begin{bmatrix} I & D_{21}^{\mathrm{T}} \\ D_{21} & D_{21} D_{21}^{\mathrm{T}} \end{bmatrix} \delta(t-s). \tag{3.248}$$

Suppose that the system $\dot{x}(t) = Ax(t) + B_1 w(t)$, $y(t) = C_2 x(t)$ is stabilizable and detectable, and the intensity matrix

$$\begin{bmatrix} I & D_{21}^{\mathrm{T}} \\ D_{21} & D_{21} D_{21}^{\mathrm{T}} \end{bmatrix} \tag{3.249}$$

is positive-definite. A necessary and sufficient condition for the latter is that $D_{21}D_{21}^{\mathrm{T}}$ be nonsingular. Then there exists a well-defined Kalman filter of the form

$$\dot{\hat{x}}(t) = A\hat{x}(t) + B_2u(t) + K[y(t) - C_2\hat{x}(t) - D_{22}u(t)]. \tag{3.250}$$

The gain matrix K may be solved from the formulas of § 3.3.4 (p. 120). Once the Kalman filter (3.250) is in place the optimal input for the output feedback problem is obtained as

$$u(t) = -F\hat{x}(t). \tag{3.251}$$

F is the same state feedback gain as in (3.242).

A

Matrices

This appendix lists several matrix definitions, formulae and results that are used in the lecture notes.

In what follows capital letters denote matrices, lower case letters denote column or row vectors or scalars. The element in the ith row and jth column of a matrix A is denoted by A_{ij}. Whenever sums $A+B$ and products AB etcetera are used then it is assumed that the dimensions of the matrices are compatible.

A.1 Basic matrix results

Eigenvalues and eigenvectors

A column vector $v \in \mathbb{C}^n$ is an *eigenvector* of a square matrix $A \in \mathbb{C}^{n\times n}$ if $v \neq 0$ and $Av = \lambda v$ for some $\lambda \in \mathbb{C}$. In that case λ is referred to as an *eigenvalue* of A. Often $\lambda_i(A)$ is used to denote the ith eigenvalue of A (which assumes an ordering of the eigenvalues, an ordering that should be clear from the context in which it is used). The eigenvalues are the zeros of the *characteristic polynomial*

$$\chi_A(\lambda) = \det(\lambda I_n - A), \qquad (\lambda \in \mathbb{C},)$$

where I_n denotes the $n \times n$ *identity* matrix or *unit* matrix.

An *eigenvalue decomposition* of a square matrix A is a decomposition of A of the form

$$A = VDV^{-1}, \qquad \text{where } V \text{ and } D \text{ are square and } D \text{ is diagonal.}$$

In this case the diagonal entries of D are the eigenvalues of A and the columns of V are the corresponding eigenvectors. Not every square matrix A has an eigenvalue decomposition.

The eigenvalues of a square A and of TAT^{-1} are the same for any nonsingular T. In particular $\chi_A = \chi_{TAT^{-1}}$.

Rank, trace, determinant, singular and nonsingular matrices

The *trace*, $\mathrm{tr}(A)$ of a square matrix $A \in \mathbb{C}^{n\times n}$ is defined as $\mathrm{tr}(A) = \sum_{i=1}^n A_{ii}$. It may be shown that

$$\mathrm{tr}(A) = \sum_{i=1}^{n} \lambda_i(A).$$

The *rank* of a (possibly nonsquare) matrix A is the maximal number of linearly independent rows (or, equivalently, columns) in A. It also equals the rank of the square matrix $A^{\mathrm{H}}A$ which in turn equals the number of nonzero eigenvalues of $A^{\mathrm{H}}A$.

The *determinant* of a square matrix $A \in \mathbb{C}^{n\times n}$ is usually defined (but not calculated) recursively by

$$\det(A) = \begin{cases} \sum_{j=1}^{n} (-1)^{j+1} A_{1j} \det(A_{1j}^{\text{minor}}) & \text{if } n > 1 \\ A & \text{if } n = 1 \end{cases}.$$

Here A_{ij}^{minor} is the $(n-1)\times(n-1)$-matrix obtained from A by removing its ith row and jth column. The determinant of a matrix equals the product of its eigenvalues, $\det(A) = \prod_{i=1}^{n} \lambda_i(A)$.

A square matrix is *singular* if $\det(A) = 0$ and is *regular* or *nonsingular* if $\det(A) \neq 0$. For square matrices A and B of the same dimension we have

$$\det(AB) = \det(A)\det(B).$$

Symmetric, Hermitian and positive definite matrices, the transpose and unitary matrices

A matrix $A \in \mathbb{R}^{n\times n}$ is *(real) symmetric* if $A^{\mathrm{T}} = A$. Here A^{T} is the *transpose* of A is defined elementwise as $(A^{\mathrm{T}})_{ij} = A_{ji}$, $(i, j = 1, \ldots, n)$.

A matrix $A \in \mathbb{C}^{n\times n}$ is *Hermitian* if $A^{\mathrm{H}} = A$. Here A^{H} is the *complex conjugate transpose* of A defined as $(A^{\mathrm{H}})_{ij} = \overline{A}_{ji}$ $(i, j = 1, \ldots, n)$. Overbars $\overline{x + \mathrm{j}y}$ of a complex number $x + \mathrm{j}y$ denote the complex conjugate: $\overline{x + \mathrm{j}y} = x - \mathrm{j}y$.

Every real-symmetric and Hermitian matrix A has an eigenvalue decomposition $A = VDV^{-1}$ and they have the special property that the matrix V may be chosen *unitary* which is that the columns of V have unit length and are mutually orthogonal: $V^{\mathrm{H}}V = I$.

A symmetric or Hermitian matrix A is said to be *nonnegative definite* or *positive semi-definite* if $x^{\mathrm{H}}Ax \geq 0$ for all column vectors x. We denote this by

$$A \geq 0.$$

A symmetric or Hermitian matrix A is said to be *positive definite* if $x^{\mathrm{H}}Ax > 0$ for all *nonzero* column vectors x. We denote this by

$$A > 0.$$

For Hermitian matrices A and B the inequality $A \geq B$ is defined to mean that $A - B \geq 0$.

Lemma A.1.1. *Let $A \in \mathbb{C}^{n\times n}$ be a Hermitian matrix. Then*

1. *All eigenvalues of A are real valued,*
2. $A \geq 0 \iff \lambda_i(A) \geq 0 \quad (\forall\, i = 1, \ldots, n)$,
3. $A > 0 \iff \lambda_i(A) > 0 \quad (\forall\, i = 1, \ldots, n)$,
4. *If T is nonsingular then $A \geq 0$ if and only $T^H AT \geq 0$.*

□

A.2 Three matrix lemmas

Lemma A.2.1. *Suppose A and B^H are matrices of the same dimension $n \times m$. Then for any $\lambda \in \mathbb{C}$ there holds*

$$\det(\lambda I_n - AB) = \lambda^{n-m} \det(\lambda I_m - BA). \tag{A.1}$$

Proof. One the one hand we have

$$\begin{bmatrix} \lambda I_m & B \\ A & I_n \end{bmatrix} \begin{bmatrix} I_m & -\frac{1}{\lambda} B \\ 0 & I_n \end{bmatrix} = \begin{bmatrix} \lambda I_m & 0 \\ A & I_n - \frac{1}{\lambda} AB \end{bmatrix}$$

and on the other hand

$$\begin{bmatrix} \lambda I_m & B \\ A & I_n \end{bmatrix} \begin{bmatrix} I_m & 0 \\ -A & I_n \end{bmatrix} = \begin{bmatrix} \lambda I_m - BA & B \\ 0 & I_n \end{bmatrix}.$$

Taking determinants of both of these equations shows that

$$\lambda^m \det(I_n - \frac{1}{\lambda} AB) = \det \begin{bmatrix} \lambda I_m & B \\ A & I_n \end{bmatrix} = \det(\lambda I_m - BA).$$

■

So the *nonzero* eigenvalues of AB and BA are the same. This gives the two very useful identities:

1. $\det(I_n - AB) = \det(I_m - BA)$,
2. $\operatorname{tr}(AB) = \sum_i \lambda_i(AB) = \sum_i \lambda_i(BA) = \operatorname{tr}(BA)$.

Lemma A.2.2 (Sherman-Morrison-Woodburry & rank-one update).

$$(A + UV^H)^{-1} = A^{-1} - A^{-1}U(I + V^H A^{-1} V^H) A^{-1}$$

This formula is used mostly if $U = u$ and $V = v$ are column vectors. Then $UV^H = uv^H$ has rank one, and it shows that a rank-one update of A corresponds to a rank-one update of its inverse,

$$(A + uv^H)^{-1} = A^{-1} - \underbrace{\frac{1}{1 + v^H A^{-1} u} (A^{-1}u)(v^H A^{-1})}_{\text{rank-one}}.$$

□

Lemma A.2.3 (Schur complement). *Suppose a Hermitian matrix A is partitioned as*

$$A = \begin{bmatrix} P & Q \\ Q^H & R \end{bmatrix}$$

with P and R square. Then

$$A > 0 \iff P \text{ is invertible, } P > 0 \text{ and } R - Q^H P^{-1} Q > 0.$$

The matrix $R - Q^H P^{-1} Q$ is referred to as the Schur complement *of P (in A).* □

References

Abramowitz, M. and I. A. Stegun (1965). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover Publications, Inc., New York.

Anderson, B. D. O. and E. I. Jury (1976). A note on the Youla-Bongiorno-Lu condition. *Automatica 12*, 387–388.

Anderson, B. D. O. and J. B. Moore (1971). *Linear Optimal Control*. Prentice-Hall, Englewood Cliffs, NJ.

Anderson, B. D. O. and J. B. Moore (1990). *Optimal Control: Linear Quadratic Methods*. Prentice Hall, Englewood Cliffs, NJ.

Aplevich, J. D. (1991). *Implicit Linear Systems*, Volume 152 of *Lecture Notes in Control and Information Sciences*. Springer-Verlag, Berlin. ISBN 0-387-53537-3.

Bagchi, A. (1993). *Optimal control of stochastic systems*. International Series in Systems and Control Engineering. Prentice Hall International, Hempstead, UK.

Balas, G. J., J. C. Doyle, K. Glover, A. Packard, and R. Smith (1991). *User's Guide, μ-Analysis and Synthesis Toolbox*. The MathWorks, Natick, Mass.

Barmish, B. R. and H. I. Kang (1993). A survey of extreme point results for robustness of control systems. *Automatica 29*, 13–35.

Bartlett, A. C., C. V. Hollot, and H. Lin (1988). Root locations of an entire polytope of polynomials: It suffices to check the edges. *Math. Control Signals Systems 1*, 61–71.

Basile, G. and G. Marro (1992). *Controlled and Conditioned Invariance in Linear System Theory*. Prentice Hall, Inc., Englewood Cliffs, NJ.

Berman, G. and R. G. Stanton (1963). The asymptotes of the root locus. *SIAM Review 5*, 209–218.

Białas, S. (1985). A necessary and sufficient condition for the stability of convex combinations of polynomials or matrices. *Bulletin Polish Acad. of Sciences 33*, 473–480.

Black, H. S. (1934). Stabilized feedback amplifiers. *Bell System Technical Journal 13*, 1–18.

Blondel, V. (1994). *Simultaneous Stabilization of Linear Systems*, Volume 191 of *Lecture Notes in Control and Information Sciences*. Springer-Verlag, Berlin etc.

Bode, H. W. (1940). Relations between attenuation and phase in feedback amplifier design. *Bell System Technical Journal 19*, 421–454.

Bode, H. W. (1945). *Network Analysis and Feedback Amplifier Design*. Van Nostrand, New York.

Boyd, S. P. and C. H. Barratt (1991). *Linear Controller Design: Limits of Performance*. Prentice Hall, Englewood Cliffs, NJ.

Bristol, E. H. (1966). On a new measure of interaction for multivariable process control. *IEEE Transactions on Automatic Control 11*, 133–134.

Brockett, R. W. and M. D. Mesarovic (1965). The reproducibility of multivariable systems. *Journal of Mathematical Analysis and Applications 11*, 548–563.

Campo, P. J. and M. Morari (1994). Achievable closed-loop properties of systems under decentralized control: conditions involving the steady-state gain. *IEEE Transactions on Automatic Control 39*, 932–943.

Chen, C.-T. (1970). *Introduction to Linear Systems Theory*. Holt, Rinehart and Winston, New York.

Chiang, R. Y. and M. G. Safonov (1988). *User's Guide, Robust-Control Toolbox*. The MathWorks, South Natick, Mass.

Chiang, R. Y. and M. G. Safonov (1992). *User's Guide, Robust Control Toolbox*. The MathWorks, Natick, Mass., USA.

Choi, S.-G. and M. A. Johnson (1996). The root loci of H_∞-optimal control: A polynomial approach. *Optimal Control Applications and Methods 17*, 79–105.

Control Toolbox (1990). *User's Guide*. The MathWorks, Natick, Mass., USA.

Dahleh, M. A. and Y. Ohda (1988). A necessary and sufficient condition for robust BIBO stability. *Systems & Control Letters 11*, 271–275.

Davison, E. J. and S. H. Wang (1974). Properties and calculation of transmission zeros of linear multivariable systems. *Automatica 10*, 643–658.

Davison, E. J. and S. H. Wang (1978). An algorithm for the calculation of transmission zeros of a the system (C,A,B,D) using high gain output feedback. *IEEE Transactions on Automatic Control 23*, 738–741.

D'Azzo, J. J. and C. H. Houpis (1988). *Linear Control System Analysis and Design: Conventional and Modern*. McGraw-Hill Book Co., New York, NY. Third Edition.

Desoer, C. A. and J. D. Schulman (1974). Zeros and poles of matrix transfer functions and their dynamical interpretations. *IEEE Transactions on Circuits and Systems 21*, 3–8.

Desoer, C. A. and M. Vidyasagar (1975). *Feedback Systems: Input-Output Properties*. Academic Press, New York.

Desoer, C. A. and Y. T. Wang (1980). Linear time-invariant robust servomechanism problem: a self-contained exposition. In C. T. Leondes (Ed.), *Control and Dynamic Systems*, Volume 16, pp. 81–129. Academic Press, New York, NY.

Dorf, R. C. (1992). *Modern Control Systems*. Addison-Wesley Publishing Co., Reading, MA. Sixth Edition.

Doyle, J. C. (1979). Robustness of multiloop linear feedback systems. In *Proc. 17th IEEE Conf. Decision & Control*, pp. 12–18.

Doyle, J. C. (1982). Analysis of feedback systems with structured uncertainties. *IEE Proc., Part D 129*, 242–250.

Doyle, J. C. (1984). Lecture Notes, ONR/Honeywell Workshop on Advances in Multivariable Control, Minneapolis, Minn.

Doyle, J. C., B. A. Francis, and A. R. Tannenbaum (1992). *Feedback Control Theory*. Macmillan, New York.

Doyle, J. C., K. Glover, P. P. Khargonekar, and B. A. Francis (1989). State-space solutions to standard H_2 and H_∞ control problems. *IEEE Trans. Aut. Control 34*, 831–847.

Doyle, J. C. and G. Stein (1981). Multivariable feedback design: concepts for a classical/modern synthesis. *IEEE Transactions on Automatic Control 26*, 4–16.

Emami-Naeini, A. and P. VanDooren (1982). Computation of zeros of linear multivariable systems. *Automatica 18*, 415–430.

Engell, S. (1988). *Optimale lineare Regelung: Grenzen der erreichbaren Regelgüte in linearen zeitinvarianten Regelkreisen*, Volume 18 of *Fachberichte Messen–Steuern–Regeln*. Springer-Verlag, Berlin, etc.

Evans, W. R. (1948). Graphical analysis of control systems. *Trans. Amer. Institute of Electrical Engineers 67*, 547–551.

Evans, W. R. (1950). Control system synthesis by root locus method. *Trans. Amer. Institute of Electrical Engineers 69*, 66–69.

Evans, W. R. (1954). *Control System Dynamics*. McGraw-Hill Book Co., New York, NY.

Föllinger, O. (1958). Über die Bestimmung der Wurzelortskurve. *Regelungstechnik 6*, 442–446.

Francis, B. A. and W. M. Wonham (1975). The role of transmission zeros in linear multivariable regulators. *International Journal of Control 22*, 657–681.

Franklin, G. F., J. D. Powell, and A. Emami-Naeini (1986). *Feedback Control of Dynamic Systems*. Addison-Wesley Publ. Co., Reading, MA.

Franklin, G. F., J. D. Powell, and A. Emami-Naeini (1991). *Feedback Control of Dynamic Systems*. Addison-Wesley Publ. Co., Reading, MA. Second Edition.

Freudenberg, J. S. and D. P. Looze (1985). Right half plane poles and zeros and design tradeoffs in feedback systems. *IEEE Trans. Automatic Control 30*, 555–565.

Freudenberg, J. S. and D. P. Looze (1988). *Frequency Domain Properties of Scalar and Multivariable Feedback Systems*, Volume 104 of *Lecture Notes in Control and Information Sciences*. Springer-Verlag, Berlin, etc.

Gantmacher, F. R. (1959). *The Theory of Matrices. Volume 2*. Chelsea Publishing Company, New York, N.Y., USA.

Gantmacher, F. R. (1964). *The Theory of Matrices*. Chelsea Pub. Comp., New York, NY. Reprinted.

Glover, K. and J. C. Doyle (1988). State-space formulae for all stabilizing controllers that satisfy an H_∞-norm bound and relations to risk sensitivity. *Systems & Control Letters 11*, 167–172.

Glover, K. and J. C. Doyle (1989). A state space approach to H_∞ optimal control. In H. Nijmeijer and J. M. Schumacher (Eds.), *Three Decades of Mathematical System Theory*, Volume 135 of *Lecture Notes in Control and Information Sciences*. Springer-Verlag, Heidelberg, etc.

Glover, K., D. J. N. Limebeer, J. C. Doyle, E. M. Kasenally, and M. G. Safonov (1991). A characterization of all solutions to the four block general distance problem. *SIAM J. Control and Optimization 29*, 283–324.

Gohberg, I., P. Lancaster, and L. Rodman (1982). *Matrix polynomials*. Academic Press, New York, NY.

Golub, G. M. and C. Van Loan (1983). *Matrix Computations*. The Johns Hopkins University Press, Baltimore, Maryland.

Graham, D. and R. C. Lathrop (1953). The synthesis of 'optimum' transient response: Criteria and standard forms. *Trans. AIEE Part II (Applications and Industry) 72*, 273–288.

Green, M. and D. J. N. Limebeer (1995). *Linear Robust Control*. Prentice Hall, Englewood Cliffs.

Grimble, M. J. (1994). Two and a half degrees of freedom LQG controller and application to wind turbines. *IEEE Trans. Aut. Control 39*(1), 122–127.

Grosdidier, P. and M. Morari (1987). A computer aided methodology for the design of decentralized controllers. *Computers and Chemical Engineering 11*, 423–433.

Grosdidier, P., M. Morari, and B. R. Holt (1985). Closed-loop properties from steady-state gain information. *Industrial and Engineering Chemistry. Fundamentals 24*, 221–235.

Hagander, P. and B. Bernhardsson (1992). A simple test example for H_∞ optimal control. In *Preprints, 1992 American Control Conference, Chicago, Illinois*.

Henrici, P. (1974). *Applied and Computational Complex Analysis, Vol. 1*. Wiley-Interscience, New York.

Hinrichsen, D. and A. J. Pritchard (1986). Stability radii of linear systems. *Systems & Control Letters 7*, 1–10.

Horowitz, I. (1982, November). Quantitative feedback theory. *IEE Proc. Pt. D 129*, 215–226.

Horowitz, I. and A. Gera (1979). Blending of uncertain nonminimum-phase plants for elimination or reduction of nonminimum-phase property. *International Journal of Systems Science 10*, 1007–1024.

Horowitz, I. and M. Sidi (1972). Synthesis of feedback systems with large plant ignorance for prescribed time domain tolerances. *Int. J. Control 126*, 287–309.

Horowitz, I. M. (1963). *Synthesis of of Feedback Systems*. Academic Press, New York.

Hovd, M. and S. Skogestad (1992). Simple frequency-dependent tools for control system analysis, structure selection and design. *Automatica 28*, 989–996.

Hovd, M. and S. Skogestad (1994). Pairing criteria for decentralized control of unstable plants. *Industrial and Engineering Chemistry Research 33*, 2134–2139.

Hsu, C. H. and C. T. Chen (1968). A proof of the stability of multivariable feedback systems. *Proceedings of the IEEE 56*, 2061–2062.

James, H. M., N. B. Nichols, and R. S. Philips (1947). *Theory of Servomechanisms*. McGraw-Hill, New York.

Kailath, T. (1980a). *Linear Systems*. Prentice Hall, Englewood Cliffs, N. J.

Kailath, T. (1980b). *Linear Systems*. Prentice Hall, Inc., Englewood Cliffs, NJ.

Kalman, R. E. and R. S. Bucy (1961). New results in linear filtering and prediction theory. *J. Basic Engineering, Trans. ASME Series D 83*, 95–108.

Karcanias, N. and C. Giannakopoulos (1989). Necessary and sufficient conditions for zero assignment by constant squaring down. *Linear Algebra and its Applications 122/123/124*, 415–446.

Kharitonov, V. L. (1978a). Asymptotic stability of an equilibrium position of a family of systems of linear differential equations. *Differential'nye Uraveniya 14*, 1483–1485.

Kharitonov, V. L. (1978b). On a generalization of a stability criterion. *Akademii Nauk Kahskoi SSR, Fiziko-matematicheskaia 1*, 53–57.

Korn, U. and H. H. Wilfert (1982). *Mehrgrössenregelungen. Moderne Entwurfsprinzipien im Zeit- und Frequenzbereich*. Verlag Technik, Berlin.

Krall, A. M. (1961). An extension and proof of the root-locus method. *Journal Soc. Industr. Applied Mathematatics 9*, 644–653.

Krall, A. M. (1963). A closed expression for the root locus method. *Journal Soc. Industr. Applied Mathematics 11*, 700–704.

Krall, A. M. (1970). The root-locus method: A survey. *SIAM Review 12*, 64–72.

Krall, A. M. and R. Fornaro (1967). An algorithm for generating root locus diagrams. *Communications of the ACM 10*, 186–188.

Krause, J. M. (1992). Comments on Grimble's comments on Stein's comments on rolloff of H^∞ optimal controllers. *IEEE Trans. Auto. Control 37*, 702.

Kwakernaak, H. (1976). Asymptotic root loci of optimal linear regulators. *IEEE Trans. Aut. Control 21*(3), 378–382.

Kwakernaak, H. (1983). Robustness optimization of linear feedback systems. In *Preprints, 22nd IEEE Conference of Decision and Control, San Antonio, Texas, USA.*

Kwakernaak, H. (1985). Minimax frequency domain performance and robustness optimization of linear feedback systems. *IEEE Trans. Auto. Control 30*, 994–1004.

Kwakernaak, H. (1986). A polynomial approach to minimax frequency domain optimization of multivariable systems. *Int. J. Control 44*, 117–156.

Kwakernaak, H. (1991). The polynomial approach to H_∞-optimal regulation. In E. Mosca and L. Pandolfi (Eds.), *H_∞-Control Theory*, Volume 1496 of *Lecture Notes in Mathematics*, pp. 141–221. Springer-Verlag, Heidelberg, etc.

Kwakernaak, H. (1993). Robust control and H^∞-optimization. *Automatica 29*, 255–273.

Kwakernaak, H. (1995). Symmetries in control system design. In A. Isidori (Ed.), *Trends in Control — A European Perspective*, pp. 17–51. Springer, Heidelberg, etc.

Kwakernaak, H. (1996). Frequency domain solution of the standard H_∞ problem. In M. J. Grimble and V. Kučera (Eds.), *Polynomial Methods for Control Systems Design*. Springer-Verlag.

Kwakernaak, H. and R. Sivan (1972). *Linear Optimal Control Systems*. Wiley, New York.

Kwakernaak, H. and R. Sivan (1991). *Modern Signals and Systems*. Prentice Hall, Englewood Cliffs, NJ.

Kwakernaak, H. and J. H. Westdijk (1985). Regulability of a multiple inverted pendulum system. *Control – Theory and Advanced Technology 1*, 1–9.

Landau, I. D., F. Rolland, C. Cyrot, and A. Voda (1993). Régulation numérique robuste: Le placement de pôles avec calibrage de la fonction de sensibilité. In A. Oustaloup (Ed.), *Summer School on Automatic Control of Grenoble*. Hermes. To be published.

Laub, A. J. and B. C. Moore (1978). Calculation of transmission zeros using QZ techniques. *Automatica 14*, 557–566.

Le, D. K., O. D. I. Nwokah, and A. E. Frazho (1991). Multivariable decentralized integral controllability. *International Journal of Control 54*, 481–496.

Le, V. X. and M. G. Safonov (1992). Rational matrix GCD's and the design of squaring-down compensators- a state-space theory. *IEEE Transactions on Automatic Control 37*, 384–392.

Lin, J.-L., I. Postlethwaite, and D.-W. Gu (1993). μ–K Iteration: A new algorithm for μ-synthesis. *Automatica 29*, 219–224.

Luenberger, D. G. (1969). *Optimization by Vector Space Methods*. John Wiley, New York.

Lunze, J. (1985). Determination of robust multivariable I-controllers by means of experiments and simulation. *Systems Analysis Modelling Simulation 2*, 227–249.

Lunze, J. (1988). *Robust Multivariable Feedback Control.* Prentice Hall International Ltd, London, UK.

Lunze, J. (1989). *Robust Multivariable Feedback Control.* International Series in Systems and Control Engineering. Prentice Hall International, Hempstead, UK.

MacDuffee, C. C. (1956). *The Theory of Matrices*. Chelsea Publishing Company, New York, N.Y., USA.

MacFarlane, A. G. J. and N. Karcanias (1976). Poles and zeros of linear multivariable systems: a survey of the algebraic, geometric and complex-variable theory. *International Journal of Control 24*, 33–74.

Maciejowski, J. M. (1989). *Multivariable Feedback Design*. Addison-Wesley Publishing Company, Wokingham, UK.

McAvoy, T. J. (1983). *Interaction Analysis. Principles and Applications*. Instrument Soc. Amer., Research Triangle Park, NC.

McFarlane, D. C. and K. Glover (1990). *Robust Controller Design Using Normalized Coprime Factor Plant Descriptions*, Volume 146. Springer-Verlag, Berlin, etc.

McMillan, B. (1952). Introduction to formal realizability theory - I, II. *Bell System Technical Journal 31*, 217–279, 541–600.

Middleton, R. H. (1991). Trade-offs in linear control system design. *Automatica 27*, 281–292.

Minnichelli, R. J., J. J. Anagnost, and C. A. Desoer (1989). An elementary proof of Kharitonov's stability theorem with extensions. *IEEE Trans. Aut. Control 34*, 995–1001.

Morari, M. (1985). Robust stability of systems with integral control. *IEEE Transactions on Automatic Control 30*, 574–577.

Morari, M. and E. Zafiriou (1989). *Robust Process Control*. Prentice Hall International Ltd, London, UK.

Morse, A. S. (1973). Structural invariants of linear multivariable systems. *SIAM Journal on Control 11*, 446–465.

Nett, C. N. and V. Manousiouthakis (1987). Euclidean condition and block relative gain: connections, conjectures, and clarifications. *IEEE Transactions on Automatic Control 32*, 405–407.

Noble, B. (1969). *Applied Linear Algebra*. Prentice Hall, Englewood Cliffs, NJ.

Nwokah, O. D. I., A. E. Frazho, and D. K. Le (1993). A note on decentralized integral controllability. *International Journal of Control 57*, 485–494.

Nyquist, H. (1932). Regeneration theory. *Bell System Technical Journal 11*, 126–147.

Ogata, K. (1970). *Modern Control Engineering*. Prentice Hall, Englewood Cliffs, NJ.

Packard, A. and J. C. Doyle (1993). The complex structured singular value. *Automatica 29*, 71–109.

Postlethwaite, I., J. M. Edmunds, and A. G. J. MacFarlane (1981). Principal gains and principal phases in the analysis of linear multivariable feedback systems. *IEEE Transactions on Automatic Control 26*, 32–46.

Postlethwaite, I. and A. G. J. MacFarlane (1979a). *A Complex Variable Approach to the Analysis of Linear Multivariable Feedback Systems*, Volume 12 of *Lecture Notes in Control and Information Sciences*. Springer-Verlag, Berlin, etc.

Postlethwaite, I. and A. G. J. MacFarlane (1979b). *A Complex Variable Approach to the Analysis of Linear Multivariable Feedback Systems.* Springer Verlag, Berlin.

Postlethwaite, I., M. C. Tsai, and D.-W. Gu (1990). Weighting function selection in H^∞ design. In *Proceedings, 11th IFAC World Congress, Tallinn*. Pergamon Press, Oxford, UK.

Qiu, L., B. Bernhardsson, A. Rantzer, E. J. Davison, P. M. Young, and J. C. Doyle (1995). A formula for computation of the real stability radius. *Automatica 31*, 879–890.

Raisch, J. (1993). *Mehrgrössenregelung im Frequenzbereich*. R.Oldenbourg Verlag, München, BRD.

Rosenbrock, H. H. (1970). *State-space and Multivariable Theory*. Nelson, London.

Rosenbrock, H. H. (1973). The zeros of a system. *International Journal of Control 18*, 297–299.

Rosenbrock, H. H. (1974a). *Computer-aided Control System Design*. Academic Press, London.

Rosenbrock, H. H. (1974b). Correction on "The zeros of a system". *International Journal of Control 20*, 525–527.

Saberi, A., B. M. Chen, and P. Sannuti (1993). *Loop transfer recovery: Analysis and design.* Prentice Hall, Englewood Cliffs, N. J.

Saberi, A., P. Sannuti, and B. M. Chen (1995). *H_2 optimal control*. Prentice Hall, Englewood Cliffs, N. J. ISBN 0-13-489782-X.

Safonov, M. G. and M. Athans (1981). A multiloop generalization of the circle criterion for stability margin analysis. *IEEE Trams. Automatic Control 26*, 415–422.

Sain, M. K. and C. B. Schrader (1990). The role of zeros in the performance of multiinput,multioutput feedback systems. *IEEE Transactions on Education 33*, 244–257.

Savant, C. J. (1958). *Basic Feedback Control System Design*. McGraw-Hill Book Co., New York, NY.

Schrader, C. B. and M. K. Sain (1989). Research on system zeros: A survey. *International Journal of Control 50*, 1407–1433.

Sebakhy, O. A., M. ElSingaby, and I. F. ElArabawy (1986). Zero placement and squaring problem: a state space approach. *International Journal of Systems Science 17*, 1741–1750.

Shaked, U. (1976). Design techniques for high feedback gain stability. *International Journal of Control 24*, 137–144.

Silverman, L. M. (1969). Inversion of multivariable linear systems. *IEEE Transactions on Automatic Control 14*, 270–276.

Silverman, L. M. and H. J. Payne (1971). Input-output structure of linear systems with application to the decoupling problem. *SIAM Journal on Control 9*, 199–233.

Sima, V. (1996). *Algorithms for linear-quadratic optimization*. Marcel Dekker, New York-Basel. ISBN 0-8247-9612-8.

Skogestad, S., M. Hovd, and P. Lundström (1991). Simple frequency-dependent tools for analysis of inherent control limitations. *Modeling Identification and Control 12*, 159–177.

Skogestad, S. and M. Morari (1987). Implications of large RGA elements on control performance. *Industrial and Engineering Chemistry Research 26*, 2323–2330.

Skogestad, S. and M. Morari (1992). Variable selection for decentralized control. *Modeling Identification and Control 13*, 113–125.

Skogestad, S. and I. Postlethwaite (1995). *Multivariable Feedback Control. Analysis and Design*. John Wiley and Sons Ltd., Chichester, Sussex, UK.

Stoorvogel, A. A. (1992). *The H-Infinity Control Problem: A State Space Approach*. Prentice Hall, Englewood Cliffs, NJ.

Stoorvogel, A. A. and J. H. A. Ludlage (1994). Squaring down and the problems of almost zeros for continuous time systems. *Systems and Control Letters 23*, 381–388.

Svaricek, F. (1995). *Zuverlässige numerische Analyse linearer Regelungssysteme*. B.G.Teubner Verlagsgesellschaft, Stuttgart, BRD.

Thaler, G. J. (1973). *Design of Feedback Systems*. Dowden, Hutchinson, and Ross, Stroudsburg, PA.

Tolle, H. (1983). *Mehrgrössen-Regelkreissynthese. Band I. Grundlagen und Frequenzbereichsverfahren*. R.Oldenbourg Verlag, München, BRD.

Tolle, H. (1985). *Mehrgrössen-Regelkreissynthese. Band II: Entwurf im Zustandsraum*. R.Oldenbourg Verlag, München, BRD.

Trentelman, H. L. and A. A. Stoorvogel (1993, November). Control theory for linear systems. Vakgroep Wiskunde, RUG. Course for the Dutch Graduate Network on Systems and Control, Winter Term 1993–1994.

Truxal, J. G. (1955). *Automatic Feedback Control System Synthesis*. McGraw-Hill Book Co., New York, NY.

Van de Vegte, J. (1990). *Feedback control systems*. Prentice-Hall, Englewood Cliffs, N. J. Second Edition.

VandenBoom, A. and F. Brown (1992). SLICOT, a subroutine library in control and systems theory. In H. A. Barker (Ed.), *Computer Aided Design in Control Systems, Selected Papers from the IFAC Symposium, Swansea, UK, 15-17 july 1991.*, pp. 89–94. Pergamon Press, Oxford, England.

VanDooren, P. (1979). The computation of Kronecker's canonical form of a singular pencil. *Linear Algebra and its Applications 27*, 103–140.

Vardulakis, A. I. G. (1991). *Linear Multivariable Control: Algebraic Analysis and Synthesis Methods*. John Wiley and Sons Ltd., Chichester, Sussex, UK.

Verma, M. and E. Jonckheere (1984). L_∞-compensation with mixed sensitivity as a broadband matching problem. *Systems & Control Letters 14*, 295–306.

Vidyasagar, M., H. Schneider, and B. A. Francis (1982). Algebraic and topological aspects of feedback stabilization. *IEEE Trans. Aut. Control 27*, 880–894.

Vidysagar, M. (1985). *Control Systems Synthesis—A Factorization Approach*. MIT Press, Cambridge, MA.

Wilson, G. T. (1978). A convergence theorem for spectral factorization. *J. Multivariate Anal. 8*, 222–232.

Wolovich, W. A. (1974). *Linear Multivariable Systems*. Springer Verlag, New York, NY.

Wong, E. (1983). *Introduction to random processes*. Springer-Verlag, New York.

Wonham, W. M. (1979a). *Linear Multivariable Control: A Geometric Approach*. Springer-Verlag, New York, etc.

Wonham, W. M. (1979b). *Linear Multivariable Control: a Geometric Approach. Second Edition*. Springer Verlag, New York, NY.

Youla, D. C., J. J. Bongiorno, and C. N. Lu (1974). Single-loop feedback-stabilization of linear multivariable dynamical plants. *Automatica 10*, 159–173.

Young, P. M., M. P. Newlin, and J. C. Doyle (1995a). Computing bounds for the mixed μ problem. *Int. J. of Robust and Nonlinear Control 5*, 573–590.

Young, P. M., M. P. Newlin, and J. C. Doyle (1995b). Let's get real. In B. A. Francis and P. P. Khargonekar (Eds.), *Robust Control Theory*, Volume 66 of *IMA Volumes in Mathematics and its Applications*, pp. 143–173. Springer-Verlag.

Zames, G. (1981). Feedback and optimal sensitivity: Model reference transformations, multiplicative seminorms, and approximate inverses. *IEEE Trans. Aut. Control 26*, 301–320.

Zhou, K., J. C. Doyle, and K. Glover (1996). *Robust and Optimal Control*. Prentice Hall, Englewood Cliffs. ISBN 0-13-456567-3.

Ziegler, J. G. and N. B. Nichols (1942). Optimum settings for automatic controllers. *Trans. ASME, J. Dyn. Meas. and Control 64*, 759–768.

Index