

Design Methods for Control Systems

Okko H. Bosgra

*Delft University of Technology
Delft, The Netherlands*

Huibert Kwakernaak

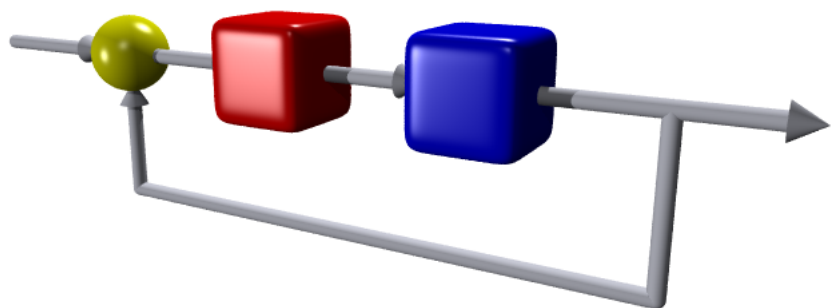
*Emeritus Professor
University of Twente
Enschede, The Netherlands*

Gjerrit Meinsma

*University of Twente
Enschede, The Netherlands*

*Notes for a course of the
Dutch Institute of Systems and Control
Winter term 2005–2006*





Preface

Part of these notes were developed for a course of the Dutch Network on Systems and Control with the title “Robust control and \mathcal{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 were added and Chapter 4 was revised. A couple of mistakes were corrected in the 2001–2002, 2002–2003 and 2003–2004 issue. H_∞ theory was updated in 2004–2005 and the main modification in 2005–2006 was that chapters 3 and 4 were interchanged.

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	1
1.1. Introduction	1
1.2. Basic feedback theory	3
1.3. Closed-loop stability	11
1.4. Stability robustness	20
1.5. Frequency response design goals	27
1.6. Loop shaping	34
1.7. Limits of performance	40
1.8. Two-degrees-of-freedom feedback systems	47
1.9. Conclusions	51
1.10. Appendix: Proofs	52
2. Classical Control System Design	59
2.1. Introduction	59
2.2. Steady state error behavior	60
2.3. Integral control	64
2.4. Frequency response plots	69
2.5. Classical control system design	80
2.6. Lead, lag, and lag-lead compensation	82
2.7. The root locus approach to parameter selection	88
2.8. The Guillemin-Truxal design procedure	90
2.9. Quantitative feedback theory (QFT)	93
2.10. Concluding remarks	101
3. Multivariable Control System Design	103
3.1. Introduction	103
3.2. Poles and zeros of multivariable systems	108
3.3. MIMO structural requirements and design methods	117
3.4. Appendix: Proofs and Derivations	131
3.5. Exercises	133
4. LQ, LQG and \mathcal{H}_2 Control System Design	135
4.1. Introduction	135
4.2. LQ theory	136
4.3. LQG Theory	145
4.4. \mathcal{H}_2 optimization	152
4.5. Feedback system design by \mathcal{H}_2 optimization	155
4.6. Examples and applications	161
4.7. Appendix: Proofs	167
4.8. Exercises	178

5. Uncertainty Models and Robustness	183
5.1. Introduction	183
5.2. Parametric robustness analysis	184
5.3. The basic perturbation model	195
5.4. The small gain theorem	198
5.5. Stability robustness of feedback systems	205
5.6. Structured singular value robustness analysis	216
5.7. Combined performance and stability robustness	223
5.8. Appendix: Proofs	230
5.9. Exercises	233
6. \mathcal{H}_∞-Optimization and μ-Synthesis	239
6.1. Introduction	239
6.2. The mixed sensitivity problem	240
6.3. The standard \mathcal{H}_∞ problem	248
6.4. Suboptimal solutions and an example	250
6.5. State space solution of the standard \mathcal{H}_∞ problem	253
6.6. Optimal solutions to the \mathcal{H}_∞ problem	255
6.7. Integral control and high-frequency roll-off	258
6.8. μ -Synthesis	266
6.9. An application of μ -synthesis	270
6.10. Appendix: Proofs	281
6.11. Exercises	281
A. Matrices	287
A.1. Basic matrix results	287
A.2. Three matrix lemmas	288
B. Norms of signals and systems	291
B.1. Norms of vector-valued signals	291
B.2. Singular values of vectors and matrices	292
B.3. Norms of signals	293
B.4. Norms of linear operators and systems	293
B.5. BIBO and internal stability	296
B.6. Appendix: Proofs	299
B.7. Problems	301
C. Bibliography	305
Index	315

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

$$\text{targets} \quad \left\{ \begin{array}{l} \bullet \text{ closed-loop stability,} \\ \bullet \text{ disturbance attenuation,} \\ \bullet \text{ stability robustness,} \end{array} \right.$$

within the limitations set by

$$\text{limitations} \quad \left\{ \begin{array}{l} \bullet \text{ plant capacity,} \\ \bullet \text{ corruption by measurement noise.} \end{array} \right.$$

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

$$\text{further targets} \quad \left\{ \begin{array}{l} \bullet \text{ satisfactory closed-loop response,} \\ \bullet \text{ robustness of the closed-loop response.} \end{array} \right.$$

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 level¹. 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.

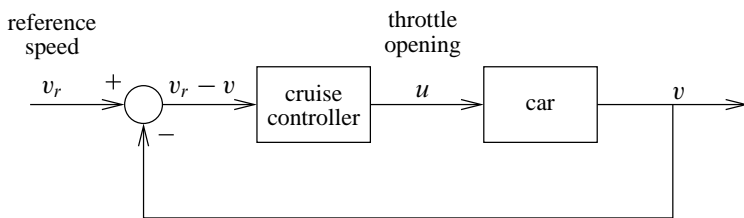


Figure 1.1: Block diagram of the cruise control system

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

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

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

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

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

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

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

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

$$w = \frac{v}{v_{\text{max}}} \quad (1.4)$$

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

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

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

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

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

$$\dot{w} = -\frac{1}{\theta}w + \frac{1}{T}u, \quad t \geq 0, \quad (1.7)$$

with

$$\theta = \frac{T}{2w_0}. \quad (1.8)$$

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]. \square

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

$$u(t) = 1, \quad t \geq 0, \quad (1.9)$$

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

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

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

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

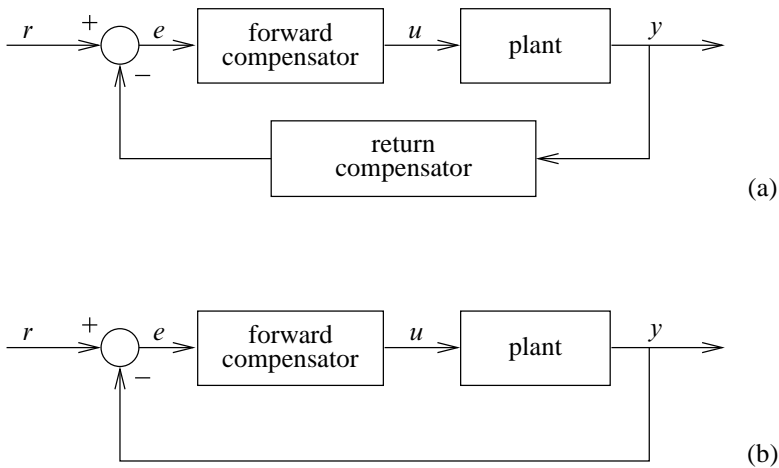


Figure 1.2: Feedback configurations: (a) General. (b) Unit feedback

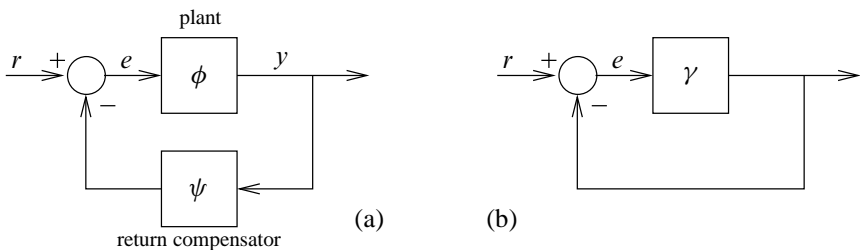


Figure 1.3: (a) Feedback configuration with input-output maps. (b) Equivalent unit feedback configuration

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*.

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

$$y = \phi(e), \quad e = r - \psi(y). \quad (1.11)$$

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

$$e + \gamma(e) = r. \quad (1.12)$$

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

$$e + \gamma(e) = r \quad (1.13)$$

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

$$\|\gamma(e)\| \gg \|e\|, \quad (1.14)$$

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

$$\gamma(e) \approx r. \quad (1.15)$$

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

$$\|e\| \ll \|r\|. \quad (1.16)$$

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

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

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 solution² 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 stable³. 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 \mathcal{E} . The class usually consists of band- and amplitude-limited signals, and depends on the “capacity” of the plant.

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

³A 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

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

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

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

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

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

that is,

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

Stability is ensured as long as

$$\frac{1}{\theta} + \frac{g}{T} > 0. \quad (1.22)$$

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

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

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 $H_{cl}(j\omega)$, $\omega \in \mathbb{R}$, for different values of the gain g . If the gain g is large then $H_{cl}(j\omega) \approx 1$ for low frequencies. The larger the gain, the larger the frequency region is over which this holds. \square

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

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

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

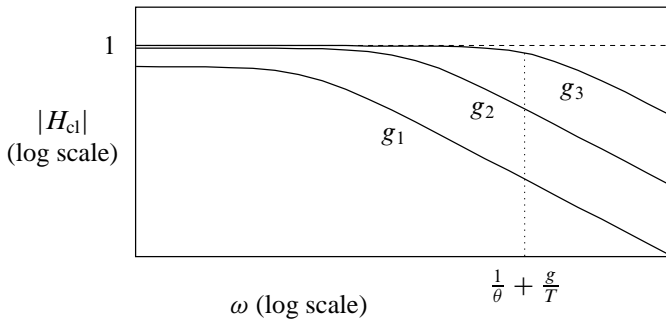


Figure 1.4: Magnitude plots of the closed-loop frequency response function for three values of the gain with $g_1 < g_2 < g_3$

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

$$\frac{1}{\theta_{cl}} = \frac{1}{\theta} + \frac{g}{T}. \quad (1.25)$$

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

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

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

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

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

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.

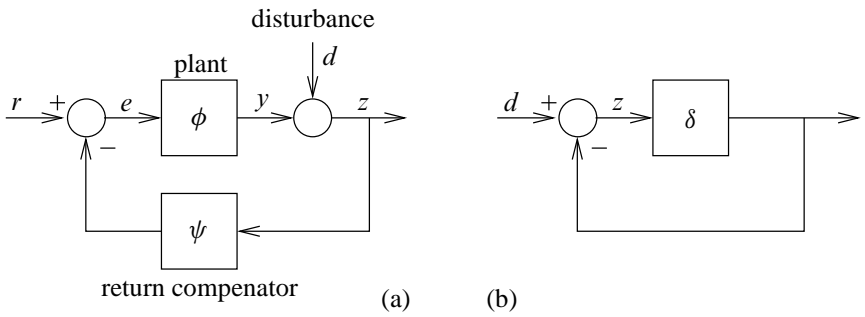


Figure 1.5: (a) Feedback system with disturbance. (b) Equivalent unit feedback configuration in the absence of the control input r

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

$$\theta_{cl} = \frac{\theta}{1 + \frac{g\theta}{T}}. \quad (1.28)$$

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. \square

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

$$T\dot{w} = u - w^2. \quad (1.29)$$

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

$$u - u_0 = g(r - w). \quad (1.30)$$

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. \square

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

$$z = d - \delta(z), \quad (1.31)$$

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

$$\|z\| \ll \|d\|. \quad (1.32)$$

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

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

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

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

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

$$w_{\text{cl}} = \frac{1}{s + \frac{1}{\theta_{\text{cl}}}} d. \quad (1.35)$$

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

$$w_{\text{ol}} = \frac{1}{s + \frac{1}{\theta}} d. \quad (1.36)$$

This signal w_{ol} actually is the “equivalent disturbance at the output” of Fig. 1.5(a). Comparison of (1.34) and (1.35) shows that

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

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 $S(j\omega)$. The plot shows that the open-loop disturbances are attenuated by a factor

$$\frac{\theta_{\text{cl}}}{\theta} = \frac{1}{1 + \frac{g\theta}{T}} \quad (1.38)$$

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 $1/\theta_{cl}$.

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

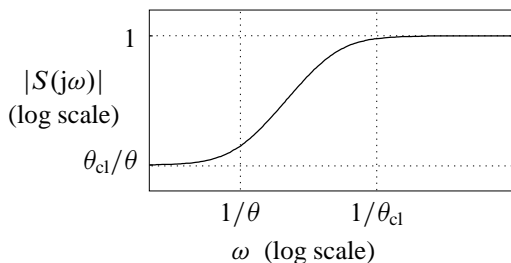


Figure 1.6: Magnitude plot of the sensitivity function of the proportional cruise control system

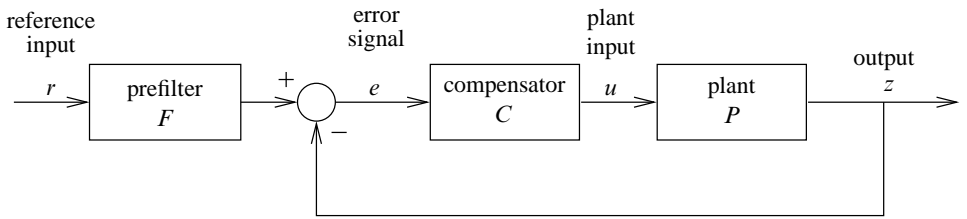


Figure 1.7: Two-degree-of-freedom feedback system configuration

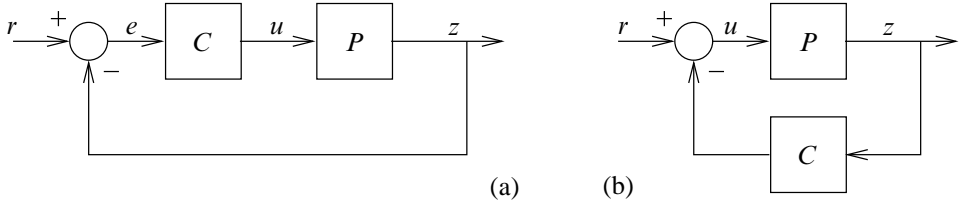


Figure 1.8: Single-degree-of-freedom feedback system configurations

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

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

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

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. \square

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 Appendix B. 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 modes⁴ 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 modes⁵ then it is stable in the sense of Definition 1.3.1. \square

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. \square

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).

⁴A 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 $t \rightarrow \infty$.

⁵The 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 .

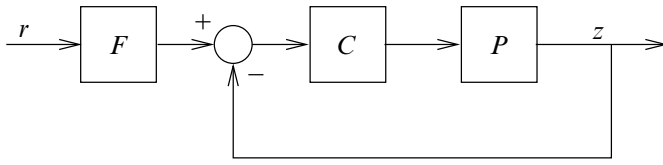


Figure 1.9: Two-degree-of-freedom feedback system

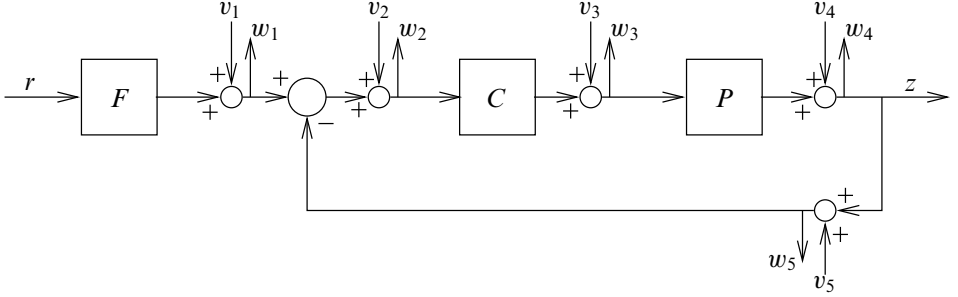


Figure 1.10: Two-degree-of-freedom system with internal inputs and outputs added

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.

We say that a controller *stabilizes* a loop, or, *is stabilizing*, if the closed loop with that controller is stable or internally stable.

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

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

$$y(t) = Cx(t) + Du(t), \quad (1.42)$$

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

$$\chi(s) = \det(sI - A). \quad (1.43)$$

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 $\begin{bmatrix} x \\ z \end{bmatrix}$ 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. \square

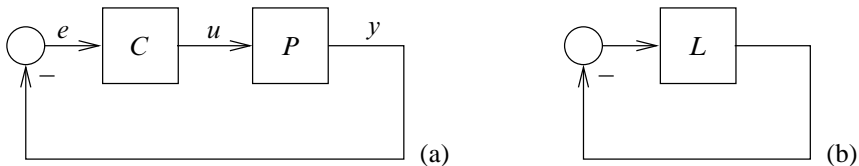


Figure 1.11: MIMO or SISO feedback systems

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 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, $L(\infty)$ exists.

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

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

We call χ_{cl} the *closed-loop characteristic polynomial*.

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

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

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

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

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

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

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

$$D(s)X(s) + N(s)Y(s). \quad (1.48)$$

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 coefficient⁶.

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

$$\chi = DX + NY \quad (1.49)$$

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, 1980).

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.

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

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

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

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

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 $\frac{1}{2}\sqrt{2}(-1 \pm j)$ to obtain a closed-loop bandwidth of 1 [rad/s], and place a non-dominant pair at $5\sqrt{2}(-1 \pm j)$. Hence,

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

Write $X(s) = x_2s^2 + x_1s + x_0$ and $Y(s) = y_1s + y_0$. Then

$$\begin{aligned} D(s)X(s) + N(s)Y(s) &= s^2(x_2s^2 + x_1s + x_0) + (y_1s + y_0) \\ &= x_2s^4 + x_1s^3 + x_0s^2 + y_1s + y_0. \end{aligned} \quad (1.52)$$

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

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

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

□

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

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

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

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

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

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

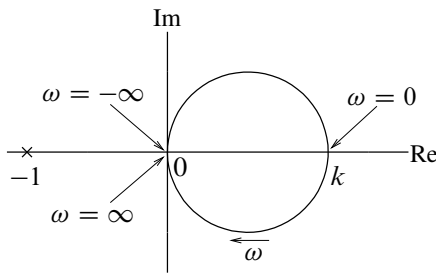


Figure 1.12: Nyquist plot of the loop gain transfer function $L(s) = k/(1 + s\theta)$

This in turn may be represented as

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

and solved for \mathbf{x} and \mathbf{y} . If the polynomials D and N are coprime⁸ 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 C 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

$$L(j\omega), \quad \omega \in \mathbb{R}. \quad (1.60)$$

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

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

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

⁸That is, they have no nontrivial common factors.

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

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

the closed-loop system is stable if and only if the Nyquist plot of L does not encircle the point -1 . □

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 $L = PC$. 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

$$\begin{aligned} & \text{the number of unstable closed-loop poles} \\ & \quad = \\ & \text{the number of times the Nyquist plot of } \det(I + L) \text{ encircles the origin clockwise}^{11} \\ & \quad + \\ & \text{the number of unstable open-loop poles.} \end{aligned}$$

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.

¹¹This means the number of clockwise encirclements minus the number of anticlockwise encirclements. I.e., this number may be negative.

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 et al. (1974); see also Anderson and Jury (1976) and Blondel (1994).

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. \square

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

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

possesses the parity interlacing property while

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

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

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 .

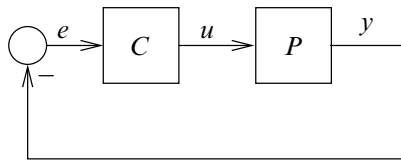


Figure 1.13: Feedback system configuration

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

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

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 $L(j\omega_m)$, 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).

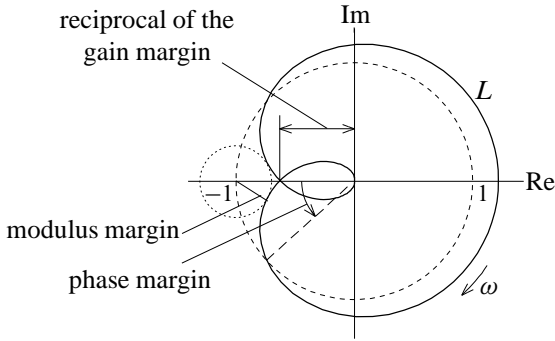


Figure 1.14: Robustness margins

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 et al. (1993) introduced two more margins.

Modulus margin¹² 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 .

¹²French: *marge de module*.

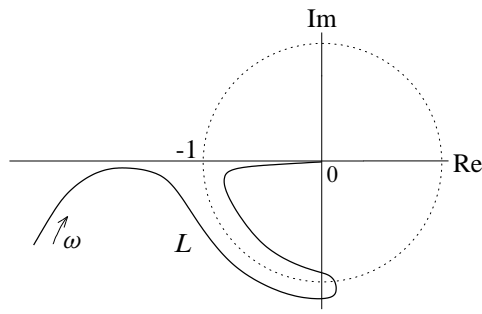


Figure 1.15: This Nyquist plot has good gain and phase margins but a small simultaneous perturbation of gain and phase destabilizes the system

Delay margin ¹³ 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

$$\tau_m = \min_{\omega_*} \frac{\phi_*}{\omega_*}. \quad (1.65)$$

Here ω_* ranges over all nonnegative frequencies at which the Nyquist plot intersects the unit circle, and ϕ_* denotes the corresponding phase $\phi_* = \arg L(j\omega_*)$. In particular $\tau_m \leq \frac{\phi_m}{\omega_m}$.

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

$$k_m \geq \frac{1}{1 - s_m}, \quad \phi_m \geq 2 \arcsin \frac{s_m}{2}. \quad (1.66)$$

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 et al., 1993). The converse is not true in general. \square

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.

¹³French: *marge de retard*.

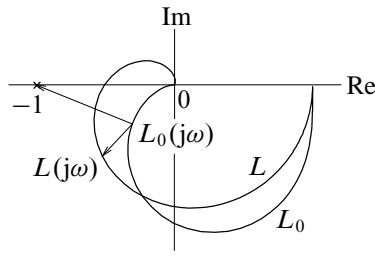


Figure 1.16: Nominal and perturbed Nyquist plots

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 $|L(j\omega) - L_0(j\omega)|$ between any point $L(j\omega)$ and 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

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

This is equivalent to

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

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

$$T_0 = \frac{L_0}{1 + L_0}. \quad (1.69)$$

T_0 bears its name because its complement

$$1 - T_0 = \frac{1}{1 + L_0} = S_0 \quad (1.70)$$

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

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

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

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

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

This result is discussed more extensively in Section 5.5, 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

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

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 be bounded in the form

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

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

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

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) (Vidyasagar, 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

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

with W a given function, if and only if

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

□

Again, the MIMO version is presented in Section 5.5. 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

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

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

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

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

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

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

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

with W a given function, if and only if

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

□

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

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

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

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

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

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 $1/T_0$.

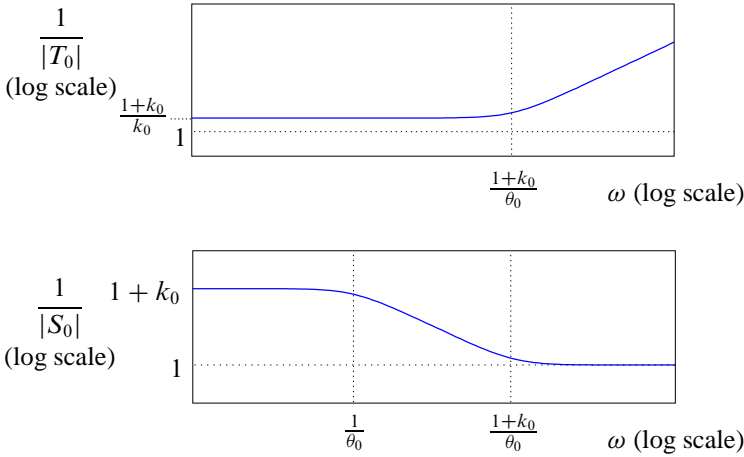


Figure 1.17: Magnitude plots of $1/T_0$ and $1/S_0$

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. \square

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 :

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

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

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

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 .

□

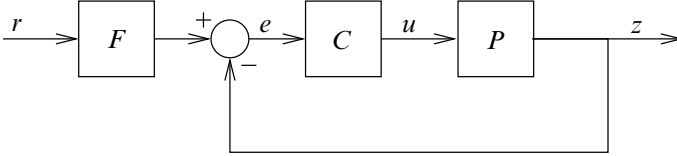


Figure 1.18: Two-degree-of-freedom feedback system

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

- 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 .

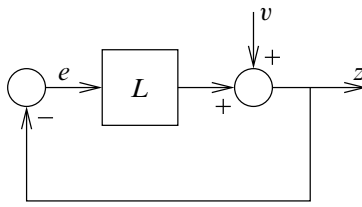


Figure 1.19: Feedback loop with disturbance

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

$$z = \underbrace{\frac{1}{1+L}}_S v = Sv, \quad (1.88)$$

where S is the sensitivity function of the closed-loop system. The smaller $|S(j\omega)|$ 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 1.7.

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 shape¹⁴ for the complementary sensitivity T corresponding to the sensitivity 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

¹⁴Note that S and T are complementary, not $|S|$ and $|T|$.

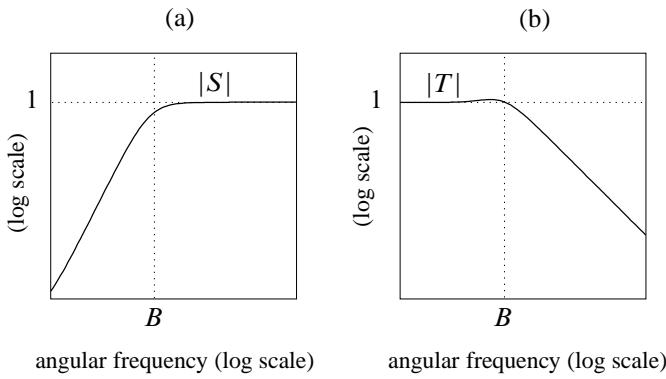


Figure 1.20: (a) "Ideal" sensitivity function. (b) A corresponding complementary sensitivity function

frequency. For this reason it sometimes is useful to replace the requirement that S be small with the requirement that

$$|S(j\omega)V(j\omega)| \quad (1.89)$$

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 $V = P$.

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

$$P(s) = \frac{\omega_0^2}{s^2 + 2\zeta_0\omega_0s + \omega_0^2}. \quad (1.90)$$

Choose the compensator

$$C(s) = \frac{k}{\omega_0^2} \frac{s^2 + 2\zeta_0\omega_0s + \omega_0^2}{s(s + \alpha)}. \quad (1.91)$$

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

$$z = \underbrace{\frac{PC}{1+PC}}_H F r. \quad (1.92)$$

The closed-loop transfer function H may be expressed as

$$H = \underbrace{\frac{L}{1+L}}_T F = TF, \quad (1.93)$$

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

$$u = \underbrace{\frac{C}{1+CP}}_M (Fr - m - v). \quad (1.94)$$

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.

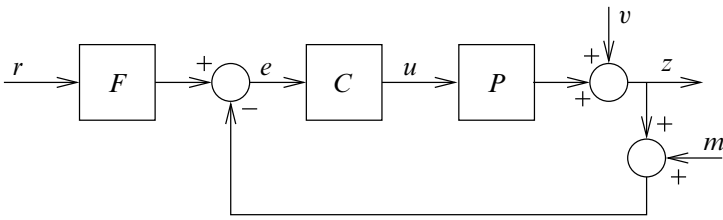


Figure 1.21: Two-degree-of-freedom system with disturbances and measurement noise

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

$$T = MP. \quad (1.95)$$

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 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 $1 - e^{-t/\theta}$, $t \geq 0$, 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

$$z = \underbrace{\frac{1}{1+PC}}_S v + \underbrace{\frac{PC}{1+PC}}_T Fr - \underbrace{\frac{PC}{1+PC}}_T m. \quad (1.96)$$

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

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

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

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

$$\frac{\frac{1}{S} - \frac{1}{S_0}}{\frac{1}{S_0}} = \frac{S_0 - S}{S} = T_0 \frac{L - L_0}{L_0}. \quad (1.99)$$

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

$$\frac{\frac{1}{T} - \frac{1}{T_0}}{\frac{1}{T_0}} = \frac{T_0 - T}{T} = S_0 \frac{L_0 - L}{L} = S_0 \frac{\frac{1}{L} - \frac{1}{L_0}}{\frac{1}{L_0}}. \quad (1.100)$$

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 $L(j\omega)$, $\omega \in \mathbb{R}$, 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

$$S = \frac{1}{1+L}, \quad T = \frac{L}{1+L} \quad (1.101)$$

shows that these targets may be achieved simultaneously by making the loop gain L large, that is, by making $|L(j\omega)| \gg 1$ 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 $|L(j\omega)| \ll 1$ 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 .

Crossover region. In the crossover region we have $|L(j\omega)| \approx 1$. 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

$$S = \frac{1}{1+L} \quad (1.102)$$

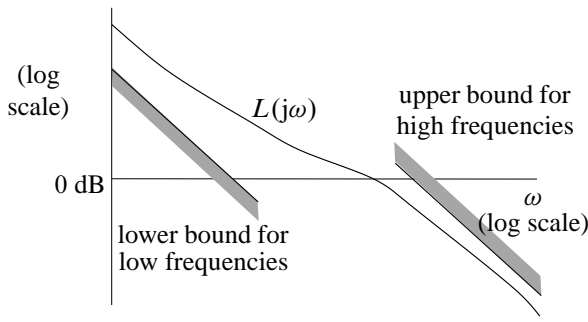


Figure 1.22: Robustness bounds on L in the Bode magnitude plot

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

$$T = \frac{L}{1+L} = \frac{1}{1+\frac{1}{L}} \quad (1.103)$$

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 $|L(j\omega)|$, $\omega \in \mathbb{R}$, and the phase $\arg L(j\omega)$, $\omega \in \mathbb{R}$, of a minimum phase¹⁵ linear time-invariant system with real-rational¹⁶ transfer function L are uniquely related. If on a log-log scale the plot of the magnitude $|L(j\omega)|$ versus ω has an approximately constant slope n [decade/decade] then

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

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

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

$$L(j\omega) \approx c(j\omega)^n, \quad (1.105)$$

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 ? \square

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

¹⁵A rational transfer function is minimum phase if all its poles and zeros have strictly negative real parts (see Exercise 1.6.3).

¹⁶That is, $L(s)$ is a rational function in s with real coefficients.

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

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

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

$$u = \log \frac{\omega}{\omega_0}. \quad (1.107)$$

W is the function

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

L_{ω_0} , finally, is given by

$$L_{\omega_0}(ju) = L(j\omega), \quad u = \log \frac{\omega}{\omega_0}. \quad (1.109)$$

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

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 its weight near 0, so that $\arg L(j\omega_0)$ is determined by

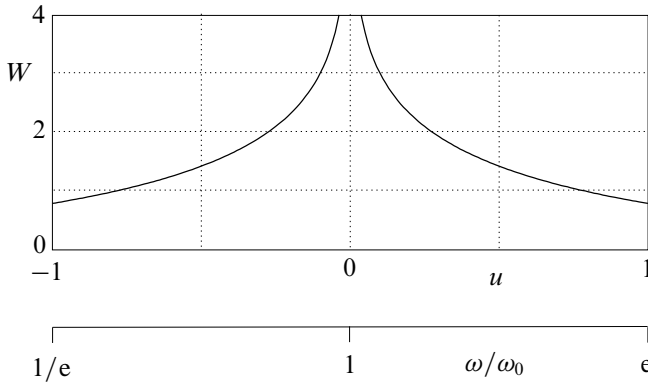


Figure 1.23: Weighting function W in Bode's gain-phase relationship

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.

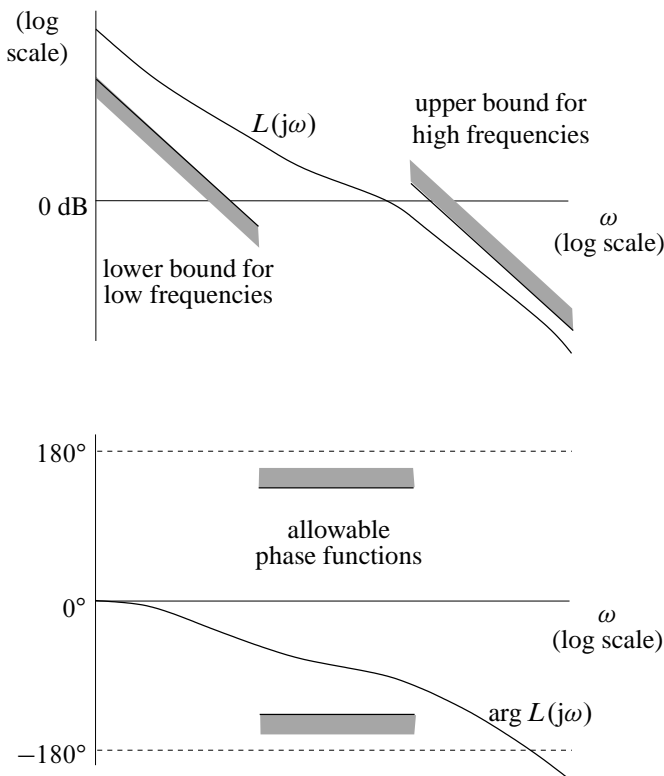


Figure 1.24: Allowable regions for gain and phase of the loop gain L

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 decrease 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 stable¹⁷. Then $L = L_m \cdot L_z$, where L_m is minimum phase and L_z is an all-pass function¹⁸ such that $|L_z(j\omega)| = 1$ for all ω . It follows that $|L(j\omega)| = |L_m(j\omega)|$ and

$$\arg L(j\omega) = \arg L_m(j\omega) + \arg L_z(j\omega) \leq \arg L_m(j\omega). \quad (1.110)$$

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

¹⁷That is, L has right-half plane zeros but no right-half plane poles.

¹⁸That is, $|L_z(j\omega)|$ is constant for all $\omega \in \mathbb{R}$.

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

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

be a rational transfer function with all its poles p_1, p_2, \dots, 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 i th 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. \square

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 plane¹⁹. Then if L has at least two more poles than zeros the sensitivity function

$$S = \frac{1}{1 + L} \quad (1.112)$$

satisfies

$$\int_0^\infty \log |S(j\omega)| d\omega = 0. \quad (1.113)$$

\square

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 two²⁰.” 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

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

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

¹⁹That is, no poles with strictly positive real part.

²⁰In adaptive control the expression is “ L has relative degree greater than or equal to 2.”

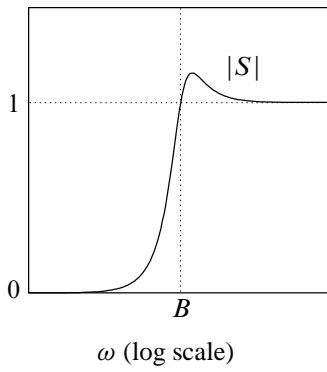


Figure 1.25: Low frequency disturbance attenuation may only be achieved at the cost of high frequency disturbance amplification

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

Exercise 1.6.5 (Bode integral).

1. 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?
2. 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

$$|S(j\omega)| \leq \alpha < 1, \quad 0 \leq \omega \leq \omega_L, \quad (1.115)$$

and

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

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

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

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 two²¹. Prove that

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

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 ? \square

1.7. Limits of performance

1.7.1. Introduction

In this section²² 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 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).

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

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

²¹This means that $1/L(s)$ behaves as $\mathcal{O}(s^2)$ for $s \rightarrow 0$, (see § 2.3).

²²The title has been taken from Engell (1988) and Boyd and Barratt (1991).

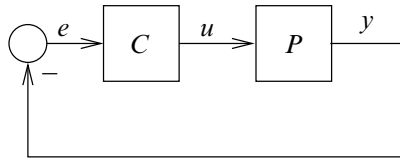


Figure 1.26: SISO feedback system

B_{poles} is the *Blaschke product*

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

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

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

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

with w_z the function

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

We arrange (1.121) in turn as

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

with W_z the function

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

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 $\omega = |z|$.

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

$$\frac{z + j\omega}{z - j\omega}. \quad (1.125)$$

\square

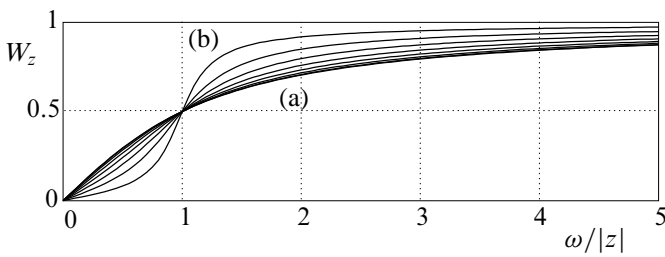


Figure 1.27: The function W_z for values of $\arg z$ increasing from (a) 0 to (b) almost $\pi/2$

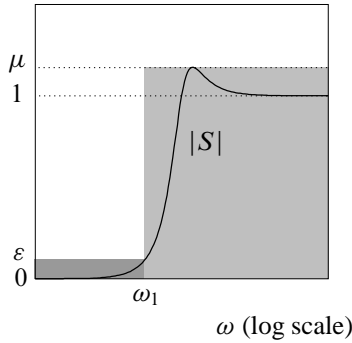


Figure 1.28: Bounds on $|S|$

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. \square

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 $dW_z(j\omega) = w_z(j\omega)d\omega$ 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 particular, a low-frequency band—it necessarily peaks in another band. Suppose that we wish $|S(j\omega)|$ 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

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

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

to satisfy

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

Evaluation of the left-hand side leads to

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

Resolution of (1.128) results in the inequality

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

We note this:

- For a fixed zero $z = x + jy$ 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

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

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).

1. 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.
2. 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.

3. 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.

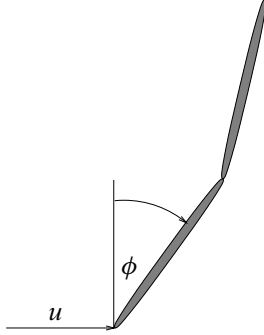


Figure 1.29: Double inverted pendulum system

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

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

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

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

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 $\sup_{\omega \in \mathbb{R}} |S(j\omega)|$ of the sensitivity function S of the closed-loop system. This compensator transfer function is

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

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.

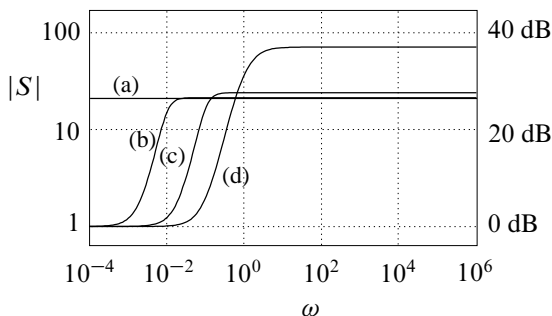


Figure 1.30: Sensitivity functions for the double pendulum system

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

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

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. 20). Hence, no stabilizing compensator exists that by itself is stable. □

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

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

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

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.

2. 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

$$T = \frac{L}{1 + L} = \frac{1}{1 + \frac{1}{L}}. \quad (1.136)$$

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)

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

with B_{zeros} the Blaschke product

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

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).

1. 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.
2. 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.

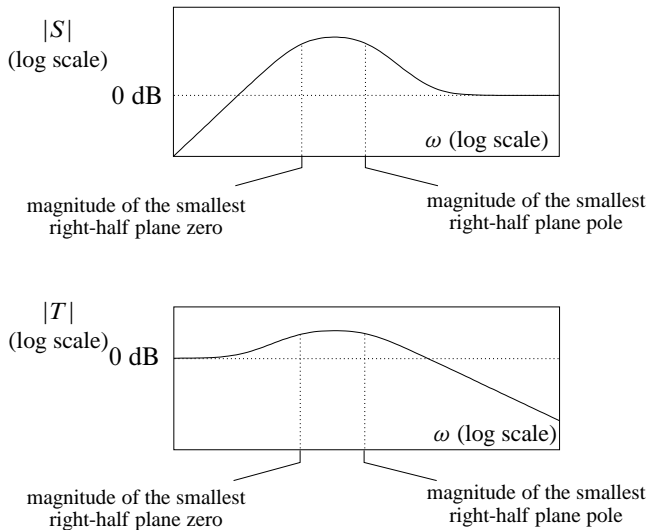


Figure 1.31: Right-half plane zeros and poles constrain S and T

□

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

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

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.

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.

□

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

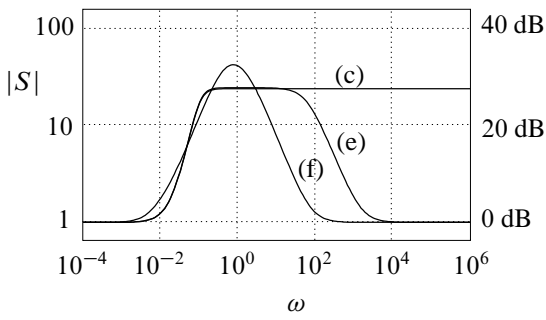


Figure 1.32: More sensitivity functions for the double pendulum system

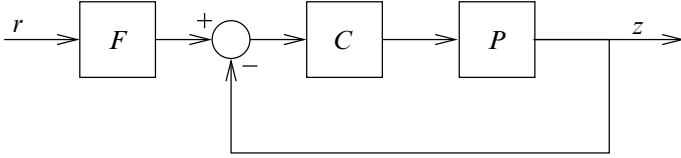


Figure 1.33: Two-degrees-of-freedom feedback system configuration

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 polynomial fraction form

$$P = \frac{N}{D}, \quad C = \frac{Y}{X}. \quad (1.140)$$

The feedback loop is stable if and only if the roots of the closed-loop characteristic polynomial $D_{cl} = DX + NY$ 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

$$H = \frac{PC}{1 + PC} F = \frac{NY}{D_{cl}} F. \quad (1.141)$$

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

$$H_0 = \frac{NY}{D_{cl}}. \quad (1.142)$$

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

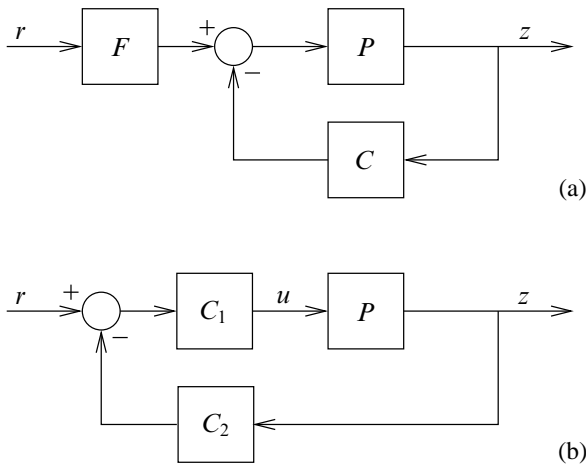


Figure 1.34: Further two-degrees-of-freedom feedback system configurations

closed-loop transfer function

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

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

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

To match the closed-loop characteristics of the configurations of Figs. 1.33 and 1.34(a) we need $C = C_1 C_2$. This implies that $X_1 X_2 = X$ and $Y_1 Y_2 = Y$. The closed-loop transfer function now is

$$H = \frac{PC_1}{1 + PC} = \frac{NX_2 Y_1}{D_{cl}}. \quad (1.145)$$

Inspection shows that the numerator of H is independent of the compensator if we let $X_2 Y_1 = 1$, so that

$$C_1 = \frac{Y_1}{X_1} = \frac{1}{X_1 X_2} = \frac{1}{X}, \quad C_2 = \frac{Y_2}{X_2} = Y_1 Y_2 = Y. \quad (1.146)$$

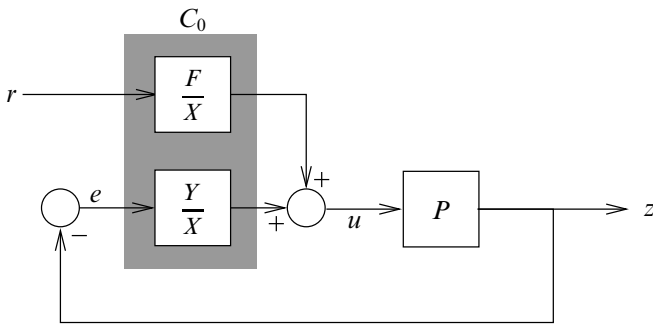


Figure 1.35: $1\frac{1}{2}$ -degrees-of-freedom control system

The closed-loop transfer function now is

$$H = \frac{N}{D_{cl}}. \quad (1.147)$$

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

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

This implies

$$Xu = r + Ye. \quad (1.149)$$

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 disadvantage may be overcome by modifying (1.149) to

$$Xu = Fr + Ye, \quad (1.150)$$

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

$$H = \frac{NF}{D_{cl}}. \quad (1.151)$$

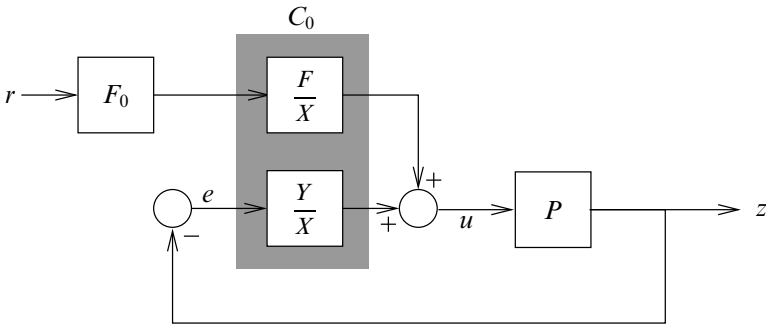


Figure 1.36: $2\frac{1}{2}$ -degrees-of-freedom configuration

The design of the prefilter amounts to choosing the polynomial F . This might be called a $1\frac{1}{2}$ -degrees-of-freedom control system²³. By application of a further prefilter F_0 as in Fig. 1.36 the closed-loop transfer function becomes

$$H = \frac{NF}{D_{cl}} F_0. \quad (1.152)$$

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

$$X(s) = s^n + a_{n-1}s^{n-1} + a_{n-2}s^{n-2} + \cdots + a_0, \quad (1.153)$$

$$F(s) = b_n s^n + b_{n-1}s^{n-1} + b_{n-2}s^{n-2} + \cdots + b_0, \quad (1.154)$$

$$Y(s) = c_n s^n + c_{n-1}s^{n-1} + c_{n-2}s^{n-2} + \cdots + c_0. \quad (1.155)$$

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

²³Half degrees of freedom were introduced in control engineering terminology by Grimble (1994), though with a different connotation than the one used here.

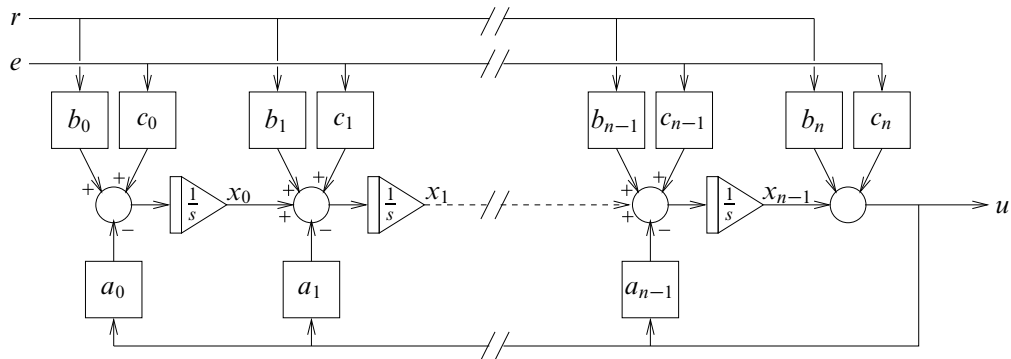


Figure 1.37: Realization of the $1\frac{1}{2}$ -degrees-of-freedom compensator $Xu = Fr + Ye$

- small at high frequencies.

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

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

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

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

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

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

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

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

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

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

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 theory²⁴.

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

$$\begin{aligned} &(\text{the number of times } R(s) \text{ encircles the origin in clockwise direction as } s \text{ traverses } \mathcal{C}) \\ &= \\ &(\text{the number of zeros of } R \text{ inside } \mathcal{C}) - (\text{the number of poles of } R \text{ inside } \mathcal{C}). \end{aligned}$$

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 \mathcal{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 $\det(I + L(j\infty))$. ■

²⁴See Henrici (1974). The generalized form in which we state the principle may be found in Postlethwaite and MacFarlane (1979).

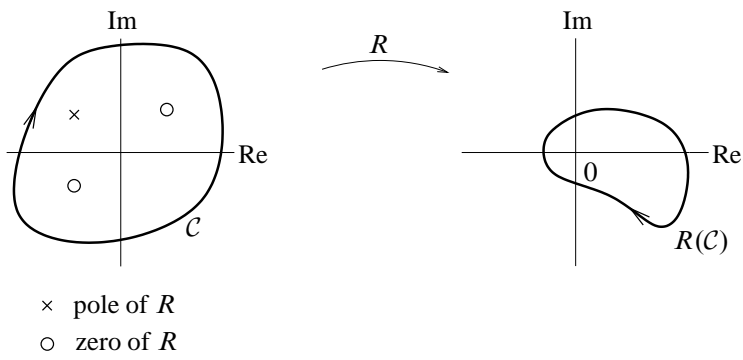


Figure 1.38: Principle of the argument. Left: a closed contour C in the complex plane. Right: the image $R(C)$ of C under a rational function R

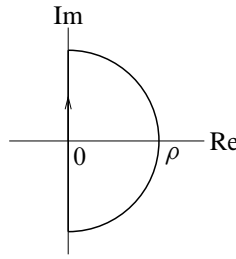


Figure 1.39: Nyquist contour

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

$$\int_0^\infty \log |S(j\omega)| d\omega = 0. \quad (1.161)$$

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

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

As a result,

$$(\omega_H - \omega_L) \sup_{\omega_L \leq \omega \leq \omega_H} \log |S(j\omega)| \geq \omega_L \log \frac{1}{\alpha} - \int_{\omega_H}^\infty \log |S(j\omega)| d\omega. \quad (1.163)$$

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

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

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

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

The assumption that

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

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

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

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

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

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 $\mathbb{C} \rightarrow \mathbb{C}$ that is analytic in the closed right-half plane²⁵ and is such that

$$\lim_{R \rightarrow \infty} \frac{|F(Re^{j\theta})|}{R} = 0 \quad (1.169)$$

for all $\theta \in [-\frac{\pi}{2}, \frac{\pi}{2}]$. Then the value of $F(s)$ at any point $s = x + jy$ in the open right-half plane²⁶ can be recovered from the values of $F(j\omega)$, $\omega \in \mathbb{R}$, by the integral relation

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

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 $s = x + jy$ we may write

$$F(s) = \int_{-\infty}^{\infty} f(t) e^{-(x+jy)t} dt = \int_{-\infty}^{\infty} f(t) e^{-x|t|} e^{-jy t} dt. \quad (1.171)$$

²⁵For rational F this means that F has no poles in $\text{Re}(s) \geq 0$.

²⁶That is, for $\text{Re}(s) > 0$.

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

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

Thus, we have

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

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

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

■

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 polynomials²⁷. Then

$$S = \frac{D}{D + N}. \quad (1.175)$$

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

$$\tilde{S} = B_{\text{poles}}^{-1} S. \quad (1.176)$$

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

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

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

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

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

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

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

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

which is what we set out to prove.

²⁷That 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)

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

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,

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

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 et al. (1947), Evans (1954), Truxal (1955), Savant (1958), Horowitz (1963), Ogata (1970), Thaler (1973), Franklin et al. (1986), D’Azzo and Houpis (1988), Van de Vegte (1990), Franklin et al. (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. 80). 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. 90). 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 $t \rightarrow \infty$ when the reference input r is a polynomial time function of degree n . Thus,

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

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 ε defined by

$$\varepsilon(t) = r(t) - z(t), \quad t \geq 0. \quad (2.2)$$

The steady-state error, if it exists, is

$$\varepsilon_\infty = \lim_{t \rightarrow \infty} \varepsilon(t). \quad (2.3)$$

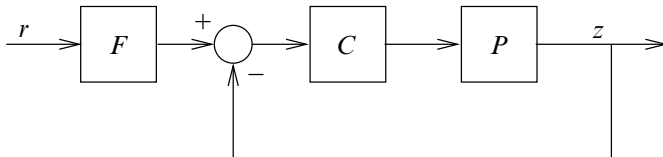


Figure 2.1: Single-loop control system configuration

The Laplace transform of the control error is

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

The function

$$H = \frac{PCF}{1 + PC} = \frac{L}{1 + L}F \quad (2.5)$$

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

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

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

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

and the steady-state error is

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

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

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

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

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

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.

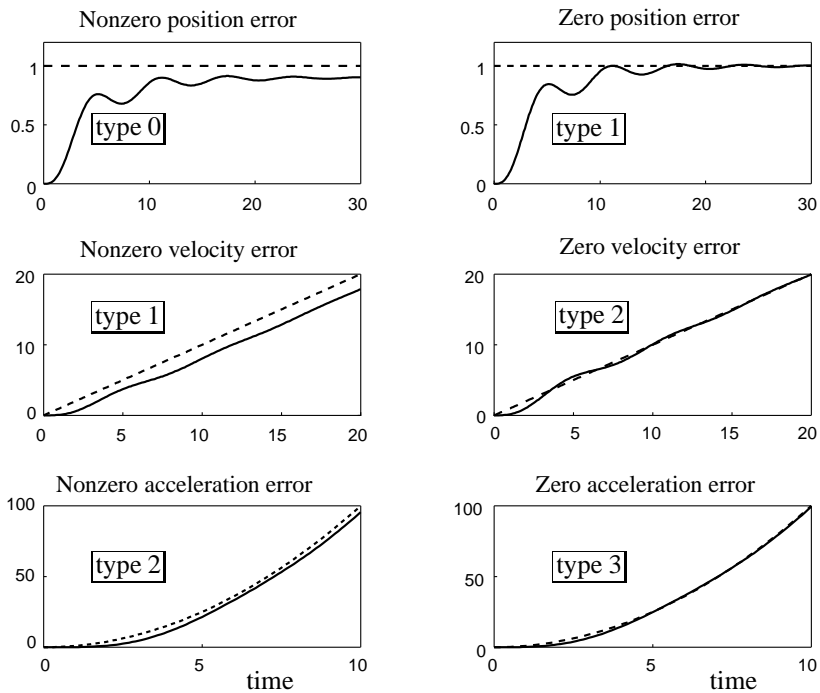


Figure 2.2: System type and steady-state errors

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.

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

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

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

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

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

The number $K_v = \lim_{s \downarrow 0} sL(s)$ is the *velocity 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	$\frac{1}{K_a}$

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

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

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

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

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

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

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 . \square

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

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

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

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

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

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

and

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

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

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

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

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

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

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

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

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*.

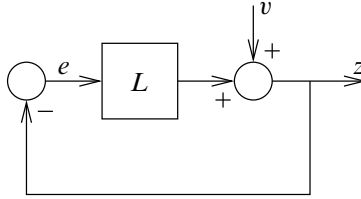


Figure 2.3: Feedback loop

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

$$\lim_{t \rightarrow \infty} z(t) = 0. \quad (2.23)$$

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, 1979) 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

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

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

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

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

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

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

gives considerably more flexibility.

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

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

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

$$\frac{sT_d}{1 + sT}, \quad (2.28)$$

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 et al. (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 g_0 and the period of oscillation as T_0 . Then the parameters of the PID-controller are given by

$$\text{P-controller: } g = 0.5g_0, \quad T_i = \infty, \quad T_d = 0,$$

$$\text{PI-controller: } g = 0.45g_0, \quad T_i = 0.85T_0, \quad T_d = 0,$$

$$\text{PID-controller: } g = 0.6g_0, \quad T_i = 0.5T_0, \quad T_d = 0.125T_0.$$

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

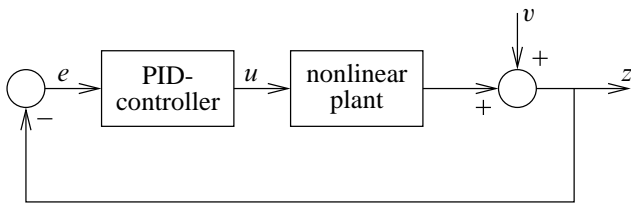


Figure 2.4: Nonlinear plant with integral control

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

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

If the system is controlled with pure integral control

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

then the loop gain and sensitivity functions are

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

The roots of the denominator polynomial

$$s^2 + \frac{1}{\theta}s + \frac{1}{TT_i} \quad (2.32)$$

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 $s^2 + 2\zeta_0\omega_0s + \omega_0^2$, with ω_0 the resonance frequency and ζ_0 the relative damping. It easily follows that

$$\omega_0 = \frac{1}{\sqrt{TT_i}}, \quad \zeta_0 = \frac{1/\theta}{2\omega_0} = \frac{\sqrt{TT_i}}{2\theta}. \quad (2.33)$$

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

$$T_i = \frac{2\theta^2}{T}. \quad (2.34)$$

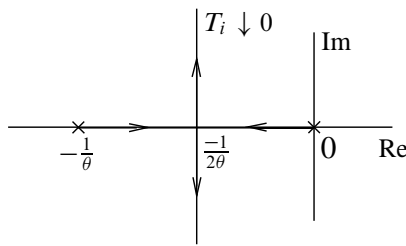


Figure 2.5: Root loci for the cruise control system. The \times s mark the open-loop poles

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.) \square

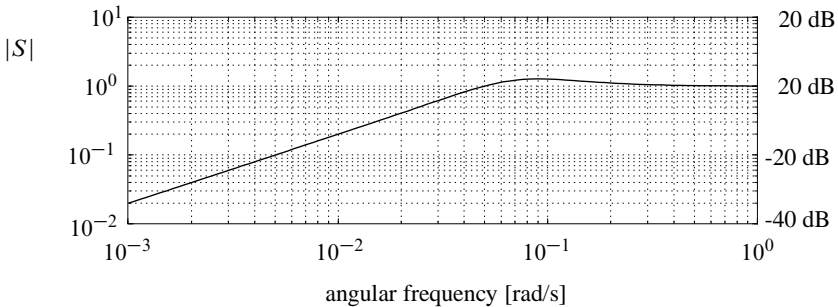


Figure 2.6: Magnitude plot of the sensitivity function of the cruise control system with integrating action

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. \square

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 every constant disturbance if and only if $L^{-1}(0) = 0$. \square

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

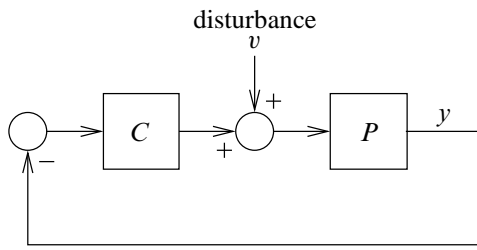


Figure 2.7: Plant with input disturbance

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

$$R = \frac{P}{1 + PC}. \quad (2.35)$$

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 if the integrating action is localized in the compensator. \square

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

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

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

Write

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

with k a constant, z_1, z_2, \dots, z_m the zeros, and p_1, p_2, \dots, p_n the poles of the system. Then for any $s = j\omega$ the magnitude and phase of $L(j\omega)$ 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 $\text{Im } p_i$. This results in large magnitude $|L(j\omega)|$ at this frequency, and explains why a pole that is close to the imaginary axis leads to a resonance peak in the frequency response.

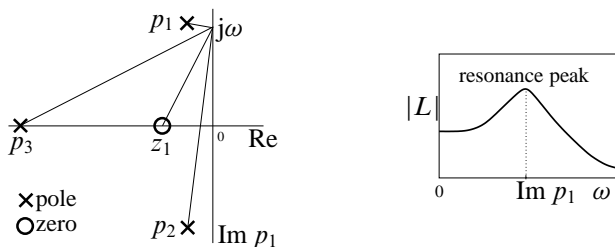


Figure 2.8: Frequency response determined from the pole-zero pattern

2.4.2. Bode plots

A frequency response function $L(j\omega)$, $\omega \in \mathbb{R}$, 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.

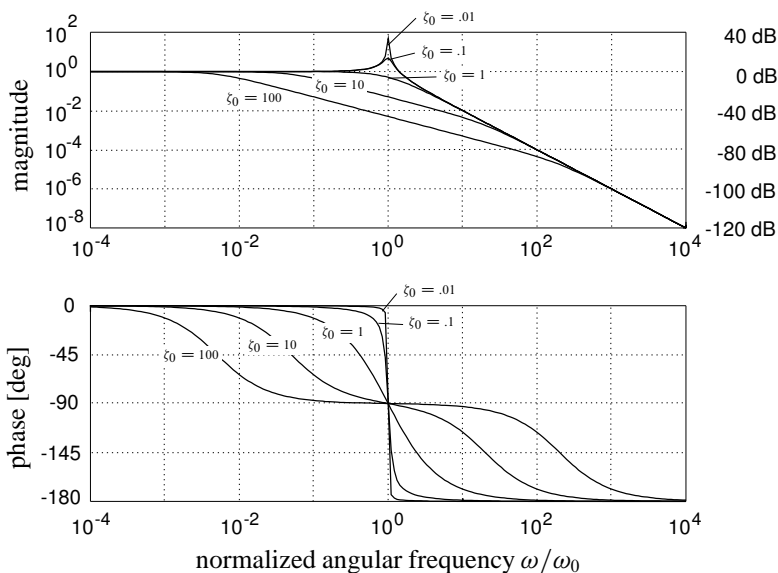


Figure 2.9: Bode diagram of the transfer function $\omega_0^2 / (s^2 + 2\zeta_0\omega_0 s + \omega_0^2)$ for various values of ζ_0

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

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

It follows for the logarithm of the amplitude and for the phase

$$\log |L(j\omega)| = \log |k| + \sum_{i=1}^m \log |j\omega - z_i| - \sum_{i=1}^n \log |j\omega - p_i|, \quad (2.39)$$

$$\arg L(j\omega) = \arg k + \sum_{i=1}^m \arg(j\omega - z_i) - \sum_{i=1}^n \arg(j\omega - p_i). \quad (2.40)$$

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

$$\log |j\omega + \omega_0| \approx \begin{cases} \log |\omega_0| & \text{for } 0 \leq \omega \ll |\omega_0|, \\ \log \omega & \text{for } \omega \gg |\omega_0|, \end{cases} \quad (2.41)$$

$$\arg(j\omega + \omega_0) \approx \begin{cases} \arg(\omega_0) & \text{for } 0 \leq \omega \ll |\omega_0|, \\ 90^\circ & \text{for } \omega \gg |\omega_0|. \end{cases} \quad (2.42)$$

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).

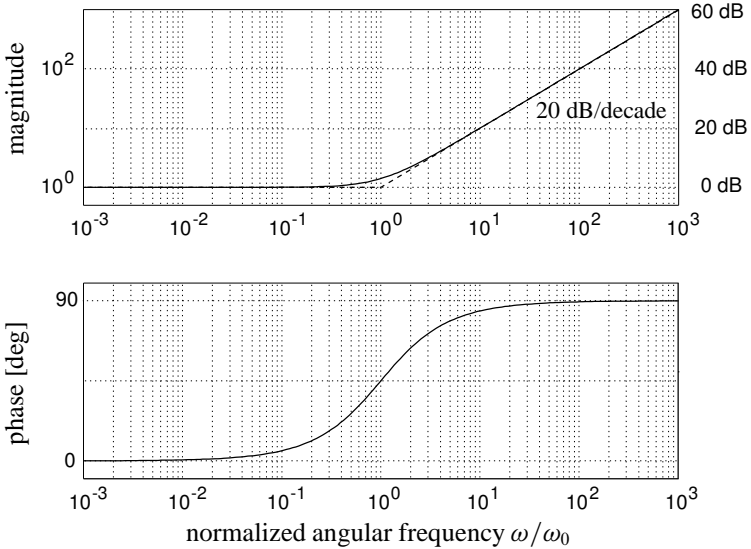


Figure 2.10: Bode amplitude and phase plots for the factor $s + \omega_0$. Dashed: low- and high-frequency asymptotes

Factors corresponding to complex conjugate pairs of poles or zeros are best combined to a **second-order factor of the form**

$$s^2 + 2\zeta_0\omega_0s + \omega_0^2. \quad (2.43)$$

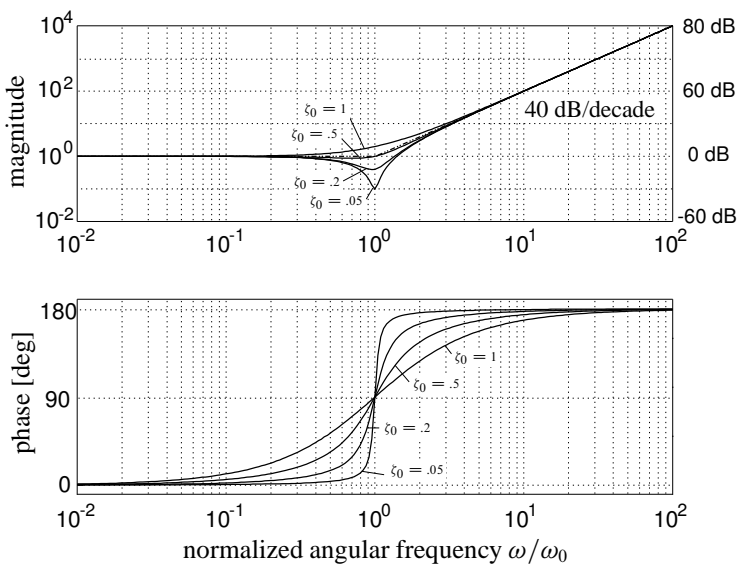


Figure 2.11: Bode plots for a second-order factor $s^2 + 2\zeta_0\omega_0s + \omega_0^2$ for different values of the relative damping ζ_0 . Dashed: low- and high-frequency asymptotes

Asymptotically,

$$\log |(j\omega)^2 + 2\zeta_0\omega_0(j\omega) + \omega_0^2| \approx \begin{cases} 2 \log |\omega_0| & \text{for } 0 \leq \omega \ll |\omega_0|, \\ 2 \log \omega & \text{for } \omega \gg |\omega_0|, \end{cases} \quad (2.44)$$

$$\arg((j\omega)^2 + 2\zeta_0\omega_0(j\omega) + \omega_0^2) \approx \begin{cases} 0 & \text{for } 0 \leq \omega \ll |\omega_0|, \\ 180^\circ & \text{for } \omega \gg |\omega_0|. \end{cases} \quad (2.45)$$

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 $s^2 + 2\zeta_0\omega_0s + \omega_0^2$. 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

$$\omega_0 \left(-\zeta_0 \pm j\sqrt{1 - \zeta_0^2} \right). \quad (2.46)$$

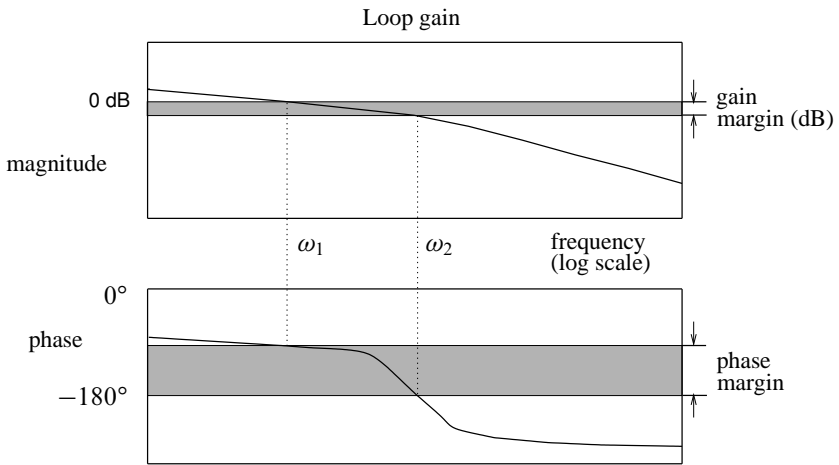


Figure 2.12: Gain and phase margins from the Bode plot of the loop gain

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 $\arccos \zeta_0$.
3. Assume that $0 < \zeta_0 < 1$. At which frequency has the amplitude plot of the factor $(j\omega)^2 + 2\zeta_0\omega_0(j\omega) + \omega_0^2$, $\omega \in \mathbb{R}$, its minimum (and, hence, has the amplitude plot of its reciprocal its maximum)? Note that this frequency is not precisely ω_0 .

□

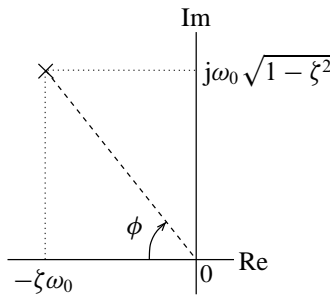


Figure 2.13: Complex conjugate root pair of the factor $s^2 + 2\zeta_0\omega_0 s + \omega_0^2$

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.

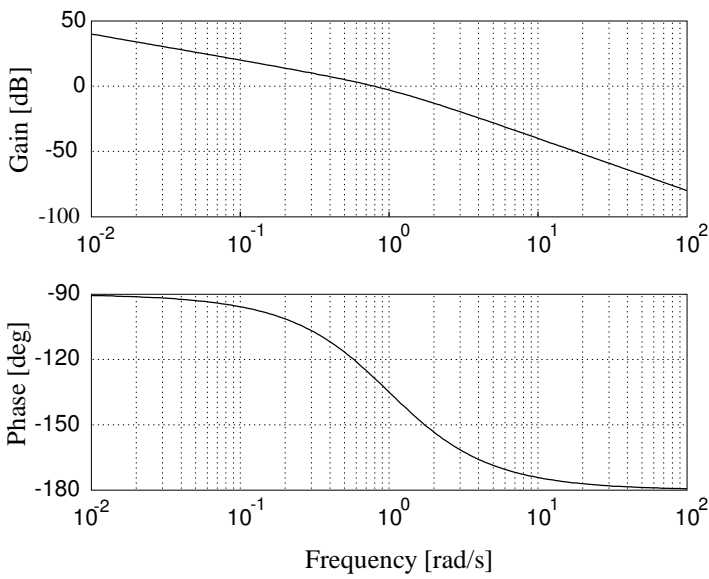


Figure 2.14: Bode diagram of a loop gain L

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. \square

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 $\omega \rightarrow \infty$ the Nyquist plot approaches the origin. Write L in terms of its poles and zeros as in (2.37). Then asymptotically

$$L(j\omega) \approx \frac{k}{(j\omega)^{n-m}} \quad \text{for} \quad \omega \rightarrow \infty. \quad (2.47)$$

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

$$L(j\omega) \approx \frac{c}{(j\omega)^k} \quad \text{for} \quad \omega \downarrow 0, \quad (2.48)$$

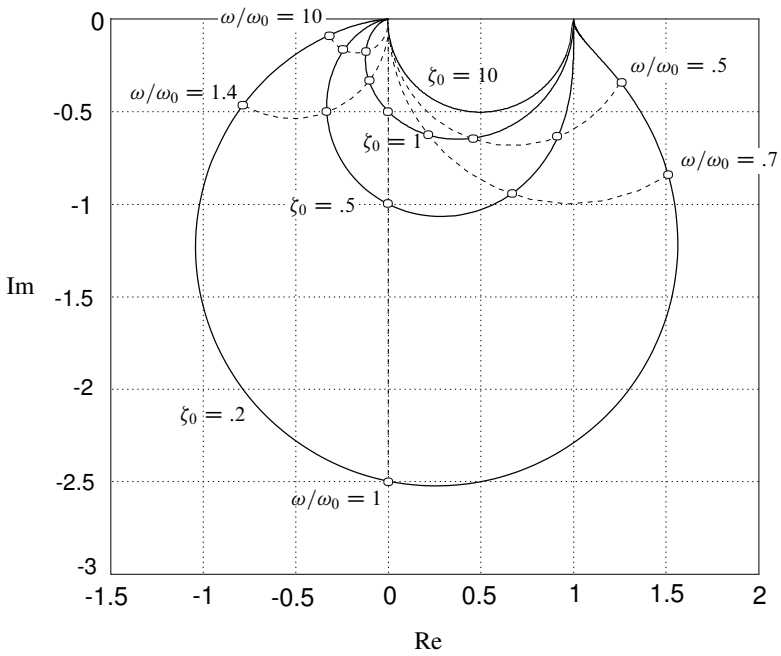


Figure 2.15: Nyquist plots of the transfer function $\omega_0^2/(s^2 + 2\zeta_0\omega_0s + \omega_0^2)$ for different values of the relative damping ζ_0

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 .
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 $L(s) = \omega_0^2/(\omega_0^2 + 2\zeta\omega_0s + s^2)$ is independent of ω_0 .

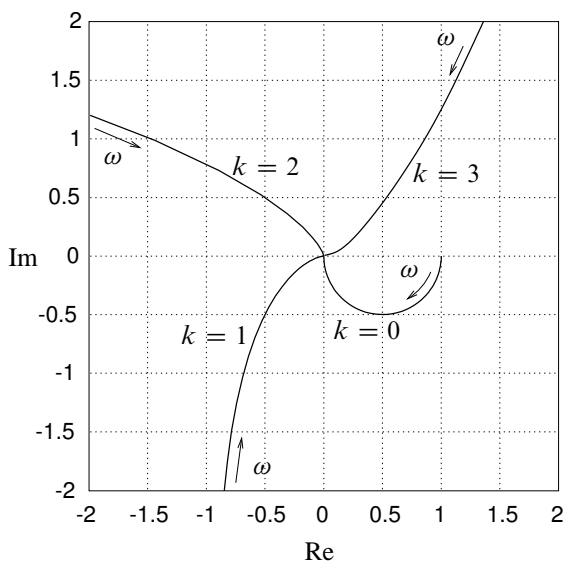
□

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

$$H = T = \frac{L}{1 + L}. \quad (2.49)$$

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 .



$$L(s) = \frac{1}{s^k(1+s)}$$

Figure 2.16: Nyquist plots of the loop gain for different values of system type

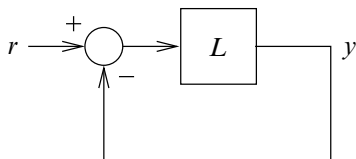


Figure 2.17: Unit feedback loop

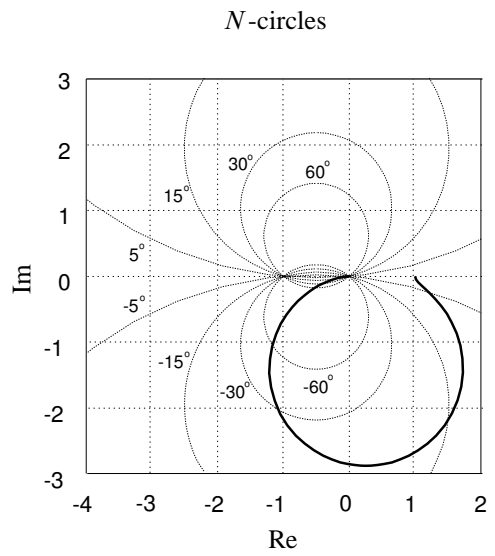
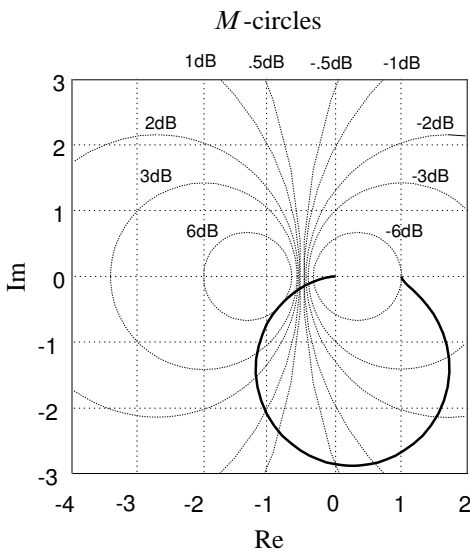


Figure 2.18: M -circles (left) and N -circles (right)

An M -circle is the locus of points z in the complex plane where the magnitude of the complex number

$$\frac{z}{1+z} \quad (2.50)$$

is constant and equal to M . An M -circle has center and radius

$$\text{center} \left(\frac{M^2}{1-M^2}, 0 \right), \quad \text{radius} \left| \frac{M}{1-M^2} \right|. \quad (2.51)$$

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

$$\text{center} \left(-\frac{1}{2}, \frac{1}{2N} \right), \quad \text{radius} \frac{1}{2} \sqrt{1 + \frac{1}{N^2}}. \quad (2.52)$$

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.
- 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 et al., 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 - and 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. □

Exercise 2.4.6 (Lag-lead compensator). Consider a compensator with the second-order transfer function

$$C(s) = \frac{(1 + sT_1)(1 + sT_2)}{(1 + sT_1)(1 + sT_2) + sT_{12}}. \quad (2.53)$$

T_1 , T_2 and T_{12} are time constants. The corresponding frequency response function is

$$C(j\omega) = \frac{(1 - \omega^2 T_1 T_2) + j\omega(T_1 + T_2)}{(1 - \omega^2 T_1 T_2) + j\omega(T_1 + T_2 + T_{12})}, \quad \omega \in \mathbb{R}. \quad (2.54)$$

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 C has the shape of a circle in the right half of the complex plane with its center on the real axis. Since $C(0) = C(j\infty) = 1$ the plot begins and ends in the point 1. □

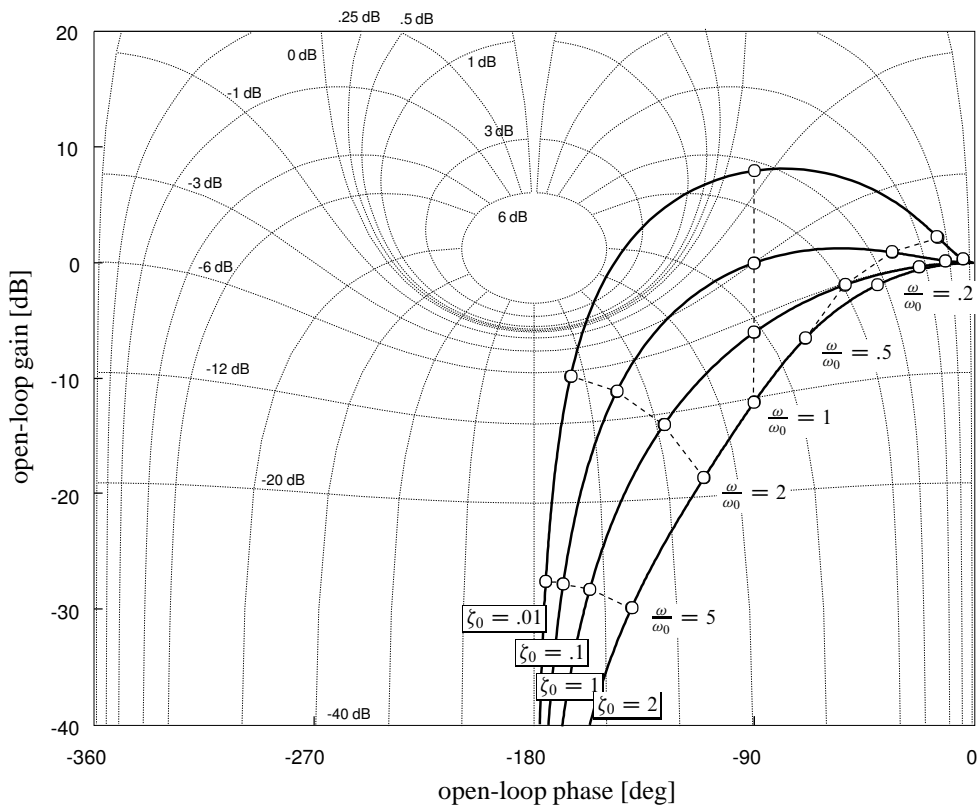


Figure 2.19: Nichols plots of the transfer function $\omega_0^2/(s^2 + 2\zeta_0\omega_0s + \omega_0^2)$ for different values of ζ_0

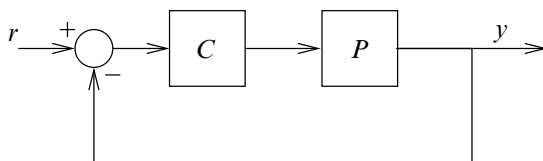


Figure 2.20: Basic feedback system

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-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.

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:

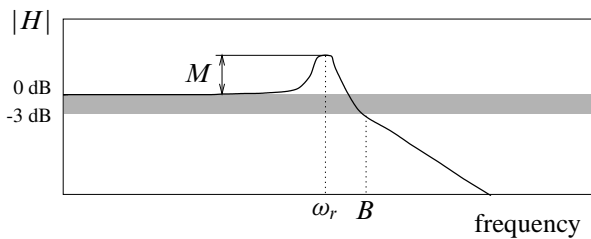


Figure 2.21: Frequency-domain performance quantities

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:

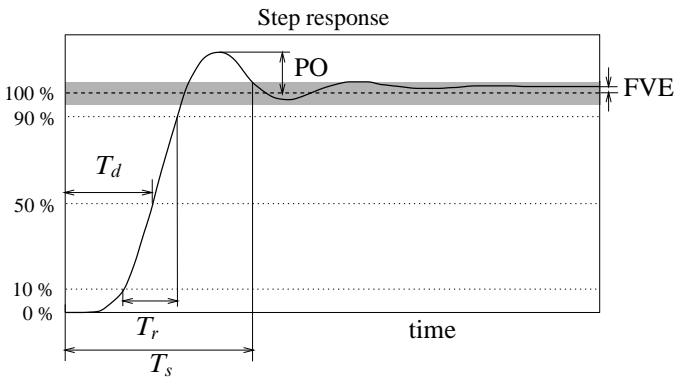


Figure 2.22: Time-domain quantities

Delay time T_d . 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 $P(j\omega)$, $\omega \in \mathbb{R}$, 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. 90) 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 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

$$C(j\omega) = \alpha \frac{1 + j\omega T}{1 + j\omega \alpha T}. \quad (2.55)$$

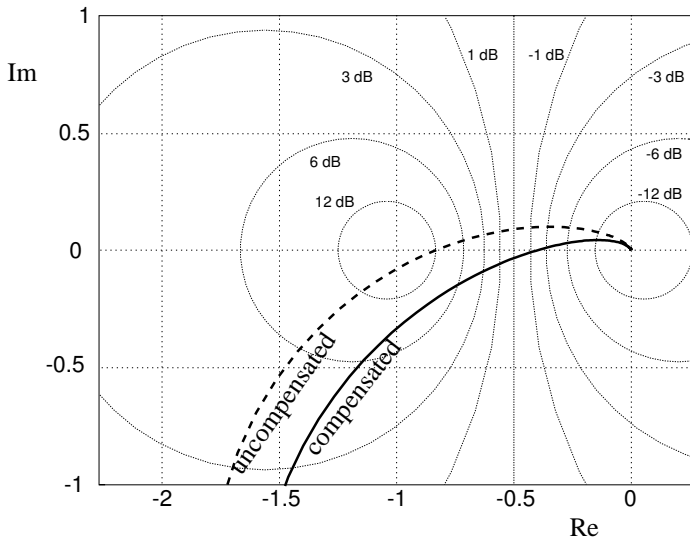


Figure 2.23: Nyquist plot of uncompensated and compensated plant

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.

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. 85).

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

$$\omega_{\text{peak}} T = 1/\sqrt{\alpha}, \quad (2.56)$$

and that the peak phase lead or lag equals

$$\phi_{\text{max}} = \arctan \frac{1}{2} \left| \frac{1}{\sqrt{\alpha}} - \sqrt{\alpha} \right|. \quad (2.57)$$

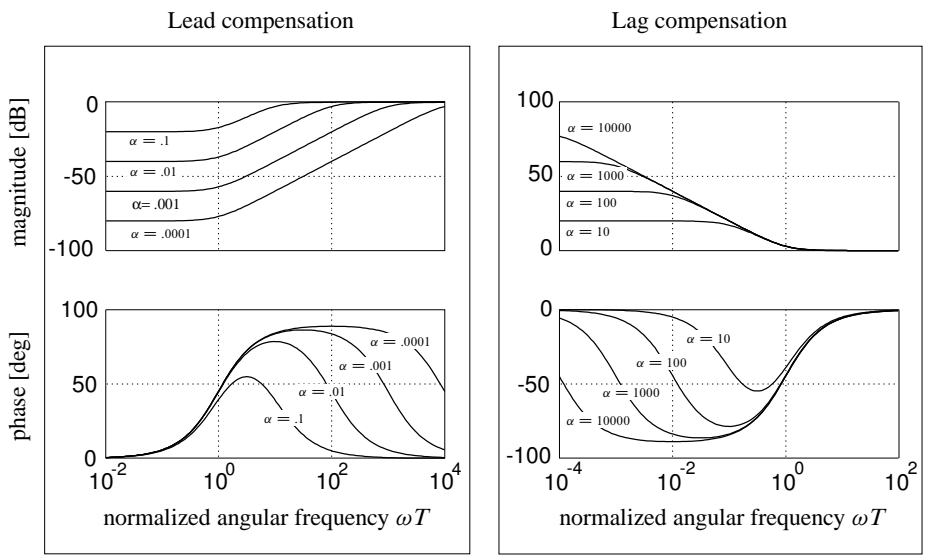


Figure 2.24: Log magnitude and phase of lead and lag compensators $C(s) = \alpha \frac{1+sT}{1+s\alpha T}$

2. Show that the width of the window over which phase lead or lag is effected is roughly $|\log_{10} \alpha|$ decades.
3. Show that the low frequency gain loss or boost is $20 |\log_{10} \alpha|$ dB.

Figure 2.25 shows plots of the peak phase lead or lag, the window width, and the low-frequency gain loss or boost. □

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

$$C(j\omega) = \alpha \frac{1 + j\omega T}{1 + j\omega \alpha T} \quad (2.58)$$

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 $\alpha \rightarrow \infty$ the compensator frequency response function becomes

$$C(j\omega) = 1 + \frac{1}{j\omega T}. \quad (2.59)$$

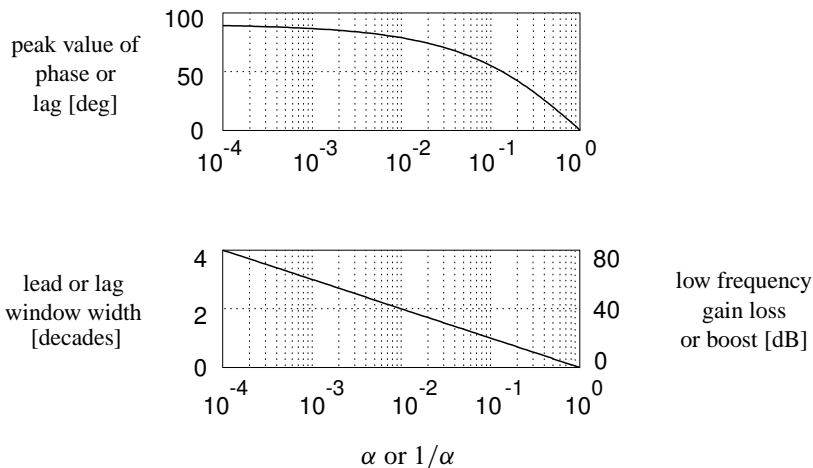


Figure 2.25: Peak phase lead or lag

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

$$C(s) = k \frac{1 + sT_i}{sT_i}. \quad (2.60)$$

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. \square

2.6.4. Lag-lead compensation

We illustrate the design of a lag-lead compensator by an example. Note the successive design steps.

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

Example 2.6.3 (Lag-lead compensator). Consider the simple second-order plant with transfer function

$$P(s) = \frac{\omega_0^2}{s^2 + 2\zeta_0\omega_0s + \omega_0^2}, \quad (2.61)$$

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

$$C_0(s) = k_0 \frac{1 + sT_i}{sT_i}. \quad (2.62)$$

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 system is unstable.

Step 2: Phase lead compensation. We stabilize the closed loop by lead compensation of the form

$$C_1(s) = k_1\alpha \frac{1 + sT}{1 + s\alpha T}. \quad (2.63)$$

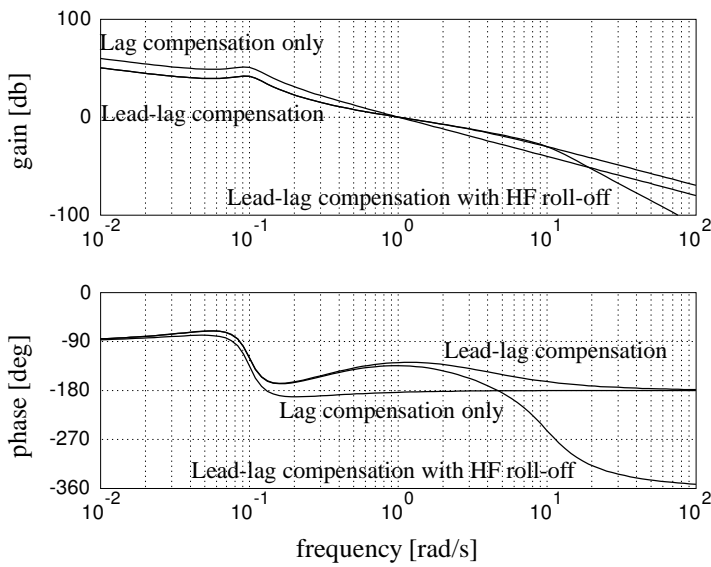


Figure 2.26: Bode diagrams of the loop gain

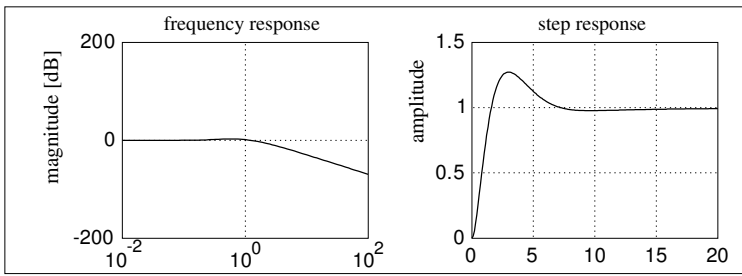
Phase advance is needed in the frequency region between, say, 0.1 and 10 [rad/s]. Inspection of Fig. 2.24 or 2.25 and some experimenting leads to the choice $\alpha = 0.1$ and $T = 3$ [rad/s]. Setting $k_1 = 3.3$ makes the crossover frequency equal to 1 [rad/s]. The resulting Bode diagram of the loop 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.

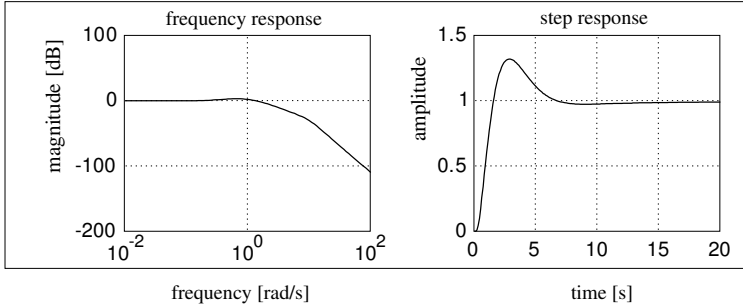
Step 3. High-frequency roll-off. For measurement noise reduction and high-frequency robustness we provide high-frequency roll-off of the compensator by including additional lag compensation of the form

$$C_2(s) = \frac{\omega_1^2}{s^2 + 2\zeta_1\omega_1s + \omega_1^2}. \quad (2.64)$$

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.



lag-lead compensation



lag-lead compensation with extra HF roll-off

Figure 2.27: Closed-loop frequency and step responses

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

$$L(s) = k \frac{(s - z_1)(s - z_2) \cdots (s - z_m)}{(s - p_1)(s - p_2) \cdots (s - p_n)}. \quad (2.65)$$

For physically realizable systems the loop transfer function L is proper, that is, $m \leq n$. The roots z_1, z_2, \dots, z_m of the numerator polynomial are the *open-loop zeros* of the system. The roots p_1, p_2, \dots, 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,

$$(s - p_1)(s - p_2) \cdots (s - p_n) + k(s - z_1)(s - z_2) \cdots (s - z_m) = 0. \quad (2.66)$$

Under the assumption that $m \leq n$ 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, \dots, p_n .
2. If $k \rightarrow \infty$ then m of the closed-loop poles approach the (finite) open-loop zeros z_1, z_2, \dots, 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

$$\alpha_i = \frac{2i + 1}{n - m} \pi \quad [\text{rad}] \quad i = 0, 1, \dots, n - m - 1. \quad (2.67)$$

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

$$s_0 = \frac{(\text{sum of open-loop poles}) - (\text{sum of open-loop zeros})}{n - m}. \quad (2.68)$$

6. As we consider real-rational functions only the loci are symmetrical about the real axis.

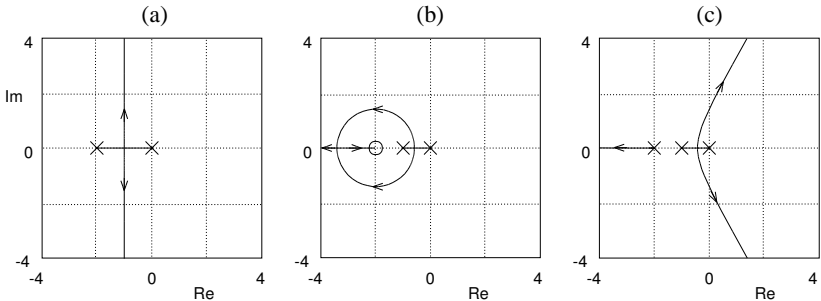


Figure 2.28: Examples of root loci:

$$(a) L(s) = \frac{k}{s(s+2)}, \quad (b) L(s) = \frac{k(s+2)}{s(s+1)}, \quad (c) L(s) = \frac{k}{s(s+1)(s+2)}$$

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 et al. (1986), Franklin et al. (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.

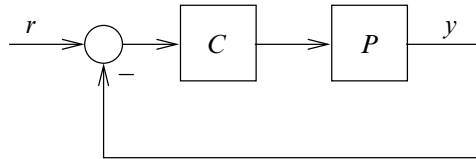


Figure 2.29: Unit feedback system

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

$$H = \frac{PC}{1 + PC}. \quad (2.69)$$

Solving for the compensator transfer function C we obtain

$$C = \frac{1}{P} \frac{H}{1 - H}. \quad (2.70)$$

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

$$L(s) = \frac{N(s)}{s^k D(s)}, \quad (2.71)$$

with N and D polynomials that have no roots at 0. It follows that

$$H(s) = \frac{L(s)}{1 + L(s)} = \frac{N(s)}{s^k D(s) + N(s)} = \frac{a_m s^m + \cdots + a_k s^k + b_{k-1} s^{k-1} + \cdots + b_0}{c_n s^n + \cdots + c_k s^k + b_{k-1} s^{k-1} + \cdots + b_0}. \quad (2.72)$$

Conversely, choosing the first k coefficients b_j in the numerator polynomial equal to that of the denominator polynomial 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

$$H(s) = \frac{b_0}{s^n + b_{n-1} s^{n-1} + \cdots + b_0}, \quad (2.73)$$

which implies a zero steady-state error for step inputs $r(t) = \mathbb{1}(t)$.

Exercise 2.8.1 (Zero steady-state error). Prove this. □

One way to choose the coefficients b_0, b_1, \dots, 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 $s := s/\omega_0$.

Another popular choice is to choose the coefficients such that the integral of the time multiplied absolute error

$$\int_0^\infty t|e(t)| dt \quad (2.74)$$

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.

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

$$P(s) = \frac{\frac{1}{T}}{s + \frac{1}{\theta}}, \quad (2.75)$$

with $T = \theta = 10$ [s]. We specify the desired closed-loop transfer function

$$H(s) = \frac{\omega_0^2}{s^2 + 1.4\omega_0 s + \omega_0^2}. \quad (2.76)$$

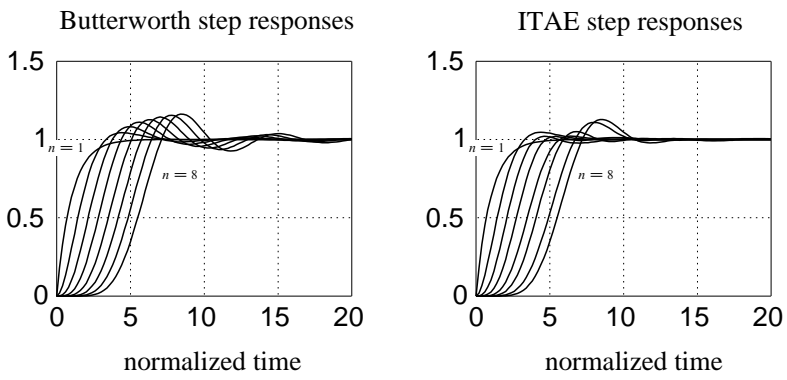


Figure 2.30: Step responses for Butterworth (left) and ITAE (right) denominator polynomials

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	$s^8 + 5.13s^7 + 13.14s^6 + 21.85s^5 + 25.69s^4 + 21.85s^3 + 13.14s^2 + 5.13s + 1$

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	$s^8 + 5.20s^7 + 12.80s^6 + 21.60s^5 + 25.75s^4 + 22.20s^3 + 13.30s^2 + 5.15s + 1$

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

$$C(s) = \frac{1}{P(s)} \frac{H(s)}{1 - H(s)} = \frac{\omega_0^2 T(s + \frac{1}{\theta})}{s(s + 1.4s)}. \quad (2.77)$$

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 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 cancellation phenomenon is typical for naïve applications of the Guillemin-Truxal method. Inspection of (2.70) shows that cancellation 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.

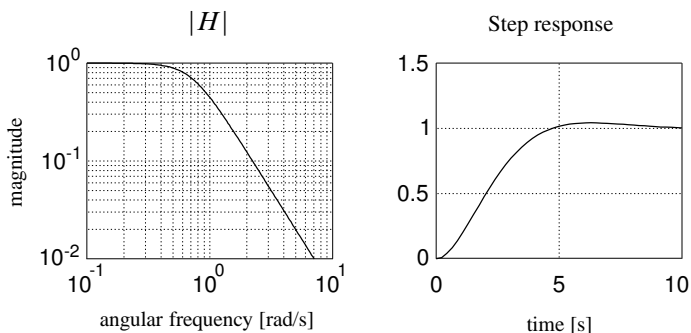


Figure 2.31: Closed-loop transfer function H and closed-loop step response of a Guillemin-Truxal design for the cruise control system

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 quantitative 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 $(-180^\circ, 0 \text{ dB})$, 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.

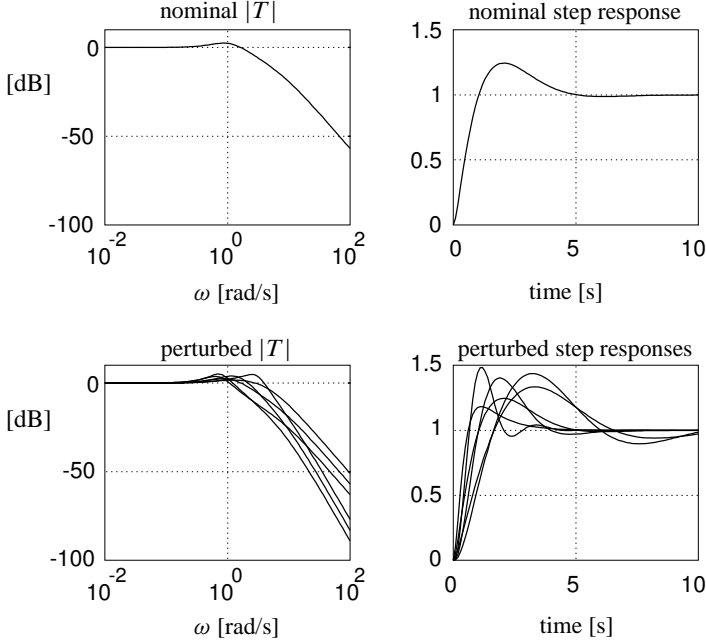


Figure 2.32: Nominal and perturbed complementary sensitivity functions and step responses for the nominal design

Example 2.9.1 (Uncertain second-order system). As an example we consider the plant with transfer function

$$P(s) = \frac{g}{s^2(1 + s\theta)}. \quad (2.78)$$

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

$$C(s) = \frac{k + T_d s}{1 + T_0 s}, \quad (2.79)$$

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.

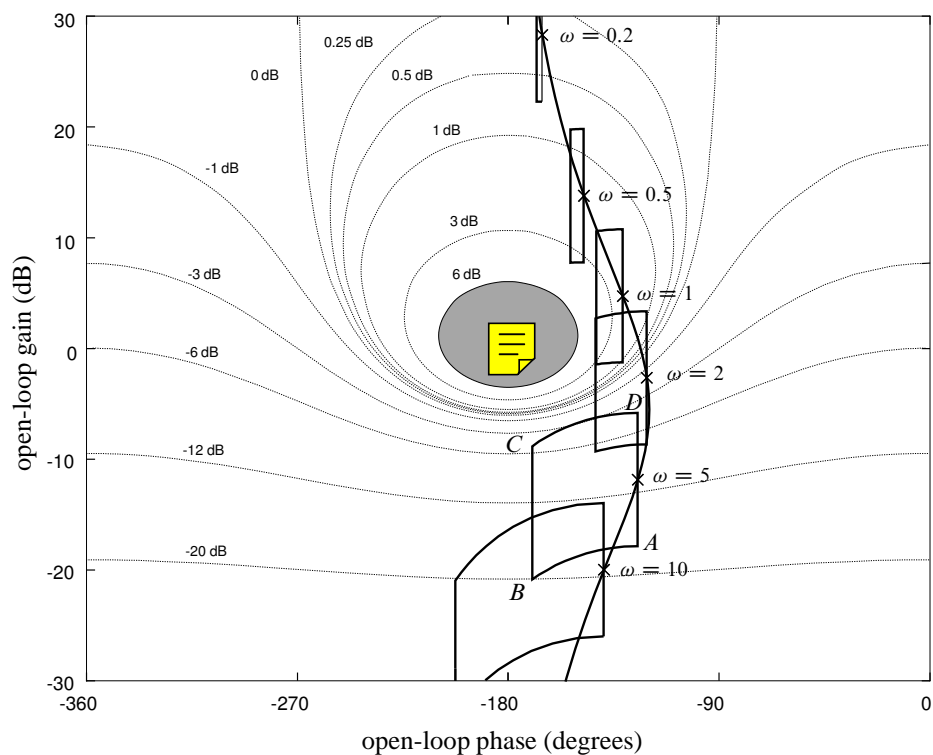


Figure 2.33: Nominal Nichols plot and uncertainty regions. (Shaded: forbidden region)

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 $L(j\omega)$ 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 $\omega = 5$, correspond to extreme values of the parameters as follows:

- 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 $|H(j\omega)|$, 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

$$\log |H| = \log |T| + \log |F|. \quad (2.80)$$

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.

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 $\omega = 2, 5$ 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 $L(j\omega)$ is on or above the corresponding performance boundary;

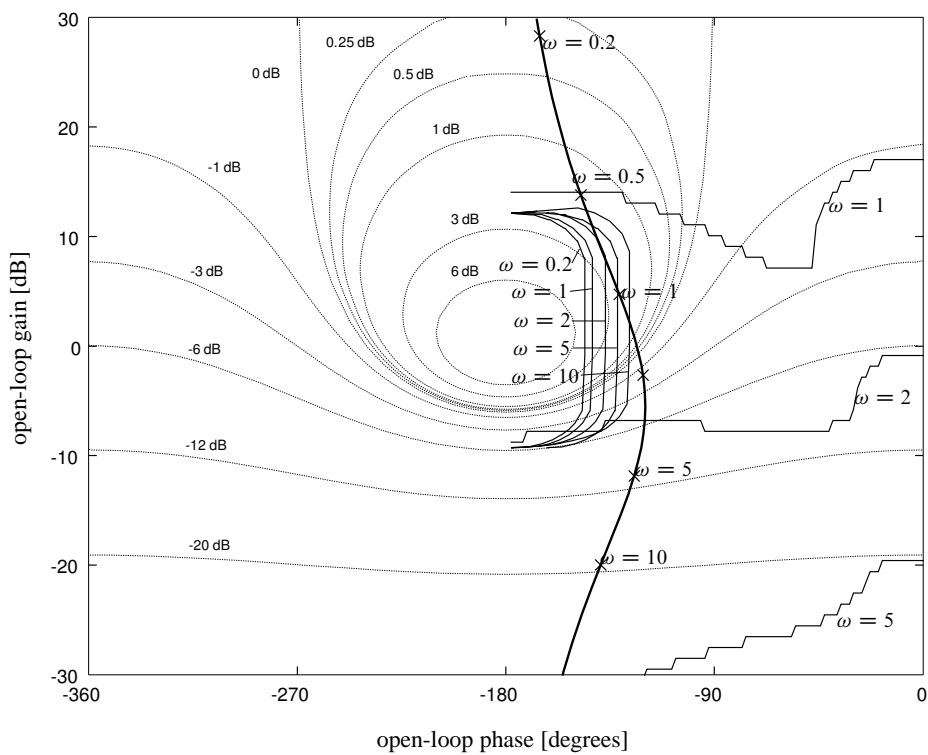


Figure 2.34: Performance and robustness boundaries

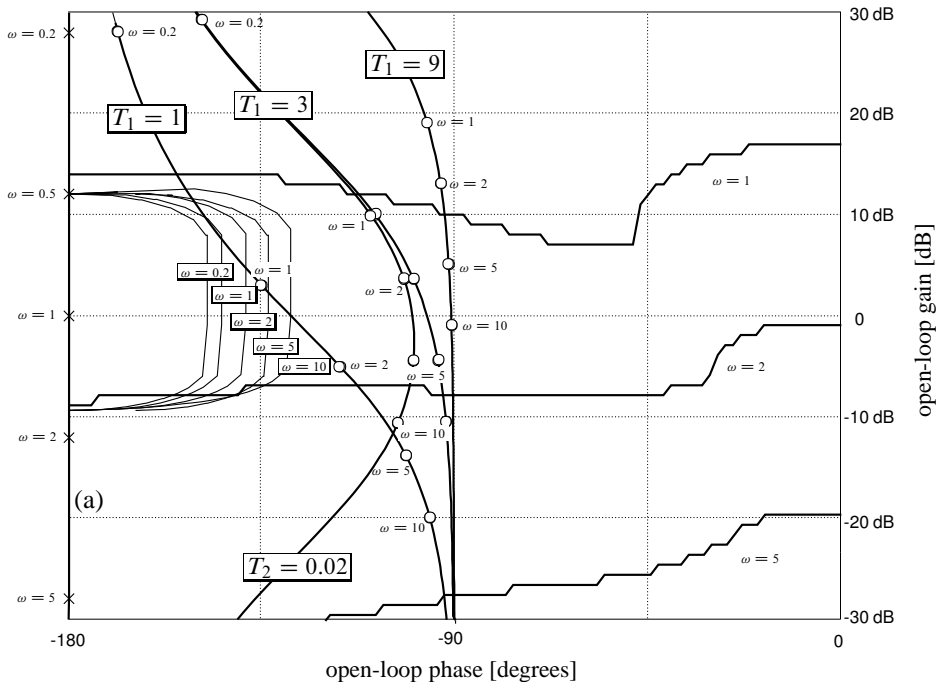


Figure 2.35: Redesign of the loop gain

2. at each critical frequency ω the corresponding loop gain $L(j\omega)$ 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 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

$$C(s) = 1 + sT_1. \quad (2.81)$$

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

$$C(s) = \frac{1 + sT_1}{1 + sT_2}. \quad (2.82)$$

The resulting loop gain for $T_2 = 0.02$ is also included in Fig. 2.35. It very nearly satisfies the requirements¹. Figure 2.36 gives plots of the resulting nominal and perturbed step responses and complementary sensitivity functions. The robustness improvement is evident. \square

¹The requirements may be completely satisfied by adding a few dB to the loop gain.

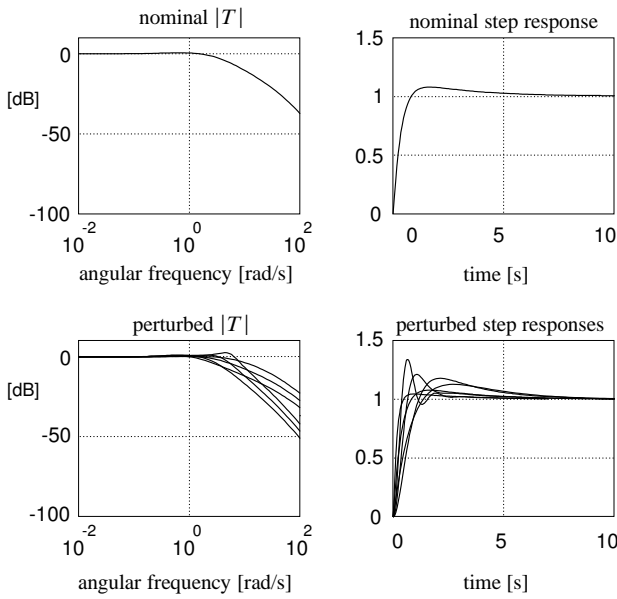


Figure 2.36: Nominal and perturbed complementary sensitivity functions and step responses of the revised design

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. □

2.9.5. Prefilter design

Once the feedback compensator 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 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

$$H = \frac{NF}{D_{cl}} F_0. \quad (2.83)$$

$P = N/D$ is the plant transfer function and $D_{cl} = DX + NY$ 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. 97) results in the closed-loop characteristic polynomial

$$D_{cl}(s) = 0.02s^3 + s^2 + 3s + 1. \quad (2.84)$$

Its roots are -0.3815 , -2.7995 , and -46.8190 . Completing the design amounts to choosing the

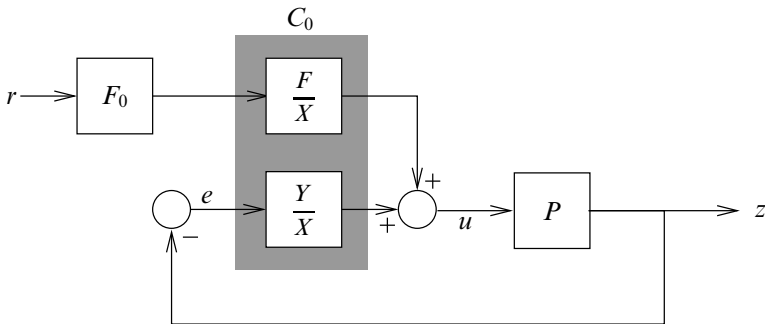


Figure 2.37: $2\frac{1}{2}$ -degree-of-freedom feedback system

correct polynomial F and transfer function F_0 to provide the necessary compensation in

$$H(s) = \frac{F(s)}{0.02(s + 0.3815)(s + 2.7995)(s + 46.8190)} F_0(s). \quad (2.85)$$

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

$$F_0(s) = \frac{\omega_0^2}{s^2 + 2\zeta_0\omega_0s + \omega_0^2}, \quad (2.86)$$

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. \square

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 nonlocal plants. Horowitz (1982) provides references and a review. Recently a MATLAB toolbox for QFT design has appeared².

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.

²Quantitative Feedback Theory Toolbox, The MathWorks Inc., Natick, MA, USA, 1995 release.

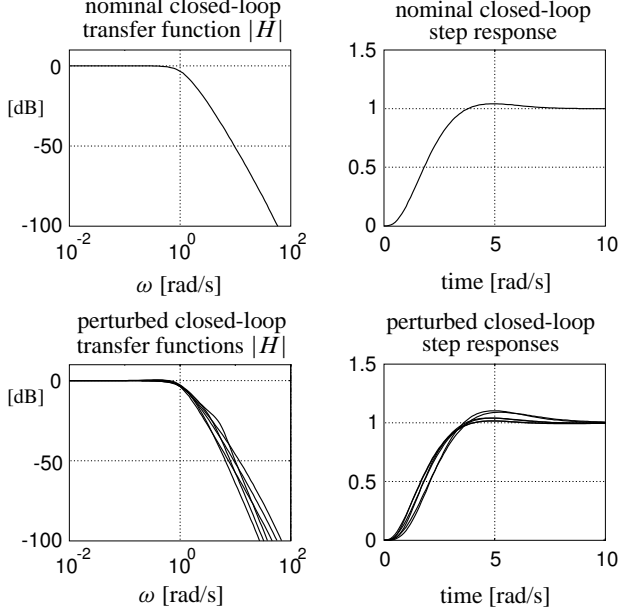


Figure 2.38: Nominal and perturbed step responses and closed-loop transfer functions of the final QFT design

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.

3. Multivariable Control System Design

Overview – Design of controllers for multivariable systems requires an assessment of structural properties of transfer matrices. The zeros and gains in multivariable systems have directions.

With norms of multivariable signals and systems it is possible to obtain bounds for gains, bandwidth and other system properties.

A possible approach to multivariable controller design is to reduce the problem to a series of single loop controller design problems. Examples are decentralized control and decoupling control.

The internal model principle applies to multivariable systems and, for example, may be used to design for multivariable integral action.

3.1. Introduction

Many complex engineering systems are equipped with several actuators that may influence their static and dynamic behavior. Commonly, in cases where some form of automatic control is required over the system, also several sensors are available to provide measurement information about important system variables that may be used for feedback control purposes. Systems with more than one actuating control input and more than one sensor output may be considered as *multivariable* systems or *multi-input-multi-output (MIMO)* systems. The control objective for multivariable systems is to obtain a desirable behavior of several output variables by simultaneously manipulating several input channels.

3.1.1. Examples of multivariable feedback systems

The following two examples discuss various phenomena that specifically occur in MIMO feedback systems and not in SISO systems, such as interaction between loops and multivariable non-minimum phase behavior.

Example 3.1.1 (Two-tank liquid flow process). Consider the flow process of Fig. 3.1. The incoming flow ϕ_1 and recycle flow ϕ_4 act as manipulable input variables to the system, and the outgoing flow ϕ_3 acts as a disturbance. The control objective is to keep the liquid levels h_1 and h_2 between acceptable limits by applying feedback from measurements of these liquid levels, while accommodating variations in the output flow ϕ_3 . As derived in Appendix 3.4, a dynamic

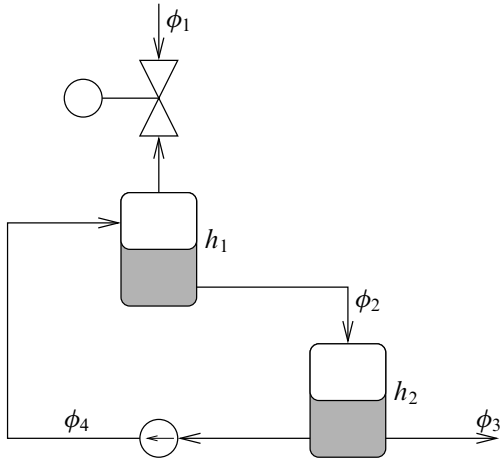


Figure 3.1: Two-tank liquid flow process with recycle

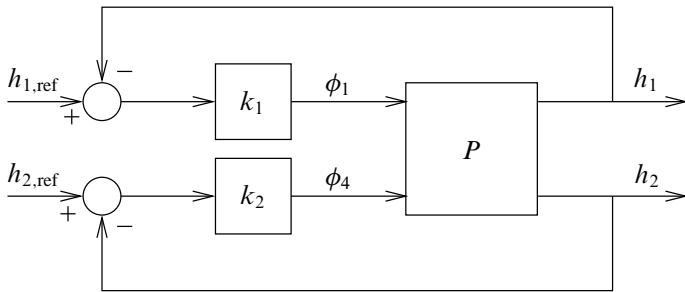


Figure 3.2: Two-loop feedback control of the two-tank liquid flow process

model, linearized about a given steady state, is

$$\begin{bmatrix} \dot{h}_1(t) \\ \dot{h}_2(t) \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} h_1(t) \\ h_2(t) \end{bmatrix} + \begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \phi_1(t) \\ \phi_4(t) \end{bmatrix} + \begin{bmatrix} 0 \\ -1 \end{bmatrix} \phi_3(t). \quad (3.1)$$

Here the h_i and ϕ_i denote the deviations from the steady state levels and flows. Laplace transformation yields

$$\begin{bmatrix} H_1(s) \\ H_2(s) \end{bmatrix} = \begin{bmatrix} s+1 & 0 \\ -1 & s \end{bmatrix}^{-1} \left(\begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \Phi_1(s) \\ \Phi_4(s) \end{bmatrix} + \begin{bmatrix} 0 \\ -1 \end{bmatrix} \Phi_3(s) \right),$$

which results in the transfer matrix relationship

$$\begin{bmatrix} H_1(s) \\ H_2(s) \end{bmatrix} = \begin{bmatrix} \frac{1}{s+1} & \frac{1}{s+1} \\ \frac{1}{s(s+1)} & -\frac{1}{s+1} \end{bmatrix} \begin{bmatrix} \Phi_1(s) \\ \Phi_4(s) \end{bmatrix} + \begin{bmatrix} 0 \\ -\frac{1}{s} \end{bmatrix} \Phi_3(s). \quad (3.2)$$

□

The example demonstrates the following typical MIMO system phenomena:

- Each of the manipulable inputs ϕ_1 and ϕ_4 affects each of the outputs to be controlled h_1 and h_2 , as a consequence the transfer matrix

$$P(s) = \begin{bmatrix} \frac{1}{s+1} & \frac{1}{s+1} \\ \frac{1}{s(s+1)} & -\frac{1}{s+1} \end{bmatrix} \quad (3.3)$$

has nonzero entries both at the diagonal and at the off-diagonal entries.

- Consequently, if we control the two output variables h_1 and h_2 using two separate control loops, it is not immediately clear which input to use to control h_1 , and which one to control h_2 . This issue is called the *input/output pairing* problem.
- A possible feedback scheme using two separate loops is shown in Fig. 3.2. Note that in this control scheme, there exists a coupling between both loops due to the non-diagonal terms in the transfer matrix (3.3). As a result, the plant transfer function in the open upper loop from ϕ_1 to h_1 , with the lower loop closed with proportional gain k_2 , depends on k_2 ,

$$P_{11}(s)|_{llc} = \frac{1}{s+1} \left(1 - \frac{k_2}{s(s+1-k_2)} \right). \quad (3.4)$$

Here *llc* indicates that the lower loop is closed. Owing to the negative steady state gain in the lower loop, negative values of k_2 stabilize the lower loop. The dynamics of $P_{11}(s)|_{llc}$ changes with k_2 :

$$\begin{aligned} k_2 = 0 : & \quad P_{11}(s)|_{llc} = \frac{1}{s+1}, \\ k_2 = -1 : & \quad P_{11}(s)|_{llc} = \frac{s+1}{s(s+2)}, \\ k_2 = -\infty : & \quad P_{11}(s)|_{llc} = \frac{1}{s}. \end{aligned}$$

The phenomenon that the loop gain in one loop also depends on the loop gain in another loop is called *interaction*. The *multiple loop* control structure used here does not explicitly acknowledge interaction phenomena. A control structure using individual loops, such as multiple loop control, is an example of a *decentralized* control structure, as opposed to a *centralized* control structure where all measured information is available for feedback in all feedback channels.

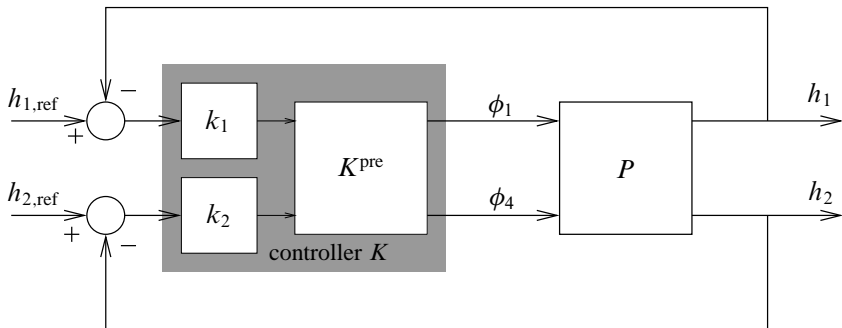


Figure 3.3: Decoupling feedback control of two-tank liquid flow process

- A different approach to multivariable control is to try to compensate for the plant interaction. In Example 3.1.1 a precompensator K^{pre} in series with the plant can be found which removes the interaction completely. The precompensator K^{pre} is to be designed such that the transfer matrix PK^{pre} is diagonal. Figure 3.3 shows the feedback structure. Suppose that the precompensator K^{pre} is given the structure

$$K^{\text{pre}}(s) = \begin{bmatrix} 1 & K_{12}^{\text{pre}}(s) \\ K_{21}^{\text{pre}}(s) & 1 \end{bmatrix} \quad (3.5)$$

then PK^{pre} is diagonal by selecting

$$K_{12}^{\text{pre}}(s) = -1, \quad K_{21}^{\text{pre}}(s) = \frac{1}{s}. \quad (3.6)$$

The compensated plant PK^{pre} for which a multi-loop feedback is to be designed as a next step, reads

$$P(s)K^{\text{pre}}(s) = \begin{bmatrix} \frac{1}{s} & 0 \\ 0 & -\frac{1}{s} \end{bmatrix}. \quad (3.7)$$

Closing both loops with proportional gains $k_1 > 0$ and $k_2 < 0$ leads to a stable system. In doing so, a multivariable controller K has been designed having transfer matrix

$$K(s) = K^{\text{pre}}(s) \begin{bmatrix} k_1 & 0 \\ 0 & k_2 \end{bmatrix} = \begin{bmatrix} k_1 & -k_2 \\ k_1/s & k_2 \end{bmatrix}. \quad (3.8)$$

This approach of removing interaction before actually closing multiple individual loops is called *decoupling control*.

The next example shows that a multivariable system may show non-minimum phase behavior.

Example 3.1.2 (Multivariable non-minimum phase behavior). Consider the two input, two output linear system

$$\begin{bmatrix} y_1(s) \\ y_2(s) \end{bmatrix} = \begin{bmatrix} \frac{1}{s+1} & \frac{2}{s+3} \\ \frac{1}{s+1} & \frac{1}{s+1} \end{bmatrix} \begin{bmatrix} u_1(s) \\ u_2(s) \end{bmatrix}. \quad (3.9)$$

If the lower loop is closed with constant gain, $u_2(s) = -k_2 y_2(s)$, then for high gain values ($k_2 \rightarrow \infty$) the lower feedback loop is stable but has a zero in the right-half plane,

$$y_1(s) = \left(\frac{1}{s+1} - \frac{2}{s+3} \right) u_1(s) = -\frac{s-1}{(s+1)(s+3)} u_1(s). \quad (3.10)$$

Thus under high gain feedback of the lower loop, the upper part of the system exhibits non-minimum phase behavior. Conversely if the upper loop is closed under high gain feedback, $u_1 = -k_1 y_1$ with $k_1 \rightarrow \infty$, then

$$y_2(s) = -\frac{s-1}{(s+1)(s+3)}u_2(s). \quad (3.11)$$

Apparently, the non-minimum phase behavior of the system is not connected to one particular input-output relation, but shows up in both relations. One loop can be closed with high gains under stability of the loop, and the other loop is then restricted to have limited gain due to the non-minimum phase behavior. Analysis of the transfer matrix

$$P(s) = \begin{bmatrix} \frac{1}{s+1} & \frac{2}{s+3} \\ \frac{1}{s+1} & \frac{1}{s+1} \end{bmatrix} = \frac{1}{(s+1)(s+3)} \begin{bmatrix} s+3 & 2s+2 \\ s+3 & s+3 \end{bmatrix} \quad (3.12)$$

shows that it loses rank at $s = 1$. In the next section it will be shown that $s = 1$ is an unstable *transmission zero* of the multivariable system and this limits the closed loop behavior irrespective of the controller used. Consider the method of decoupling precompensation. Express the precompensator as

$$K^{\text{pre}}(s) = \begin{bmatrix} K_{11}(s) & K_{12}(s) \\ K_{21}(s) & K_{22}(s) \end{bmatrix}. \quad (3.13)$$

Decoupling means that PK^{pre} is diagonal. Inserting the matrix (3.12) into PK^{pre} , a solution for a decoupling K^{pre} is

$$K_{11}(s) = 1, \quad K_{22}(s) = 1, \quad K_{12}(s) = -2\frac{s+1}{s+3}, \quad K_{21}(s) = -1. \quad (3.14)$$

Then the compensated system is

$$P(s)K^{\text{pre}}(s) = \begin{bmatrix} \frac{1}{s+1} & \frac{2}{s+3} \\ \frac{1}{s+1} & \frac{1}{s+1} \end{bmatrix} \begin{bmatrix} 1 & -2\frac{s+1}{s+3} \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} -\frac{s-1}{(s+1)(s+3)} & 0 \\ 0 & -\frac{s-1}{(s+1)(s+3)} \end{bmatrix}. \quad (3.15)$$

The requirement of decoupling has introduced a right-half plane zero in each of the loops, so that either loop now only can have a restricted gain in the feedback loop. \square

The example suggests that the occurrence of non-minimum phase phenomena is a property of the transfer matrix and exhibits itself in a matrix sense, not connected to a particular input-output pair. It shows that a zero of a multivariable system has no relationship with the possible zeros of individual entries of the transfer matrix.

In the example the input-output pairing has been the natural one: output i is connected by a feedback loop to input i . This is however quite an arbitrary choice, as it is the result of our model formulation that determines which inputs and which outputs are ordered as one, two and so on. Thus the selection of the most useful input-output pairs is a *non-trivial* issue in *multiloop control* or *decentralized control*, i.e. in control configurations where one has individual loops as in the example. A classical approach towards dealing with multivariable systems is to bring a multivariable system to a structure that is a collection of one input, one output control problems. This approach of *decoupling control* may have some advantages in certain practical situations, and was thought to lead to a simpler design approach. However, as the above example showed, decoupling may introduce additional restrictions regarding the feedback properties of the system.

3.1.2. Survey of developments in multivariable control

The first papers on multivariable control systems appeared in the fifties and considered aspects of noninteracting control. In the sixties, the work of [Rosenbrock \(1970\)](#) considered matrix techniques to study questions of rational and polynomial representation of multivariable systems. The polynomial representation was also studied by [Wolovich \(1974\)](#). The books by [Kailath \(1980\)](#) and [Vardulakis \(1991\)](#) provide a broad overview. The use of Nyquist techniques for multivariable control design was developed by [Rosenbrock \(1974a\)](#). The generalization of the Nyquist criterion and of root locus techniques to the multivariable case can be found in the work of [Postlethwaite and MacFarlane \(1979\)](#). The geometric approach to multivariable state-space control design is contained in the classical book by [Wonham \(1979\)](#) and in the book by [Basile and Marro \(1992\)](#). A survey of classical design methods for multivariable control systems can be found in [Korn and Wilfert \(1982\)](#), [Lunze \(1988\)](#) and in the two books by [Tolle \(1983\)](#), 1985. Modern approaches to frequency domain methods can be found in [Raisch \(1993\)](#), [Maciejowski \(1989\)](#), and [Skogestad and Postlethwaite \(1995\)](#). Interaction phenomena in multivariable process control systems are discussed in terms of a process control formulation in [McAvoy \(1983\)](#). A modern, process-control oriented approach to multivariable control is presented in [Morari and Zafriou \(1989\)](#). The numerical properties of several computational algorithms relevant to the area of multivariable control design are discussed in [Svaricek \(1995\)](#).

3.2. Poles and zeros of multivariable systems

In this section, a number of structural properties of multivariable systems are discussed. These properties are important for the understanding of the behavior of the system under feedback. The systems are analyzed both in the time domain and frequency domain.

3.2.1. Polynomial and rational matrices

Let P be a proper real-rational transfer matrix. *Real-rational* means that every entry P_{ij} of P is a ratio of two polynomials having real coefficients. Any rational matrix P may be written as a fraction

$$P = \frac{1}{d}N$$

where N is a *polynomial matrix*¹ and d is the monic least common multiple of all denominator polynomials of the entries of P . Polynomial matrices will be studied first. Many results about polynomial matrices may be found in, e.g., the books by [MacDuffee \(1956\)](#), [Gohberg et al. \(1982\)](#) and [Kailath \(1980\)](#).

In the scalar case a polynomial has no zeros if and only if it is a nonzero constant, or, to put it differently, if and only if its inverse is polynomial as well. The matrix generalization is as follows.

Definition 3.2.1 (Unimodular polynomial matrix). A polynomial matrix is *unimodular* if it square and its inverse exists and is a polynomial matrix. □

It may be shown that a square polynomial matrix U is unimodular if and only if $\det U$ is a nonzero constant. Unimodular matrices are considered having no zeros and, hence, multiplication by unimodular matrices does not affect the zeros.

¹That is, a matrix whose entries are polynomials.

Summary 3.2.2 (Smith form of a polynomial matrix). For every polynomial matrix N there exist unimodular U and V such that

$$UNV = \underbrace{\begin{bmatrix} \varepsilon'_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \varepsilon'_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \ddots & 0 & 0 & 0 \\ 0 & 0 & 0 & \varepsilon'_r & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}}_S \quad (3.16)$$

where ε'_i , ($i = 1, \dots, r$) are monic polynomials with the property that $\varepsilon_i/\varepsilon_{i-1}$ is polynomial. The matrix S is known as the *Smith form* of N and the polynomials ε'_i the *invariant polynomials* of N .

In this case r is the *normal rank* of the polynomial matrix N , and the *zeros* of N are defined as the zeros of one or more of its invariant polynomials. \square

The Smith form S of N may be obtained by applying to N a sequence of *elementary row and column operations*, which are:

- multiply a row/column by a nonzero constant;
- interchange two rows or two columns;
- add a row/column multiplied by a polynomial to another row/column.

Example 3.2.3 (Elementary row/column operations). The following sequence of elementary operations brings the polynomial matrix

$$N(s) = \begin{bmatrix} s+1 & s-1 \\ s+2 & s-2 \end{bmatrix}$$

to diagonal Smith-form:

$$\begin{bmatrix} s+1 & s-1 \\ s+2 & s-2 \end{bmatrix} \xrightarrow{(1)} \begin{bmatrix} s+1 & s-1 \\ 1 & -1 \end{bmatrix} \xrightarrow{(2)} \begin{bmatrix} s+1 & 2s \\ 1 & 0 \end{bmatrix} \xrightarrow{(3)} \begin{bmatrix} 0 & 2s \\ 1 & 0 \end{bmatrix} \xrightarrow{(4)} \begin{bmatrix} 1 & 0 \\ 0 & s \end{bmatrix}.$$

Here, in step (1) we subtracted row 1 from row 2; in step (2) we added the first column to the second column; in step (3), $s+1$ times the second row was subtracted from row 1. Finally, in step (4), we interchanged the rows and then divided the (new) second row by a factor 2.

Elementary *row* operations correspond the *premultiplication* by unimodular matrices, and elementary *column* operations correspond the *postmultiplication* by unimodular matrices. For the above four elementary operations these are

- (1) premultiply by $U_{L1}(s) := \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}$;
- (2) postmultiply by $V_{R2}(s) := \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$;
- (3) premultiply by $U_{L3}(s) := \begin{bmatrix} 1 & -(s+1) \\ 0 & 1 \end{bmatrix}$;

(4) premultiply by $U_{L4}(s) := \begin{bmatrix} 0 & 1 \\ 1/2 & 0 \end{bmatrix}$.

Instead of applying sequentially multiplication on N we may also first combine the sequence of unimodular matrices into two unimodular matrices $U = U_{L4}U_{L3}U_{L1}$ and $V = V_{R2}$, and then apply the U and V ,

$$\underbrace{\begin{bmatrix} -1 & 1 \\ 1+s/2 & -1/2-s/2 \end{bmatrix}}_{U(s) = U_{L4}(s)U_{L3}(s)U_{L1}(s)} \underbrace{\begin{bmatrix} s+1 & s-1 \\ s+2 & s-2 \end{bmatrix}}_{N(s)} \underbrace{\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}}_{V(s) = V_{R2}(s)} = \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & s \end{bmatrix}}_{S(s)}.$$

This clearly shows that S is the Smith-form of N . The polynomial matrix N has rank 2 and has one zero at $s = 0$. \square

Now consider a rational matrix P . We write P as

$$P = \frac{1}{d}N \quad (3.17)$$

where N is a polynomial matrix and d a scalar polynomial. We immediately get a generalization of the Smith form.

Summary 3.2.4 (Smith-McMillan form of a rational matrix). For every rational matrix P there exist unimodular polynomial matrices U and V such that

$$UPV = \underbrace{\begin{bmatrix} \frac{\varepsilon_1}{\psi_1} & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{\varepsilon_2}{\psi_2} & 0 & 0 & 0 & 0 \\ 0 & 0 & \ddots & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{\varepsilon_r}{\psi_r} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}}_M \quad (3.18)$$

where the ε_i and ψ_i are coprime such that

$$\frac{\varepsilon_i}{\psi_i} = \frac{\varepsilon'_i}{d} \quad (3.19)$$

and the ε'_i , ($i = 1, \dots, r$) are the invariant polynomials of $N = dP$. The matrix M is known as the *Smith-McMillan form* of P .

In this case r is the *normal rank* of the rational matrix P . The *transmission zeros* of P are defined as the zeros of $\prod_{i=1}^r \varepsilon_i$ and the *poles* are the zeros of $\prod_{i=1}^r \psi_i$. \square

Definitions for other notions of zeros may be found in [Rosenbrock \(1970\)](#), [Rosenbrock \(1973\)](#) and [Rosenbrock \(1974b\)](#), and in [Schrader and Sain \(1989\)](#).

Example 3.2.5 (Smith-McMillan form). Consider the transfer matrix

$$P(s) = \begin{bmatrix} \frac{1}{s} & \frac{1}{s(s+1)} \\ \frac{1}{s+1} & \frac{1}{s+1} \end{bmatrix} = \frac{1}{s(s+1)} \begin{bmatrix} s+1 & 1 \\ s & s \end{bmatrix} = \frac{1}{d(s)}N(s).$$

The Smith form of N is obtained by the following sequence of elementary operations:

$$\begin{bmatrix} s+1 & 1 \\ s & s \end{bmatrix} \Rightarrow \begin{bmatrix} s+1 & 1 \\ -s^2 & 0 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & s+1 \\ 0 & s^2 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 0 \\ 0 & s^2 \end{bmatrix}$$

Division of each diagonal entry by d yields the Smith-McMillan form

$$\begin{bmatrix} \frac{1}{s(s+1)} & 0 \\ 0 & \frac{s}{s+1} \end{bmatrix}.$$

The set of transmission zeros is $\{0\}$, the set of poles is $\{-1, -1, 0\}$. This shows that transmission zeros of a multivariable transfer matrix may coincide with poles without being canceled if they occur in different diagonal entries of the Smith-McMillan form. In the determinant they do cancel. Therefore from the determinant

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

we may not always uncover *all* poles and transmission zeros of P . (Generally P need not be square so its determinant may not even exist.) \square

An $s_0 \in \mathbb{C}$ is a *pole* of P if and only if it is a pole of one or more entries P_{ij} of P . An $s_0 \in \mathbb{C}$ that is not a pole of P is a transmission zero of P if and only if the rank of $P(s_0)$ is strictly less than the normal rank r as defined by the Smith-McMillan form of P . For square invertible matrices P there further holds that s_0 is a transmission zero of P if and only if it is pole of P^{-1} . For example

$$P(s) = \begin{bmatrix} 1 & 1/s \\ 0 & 1 \end{bmatrix}$$

has a transmission zero at $s = 0$ — even though $\det P(s) = 1$ — because $P^{-1}(s) = \begin{bmatrix} 1 & -1/s \\ 0 & 1 \end{bmatrix}$ has a pole at $s = 0$.

3.2.2. Squaring down

If the normal rank of an $n_y \times n_u$ plant P is r then at most r entries of the output $y = Pu$ can be given independent values by manipulating the input u . This generally means that r entries of the output are enough for feedback control. Let K be any $n_u \times n_y$ controller transfer matrix, then the sensitivity matrix S and complementary sensitivity matrix T as previously defined are for the multivariable case (see Exercise 1.5.4)

$$S = (I + PK)^{-1}, \quad T = (I + PK)^{-1} PK. \quad (3.20)$$

If $r < n_y$, then the $n_y \times n_y$ loop gain PK is singular, so that S has one or more eigenvalues $\lambda = 1$ for any s in the complex plane, in particular everywhere on the imaginary axis. In such cases $\|S(j\omega)\|$ in whatever norm does not converge to zero as $\omega \rightarrow 0$. This indicates an undesired situation which is to be prevented by ascertaining the rank of P to equal the number of the be controller outputs. This must be realized by proper selection and application of actuators and sensors in the feedback control system.

Feedback around a nonsquare P can only occur in conjunction with a compensator K which makes the series connection PK square, as required by unity feedback. Such controllers K are said to *square down* the plant P . We investigate down squaring.

Example 3.2.6 (Squaring down). Consider

$$P(s) = \begin{bmatrix} \frac{1}{s} & \frac{1}{s^2} \end{bmatrix}.$$

Its Smith-McMillan form is $M(s) = [1/s^2 \ 0]$. The plant has no transmission zeros and has a double pole at $s = 0$. As pre-compensator we propose

$$K_0(s) = \begin{bmatrix} 1 \\ a \end{bmatrix}.$$

This results in the squared-down system

$$P(s)K_0(s) = \frac{s+a}{s^2}.$$

Apparently squaring down may introduces transmission zeros. In this example the choice $a > 0$ creates a zero in the left-half plane, so that subsequent high gain feedback can be applied, allowing a stable closed-loop system with high loop gain. \square

In the example, P does not have transmission zeros. A general 1×2 plant $P(s) = [a(s) \ b(s)]$ has transmission zeros if and only if a and b have common zeros. If a and b are in some sense randomly chosen then it is very unlikely that a and b have common zeros. Generally it holds that nonsquare plants have no transmission zeros assuming the entries of P_{ij} are in some sense uncorrelated with other entries. However, many physical systems bear structure in the entries of their transfer matrix, so that these cannot be considered as belonging to the class of generic systems. Many physically existing nonsquare systems actually turn out to possess transmission zeros.

The theory of transmission zero placement by squaring down is presently not complete, although a number of results are available in the literature. [Sain and Schrader \(1990\)](#) describe the general problem area, and results regarding the squaring problem are described in [Horowitz and Gera \(1979\)](#), [Karcianas and Giannakopoulos \(1989\)](#), [Le and Safonov \(1992\)](#), [Sebakhy et al. \(1986\)](#), [Stoorvogel and Ludlage \(1994\)](#), and [Shaked \(1976\)](#).

Example 3.2.7 (Squaring down). Consider the transfer matrix P and its Smith-McMillan form

$$P(s) = \begin{bmatrix} \frac{1}{s} & \frac{1}{s(s^2-1)} \\ 1 & \frac{1}{s^2-1} \\ 0 & \frac{s}{s^2-1} \end{bmatrix} = \frac{1}{s(s^2-1)} \begin{bmatrix} s^2-1 & 1 \\ s(s^2-1) & s \\ 0 & s^2 \end{bmatrix} \quad (3.21)$$

$$= \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ s & 0 & 1 \\ s^2 & -1 & 0 \end{bmatrix}}_{U^{-1}(s)} \underbrace{\begin{bmatrix} \frac{1}{s(s^2-1)} & 0 \\ 0 & s \\ 0 & 0 \end{bmatrix}}_{M(s)} \underbrace{\begin{bmatrix} s^2-1 & 1 \\ 1 & 0 \end{bmatrix}}_{V^{-1}(s)}. \quad (3.22)$$

There is one transmission zero $\{0\}$ and the set of poles is $\{-1, 1, 0\}$. The postcompensator K_0 is now 2×3 , which we parameterize as

$$K_0 = \begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix}.$$

We obtain the newly formed set of transmission zeros as the set of zeros of

$$\det \begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix} \begin{bmatrix} 1 & 0 \\ s & 0 \\ s^2 & -1 \end{bmatrix} = (cd - af) + (ce - bf)s,$$

where the second matrix in this expression are the first two columns of U^{-1} . Thus one single transmission zero can be assigned to a desired location. Choose this zero at $s = -1$. This leads for example to the choice $a = 0$, $b = 0$, $c = 1$, $d = 1$, $e = 1$, $f = 0$,

$$K_0(s) := \begin{bmatrix} 0 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$$

and the squared-down system

$$K_0(s)P(s) = \begin{bmatrix} 0 & \frac{s}{s^2-1} \\ \frac{s+1}{s} & \frac{1}{s(s-1)} \end{bmatrix} = \frac{1}{s(s^2-1)} \begin{bmatrix} 0 & s^2 \\ (s+1)(s^2-1) & s+1 \end{bmatrix}.$$

This squared-down matrix may be shown to have Smith-McMillan form

$$\begin{bmatrix} \frac{1}{s(s^2-1)} & 0 \\ 0 & s(s+1) \end{bmatrix}$$

and consequently the set of transmission zeros is $\{0, -1\}$, the set of poles remain unchanged as $\{-1, 1, 0\}$. These values are as expected, i.e., the zero at $s = 0$ has been retained, and the new zero at -1 has been formed. Note that $s = 0$ is both a zero and a pole. \square

3.2.3. Transmission zeros of state-space realizations

Any proper $n_y \times n_u$ rational matrix P has a *state-space realization*

$$P(s) = C(sI_n - A)^{-1}B + D, \quad A \in \mathbb{R}^{n \times n}, B \in \mathbb{R}^{n \times n_u}, C \in \mathbb{R}^{n_y \times n}, D \in \mathbb{R}^{n_y \times n_u}.$$

Realizations are commonly denoted as a quadruple (A, B, C, D) . There are many realizations (A, B, C, D) that define the same transfer matrix P . For one, the *order* n is not fixed, A realization of order n of P is *minimal* if no realization of P exists that has a lower order. Realizations are minimal if and only if the order n equals the degree of the polynomial $\psi_1\psi_2 \dots \psi_r$ formed from the denominator polynomials in (3.18), see [Rosenbrock \(1970\)](#). The poles of P then equal the eigenvalues of A , which, incidentally, shows that computation of poles is a standard eigenvalue problem. Numerical computation of the transmission zeros may be more difficult. In some special cases, computation can be done by standard operations, in other cases specialized numerical algorithms have to be used.

Lemma 3.2.8 (Transmission zeros of a minimal state-space system). Let (A, B, C, D) be a minimal state-space realization of order n of a proper real-rational $n_y \times n_u$ transfer matrix P of rank r . Then the transmission zeros $s_0 \in \mathbb{C}$ of P as defined in Summary 3.2.4 are the zeros s_0 of the polynomial matrix

$$\begin{bmatrix} A - s_0 I & B \\ C & D \end{bmatrix}. \quad (3.23)$$

That is, s_0 is transmission zero if and only if $\text{rank} \begin{bmatrix} A - s_0 I & B \\ C & D \end{bmatrix} < n + r$. \square

The proof of this result may be found in Appendix 3.4. Several further properties may be derived from this result.

Summary 3.2.9 (Invariance of transmission zeros). The zeros s_0 of (3.23) are invariant under the following operations:

- nonsingular *input space* transformations (T_2) and *output space* transformations (T_3):

$$\begin{bmatrix} A - sI & B \\ C & D \end{bmatrix} \Rightarrow \begin{bmatrix} A - sI & BT_2 \\ T_3C & T_3DT_2 \end{bmatrix}$$

- nonsingular *state space* transformations (T_1):

$$\begin{bmatrix} A - sI & B \\ C & D \end{bmatrix} \Rightarrow \begin{bmatrix} T_1AT_1^{-1} - sI & T_1B \\ CT_1^{-1} & D \end{bmatrix}$$

- *Static output feedback* operations, i.e. for the system $\dot{x} = Ax + Bu$, $y = Cx + Du$, we apply the feedback law $u = -Ky + v$ where v acts as new input. Assuming that $\det(I + DK) \neq 0$, this involves the transformation:

$$\begin{bmatrix} A - sI & B \\ C & D \end{bmatrix} \Rightarrow \begin{bmatrix} A - BK(I + DK)^{-1}C - sI & B(I + KD)^{-1} \\ (I + DK)^{-1}C & (I + DK)^{-1}D \end{bmatrix}$$

- *State feedback* (F) and *output injection* (L) operations,

$$\begin{bmatrix} A - sI & B \\ C & D \end{bmatrix} \Rightarrow \begin{bmatrix} A - BF - sI & B \\ C - DF & D \end{bmatrix} \Rightarrow \begin{bmatrix} A - BF - LC + LDF - sI & B - LD \\ C - DF & D \end{bmatrix} \quad \square$$

The transmission zeros have a clear interpretation as that of *blocking* certain exponential inputs, [Desoer and Schulman \(1974\)](#), [MacFarlane and Karcianias \(1976\)](#). Under constant high gain output feedback $u = Ky$ the finite closed loop poles generally approach the transmission zeros of the transfer matrix P . In this respect the transmission zeros of a MIMO system play a similar role in determining performance limitations as do the zeros in the SISO case. See e.g. [Francis and Wonham \(1975\)](#). If P is square and invertible, then the transmission zeros are the zeros of the polynomial

$$\det \begin{bmatrix} A - sI & B \\ C & D \end{bmatrix} \quad (3.24)$$

If in addition D is invertible, then

$$\det \begin{bmatrix} A - sI & B \\ C & D \end{bmatrix} \begin{bmatrix} I & 0 \\ -D^{-1}C & I \end{bmatrix} = \det \begin{bmatrix} (A - BD^{-1}C) - sI & B \\ 0 & D \end{bmatrix}. \quad (3.25)$$

So then the transmission zeros are the eigenvalues of $A - BD^{-1}C$. If D is not invertible then computation of transmission zeros is less straightforward. We may have to resort to computation of the Smith-McMillan form, but preferably we use methods based on state-space realizations.

Example 3.2.10 (Transmission zeros of the liquid flow system). Consider the Example 3.1.1 with plant transfer matrix

$$P(s) = \begin{bmatrix} \frac{1}{s+1} & \frac{1}{s+1} \\ \frac{1}{s(s+1)} & -\frac{1}{s+1} \end{bmatrix} = \frac{1}{s(s+1)} \begin{bmatrix} s & s \\ 1 & -s \end{bmatrix}.$$

Elementary row and column operations successively applied to the polynomial part lead to the Smith form

$$\begin{bmatrix} s & s \\ 1 & -s \end{bmatrix} \xrightarrow{(1)} \begin{bmatrix} s+1 & 0 \\ 1 & -s \end{bmatrix} \xrightarrow{(2)} \begin{bmatrix} s+1 & s(s+1) \\ 1 & 0 \end{bmatrix} \xrightarrow{(3)} \begin{bmatrix} 1 & 0 \\ 0 & s(s+1) \end{bmatrix}.$$

The Smith-McMillan form hence is

$$M(s) = \frac{1}{s(s+1)} \begin{bmatrix} 1 & 0 \\ 0 & s(s+1) \end{bmatrix} = \begin{bmatrix} \frac{1}{s(s+1)} & 0 \\ 0 & 1 \end{bmatrix}.$$

The liquid flow system has no zeros but has two poles $\{0, -1\}$.

We may compute the poles and zeros also from the state space representation of P ,

$$P(s) = C(sI - A)^{-1}B + D, \quad \left[\begin{array}{c|c} A - sI & B \\ \hline C & D \end{array} \right] = \left[\begin{array}{cc|cc} -1-s & 0 & 1 & 1 \\ 1 & -s & 0 & -1 \\ \hline -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{array} \right].$$

The realization is minimal because both B and C are square and nonsingular. The poles are therefore the eigenvalues of A , which indeed are $\{0, -1\}$ as before. There are no transmission zeros because $\det \begin{bmatrix} A-sI & B \\ C & D \end{bmatrix}$ is a nonzero constant (verify this). \square

Example 3.2.11 (Transmission zeros via state space and transfer matrix representation).

Consider the system with state space realization (A, B, C, D) , with

$$A = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad C = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad D = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

The system's transfer matrix P equals

$$P(s) = C(sI - A)^{-1}B + D = \begin{bmatrix} 1/s & 1/s^2 \\ 1/s^2 & 1/s^3 \end{bmatrix} = \frac{1}{s^3} \begin{bmatrix} s^2 & s \\ s & 1 \end{bmatrix}.$$

Elementary operations on the numerator polynomial matrix results in

$$\begin{bmatrix} s^2 & s \\ s & 1 \end{bmatrix} \Rightarrow \begin{bmatrix} 0 & s \\ 0 & 1 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

so the Smith-McMillan form of P is

$$M(s) = \frac{1}{s^3} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1/s^3 & 0 \\ 0 & 0 \end{bmatrix}.$$

The system therefore has three poles, all at zero: $\{0, 0, 0\}$. As the matrix $A \in \mathbb{R}^{3 \times 3}$ has an equal number of eigenvalues, it must be that the realization is minimal and that its eigenvalues coincide with the poles of the system. From the Smith-McMillan form we see that there are no transmission zeros. This may also be verified from the matrix pencil

$$\left[\begin{array}{c|c} A - sI & B \\ \hline C & D \end{array} \right] = \left[\begin{array}{ccc|cc} -s & 0 & 0 & 0 & 1 \\ 1 & -s & 0 & 1 & 0 \\ 0 & 1 & -s & 0 & 0 \\ \hline 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \end{array} \right].$$

Elementary row and column operations do not affect the rank of this matrix pencil. Therefore the rank equals that of

$$\begin{bmatrix} -s & 0 & 0 & 0 & 1 \\ 1 & -s & 0 & 1 & 0 \\ 0 & 1 & -s & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \end{bmatrix} \Rightarrow \begin{bmatrix} -s & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \end{bmatrix} \Rightarrow \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \end{bmatrix}.$$

As s has disappeared from the matrix pencil it is direct that the rank of the matrix pencil does not depend on s . The system hence has no transmission zeros. \square

3.2.4. Numerical computation of transmission zeros

If a transfer matrix P is given, the numerical computation of transmission zeros is in general based on a minimal state space realization of P because reliable numerical algorithms for rational and polynomial matrix operations are not yet numerically as reliable as state space methods. We briefly discuss two approaches.

Algorithm 3.2.12 (Transmission zeros via high gain output feedback). The approach has been proposed by [Davison and Wang \(1974\)](#), [1978](#). Let (A, B, C, D) be a minimal realization of P and assume that P is either left-invertible or right-invertible. Then:

1. Determine an arbitrary full rank output feedback matrix $K \in \mathbb{R}^{n_u \times n_y}$, e.g. using a random number generator.
2. Determine the eigenvalues of the matrix

$$Z_\rho := A + BK\left(\frac{1}{\rho}I - DK\right)^{-1}C$$

for various large real values of ρ , e.g. $\rho = 10^{10}, \dots, 10^{20}$ at double precision computation. Now as ρ goes to ∞ we observe that some eigenvalues go to infinity while others converge (the so called *finite* eigenvalues).

3. For square systems, the transmission zeros equal the finite eigenvalues of Z_ρ . For non-square systems, the transmission zeros equal the finite eigenvalues of Z_ρ for *almost* all choices of the output feedback matrix K .
4. In cases of doubt, vary the values of ρ and K and repeat the calculations. \square

See Section 3.4 for a sketch of the proof.

Algorithm 3.2.13 (Transmission zeros via generalized eigenvalues). The approach has been proposed in [Laub and Moore \(1978\)](#) and makes use of the QZ algorithm for solving the generalized eigenvalue problem. Let (A, B, C, D) be a minimal realization of P of order n having the additional property that it is left-invertible (if the system is right-invertible, then use the dual system (A^T, C^T, B^T, D^T)). Then:

1. Define the matrices M and L as

$$M = \begin{bmatrix} I_n & 0 \\ 0 & 0 \end{bmatrix}, \quad L = \begin{bmatrix} A & B \\ -C & D \end{bmatrix}.$$

2. Compute a solution to the *generalized eigenvalue* problem i.e. determine all values $s \in \mathbb{C}$ and $r \in \mathbb{C}^{n+n_u}$ satisfying

$$[sM - L]r = 0$$

3. The set of *finite* values for s are the transmission zeros of the system.

Although reliable numerical algorithms exist for the solution of generalized eigenvalue problems, a major problem is to decide which values of s belong to the finite transmission zeros and which values are to be considered as infinite. However, this problem is inherent in all numerical approaches to transmission zero computation. \square

3.3. MIMO structural requirements and design methods

During both the construction and the design of a control system it is mandatory that structural requirements of plant and controller and the control configuration are taken into account. For example, the choice of placement of the actuators and sensors may affect the number of right-half plane zeros of the plant and thereby may limit the closed loop performance, irrespective which controller is used. In this section a number of such requirements will be identified. These requirements show up in the MIMO design methods that are discussed in this section.

3.3.1. Output controllability and functional reproducibility

Various concepts of *controllability* are relevant for the design of control systems. A minimal requirement for almost any control system, to be included in any control objective, is the requirement that the output of the system can be steered to any desired position in output space by manipulating the input variables. This property is formalized in the concept of *output controllability*.

Summary 3.3.1 (Output controllability). A time-invariant linear system with input $u(t)$ and output $y(t)$ is said to be *output controllable* if for any $y_1, y_2 \in \mathbb{R}^p$ there exists an input $u(t)$, $t \in [t_1, t_2]$ with $t_1 < t_2$ that brings the output from $y(t_1) = y_1$ to $y(t_2) = y_2$. In case the system has a strictly proper transfer matrix and is described by a state-space realization (A, B, C) , the system is *output controllable* if and only if the constant matrix

$$\begin{bmatrix} CB & CAB & CA^2B & \dots & CA^{n-1}B \end{bmatrix} \quad (3.26)$$

has full row rank □

If in the above definition $C = I_n$ then we obtain the definition of state controllability. The concept of output controllability is generally weaker: a system (A, B, C) may be output controllable and yet not be (state) controllable². The property of output controllability is an *input-output property* of the system while (state) controllability is not.

Note that the concept of output controllability only requires that the output can be given a desired value at each instant of time. A stronger requirement is to demand the output to be able to follow any preassigned trajectory in time over a given time interval. A system capable of satisfying this requirement is said to be *output functional reproducible* or *functional controllable*. Functional controllability is a necessary requirement for output regulator and servo/tracking problems. [Brockett and Mesarovic \(1965\)](#) introduced the notion of *reproducibility*, termed *output controllability* by [Rosenbrock \(1970\)](#).

Summary 3.3.2 (Output functional reproducibility). A system having proper real-rational $n_y \times n_u$ transfer matrix is said to be *functionally reproducible* if $\text{rank } P = n_y$. In particular for functionally reproducible it is necessary that $n_y \leq n_u$. □

Example 3.3.3 (Output controllability and Functional reproducibility). Consider the linear time-invariant state-space system of Example 3.2.11,

$$A = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad C = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (3.27)$$

²Here we assume that C has full row rank, which is a natural assumption because otherwise the entries of y are linearly dependent.

This forms a minimal realization of the transfer matrix

$$P(s) = C(sI - A)^{-1}B = \begin{bmatrix} \frac{1}{s} & \frac{1}{s^2} \\ \frac{1}{s^2} & \frac{1}{s^3} \end{bmatrix}. \quad (3.28)$$

The system (A, B, C) is controllable and observable, and C has full rank. Thus the system also is output controllable. However, the rank of P is 1. Thus the system is *not* functionally reproducible. \square

3.3.2. Decoupling control

A classical approach to multivariable control design consists of the design of a precompensator that brings the system transfer matrix to diagonal form, with subsequent design of the actual feedback loops for the various single-input, single-output channels separately. This allows the tuning of individual controllers in separate feedback loops, and it is thought to provide an acceptable control structure providing ease of survey for process operators and maintenance personnel. The subject of noninteracting or decoupling control as discussed in this section is based on the works of [Silverman \(1970\)](#), [Williams and Antsaklis \(1986\)](#). The presentation follows that of [Williams and Antsaklis \(1996\)](#).

In this section we investigate whether and how a square P can be brought to diagonal form by applying feedback and/or static precompensation. Suppose that P has state-space representation

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

with u and y having equally many entries, $n_y = n_u = m$, i.e., $P(s) = C(sI - A)^{-1}B + D$ square. Assume in what follows that P is invertible. The inverse transfer matrix P^{-1} —necessarily a rational matrix—may be viewed as a precompensator of P that diagonalizes the series connection PP^{-1} . If the direct feedthrough term $D = P(\infty)$ of P is invertible, then the inverse P^{-1} is proper and has realization

$$P^{-1}(s) = -D^{-1}C(sI - A + BD^{-1}C)^{-1}BD^{-1} + D^{-1}. \quad (3.29)$$

If D is singular then the inverse P^{-1} —assuming it exists—is not proper. In this case we proceed as follows. Define the indices $f_i \geq 0$ ($i = 1, \dots, m$) such that D_f defined as

$$D_f(s) = \text{diag}(s^{f_1}, s^{f_2}, \dots, s^{f_m}) \quad (3.30)$$

is such that

$$\mathcal{D} := \lim_{|s| \rightarrow \infty} D_f(s)P(s) \quad (3.31)$$

is defined and every row of \mathcal{D} has at least one nonzero entry. This identifies the indices f_i uniquely. The f_i equal the maximum of the relative degrees of the entries in the i th row of P . Indeed, then by construction the largest relative degree in each row of D_fP is zero, and as a consequence $\lim_{|s| \rightarrow \infty} D_f(s)P(s)$ has a nonzero value at precisely the entries of relative degree zero. The indices f_i may also be determined using the realization of P : If the i th row of D is not identical to zero then $f_i = 0$, otherwise

$$f_i = \min\{k > 0 \mid \text{row } i \text{ of } CA^{k-1}B \text{ is not identical to zero}\}.$$

It may be shown that $f_i \leq n$, where n is the state dimension. The indices f_i so defined are known as the *decoupling indices*.

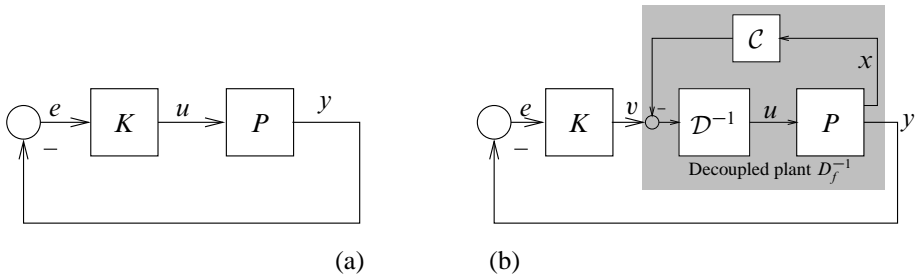


Figure 3.4: (a) closed loop with original plant, (b) with decoupled plant

Summary 3.3.4 (Decoupling state feedback). If \mathcal{D} as defined in (3.31) is nonsingular, then the regular state feedback

$$u(t) = \mathcal{D}^{-1}(-\mathcal{C}x(t) + v(t))$$

where

$$\mathcal{C} = \begin{bmatrix} C_{1*}A^{f_1} \\ C_{2*}A^{f_2} \\ \vdots \\ C_{m*}A^{f_m} \end{bmatrix}$$

renders the system with output y and new input v decoupled with diagonal transfer matrix $D_f^{-1}(s) = \text{diag}(s^{-f_1}, s^{-f_2}, \dots, s^{-f_m})$. The compensated plant is shown in Fig. 3.4(b). \square

Here C_{i*} is denotes the i th row of the matrix C . A proof is given in Appendix 3.4. The decoupled plant D_f^{-1} has all its poles at $s = 0$ and it has no transmission zeros. We also know that regular state-feedback does not change the zeros of the matrix $\begin{bmatrix} A^{-s}I & B \\ C & D \end{bmatrix}$. This implies that after state feedback all non-zero transmission zeros of the system are canceled by poles. In other words, after state feedback the realization has unobservable modes at the open loop transmission zeros. This is undesirable if P has unstable transmission zeros as it precludes closed loop stability using output feedback. If P has no unstable transmission zeros, then we may proceed with the decoupled plant D_f^{-1} and use (diagonal) K to further the design, see Fig. 3.4. A decoupled plant that is stable may offer better perspectives for further performance enhancement by multiloop feedback in individual channels. To obtain a stable decoupled plant we take instead of (3.30) a D_f of the form

$$D_f = \text{diag}(p_1, \dots, p_m)$$

where the p_i are strictly Hurwitz³ if monic⁴ polynomials of degree f_i . The formulae are now more messy, but the main results holds true also for this choice of D_f :

Summary 3.3.5 (Decoupling, stabilizing state feedback). Suppose the transfer matrix P of the plant is proper, square $m \times m$ and invertible, and suppose it has no unstable transmission zeros. Let (A, B, C, D) be a minimal realization of P and let f_i be the decoupling indices and suppose that p_i are Hurwitz polynomials of degree f_i . Then there is a regular state feedback $u(t) = Fx(t) + Wv(t)$ for which the loop from $v(t)$ to $y(t)$ is decoupled. Its realization is controllable and detectable and has a stable transfer matrix $\text{diag}(\frac{1}{p_1}, \dots, \frac{1}{p_m})$. \square

³A strictly Hurwitz polynomial is a polynomial whose zeros have strictly negative real part.

⁴A monic polynomial is a polynomial whose highest degree coefficient equals 1.

Example 3.3.6 (Diagonal decoupling by state feedback precompensation). Consider a plant with transfer matrix

$$P(s) = \begin{bmatrix} 1/s & 1/s^2 \\ -1/s^2 & -1/s^2 \end{bmatrix}.$$

A minimal realization of P is

$$\left[\begin{array}{c|c} A & B \\ \hline C & D \end{array} \right] = \left[\begin{array}{cccc|cc} 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ \hline 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{array} \right].$$

Its transmission zeros follow after a sequence of elementary row and column operations

$$P(s) = \frac{1}{s^2} \begin{bmatrix} s & 1 \\ 1 & -1 \end{bmatrix} \Rightarrow \frac{1}{s^2} \begin{bmatrix} s+1 & 1 \\ 0 & 1 \end{bmatrix} \Rightarrow \frac{1}{s^2} \begin{bmatrix} 1 & 0 \\ 0 & s+1 \end{bmatrix} = \begin{bmatrix} 1/s^2 & 0 \\ 0 & (s+1)/s^2 \end{bmatrix}.$$

Thus P has one transmission zeros at $s = -1$. It is a stable zero hence we may proceed. The direct feedthrough term D is the zero matrix so the decoupling indices f_i are all greater than zero. We need to compute $C_{i*}A^{k-1}B$.

$$C_{1*}B = \begin{bmatrix} 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 1 & -1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \end{bmatrix}$$

it is not identically zero so that $f_1 = 1$. Likewise

$$C_{2*}B = \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 1 & -1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \end{bmatrix}.$$

It is zero hence we must compute $C_{2*}AB$,

$$C_{2*}AB = \begin{bmatrix} 1 & -1 \end{bmatrix}.$$

As it is nonzero we have $f_2 = 2$. This gives us

$$\mathcal{D} = \begin{bmatrix} C_{1*}B \\ C_{2*}AB \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & -1 \end{bmatrix}.$$

The matrix \mathcal{D} is nonsingular so a decoupling state feedback exists. Take

$$D_f(s) = \begin{bmatrix} s+1 & 0 \\ 0 & (s+1)^2 \end{bmatrix}.$$

Its diagonal entries have the degrees $f_1 = 1$ and $f_2 = 2$ as required. A realization of

$$D_f(s)P(s) = \begin{bmatrix} (s+1)/s & (s+1)/s^2 \\ (s+1)^2/s^2 & -(s+1)^2/s^2 \end{bmatrix}$$

is (A, B, C, D) with

$$C = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 2 & 1 \end{bmatrix}, \quad D = \begin{bmatrix} 1 & 0 \\ 1 & -1 \end{bmatrix}.$$

Now the regular state feedback

$$u(t) = \begin{bmatrix} 1 & 0 \\ 1 & -1 \end{bmatrix}^{-1} \left(-\begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 2 & 1 \end{bmatrix} x(t) + v(t) \right)$$

renders the system with input v and output y , decoupled. After applying this state feedback we arrive at the state-space system

$$\begin{aligned} \dot{x}(t) &= A + BD^{-1}(-Cx(t) + v(t)) = \begin{bmatrix} -1 & -1 & -2 & 1 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -2 & -1 \\ 0 & 0 & 1 & 0 \end{bmatrix} x(t) + \begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} v(t) \\ y(t) &= Cx(t) + Dv(t) = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} x(t) \end{aligned} \quad (3.32)$$

and its transfer matrix by construction is

$$D_f^{-1}(s) = \begin{bmatrix} 1/(s+1) & 0 \\ 0 & 1/(s+1)^2 \end{bmatrix}.$$

Note that a minimal realization of D_f^{-1} has order 3 whereas the realization (3.32) has order 4. Therefore (3.32) is not minimal. This is typical for decoupling procedures. \square

3.3.3. Directional properties of gains

In contrast to a SISO system, a MIMO system generally does not have a unique gain. A trivial example is the 2×2 system with constant transfer matrix

$$P(s) = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}.$$

The gain is in between 1 and 2 depending on the direction of the input.

There are various ways to define gains for MIMO systems. A natural generalization of the SISO gain $|y(j\omega)|/|u(j\omega)|$ from input $u(j\omega)$ to output $y(j\omega)$ is to use norms instead of merely absolute values. We take the 2-norm. This subsection assumes knowledge of *norms*, see Appendix B.

Summary 3.3.7 (Bounds on gains). For any given input u and fixed frequency ω there holds

$$\underline{\sigma}(P(j\omega)) \leq \frac{\|P(j\omega)u(j\omega)\|_2}{\|u(j\omega)\|_2} \leq \overline{\sigma}(P(j\omega)) \quad (3.33)$$

and the lower bound $\underline{\sigma}(P(j\omega))$ and upper bound $\overline{\sigma}(P(j\omega))$ are achieved for certain inputs u . \square

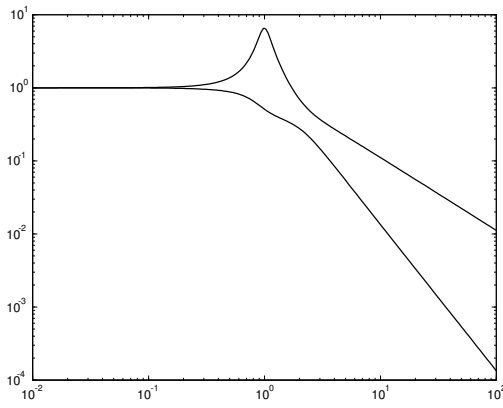


Figure 3.5: Lower and upper bound on the gain of $T(j\omega)$

Example 3.3.8 (Upper and lower bounds on gains). Consider the plant with transfer matrix

$$P(s) = \begin{bmatrix} \frac{1}{s} & \frac{1}{s^2 + \sqrt{2}s + 1} \\ -\frac{1}{2(s+1)} & \frac{1}{s} \end{bmatrix}.$$

In unity feedback the complementary sensitivity matrix $T = P(I + P)^{-1}$ has dimension 2×2 . So at each frequency ω the matrix $T(j\omega)$ has two singular values. Figure 3.5 shows the two singular values $\bar{\sigma}(j\omega)$ and $\underline{\sigma}(j\omega)$ as a function of frequency. Near the crossover frequency the singular values differ considerably. \square

Let

$$P(j\omega) = Y(j\omega)\Sigma(j\omega)U^*(j\omega) \quad (3.34)$$

be an SVD (at each frequency) of $P(j\omega)$, that is, $Y(j\omega)$ and $U(j\omega)$ are unitary and

$$\Sigma(j\omega) = \text{diag}(\sigma_1(j\omega), \sigma_2(j\omega), \dots, \sigma_{\min(n_y, n_u)}(j\omega))$$

with

$$\sigma_1(j\omega) \geq \sigma_2(j\omega) \geq \dots \geq \sigma_{\min(n_y, n_u)}(j\omega) \geq 0.$$

The columns $u_k(j\omega)$ of $U(j\omega)$ the *input principal directions*, [Postlethwaite et al. \(1981\)](#) and they have the special property that their gains are precisely the corresponding singular values

$$\frac{\|P(j\omega)u_k(j\omega)\|_2}{\|u_k(j\omega)\|_2} = \sigma_k(j\omega)$$

and the response $y_k(j\omega) = P(j\omega)u_k(j\omega)$ in fact equals $\sigma_k(j\omega)$ times the k th column of $Y(j\omega)$. The columns of $Y(j\omega)$ are the *output principal directions* and the $\sigma_k(j\omega)$ the *principal gains*.

The *condition number* κ defined as

$$\kappa(P(j\omega)) = \frac{\bar{\sigma}(P(j\omega))}{\underline{\sigma}(P(j\omega))} \quad (3.35)$$

is a possible measure of the difficulty of controlling the system. If for a plant all principal gains are the same, such as when $P(j\omega)$ is unitary, then $\kappa(j\omega) = 1$. If $\kappa(j\omega) \gg 1$ then the loop

gain $L = PK$ around the crossover frequency *may* be difficult to shape. Note that the condition number is not invariant under scaling. Describing the same system in terms of a different set of physical units leads to a different condition number. A high condition number indicates that the system is “close” to losing its full rank, i.e. close to not satisfying the property of functional controllability.

Example 3.3.9 (Multivariable control system—design goals). The performance of a multivariable feedback system can be assessed using the notion of principal gains as follows. The basic idea follows the ideas previously discussed for single input, single output systems.

The disturbance rejecting properties of the loop transfer require the sensitivity matrix to be small:

$$\overline{\sigma}(S(j\omega)) \ll 1.$$

At the same time the restriction of the propagation of measurement noise requires

$$\overline{\sigma}(T(j\omega)) \ll 1.$$

As in the SISO case these two requirements are conflicting. Indeed, using the triangle inequality we have the bounds

$$|1 - \overline{\sigma}(S(j\omega))| \leq \overline{\sigma}(T(j\omega)) \leq 1 + \overline{\sigma}(S(j\omega)) \quad (3.36)$$

and

$$|1 - \overline{\sigma}(T(j\omega))| \leq \overline{\sigma}(S(j\omega)) \leq 1 + \overline{\sigma}(T(j\omega)). \quad (3.37)$$

It is useful to try to make the difference between $\underline{\sigma}(T)$ and $\overline{\sigma}(T)$, and between $\underline{\sigma}(S)$ and $\overline{\sigma}(S)$, not too large in the cross over region. Making them equal would imply that the whole plant behaves identical in all directions and this is in general not possible.

We must also be concerned to make the control inputs not too large. Thus the transfer matrix $(I + KP)^{-1}K$ from r to u must not be too large.

$$\begin{aligned} \frac{\|u(j\omega)\|_2}{\|r(j\omega)\|_2} &= \frac{\|(I + K(j\omega)P(j\omega))^{-1}K(j\omega)r(j\omega)\|_2}{\|r(j\omega)\|_2} \\ &\leq \overline{\sigma}((I + K(j\omega)P(j\omega))^{-1}K(j\omega)) \\ &\leq \overline{\sigma}((I + K(j\omega)P(j\omega))^{-1})\overline{\sigma}(K(j\omega)) \\ &\leq \frac{\overline{\sigma}(K(j\omega))}{\underline{\sigma}(I + K(j\omega)P(j\omega))} \\ &\leq \frac{\overline{\sigma}(K(j\omega))}{1 - \overline{\sigma}(K(j\omega)P(j\omega))} \\ &\leq \frac{\overline{\sigma}(K(j\omega))}{1 - \overline{\sigma}(K(j\omega))\overline{\sigma}(P(j\omega))}. \end{aligned} \quad (3.38)$$

Here several properties of Exercise B.2 are used and in the last two inequalities it is assumed that $\overline{\sigma}(K(j\omega))\overline{\sigma}(P(j\omega)) < 1$. The upperbound shows that where $\overline{\sigma}(P(j\omega))$ is not too large that $\overline{\sigma}(K(j\omega)) \ll 1$ guarantees a small gain from $r(j\omega)$ to $u(j\omega)$. The upper bound (3.38) of this gain is easily determined numerically.

The direction of an important disturbance can be taken into consideration to advantage. Let $v(j\omega)$ be a disturbance acting additively on the output of the system. Complete rejection of the disturbance would require an input

$$u(j\omega) = -P^{-1}(j\omega)v(j\omega). \quad (3.39)$$

Thus

$$\frac{\|u(j\omega)\|_2}{\|v(j\omega)\|_2} = \frac{\|P^{-1}(j\omega)v(j\omega)\|_2}{\|v(j\omega)\|_2} \quad (3.40)$$

measures the magnitude of u needed to reject a unit magnitude disturbance acting in the direction v . The most and least favorable disturbance directions are those for which v is into the direction of the principal output directions corresponding to $\overline{\sigma}(P(j\omega))$ and $\underline{\sigma}(P(j\omega))$, respectively. This leads to the *disturbance condition number* of the plant,

$$\kappa_v(P(j\omega)) = \frac{\|P^{-1}(j\omega)v(j\omega)\|_2}{\|v(j\omega)\|_2} \overline{\sigma}(P(j\omega)) \quad (3.41)$$

which measures the input needed to reject the disturbance v relative to the input needed to reject the disturbance acting in the most favorable direction. It follows that

$$1 \leq \kappa_v(P) \leq \kappa(P) \quad (3.42)$$

and $\kappa_v(P)$ generalizes the notion of condition number of the plant defined in (3.35). \square

3.3.4. Decentralized control structures

Dual to decoupling control where the aim is make the plant P diagonal, is *decentralized control* where it is the controller K that is restricted to be diagonal or block diagonal,

$$K = \text{diag}\{K_i\} = \begin{bmatrix} K_1 & & & 0 \\ & K_2 & & \\ & & \ddots & \\ 0 & & & K_m \end{bmatrix}. \quad (3.43)$$

This control structure is called *multiloop control* and it assumes that there as many control inputs as control outputs. In a multiloop control structure it is in general desirable to make an ordering of the input and output variables such that the interaction between the various control loops is as small as possible. This is the *input-output pairing problem*.

Example 3.3.10 (Relative gain of a two-input-two-output system). Consider a two-input-two-output system

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = P \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

and suppose we decide to close the loop by a diagonal controller $u_j = K_j(r_j - y_j)$, that is, the first component of u is only controlled by the first component of y , and similarly u_2 is controlled only by y_2 . Suppose we leave the second control loop open for the moment and that we vary the first control loop. If the open loop gain from u_2 to y_2 does not change a lot as we vary the first loop $u_1 = K_1(r_1 - y_1)$, it is save to say that we can design the second control loop independent from the first, or to put it differently, that the second control loop is insensitive to tuning of the first. This is desirable. In summary, we want the ratio

$$\lambda := \frac{\text{gain from } u_2 \text{ to } y_2 \text{ if first loop is open}}{\text{gain from } u_2 \text{ to } y_2 \text{ if first loop is closed}} \quad (3.44)$$

preferably close to 1. To make this λ more explicit we assume now that the reference signal r is constant and that all signals have settled to their constant steady state values. Now if the first loop is open then the first input entry u_1 is zero (or constant). However if the first loop is *closed* then—assuming perfect control—it is the output y_1 that is constant (equal to r_1). This allows us to express (3.44) as

$$\lambda = \frac{dy_2/du_2|_{u_1}}{dy_2/du_2|_{y_1}}$$

where $|_{u_1}$ expresses that u_1 is considered constant in the differentiation. This expression for λ exists and may be computed if $P(0)$ is invertible:

$$\frac{dy_2}{du_2} \Big|_{u_1} = \frac{d P_{21}(0)u_1 + P_{22}(0)u_2}{du_2} \Big|_{u_1} = P_{22}(0)$$

and because $u = P^{-1}(0)y$ we also have

$$\frac{dy_2}{du_2} \Big|_{y_1} = \frac{1}{du_2/dy_2} \Big|_{y_1} = \frac{1}{\frac{d P_{21}^{-1}(0)y_1 + P_{22}^{-1}(0)y_2}{dy_2}} \Big|_{y_1} = \frac{1}{P_{22}^{-1}(0)} = P_{22}^{-1}(0).$$

The relative gain λ hence equals $P_{22}(0)P_{22}^{-1}(0)$. □

For general square MIMO systems P we want to consider the *relative gain array (RGA)* which is the (rational matrix) Λ defined element-wise as

$$\Lambda_{ij} = P_{ij} P_{ji}^{-1}$$

or, equivalently as a matrix, as

$$\Lambda = P \circ (P^{-1})^T \tag{3.45}$$

where \circ denotes the *Hadamard product* which is the entry-wise product of two matrices of the same dimensions.

Summary 3.3.11 (Properties of the RGA).

1. In constant steady state there holds that

$$\Lambda_{ij}(0) = \frac{\frac{dy_j}{du_j} \Big|_{\text{all loops } (u_k, y_k), k \neq j \text{ open}}}{\frac{dy_j}{du_j} \Big|_{\text{all loops } (u_k, y_k), k \neq j \text{ closed}}}$$

2. The sum of the elements of each row or each column of Λ is 1
3. Any permutation of rows and columns in P results in the same permutations of rows and columns in Λ
4. Λ is invariant under diagonal input and output scaling of P
5. If P is diagonal or upper or lower triangular then $\Lambda = I_m$

6. The norm $\overline{\sigma}(\Lambda(j\omega))$ of the RGA is believed to be closely related to the minimized condition number of $P(j\omega)$ under diagonal scaling, and thus serves as an indicator for badly conditioned systems. See [Nett and Manousiouthakis \(1987\)](#), [Grosdidier et al. \(1985\)](#), and [Skogestad and Morari \(1987\)](#).
7. The RGA measures the sensitivity of P to relative element-by-element uncertainty. It indicates that P becomes singular by an element perturbation from P_{ij} to $P_{ij}[1 - \lambda_{ij}^{-1}]$.

If Λ deviates a lot from I_m then this is an indication that interaction is present in the system. \square

Example 3.3.12 (Relative Gain Array for 2×2 systems). For the two-input, two-output case

$$P = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix}, \quad (P^{-1})^T = \frac{1}{P_{11}P_{22} - P_{21}P_{12}} \begin{bmatrix} P_{22} & -P_{21} \\ -P_{12} & P_{11} \end{bmatrix},$$

the relative gain array $\Lambda = P \circ P^{-T}$ equals

$$\Lambda = \begin{bmatrix} \lambda & 1 - \lambda \\ 1 - \lambda & \lambda \end{bmatrix}, \quad \lambda := \frac{P_{11}P_{22}}{P_{11}P_{22} - P_{21}P_{12}}.$$

We immediately see that rows and columns add up to one, as claimed. Some interesting special cases are

$$\begin{aligned} P = \begin{bmatrix} P_{11} & 0 \\ 0 & P_{22} \end{bmatrix} &\Rightarrow \Lambda = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\ P = \begin{bmatrix} P_{11} & 0 \\ P_{21} & P_{22} \end{bmatrix} &\Rightarrow \Lambda = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\ P = \begin{bmatrix} 0 & P_{12} \\ P_{21} & 0 \end{bmatrix} &\Rightarrow \Lambda = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \\ P = \begin{bmatrix} p & p \\ -p & p \end{bmatrix} &\Rightarrow \Lambda = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix} \end{aligned}$$

In the first and second example, the relative gain suggests to pair u_1 with y_1 , and u_2 with y_1 . In the first example this pairing is also direct from the fact that P is diagonal. In the third example the RGA is *antidiagonal*, so we want to pair u_1 with y_2 and u_2 with y_1 . In the fourth example all entries of the RGA are the same, hence no pairing can be deduced from the RGA.

If P is close to singular then several entries of Λ are large. The corresponding pairings are to be avoided, and possibly no sensible pairing exists. For example

$$P = \begin{bmatrix} a & a \\ a & a(1 + \delta) \end{bmatrix} \Rightarrow \Lambda = \begin{bmatrix} \frac{1+\delta}{\delta} & \frac{-1}{\delta} \\ \frac{-1}{\delta} & \frac{1+\delta}{\delta} \end{bmatrix}$$

with $\delta \approx 0$. \square

Although the RGA has been derived originally by [Bristol \(1966\)](#) for the evaluation of $P(s)$ at steady-state $s = 0$ assuming stable steady-state behavior, the RGA may be useful over the complete frequency range. The following *i/o pairing rules* have been in use on the basis of an heuristic understanding of the properties of the RGA:

1. Prefer those pairing selections where $\Lambda(j\omega)$ is close to unity around the crossover frequency region. This prevents undesirable stability interaction with other loops.

2. Avoid pairings with negative steady-state or low-frequency values of $\lambda_{ii}(j\omega)$ on the diagonal.

The following result by [Hovd and Skogestad \(1992\)](#) makes more precise the statement in [Bristol \(1966\)](#) relating negative entries in the RGA, and nonminimum-phase behavior. See also the counterexample against Bristol's claim in [Grosdidier and Morari \(1987\)](#).

Summary 3.3.13 (Relative Gain Array and RHP transmission zeros). Suppose that P has stable elements having no poles nor zeros at $s = 0$. Assume that the entries of Λ are nonzero and finite for $s \rightarrow \infty$. If Λ_{ij} shows different signs when evaluated at $s = 0$ and $s = \infty$, then at least one of the following statements holds:

1. The entry P_{ij} has a zero in the right half plane
2. P has a transmission zero in the right half plane
3. The subsystem of P with input j and output i removed has a transmission zero in the right half plane

□

As decentralized or multiloop control structures are widely applied in industrial practice, a requirement arises regarding the ability to take one loop out of operation without destabilizing the other loops. In a problem setting employing integral feedback, this leads to the notion of decentralized integral controllability [Campo and Morari \(1994\)](#).

Definition 3.3.14 (Decentralized integral controllability (DIC)). The system P is DIC if there exists a decentralized (multiloop) controller having integral action in each loop such that the feedback system is stable and remains stable when each loop i is individually detuned by a factor ε_i for $0 \leq \varepsilon_i \leq 1$.

□

The definition of DIC implies that the system P must be open-loop stable. It is also assumed that the integrator in the control loop is put out of order when $\varepsilon_i = 0$ in loop i . The steady-state RGA provides a tool for testing on DIC. The following result has been shown by [Grosdidier et al. \(1985\)](#).

Summary 3.3.15 (Steady-state RGA and stability). Let the plant P be stable and square, and consider a multiloop controller K with diagonal transfer matrix having integral action in each diagonal element. Assume further that the loop gain PK is strictly proper. If the RGA $\Lambda(s)$ of the plant contains a negative diagonal value for $s = 0$ then the closed-loop system satisfies at least one of the following properties:

1. The overall closed-loop system is unstable
2. The loop with the negative steady-state relative gain is unstable by itself
3. The closed-loop system is unstable if the loop corresponding to the negative steady-state relative gain is opened

□

A further result regarding stability under multivariable integral feedback has been provided by [Lunze \(1985\)](#); see also [Grosdidier et al. \(1985\)](#), and [Morari \(1985\)](#).

Summary 3.3.16 (Steady-state gain and DIC). Consider the closed loop system consisting of the open-loop stable plant P and the controller K having transfer matrix

$$K(s) := \frac{\alpha}{s} K_{ss} \quad (3.46)$$

Assume unity feedback around the loop and assume that the summing junction in the loop introduces a minus sign in the loop. Then a necessary condition for the closed loop system to be stable for all α in an interval $0 \leq \alpha \leq \bar{\alpha}$ with $\bar{\alpha} > 0$ is that the matrix $P(0)K_{ss}$ must have all its eigenvalues in the open right half plane. \square

Further results on decentralized i/o pairing can be found in [Skogestad and Morari \(1992\)](#), and in [Hovd and Skogestad \(1994\)](#). The inherent control limitations of decentralized control have been studied by [Skogestad et al. \(1991\)](#). Further results on decentralized integral controllability can be found in [Le et al. \(1991\)](#) and in [Nwokah et al. \(1993\)](#).

3.3.5. Internal model principle and the servomechanism problem

The *servomechanism problem* is to design a controller – called *servocompensator* – that renders the plant output as insensitive as possible to disturbances while at the same asymptotically track certain reference inputs r . We discuss a state space solution to this problem; a solution that works for SISO as well as MIMO systems. Consider a plant with state space realization

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t) & u(t) \in \mathbb{R}^{n_u}, \quad y \in \mathbb{R}^{n_y}, \quad x(0) = x_0 \in \mathbb{R}^n \\ y(t) &= Cx(t) + v(t) \end{aligned} \quad (3.47)$$

and assume that v is a disturbance that is generated by a dynamic system having all of its poles on the imaginary axis

$$\begin{aligned} \dot{x}_v(t) &= A_v x_v(t), & x_v(0) = x_{v0} \\ v(t) &= C_v x_v(t). \end{aligned} \quad (3.48)$$

The disturbance is *persistent* because the eigenvalues of A_v are assumed to lie on the imaginary axis. Without loss of generality we further assume that (A_v, C_v) is an observable pair. Then a possible choice of the *servocompensator* is the compensator (controller) with realization

$$\begin{aligned} \dot{x}_s(t) &= A_s x_s(t) + B_s e(t), & x_s(0) = x_{s0} \\ e(t) &= r(t) - y(t) \end{aligned} \quad (3.49)$$

where A_s is block-diagonal with n_y blocks

$$A_s = \text{diag}(A_v, A_v, \dots, A_v) \quad (3.50)$$

and where B_s is of compatible dimensions and such that (A_s, B_s) is controllable, but otherwise arbitrary. The composite system, consisting of the plant (3.47) and the servocompensator (3.49) yields

$$\begin{aligned} \begin{bmatrix} \dot{x}(t) \\ \dot{x}_s(t) \end{bmatrix} &= \begin{bmatrix} A & 0 \\ -B_s C & A_s \end{bmatrix} \begin{bmatrix} x(t) \\ x_s(t) \end{bmatrix} + \begin{bmatrix} B \\ 0 \end{bmatrix} u(t) + \begin{bmatrix} 0 \\ B_s \end{bmatrix} (r(t) - v(t)) \\ y(t) &= \begin{bmatrix} C & 0 \end{bmatrix} \begin{bmatrix} x(t) \\ x_s(t) \end{bmatrix} + v(t). \end{aligned}$$

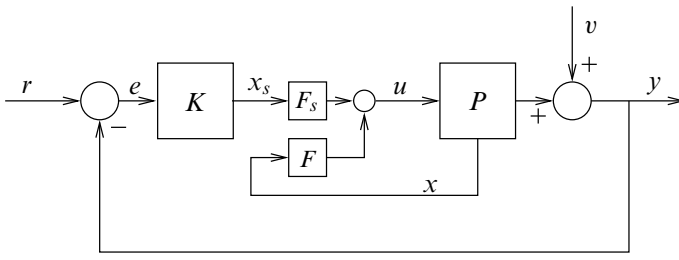


Figure 3.6: Servomechanism configuration

As it stands the closed loop is not stable, but it may be stabilized by an appropriate state feedback

$$u(t) = Fx(t) + F_s x_s(t) \quad (3.51)$$

where F and F_s are chosen by any method for designing stabilizing state feedback, such as LQ theory. The resulting closed loop is shown in Fig. 3.6.

Summary 3.3.17 (Servocompensator). Suppose that the following conditions are satisfied.

1. (A, B) is stabilizable and (A, C) is detectable,
2. $n_u \geq n_y$, i.e. there are at least as many inputs as there are outputs
3. $\text{rank} \begin{bmatrix} A - \lambda_i I & B \\ C & D \end{bmatrix} = n + n_y$ for every eigenvalue λ_i of A_v .

Then F, F_s can be selected such that the closed loop system matrix

$$\begin{bmatrix} A + BF & BF_s \\ -B_s C & A_s \end{bmatrix} \quad (3.52)$$

has all its eigenvalues in the open left-half plane. In that case the controller (3.49),(3.51) is a servocompensator in that $e(t) \rightarrow 0$ as $t \rightarrow \infty$ for any of the persistent v that satisfy (3.48). \square

The system matrix A_s of the servocompensator contains several copies of the system matrices A_v that defines v . This means that the servocompensator contains the mechanism that generates v ; it is an example of the more general *internal model principle*. The conditions as mentioned imply the following, see Davison (1996) and Desoer and Wang (1980).

- No transmission zero of the system should coincide with one of the imaginary axis poles of the disturbance model
- The system should be functionally controllable
- The variables r and v occur in the equations in a completely interchangeable role, so that the disturbance model used for v can also be utilized for r . That is, for any r that satisfies

$$\dot{x}_v(t) = A_v x_v(t), \quad r(t) = C_r x_v(t)$$

the error signal $e(t)$ converges to zero. Hence the output $y(t)$ approaches $r(t)$ as $t \rightarrow \infty$.

- Asymptotically, when $e(t)$ is zero, the disturbance $v(t)$ is still present and the servocompensator acts in an open loop fashion as the autonomous generator of a compensating disturbance at the output $y(t)$, of equal form as $v(t)$ but of opposite sign.

- If the state of the system is not available for feedback, then the use of a state estimator is possible without altering the essentials of the preceding results.

Example 3.3.18 (Integral action). If the servocompensator is taken to be an integrator $\dot{x}_s = e$ then constant disturbances v are asymptotically rejected and constant reference inputs r asymptotically tracked. We shall verify the conditions of Summary 3.3.17 for this case. As before the plant P is assumed strictly proper with state space realization and corrupted with noise v ,

$$\dot{x}(t) = Ax(t) + Bu(t), \quad y(t) = Cx(t) + v(t).$$

Combined with the integrating action of the servocompensator $\dot{x}_s(t) = e(t)$, with $e(t) = r(t) - y(t)$, we obtain

$$\begin{aligned} \begin{bmatrix} \dot{x}(t) \\ \dot{x}_s(t) \end{bmatrix} &= \begin{bmatrix} A & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x(t) \\ x_s(t) \end{bmatrix} + \begin{bmatrix} B \\ 0 \end{bmatrix} u(t) + \begin{bmatrix} 0 \\ I \end{bmatrix} e(t) \\ &= \begin{bmatrix} A & 0 \\ -C & 0 \end{bmatrix} \begin{bmatrix} x \\ x_s \end{bmatrix} + \begin{bmatrix} B \\ 0 \end{bmatrix} u(t) + \begin{bmatrix} 0 \\ I \end{bmatrix} (r(t) - v(t)). \end{aligned}$$

Closing the loop with state feedback $u(t) = Fx(t) + F_s x_s(t)$ renders the closed loop state space representation

$$\begin{aligned} \begin{bmatrix} \dot{x}(t) \\ \dot{x}_s(t) \end{bmatrix} &= \begin{bmatrix} A + BF & BF_s \\ -C & 0 \end{bmatrix} \begin{bmatrix} x(t) \\ x_s(t) \end{bmatrix} + \begin{bmatrix} 0 \\ I \end{bmatrix} (r(t) - v(t)) \\ y(t) &= \begin{bmatrix} C & 0 \end{bmatrix} \begin{bmatrix} x(t) \\ x_s(t) \end{bmatrix} + v(t). \end{aligned}$$

All closed-loop poles can be arbitrarily assigned by state feedback $u(t) = Fx(t) + F_s x_s(t)$ if and only if $(\begin{bmatrix} A & 0 \\ C & 0 \end{bmatrix}, \begin{bmatrix} B \\ 0 \end{bmatrix})$ is controllable. This is the case if and only if

$$\begin{bmatrix} B & AB & A^2B & \dots \\ 0 & CB & CAB & \dots \end{bmatrix}$$

has full row rank. The above matrix may be decomposed as

$$\begin{bmatrix} A & B \\ C & 0 \end{bmatrix} \begin{bmatrix} 0 & B & AB & \dots \\ I_m & 0 & 0 & \dots \end{bmatrix}$$

and therefore has full row rank if and only if

1. (A, B) is controllable,
2. $n_u \geq n_y$,
3. $\text{rank} \begin{bmatrix} A & B \\ C & 0 \end{bmatrix} = n + n_y$.

(In fact the third condition implies the second.) The three conditions are exactly what Summary 3.3.17 states. If the full state x is not available for feedback then we may replace the feedback law $u(t) = Fx(t) + F_s x_s(t)$ by an approximating observer,

$$\begin{aligned} \dot{\hat{x}}(t) &= (A - KC)\hat{x}(t) + Bu(t) + Ky(t) \\ u(t) &= F\hat{x}(t) + F_s x_s(t). \end{aligned}$$

3.4. Appendix: Proofs and Derivations

Derivation of process dynamic model of the two-tank liquid flow process. A dynamic process model for the liquid flow process of figure 3.1 is derived under the following assumptions:

- The liquid is incompressible;
- The pressure in both vessels is atmospheric;
- The flow ϕ_2 depends instantaneously on the static pressure defined by the liquid level H_1 ;
- The flow ϕ_4 is the instantaneous result of a flow-controlled recycle pump;
- ϕ_3 is determined by a valve or pump outside the system boundary of the process under consideration, independent of the variables under consideration in the model.

Let ρ denote density of the liquid expressed in kg/m^3 , and let A_1 and A_2 denote the cross-sectional areas (in m^2) of each vessel, then the mass balances over each vessel and a relation for outflow are:

$$A_1 \rho \dot{h}_1(t) = \rho[\phi_1(t) + \phi_4(t) - \phi_2(t)] \quad (3.53)$$

$$A_2 \rho \dot{h}_2(t) = \rho[\phi_2(t) - \phi_4(t) - \phi_3(t)] \quad (3.54)$$

The flow rate $\phi_2(t)$ of the first tank is assumed to be a function of the level h_1 of this tank. The common model is that

$$\phi_2(t) = k \sqrt{h_1(t)}. \quad (3.55)$$

where k is a valve dependent constant. Next we linearize these equations around an assumed equilibrium state $(h_{k,0}, \phi_{i,0})$. Define the deviations from the equilibrium values with tildes, that is,

$$\begin{aligned} \phi_i(t) &= \phi_{i,0} + \tilde{\phi}_i(t), & i &= 1 \dots 4 \\ h_k(t) &= h_{k,0} + \tilde{h}_k(t), & k &= 1, 2. \end{aligned} \quad (3.56)$$

Linearization of equation (3.55) around $(h_{k,0}, \phi_{i,0})$ gives

$$\tilde{\phi}_2(t) = \frac{k}{\sqrt{h_{1,0}}} \frac{1}{2} \dot{\tilde{h}}_1(t) = \frac{\phi_{2,0}}{2h_{1,0}} \dot{\tilde{h}}_1(t). \quad (3.57)$$

Inserting this relation into equations (3.53), (3.54) yields the linearized equations

$$\begin{bmatrix} \dot{\tilde{h}}_1(t) \\ \dot{\tilde{h}}_2(t) \end{bmatrix} = \begin{bmatrix} -\frac{\phi_{2,0}}{2A_1 h_{1,0}} & 0 \\ \frac{\phi_{2,0}}{2A_1 h_{1,0}} & 0 \end{bmatrix} \begin{bmatrix} \tilde{h}_1(t) \\ \tilde{h}_2(t) \end{bmatrix} + \begin{bmatrix} \frac{1}{A_1} & \frac{1}{A_1} \\ 0 & -\frac{1}{A_2} \end{bmatrix} \begin{bmatrix} \tilde{\phi}_1(t) \\ \tilde{\phi}_4(t) \end{bmatrix} + \begin{bmatrix} 0 \\ -\frac{1}{A_2} \end{bmatrix} \tilde{\phi}_3(t) \quad (3.58)$$

Assuming that $\phi_{4,0} = \phi_{1,0}$, and thus $\phi_{2,0} = 2\phi_{1,0}$, and taking numerical values $\phi_{1,0} = 1$, $A_1 = A_2 = 1$, and $h_{1,0} = 1$ leads to Equation (3.1). ■

Proof of Algorithm 3.2.12. We only prove the square case, i.e., where D and K are square. As ρ goes to infinity, the zeros s of

$$\begin{bmatrix} A - sI & B & 0 \\ C & D & I \\ 0 & \frac{1}{\rho}I & K \end{bmatrix}$$

converge to the transmission zeros. Now apply the following two elementary operations to the above matrix: subtract ρK times the second column from the third column and then interchange these columns. Then we get the matrix

$$\begin{bmatrix} A - sI & -\rho BK & B \\ C & I - \rho DK & D \\ 0 & 0 & \frac{1}{\rho}I \end{bmatrix}.$$

In this form it is clear that its zeros are the zeros of $\begin{bmatrix} A-sI & -\rho B \\ C & I-\rho DK \end{bmatrix}$. However, since

$$\begin{bmatrix} A-sI & -\rho BK \\ C & I-\rho DK \end{bmatrix} \underbrace{\begin{bmatrix} I & 0 \\ -(I-\rho DK)^{-1}C & I \end{bmatrix}}_{\text{nonsingular}} = \begin{bmatrix} A + BK(\frac{1}{\rho}I - DK)^{-1}C - sI & \rho B \\ 0 & I - \rho DK \end{bmatrix}$$

we see that these zeros are simply the eigenvalues of $A + BK(\frac{1}{\rho}I - DK)^{-1}C$. ■

Proof of Summary 3.3.4. We make use of the fact that if

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t), & x(0) &= 0, \\ y(t) &= Cx(t) \end{aligned}$$

then the derivative $\dot{y}(t)$ satisfies

$$\dot{y}(t) = C\dot{x}(t) = CAx(t) + CBu(t)$$

The matrix $D_f(s) = \text{diag}(s^{f_1}, \dots, s^{f_m})$ is a diagonal matrix of *differentiators*. Since the plant $y = Pu$ satisfies

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t), & x(0) &= 0, \\ y(t) &= Cx(t) + Du(t) \end{aligned} \tag{3.59}$$

it follows that the component-wise differentiated v defined as $v := D_f y$ satisfies

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t), & x(0) &= 0, \\ v(t) &= \begin{bmatrix} \frac{d^{f_1}}{dt^{f_1}} y_1(t) \\ \frac{d^{f_2}}{dt^{f_2}} y_2(t) \\ \vdots \\ \frac{d^{f_m}}{dt^{f_m}} y_m(t) \end{bmatrix} = \underbrace{\begin{bmatrix} C_{1*} A^{f_1} \\ C_{2*} A^{f_2} \\ \vdots \\ C_{m*} A^{f_m} \end{bmatrix}}_C x(t) + Du(t) \end{aligned}$$

By assumption \mathcal{D} is nonsingular, so $u(t) = \mathcal{D}^{-1}(-Cx(t) + v(t))$. Inserting this in the state realization (3.59) yields

$$\begin{aligned} \dot{x}(t) &= (A - B\mathcal{D}^{-1}C)x(t) + B\mathcal{D}^{-1}v(t), & x(0) &= 0, \\ y(t) &= (C - D\mathcal{D}^{-1}C)x(t) + D\mathcal{D}^{-1}v(t) \end{aligned}$$

Since $v = D_f y$ it must be that the above is a realization of the D_f^{-1} . ■

Proof of Lemma 3.2.8 (sketch). We only prove it for the case that none of the transmission zeros s_0 are poles. Then $(A - s_0 I_n)^{-1}$ exists for any transmission zero, and s_0 is a transmission zero if and only if $P(s_0)$ drops rank. Since

$$\begin{bmatrix} A - s_0 I_n & B \\ C & D \end{bmatrix} \begin{bmatrix} I_n & (s_0 I_n - A)^{-1} B \\ 0 & I \end{bmatrix} = \begin{bmatrix} A - s_0 I_n & 0 \\ C & P(s_0) \end{bmatrix}$$

we see that the rank of $\begin{bmatrix} A - s_0 I_n & B \\ C & D \end{bmatrix}$ equals n plus the rank of $P(s_0)$. Hence s_0 is a transmission zero of P if and only if $\begin{bmatrix} A - s_0 I_n & B \\ C & D \end{bmatrix}$ drops rank. ■

3.5. Exercises

3.1 Poles and transmission zeros.

- Find the poles and transmission zeros of $\frac{s}{s^2}$ (trick question).
- Find the poles and transmission zeros of

$$\begin{bmatrix} \frac{1}{s+3} & \frac{2}{(s-2)(s+3)} \\ \frac{-2}{(s-2)(s+3)} & \frac{s+2}{(s-2)(s+3)} \end{bmatrix}$$

using the Smith-McMillan form.

- Find the transmission zeros of

$$\begin{bmatrix} \frac{1}{s+3} & \frac{2}{(s-2)(s+3)} \\ \frac{-2}{(s-2)(s+3)} & \frac{s+2}{(s-2)(s+3)} \end{bmatrix}$$

using the inverse of this rational matrix.

3.2 Blocking property of transmission zeros Consider the system $y = Pu$ and assume that P has minimal realization $P(s) = C(sI - A)^{-1}B + D$ and that P has full column rank, that is $\text{rank } P = n_u$.

- Show that with any transmission zero s_0 there is a $u_0 \in \mathbb{C}^{n_u}$ such that for the exponential input $u(t) = u_0 e^{s_0 t}$ and appropriate initial state $x(0)$ the output $y(t)$ is identically zero for all time.
- Show the converse: If an exponential input $u(t) = u_0 e^{s_0 t}$ and initial state $x(0)$ exist such that $y(t) = 0$ for all t then s_0 is a transmission zero.

3.3 Non-minimum phase loop gain. In Example 3.1.2 one decoupling precompensator is found that diagonalizes the loop gain PK with the property that both diagonal elements of PK have a right half-plane zero at $s = 1$. Show that every internally stabilizing decoupling controller K has this property (unless K is singular). [See Definition 1.3.4.]

3.4 Internal model control & MIMO disturbance rejection. Consider the MIMO system shown in Fig. 3.7 and suppose that the plant P is stable.

- Show that the closed loop is internally stable if and only if $Q := K(I + PK)^{-1}$ is a stable transfer matrix.
- Express K and $S := (I + PK)^{-1}$ in terms of P and Q . Why is it useful to express K and S in terms of P and Q ?

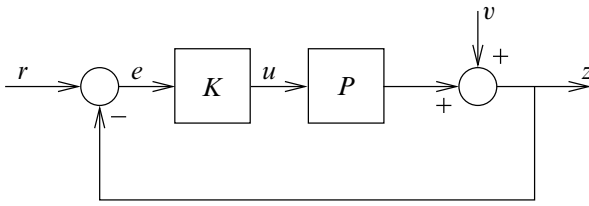


Figure 3.7: Disturbance rejection for MIMO systems

- c) Let P be the 2×2 system $P(s) = \frac{1}{s+1} \begin{bmatrix} 1 & s \\ -s & 1 \end{bmatrix}$ and suppose that v is a persistent harmonic disturbance $v(t) = \begin{bmatrix} 1 \\ 2 \end{bmatrix} e^{i\omega_0 t}$ for some known ω_0 .
- Assume $\omega_0 \neq \pm 1$. Find a stabilizing K (using the parameterization by stable Q) such that in closed loop the effect of $v(t)$ on $z(t)$ is asymptotically rejected (i.e., $e(t) \rightarrow 0$ as $t \rightarrow \infty$ for $r(t) = 0$).
 - Does there exist such a stabilizing K if $\omega_0 = \pm 1$? (Explain.)

3.5 *Realization of inverse.* Verify Eqn. (3.29).

3.6 *An invertible system that is not decouplable by state feedback.* Consider the plant with transfer matrix

$$P(s) = \begin{bmatrix} 1/(s+1) & 1/s \\ 1/(s-1) & 1/s \end{bmatrix}.$$

- Find the transmission zeros.
- Show that the methods of Summary 3.3.4 and Summary 3.3.5 are not applicable here.
- Find a stable precompensator K_0 that renders the product PK_0 diagonal.

3.7 Determine the zeros of

$$P(s) = \begin{bmatrix} \frac{s^2 + 6s + 7}{(s+2)(s+3)} & \frac{s+1}{s+2} \\ \frac{1}{s+3} & 1 \end{bmatrix}.$$

Does the method of § 3.3.2 apply to this plant?

4. LQ, LQG and \mathcal{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.

In LQG disturbances and measurement noise are modeled as stochastic processes. The \mathcal{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.

4.1. 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 \mathcal{H}_2 theory [Doyle et al. \(1989\)](#), in the wake of the attention for the so-called \mathcal{H}_∞ control theory.

In the present chapter we present a short overview of a number of results of LQG and \mathcal{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 4.2](#) (p. 136). 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 4.3](#) (p. 145) 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 4.4 (p. 152) it is demonstrated that LQG optimization amounts to minimization of the \mathcal{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 \mathcal{H}_2 interpretation naturally leads to \mathcal{H}_2 optimization with frequency dependent weighting functions. These permit a great deal of extra flexibility and make \mathcal{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 \mathcal{H}_2 framework. In Section 4.6 (p. 161) we present both a SISO and a MIMO design example.

In Section 4.7 (p. 167) a number of proofs for this chapter are collected.

4.2. LQ theory

4.2.1. 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

$$\begin{aligned}\dot{x}(t) &= Ax(t) + Bu(t), \\ z(t) &= Dx(t),\end{aligned}\quad t \geq 0. \quad (4.1)$$

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

$$\mathcal{J} = \int_0^\infty [z^T(t)Qz(t) + u^T(t)Ru(t)] dt. \quad (4.2)$$

Q and R are symmetric weighting matrices, that is, $Q = Q^T$ and $R = R^T$. Often it is adequate to let the two matrices simply be diagonal.

The two terms $z^T(t)Qz(t)$ and $u^T(t)Ru(t)$ are quadratic forms in the components of the output z and the input u , respectively. The first term in the integral criterion (4.2) 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-definite¹. 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 (4.2) is minimal along all possible trajectories of the system is the *optimal linear regulator problem*.

4.2.2. 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

¹An $n \times n$ symmetric matrix R is nonnegative-definite if $x^T R x \geq 0$ for every n -dimensional vector x . R is positive-definite if $x^T R x > 0$ for all nonzero x .

generated by choosing the input for $t \geq 0$ as

$$u(t) = -Fx(t). \quad (4.3)$$

This solution requires that the state $x(t)$ be fully accessible for measurement at all times. We return to this unreasonable assumption in § 4.3 (p. 145). The $k \times n$ state feedback gain matrix F is given by

$$F = R^{-1} B^T X. \quad (4.4)$$

The symmetric $n \times n$ matrix X is the nonnegative-definite solution of the *algebraic matrix Riccati equation* (ARE)

$$A^T X + XA + D^T QD - XBR^{-1} B^T X = 0. \quad (4.5)$$

The proof is sketched in § 4.7 (p. 167), the appendix to this chapter. The solution of the algebraic Riccati equation is discussed in § 4.2.9 (p. 144).

We summarize a number of well-known important facts about the solution of the LQ problem. An outline of the proof is given in § 4.7.1 (p. 167).

Summary 4.2.1 (Properties of the solution of the optimal linear regulator problem).

Assumptions:

- The system (4.1) is stabilizable² and detectable³. 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)

$$A^T X + XA + D^T QD - XBR^{-1} B^T X = 0 \quad (4.6)$$

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 (4.2) is $\mathcal{J}_{\min} = x^T(0)Xx(0)$.
3. The minimal value of the performance index is achieved by the feedback control law

$$u(t) = -Fx(t), \quad t \geq 0, \quad (4.7)$$

with $F = R^{-1} B^T X$.

4. The closed-loop system

$$\dot{x}(t) = (A - BF)x(t), \quad t \geq 0, \quad (4.8)$$

is stable, that is, all the eigenvalues of the matrix $A - BF$ have strictly negative real parts.

²That 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.

³That is, there exists a matrix K such that the system $\dot{e}(t) = (A - KD)e(t)$ is stable.

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.

4.2.3. 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

$$Q = \begin{bmatrix} Q_1 & 0 & 0 & \cdots & 0 \\ 0 & Q_2 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & \cdots & 0 & Q_m \end{bmatrix}, \quad R = \begin{bmatrix} R_1 & 0 & 0 & \cdots & 0 \\ 0 & R_2 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & \cdots & 0 & R_k \end{bmatrix}, \quad (4.9)$$

where Q and R have positive diagonal entries such that

$$\sqrt{Q_i} = \frac{1}{z_i^{\max}}, \quad i = 1, 2, \dots, m, \quad \sqrt{R_i} = \frac{1}{u_i^{\max}}, \quad i = 1, 2, \dots, k. \quad (4.10)$$

The number z_i^{\max} denotes the maximally acceptable deviation value for the i th component of the output z . The other quantity u_i^{\max} has a similar meaning for the i th 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.

4.2.4. 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

$$\mathcal{J}_\alpha = \int_0^\infty e^{2\alpha t} [z^T(t)Qz(t) + u^T(t)Ru(t)] dt, \quad (4.11)$$

with α a real number. Define

$$x_\alpha(t) = x(t)e^{\alpha t}, \quad u_\alpha(t) = u(t)e^{\alpha t}, \quad t \geq 0. \quad (4.12)$$

These signals satisfy

$$\dot{x}_\alpha(t) = (A + \alpha I)x_\alpha(t) + Bu_\alpha(t) \quad (4.13)$$

and

$$\mathcal{J}_\alpha = \int_0^\infty [z_\alpha^T(t)Qz_\alpha(t) + u_\alpha^T(t)Ru_\alpha(t)] dt. \quad (4.14)$$

Consequently, the minimizing u_α is

$$u_\alpha(t) = -R^{-1}B^T X_\alpha x_\alpha(t), \quad (4.15)$$

or

$$u(t) = -F_\alpha x(t), \quad (4.16)$$

with $F_\alpha = R^{-1}B^T X_\alpha$. X_α is the positive-definite solution of the modified algebraic Riccati equation

$$(A^T + \alpha I)X_\alpha + X_\alpha(A + \alpha I) + D^T Q D - X_\alpha B R^{-1} B^T X_\alpha = 0. \quad (4.17)$$

The stabilizing property of the optimal solution of the modified problem implies that

$$\operatorname{Re} \lambda_i(A + \alpha I - B F_\alpha) < 0, \quad i = 1, 2, \dots, n, \quad (4.18)$$

with $\lambda_i(A + \alpha I - B F_\alpha)$ denoting the i th eigenvalue. Application of the control law (4.16) to the system (4.1) creates a closed-loop system matrix $\bar{A}_\alpha = A - B F_\alpha$. It follows from (4.18) that its eigenvalues satisfy

$$\operatorname{Re} \lambda_i(\bar{A}_\alpha) < -\alpha. \quad (4.19)$$

Thus, choosing α positive results in an optimal closed-loop system with a *prescribed degree of stability*.

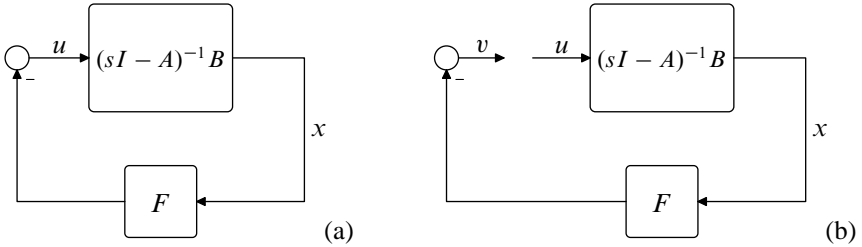


Figure 4.1: State feedback

4.2.5. Return difference equality and inequality

Figure 4.1(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. 4.1(b) then the loop gain is

$$L(s) = F(sI - A)^{-1}B. \quad (4.20)$$

The quantity

$$J(s) = I + L(s) \quad (4.21)$$

is known as the *return difference*, $J(s)u$ is the difference between the signal u in Fig. 4.1(b) and the “returned” signal $v = -L(s)u$.

Several properties of the closed-loop system may be related to the return difference. Consider

$$\det J(s) = \det[I + L(s)] = \det[I + F(sI - A)^{-1}B]. \quad (4.22)$$

Using the well-known matrix equality $\det(I + MN) = \det(I + NM)$ we obtain

$$\begin{aligned}\det J(s) &= \det[I + (sI - A)^{-1}BF] \\ &= \det(sI - A)^{-1} \det(sI - A + BF) \\ &= \frac{\det(sI - A + BF)}{\det(sI - A)} = \frac{\chi_{\text{cl}}(s)}{\chi_{\text{ol}}(s)}.\end{aligned}\tag{4.23}$$

The quantities $\chi_{\text{ol}}(s) = \det(sI - A)$ and $\chi_{\text{cl}}(s) = \det(sI - A + BF)$ are the open- and the closed-loop characteristic polynomial, respectively. We found the same result in § 1.3 (p. 11) in the form $\chi_{\text{cl}}(s) = \chi_{\text{ol}}(s) \det[I + L(s)]$.

Suppose that the gain matrix F is optimal as in Summary 4.2.1. It is proved in § 4.7.2 (p. 171) by manipulation of the algebraic Riccati equation (4.6) that the corresponding return difference satisfies the equality

$$J^T(-s)RJ(s) = R + G^T(-s)QG(s).\tag{4.24}$$

$G(s) = D(sI - A)^{-1}B$ is the *open-loop transfer matrix* of the system (4.1).

The relation (4.24) is known as the *return difference equality* or as the *Kalman-Yakubovič-Popov (KYP)* equality, after its discoverers. In Subsection 4.2.6 (p. 140) 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

$$J^T(-j\omega)RJ(j\omega) \geq R \quad \text{for all } \omega \in \mathbb{R}.\tag{4.25}$$

In Subsection 4.2.7 (p. 143) we apply the return difference inequality to establish a well-known robustness property of the optimal state feedback system.

4.2.6. Asymptotic performance weighting

For simplicity we first consider the case that (4.1) is a SISO system. To reflect this in the notation we rewrite the system equations (4.1) in the form

$$\begin{aligned}\dot{x}(t) &= Ax(t) + bu(t), \\ z(t) &= dx(t),\end{aligned}\tag{4.26}$$

with b a column vector and d a row vector. Similarly, we represent the optimal state feedback controller as

$$u(t) = -f x(t),\tag{4.27}$$

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

$$\mathcal{J} = \int_0^\infty [z^2(t) + \rho u^2(t)] dt,\tag{4.28}$$

with ρ a positive number. This amounts to setting $Q = 1$ and $R = \rho$.

⁴If P and Q are symmetric matrices of the same dimensions then $P \geq Q$ means that $x^T P x \geq x^T Q x$ for every real n -dimensional vector x .

Under these assumptions the return difference equality (4.24) reduces to

$$J(-s)J(s) = 1 + \frac{1}{\rho}G(s)G(-s). \quad (4.29)$$

From (4.25) we have

$$J(s) = \frac{\chi_{cl}(s)}{\chi_{ol}(s)}, \quad (4.30)$$

with χ_{cl} the closed-loop characteristic polynomial and χ_{ol} the open-loop characteristic polynomial. We furthermore write

$$G(s) = \frac{k\psi(s)}{\chi_{ol}(s)}. \quad (4.31)$$

The constant k is chosen such that the polynomial ψ is *monic*⁵. From (4.29–4.31) we now obtain

$$\chi_{cl}(-s)\chi_{cl}(s) = \chi_{ol}(-s)\chi_{ol}(s) + \frac{k^2}{\rho}\psi(-s)\psi(s). \quad (4.32)$$

The left-hand side $\chi_{cl}(-s)\chi_{cl}(s)$ of this relation defines a polynomial whose roots consists of the closed-loop poles (the roots of $\chi_{cl}(s)$) together with their *mirror images* with respect to the imaginary axis (the roots of $\chi_{cl}(-s)$). 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 (4.32) 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 $\rho \rightarrow \infty$ then the closed-loop poles and their mirror images approach the roots of $\chi_{ol}(s)\chi_{ol}(-s)$. This means that the closed-loop poles approach

- 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 $\rho \rightarrow \infty$ 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 $\rho \rightarrow \infty$ 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.

⁵That is, the coefficient of the term of highest degree is 1.

- 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 (4.32) reduces to

$$(-s)^n s^n + \frac{k^2}{\rho} (-s)^q s^q, \quad (4.33)$$

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

$$s^{2(n-q)} + (-1)^{n-q} \frac{k^2}{\rho} = 0. \quad (4.34)$$

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

$$\omega_c = \left(\frac{k^2}{\rho} \right)^{\frac{1}{2(n-q)}}. \quad (4.35)$$

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). 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 $\rho \rightarrow \infty$ 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 $\det G^T(-s)QG(s)$.

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 increases without bound as $\rho \downarrow 0$. Correspondingly, the solution X of the Riccati equation approaches the zero matrix.

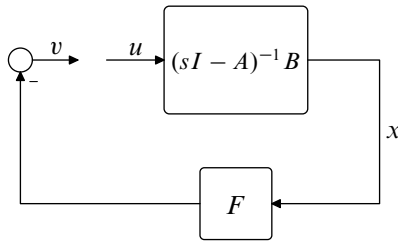


Figure 4.2: Loop gain for state feedback

4.2.7. Guaranteed gain and phase margins

If the state feedback loop is opened at the plant input as in Fig. 4.2 then the loop gain is $L(s) = F(sI - A)^{-1}B$. For the single-input case discussed in Subsection 4.2.6 (see Eqn. (4.29)) the return difference inequality (4.25) takes the form

$$|1 + L(j\omega)| \geq 1, \quad \omega \in \mathbb{R}. \quad (4.36)$$

This inequality implies that the Nyquist plot of the loop gain stays outside the circle with center at -1 and radius 1. Figure 4.3 shows two possible behaviors of the Nyquist plot.

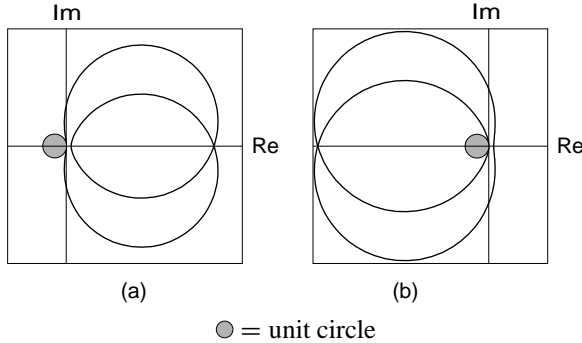


Figure 4.3: Examples of Nyquist plots of optimal loop gains. (a): Open-loop stable plant. (b): Open-loop unstable plant with one right-half plane pole

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 ∞ . 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 (4.25). 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 4.7.3 (p. 172) of § 4.7, the appendix to this chapter, that the closed-loop *remains* stable provided

$$RW(j\omega) + W^T(-j\omega)R > R, \quad \omega \in \mathbb{R}. \quad (4.37)$$

If both R and W are diagonal then this reduces to

$$W(j\omega) + W^T(-j\omega) > I, \quad \omega \in \mathbb{R}. \quad (4.38)$$

This shows that if the i th 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 i th diagonal entry is $W_i(j\omega) = e^{j\phi}$ 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.

4.2.8. Cross term in the performance index

In the optimal regulator problem for the stabilizable and detectable system

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t), \\ z(t) &= Dx(t), \end{aligned} \quad t \geq 0, \quad (4.39)$$

we may consider the generalized quadratic performance index

$$\mathcal{J} = \int_0^\infty \begin{bmatrix} z^T(t) & u^T(t) \end{bmatrix} \begin{bmatrix} Q & S \\ S^T & R \end{bmatrix} \begin{bmatrix} z(t) \\ u(t) \end{bmatrix} dt. \quad (4.40)$$

We assume that

$$\begin{bmatrix} Q & S \\ S^T & R \end{bmatrix} \quad (4.41)$$

is positive-definite. Define $v(t) = u(t) + R^{-1}S^T z(t)$. Then minimization of \mathcal{J} is equivalent to minimizing

$$\mathcal{J} = \int_0^\infty [z^T(t)(Q - SR^{-1}S^T)z(t) + v^T(t)Rv(t)] dt \quad (4.42)$$

for the system

$$\dot{x}(t) = (A - BR^{-1}S^T D)x(t) + Bv(t). \quad (4.43)$$

The condition that (4.41) be positive-definite is equivalent to the condition that both R and $Q - SR^{-1}S^T$ be positive-definite (see Exercise 4.6, p. 180). Thus we satisfy the conditions of Summary 4.2.1. The Riccati equation now is

$$A^T X + XA + D^T QD - (XB + D^T S)R^{-1}(B^T X + S^T D) = 0. \quad (4.44)$$

The optimal input for the system (4.39) is $u(t) = -Fx(t)$, with

$$F = R^{-1}(B^T X + S^T D). \quad (4.45)$$

4.2.9. Solution of the ARE

There are several algorithms for the solution of the algebraic Riccati equation (4.6) or (4.44). For all but the simplest problem recourse needs to be taken to numerical computer calculation. Equation (4.44) is the most general form of the Riccati equation. By redefining $D^T QD$ as Q and $D^T S$ as S the ARE (4.44) reduces to

$$A^T X + XA + Q - (XB + S)R^{-1}(B^T X + S^T) = 0. \quad (4.46)$$

The most dependable solution method relies on the *Hamiltonian matrix*

$$\mathcal{H} = \begin{bmatrix} A - BR^{-1}S^T & -BR^{-1}B^T \\ -Q + SR^{-1}S^T & -(A - BR^{-1}S^T)^T \end{bmatrix} \quad (4.47)$$

associated with the LQ problem. Under the assumptions of Summary 4.2.1 (p. 137) or § 4.2.7 (p. 143) the Hamiltonian matrix \mathcal{H} has no eigenvalues on the imaginary axis. If λ is an eigenvalue of the $2n \times 2n$ matrix \mathcal{H} then $-\lambda$ is also an eigenvalue. Hence, \mathcal{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 \mathcal{H} corresponding to the eigenvalues with strictly negative real parts. Partition

$$E = \begin{bmatrix} E_1 \\ E_2 \end{bmatrix}, \quad (4.48)$$

with E_1 and E_2 both square. It is proved in § 4.7.4 (p. 173) that

$$X = E_2 E_1^{-1} \quad (4.49)$$

is the desired solution of the algebraic Riccati equation.

E may efficiently be computed by Schur decomposition (Golub and Van Loan, 1983)

$$\mathcal{H} = UTU^H \quad (4.50)$$

of the Hamiltonian matrix \mathcal{H} . U is unitary, that is, $UU^H = U^H U = I$. 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 \mathcal{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 \mathcal{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).

4.2.10. Concluding remarks

The LQ paradigm might 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 4.3 (p. 145) 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.

4.3. LQG Theory

4.3.1. Introduction

In this section we review what is known as LQG theory. LQG stands for Linear Quadratic Gaussian. 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

$$\left. \begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t) + Gv(t) \\ y(t) &= Cx(t) + w(t) \\ z(t) &= Dx(t) \end{aligned} \right\} \quad t \in \mathbb{R}. \quad (4.51)$$

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

$$\left. \begin{aligned} Ev(t)v^T(s) &= V\delta(t-s) \\ Ev(t)w^T(s) &= 0 \\ Ew(t)w^T(s) &= W\delta(t-s) \end{aligned} \right\} \quad t, s \in \mathbb{R}. \quad (4.52)$$

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

$$z^T(t)Qz(t) + u^T(t)Ru(t), \quad t \geq 0, \quad (4.53)$$

is a random process. The problem of controlling the system such that the integrated expected value

$$\int_0^T E[z^T(t)Qz(t) + u^T(t)Ru(t)] dt \quad (4.54)$$

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 $T \rightarrow \infty$. At any time t the entire past measurement signal $y(s)$, $s \leq t$, is assumed to be available for feedback. Figure 4.4 clarifies the situation.

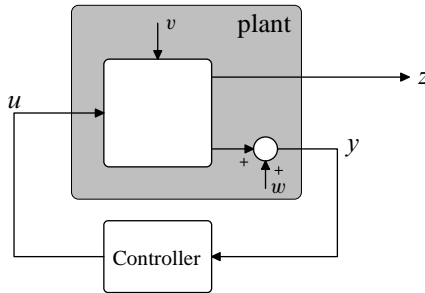


Figure 4.4: LQG feedback

4.3.2. Observers

Consider the observed system

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t), \\ y(t) &= Cx(t), \end{aligned} \quad t \in \mathbb{R}. \quad (4.55)$$

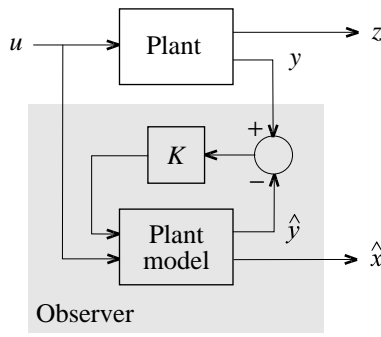


Figure 4.5: Structure of an observer

This is the system (4.51) but without the state noise v and the measurement noise w . The state x of the system (4.55) 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

$$\dot{\hat{x}}(t) = A\hat{x}(t) + Bu(t) + K[y(t) - C\hat{x}(t)], \quad t \in \mathbb{R}. \quad (4.56)$$

The signal \hat{x} is meant to be an estimate of the state $x(t)$. It satisfies the state differential equation of the system (4.55) 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 (4.56) provides a correction that is active as soon as the observation error is nonzero. Figure 4.5 shows the structure of the observer. Define

$$e(t) = \hat{x}(t) - x(t) \quad (4.57)$$

as the *state estimation error*. Differentiation of e yields after substitution of (4.56) and (4.55) that the error satisfies the differential equation

$$\dot{e}(t) = (A - KC)e(t), \quad t \in \mathbb{R}. \quad (4.58)$$

If the system (4.55) is detectable then there always exists a gain matrix K such that the error system (4.58) is stable. If the error system is stable then $e(t) \rightarrow 0$ as $t \rightarrow \infty$ for any initial error $e(0)$. Hence,

$$\hat{x}(t) \xrightarrow{t \rightarrow \infty} x(t), \quad (4.59)$$

so that the estimated state converges to the actual state.

4.3.3. The Kalman filter

Suppose that we connect the observer

$$\dot{\hat{x}}(t) = A\hat{x}(t) + Bu(t) + K[y(t) - C\hat{x}(t)], \quad t \in \mathbb{R}. \quad (4.60)$$

to the noisy system

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t) + Gv(t), \\ y(t) &= Cx(t) + w(t), \end{aligned} \quad t \in \mathbb{R}. \quad (4.61)$$

Differentiation of $e(t) = \hat{x}(t) - x(t)$ leads to the error differential equation

$$\dot{e}(t) = (A - KC)e(t) - Gv(t) + Kw(t), \quad t \in \mathbb{R}. \quad (4.62)$$

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 § 4.7.7 (p. 176) that as $t \rightarrow \infty$ the *error covariance matrix*

$$Ee(t)e^T(t) \quad (4.63)$$

converges to a constant steady-state value Y that satisfies the linear matrix equation

$$(A - KC)Y + Y(A - KC)^T + GVG^T + KWK^T = 0. \quad (4.64)$$

This type of matrix equation is known as a *Lyapunov equation*. It is made plausible in Subsection 4.7.5 that as a function of the gain matrix K the steady-state error covariance matrix Y is *minimal* if K is chosen as

$$K = YC^TW^{-1}. \quad (4.65)$$

“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 (4.65) minimizes the steady-state mean square state reconstruction error $\lim_{t \rightarrow \infty} Ee^T(t)e(t)$. As a matter of fact, the gain minimizes the weighted mean square construction error $\lim_{t \rightarrow \infty} Ee^T(t)W_e e(t)$ for any nonnegative-definite weighting matrix W_e .

Substitution of the optimal gain matrix (4.65) into the Lyapunov equation (4.64) yields

$$AY + YA^T + GVG^T - YC^TW^{-1}CY = 0. \quad (4.66)$$

This is another matrix Riccati equation. The observer

$$\dot{\hat{x}}(t) = A\hat{x}(t) + Bu(t) + K[y(t) - C\hat{x}(t)], \quad t \in \mathbb{R}, \quad (4.67)$$

with the gain chosen as in (4.65) and the covariance matrix Y the nonnegative-definite solution of the Riccati equation (4.66) 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 4.2.1 (p. 137) for the Riccati equation associated with the regulator problem⁶.

Summary 4.3.1 (Properties of the Kalman filter).

Assumptions:

- The system

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Gv(t), \\ y(t) &= Cx(t), \end{aligned} \quad t \in \mathbb{R}, \quad (4.68)$$

is stabilizable and detectable.

⁶The optimal regulator and the Kalman filter are dual in the following sense. Given the regulator problem of § 4.2 (p. 136), 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 (4.6) becomes the observer Riccati equation (4.66), 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 4.2.1 (p. 137) by duality:

1. The algebraic Riccati equation

$$AY + YA^T + GVG^T - YC^T W^{-1} CY = 0 \quad (4.69)$$

has a unique nonnegative-definite symmetric solution Y . If the system (4.68) 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 $\lim_{t \rightarrow \infty} E e^T(t) W_e e(t)$ is $\text{tr } Y W_e$.
3. The minimal value of the mean square reconstruction error is achieved by the observer gain matrix $K = YC^T W^{-1}$.
4. The error system

$$\dot{e}(t) = (A - KC)e(t), \quad t \in \mathbb{R}, \quad (4.70)$$

is stable, that is, all the eigenvalues of the matrix $A - KC$ have strictly negative real parts. As a consequence also the observer

$$\dot{\hat{x}}(t) = A\hat{x}(t) + Bu(t) + K[y(t) - C\hat{x}(t)], \quad t \in \mathbb{R}, \quad (4.71)$$

is stable.

□

The reasons for the assumptions may be explained as follows. If the system (4.68) 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.

4.3.4. 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

$$E \begin{bmatrix} v(t) \\ w(t) \end{bmatrix} \begin{bmatrix} v^T(s) & w^T(s) \end{bmatrix} = \begin{bmatrix} V & U \\ U^T & W \end{bmatrix} \delta(t-s), \quad t, s \in \mathbb{R}. \quad (4.72)$$

Assume that

$$\begin{bmatrix} V & U \\ U^T & W \end{bmatrix} \quad (4.73)$$

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

$$K = (YC^T + GU)W^{-1}, \quad (4.74)$$

where the steady-state error covariance matrix Y is the positive-definite solution of the Riccati equation

$$AY + YA^T + GVG^T - (YC^T + GU)W^{-1}(CY + U^T G^T) = 0. \quad (4.75)$$

⁷The quantity $\text{tr } M = \sum_{i=1}^m M_{ii}$ is called the *trace* of the $m \times m$ matrix M with entries M_{ij} , $i, j = 1, 2, \dots, m$.

4.3.5. Solution of the stochastic linear regulator problem

The stochastic linear regulator problem consists of minimizing

$$\int_0^T E[z^T(t)Qz(t) + u^T(t)Ru(t)] dt \quad (4.76)$$

for the system

$$\left. \begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t) + Gv(t) \\ y(t) &= Cx(t) + w(t) \\ z(t) &= Dx(t) \end{aligned} \right\} \quad t \in \mathbb{R}. \quad (4.77)$$

We successively consider the situation of no state noise, state feedback, and output feedback.

No state noise. From § 4.2 (p. 167) we know that if the disturbance v is absent and the state $x(t)$ may be directly and accurately accessed for measurement, then for $T \rightarrow \infty$ the performance index is minimized by the state feedback law

$$u(t) = -Fx(t), \quad (4.78)$$

with the feedback gain F as in Summary 4.2.1 (p. 137).

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 (4.76) does not converge to a finite number as $T \rightarrow \infty$. It is proved in Subsection 4.7.6 (p. 175) that the state feedback law (4.78) minimizes the *rate* at which (4.76) approaches ∞ , that is, it minimizes

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T E[z^T(t)Qz(t) + u^T(t)Ru(t)] dt. \quad (4.79)$$

This limit equals the *steady-state mean square error*

$$\lim_{t \rightarrow \infty} E[z^T(t)Qz(t) + u^T(t)Ru(t)]. \quad (4.80)$$

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 (4.78) with the *estimated state* $\hat{x}(t)$. Thus, the optimal controller is given by

$$\begin{aligned} \dot{\hat{x}}(t) &= A\hat{x}(t) + Bu(t) + K[y(t) - C\hat{x}(t)], \\ u(t) &= -F\hat{x}(t), \end{aligned} \quad t \in \mathbb{R}. \quad (4.81)$$

The controller minimizes the steady-state mean square error (4.80) under output feedback. The feedback gain F and the observer gain K follow from Summaries 4.2.1 (p. 137) and 4.3.1 (p. 148), respectively. Figure 4.6 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*.

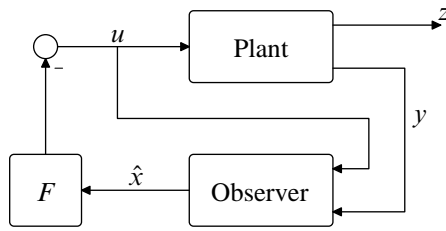


Figure 4.6: Observer based feedback control

The closed-loop system that results from interconnecting the plant (4.77) with the compensator (4.81) is stable—under the assumptions of Summaries 4.2.1 and 4.3.1, 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 $\hat{x}(t) = x(t) + e(t)$

$$\dot{x}(t) = (A - BF)x(t) - BFe(t) + Gv(t). \quad (4.82)$$

Together with (4.62) we thus have

$$\begin{bmatrix} \dot{x}(t) \\ \dot{e}(t) \end{bmatrix} = \begin{bmatrix} A - BF & -BF \\ 0 & A - KC \end{bmatrix} \begin{bmatrix} x(t) \\ e(t) \end{bmatrix} + \begin{bmatrix} Gv(t) \\ -Gv(t) + Kw(t) \end{bmatrix}. \quad (4.83)$$

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 (4.77) has order n then the compensator also has order n . Hence, there are $2n$ closed-loop poles.

4.3.6. 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. 4.7 we obtain the loop gain

$$L_\sigma(s) = C_e(s)P(s) = F(sI - A + BF + K_\sigma C)^{-1}K_\sigma C(sI - A)^{-1}B. \quad (4.84)$$

(Compare Exercise 4.11, p. 181.) To emphasize the dependence on σ the observer gain and the 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 4.2.6 (p. 140) 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 § 4.7.7 (p. 176) that as $\sigma \downarrow 0$ the loop gain L_σ approaches the expression

$$L_0(s) = F(sI - A)^{-1}B. \quad (4.85)$$

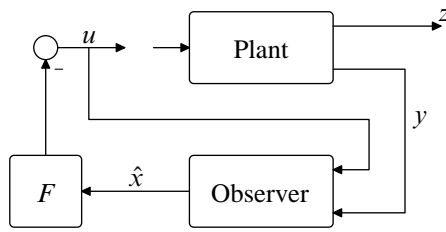


Figure 4.7: Breaking the loop

The asymptotic loop gain L_0 is precisely the loop gain for full state feedback. Accordingly, the guaranteed gain and phase margins of Subsection § 4.2.7 (p. 143) 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 et al. (1993). We use the method in the design examples of § 4.6 (p. 155).

4.4. \mathcal{H}_2 optimization

4.4.1. Introduction

In this section we define the LQG problem as a special case of a larger class of problems, which has become known as \mathcal{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 \mathcal{H}_2 -norm. To introduce this point of view, consider the stable system

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bv(t), \\ y(t) &= Cx(t), \end{aligned} \quad t \in \mathbb{R}. \quad (4.86)$$

The system has the transfer matrix $H(s) = C(sI - A)^{-1}B$. Suppose that the signal v is white noise with covariance function $Ev(t)v^T(s) = V\delta(t - s)$. Then the output y of the system is a stationary stochastic process with spectral density matrix

$$S_y(f) = H(j2\pi f)VH^T(-j2\pi f), \quad f \in \mathbb{R}. \quad (4.87)$$

As a result, the mean square output is

$$Ey^T(t)y(t) = \text{tr} \int_{-\infty}^{\infty} S_y(f) df = \text{tr} \int_{-\infty}^{\infty} H(j2\pi f)VH^T(j2\pi f) df. \quad (4.88)$$

Here we introduce the notation $H^\sim(s) = H^T(-s)$. The quantity

$$\|H\|_2 = \sqrt{\text{tr} \int_{-\infty}^{\infty} H(j2\pi f)H^\sim(j2\pi f) df} = \sqrt{\frac{1}{2\pi} \text{tr} \int_{-\infty}^{\infty} H(j\omega)H^\sim(j\omega) d\omega} \quad (4.89)$$

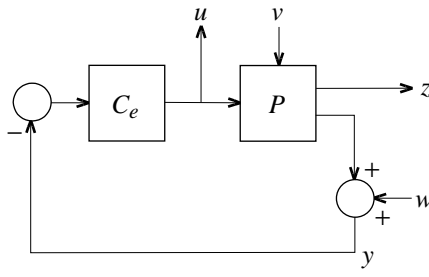


Figure 4.8: Feedback system with stochastic inputs and outputs

is called the \mathcal{H}_2 -norm of the system. If the white noise v has intensity $V = I$ then the mean square output $E y^T(t)y(t)$ equals precisely the square of the \mathcal{H}_2 -norm of the system.

\mathcal{H}_2 refers to the space of square integrable functions on the imaginary axis whose inverse Fourier transform is zero for negative time.

4.4.2. \mathcal{H}_2 optimization

In this subsection we rewrite the time domain LQG problem into an equivalent frequency domain \mathcal{H}_2 optimization problem. While the LQG problem requires state space realizations, the \mathcal{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

$$\lim_{t \rightarrow \infty} E[z^T(t)z(t) + u^T(t)u(t)]. \quad (4.90)$$

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

$$\dot{x} = Ax + Bu + Gv, \quad (4.91)$$

$$z = Dx, \quad (4.92)$$

$$y = Cx + w \quad (4.93)$$

we have in terms of transfer matrices

$$z = \underbrace{D(sI - A)^{-1}G}_{P_{11}(s)} v + \underbrace{D(sI - A)^{-1}B}_{P_{12}(s)} u, \quad (4.94)$$

$$y = \underbrace{C(sI - A)^{-1}G}_{P_{21}(s)} v + \underbrace{C(sI - A)^{-1}B}_{P_{22}(s)} u + w. \quad (4.95)$$

Interconnecting the system as in Fig. 4.8 with a compensator C_e we have the signal balance $u = -C_e y = -C_e(P_{21}v + P_{22}u + w)$, so that

$$u = \underbrace{-(I + C_e P_{22})^{-1}C_e P_{21}}_{H_{21}(s)} v - \underbrace{-(I + C_e P_{22})^{-1}C_e}_{H_{22}(s)} w. \quad (4.96)$$

From $z = P_{11}v + P_{12}u$ we obtain

$$z = \underbrace{P_{11} - P_{12}(I + C_e P_{22})^{-1}C_e P_{21}}_{H_{11}(s)} v - \underbrace{P_{12}(I + C_e P_{22})^{-1}C_e}_{H_{12}(s)} w. \quad (4.97)$$

A more compact notation is

$$\begin{bmatrix} z \\ u \end{bmatrix} = \underbrace{\begin{bmatrix} H_{11}(s) & H_{12}(s) \\ H_{21}(s) & H_{22}(s) \end{bmatrix}}_{H(s)} \begin{bmatrix} v \\ w \end{bmatrix}. \quad (4.98)$$

From this we find for the steady-state mean square error

$$\lim_{t \rightarrow \infty} E(z^T(t)z(t) + u^T(t)u(t)) = \lim_{t \rightarrow \infty} E\left(\begin{bmatrix} z(t) \\ u(t) \end{bmatrix}^T \begin{bmatrix} z(t) \\ u(t) \end{bmatrix}\right) \quad (4.99)$$

$$= \operatorname{tr} \int_{-\infty}^{\infty} H(j2\pi f) H^*(j2\pi f) df \quad (4.100)$$

$$= \|H\|_2^2. \quad (4.101)$$

Hence, solving the LQG problem amounts to minimizing the \mathcal{H}_2 norm of the closed-loop system of Fig. 4.8 with (v, w) as input and (z, u) as output.

The configuration of Fig. 4.8 is a special case of the configuration of Fig. 4.9. In the latter diagram w is the *external input* (v and w in Fig. 4.8). The signal z is the *error signal*, which ideally should be zero (z and u in Fig. 4.8). Furthermore, u is the *control input*, and y the *observed output*. The block G is the *generalized plant*, and C_e the compensator. Note that the sign reversion at the compensator input in Fig. 4.8 has been absorbed into the compensator C_e .

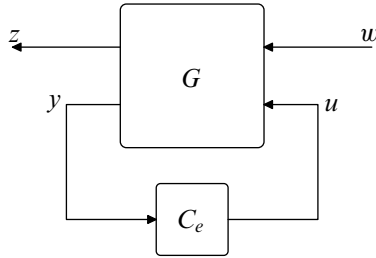


Figure 4.9: The standard \mathcal{H}_2 problem

4.4.3. The standard \mathcal{H}_2 problem and its solution

The *standard* \mathcal{H}_2 optimization problem is the problem of choosing the compensator C_e in the block diagram of Fig. 4.9 such that it

1. stabilizes the closed-loop system, and
2. minimizes the \mathcal{H}_2 -norm of the closed-loop system (with w as input and z as output).

We represent the generalized plant G of Fig. 4.9 in state space form as

$$\dot{x}(t) = Ax(t) + B_1 w(t) + B_2 u(t), \quad (4.102)$$

$$z(t) = C_1 x(t) + D_{11} w(t) + D_{12} u(t), \quad (4.103)$$

$$y(t) = C_2 x(t) + D_{21} w(t) + D_{22} u(t). \quad (4.104)$$

The \mathcal{H}_2 problem may be solved by reducing it to an LQG problem. This is done in § 4.7.8 (p. 177). The derivation necessitates the introduction of some assumptions, which are listed in the summary that follows. They are natural assumptions for LQG problems.

Summary 4.4.1 (Solution of the \mathcal{H}_2 problem). Consider the standard \mathcal{H}_2 optimization problem for the generalized plant

$$\dot{x}(t) = Ax(t) + B_1w(t) + B_2u(t), \quad (4.105)$$

$$z(t) = C_1x(t) + D_{12}u(t), \quad (4.106)$$

$$y(t) = C_2x(t) + D_{21}w(t) + D_{22}u(t). \quad (4.107)$$

Assumptions:

- The system $\dot{x}(t) = Ax(t) + B_2u(t)$, $y(t) = C_2x(t)$ is stabilizable and detectable.
- The matrix $\begin{bmatrix} A-sI & B_1 \\ C_2 & D_{21} \end{bmatrix}$ has full row rank for every $s = j\omega$, and D_{21} has full row rank.
- The matrix $\begin{bmatrix} A-sI & B_2 \\ C_1 & D_{12} \end{bmatrix}$ has full column rank for every $s = j\omega$, and D_{12} has full column rank.

Under these assumptions the optimal output feedback controller $u = C_e y$ is

$$\dot{\hat{x}}(t) = A\hat{x}(t) + B_2u(t) + K[y(t) - C_2\hat{x}(t) - D_{22}u(t)] \quad (4.108)$$

$$u(t) = -F\hat{x}(t). \quad (4.109)$$

The observer and state feedback gain matrices are

$$F = (D_{12}^T D_{12})^{-1} (B_2^T X + D_{12}^T C_1), \quad K = (YC_2^T + B_1 D_{21}^T) (D_{21} D_{21}^T)^{-1}. \quad (4.110)$$

The symmetric matrices X and Y are the unique positive-definite solutions of the algebraic Riccati equations

$$\begin{aligned} A^T X + XA + C_1^T C_1 - (XB_2 + C_1^T D_{12})(D_{12}^T D_{12})^{-1}(B_2^T X + D_{12}^T C_1) &= 0, \\ AY + AY^T + B_1 B_1^T - (YC_2^T + B_1 D_{21}^T)(D_{21} D_{21}^T)^{-1}(C_2 Y + D_{21} B_1^T) &= 0. \end{aligned} \quad (4.111)$$

□

The condition that D_{12} has full column rank means that there is “direct feedthrough” from the input u to the error signal z . Dually, the condition that D_{21} has full row rank means that the some noise w is directly fed through to the observed output y .

The \mathcal{H}_2 optimization problem and its solution are discussed at length in [Saber et al. \(1995\)](#). In §§ 4.5 (p. 155) and 4.6 (p. 161) we discuss the application of \mathcal{H}_2 optimization to control system design.

4.5. Feedback system design by \mathcal{H}_2 optimization

4.5.1. Introduction

In this section we review how LQG and \mathcal{H}_2 optimization may be used to design SISO and MIMO linear feedback systems.

4.5.2. 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

$$\lim_{t \rightarrow \infty} E[z^T(t)Qz(t) + u^T(t)Ru(t)] \quad (4.112)$$

for the system

$$\left. \begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t) + Gv(t) \\ z(t) &= Dx(t) \\ y(t) &= Cx(t) + w(t) \end{aligned} \right\} \quad t \in \mathbb{R}. \quad (4.113)$$

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 § 4.2.3 (p. 138).
 - 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 § 4.2.6, p. 140)).
In the MIMO case one may let $R = \rho R_0$, where the fixed matrix R_0 is selected according to the rules of § 4.2.3 (p. 138) 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 § 4.3.6 (p. 151).
 - 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 § 4.2.3 (p. 138). 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.

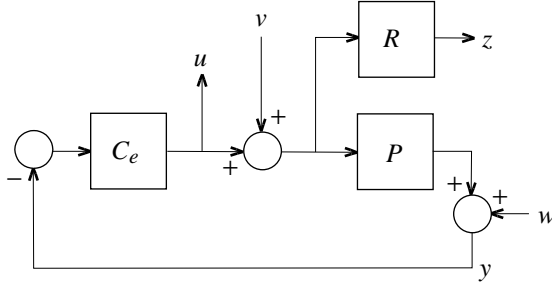


Figure 4.10: Block diagram for loop recovery design

4.5.3. \mathcal{H}_2 Interpretation

In this subsection we discuss the interpretation of the LQG problem as an \mathcal{H}_2 optimization problem.

If we choose $G = B$ to take advantage of loop transfer recovery then the open-loop equations (4.94–4.95) may be rewritten as

$$z = R(s)(u + v), \quad (4.114)$$

$$y = P(s)(u + v) + w, \quad (4.115)$$

where

$$P(s) = C(sI - A)^{-1}B, \quad R(s) = D(sI - A)^{-1}B. \quad (4.116)$$

The corresponding block diagram is represented in Fig. 4.10. If P is invertible then by block diagram substitution Fig. 4.10 may be redrawn as in Fig. 4.11, 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. 4.11. By setting up two appropriate signal balances it is easy to find that

$$z = W_0SPv - W_0Tw, \quad (4.117)$$

$$u = -T'v - Uw. \quad (4.118)$$

Here

$$S = (I + PC_e)^{-1}, \quad T = (I + PC_e)^{-1}PC_e, \quad (4.119)$$

$$T' = (I + C_eP)^{-1}C_eP, \quad U = C_e(I + PC_e)^{-1}. \quad (4.120)$$

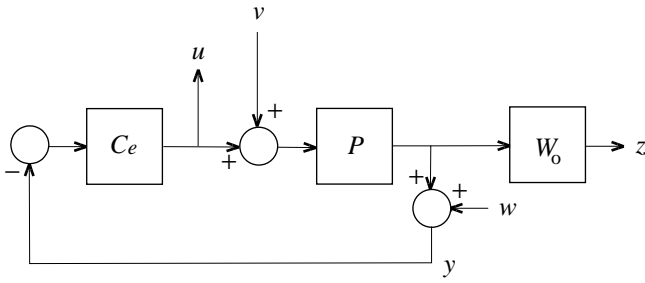


Figure 4.11: Equivalent block diagram

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 $T' = T$. From (4.117–4.118) we find for the performance index

$$\|H\|_2^2 = \text{tr} \int_{-\infty}^{\infty} (W_0 S P P^{\sim} S^{\sim} W_0^{\sim} + W_0 T T^{\sim} W_0^{\sim} + T' T'^{\sim} + U U^{\sim}) df. \quad (4.121)$$

The argument $j2\pi f$ 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 (4.121) arise from the LQG problem and are not very flexible. For more freedom we generalize the block diagram of Fig. 4.11 to that of Fig. 4.12. 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

$$z_1 = W_1 S P V_1 v - W_1 T V_2 w, \quad (4.122)$$

$$z_2 = -W_2 T' V_1 v - W_2 U V_2 w. \quad (4.123)$$

As a result, the performance index now takes the form

$$\begin{aligned} \|H\|_2^2 = \text{tr} \int_{-\infty}^{\infty} & (W_1 S P V_1 V_1^{\sim} P^{\sim} S^{\sim} W_1^{\sim} + W_1 T V_2 V_2^{\sim} T^{\sim} W_1^{\sim} \\ & + W_2 T' V_1 V_1^{\sim} T'^{\sim} W_2^{\sim} + W_2 U V_2 V_2^{\sim} U^{\sim} W_2^{\sim}) df. \end{aligned} \quad (4.124)$$

In the next subsections some applications of this generalized problem are discussed.

4.5.4. 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 (4.124). 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

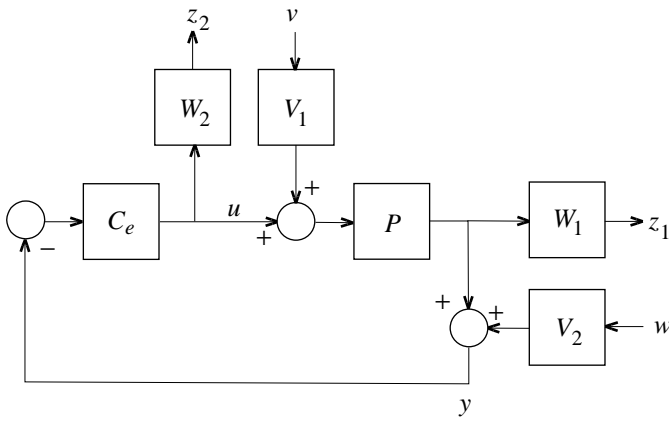


Figure 4.12: Generalized configuration

the compensator. Inspection of (4.124) 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

$$V_1(s) = \frac{s + \alpha}{s}. \quad (4.125)$$

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 (4.124) 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

$$W_2(s) = \frac{s}{s + \alpha} W_{2o}(s), \quad (4.126)$$

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 choosing V_1 as in (4.125) is that it makes the open-loop system unstabilizable, because of the integrator outside the loop. This violates one of the assumptions of § 4.4.3 (p. 154) required for the solution of the \mathcal{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) \rightarrow (b) \rightarrow (c) of Fig. 4.13. 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 (4.126). W_{2o} remains.

Additionally an extra factor $\frac{s}{s+\alpha}$ is included in the compensator. If the modified problem leads

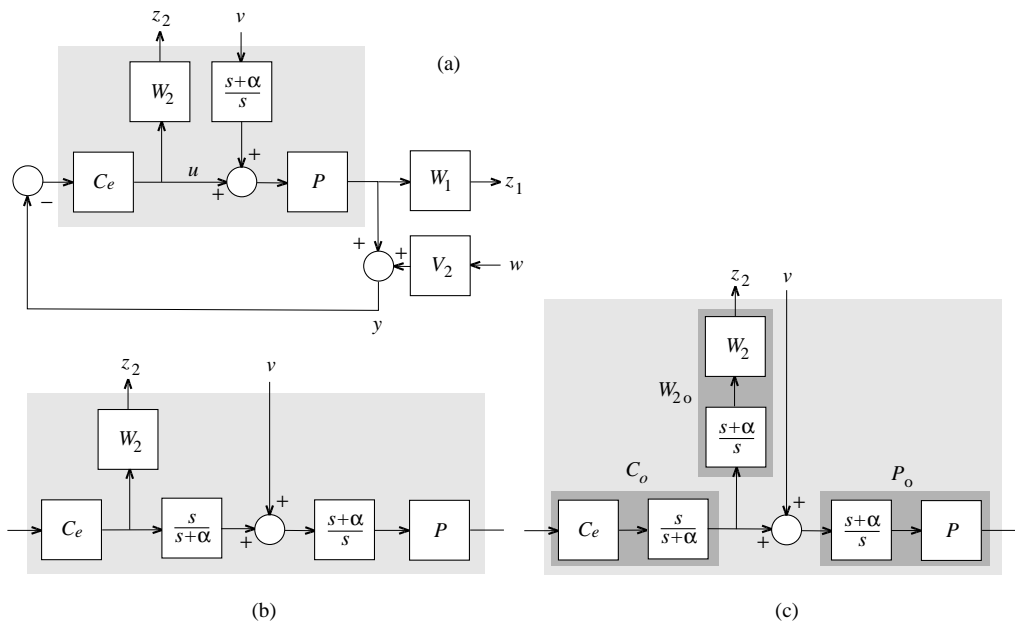


Figure 4.13: Block diagram substitutions for integral control (a)→(b)→(c)

to an optimal compensator C_0 then the optimal compensator for the original problem is

$$C(s) = \frac{s + \alpha}{s} C_0(s). \quad (4.127)$$

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 § 4.6.3 (p. 163). The method is explained in greater generality in § 6.7 (p. 258) in the context of \mathcal{H}_∞ optimization.

4.5.5. 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. For some applications it may be desirable to have a steeper high-frequency roll-off. Inspection of (4.124) 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

$$W_2(s) = \rho(1 + rs), \quad (4.128)$$

with r a positive constant. Then by inspecting the fourth term in the integrand of (4.124) 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.

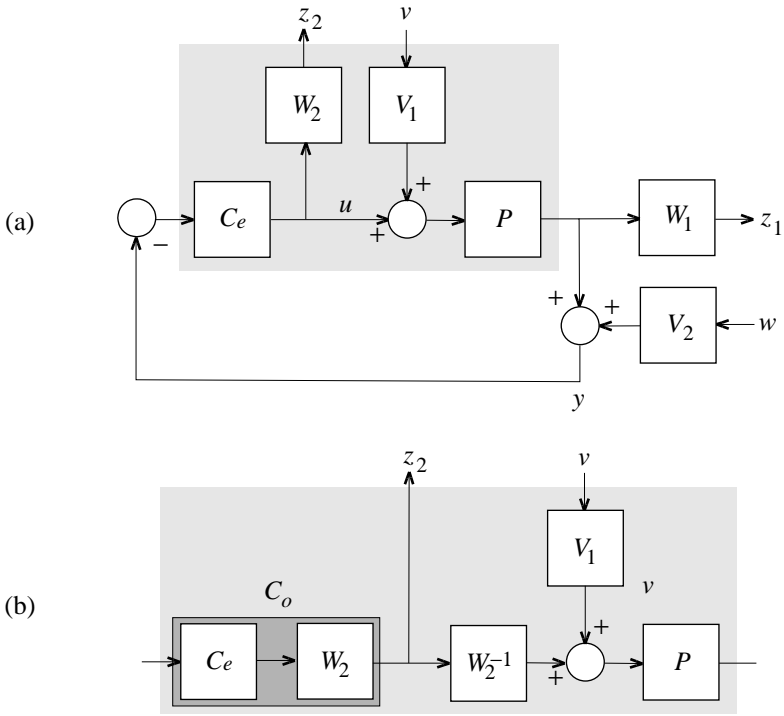


Figure 4.14: Block diagram substitution for high-frequency roll-off

The difficulty that there is no state space realization for the block W_2 with the transfer function (4.128) may be avoided by the block diagram substitution of Fig. 4.14. If the modified problem is solved by the compensator C_0 then the optimal compensator for the original problem is

$$C_e(s) = \frac{C_0(s)}{W_2(s)} = \frac{C_0(s)}{\rho(1 + rs)}. \quad (4.129)$$

The extra roll-off is apparent. Even more roll-off may be obtained by letting $W_2(s) = \mathcal{O}(s^m)$ as $|s| \rightarrow \infty$, with $m \geq 2$.

For a more general exposition of the block diagram substitution method see § 6.7 (p. 258).

4.6. Examples and applications

4.6.1. Introduction

In this section we present two design applications of \mathcal{H}_2 theory: A simple SISO system and a not very complicated MIMO system.

4.6.2. LQG design of a double integrator plant

We consider the double integrator plant

$$P(s) = \frac{1}{s^2}. \quad (4.130)$$

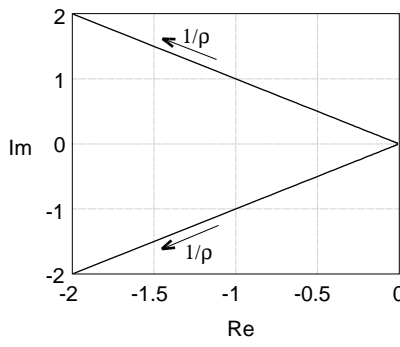


Figure 4.15: Loci of the closed-loop poles

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

$$\dot{x} = \underbrace{\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}}_A x + \underbrace{\begin{bmatrix} 0 \\ 1 \end{bmatrix}}_B u, \quad y = \underbrace{\begin{bmatrix} 1 & 0 \end{bmatrix}}_C x. \quad (4.131)$$

Completing the system equations we have

$$\dot{x} = Ax + Bu + Gv, \quad (4.132)$$

$$y = Cx + w, \quad (4.133)$$

$$z = Dx. \quad (4.134)$$

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 § 4.2.6 (p. 140) we have $k = 1$, $\psi(s) = 1$ and $\chi_{ol}(s) = s^2$. It follows from (4.32) that the closed-loop characteristic polynomial χ_{cl} for state feedback satisfies

$$\chi_{cl}(-s)\chi_{cl}(s) = \chi_{ol}(-s)\chi_{ol}(s) + \frac{k^2}{\rho}\psi(-s)\psi(s) = s^4 + \frac{1}{\rho}. \quad (4.135)$$

The roots of the polynomial on the right-hand side are $\frac{1}{2}\sqrt{2}(\pm 1 \pm j)/\rho^{\frac{1}{4}}$. To determine the closed-loop poles we select those two roots that have negative real parts. They are given by

$$\frac{1}{2}\sqrt{2}(-1 \pm j)/\rho^{\frac{1}{4}}. \quad (4.136)$$

Figure 4.15 shows the loci of the closed-loop poles as ρ varies. As the magnitude of the closed-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 4.8(c) (p. 180) the observer characteristic polynomial χ_f satisfies

$$\chi_f(-s)\chi_f(s) = \chi_{ol}(s)\chi_{ol}(-s)\left[1 + \frac{1}{\sigma}M(s)M(-s)\right] = s^4 + \frac{1}{\sigma^4}, \quad (4.137)$$

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

$$\frac{1}{2}\sqrt{2}(-1 \pm j)/\sigma^{\frac{1}{4}}. \quad (4.138)$$

By the rule of thumb of § 4.5.2 (p. 156) 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 computation⁸ it is found that the optimal compensator transfer function is

$$C_e(s) = \frac{155.6(s + 0.6428)}{s^2 + 15.56s + 121.0}. \quad (4.139)$$

This is a lead compensator.

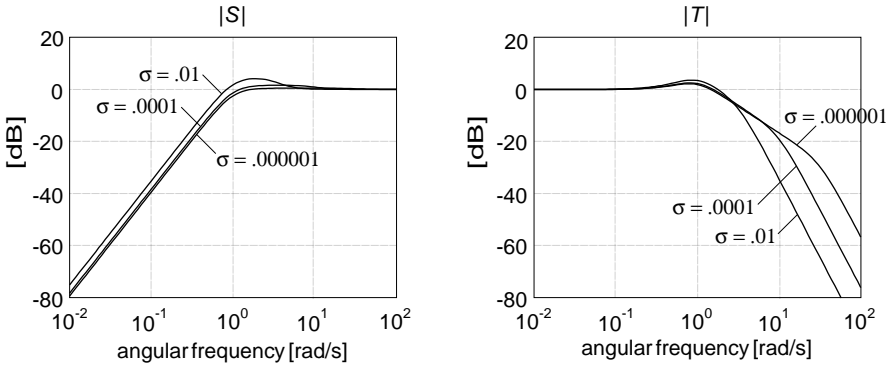


Figure 4.16: Sensitivity function and complementary sensitivity function for the \mathcal{H}_2 design

Figure 4.16 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 4.17 gives the closed-loop step response. The value $\sigma = .000001$ gives the best response but that for $\sigma = .0001$ is very close.

4.6.3. A MIMO system

As a second example we consider the two-input two-output plant with transfer matrix (Kwakernaak, 1986)

$$P(s) = \begin{bmatrix} \frac{1}{s^2} & \frac{1}{s+2} \\ 0 & \frac{1}{s+2} \end{bmatrix}. \quad (4.140)$$

⁸This may be done very conveniently with MATLAB using the Control Toolbox (Control Toolbox, 1990).

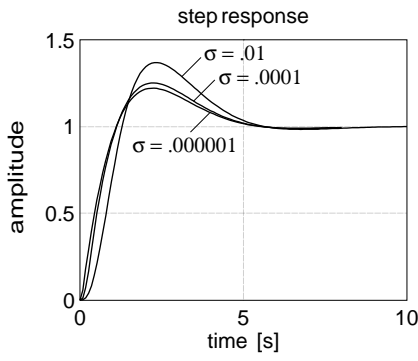


Figure 4.17: Closed-loop step response for the \mathcal{H}_2 design

Figure 4.18 shows the block diagram. The plant is triangularly coupled. It is easy to see that it may be represented in state space form as

$$\dot{x} = \underbrace{\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -2 \end{bmatrix}}_A x + \underbrace{\begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}}_B u, \quad y = \underbrace{\begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}}_C x. \quad (4.141)$$

The first two components of the state represent the block $1/s^2$ in Fig. 4.18, and the third the block $1/(s+2)$.

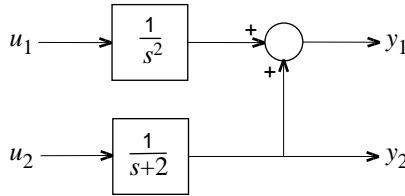


Figure 4.18: MIMO system

The plant has no right-half plane poles or zeros, so that there are no fundamental limitations to its performance. 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

$$\dot{x} = Ax + Bu + Gv, \quad (4.142)$$

$$y = Cx + w, \quad (4.143)$$

$$z = Dx. \quad (4.144)$$

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

$$Q = I, \quad R = \rho I, \quad V = I, \quad W = \sigma I, \quad (4.145)$$

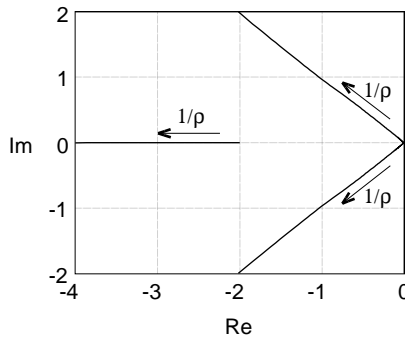


Figure 4.19: Loci of the regulator poles

with the positive scalars ρ and σ to be determined. First we consider the design of the regulator 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. 4.19. 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.. 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 \mathcal{H}_2 solution may now be found. Figure 4.20 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 corresponds 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 § 4.5.4 (p. 158) amounts to including an extra block

$$\frac{s + \alpha}{s} \quad (4.146)$$

in the second channel, as indicated in Fig. 4.21. After completion of the design the extra block is absorbed into the compensator. Representing the extra block by the state space realization $\dot{x}_4 = u'_2$, $u_2 = \alpha x_4 + u'_2$ we obtain the modified plant

$$\dot{x} = \underbrace{\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -2 & \alpha \\ 0 & 0 & 0 & 0 \end{bmatrix}}_A x + \underbrace{\begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}}_B u, \quad y = \underbrace{\begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}}_C x. \quad (4.147)$$

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 .

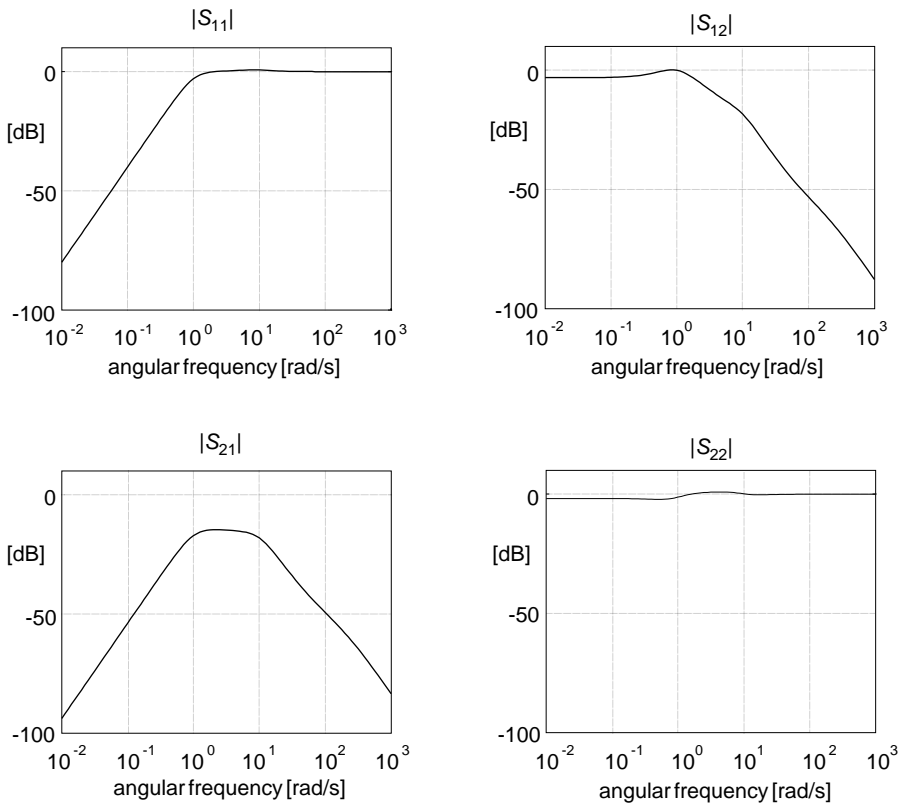


Figure 4.20: Magnitudes of the entries of the sensitivity matrix S

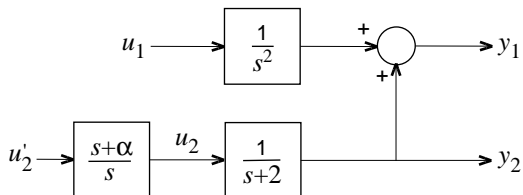


Figure 4.21: Expanded plant for integrating action in the second channel

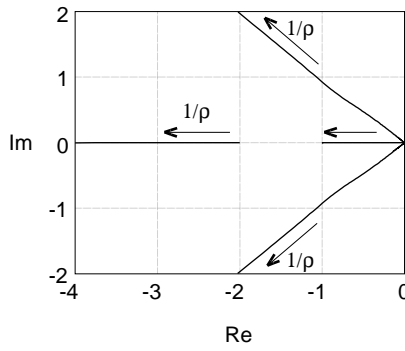


Figure 4.22: Loci of the regulator poles of the extended system

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 4.22 shows the loci of the regulator poles with ρ as parameter. 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 , $-0.8141 \pm j 0.7394$ and -0.6079 . The latter pole turns out to be nondominant. The pole pair $-0.8141 \pm j 0.7394$, 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 , $-7.076 \pm j 7.067$ and -1 . The latter pole is nondominant and the pole pair $-7.076 \pm j 7.067$ has magnitude 10.

Figure 4.23 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. 4.24, which shows the entries s_{ij} , $i, j = 1, 2$ of the closed-loop response to unit steps on the two inputs.

4.7. Appendix: Proofs

In this appendix we provide sketches of several of the proofs for this chapter.

4.7.1. Outline of the solution of the regulator problem

We consider the problem of minimizing

$$\int_0^\infty [z^T(t)Qz(t) + u^T(t)Ru(t)] dt \quad (4.148)$$

for the stabilizable and detectable system

$$\dot{x}(t) = Ax(t) + Bu(t), \quad z(t) = Dx(t). \quad (4.149)$$

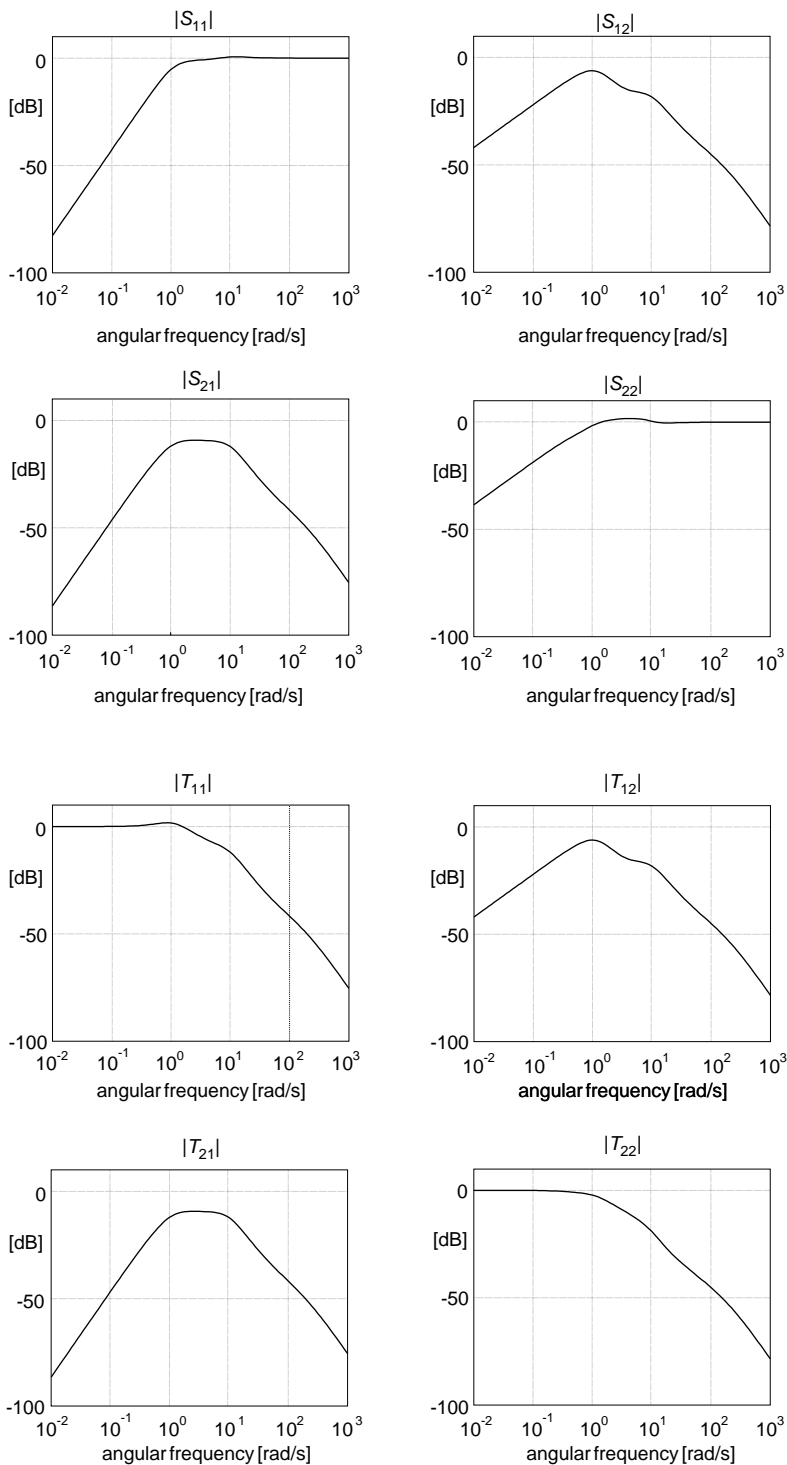


Figure 4.23: Magnitudes of the entries of the sensitivity matrix S and the complementary sensitivity matrix T

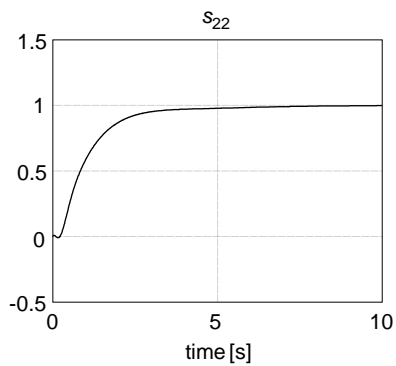
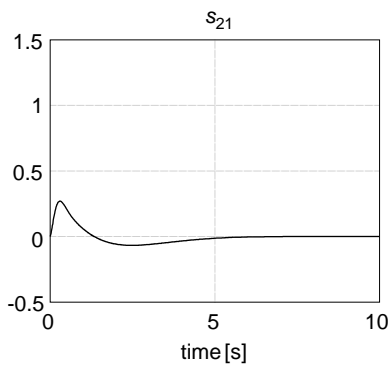
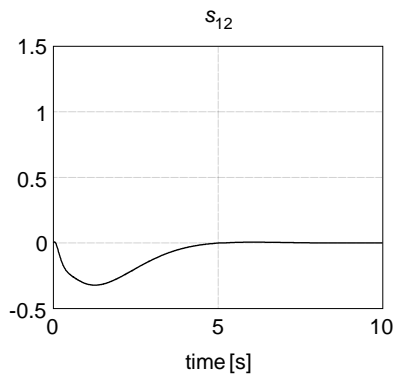
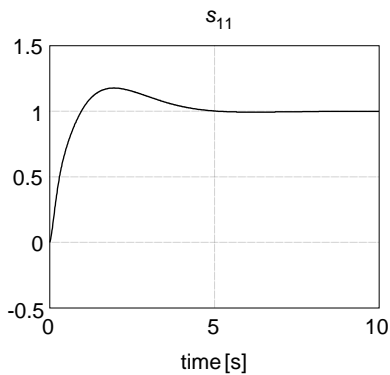


Figure 4.24: Entries of the closed-loop unit step response matrix

Lyapunov equation. We first study how to compute the quantity

$$\int_0^{t_1} x^T(s) Q(s) x(s) ds + x^T(t_1) X_1 x(t_1) \quad (4.150)$$

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

$$\begin{aligned} \int_0^{t_1} x^T(s) Q(s) x(s) ds + x^T(t_1) X_1 x(t_1) \\ = x^T(0) \left(\int_0^{t_1} \Phi^T(s, 0) Q(s) \Phi(s, 0) ds + \Phi^T(t_1, 0) X_1 \Phi(t_1, 0) \right) x(0). \end{aligned} \quad (4.151)$$

Define the time-dependent matrix

$$X(t) = \int_t^{t_1} \Phi^T(s, t) Q(s) \Phi(s, t) ds + \Phi^T(t_1, t) X_1 \Phi(t_1, t). \quad (4.152)$$

Then

$$\int_0^{t_1} x^T(s) Q(s) x(s) ds + x^T(t_1) X_1 x(t_1) = x^T(0) X(0) x(0). \quad (4.153)$$

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

$$-\dot{X}(t) = A^T(t) X(t) + X(t) A(t) + Q(t), \quad X(t_1) = X_1. \quad (4.154)$$

This equation is a *Lyapunov matrix differential equation*. If A is constant and stable and R is constant then as $t_1 \rightarrow \infty$ 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*

$$0 = A^T \bar{X} + \bar{X} A + Q. \quad (4.155)$$

Solution of the regulator problem. We consider how to determine the time-dependent gain $F(t)$ such that the state feedback $u(t) = -F(t)x(t)$, $0 \leq t \leq t_1$, minimizes the performance criterion

$$\begin{aligned} \int_0^{t_1} [z^T(t) Q z(t) + u^T(t) R u(t)] dt + x^T(t_1) X_1 x(t_1) \\ = \int_0^{t_1} x^T(t) \left(D^T Q D + F^T(t) R F(t) \right) x(t) dt + x^T(t_1) X_1 x(t_1) \end{aligned} \quad (4.156)$$

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

$$\int_0^\infty [z^T(t) Q z(t) + u^T(t) R u(t)] dt + x^T(t_1) X_1 x(t_1) = x^T(0) X(0) x(0), \quad (4.157)$$

where $X(t)$ is the solution of

$$-\dot{X}(t) = [A - BF(t)]^T X(t) + X(t) [A - BF(t)] + D^T Q D + F^T(t) R F(t), \quad (4.158)$$

$X(t_1) = X_1$. Completion of the square on the right-hand side (with respect to $F(t)$) results in

$$\begin{aligned} -\dot{X}(t) &= [F(t) - R^{-1} B^T X(t)]^T R [F(t) - R^{-1} B^T X(t)] \\ &\quad + A^T X(t) + X(t) A + D^T Q D - X(t) B R^{-1} B^T X(t), \quad X(t_1) = X_1. \end{aligned} \quad (4.159)$$

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

$$F(t) = -R^{-1} B^T X(t). \quad (4.160)$$

Correspondingly, (4.159) reduces to the *matrix differential Riccati equation*

$$-\dot{X}(t) = A^T X(t) + X(t)A + D^T Q D - X(t) B R^{-1} B^T X(t), \quad X(t_1) = X_1. \quad (4.161)$$

If the system is stabilizable then the minimum of the performance criterion on the left-hand side of (4.156) 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 $t_1 \rightarrow \infty$. Hence, also $X(t)$ has a well-defined limit as $t_1 \rightarrow \infty$. 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

$$0 = A^T \bar{X} + \bar{X} A + D^T Q D - \bar{X} B R^{-1} B^T \bar{X}. \quad (4.162)$$

Correspondingly the optimal feedback gain is time-invariant and equal to

$$F = -R^{-1} B^T \bar{X}. \quad (4.163)$$

The left-hand side of (4.156) can only converge to a finite limit if $z(t) = Dx(t) \rightarrow 0$ as $t \rightarrow \infty$. By detectability this implies that $x(t) \rightarrow 0$, 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 $t \geq 0$.

4.7.2. Kalman-Yakubovič-Popov equality

The Kalman-Yakubovič-Popov equality is a generalization of the return difference equality that we use in § 4.2.5 (p. 139). The KYP equality establishes the connection between factorizations and algebraic Riccati equations.

Summary 4.7.1 (Kalman-Yakubovič-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

$$0 = A^T X + XA + C^T Q C - (XB + C^T Q D)(D^T Q D + R)^{-1}(B^T X + D^T Q C) \quad (4.164)$$

has a symmetric solution X . Then

$$R + G^\sim(s) Q G(s) = J^\sim(s) R_D J(s). \quad (4.165)$$

The constant symmetric matrix R_D and the rational matrix function J are given by

$$R_D = R + D^T Q D, \quad J(s) = I + F(sI - A)^{-1} B, \quad (4.166)$$

with $F = R_D^{-1}(B^T X + D^T Q C)$. The zeros of the numerator of $\det J$ are the eigenvalues of the matrix $A - BF$. \square

We use the notation $G^\sim(s) = G^T(-s)$.

The KYP 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

$$\int_0^\infty [y^T(t) Q y(t) + u^T(t) R u(t)] dt. \quad (4.167)$$

The equation (4.164) is the algebraic Riccati equation associated with this problem, and $u(t) = -Fx(t)$ is the corresponding optimal state feedback law.

The KYP equality is best known for the case $D = 0$ (see for instance Kwakernaak and Sivan (1972)). It then reduces to the *return difference equality*

$$J^\sim(s) R J(s) = R + G^\sim(s) Q G(s). \quad (4.168)$$

Kalman-Yakubovič-Popov equality. The proof starts from the algebraic Riccati equation

$$0 = A^T X + XA + C^T Q C - (XB + C^T Q D) R_D^{-1} (B^T X + D^T Q C), \quad (4.169)$$

with $R_D = R + D^T Q D$. From the relation $F = R_D^{-1} (B^T X + D^T Q C)$ we have $B^T X + D^T Q C = R_D F$, so that the Riccati equation may be written as

$$0 = A^T X + XA + C^T Q C - F^T R_D F. \quad (4.170)$$

This in turn we rewrite as

$$0 = -(sI - A^T)X - X(sI - A) + C^T Q C - F^T R_D F. \quad (4.171)$$

Premultiplication by $B^T(-sI - A^T)^{-1}$ and postmultiplication by $(sI - A)^{-1}B$ results in

$$\begin{aligned} 0 = & -B^T X(sI - A)^{-1}B - B^T(-sI - A^T)^{-1}XB \\ & + B^T(-sI - A^T)^{-1}(C^T Q C - F^T R_D F)(sI - A)^{-1}B. \end{aligned} \quad (4.172)$$

Substituting $B^T X = R_D F - D^T Q C$ we find

$$\begin{aligned} 0 = & (D^T Q C - R_D F)(sI - A)^{-1}B + B^T(-sI - A^T)^{-1}(C^T Q D - F^T R_D) \\ & + B^T(-sI - A^T)^{-1}(C^T Q C - F^T R_D F)(sI - A)^{-1}B. \end{aligned} \quad (4.173)$$

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

$$R + G^\sim(s) Q G(s) = J^\sim(s) R_D J(s). \quad (4.174)$$

■

4.7.3. Robustness under state feedback

We consider an open-loop stable system with loop gain matrix L that satisfies the return difference inequality

$$(I + L)^\sim R(I + L) \geq R \quad \text{on the imaginary axis.} \quad (4.175)$$

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

$$RW + W^\sim R > R \quad \text{on the imaginary axis.} \quad (4.176)$$

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

$$L^{-1} + (L^{-1})^\sim + I \geq 0 \quad \text{on the imaginary axis.} \quad (4.177)$$

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

$$I + [(1 - \varepsilon)I + \varepsilon W]L \quad (4.178)$$

cross the imaginary axis. Hence, the perturbed system is stable if only if

$$L^{-1} + (1 - \varepsilon)I + \varepsilon W = M_\varepsilon \quad (4.179)$$

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 (4.177) yields

$$M_\varepsilon + M_\varepsilon^\sim \geq 2(1 - \varepsilon)I + \varepsilon(W + W^\sim) = (2 - \varepsilon)I + \varepsilon(W + W^\sim - I). \quad (4.180)$$

Inspection shows that if

$$W + W^\sim > I \quad (4.181)$$

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.

4.7.4. Riccati equations and the Hamiltonian matrix

We consider the algebraic Riccati equation

$$A^T X + XA + Q - (XB + S)R^{-1}(B^T X + S^T) = 0 \quad (4.182)$$

and the associated Hamiltonian matrix

$$\mathcal{H} = \begin{bmatrix} A - BR^{-1}S^T & -BR^{-1}B^T \\ -Q + SR^{-1}S^T & -(A - BR^{-1}S^T)^T \end{bmatrix}. \quad (4.183)$$

Summary 4.7.2 (Riccati equation and the Hamiltonian matrix).

1. If λ is an eigenvalue of \mathcal{H} then also $-\lambda$ is an eigenvalue of \mathcal{H} .
2. Given a solution X of the Riccati equation, define $F = R^{-1}(B^T X + S^T)$. Then

$$\mathcal{H} \begin{bmatrix} I \\ X \end{bmatrix} = \begin{bmatrix} I \\ X \end{bmatrix} (A - BF). \quad (4.184)$$

3. If λ is an eigenvalue of $A - BF$ corresponding to the eigenvector x then λ is also an eigenvalue of \mathcal{H} , corresponding to the eigenvector

$$\begin{bmatrix} I \\ X \end{bmatrix} x. \quad (4.185)$$

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 4.2.1 (p. 137)—then the eigenvalues of \mathcal{H} consist of these n eigenvalues of $A - BF$ and their negatives.

4. Assume that \mathcal{H} has no eigenvalues with zero real part. Then there is a similarity transformation U that brings \mathcal{H} into upper triangular form T such that

$$\mathcal{H} = U T U^{-1} = U \begin{bmatrix} T_{11} & T_{12} \\ 0 & T_{22} \end{bmatrix} U, \quad (4.186)$$

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

$$U = \begin{bmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{bmatrix}, \quad (4.187)$$

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

$$X = U_{21}U_{11}^{-1} \quad (4.188)$$

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 4.2.1 (p. 137) 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

$$\mathcal{H} = \begin{bmatrix} A & Q \\ R & -A^T \end{bmatrix}, \quad (4.189)$$

with all blocks square, and Q and R symmetric. We have for the characteristic polynomial of \mathcal{H}

$$\begin{aligned} \det(\lambda I - \mathcal{H}) &= \det \begin{bmatrix} \lambda I - A & -Q \\ -R & \lambda I + A^T \end{bmatrix} \stackrel{(1)}{=} \det \begin{bmatrix} -\lambda I + A & Q \\ R & -\lambda I - A^T \end{bmatrix} \\ &\stackrel{(2)}{=} (-1)^n \det \begin{bmatrix} R & -\lambda I - A^T \\ -\lambda I + A & Q \end{bmatrix} \stackrel{(3)}{=} \det \begin{bmatrix} -\lambda I - A^T & R \\ Q & -\lambda I + A \end{bmatrix} \\ &\stackrel{(4)}{=} \det \begin{bmatrix} -\lambda I + A & Q \\ R & -\lambda I - A^T \end{bmatrix} \stackrel{(5)}{=} \det \begin{bmatrix} -\lambda I + A & -Q \\ -R & -\lambda I - A^T \end{bmatrix} \end{aligned} \quad (4.190)$$

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 first 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 $\bar{\lambda}$. Hence, if λ is an eigenvalue, so is $\bar{\lambda}$.

2. Using the Riccati equation we obtain from (4.183)

$$\mathcal{H} \begin{bmatrix} I \\ X \end{bmatrix} = \begin{bmatrix} A - BF \\ -Q + SR^{-1}S^T - (A - BR^{-1}S^T)^T X \end{bmatrix} \quad (4.191)$$

$$= \begin{bmatrix} A - BF \\ XA - XBF \end{bmatrix} = \begin{bmatrix} I \\ X \end{bmatrix} (A - BF). \quad (4.192)$$

3. If $(A - BF)x = \lambda x$ then

$$H \begin{bmatrix} I \\ X \end{bmatrix} x = \begin{bmatrix} I \\ X \end{bmatrix} (A - BF)x = \lambda \begin{bmatrix} I \\ X \end{bmatrix} x. \quad (4.193)$$

4. From $\mathcal{H}U = UT$ we obtain

$$\mathcal{H} \begin{bmatrix} U_{11} \\ U_{21} \end{bmatrix} = \begin{bmatrix} U_{11} \\ U_{21} \end{bmatrix} T_{11}. \quad (4.194)$$

After multiplying on the right by U_{11}^{-1} it follows that

$$\mathcal{H} \begin{bmatrix} I \\ U_{21}U_{11}^{-1} \end{bmatrix} = \begin{bmatrix} I \\ U_{21}U_{11}^{-1} \end{bmatrix} U_{11}T_{11}U_{11}^{-1}. \quad (4.195)$$

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. ■

4.7.5. The Kalman filter

In this subsection we sketch 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, $E v(t)v^T(s) = V\delta(t - s)$. The state of the system

$$x(t) = \int_{-\infty}^t e^{A(t-s)} v(s) ds, \quad t \in \mathbb{R}, \quad (4.196)$$

is a stationary stochastic process with covariance matrix

$$\begin{aligned} Y &= E x(t) x^T(t) = \int_{-\infty}^t \int_{-\infty}^t e^{A(t-s_1)} (E v(s_1) v^T(s_2)) e^{A^T(t-s_2)} ds_1 ds_2 \\ &= \int_{-\infty}^t e^{A(t-s)} V e^{A^T(t-s)} ds = \int_0^\infty e^{A\tau} V e^{A^T\tau} d\tau. \end{aligned} \quad (4.197)$$

It follows that

$$\begin{aligned} AY + YA^T &= \int_0^\infty \left(A e^{A\tau} V e^{A^T\tau} + e^{A\tau} V e^{A^T\tau} A^T \right) d\tau \\ &= \int_0^\infty \frac{d}{d\tau} \left(e^{A\tau} V e^{A^T\tau} \right) d\tau = e^{A\tau} V e^{A^T\tau} \Big|_0^\infty = -V. \end{aligned} \quad (4.198)$$

Hence, the covariance matrix Y is the unique solution of the Lyapunov equation

$$AY + YA^T + V = 0. \quad (4.199)$$

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

$$\dot{\hat{x}}(t) = A\hat{x}(t) + Bu(t) + K[y(t) - C\hat{x}(t)] \quad (4.200)$$

satisfies the differential equation

$$\dot{e}(t) = (A - KC)e(t) - Gv(t) + Kw(t). \quad (4.201)$$

The noise process $-Gv(t) + Kw(t)$ is white noise with intensity $G^T V G + K^T W K$. Hence, if the error system is stable then the error covariance matrix $Y = E e(t) e^T(t)$ is the unique solution of the Lyapunov equation

$$(A - KC)Y + Y(A - KC)^T + G^T V G + K^T W K = 0. \quad (4.202)$$

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 (4.202) as

$$(K - YC^T W^{-1})W(K - YC^T W^{-1})^T + AY + YA^T + GVG^T - YC^T W^{-1}CY = 0. \quad (4.203)$$

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 (4.203) shows that this implies

$$K = YC^T W^{-1}. \quad (4.204)$$

With this gain the observer reduces to the Kalman filter. The minimal error variance matrix Y satisfies the Riccati equation

$$AY + YA^T + GVG^T - YC^T W^{-1}CY = 0. \quad (4.205)$$

4.7.6. 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

$$E \left(z^T(t) Q z(t) + u^T(t) R u(t) \right) \quad (4.206)$$

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

$$Y = E x(t)x^T(t) = \int_0^\infty e^{(A-BF)s} V e^{(A-BF)^T s} ds. \quad (4.207)$$

Hence we have for the steady-state mean square error

$$\begin{aligned} E \left(z^T(t) Q z(t) + u^T(t) R u(t) \right) &= E \left(x^T(t) D^T Q D x(t) + x^T(t) F^T R F x(t) \right) \\ &= E \operatorname{tr} \left(x(t)x^T(t) D^T Q D + x(t)x^T(t) F^T R F \right) = \operatorname{tr} Y \left(D^T Q D + F^T R F \right). \end{aligned} \quad (4.208)$$

We rewrite this in the form

$$\begin{aligned} E \left(z^T(t) Q z(t) + u^T(t) R u(t) \right) &= \operatorname{tr} Y \left(D^T Q D + F^T R F \right) \\ &= \operatorname{tr} \int_0^\infty e^{(A-BF)s} V e^{(A-BF)^T s} ds \left(D^T Q D + F^T R F \right) \\ &= \operatorname{tr} V \underbrace{\int_0^\infty e^{(A-BF)^T s} \left(D^T Q D + F^T R F \right) e^{(A-BF)s} ds}_X = \operatorname{tr} V X. \end{aligned} \quad (4.209)$$

X is the solution of the Lyapunov equation

$$(A - BF)^T X + X(A - BF) + D^T Q D + F^T R F = 0. \quad (4.210)$$

X and, hence, $\operatorname{tr} V X$, is minimized by choosing $F = R^{-1} B^T X$, with X the solution of the Riccati equation $A^T X + X A + D^T Q D - X B R^{-1} B^T X = 0$.

4.7.7. 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

$$y = C(sI - A)^{-1}B(u + v) = G(s)(u + v). \quad (4.211)$$

Because by assumption G^{-1} is stable the input noise v may be recovered with arbitrary precision by approximating the inverse relation

$$v = G^{-1}(s)y - u \quad (4.212)$$

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

$$A Y_\sigma + Y_\sigma A^T + B V B^T - Y_\sigma C^T W^{-1} C Y_\sigma = 0. \quad (4.213)$$

We rewrite this as

$$A Y_\sigma + Y_\sigma A^T + B V B^T - \sigma K_\sigma W_0 K_\sigma^T = 0, \quad (4.214)$$

with $K_\sigma = Y_\sigma C^T W^{-1}$ the gain. Inspection shows that if $Y_\sigma \downarrow 0$ then necessarily $K_\sigma \rightarrow \infty$. In fact we may write

$$K_\sigma \approx \frac{1}{\sqrt{\sigma}} B U_\sigma \quad \text{as } \sigma \downarrow 0, \quad (4.215)$$

where U_σ is a square nonsingular matrix (which may depend on σ) such that $U_\sigma W_0 U_\sigma^T = V$.

We study the asymptotic behavior of the loop gain

$$\begin{aligned}
L_\sigma(s) &= F(sI - A + BF + K_\sigma C)^{-1} K_\sigma C(sI - A)^{-1} B \\
&\approx F(sI - A + BF + \frac{1}{\sqrt{\sigma}} B U_\sigma C)^{-1} \frac{1}{\sqrt{\sigma}} B U_\sigma C(sI - A)^{-1} B \\
&\approx F(sI - A + \frac{1}{\sqrt{\sigma}} B U_\sigma C)^{-1} \frac{1}{\sqrt{\sigma}} B U_\sigma C(sI - A)^{-1} B \\
&= F(sI - A)^{-1} \left(I + \frac{1}{\sqrt{\sigma}} B U_\sigma C(sI - A)^{-1} \right)^{-1} \frac{1}{\sqrt{\sigma}} B U_\sigma C(sI - A)^{-1} B \\
&\stackrel{(1)}{=} F(sI - A)^{-1} B \frac{1}{\sqrt{\sigma}} U_\sigma \left(I + \frac{1}{\sqrt{\sigma}} C(sI - A)^{-1} B U_\sigma \right)^{-1} C(sI - A)^{-1} B \\
&= F(sI - A)^{-1} B U_\sigma \left(I \sqrt{\sigma} + C(sI - A)^{-1} B U_\sigma \right)^{-1} C(sI - A)^{-1} B.
\end{aligned} \tag{4.216}$$

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

$$L_\sigma(s) \xrightarrow{\sigma \downarrow 0} F(sI - A)^{-1} B, \tag{4.217}$$

which is the loop gain under full state feedback.

4.7.8. Solution of the \mathcal{H}_2 optimization problem

We solve the standard \mathcal{H}_2 optimization problem of § 4.4.3 (p. 154) as if it is an LQG problem, that is, we set out to minimize the steady-state value of

$$E z^T(t) z(t) \tag{4.218}$$

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

$$\dot{x}(t) = Ax(t) + B_1 w(t) + B_2 u(t), \tag{4.219}$$

$$z(t) = C_1 x(t) + D_{11} w(t) + D_{12} u(t). \tag{4.220}$$

If $D_{11} \neq 0$ then the output z has a white noise component that may well make the mean square output (4.218) infinite. We therefore assume that $D_{11} = 0$. Under this assumption we have

$$z(t) = C_1 x(t) + D_{12} u(t) = \begin{bmatrix} I & D_{12} \end{bmatrix} \begin{bmatrix} C_1 x(t) \\ u(t) \end{bmatrix} = \begin{bmatrix} I & D_{12} \end{bmatrix} \begin{bmatrix} z_0(t) \\ u(t) \end{bmatrix}, \tag{4.221}$$

where $z_0(t) = C_1 x(t)$. As a result,

$$\begin{aligned}
E z^T(t) z(t) &= E \begin{bmatrix} z_0^T(t) & u^T(t) \end{bmatrix} \begin{bmatrix} I \\ D_{12}^T \end{bmatrix} \begin{bmatrix} I & D_{12} \end{bmatrix} \begin{bmatrix} z_0 \\ u(t) \end{bmatrix} \\
&= E \begin{bmatrix} z_0^T(t) & u^T(t) \end{bmatrix} \begin{bmatrix} I & D_{12} \\ D_{12}^T & D_{12}^T D_{12} \end{bmatrix} \begin{bmatrix} z_0 \\ u(t) \end{bmatrix}.
\end{aligned} \tag{4.222}$$

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

$$\begin{bmatrix} I & D_{12} \\ D_{12}^T & D_{12}^T D_{12} \end{bmatrix} \tag{4.223}$$

is positive-definite. A necessary and sufficient condition for the latter is that $D_{12}^T D_{12}$ be nonsingular. The solution to the regulator problem is a state feedback law of the form

$$u(t) = -Fx(t). \quad (4.224)$$

The gain matrix F may easily be found from the results of § 4.2.8 (p. 144) and is given in Summary 4.4.1 (p. 148).

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

$$\dot{x}(t) = Ax(t) + B_1 w(t) + B_2 u(t), \quad (4.225)$$

$$y(t) = C_2 x(t) + D_{21} w(t) + D_{22} u(t). \quad (4.226)$$

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

$$\dot{x}(t) = Ax(t) + B_1 w(t) + B_2 u(t), \quad (4.227)$$

$$y(t) - D_{22}u(t) = C_2 x(t) + v(t) \quad (4.228)$$

defines a stochastic system with cross correlated noise terms. We have

$$E \begin{bmatrix} w(t) \\ v(t) \end{bmatrix} \begin{bmatrix} w^T(s) & v^T(s) \end{bmatrix} = E \begin{bmatrix} I \\ D_{21} \end{bmatrix} w(t) w^T(s) \begin{bmatrix} I & D_{21}^T \end{bmatrix} \quad (4.229)$$

$$= \begin{bmatrix} I & D_{21}^T \\ D_{21} & D_{21} D_{21}^T \end{bmatrix} \delta(t-s). \quad (4.230)$$

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

$$\begin{bmatrix} I & D_{21}^T \\ D_{21} & D_{21} D_{21}^T \end{bmatrix} \quad (4.231)$$

is positive-definite. A necessary and sufficient condition for the latter is that $D_{21} D_{21}^T$ be nonsingular. Then there exists a well-defined Kalman filter of the form

$$\dot{\hat{x}}(t) = A\hat{x}(t) + B_2 u(t) + K[y(t) - C_2 \hat{x}(t) - D_{22} u(t)]. \quad (4.232)$$

The gain matrix K may be solved from the formulas of § 4.3.4 (p. 149). Once the Kalman filter (4.232) is in place the optimal input for the output feedback problem is obtained as

$$u(t) = -F\hat{x}(t). \quad (4.233)$$

F is the same state feedback gain as in (4.224).

4.8. Exercises

4.1 *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

$$\int_0^\infty [x^2(t) + \rho u^2(t)] dt \quad (4.234)$$

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 .

4.2 *Cruise control system.* Modify the criterion (4.234) to

$$\int_0^\infty e^{2\alpha t} [x^2(t) + \rho u^2(t)] dt, \quad (4.235)$$

with α a positive constant. Rework Exercise 4.1 while explaining and interpreting the effect of α .

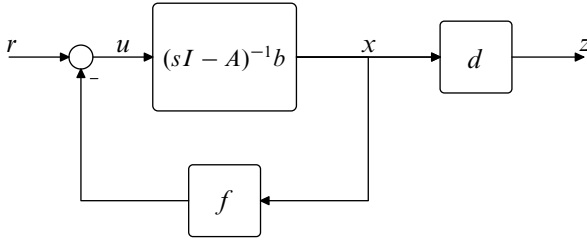


Figure 4.25: State feedback with reference input r and output z

4.3 *Closed-loop frequency response characteristics.* Recall these statements on page 142:

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).

Verify these statements by considering the SISO configuration of Fig. 4.25. Let $L(s) = f(sI - A)^{-1}b$ and $G(s) = d(sI - A)^{-1}b$.

a) Prove that

$$u = \frac{1}{1 + L(s)} r, \quad z = \underbrace{\frac{G(s)}{1 + L(s)}}_{H(s)} r. \quad (4.236)$$

It follows that the closed-loop transfer function H is

$$H(s) = \frac{k\psi(s)}{\chi_{cl}(s)}. \quad (4.237)$$

b) 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

$$H(s) \approx \frac{k}{B_{n-q}(s/\omega_c)}. \quad (4.238)$$

B_k denotes a Butterworth polynomial of order k (see Table 2.3, p. 92). Hence, the closed-loop bandwidth is ω_c .

- c) 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

$$H(s) \approx \frac{s - \zeta}{s + \zeta} \frac{k}{B_{n-q}(s/\omega_c)}. \quad (4.239)$$

Argue that this means that asymptotically the bandwidth is ζ (that is, the frequency response function $H(j\omega)/H(0) - 1$, $\omega \in \mathbb{R}$, is small over the frequency range $[0, \zeta]$).

4.4 *Phase margin.* In Subsection 4.2.7 it is claimed that the phase margin is at least 60° . Prove it.

4.5 *LQ problem for system with direct feedthrough.* The system

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t), \\ z(t) &= Dx(t) + Eu(t), \end{aligned} \quad t \geq 0, \quad (4.240)$$

has what is called “direct feedthrough” because of the term with u in the output z . Show that the problem of minimizing

$$\mathcal{J} = \int_0^\infty [z^T(t)Qz(t) + u^T(t)Ru(t)] dt \quad (4.241)$$

for this system may be converted into the cross term problem of this subsection.

4.6 *Positive-definiteness and Schur complement.* Prove that the condition that the matrix (4.41) be positive-definite is equivalent to either of the following two conditions:

- a) Both R and $Q - SR^{-1}S^T$ are positive-definite. $Q - SR^{-1}S^T$ is called the *Schur complement* of R . *Hint:*

$$\begin{bmatrix} Q & S \\ S^T & R \end{bmatrix} = \begin{bmatrix} I & SR^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} Q - SR^{-1}S^T & 0 \\ 0 & R \end{bmatrix} \begin{bmatrix} I & 0 \\ R^{-1}S^T & I \end{bmatrix}. \quad (4.242)$$

- b) Both Q and $R - S^TQ^{-1}S$ are positive-definite. $R - S^TQ^{-1}S$ is the Schur complement of Q .

4.7 *Cruise control system.* Use the method of this Subsection 4.2.9 to solve the ARE that arises in Exercise 4.1.

4.8 *Asymptotic results.* The asymptotic results for the regulator problem may be “dualized” to the Kalman filter.

- a) Define the “observer return difference”

$$J_f(s) = I + C(sI - A)^{-1}K. \quad (4.243)$$

Prove that

$$\det J_f(s) = \frac{\chi_f(s)}{\chi_{ol}(s)}, \quad (4.244)$$

where $\chi_{ol}(s) = \det(sI - A)$ is the system characteristic polynomial and $\chi_f(s) = \det(sI - A + KC)$ the observer characteristic polynomial.

b) Prove that the return difference J_f of the Kalman filter satisfies

$$J_f(s)WJ_f^T(-s) = W + M(s)VM^T(-s), \quad (4.245)$$

where M is the open-loop transfer matrix $M(s) = C(sI - A)^{-1}G$.

c) Consider the SISO case with $V = 1$ and $W = \sigma$. J_f and M are now scalar functions. Prove that

$$\chi_f(s)\chi_f(-s) = \chi_{ol}(s)\chi_{ol}(-s)[1 + \frac{1}{\sigma}M(s)M(-s)]. \quad (4.246)$$

d) Write

$$M(s) = g \frac{\zeta(s)}{\chi_{ol}(s)}, \quad (4.247)$$

with ζ a monic polynomial and g a constant.

- Prove that as $\sigma \rightarrow \infty$ 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 $\sigma \downarrow 0$.

4.9 *Cross correlated noises.* Prove the claim of Subsection 4.3.4.

4.10 *Closed-loop eigenvalues.*

- Prove that the eigenvalues of (4.83) are the eigenvalues of the closed-loop system.
- 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 (4.81) by itself is stable.

4.11 *Compensator transfer function.* The configuration of Fig. 4.6 may be rearranged as in Fig. 4.26. Show that the equivalent compensator C_e has the transfer matrix $C_e(s) = F(sI - A + BF + KC)^{-1}K$.

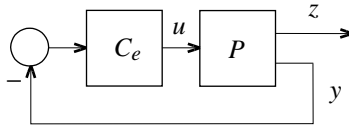


Figure 4.26: Equivalent unit feedback configuration

4.12 *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 $\rho \downarrow 0$.

- a) Make it plausible that the loop gain approaches

$$L_0(s) = C(sI - A)^{-1}K. \quad (4.248)$$

- b) Show that the corresponding return difference $J_0(s) = I + L_0(s)$ satisfies the return difference inequality

$$J_0(j\omega)WJ_0^T(-j\omega) \geq W, \quad \omega \in \mathbb{R}. \quad (4.249)$$

- c) Show that gain and phase margins apply that agree with those found in Subsection 4.2.7 (p. 143).

- 4.13 *Lyapunov equation.* Prove that the \mathcal{H}_2 -norm of the stable system (4.86) is given by $\|H\|_2^2 = \text{tr } CYC^T$, where the matrix Y is the unique symmetric solution of the Lyapunov equation $AY + YA^T + BB^T = 0$.

- 4.14 *Generalized plant for the LQG problem.* Show that for the LQG problem the generalized plant of the corresponding standard \mathcal{H}_2 problem in state space form may be represented as

$$\dot{x}(t) = Ax(t) + \begin{bmatrix} G & 0 \end{bmatrix} \begin{bmatrix} v(t) \\ w(t) \end{bmatrix} + Bu(t), \quad (4.250)$$

$$\begin{bmatrix} z(t) \\ u(t) \\ y(t) \end{bmatrix} = \begin{bmatrix} D \\ 0 \\ C \end{bmatrix} x(t) + \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} v(t) \\ w(t) \end{bmatrix} + \begin{bmatrix} 0 \\ I \\ 0 \end{bmatrix} u(t). \quad (4.251)$$

- 4.15 *Transfer matrices.* Derive (4.117–4.120).

- 4.16 *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.

- 4.17 *Poles and zeros.* What are the open-loop poles and zeros of the system of Fig. 4.18.

- 4.18 *Identical loci.* On page 165 it is claimed that “Like in the double integrator example, they are identical to those of the regulator poles.”

Check that this is a consequence of choosing $Q = V = I$, $R = \rho I$, $W = \sigma I$, $D = C$ and $G = B$.

- 4.19 *The case $R \neq I$.* Subsection 4.7.3 considers the case $R = I$. 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 Cholesky factorization—as $R = R_0^T R_0$. Work this out to prove that the closed-loop system remains stable under perturbation satisfying $RW + W^*R > R$ on the imaginary axis.

- 4.20 Consider the plant

$$P(s) = \frac{1000}{s(s + 10s + 300^2)}.$$

Design a stabilizing controller $C(s)$ that achieves a cross-over frequency of 7Hz (i.e. the 0dB point of $L = PC$ is at 7Hz), and such that $|S(j\omega)| < 2$ for all frequencies.

5. Uncertainty Models and Robustness

Overview – Various techniques exist to examine the stability robustness of control systems subject to parametric uncertainty.

Parametric and nonparametric uncertainty with varying degree of structure may be captured by the basic perturbation model. The size of the perturbation is characterized by bounds on the norm of the perturbation. The small gain theorem provides the tool to analyze the stability robustness of this model.

These methods allow to generalize the various stability robustness results for SISO systems of Chapter 2 in several ways.

5.1. Introduction

In this chapter we discuss various paradigms for representing uncertainty about the *dynamic properties* of a plant. Moreover, we present methods to analyze the effect of uncertainty on closed-loop stability and performance.

Section 5.2 is devoted to *parametric* uncertainty models. The idea is to assume that the equations that describe the dynamics of the plant and compensator (in particular, their transfer functions or matrices) are known, but that there is uncertainty about the precise values of various parameters in these equations. The uncertainty is characterized by an *interval* of possible values. We discuss some methods to analyze closed-loop stability under this type of uncertainty. The most famous result is *Kharitonov's theorem*.

In § 5.3 we introduce the so-called *basic perturbation model* for linear systems, which admits a much wider class of perturbations, including *unstructured perturbations*. Unstructured perturbations may involve changes in the order of the dynamics and are characterized by *norm bounds*. Norm bounds are bounds on the norms of operators corresponding to systems.

In § 5.4 we review the small gain theorem. With the help of that we establish sufficient and necessary conditions for the robust stability of the basic perturbation model. Next, in § 5.5 the basic stability robustness result is applied to proportional, proportional inverse perturbations and fractional perturbations of feedback loops.

The perturbation models of § 5.5 are relatively crude. Doyle's *structured singular value* allows much finer structuring. It is introduced in § 5.6. An appealing feature of structured singular value analysis is that it allows studying *combined* stability and performance robustness in a unified framework. This is explained in § 5.7. In § 5.8 a number of proofs for this chapter are presented.

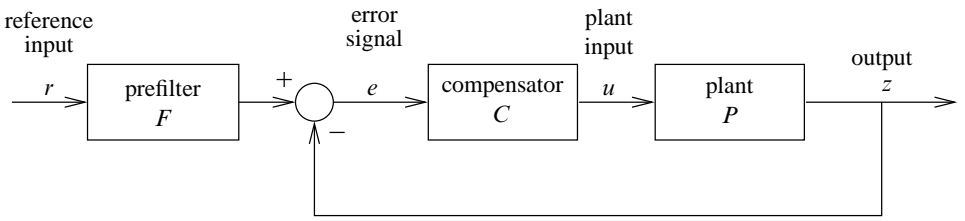


Figure 5.1: Feedback system

5.2. Parametric robustness analysis

In this section we review the *parametric* approach to robustness analysis. In this view of uncertainty the plant and compensator transfer functions are assumed to be given, but contain several parameters whose values are not precisely known. We illustrate this by an example.

Example 5.2.1 (Third-order uncertain plant). Consider a feedback configuration as in Fig. 5.1 where the plant transfer function is given by

$$P(s) = \frac{g}{s^2(1 + \theta s)}. \quad (5.1)$$

The gain g is not precisely known but is nominally equal to $g_0 = 1 \text{ [s}^{-2}\text{]}$. The number θ is a parasitic time constant and nominally 0 [s]^1 . A system of this type can successfully be controlled by a PD controller with transfer function $C(s) = k + T_d s$. The latter transfer function is not proper so we modify it to

$$C(s) = \frac{k + T_d s}{1 + T_0 s}, \quad (5.2)$$

where the time constant T_0 is small so that the PD action is not affected at low frequencies. With these plant and compensator transfer functions, the return difference $J = 1 + L = 1 + PC$ is given by

$$J(s) = 1 + \frac{g}{s^2(1 + \theta s)} \frac{k + T_d s}{1 + T_0 s} = \frac{\theta T_0 s^4 + (\theta + T_0)s^3 + s^2 + gT_d s + gk}{s^2(1 + \theta s)(1 + T_0 s)}. \quad (5.3)$$

Hence, the stability of the feedback system is determined by the locations of the roots of the closed-loop characteristic polynomial

$$\chi(s) = \theta T_0 s^4 + (\theta + T_0)s^3 + s^2 + gT_d s + gk. \quad (5.4)$$

Presumably, the compensator can be constructed with sufficient precision so that the values of the parameters k , T_d , and T_0 are accurately known, and such that the closed-loop system is stable at the nominal plant parameter values $g = g_0 = 1$ and $\theta = 0$. The question is whether the system remains stable under variations of g and θ , that is, whether the roots of the closed-loop characteristic polynomial remain in the open left-half complex plane. \square

Because we pursue this example at some length we use the opportunity to demonstrate how the compensator may be designed by using the classical *root locus* design tool. We assume that the ground rules for the construction of root loci are known.

¹All physical units are SI. From this point on they are usually omitted.

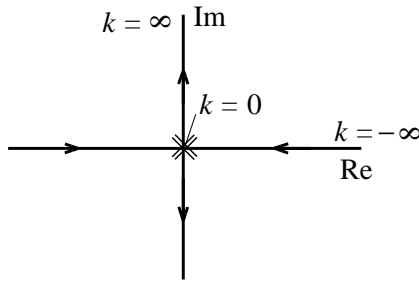


Figure 5.2: Root locus for compensator consisting of a simple gain

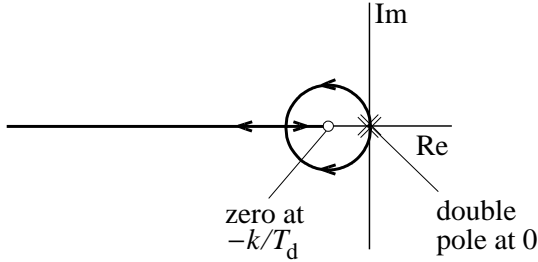


Figure 5.3: Root locus for a PD compensator

Example 5.2.2 (Compensator design by root locus method). Given the nominal plant with transfer function

$$P_0(s) = \frac{g_0}{s^2}, \quad (5.5)$$

the simplest choice for a compensator would be a simple gain $C(s) = k$. Varying k from 0 to ∞ or from 0 to $-\infty$ results in the root locus of Fig. 5.2. For no value of k stability is achieved.

Modification to a PD controller with transfer function $C(s) = k + sT_d$ amounts to addition of a zero at $-k/T_d$. Accordingly, keeping $-k/T_d$ fixed while varying k the root locus of Fig. 5.2 is altered to that of Fig. 5.3 (only the part for $k \geq 0$ is shown). We assume that a closed-loop bandwidth of 1 is required. This may be achieved by placing the two closed-loop poles at $\frac{1}{2}\sqrt{2}(-1 \pm j)$. The distance of this pole pair from the origin is 1, resulting in the desired closed-loop bandwidth. Setting the ratio of the imaginary to the real part of this pole pair equal to 1 ensures an adequate time response with a good compromise between rise time and overshoot.

The closed-loop characteristic polynomial corresponding to the PD compensator transfer function $C(s) = k + sT_d$ is easily found to be given by

$$s^2 + g_0T_d s + g_0k. \quad (5.6)$$

Choosing $g_0T_d = \sqrt{2}$ and $g_0k = 1$ (i.e., $T_d = \sqrt{2}$ and $k = 1$) places the closed-loop poles at the desired locations. The zero in Fig. 5.3 may now be found at $-k/T_d = -\frac{1}{2}\sqrt{2} = -0.7071$.

The final step in the design is to make the compensator transfer function proper by changing it to

$$C(s) = \frac{k + T_d s}{1 + T_0 s}. \quad (5.7)$$

This amounts to adding a pole at the location $-1/T_0$. Assuming that T_0 is small the root locus now takes the appearance shown in Fig. 5.4. The corresponding closed-loop characteristic polynomial is

$$T_0 s^3 + s^2 + g_0 T_d s + g_0 k. \quad (5.8)$$

Keeping the values $T_d = \sqrt{2}$ and $k = 1$ and choosing T_0 somewhat arbitrarily equal to $1/10$ (which places the additional pole in Fig. 5.4 at -10) results in the closed-loop poles

$$-0.7652 \pm j0.7715, \quad -8.4697. \quad (5.9)$$

The dominant pole pair at $\frac{1}{2}\sqrt{2}(-1 \pm j)$ has been slightly shifted and an additional non-dominant pole at -8.4697 appears. \square

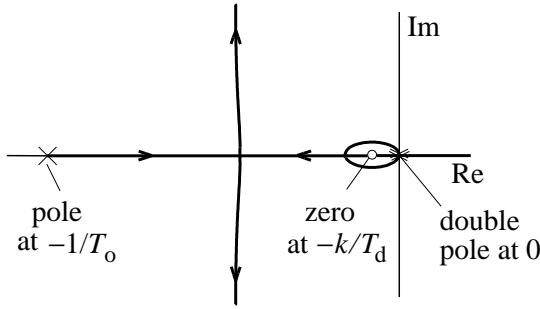


Figure 5.4: Root locus for the modified PD compensator

5.2.1. Routh-Hurwitz Criterion

In the parametric approach, stability robustness analysis comes down to investigating the roots of a characteristic polynomial of the form

$$\chi(s) = \chi_n(p)s^n + \chi_{n-1}(p)s^{n-1} + \cdots + \chi_0(p), \quad (5.10)$$

whose coefficients $\chi_n(p)$, $\chi_{n-1}(p)$, \cdots , $\chi_0(p)$ depend on the parameter vector p . Usually it is not possible to determine the dependence of the roots on p explicitly.

Sometimes, if the problem is not very complicated, the Routh-Hurwitz criterion may be invoked for testing stability. For completeness we summarize this celebrated result (see for instance [Chen \(1970\)](#)). Recall that a polynomial is *Hurwitz* if all its roots have zero or negative real part. It is *strictly Hurwitz* if all its roots have strictly negative real part.

Theorem 5.2.3 (Routh-Hurwitz criterion). A polynomial

$$\chi(s) = a_0 s^n + a_1 s^{n-1} + \cdots + a_{n-1} s + a_n \quad (5.11)$$

with real coefficients a_0, a_1, \dots, a_n is stable if and only if the $n + 1$ entries in the first column of the *Hurwitz tableau* exist, are nonzero and have the same sign.

The *Hurwitz tableau* is of the form

$$\begin{array}{ccccc} a_0 & a_2 & a_4 & a_6 & \cdots \\ a_1 & a_3 & a_5 & a_7 & \cdots \\ b_0 & b_2 & b_4 & \cdots & \\ b_1 & b_3 & b_5 & \cdots & \\ \cdots & & & & \end{array} \quad (5.12)$$

The first two rows of the tableau are directly taken from the “even” and “odd” coefficients of the polynomial χ , respectively. The third row is constructed from the two preceding rows by letting

$$[b_0 \quad b_2 \quad b_4 \quad \cdots] = [a_2 \quad a_4 \quad a_6 \quad \cdots] - \frac{a_0}{a_1} [a_3 \quad a_5 \quad a_7 \quad \cdots]. \quad (5.13)$$

The fourth row b_1, b_3, \dots is formed from the two preceding rows in the same way the third row is formed from the first and second rows. All further rows are constructed in this manner. Missing entries at the end of the rows are replaced with zeros. We stop after $n + 1$ rows. \square

In principle, the Routh-Hurwitz criterion allows establishing the *stability region* of the parameter dependent polynomial χ as given by (5.10), that is, the set of all parameter values p for which the roots of χ all have strictly negative real part. In practice, this is often not simple.

Example 5.2.4 (Stability region for third-order plant). By way of example we consider the third-order plant of Examples 5.2.1 and 5.2.2. Using the numerical values established in Example 5.2.2 we have from Example 5.2.1 that the closed-loop characteristic polynomial is given by

$$\chi(s) = \frac{\theta}{10} s^4 + (\theta + \frac{1}{10}) s^3 + s^2 + g\sqrt{2}s + g. \quad (5.14)$$

The parameter vector p has components g and θ . The Hurwitz tableau may easily be found to be given by

$$\begin{array}{ccc} \frac{\theta}{10} & 1 & g \\ \theta + \frac{1}{10} & g\sqrt{2} & \\ b_0 & g & \\ b_1 & & \\ g & & \end{array} \quad (5.15)$$

where

$$b_0 = \frac{\theta + \frac{1}{10} - \frac{\sqrt{2}}{10}g\theta}{\theta + \frac{1}{10}}, \quad (5.16)$$

$$b_1 = \frac{(\theta + \frac{1}{10} - \frac{\sqrt{2}}{10}g\theta)\sqrt{2} - (\theta + \frac{1}{10})^2}{\theta + \frac{1}{10} - \frac{\sqrt{2}}{10}g\theta}g. \quad (5.17)$$

Inspection of the coefficients of the closed-loop characteristic polynomial (5.14) shows that a necessary condition for closed-loop stability is that both g and θ be positive. This condition ensures the first, second and fifth entries of the first column of the tableau to be positive. The third entry b_0 is positive if $g < g_3(\theta)$, where g_3 is the function

$$g_3(\theta) = \frac{5\sqrt{2}(\theta + \frac{1}{10})}{\theta}. \quad (5.18)$$

The fourth entry b_1 is positive if $g < g_4(\theta)$, with g_4 the function

$$g_4(\theta) = \frac{5(\theta + \frac{1}{10})(\sqrt{2} - \frac{1}{10} - \theta)}{\theta}. \quad (5.19)$$

Figure 5.5 shows the graphs of g_3 and g_4 and the resulting stability region.

The case $\theta = 0$ needs to be considered separately. Inspection of the root locus of Fig. 5.4 (which applies if $\theta = 0$) shows that for $\theta = 0$ closed-loop stability is obtained for all $g > 0$. \square

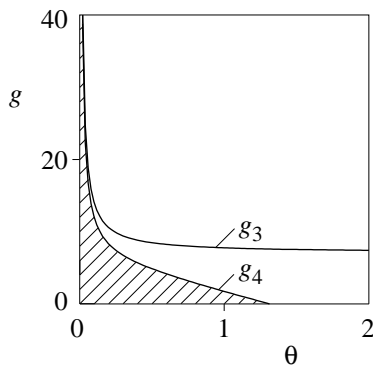


Figure 5.5: Stability region

Exercise 5.2.5 (Stability margins). The stability of the feedback system of Example 5.2.4 is quite robust with respect to variations in the parameters θ and g . Inspect the Nyquist plot of the nominal loop gain to determine the various stability margins of § 1.4.2 and Exercise 1.4.9(b) of the closed-loop system. \square

5.2.2. Gridding

If the number of parameters is larger than two or three then it seldom is possible to establish the stability region analytically as in Example 5.2.4. A simple but laborious alternative method is

known as *gridding*. It consists of covering the relevant part of the parameter space with a grid, and testing for stability in each grid point. The grid does not necessarily need to be rectangular.

Clearly this is a task that easily can be coded for a computer. The number of grid points increases exponentially with the dimension of the parameter space, however, and the computational load for a reasonably fine grid may well be enormous.

Example 5.2.6 (Gridding the stability region for the third-order plant). Figure 5.6 shows the results of gridding the parameter space in the region defined by $0.5 \leq g \leq 4$ and $0 \leq \theta \leq 1$. The small circles indicate points where the closed-loop system is stable. Plus signs correspond to unstable points. Each point may be obtained by applying the Routh-Hurwitz criterion or any other stability test. □

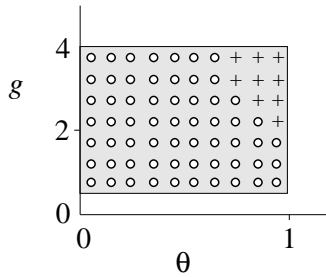


Figure 5.6: Stability region obtained by gridding

5.2.3. Kharitonov's theorem

Gridding is straightforward but may involve a tremendous amount of repetitious computation. In 1978, [Kharitonov](#) (see also [Kharitonov \(1978b\)](#)) published a result that subsequently attracted much attention in the control literature² because it may save a great deal of work. Kharitonov's theorem deals with the stability of a system with closed-loop characteristic polynomial

$$\chi(s) = \chi_0 + \chi_1 s + \cdots + \chi_n s^n. \quad (5.20)$$

Each of the coefficients χ_i is known to be bounded in the form

$$\underline{\chi}_i \leq \chi_i \leq \overline{\chi}_i, \quad (5.21)$$

with $\underline{\chi}_i$ and $\overline{\chi}_i$ given numbers for $i = 0, 1, \dots, n$. Kharitonov's theorem allows verifying whether all these characteristic polynomials are strictly Hurwitz by checking only *four* special polynomials out of the infinite family.

²For a survey see [Barmish and Kang \(1993\)](#).

Theorem 5.2.7 (Kharitonov's theorem). Each member of the infinite family of polynomials

$$\chi(s) = \chi_0 + \chi_1 s + \cdots + \chi_n s^n : \quad \text{with} \quad \underline{\chi}_i \leq \chi_i \leq \overline{\chi}_i, \quad i = 0, 1, 2, \dots, n, \quad (5.22)$$

is stable if and only if each of the four Kharitonov polynomials

$$k_1(s) = \underline{\chi}_0 + \underline{\chi}_1 s + \overline{\chi}_2 s^2 + \overline{\chi}_3 s^3 + \underline{\chi}_4 s^4 + \underline{\chi}_5 s^5 + \overline{\chi}_6 s^6 + \cdots, \quad (5.23)$$

$$k_2(s) = \overline{\chi}_0 + \overline{\chi}_1 s + \underline{\chi}_2 s^2 + \underline{\chi}_3 s^3 + \overline{\chi}_4 s^4 + \overline{\chi}_5 s^5 + \underline{\chi}_6 s^6 + \cdots, \quad (5.24)$$

$$k_3(s) = \overline{\chi}_0 + \underline{\chi}_1 s + \underline{\chi}_2 s^2 + \overline{\chi}_3 s^3 + \overline{\chi}_4 s^4 + \underline{\chi}_5 s^5 + \underline{\chi}_6 s^6 + \cdots, \quad (5.25)$$

$$k_4(s) = \underline{\chi}_0 + \overline{\chi}_1 s + \overline{\chi}_2 s^2 + \underline{\chi}_3 s^3 + \underline{\chi}_4 s^4 + \overline{\chi}_5 s^5 + \overline{\chi}_6 s^6 + \cdots \quad (5.26)$$

is stable. □

Note the repeated patterns of under- and overbars. A simple proof of Kharitonov's theorem is given by [Minnichelli et al. \(1989\)](#). We see what we can do with this result for our example.

Example 5.2.8 (Application of Kharitonov's theorem). From Example 5.2.1 we know that the stability of the third-order uncertain feedback system is determined by the characteristic polynomial

$$\chi(s) = \theta T_0 s^4 + (\theta + T_0) s^3 + s^2 + g T_d s + g k, \quad (5.27)$$

where in Example 5.2.2 we took $k = 1$, $T_d = \sqrt{2}$ and $T_0 = 1/10$. Suppose that the variations of the plant parameters g and θ are known to be bounded by

$$\underline{g} \leq g \leq \overline{g}, \quad 0 \leq \theta \leq \overline{\theta}. \quad (5.28)$$

Inspection of (5.27) shows that the coefficients χ_0 , χ_1 , χ_2 , χ_3 , and χ_4 are correspondingly bounded by

$$\begin{aligned} \underline{g} &\leq \chi_0 \leq \overline{g}, \\ \underline{g}\sqrt{2} &\leq \chi_1 \leq \overline{g}\sqrt{2}, \\ 1 &\leq \chi_2 \leq 1, \\ T_0 &\leq \chi_3 \leq T_0 + \overline{\theta}, \\ 0 &\leq \chi_4 \leq T_0 \overline{\theta}. \end{aligned} \quad (5.29)$$

We assume that

$$0 \leq \theta \leq 0.2, \quad (5.30)$$

so that $\overline{\theta} = 0.2$, but consider two cases for the gain g :

1. $0.5 \leq g \leq 5$, so that $\underline{g} = 0.5$ and $\overline{g} = 5$. This corresponds to a variation in the gain by a factor of ten. Inspection of Fig. 5.5 shows that this region of variation is well within the

stability region. It is easily found that the four Kharitonov polynomials are

$$k_1(s) = 0.5 + 0.5\sqrt{2}s + s^2 + 0.3s^3, \quad (5.31)$$

$$k_2(s) = 5 + 5\sqrt{2}s + s^2 + 0.1s^3 + 0.02s^4, \quad (5.32)$$

$$k_3(s) = 5 + 0.5\sqrt{2}s + s^2 + 0.3s^3 + 0.02s^4, \quad (5.33)$$

$$k_4(s) = 0.5 + 5\sqrt{2} + s^2 + 0.1s^3. \quad (5.34)$$

By the Routh-Hurwitz test or by numerical computation of the roots using MATLAB or another computer tool it may be found that k_1 and k_4 are strictly Hurwitz, while k_2 and k_3 are not Hurwitz. This shows that the polynomial $\chi_0 + \chi_1s + \chi_2s^2 + \chi_3s^3 + \chi_4s^4$ is not strictly Hurwitz for all variations of the coefficients within the bounds (5.29). This does *not* prove that the closed-loop system is not stable for all variations of g and θ within the bounds (5.28), however, because the coefficients χ_i of the polynomial do not vary *independently* within the bounds (5.29).

2. $0.5 \leq g \leq 2$, so that $\underline{g} = 0.5$ and $\overline{g} = 2$. The gain now only varies by a factor of four. Repeating the calculations we find that each of the four Kharitonov polynomials is strictly Hurwitz, so that the closed-loop system is stable for all parameter variations.

□

5.2.4. The edge theorem

Example 5.2.8 shows that Kharitonov's theorem usually only yields sufficient but not necessary conditions for robust stability in problems where the coefficients of the characteristic polynomial do not vary independently. We therefore consider characteristic polynomials of the form

$$\chi(s) = \chi_0(p)s^n + \chi_1(p)s^{n-1} + \cdots + \chi_n(p), \quad (5.35)$$

where the parameter vector p of uncertain coefficients p_i , $i = 1, 2, \dots, N$, enters *linearly* into the coefficients $\chi_i(p)$, $i = 0, 1, \dots, n$. Such problems are quite common; indeed, the example we are pursuing is of this type. We may rearrange the characteristic polynomial in the form

$$\chi(s) = \phi_0(s) + \sum_{i=1}^N a_i \phi_i(s), \quad (5.36)$$

where the polynomials ϕ_i , $i = 0, 1, \dots, N$ are fixed and given. Assuming that each of the parameters a_i lies in a bounded interval of the form $\underline{a}_i \leq a_i \leq \overline{a}_i$, $i = 1, 2, \dots, N$, the family of polynomials (5.36) forms a *polytope* of polynomials. Then the *edge theorem* (Bartlett et al., 1988) states that to check whether each polynomial in the polytope is Hurwitz it is sufficient to verify whether the polynomials on each of the *exposed edges* of the polytope are Hurwitz. The exposed edges are obtained by fixing $N - 1$ of the parameters p_i at their minimal or maximal value, and varying the remaining parameter over its interval.

Theorem 5.2.9 (Edge theorem). Let \mathcal{D} be a simply connected domain in the complex plane. Then all the roots of each polynomial (5.36) are contained in \mathcal{D} if and only if the roots of all polynomials on the edges of

$$\chi(s) = \phi_0(s) + \sum_{i=1}^N a_i \phi_i(s) \quad (5.37)$$

are in \mathcal{D} . □

A simply connected domain is a domain such that every simple closed contour (i.e., a contour that does not intersect itself) inside the domain encloses only points of the domain. For our purposes \mathcal{D} is the open left-half complex plane. The edges of (5.37) are obtained by fixing $N - 1$ of the parameters a_i at their minimal or maximal value, and varying the remaining parameter over its interval. There are $N2^{N-1}$ edges. Although obviously application of the edge theorem involves much more work than that of Kharitonov's theorem it produces more results.

Example 5.2.10 (Application of the edge theorem). We apply the edge theorem to the third-order uncertain plant. From Example 5.2.1 we know that the closed-loop characteristic polynomial is given by

$$\chi(s) = \theta T_0 s^4 + (\theta + T_0)s^3 + s^2 + g T_d s + g k \quad (5.38)$$

$$= (T_0 s^3 + s^2) + \theta(T_0 s^4 + s^3) + g(T_d s + k). \quad (5.39)$$

Assuming that $0 \leq \theta \leq \bar{\theta}$ and $\underline{g} \leq g \leq \bar{g}$ this forms a polytope. By the edge theorem, we need to check the locations of the roots of the four “exposed edges”

$$\begin{aligned} \theta = 0 : & \quad T_0 s^3 + s^2 + g T_d s + g k, & \underline{g} \leq g \leq \bar{g}, \\ \theta = \bar{\theta} : & \quad \bar{\theta} T_0 s^4 + (\bar{\theta} + T_0)s^3 + s^2 + g T_d s + g k, & \underline{g} \leq g \leq \bar{g}, \\ g = \underline{g} : & \quad \theta T_0 s^4 + (\theta + T_0)s^3 + s^2 + \underline{g} T_d s + \underline{g} k, & 0 \leq \theta \leq \bar{\theta}, \\ g = \bar{g} : & \quad \theta T_0 s^4 + (\theta + T_0)s^3 + s^2 + \bar{g} T_d s + \bar{g} k, & 0 \leq \theta \leq \bar{\theta}. \end{aligned} \quad (5.40)$$

Figure 5.7 shows the patterns traced by the various root loci so defined with T_0 , T_d and k as determined in Example 5.2.2, and $\bar{\theta} = 0.2$, $\underline{g} = 0.5$, $\bar{g} = 5$. This is the case where in Example 5.2.10 application of Kharitonov's theorem was not successful in demonstrating robust stability. Figure 5.7 shows that the four root loci are all contained within the left-half complex plane. By the edge theorem, the closed-loop system is stable for all parameter perturbations that are considered. □

5.2.5. Testing the edges

Actually, to apply the edge theorem it is not necessary to “grid” the edges, as we did in Example 5.2.10. By a result of Białas (1985) the stability of convex combinations of stable polynomials³ p_1 and p_2 of the form

$$p = \lambda p_1 + (1 - \lambda)p_2, \quad \lambda \in [0, 1], \quad (5.41)$$

³Stable polynomial is the same as strictly Hurwitz polynomial.

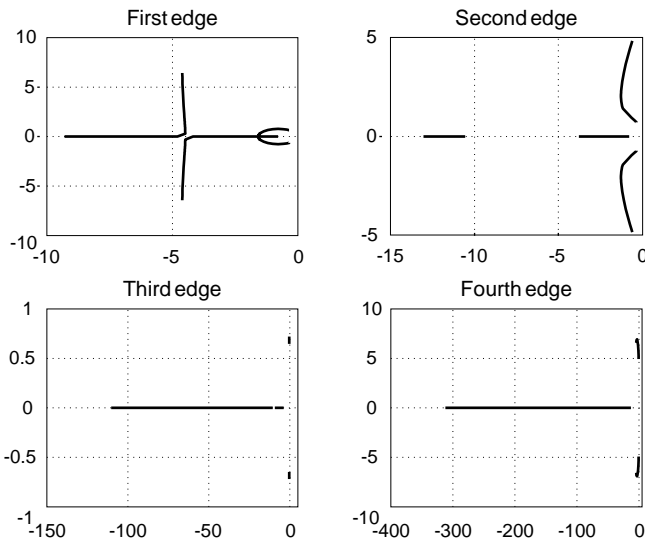


Figure 5.7: Loci of the roots of the four exposed edges

may be established by a single test. Given a polynomial

$$q(s) = q_0 s^n + q_1 s^{n-1} + q_2 s^{n-2} + \cdots + q_n, \quad (5.42)$$

define its *Hurwitz matrix* Q (Gantmacher, 1964) as the $n \times n$ matrix

$$Q = \begin{bmatrix} q_1 & q_3 & q_5 & \cdots & \cdots & \cdots \\ q_0 & q_2 & q_4 & \cdots & \cdots & \cdots \\ 0 & q_1 & q_3 & q_5 & \cdots & \cdots \\ 0 & q_0 & q_2 & q_4 & \cdots & \cdots \\ 0 & 0 & q_1 & q_3 & q_5 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix}. \quad (5.43)$$

Białaś' result may be rendered as follows.

Summary 5.2.11 (Białaś' test). Suppose that the polynomials p_1 and p_2 are strictly Hurwitz with their leading coefficient nonnegative and the remaining coefficients positive. Let P_1 and P_2 be their Hurwitz matrices, and define the matrix

$$W = -P_1 P_2^{-1}. \quad (5.44)$$

Then each of the polynomials

$$p = \lambda p_1 + (1 - \lambda) p_2, \quad \lambda \in [0, 1], \quad (5.45)$$

is strictly Hurwitz iff the real eigenvalues of W all are strictly negative. \square

Note that no restrictions are imposed on the non-real eigenvalues of W .

Example 5.2.12 (Application of Białaś' test). We apply Białaś' test to the third-order plant. In Example 5.2.10 we found that by the edge theorem we need to check the locations of the roots

of the four “exposed edges”

$$\begin{aligned}
 T_0 s^3 + s^2 + \underline{g} T_d s + \underline{g} k, & \quad \underline{g} \leq g \leq \bar{g}, \\
 \bar{\theta} T_0 s^4 + (\bar{\theta} + T_0) s^3 + s^2 + \underline{g} T_d s + \underline{g} k, & \quad \underline{g} \leq g \leq \bar{g}, \\
 \theta T_0 s^4 + (\theta + T_0) s^3 + s^2 + \underline{g} T_d s + \underline{g} k, & \quad 0 \leq \theta \leq \bar{\theta}, \\
 \theta T_0 s^4 + (\theta + T_0) s^3 + s^2 + \bar{g} T_d s + \bar{g} k, & \quad 0 \leq \theta \leq \bar{\theta}.
 \end{aligned} \tag{5.46}$$

The first of these families of polynomials is the convex combination of the two polynomials

$$p_1(s) = T_0 s^3 + s^2 + \underline{g} T_d s + \underline{g} k, \tag{5.47}$$

$$p_2(s) = T_0 s^3 + s^2 + \bar{g} T_d s + \bar{g} k, \tag{5.48}$$

which both are strictly Hurwitz for the given numerical values. The Hurwitz matrices of these two polynomials are

$$P_1 = \begin{bmatrix} 1 & \underline{g}k & 0 \\ T_0 & \underline{g}T_d & 0 \\ 0 & 1 & \underline{g}k \end{bmatrix}, \quad P_2 = \begin{bmatrix} 1 & \bar{g}k & 0 \\ T_0 & \bar{g}T_d & 0 \\ 0 & 1 & \bar{g}k \end{bmatrix}, \tag{5.49}$$

respectively. Numerical evaluation, with $T_0 = 1/10$, $T_d = \sqrt{2}$, $k = 1$, $\underline{g} = 0.5$, and $\bar{g} = 5$, yields

$$W = -P_1 P_2^{-1} = \begin{bmatrix} -1.068 & 0.6848 & 0 \\ -0.09685 & -0.03152 & 0 \\ 0.01370 & -0.1370 & -0.1 \end{bmatrix}. \tag{5.50}$$

The eigenvalues of W are -1 , -0.1 , and -0.1 . They are all real and negative. Hence, by Białas’ test, the system is stable on the edge under investigation.

Similarly, it may be checked that the system is stable on the other edges. By the edge theorem, the system is stable for the parameter perturbations studied. \square

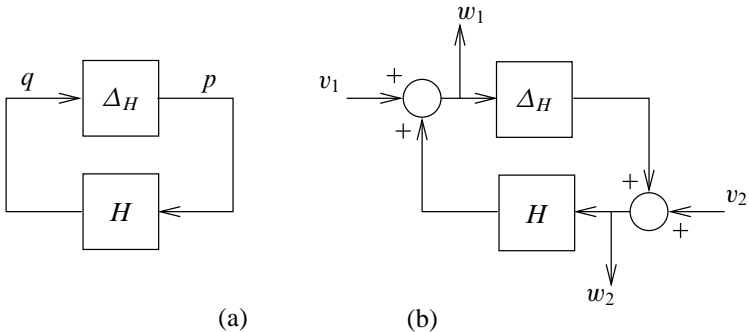


Figure 5.8: (a) Basic perturbation model. (b) Arrangement for internal stability



(b) Additive uncertainty. (c) Multiplicative uncertainty

5.3. The basic perturbation model

This section presents a quite general uncertainty model. It deals with both parametric—or *structured*—and *unstructured* uncertainties, although it is more suited to the latter. Unstructured uncertainties are uncertainties that may involve essential changes in the dynamics of the system. The common assumption in Robust Control is that the uncertainties may be modeled as an I/O-system separate from the rest of the system. The simplest such system is shown in Fig. 5.8. It is sometimes called the *basic perturbation model (BPM)*. Here H is the system whose stability robustness is under investigation. Its transfer matrix H is sometimes called the *interconnection matrix*. The block “ Δ_H ” represents a uncertainty of the dynamics of the system. This uncertainty model is simple but the same time powerful⁴.

Lemma 5.3.1 (Internal stability of the basic perturbation model). Suppose that H and Δ_H are stable systems. The following three statements are equivalent.

1. The BPM of Fig. 5.8(a) is internally stable.
2. $(I - H\Delta_H)^{-1}$ exists and is stable.
3. $(I - \Delta_H H)^{-1}$ exists and is stable.

Many robust stability problems can be brought back to robust stability of the BPM.

Example 5.3.2 (The additive uncertainty model). Figure 5.9(a) shows a feedback loop with loop gain L . After perturbing the loop gain from L to $L + \Delta_L$, the feedback system may be represented as in Fig. 5.9(b). The block within the dashed lines is the unperturbed system, denoted H , and Δ_L is the uncertainty Δ_H in the basic model.

To compute H , denote the input to the uncertainty block Δ_L as q and its output as p , as in Fig. 5.8. Inspection of the diagram of Fig. 5.9(b) shows that with the uncertainty block Δ_L taken away, the system satisfies the signal balance equation $q = -p - Lq$, so that

$$q = -(I + L)^{-1} p. \quad (5.51)$$

It follows that the interconnection matrix H is

$$H = -(I + L)^{-1} = -S, \quad (5.52)$$

⁴It may be attributed to Doyle (1984).

with S the sensitivity matrix of the feedback system. The model of Fig. 5.9(b) is known as the *additive* uncertainty model. \square

Example 5.3.3 (The multiplicative uncertainty model). An alternative uncertainty model, called *multiplicative* or *proportional* uncertainty model, is shown in Fig. 5.9(c). The transfer matrix of the perturbed loop gain is

$$(1 + \Delta_L)L, \quad (5.53)$$

which is equivalent to an additive uncertainty $\Delta_L L$. The quantity Δ_L may be viewed as the *relative size* of the uncertainty. From the signal balance $q = L(-p - q)$ we obtain $q = -(I + L)^{-1}Lp$, so that the interconnection matrix is

$$H = -(I + L)^{-1}L = -T. \quad (5.54)$$

T is the complementary sensitivity matrix of the feedback system. \square

In the above two examples internal stability of the BPM is equivalent to internal stability of the underlying additive and multiplicative uncertain feedback systems. So as far as robust stability is concerned we may as well consider the simple BPM and forget about the internal structure of the interconnection matrix H . This explains the interest in the BPM. The following important theorem shows that the BPM is useful for *any* interconnected system, not just the for the above two unit feedback loops.

Theorem 5.3.4 (A detectability theorem). Suppose we have a system Σ of interconnections of subsystems $C_1, C_2, \dots, P_1, P_2, \dots$ and $\Delta_1, \Delta_2, \dots$ and suppose that Σ is internally stable for zero uncertainties $\Delta_1 = 0, \Delta_2 = 0, \dots$. Then for any set of stable uncertainties $\Delta_1, \Delta_2, \dots$ the following two statements are equivalent.

1. The overall interconnected system Σ is internally stable,
2. the corresponding BPM is internally stable.

\square

Typically the controller is designed to work well for the *nominal* system, which is the system when all uncertainties are taken zero $\Delta_i = 0$. In particular then the controller stabilizes the system for all $\Delta_i = 0$. So this assumption needed in the above theorem is not stringent.

Example 5.3.5 (Parameter perturbation). Consider the third-order plant of Example 5.2.1 with transfer function

$$P(s) = \frac{g}{s^2(1 + s\theta)}. \quad (5.55)$$

The parameters g and θ are uncertain. It is not difficult to represent this transfer function by a block diagram where the parameters g and θ are found in distinct blocks. Figure 5.10 shows one way of doing this. It does not matter that one of the blocks is a pure differentiator. Note the way the factor $1 + s\theta$ in the denominator is handled by incorporating a feedback loop.

After perturbing g to $g + \Delta_g$ and θ to $\theta + \Delta_\theta$ and including a feedback compensator with transfer function C the block diagram may be arranged as in Fig. 5.11. The large block inside the dashed lines is the interconnection matrix H of the basic perturbation model.

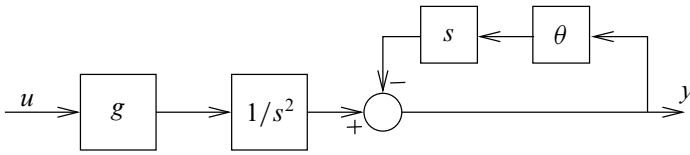


Figure 5.10: Block diagram for third-order plant

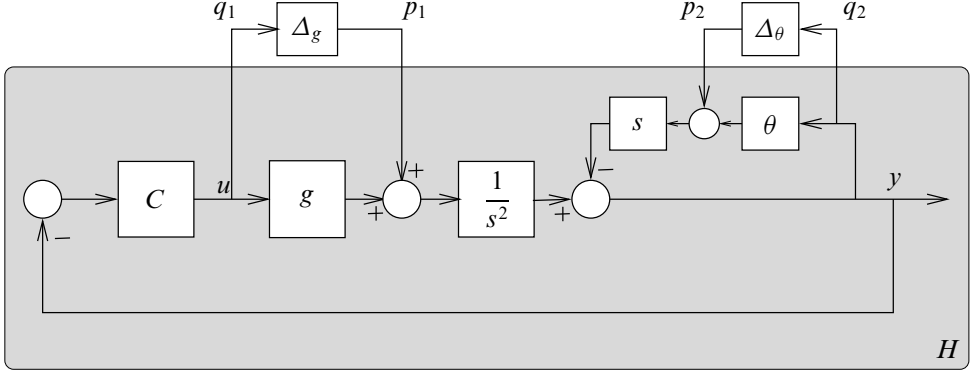


Figure 5.11: Application of the the basic uncertainty model to the third-order plant

To compute the interconnection matrix H we consider the block diagram of Fig. 5.11. Inspection shows that with the perturbation blocks Δ_g and Δ_θ omitted, the system satisfies the signal balance equation

$$q_2 = -s(p_2 + \theta q_2) + \frac{1}{s^2}(p_1 - gC(s)q_2). \quad (5.56)$$

Solution for q_2 yields

$$q_2 = \frac{1/s^2}{1 + s\theta + C(s)g/s^2} p_1 - \frac{s}{1 + s\theta + C(s)g/s^2} p_2. \quad (5.57)$$

Further inspection reveals that $q_1 = -C(s)q_2$, so that

$$q_1 = -\frac{C(s)/s^2}{1 + s\theta + C(s)g/s^2} p_1 + \frac{sC(s)}{1 + s\theta + C(s)g/s^2} p_2. \quad (5.58)$$

Consequently, the interconnection matrix is

$$H(s) = \frac{1}{1 + s\theta + \frac{C(s)g}{s^2}} \begin{bmatrix} -\frac{C(s)}{s^2} & sC(s) \\ \frac{1}{s^2} & -s \end{bmatrix} \quad (5.59)$$

$$= \frac{\frac{1}{1+s\theta}}{1 + L(s)} \begin{bmatrix} C(s) \\ -1 \end{bmatrix} \begin{bmatrix} -\frac{1}{s^2} & s \end{bmatrix}, \quad (5.60)$$

where $L = PC$ is the (unperturbed) loop gain. \square

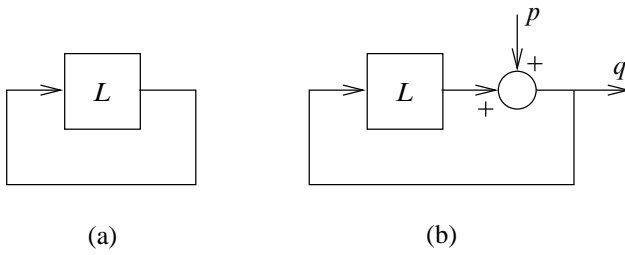


Figure 5.12: Left: feedback loop. Right: feedback loop with internal input and output.

5.4. The small gain theorem

In this section we analyze the stability of the basic uncertainty model using what is known as the *small gain theorem*. This section and the rest of the lecture notes assumes familiarity with the material of Appendix B on norms of signals and systems.

Theorem 5.4.1 (Small gain theorem). In the feedback loop of Fig. 5.12(a), suppose that $L : \mathcal{U} \rightarrow \mathcal{U}$ for some complete normed space \mathcal{U} . Then a sufficient condition for the feedback loop of Fig. 5.12(a) to be internally stable is that the input-output map L is a contraction. If L is linear then it is a contraction if and only if

$$\|L\| < 1, \quad (5.61)$$

where $\|\cdot\|$ is the norm induced by the signal norm,

$$\|L\| := \sup_{u \in \mathcal{U}} \frac{\|Lu\|_{\mathcal{U}}}{\|u\|_{\mathcal{U}}}.$$

□

A proof is given in Appendix 5.8. The small gain theorem is a classic result in systems theory (see for instance Desoer and Vidyasagar (1975)). The small gain theorem gives a *sufficient* condition for internal stability that is very simple but often also very conservative. The power of the small gain theorem is that it does not only apply to SISO linear time-invariant systems but also to MIMO systems and even nonlinear time-varying systems (for that the theorem has to be reformulated somewhat, see Subsection 5.4.2). The signal spaces of signals of finite \mathcal{L}_2 -norm $\|u\|_{\mathcal{L}_2}$ and the signals of finite amplitude $\|u\|_{\mathcal{L}_\infty}$ are two examples of complete normed spaces (see Appendix B), hence, for these norms the small gain theorem applies. Please consult Appendix B about norms of signals and systems. Here we summarize the for us two most important norms. The \mathcal{L}_2 -norm of signals is defined as

$$\|z\|_{\mathcal{L}_2} = \left(\int_{-\infty}^{\infty} z(t)^T z(t) dt \right)^{1/2}.$$

and the system norm induced by this \mathcal{L}_2 -norm:

$$\|H\|_{\infty} := \sup_u \frac{\|Hu\|_{\mathcal{L}_2}}{\|u\|_{\mathcal{L}_2}}$$

In the literature this system norm is denoted as $\|H\|_{\mathcal{H}_\infty}$ in honor of the mathematician Hardy, or as $\|H\|_{\infty}$. We adopt the latter notation. The subscript ∞ in this notation may seem awkward, but the reason is the following very useful result:

Lemma 5.4.2 (∞ -norm). For an LTI system $y = Hu$ there holds

$$\|H\|_{\infty} = \sup_{\operatorname{Re} s > 0} \overline{\sigma}(H(s)).$$

If H is rational then $\|H\|_{\infty}$ exists (that is, is finite) iff H is proper and stable and in that case

$$\|H\|_{\infty} = \sup_{\omega \in \mathbb{R}} \overline{\sigma}(H(j\omega)).$$

□

Here $\overline{\sigma}$ denotes the largest singular value (see Appendix B).

Example 5.4.3 (Small gain theorem). Consider a feedback system as in Fig. 5.12(a), where L is a linear time-invariant system with transfer function

$$L(s) = -\frac{k}{1 + s\theta}. \quad (5.62)$$

θ is a positive time constant and k a positive or negative gain. We investigate the stability of the feedback system with the help of the small gain theorem.

1. *Norm induced by the \mathcal{L}_{∞} -norm.* First consider BIBO stability in the sense of the \mathcal{L}_{∞} -norm on the input and output signals. By inverse Laplace transformation it follows that the impulse response corresponding to L is given by

$$l(t) = \begin{cases} -\frac{k}{\theta} e^{-t/\theta} & \text{for } t \geq 0, \\ 0 & \text{otherwise.} \end{cases} \quad (5.63)$$

The norm $\|L\|$ of the system induced by the \mathcal{L}_{∞} -norm is (see Summary B.4.4)

$$\|L\| = \|l\|_{\mathcal{L}_1} = \int_{-\infty}^{\infty} |l(t)| dt = |k|. \quad (5.64)$$

Hence, by the small gain theorem a sufficient condition for internal stability of the feedback system is that

$$-1 < k < 1. \quad (5.65)$$

2. *Norm induced by the \mathcal{L}_2 -norm.* For positive θ the system L is stable so according to Summary B.4.4 the \mathcal{L}_2 -induced norm exists and equals

$$\|L\|_{\infty} = \sup_{\omega \in \mathbb{R}} |L(j\omega)| = \sup_{\omega \in \mathbb{R}} \left| \frac{k}{1 + j\omega\theta} \right| = |k|. \quad (5.66)$$

Again we find from the small gain theorem that (5.65) is a sufficient condition for closed-loop stability.

In conclusion, we determine the exact stability region. The feedback equation of the system of Fig. 5.12(b) is $w = v + Lw$. Interpreting this in terms of Laplace transforms we may solve for w to obtain

$$w = \frac{1}{1 - L} v, \quad (5.67)$$

so that the closed-loop transfer function is

$$\frac{1}{1-L(s)} = \frac{1}{1 + \frac{k}{1+s\theta}} = \frac{1+s\theta}{(1+k) + s\theta}. \quad (5.68)$$

Inspection shows that the closed-loop transfer function has a single pole at $-(1+k)/\theta$. The corresponding system is BIBO stable if and only if $1+k > 0$. The closed-loop system is internally stable if and only if

$$k > -1. \quad (5.69)$$

This stability region is much larger than that indicated by (5.65). □

The small gain theorem applied to the BPM establishes that stability of H and Δ_H in combination with the small gain condition $\|H\Delta_H\| < 1$ imply internal stability of the BPM. However this by itself is not in a very useful form since typically we do not know the uncertainty Δ_H , so verifying the small gain condition $\|H\Delta_H\| < 1$ can not be done. It is useful to reformulate the small gain theorem in such a way that the known part (the interconnection matrix H) appears separate from the unknown part (the uncertainty matrix Δ_H). Different such forms exist:

Theorem 5.4.4 (Small gain theorems for the BPM). Suppose that in the basic uncertainty model of Fig. 5.8(a) both H and Δ_H are \mathcal{L}_2 -stable.

1. Sufficient for internal stability is that

$$\bar{\sigma}(\Delta_H(j\omega)) < \frac{1}{\bar{\sigma}(H(j\omega))} \quad \text{for all } \omega \in \mathbb{R} \cup \{\infty\}, \quad (5.70)$$

with $\bar{\sigma}$ denoting the largest singular value.

2. Another sufficient condition for internal stability is that

$$\|\Delta_H\|_\infty < \frac{1}{\|H\|_\infty}. \quad (5.71)$$

□

Inequalities (5.70,5.71) are a *sufficient* conditions for internal stability. It is easy to find examples that admit uncertainties that violate these conditions but at the same time do *not* destabilize the system. It may be proved, though⁵, that if robust stability is desired for *all* uncertainties satisfying (5.71) then the condition is also necessary. This means that it is always possible to find a uncertainty that violates (5.71) within an arbitrarily small margin but destabilizes the system:

Theorem 5.4.5 (Necessary and sufficient stability conditions). Suppose that in the BPM of Fig. 5.8(a) both H and Δ_H are stable. Let γ be a positive constant. Then

1. The BPM is internally for all Δ_H with $\|\Delta_H\|_\infty \leq 1/\gamma$ iff $\|H\|_\infty < \gamma$.
2. The BPM is internally for all Δ_H with $\|\Delta_H\|_\infty < 1/\gamma$ iff $\|H\|_\infty \leq \gamma$.

□

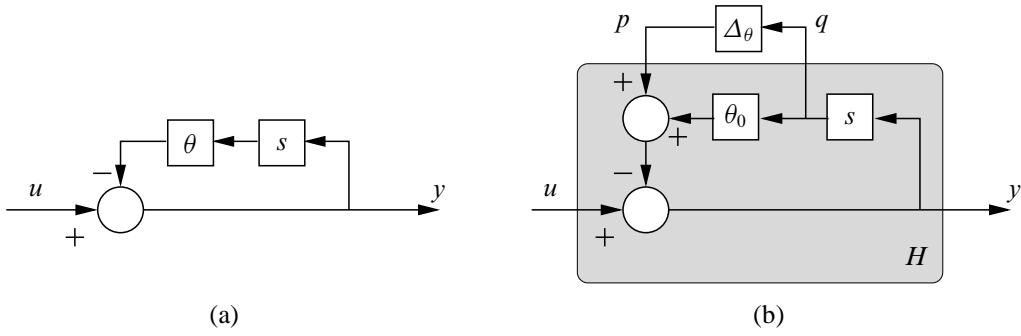


Figure 5.13: (a) Block diagram. (b) Uncertainty model

5.4.1. Examples

We consider two applications of the basic stability robustness result.

Example 5.4.6 (Stability under uncertainty). By way of example, we study which uncertainties of the real parameter θ preserve the stability of the system with transfer function

$$P(s) = \frac{1}{1 + s\theta}. \quad (5.72)$$

Arranging the system as in Fig. 5.13(a) we obtain the uncertainty model of Fig. 5.13(b), with θ_0 the nominal value of θ and Δ_θ its uncertainty. Because the system should be nominally stable we need θ_0 to be positive.

The interconnection matrix H of the standard uncertainty model is obtained from the signal balance equation $q = -s(p + \theta_0 q)$ (omitting u and y). Solution for q yields

$$q = \frac{-s}{1 + s\theta_0} p. \quad (5.73)$$

It follows that

$$H(s) = \frac{-s}{1 + s\theta_0} = \frac{-s\theta_0}{1 + s\theta_0} \frac{1}{\theta_0}, \quad (5.74)$$

so that

$$\overline{\sigma}(H(j\omega)) = |H(j\omega)| = \frac{1}{\theta_0} \sqrt{\frac{\omega^2 \theta_0^2}{1 + \omega^2 \theta_0^2}} \quad (5.75)$$

and

$$\|H\|_\infty = \sup_\omega \overline{\sigma}(H(j\omega)) = \frac{1}{\theta_0}. \quad (5.76)$$

Since $\overline{\sigma}(\Delta_H(j\omega)) = |\Delta_\theta| = \|\Delta_H\|_\infty$, both (1) and (2) of Summary 5.4.4 imply that internal stability is guaranteed for all uncertainties such that $|\Delta_\theta| < \theta_0$, or

$$-\theta_0 < \Delta_\theta < \theta_0. \quad (5.77)$$

⁵See (Vidysagar, 1985) for a proof for the rational case.

Obviously, though, the system is stable for all $\theta > 0$, that is, for all uncertainties Δ_θ such that

$$-\theta_0 < \Delta_\theta < \infty. \quad (5.78)$$

Hence, the estimate for the stability region is quite conservative. On the other hand, it is easy to find a uncertainty that violates (5.77) and destabilizes the system. The uncertainty $\Delta_\theta = -\theta_0(1 + \varepsilon)$, for instance, with ε positive but arbitrarily small, violates (5.77) with an arbitrarily small margin, and *destabilizes* the system (because it makes θ negative).

Note that the class of uncertainties such that $\|\Delta_\theta\|_\infty \leq \theta_0$ is much larger than just real uncertainties satisfying (5.77). For instance, Δ_θ could be a transfer function, such as

$$\Delta_\theta(s) = \theta_0 \frac{\alpha}{1 + s\tau}, \quad (5.79)$$

with τ any positive time constant, and α a real number with magnitude less than 1. This “parasitic” uncertainty leaves the system stable. \square

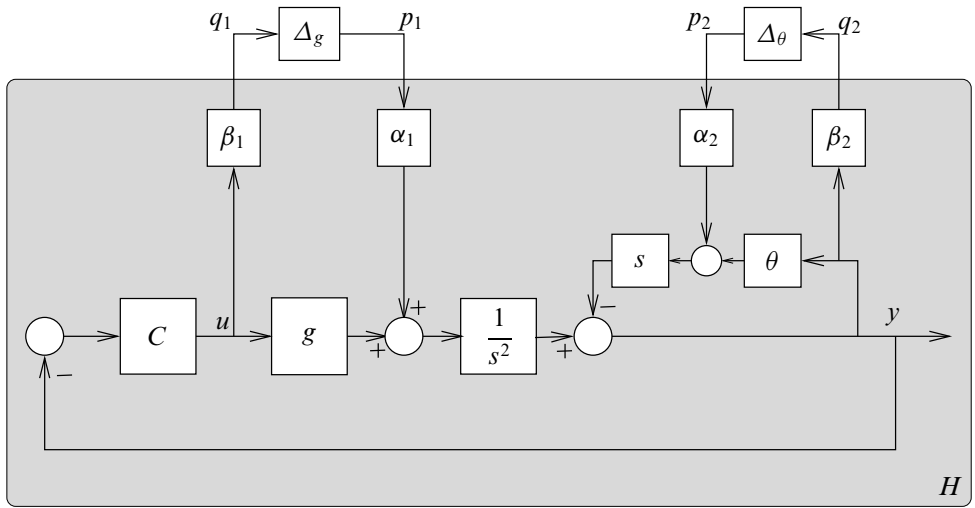


Figure 5.14: Scaled uncertainty model for the third-order plant

Example 5.4.7 (Two-parameter uncertainty). A more complicated example is the feedback system discussed in Examples 5.2.1, 5.2.2 and 5.3.5. It consists of a plant and a compensator with transfer functions

$$P(s) = \frac{g}{s^2(1 + s\theta)}, \quad C(s) = \frac{k + T_d s}{1 + T_0 s}, \quad (5.80)$$

respectively. In Example 5.3.5 we found that the interconnection matrix H with respect to uncertainties in the parameters g and θ is

$$H(s) = \frac{\frac{1}{1+s\theta_0}}{1 + L_0(s)} \begin{bmatrix} C(s) \\ -1 \end{bmatrix} \begin{bmatrix} -\frac{1}{s^2} & s \end{bmatrix}, \quad (5.81)$$

with g_0 and θ_0 denoting the nominal values of the two uncertain parameters g and θ . Before continuing the analysis we modify the uncertainty model of Fig. 5.11 to that of Fig. 5.14. This

model includes scaling factors α_1 and β_1 for the uncertainty Δ_g and scaling factors α_2 and β_2 for the uncertainty Δ_θ . The scaled uncertainties are denoted δ_g and δ_θ , respectively. The product $\alpha_1\beta_1 = \varepsilon_1$ is the largest possible uncertainty in g , while $\alpha_2\beta_2 = \varepsilon_2$ is the largest possible uncertainty in θ . It is easily verified that the interconnection matrix corresponding to Fig. 5.14 is

$$H(s) = \frac{\frac{1}{1+s\theta_0}}{1+L_0(s)} \begin{bmatrix} \beta_1 C(s) \\ -\beta_2 \end{bmatrix} \begin{bmatrix} -\frac{\alpha_1}{s^2} & \alpha_2 s \end{bmatrix}. \quad (5.82)$$

It may also easily be worked out that the largest eigenvalue of $H^T(-j\omega)H(j\omega)$ is

$$\bar{\sigma}^2(\omega) = \frac{\frac{1}{1+\omega^2\theta_0^2}}{|1+L_0(j\omega)|^2} (\beta_1^2 |C(j\omega)|^2 + \beta_2^2) \left(\frac{\alpha_1^2}{\omega^4} + \alpha_2^2 \omega^2 \right), \quad \omega \in \mathbb{R}. \quad (5.83)$$

The other eigenvalue is 0. The nonnegative quantity $\bar{\sigma}(\omega)$ is the largest singular value of $H(j\omega)$. By substitution of L_0 and C into (5.83) it follows that

$$\bar{\sigma}^2(\omega) = \frac{(\beta_1^2(k^2 + \omega^2 T_d^2) + \beta_2^2(1 + \omega^2 T_0^2))(\alpha_1^2 + \alpha_2^2 \omega^6)}{|\chi(j\omega)|^2}, \quad (5.84)$$

with χ the closed-loop characteristic polynomial

$$\chi(s) = \theta_0 T_0 s^4 + (\theta_0 + T_0) s^3 + s^2 + g_0 T_d s + g_0 k. \quad (5.85)$$

First we note that if we choose the nominal value θ_0 of the parasitic time constant θ equal to zero — which is a natural choice — and $\varepsilon_2 = \alpha_2\beta_2 \neq 0$ then $\bar{\sigma}(\infty) = \infty$, so that stability is not guaranteed. Indeed, this situation admits negative values of θ , which destabilize the closed-loop system. Hence, we need to choose θ_0 positive but such that the nominal closed-loop system is stable, of course.

Second, inspection of (5.84) shows that for fixed uncertainties $\varepsilon_1 = \alpha_1\beta_1$ and $\varepsilon_2 = \alpha_2\beta_2$ the function $\bar{\sigma}$ depends on the way the individual values of the scaling constants α_1 , β_1 , α_2 , and β_2 are chosen. On first sight this seems surprising. Reflection reveals that this phenomenon is caused by the fact that the stability robustness test is based on *full* uncertainties Δ_H . Full uncertainties are uncertainties such that all entries of the uncertainty transfer matrix Δ_H are filled with dynamical uncertainties.

We choose the numerical values $k = 1$, $T_d = \sqrt{2}$, and $T_0 = 1/10$ as in Example 5.2.2. In Example 5.2.8 we consider variations in the time constant θ between 0 and 0.2. Correspondingly, we let $\theta_0 = 0.1$ and $\varepsilon_2 = 0.1$. For the variations in the gain g we study two cases as in Example 5.2.8:

1. $0.5 \leq g \leq 5$. Correspondingly, we let $g_0 = 2.75$ and $\varepsilon_1 = 2.25$.
2. $0.5 \leq g \leq 2$. Correspondingly, we take $g_0 = 1.25$ and $\varepsilon_1 = 0.75$.

For lack of a better choice we select for the time being

$$\alpha_1 = \beta_1 = \sqrt{\varepsilon_1}, \quad \alpha_2 = \beta_2 = \sqrt{\varepsilon_2}. \quad (5.86)$$

Figure 5.15 shows the resulting plots of $\bar{\sigma}(\omega)$ for the cases (1) and (2). In case (1) the peak value is about 67.1, while in case (2) it is approximately 38.7. Both cases fail the stability robustness test by a very wide margin, although from Example 5.2.4 we know that in both situations stability is ensured. The reasons that the test fails are that (i) the test is based on full uncertainties rather than the structured uncertainties implied by the variations in the two parameters, and (ii) the test

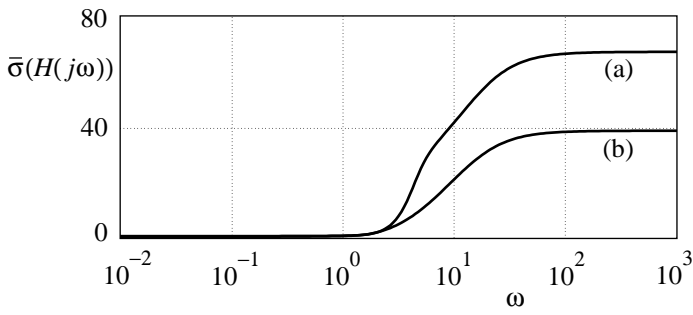


Figure 5.15: Plots of $\bar{\sigma}(H(j\omega))$ for the two-parameter plant

also allows *dynamical* uncertainties rather than the real uncertainties implied by the parameter variations.

Less conservative results may be obtained by allowing the scaling factors α_1 , β_1 , α_2 , and β_2 to be frequency-dependent, and to choose them so as to minimize $\bar{\sigma}(\omega)$ for each ω . Substituting $\beta_1 = \varepsilon_1/\alpha_1$ and $\beta_2 = \varepsilon_2/\alpha_2$ we obtain

$$\bar{\sigma}^2(\omega) = \frac{\varepsilon_1^2(k^2 + \omega^2 T_d^2) + \frac{\varepsilon_1^2(k^2 + \omega^2 T_d^2)\omega^6}{\rho} + \varepsilon_2^2(1 + \omega^2 T_0^2)\rho + \varepsilon_2^2(1 + \omega^2 T_0^2)\omega^6}{|\chi(j\omega)|^2}, \quad (5.87)$$

where $\rho = \alpha_1^2/\alpha_2^2$. It is easy to establish that for fixed ω the quantity $\bar{\sigma}^2(\omega)$ is minimized for

$$\rho = \omega^3 \frac{\varepsilon_1}{\varepsilon_2} \sqrt{\frac{k^2 + \omega^2 T_d^2}{1 + \omega^2 T_0^2}}, \quad (5.88)$$

and that for this value of ρ

$$\bar{\sigma}(\omega) = \frac{\varepsilon_1 \sqrt{k^2 + \omega^2 T_d^2} + \varepsilon_2 \omega^3 \sqrt{1 + \omega^2 T_0^2}}{|\chi(j\omega)|}, \quad \omega \geq 0. \quad (5.89)$$

Figure 5.16 shows plots of $\bar{\sigma}$ for the same two cases (a) and (b) as before. In case (a) the peak value of $\bar{\sigma}$ is about 1.83, while in case (b) it is 1. Hence, only in case (b) robust stability is established. The reason that in case (a) robust stability is not proved is that the uncertainty model allows dynamic uncertainties. Only if the size of the uncertainties is downsized by a factor of almost 2 robust stability is guaranteed.

In Example 5.6.8 in the section on the structured singular value this example is further discussed. □

5.4.2. Nonlinear perturbations

The basic stability robustness result also applies for *nonlinear* perturbations. Suppose that the perturbation Δ_H in the block diagram of Fig. 5.8(a) is a nonlinear operator that maps the signal $e(t)$, $t \in \mathbb{R}$, into the signal $(\Delta_H e)(t)$, $t \in \mathbb{R}$. Assume that there exists a positive constant k such that

$$\|\Delta_H e_1 - \Delta_H e_2\| \leq k \|e_1 - e_2\| \quad (5.90)$$

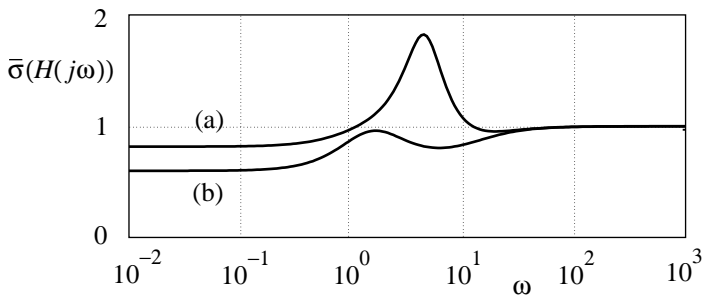


Figure 5.16: Plots of $\bar{\sigma}(H(j\omega))$ with optimal scaling for the two-parameter plant

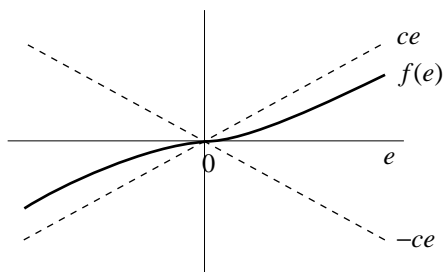


Figure 5.17: Sector bounded function

for every two input e_1, e_2 to the nonlinearity. We use the \mathcal{L}_2 -norm for signals. By application of the fixed point theorem it follows that the perturbed system of Fig. 5.8(a) is stable if

$$k\|H\|_\infty < 1. \quad (5.91)$$

Suppose for instance that the signals are scalar, and that Δ_H is a static nonlinearity described by

$$(\Delta_H e)(t) = f(e(t)), \quad (5.92)$$

with $f : \mathbb{R} \rightarrow \mathbb{R}$. Let f satisfy the inequality

$$\left| \frac{f(e)}{e} \right| \leq c, \quad e \neq 0, e \in \mathbb{R}, \quad (5.93)$$

with c a nonnegative constant. Figure 5.17 illustrates the way the function f is bounded. We call f a *sector bounded function*. It is not difficult to see that (5.90) holds with $k = c$. It follows that the perturbed system is stable for any static nonlinear perturbation satisfying (5.93) as long as $c \leq 1/\|H\|_\infty$.

Similarly, the basic stability robustness result holds for *time-varying* perturbations and perturbations that are both nonlinear and time-varying, provided that these perturbations are suitably bounded.

5.5. Stability robustness of feedback systems

In this section we apply the results of § 5.4 to various perturbation models for single-degree-of-freedom feedback systems. We successively discuss proportional, proportional inverse and

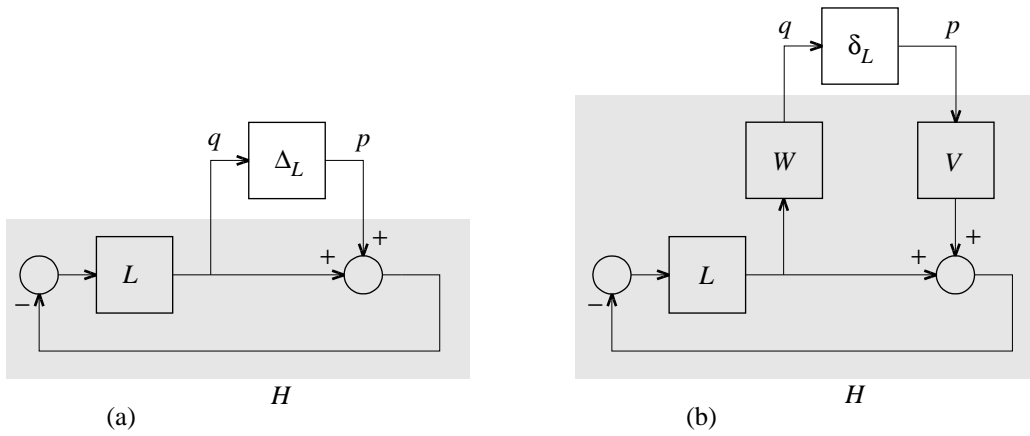


Figure 5.18: Proportional and scaled perturbations of a unit feedback loop

fractional perturbations for MIMO systems. The results of § 1.4 emerge as special cases.

5.5.1. Proportional perturbations

In § 5.3 we consider additive and multiplicative (or *proportional*) perturbation models for single-degree-of-freedom feedback systems. The proportional model is preferred because it represents the *relative* size of the perturbation. The corresponding block diagram is repeated in Fig. 5.18(a). It represents a perturbation

$$L \longrightarrow (I + \Delta_L)L \quad (5.94)$$

Since $q = -L(p + q)$, so that $q = -(I + L)^{-1}Lp$, the interconnection matrix is

$$H = -(I + L)^{-1}L = -T, \quad (5.95)$$

with $T = (I + L)^{-1}L = L(I + L)^{-1}$ the *complementary sensitivity matrix* of the closed-loop system.

To *scale* the proportional perturbation we modify the block diagram of Fig. 5.18(a) to that of Fig. 5.18(b). This diagram represents a perturbation

$$L \longrightarrow (I + V\delta_L W)L, \quad (5.96)$$

where V and W are available to scale such that $\|\delta_L\|_\infty \leq 1$. For this configuration the interconnection matrix is $H = -WTV$.

Application of the results of Summary 5.4.4 leads to the following conclusions.

Summary 5.5.1 (Robust stability of feedback systems for proportional perturbations).

Assume that the feedback system of Fig. 5.18(a) is nominally stable.

1. A sufficient condition for the closed-loop system to be stable under proportional perturbations of the form

$$L \longrightarrow (1 + \Delta_L)L \quad (5.97)$$

is that Δ_L be stable with

$$\bar{\sigma}(\Delta_L(j\omega)) < \frac{1}{\bar{\sigma}(T(j\omega))} \quad \text{for all } \omega \in \mathbb{R}. \quad (5.98)$$

with $T = (I + L)^{-1}L = L(I + L)^{-1}$ the complementary sensitivity matrix of the feedback loop. If stability is required for *all* perturbations satisfying the bound then the condition is also necessary.

2. Under *scaled* perturbations

$$L \longrightarrow (1 + V\delta_L W)L \quad (5.99)$$

the system is stable for all $\|\delta_L\|_\infty \leq 1$ if and only if

$$\|WTV\|_\infty < 1. \quad (5.100)$$

□

For SISO systems part (1) of this result reduces to Doyle’s robustness criterion of § 1.4. In fact, Summary 5.5.1 is the original MIMO version of Doyle’s robustness criterion (Doyle, 1979).

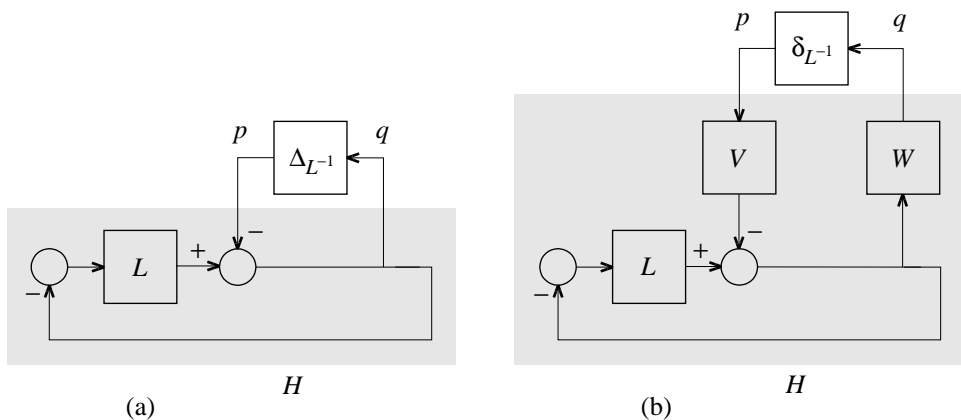


Figure 5.19: Proportional inverse perturbation of unit feedback loop

5.5.2. Proportional inverse perturbations

The result of the preceding subsection confirms the importance of the complementary sensitivity matrix (or function) T for robustness. We complement it with a “dual” result involving the sensitivity function S . To this end, consider the perturbation model of Fig. 5.19(a), where the perturbation Δ_{I-1} is included in a feedback loop. The model represents a perturbation

$$L \longrightarrow (I + \Delta_{I^{-1}})^{-1}L. \quad (5.101)$$

Assuming that L has an inverse, this may be rewritten as the perturbation

$$L^{-1} \longrightarrow L^{-1}(I + \Delta_{I^{-1}}). \quad (5.102)$$

Hence, the perturbation represents a proportional perturbation of the *inverse* loop gain. The interconnection matrix H follows from the signal balance $q = -p - Lq$ so that $q = -(I + L)^{-1}p$. Hence, the interconnection matrix is

$$H = -(I + L)^{-1} = -S. \quad (5.103)$$

$S = (I + L)^{-1}$ is the sensitivity matrix of the feedback loop.

We allow for scaling by modifying the model to that of Fig. 5.19(b). This model represents perturbations of the form

$$L^{-1} \longrightarrow L^{-1}(I + V\delta_{L^{-1}}W), \quad (5.104)$$

where V and W provide freedom to scale such that $\|\delta_{L^{-1}}\|_{\infty} \leq 1$. The interconnection matrix now is $H = -WSV$.

Application of the results of Summary 5.4.4 yields the following conclusions.

Summary 5.5.2 (Robust stability of feedback systems for proportional inverse perturbations). Assume that the feedback system of Fig. 5.19(a) and (b) is nominally stable.

1. Under *proportional inverse perturbations* of the form

$$L^{-1} \longrightarrow L^{-1}(I + \Delta_{L^{-1}}) \quad (5.105)$$

a sufficient condition for stability is that $\Delta_{L^{-1}}$ be stable with

$$\bar{\sigma}(\Delta_{L^{-1}}(j\omega)) < \frac{1}{\bar{\sigma}(S(j\omega))} \quad \text{for all } \omega \in \mathbb{R}, \quad (5.106)$$

with $S = (I + L)^{-1}$ the sensitivity matrix of the feedback loop. If stability is required for *all* perturbations satisfying the bound then the condition is also necessary.

2. Under *scaled inverse perturbations* of the form

$$L^{-1} \longrightarrow L^{-1}(I + V\delta_{L^{-1}}W) \quad (5.107)$$

the closed-loop system is stable for all $\|\delta_{L^{-1}}\|_{\infty} \leq 1$ if and only if

$$\|WSV\|_{\infty} < 1. \quad (5.108)$$

□

5.5.3. Example

We illustrate the results of Summaries 5.5.1 and 5.5.2 by application to an example.

Example 5.5.3 (Robustness of a SISO closed-loop system). In Example 5.2.1 we considered a SISO single-degree-of-freedom feedback system with plant and compensator transfer functions

$$P(s) = \frac{g}{s^2(1 + s\theta)}, \quad C(s) = \frac{k + T_d s}{1 + T_0 s}, \quad (5.109)$$

respectively. Nominally the gain g equals $g_0 = 1$, while the parasitic time constant θ is nominally 0. In Example 5.2.2 we chose the compensator parameters as $k = 1$, $T_d = \sqrt{2}$ and $T_0 = 1/10$.

We use the results of Summaries 5.5.1 and 5.5.2 to study what perturbations of the parameters g and θ leave the closed-loop system stable.

1. *Loop gain perturbation model.* Starting with the expression

$$L(s) = P(s)C(s) = \frac{g}{s^2(1+s\theta)} \frac{k+T_d s}{1+T_0 s} \quad (5.110)$$

it is easy to find that the proportional loop gain perturbations are

$$\Delta_L(s) = \frac{L(s) - L_0(s)}{L_0(s)} = \frac{\frac{g-g_0}{g_0} - s\theta}{1+s\theta}. \quad (5.111)$$

Figures 5.20(a) and (b) show the magnitude plot of the inverse $1/T_0$ of the nominal complementary sensitivity function. Inspection shows that $1/T_0$ assumes relatively small values in the low-frequency region. This is where the proportional perturbations of the loop gain need to be the smallest. For $\theta = 0$ the proportional perturbation (5.111) of the loop gain reduces to

$$\Delta_L(s) = \frac{g - g_0}{g_0}, \quad (5.112)$$

and, hence, is constant. The minimal value of the function $1/|T_0|$ is about 0.75. Therefore, for $\theta = 0$ the robustness criterion 1 of Summary 5.5.1 allows relative perturbations of g up to 0.75, so that $0.25 < g < 1.75$.

For $g = g_0$, on the other hand, the proportional perturbation (5.111) of the loop gain reduces to

$$\Delta_L(s) = \frac{-s\theta}{1+s\theta}. \quad (5.113)$$

Figure 5.20(a) shows magnitude plots of this perturbation for several values of θ . The robustness criterion (a) of Summary 5.5.1 permits values of θ up to about 1.15. For this value of θ the gain g can be neither increased nor decreased without violating the criterion.

For smaller values of the parasitic time constant θ a wider range of gains is permitted. The plots of $|\Delta_L|$ of Fig. 5.20(b) demonstrate that for $\theta = 0.2$ the gain g may vary between about 0.255 and 1.745.

The stability bounds on g and θ are conservative. The reason is of course that the perturbation model allows a much wider class of perturbations than just those caused by changes in the parameters g and θ .

2. *Inverse loop gain perturbation model.* The proportional inverse loop gain perturbation is given by

$$\Delta_{L^{-1}}(s) = \frac{L^{-1}(s) - L_0^{-1}(s)}{L_0^{-1}(s)} = \frac{g_0 - g}{g} + s\theta \frac{g_0}{g}. \quad (5.114)$$

We apply the results of Summary 5.5.2. Figure 5.21 gives the magnitude plot of the inverse $1/S_0$ of the nominal sensitivity function. Inspection shows that for the inverse loop gain perturbation model the high-frequency region is the most critical. By inspection of (5.114) we see that if $\theta \neq 0$ then $|\Delta_{L^{-1}}(\infty)| = \infty$, so that robust stability is not ensured. Apparently, this model cannot handle high-frequency perturbations caused by parasitic dynamics.

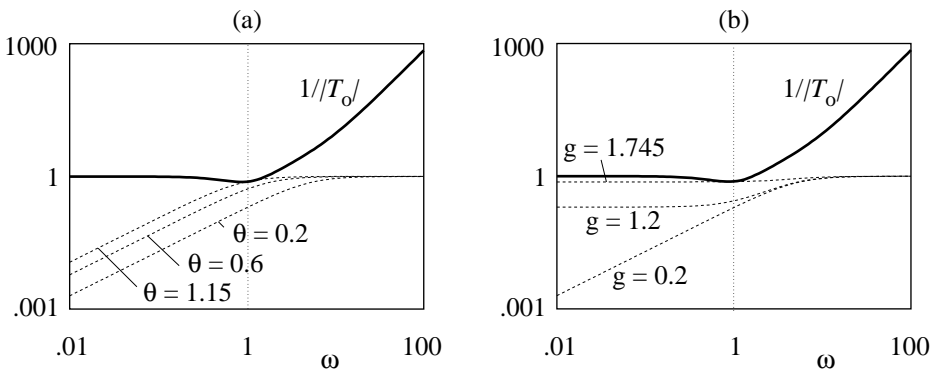


Figure 5.20: Robust stability test for the loop gain perturbation model: (a) $1/|T_o|$ and the relative perturbations for $g = 1$. (b) $1/|T_o|$ and the relative perturbations for $\theta = 0.2$

For $\theta = 0$ the proportional inverse loop gain perturbation reduces to

$$\Delta_{L^{-1}}(s) = \frac{g_0 - g}{g}, \quad (5.115)$$

and, hence, is constant. The magnitude of $1/S_0$ has a minimum of about 0.867, so that for $\theta = 0$ stability is ensured for $|\frac{g_0 - g}{g}| < 0.867$, or $0.536 < g < 7.523$. This range of variation is larger than that found by application of the proportional loop gain perturbation model.

Again, the result is conservative. It is disturbing that the model does not handle parasitic perturbations.

□

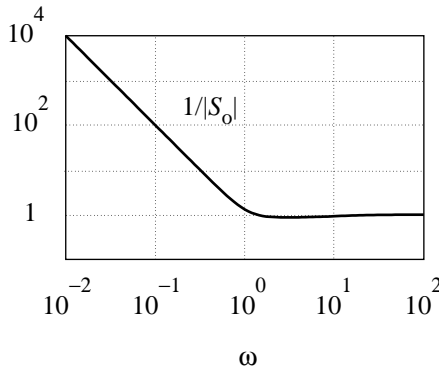


Figure 5.21: Magnitude plot of $1/S_0$

The derivation of the unstructured robust stability tests of Summary 5.5.1 is based on the small gain theorem, and presumes the perturbations Δ_L to be *BIBO stable*. This is highly restrictive if the loop gain L by itself represents an unstable system, which may easily occur (in particular, when the plant by itself is unstable).

It is well-known (Vidysagar, 1985) that the stability assumption on the perturbation may be relaxed to the assumption that both the nominal loop gain and the perturbed loop gain have the *same number of right-half plane poles*.

Likewise, the proofs of the inverse perturbation tests of Summary 5.5.2 require the perturbation of the *inverse* of the loop gain to be stable. This is highly restrictive if the loop gain by itself has right-half plane zeros. This occurs, in particular, when the plant has right-half plane zeros. The requirement, however, may be relaxed to the assumption that the nominal loop gain and the perturbed loop gain have the *same number of right-half plane zeros*.

5.5.4. Fractional representation

The stability robustness analysis of feedback systems based on perturbations of the loop gain or its inverse is simple, but often overly conservative.

Another model that is encountered in the literature⁶ relies on what we term here *fractional perturbations*. It combines, in a way, loop gain perturbations and inverse loop gain perturbations. In this analysis, the loop gain L is represented as

$$L = ND^{-1}, \quad (5.116)$$

where the *denominator* D is a square nonsingular rational or polynomial matrix, and N a rational or polynomial matrix. Any rational transfer matrix L may be represented like this in many ways. If D and N are polynomial then the representation is known as a (right)⁷ *polynomial matrix fraction representation*. If D and N are rational and proper with all their poles in the open left-half complex plane then the representation is known as a (right) *rational matrix fraction representation*.

Example 5.5.4 (Fractional representations). For a SISO system the fractional representation is obvious. Suppose that

$$L(s) = \frac{g}{s^2(1 + s\theta)}. \quad (5.117)$$

Clearly we have the polynomial fractional representation $L = ND^{-1}$ with $N(s) = g$ and $D(s) = s^2(1 + s\theta)$. The fractional representation may be made rational by letting

$$D(s) = \frac{s^2(1 + s\theta)}{d(s)}, \quad N(s) = \frac{g}{d(s)}, \quad (5.118)$$

with d any strictly Hurwitz polynomial. □

Right fractional representation may also be obtained for MIMO systems (see for instance (Vidysagar, 1985)).

5.5.5. Fractional perturbations

Figure 5.22 shows the fractional perturbation model. Inspection shows that the perturbation is given by

$$L \longrightarrow (I + \Delta_N)L(I + \Delta_D)^{-1}, \quad (5.119)$$

⁶The idea originates from Vidysagar (Vidysagar et al., 1982; Vidysagar, 1985). It is elaborated in (McFarlane and Glover, 1990).

⁷Because the denominator is on the right.

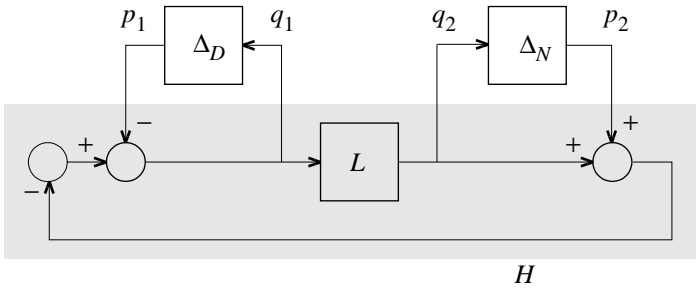


Figure 5.22: Fractional perturbation model

or

$$ND^{-1} \longrightarrow (I + \Delta_N)ND^{-1}(I + \Delta_D)^{-1}. \quad (5.120)$$

Hence, the numerator and denominator are perturbed as

$$N \longrightarrow (I + \Delta_N)N, \quad D \longrightarrow (I + \Delta_D)D. \quad (5.121)$$

Thus, Δ_D and Δ_N represent *proportional perturbations of the denominator and of the numerator*, respectively.

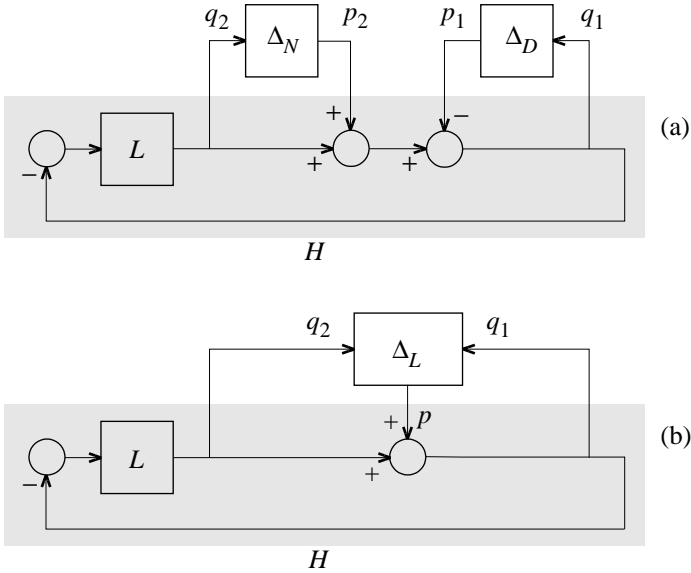


Figure 5.23: Equivalent configurations

By block diagram substitution it is easily seen that the configuration of Fig. 5.22 is equivalent to that of Fig. 5.23(a). The latter, in turn, may be rearranged as in Fig. 5.23(b). Here

$$p = -\Delta_D q_1 + \Delta_N q_2 = \Delta_L \begin{bmatrix} q_1 \\ q_2 \end{bmatrix}, \quad (5.122)$$

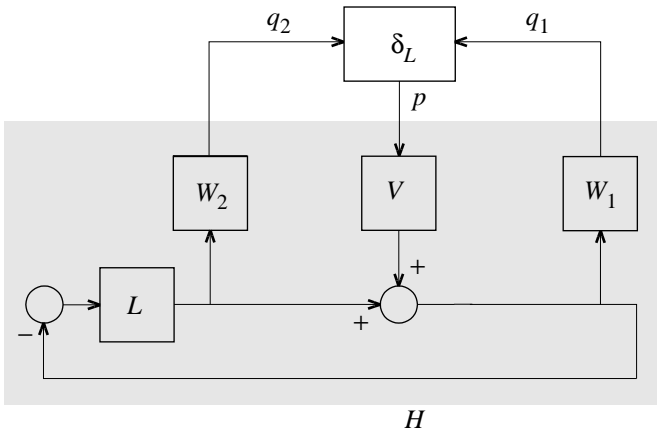


Figure 5.24: Perturbation model with scaling

with

$$\Delta_L = [-\Delta_D \quad \Delta_N]. \quad (5.123)$$

To establish the interconnection matrix H of the configuration of Fig. 5.23(b) we consider the signal balance equation $q_1 = p - Lq_1$. It follows that

$$q_1 = (I + L)^{-1} p = Sp, \quad (5.124)$$

with $S = (I + L)^{-1}$ the sensitivity matrix. Since $q_2 = -Lq_1$ we have

$$q_2 = -L(I + L)^{-1} p = -Tp, \quad (5.125)$$

with $T = L(I + L)^{-1}$ the complementary sensitivity matrix. Inspection of (5.124–5.125) shows that the interconnection matrix is

$$H = \begin{bmatrix} S \\ -T \end{bmatrix}. \quad (5.126)$$

Investigation of the frequency dependence of the greatest singular value of $H(j\omega)$ yields information about the largest possible perturbations Δ_L that leave the loop stable.

It is useful to allow for scaling by modifying the configuration of Fig. 5.23(b) to that of Fig. 5.24. This modification corresponds to representing the perturbations as

$$\Delta_D = V\delta_D W_1, \quad \Delta_N = V\delta_N W_2, \quad (5.127)$$

where V , W_1 , and W_2 are suitably chosen (stable) rational matrices such that

$$\|\delta_L\|_\infty \leq 1, \quad \text{with } \delta_L = [-\delta_D \quad \delta_N]. \quad (5.128)$$

Accordingly, the interconnection matrix changes to

$$H = \begin{bmatrix} W_1 S V \\ -W_2 T V \end{bmatrix}. \quad (5.129)$$

Application of the results of Summary 5.4.4 yields the following conclusions.

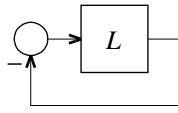


Figure 5.25: Feedback loop

Summary 5.5.5 (Numerator-denominator perturbations). In the stable feedback configuration of Fig. 5.25, suppose that the loop gain is perturbed as

$$L = ND^{-1} \longrightarrow [(I + V\delta_N W_2)N][(I + V\delta_D W_1)D]^{-1} \quad (5.130)$$

with V , W_1 and W_2 stable transfer matrices. Then the closed-loop system is stable for all stable perturbations $\delta_P = [-\delta_D \ \delta_N]$ such that $\|\delta_P\|_\infty \leq 1$ if and only if

$$\|H\|_\infty < 1. \quad (5.131)$$

Here we have

$$H = \begin{bmatrix} W_1 S V \\ -W_2 T V \end{bmatrix}, \quad (5.132)$$

with $S = (I + L)^{-1}$ the sensitivity matrix and $T = L(I + L)^{-1}$ the complementary sensitivity matrix. \square

The largest singular value of $H(j\omega)$, with H given by (5.132), equals the square root of the largest eigenvalue of

$$\begin{aligned} H^T(-j\omega)H(j\omega) &= V^T(-j\omega)[S^T(-j\omega)W_1^T(-j\omega)W_1(j\omega)S(j\omega) \\ &\quad + T^T(-j\omega)W_2^T(-j\omega)W_2(j\omega)T(j\omega)]V(j\omega). \end{aligned} \quad (5.133)$$

For SISO systems this is the scalar function

$$H^T(-j\omega)H(j\omega) = |V(j\omega)|^2[|S(j\omega)|^2|W_1(j\omega)|^2 + |T(j\omega)|^2|W_2(j\omega)|^2]. \quad (5.134)$$

5.5.6. Discussion

We consider how to arrange the fractional perturbation model. In the SISO case, without loss of generality we may take the scaling function V equal to 1. Then W_1 represents the scaling factor for the denominator perturbations and W_2 that for the numerator perturbations. We accordingly have

$$\|H\|_\infty^2 = \sup_{\omega \in \mathbb{R}} (|S(j\omega)|^2|W_1(j\omega)|^2 + |T(j\omega)|^2|W_2(j\omega)|^2). \quad (5.135)$$

For well-designed control systems the sensitivity function S is small at low frequencies while the complementary sensitivity function T is small at high frequencies. Figure 5.26 illustrates this. Hence, W_1 may be large at low frequencies and W_2 large at high frequencies without violating the robustness condition $\|H\|_\infty < 1$. This means that at low frequencies we may allow large denominator perturbations, and at high frequencies large numerator perturbations.

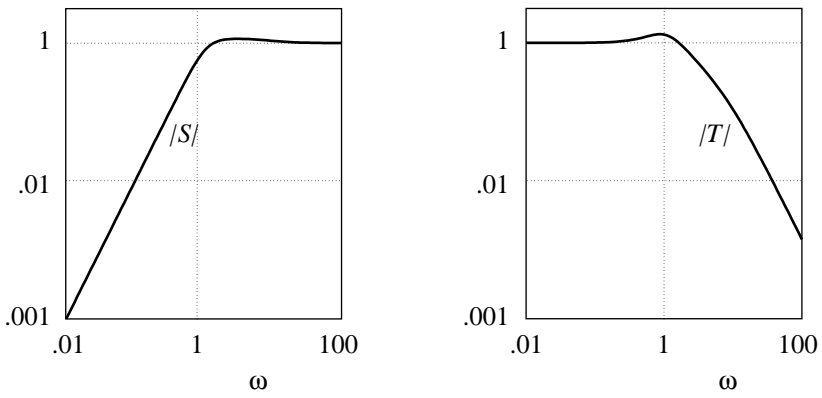


Figure 5.26: Sensitivity function and complementary sensitivity function for a well-designed feedback system

The most extreme point of view is to structure the perturbation model such that all low frequency perturbations are denominator perturbations, and all high frequency perturbations are numerator perturbations. Since we may trivially write

$$L = \frac{1}{\frac{1}{L}}, \quad (5.136)$$

modeling low frequency perturbations as pure denominator perturbations implies modeling low frequency perturbations as *inverse* loop gain perturbations. Likewise, modeling high frequency perturbations as pure numerator perturbations implies modeling high frequency perturbations as loop gain perturbations. This amounts to taking

$$|W_1(j\omega)| = \begin{cases} W_{L^{-1}}(\omega) & \text{at low frequencies,} \\ 0 & \text{at high frequencies,} \end{cases} \quad (5.137)$$

$$|W_2(j\omega)| = \begin{cases} 0 & \text{at low frequencies,} \\ W_L(\omega) & \text{at high frequencies,} \end{cases} \quad (5.138)$$

with $W_{L^{-1}}$ a bound on the size of the inverse loop perturbations, and W_L a bound on the size of the loop perturbations.

Obviously, the boundary between “low” and “high” frequencies lies in the “crossover” region, that is, near the frequency where the loop gain crosses over the zero dB line. In this frequency region neither S nor T is small.

Another way of dealing with this perturbation model is to modify the stability robustness test to checking whether for each $\omega \in \mathbb{R}$

$$|\Delta_{L^{-1}}(j\omega)| < \frac{1}{|S(j\omega)|} \quad \text{or} \quad |\Delta_L(j\omega)| < \frac{1}{|T(j\omega)|}. \quad (5.139)$$

This test amounts to verifying whether either the proportional loop gain perturbation test succeeds or the proportional inverse loop gain test. Obviously, its results are less conservative than the individual tests. Feedback systems are robustly stable for perturbations in the frequency regions where either the sensitivity is small (at low frequencies) or the complementary sensitivity is small (at high frequencies). In the *crossover region* neither sensitivity is small. Hence, the feedback system is not robust for perturbations that strongly affect the crossover region.

In the crossover region the uncertainty therefore should be limited. On the one hand this limitation restricts *structured* uncertainty — caused by load variations and environmental changes — that the system can handle. On the other hand, *unstructured* uncertainty — deriving from neglected dynamics and parasitic effects — should be kept within bounds by adequate modeling.

5.5.7. Plant perturbation models

In the previous subsections we modeled the uncertainty as an uncertainty in the loop gain L , which results in interconnection matrices in terms of this loop gain, such as S and T . It is important to realize that we have a choice in how to model the uncertainty and that we need not necessarily do that in terms of the loop gain. In particular as the uncertainty is usually present in the plant P only, and not in controller K , it makes sense to model the uncertainty as such. Table 5.1 lists several ways to model plant uncertainty.

plant	perturbed plant	interconnection matrix H	perturbation Δ
P	$P + V\Delta_P W$	$-WKS V$	Δ_P
P	$(I + V\Delta_P W)P$	$-WTV$	Δ_P
P	$P(I + V\Delta_P W)$	$-WKSPV$	Δ_P
P	$(I + V\Delta_P W)^{-1}P$	$-WSV$	Δ_P
P	$P(I + V\Delta_P W)^{-1}$	$-W(I + KP)^{-1}V$	Δ_P
$D^{-1}N$	$(D + V\Delta_D W_1)^{-1}(N + V\Delta_N W_2)$	$-[\begin{smallmatrix} W_1 S \\ W_2 K S \end{smallmatrix}]D^{-1}V$	$[\begin{smallmatrix} \Delta_D & \Delta_N \end{smallmatrix}]$
$\bar{N}\bar{D}^{-1}$	$(\bar{N} + W_1\Delta_N V)(\bar{D} + W_2\Delta_D V)^{-1}$	$-V\bar{D}^{-1}[\begin{smallmatrix} KS W_1 & (I + KP)^{-1}W_2 \end{smallmatrix}]$	$[\begin{smallmatrix} \Delta_N \\ \Delta_D \end{smallmatrix}]$

Table 5.1: A list of perturbation models with plant P and controller K

5.6. Structured singular value robustness analysis

We return to the basic perturbation model of Fig. 5.8, which we repeat in Fig. 5.27(a). As demonstrated in § 5.3, the model is very flexible in representing both structured perturbations (i.e., variations of well-defined parameters) and unstructured perturbations.

The stability robustness theorem of Summary 5.4.4 (p. 200) guarantees robustness under all perturbations Δ such that $\bar{\sigma}(\Delta(j\omega)) \cdot \bar{\sigma}(H(j\omega)) < 1$ for all $\omega \in \mathbb{R}$. Perturbations $\Delta(j\omega)$ whose norm $\bar{\sigma}(\Delta(j\omega))$ does not exceed the number $1/\bar{\sigma}(H(j\omega))$, however, are completely *unstructured*. As a result, often quite conservative estimates are obtained of the stability region for *structured* perturbations. Several examples that we considered in the preceding sections confirm this.

Doyle (1982) proposed another measure for stability robustness based on the notion of what he calls *structured singular value*. In this approach, the perturbation structure is detailed as in

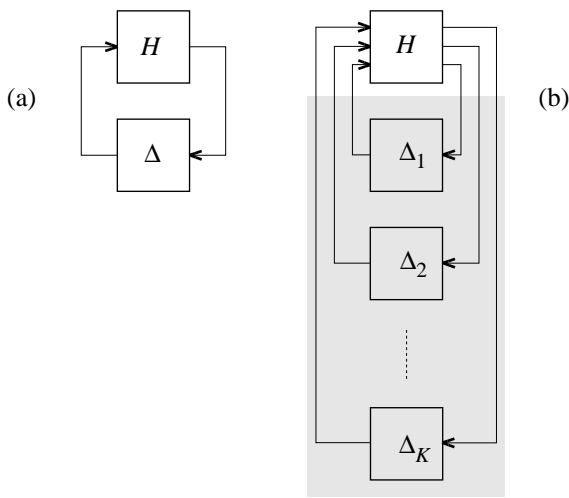


Figure 5.27: (a) Basic perturbation model. (b) Structured perturbation model

Fig. 5.27(b) (Safonov and Athans, 1981). The overall perturbation Δ has the block diagonal form

$$\Delta = \begin{bmatrix} \Delta_1 & 0 & \cdots & 0 \\ 0 & \Delta_2 & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & \Delta_K \end{bmatrix}. \quad (5.140)$$

Each diagonal block entry Δ_i has fixed dimensions, and has one of the following forms:

- $\Delta_i = \delta I$, with δ a real number. If the unit matrix I has dimension 1 this represents a real parameter variation. Otherwise, this is a *repeated real scalar perturbation*.
- $\Delta_i = \delta I$, with δ a stable⁸ scalar transfer matrix. This represents a *scalar or repeated scalar dynamic perturbation*.
- Δ_i is a (not necessarily square) stable transfer matrix. This represents a *multivariable dynamic perturbation*.

A wide variety of perturbations may be modeled this way.

We study which are the largest perturbations Δ with the given structure that do not destabilize the system of Fig. 5.27(a). Suppose that a given perturbation Δ destabilizes the system. Then by the generalized Nyquist criterion of Summary 1.3.13 the Nyquist plot of $\det(I - H\Delta)$ encircles the origin at least once, as illustrated in Fig. 5.28. Consider the Nyquist plot of $\det(I - \varepsilon H\Delta)$, with ε a real number that varies between 0 and 1. For $\varepsilon = 1$ the modified Nyquist plot coincides with that of Fig. 5.28, while for $\varepsilon = 0$ the plot reduces to the point 1. Since obviously the plot depends continuously on ε , there must exist a value of ε in the interval $(0, 1]$ such that the Nyquist plot of $\det(I - \varepsilon H\Delta)$ passes through the origin. Hence, if Δ destabilizes the perturbed system, there exist $\varepsilon \in (0, 1]$ and $\omega \in \mathbb{R}$ such that $\det(I - \varepsilon H(j\omega)\Delta(j\omega)) = 0$. Therefore, Δ does *not* destabilize the perturbed system if and only if there do not exist $\varepsilon \in (0, 1]$ and $\omega \in \mathbb{R}$ such that $\det(I - \varepsilon H(j\omega)\Delta(j\omega)) = 0$.

⁸A transfer function or matrix is “stable” if all its poles are in the open left-half plane.

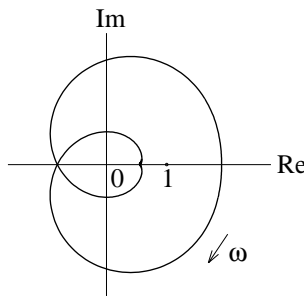


Figure 5.28: Nyquist plot of $\det(I - H\Delta)$ for a destabilizing perturbation

Fix ω , and let $\kappa(H(j\omega))$ be the *largest* real number such that $\det(I - H(j\omega)\Delta(j\omega)) \neq 0$ for all $\Delta(j\omega)$ (with the prescribed structure) such that $\bar{\sigma}(\Delta(j\omega)) < \kappa(H(j\omega))$. Then, obviously, if for a given perturbation $\Delta(j\omega)$, $\omega \in \mathbb{R}$,

$$\bar{\sigma}(\Delta(j\omega)) < \kappa(H(j\omega)) \quad \text{for all } \omega \in \mathbb{R} \quad (5.141)$$

then $\det(I - \varepsilon H(j\omega)\Delta(j\omega)) \neq 0$ for $\varepsilon \in (0, 1]$ and $\omega \in \mathbb{R}$. Therefore, the Nyquist plot of $\det(I - H\Delta)$ does not encircle the origin, and, hence, Δ does not destabilize the system. Conversely, it is always possible to find a perturbation Δ that violates (5.141) within an arbitrarily small margin that destabilizes the perturbed system. The number $\kappa(H(j\omega))$ is known as the *multivariable robustness margin* of the system H at the frequency ω (Safonov and Athans, 1981).

We note that for fixed ω

$$\begin{aligned} \kappa(H(j\omega)) &= \sup\{\gamma \mid \bar{\sigma}(\Delta(j\omega)) < \gamma \Rightarrow \det(I - H(j\omega)\Delta(j\omega)) \neq 0\} \\ &= \inf\{\bar{\sigma}(\Delta(j\omega)) \mid \det(I - H(j\omega)\Delta(j\omega)) = 0\}. \end{aligned} \quad (5.142)$$

If no structure is imposed on the perturbation Δ then $\kappa(H(j\omega)) = 1/\bar{\sigma}(H(j\omega))$. This led Doyle to terming the number $\mu(H(j\omega)) = 1/\kappa(H(j\omega))$ the *structured singular value* of the complex-valued matrix $H(j\omega)$.

Besides being determined by $H(j\omega)$ the structured singular value is of course also dependent on the perturbation structure. Given the perturbation structure of the perturbations Δ of the structured perturbation model, define by \mathcal{D} the class of *constant* complex-valued matrices of the form

$$\Delta = \begin{bmatrix} \Delta_1 & 0 & \cdots & 0 \\ 0 & \Delta_2 & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & \Delta_K \end{bmatrix}. \quad (5.143)$$

The diagonal block Δ_i has the same dimensions as the corresponding block of the dynamic perturbation, and has the following form:

- $\Delta_i = \delta I$, with δ a real number, if the dynamic perturbation is a scalar or repeated real scalar perturbation.
- $\Delta_i = \delta I$, with δ a complex number, if the dynamic perturbation is a scalar dynamic or a repeated scalar dynamic perturbation.
- Δ_i is a complex-valued matrix if the dynamic perturbation is a multivariable dynamic perturbation.

5.6.1. The structured singular value

We are now ready to define the *structured singular value* of a complex-valued matrix M :

Definition 5.6.1 (Structured singular value). Let M be an $n \times m$ complex-valued matrix, and \mathcal{D} a set of $m \times n$ structured uncertainty matrices. Then the *structured singular value* of M given the set \mathcal{D} is the number $\mu(M)$ defined by

$$\frac{1}{\mu(M)} = \inf_{\Delta \in \mathcal{D}, \det(I - M\Delta) = 0} \bar{\sigma}(\Delta). \quad (5.144)$$

If $\det(I - M\Delta) \neq 0$ for all $\Delta \in \mathcal{D}$ then $\mu(M) = 0$. □

The structured singular value $\mu(M)$ is the inverse of the norm of the smallest perturbation Δ (within the given class \mathcal{D}) that makes $I - M\Delta$ singular. Thus, the larger $\mu(M)$, the smaller the perturbation Δ is that is needed to make $I - M\Delta$ singular.

Example 5.6.2 (Structured singular value). By way of example we consider the computation of the structured singular value of a 2×2 complex-valued matrix

$$M = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix}, \quad (5.145)$$

under the real perturbation structure

$$\Delta = \begin{bmatrix} \Delta_1 & 0 \\ 0 & \Delta_2 \end{bmatrix}, \quad (5.146)$$

with $\Delta_1 \in \mathbb{R}$ and $\Delta_2 \in \mathbb{R}$. It is easily found that

$$\bar{\sigma}(\Delta) = \max(|\Delta_1|, |\Delta_2|), \quad \det(I + M\Delta) = 1 + m_{11}\Delta_1 + m_{22}\Delta_2 + m\Delta_1\Delta_2, \quad (5.147)$$

with $m = \det(M) = m_{11}m_{22} - m_{12}m_{21}$. Hence, the structured singular value of M is the inverse of

$$\kappa(M) = \inf_{\{\Delta_1 \in \mathbb{R}, \Delta_2 \in \mathbb{R}: 1 + m_{11}\Delta_1 + m_{22}\Delta_2 + m\Delta_1\Delta_2 = 0\}} \max(|\Delta_1|, |\Delta_2|). \quad (5.148)$$

Elimination of, say, Δ_2 results in the equivalent expression

$$\kappa(M) = \inf_{\Delta_1 \in \mathbb{R}} \max(|\Delta_1|, \left| \frac{1 + m_{11}\Delta_1}{m_{22} + m\Delta_1} \right|). \quad (5.149)$$

Suppose that M is numerically given by

$$M = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}. \quad (5.150)$$

Then it may be found (the details are left to the reader) that

$$\kappa(M) = \frac{-5 + \sqrt{33}}{4}, \quad (5.151)$$

so that

$$\mu(M) = \frac{4}{-5 + \sqrt{33}} = \frac{5 + \sqrt{33}}{2} = 5.3723. \quad (5.152)$$

□

5.6.2. Structured singular value and robustness

We discuss the structured singular value in more detail in § 5.6.3, but first summarize its application to robustness analysis.

Summary 5.6.3 (Structured robust stability). Given the stable unperturbed system H the perturbed system of Fig. 5.27(b) is stable for all perturbations Δ such that $\Delta(j\omega) \in \mathcal{D}$ for all $\omega \in \mathbb{R}$ if

$$\bar{\sigma}(\Delta(j\omega)) < \frac{1}{\mu(H(j\omega))} \quad \text{for all } \omega \in \mathbb{R}, \quad (5.153)$$

with μ the structured singular value with respect to the perturbation structure \mathcal{D} . If robust stability is required with respect to *all* perturbations within the perturbation class that satisfy the bound (5.153) then the condition is besides sufficient also necessary. \square

Given a structured perturbation Δ such that $\Delta(j\omega) \in \mathcal{D}$ for every $\omega \in \mathbb{R}$, we have

$$\|\Delta\|_\infty = \sup_{\omega \in \mathbb{R}} \bar{\sigma}(\Delta(j\omega)). \quad (5.154)$$

Suppose that the perturbations are *scaled*, so that $\|\Delta\|_\infty \leq 1$. Then Summary 5.6.3 implies that the perturbed system is stable if $\mu_H < 1$, where

$$\mu_H = \sup_{\omega \in \mathbb{R}} \mu(H(j\omega)). \quad (5.155)$$

With some abuse of terminology, we call μ_H the *structured singular value* of H . Even more can be said:

Theorem 5.6.4 (Structured robust stability for scaled perturbations). The perturbed system of Fig. 5.27(b) is stable for all stable perturbations Δ such that $\Delta(j\omega) \in \mathcal{D}$ for all $\omega \in \mathbb{R}$ and $\|\Delta\|_\infty \leq 1$ if and only if $\mu_H < 1$. \square

Clearly, if the perturbations are scaled to a maximum norm of 1 then a necessary and sufficient condition for robust stability is that $\mu_H < 1$.

5.6.3. Properties of the structured singular value

Before discussing the numerical computation of the structured singular value we list some of the principal properties of the structured singular value (Doyle, 1982; Packard and Doyle, 1993).

Summary 5.6.5 (Principal properties of the structured singular value). The structured singular value $\mu(M)$ of a matrix M under a structured perturbation set \mathcal{D} has the following properties:

1. *Scaling property:*

$$\mu(\alpha M) = |\alpha| \mu(M) \quad (5.156)$$

for every $\alpha \in \mathbb{R}$. If none of the perturbations are real then this holds as well for every complex α .

2. *Upper and lower bounds:* Suppose that M is square. Then

$$\rho_{\mathbb{R}}(M) \leq \mu(M) \leq \bar{\sigma}(M), \quad (5.157)$$

with $\rho_{\mathbb{R}}$ defined as in Exercise 5.10.

If none of the perturbations is real then

$$\rho(M) \leq \mu(M) \leq \bar{\sigma}(M), \quad (5.158)$$

with ρ denoting the spectral radius. The upper bounds also apply if M is not square.

3. *Preservation of the structured singular value under diagonal scaling:* Suppose that the i th diagonal block of the perturbation structure has dimensions $m_i \times n_i$. Form two block diagonal matrices D and \bar{D} whose i th diagonal blocks are given by

$$D_i = d_i I_{n_i}, \quad \bar{D}_i = d_i I_{m_i}, \quad (5.159)$$

respectively, with d_i a positive real number. I_k denotes a $k \times k$ unit matrix. Then

$$\mu(M) = \mu(DM\bar{D}^{-1}). \quad (5.160)$$

4. *Preservation of the structured singular value under unitary transformation:* Suppose that the i th diagonal block of the perturbation structure has dimensions $m_i \times n_i$. Form the block diagonal matrices Q and \bar{Q} whose i th diagonal blocks are given by the $m_i \times m_i$ unitary matrix Q_i and the $n_i \times n_i$ unitary matrix \bar{Q}_i , respectively. Then

$$\mu(M) = \mu(QM) = \mu(M\bar{Q}). \quad (5.161)$$

□

The scaling property is obvious. The upper bound in 2 follows by considering unrestricted perturbations of the form $\Delta \in \mathbb{C}^{m \times n}$ (that is, Δ is a full $m \times n$ complex matrix). The lower bound in 2 is obtained by considering restricted perturbations of the form $\Delta = \delta I$, with δ a real or complex number. The properties 3 and 4 are easily checked.

The following formula for the structured singular value of a 2×2 dyadic matrix has useful applications.

Summary 5.6.6 (Structured singular value of a dyadic matrix). The structured singular value of the 2×2 dyadic matrix

$$M = \begin{bmatrix} a_1 b_1 & a_1 b_2 \\ a_2 b_1 & a_2 b_2 \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \begin{bmatrix} b_1 & b_2 \end{bmatrix}, \quad (5.162)$$

with a_1, a_2, b_1 , and b_2 complex numbers, with respect to the perturbation structure

$$\Delta = \begin{bmatrix} \Delta_1 & 0 \\ 0 & \Delta_2 \end{bmatrix}, \quad \Delta_1 \in \mathbb{C}, \quad \Delta_2 \in \mathbb{C}, \quad (5.163)$$

is

$$\mu(M) = |a_1 b_1| + |a_2 b_2|. \quad (5.164)$$

□

The proof is given in § 5.8.

5.6.4. Numerical approximation of the structured singular value

Exact calculation of the structured singular value is often not possible practically and in any case computationally intensive. The numerical methods that are presently available for approximating the structured singular value are based on calculating upper and lower bounds for the structured singular value.

Summary 5.6.7 (Upper and lower bounds for the structured singular value).

1. *D-scaling upper bound.* Let the diagonal matrices D and \bar{D} be chosen as in (3) of Summary 5.6.5. Then with property (b) we have

$$\mu(M) = \mu(DM\bar{D}^{-1}) \leq \bar{\sigma}(DM\bar{D}^{-1}). \quad (5.165)$$

The rightmost side may be numerically minimized with respect to the free positive numbers d_i that determine D and \bar{D} .

Suppose that the perturbation structure consists of m repeated scalar dynamic perturbation blocks and M full multivariable perturbation blocks. Then if $2m + M \leq 3$ the minimized upper bound actually *equals* the structured singular value⁹.

2. *Lower bound.* Let Q be a unitary matrix as constructed in (4.) of Summary 5.6.5. With property (2.) we have (for complex perturbations only)

$$\mu(M) = \mu(MQ) \geq \rho(MQ). \quad (5.166)$$

Actually, (Doyle, 1982),

$$\mu(M) = \max_Q \rho(MQ), \quad (5.167)$$

with Q varying over the set of matrices as constructed in (4.) of Summary 5.6.5.

□

Practical algorithms for computing lower and upper bounds on the structured singular value for complex perturbations have been implemented in the MATLAB *μ -Analysis and Synthesis Toolbox* (Balas et al., 1991). The closeness of the bounds is a measure of the reliability of the calculation.

The MATLAB *Robust Control Toolbox* (Chiang and Safonov, 1992) provides routines for calculating upper bounds on the structured singular value for both complex and real perturbations.

5.6.5. Example

We apply the singular value method for the analysis of the stability robustness to an example.

Example 5.6.8 (SISO system with two parameters). By way of example, consider the SISO feedback system that was studied in Example 5.2.1 and several other places. A plant with transfer function

$$P(s) = \frac{g}{s^2(1 + s\theta)} \quad (5.168)$$

is connected in feedback with a compensator with transfer function

$$C(s) = \frac{k + sT_d}{1 + sT_0}. \quad (5.169)$$

⁹If there are *real* parametric perturbations then this result remains valid, provided that we use a generalization of D -scaling called (D, G) -scaling. This is a generalization that allows to exploit realness of perturbations.

We use the design values of the parameters established in Example 5.2.2. In Example 5.3.5 it was found that the interconnection matrix for scaled perturbations in the gain g and time constant θ is given by

$$H(s) = \frac{\frac{1}{1+s\theta_0}}{1+L_0(s)} \begin{bmatrix} \beta_1 C(s) \\ -\beta_2 \end{bmatrix} \begin{bmatrix} -\frac{\alpha_1}{s^2} & \alpha_2 s \end{bmatrix}, \quad (5.170)$$

where the subscript o indicates nominal values, and $L_0 = P_0 C$ is the nominal loop gain. The numbers α_1 , α_2 , β_1 , and β_2 are scaling factors such that $|g - g_0| \leq \varepsilon_1$ with $\varepsilon_1 = \alpha_1 \beta_1$, and $|\theta - \theta_0| \leq \varepsilon_2$ with $\varepsilon_2 = \alpha_2 \beta_2$. The interconnection matrix H has a dyadic structure. Application of the result of Summary 5.6.6 shows that its structured singular value with respect to *complex* perturbations in the parameters is

$$\begin{aligned} \mu(H(j\omega)) &= \frac{\sqrt{\frac{1}{1+\omega^2\theta_0^2}}}{|1+L_0(\omega)|} \left(\alpha_1 \beta_1 \frac{|C(j\omega)|}{\omega^2} + \alpha_2 \beta_2 \omega \right) \\ &= \frac{\varepsilon_1 \sqrt{k^2 + \omega^2 T_d^2} + \varepsilon_2 \omega^3 \sqrt{1 + \omega^2 T_0^2}}{|\chi_0(j\omega)|}, \end{aligned} \quad (5.171)$$

with $\chi_0(s) = \theta_0 T_0 s^4 + (\theta_0 + T_0) s^3 + s^2 + g_0 T_d s + g_0 k$ the nominal closed-loop characteristic polynomial.

Inspection shows that the right-hand side of this expression is identical to that of (5.89) in Example 5.4.7, which was obtained by a singular value analysis based on optimal scaling. In view of the statement in Summary 5.6.7(1) this is no coincidence.

Figure 5.29 repeats the plots of Fig. 5.16 of the structured singular value for the two cases of Example 5.4.7. For case (a) the peak value of the structured singular value is greater than 1 so that robust stability is not guaranteed. Only in case (b) robust stability is certain. Since the perturbation analysis is based on dynamic rather than parameter perturbations the results are conservative. □

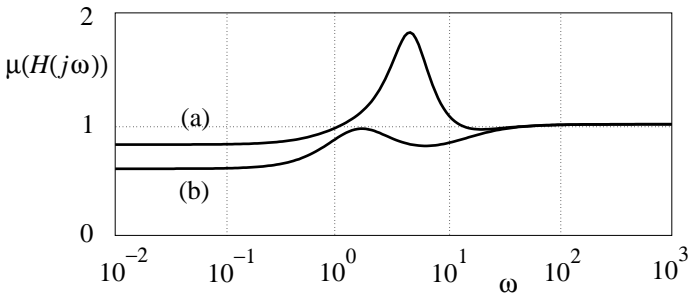


Figure 5.29: Structured singular values for the two-parameter plant for two cases

5.7. Combined performance and stability robustness

In the preceding section we introduced the structured singular value to study the stability robustness of the basic perturbation model. We continue in this section by considering *performance*

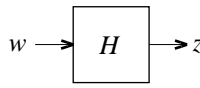


Figure 5.30: Control system

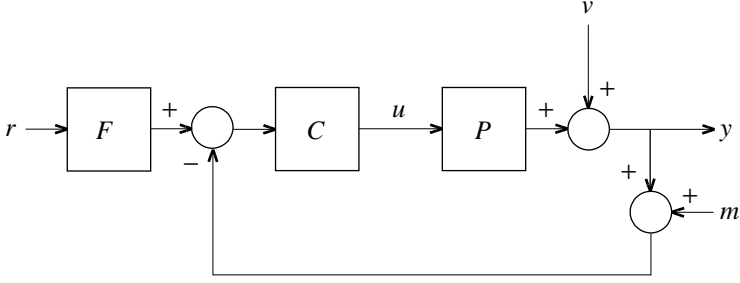


Figure 5.31: Two-degree-of-freedom feedback control system.

and its robustness. Figure 5.30 represents a control system with *external* input w , such as disturbances, measurement noise, and reference signals, and *external* output z . The output signal z represents an *error signal*, and ideally should be zero. The transfer matrix H is the transfer matrix from the external input w to the error signal z .

Example 5.7.1 (Two-degree-of-freedom feedback system). To illustrate this model, consider the two-degree-of-freedom feedback configuration of Fig. 5.31. P is the plant, C the compensator, and F a precompensator. The signal r is the reference signal, v the disturbance, m the measurement noise, u the plant input, and y the control system output. It is easy to find that

$$\begin{aligned} y &= (I + PC)^{-1} PCFr + (I + PC)^{-1} v - (I + PC)^{-1} PCm \\ &= TFr + Sv - Tm, \end{aligned} \quad (5.172)$$

where S is the sensitivity matrix of the feedback loop and T the complementary sensitivity matrix. The tracking error $z = y - r$ is given by

$$z = y - r = (TF - I)r + Sv - Tm. \quad (5.173)$$

Considering the combined signal

$$w = \begin{bmatrix} r \\ v \\ m \end{bmatrix} \quad (5.174)$$

as the external input, it follows that the transfer matrix of the control system is

$$H = \begin{bmatrix} TF - I & S & -T \end{bmatrix}. \quad (5.175)$$

□

The performance of the control system of Fig. 5.30 is ideal if $H = 0$, or, equivalently, $\|H\|_\infty = 0$. Ideals cannot always be obtained, so we settle for the norm $\|H\|_\infty$ to be “small,” rather than zero. By introducing suitable frequency dependent scaling functions (that is, by modifying H to WHV) we may arrange that “satisfactory” performance is obtained if and only if

$$\|H\|_\infty < 1. \quad (5.176)$$

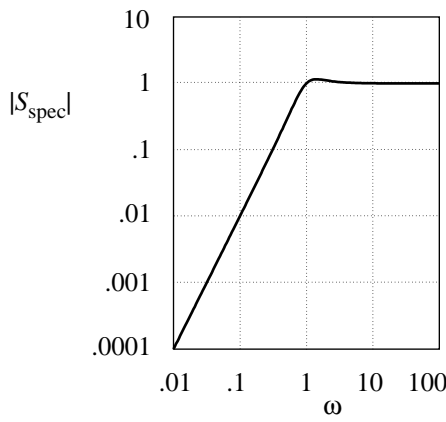


Figure 5.32: Magnitude plot of S_{spec} .

Example 5.7.2 (Performance specification). To demonstrate how performance may be specified this way, again consider the two-degree-of-freedom feedback system of Fig. 5.31. For simplicity we assume that it is a SISO system, and that the reference signal and measurement noise are absent. It follows from Example 5.7.1 that the control system transfer function reduces to the sensitivity function S of the feedback loop:

$$H = \frac{1}{1 + PC} = S. \quad (5.177)$$

Performance is deemed to be adequate if the sensitivity function satisfies the requirement

$$|S(j\omega)| < |S_{\text{spec}}(j\omega)|, \quad \omega \in \mathbb{R}, \quad (5.178)$$

with S_{spec} a specified rational function, such as

$$S_{\text{spec}}(s) = \frac{s^2}{s^2 + 2\zeta\omega_0 s + \omega_0^2}. \quad (5.179)$$

The parameter ω_0 determines the minimal bandwidth, while the relative damping coefficient ζ specifies the allowable amount of peaking. A doubly logarithmic magnitude plot of S_{spec} is shown in Fig. 5.32 for $\omega_0 = 1$ and $\zeta = \frac{1}{2}$.

The inequality (5.178) is equivalent to $|S(j\omega)V(j\omega)| < 1$ for $\omega \in \mathbb{R}$, with $V = 1/S_{\text{spec}}$. Redefining H as $H = SV$ this reduces to

$$\|H\|_{\infty} < 1. \quad (5.180)$$

□

By suitable scaling we thus may arrange that the performance of the system of Fig. 5.30 is considered satisfactory if

$$\|H\|_{\infty} < 1. \quad (5.181)$$

In what follows we exploit the observation that by (1) of Summary 5.4.4 this specification is equivalent to the requirement that the artificially perturbed system of Fig. 5.33 remains stable under all stable perturbations Δ_0 such that $\|\Delta_0\|_{\infty} \leq 1$.

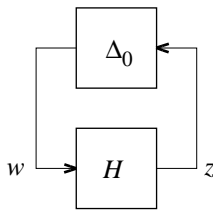


Figure 5.33: Artificially perturbed system

5.7.1. Performance robustness

Performance is said to be *robust* if $\|H\|_\infty$ remains less than 1 under perturbations. To make this statement more specific we consider the configuration of Fig. 5.34(a). The perturbation Δ may be structured, and is scaled such that $\|\Delta\|_\infty \leq 1$. We describe the system by the interconnection equations

$$\begin{bmatrix} z \\ q \end{bmatrix} = H \begin{bmatrix} w \\ p \end{bmatrix} = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} \begin{bmatrix} w \\ p \end{bmatrix}. \quad (5.182)$$

H_{11} is the nominal control system transfer matrix. We define performance to be robust if

1. the perturbed system remains stable under all perturbations, and
2. the ∞ -norm of the transfer matrix of the perturbed system remains less than 1 under all perturbations.

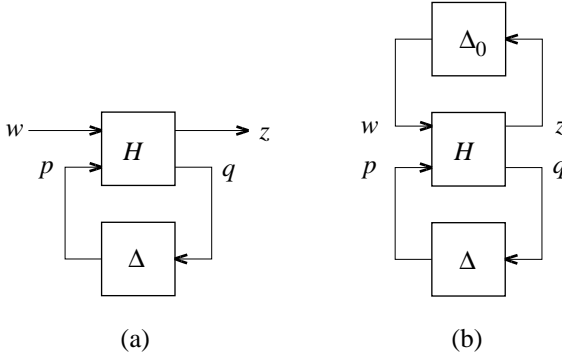


Figure 5.34: (a) Perturbed control system. (b) Doubly perturbed system

Necessary and sufficient for robust performance is that the norm of the perturbed transfer matrix from w to z in the perturbation model of Fig. 5.34(a) is less than 1 for every perturbation Δ with norm less than or equal to 1. This, in turn, is equivalent to the condition that the augmented perturbation model of Fig. 5.34(b) is stable for every “full” perturbation Δ_0 and every structured perturbation Δ , both with norm less than or equal to 1. Necessary and sufficient for this is that

$$\mu_H < 1, \quad (5.183)$$

with μ the structured singular value of H with respect to the perturbation structure defined by

$$\begin{bmatrix} \Delta_0 & 0 \\ 0 & \Delta \end{bmatrix}. \quad (5.184)$$

Summary 5.7.3 (Robust performance and stability). Robust performance of the system of Fig. 5.34(a) is achieved if and only if

$$\mu_H < 1, \quad (5.185)$$

where μ_H is the structured singular value of H under perturbations of the form

$$\begin{bmatrix} \Delta_0 & 0 \\ 0 & \Delta \end{bmatrix}. \quad (5.186)$$

Δ_0 is a “full” perturbation, and Δ structured as specified. □

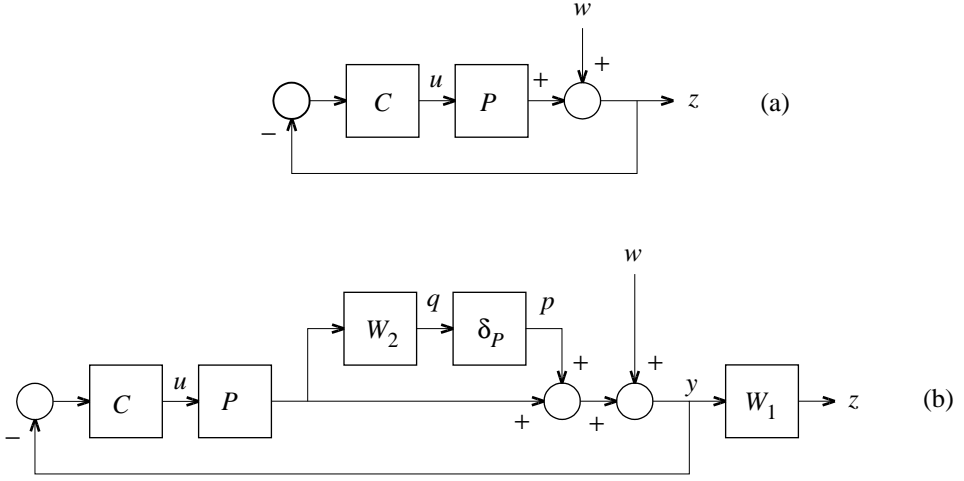


Figure 5.35: SISO feedback system. (a) Nominal. (b) Perturbed

5.7.2. SISO design for robust stability and performance

We describe an elementary application of robust performance analysis using the structured singular value (compare (Doyle et al., 1992)). Consider the feedback control system configuration of Fig. 5.35(a). A SISO plant P is connected in feedback with a compensator C .

- Performance is measured by the closed-loop transfer function from the disturbance w to the control system output z , that is, by the sensitivity function $S = 1/(1 + PC)$. Performance is considered satisfactory if

$$|S(j\omega)W_1(j\omega)| < 1, \quad \omega \in \mathbb{R}, \quad (5.187)$$

with W_1 a suitable weighting function.

- Plant uncertainty is modeled by the scaled uncertainty model

$$P \longrightarrow P(1 + \delta_P W_2), \quad (5.188)$$

with W_2 a stable function representing the maximal uncertainty, and δ_P a scaled stable perturbation such that $\|\delta_P\|_\infty \leq 1$.

The block diagram of Fig. 5.35(b) includes the weighting filter W_1 and the plant perturbation model. By inspection we obtain the signal balance equation $y = w + p - PCy$, so that

$$y = \frac{1}{1 + PC}(w + p). \quad (5.189)$$

By further inspection of the block diagram it follows that

$$z = W_1 y = W_1 S(w + p), \quad (5.190)$$

$$q = -W_2 PCy = -W_2 T(w + p). \quad (5.191)$$

Here

$$S = \frac{1}{1 + PC}, \quad T = \frac{PC}{1 + PC} \quad (5.192)$$

are the sensitivity function and the complementary sensitivity function of the feedback system, respectively. Thus, the transfer matrix H in the configuration of Fig. 5.34(b) follows from

$$\begin{bmatrix} q \\ z \end{bmatrix} = \underbrace{\begin{bmatrix} -W_2 T & -W_2 T \\ W_1 S & W_1 S \end{bmatrix}}_H \begin{bmatrix} p \\ w \end{bmatrix}. \quad (5.193)$$

H has the dyadic structure

$$H = \begin{bmatrix} W_2 T \\ W_1 S \end{bmatrix} \begin{bmatrix} -1 & 1 \end{bmatrix}. \quad (5.194)$$

With the result of Summary 5.6.6 we obtain the structured singular value of the interconnection matrix H as

$$\mu_H = \sup_{\omega \in \mathbb{R}} \mu(H(j\omega)) = \sup_{\omega \in \mathbb{R}} (|W_1(j\omega)S(j\omega)| + |W_2(j\omega)T(j\omega)|) \quad (5.195)$$

$$= \| |W_1 S| + |W_2 T| \|_{\infty}. \quad (5.196)$$

By Summary 5.7.3, robust performance and stability are achieved if and only if $\mu_H < 1$.

Example 5.7.4 (Robust performance of SISO feedback system). We consider the SISO feedback system we studied in Example 5.2.1 (p. 184) and on several other occasions, with nominal plant and compensator transfer functions

$$P_0(s) = \frac{g_0}{s^2}, \quad C(s) = \frac{k + sT_d}{1 + sT_0}, \quad (5.197)$$

respectively. We use the design values of Example 5.2.2 (p. 185). In Fig. 5.36 magnitude plots are given of the nominal sensitivity and complementary sensitivity functions of this feedback system. The nominal sensitivity function is completely acceptable, so we impose as design specification that under perturbation

$$\left| \frac{S(j\omega)}{S_0(j\omega)} \right| \leq 1 + \varepsilon, \quad \omega \in \mathbb{R}, \quad (5.198)$$

with S_0 the nominal sensitivity function and the positive number ε a tolerance. This comes down to choosing the weighting function W_1 in (5.187) as

$$W_1 = \frac{1}{(1 + \varepsilon)S_0}. \quad (5.199)$$

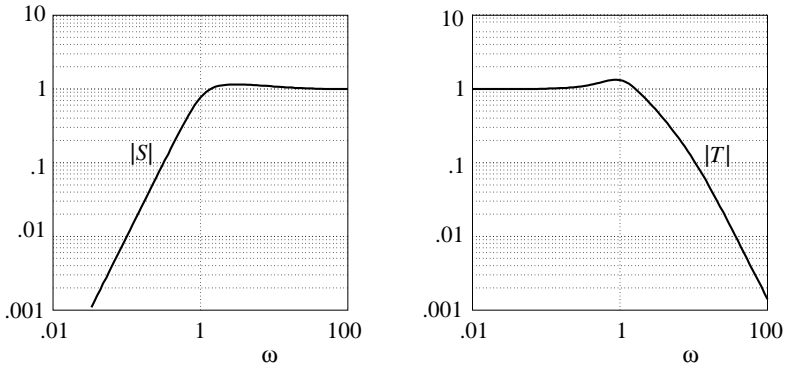
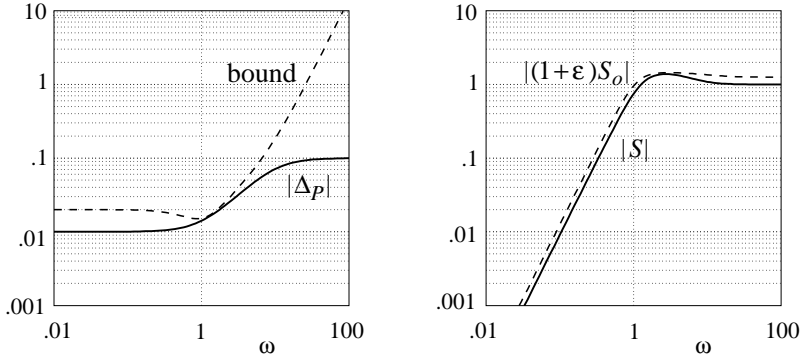


Figure 5.36: Nominal sensitivity and complementary sensitivity functions



(a) Uncertainty bound and admissible perturbation

(b) Effect of perturbation on S

Figure 5.37: The bound and its effect on S

We consider how to select the weighting function W_2 , which specifies the allowable perturbations. With W_1 chosen as in (5.199) the robust performance criterion $\mu_H < 1$ according to (5.195) reduces to

$$\frac{1}{1+\varepsilon} + |W_2(j\omega)T_0(j\omega)| < 1, \quad \omega \in \mathbb{R}, \quad (5.200)$$

with T_0 the nominal complementary sensitivity function. This is equivalent to

$$|W_2(j\omega)| < \frac{\frac{\varepsilon}{1+\varepsilon}}{|T_0(j\omega)|}, \quad \omega \in \mathbb{R}. \quad (5.201)$$

Figure 5.37(a) shows a plot of the right-hand side of (5.201) for $\varepsilon = 0.25$. This right-hand side is a frequency dependent bound for the maximally allowable scaled perturbation δ_P . The plot shows that for low frequencies the admissible proportional uncertainty is limited to $\varepsilon/(1+\varepsilon) = 0.2$, but that the system is much more robust with respect to high-frequency perturbations. In the crossover frequency region the allowable perturbation is even less than 0.2.

Suppose, as we did before, that the plant is perturbed to

$$P(s) = \frac{g}{s^2(1 + s\theta)}. \quad (5.202)$$

The corresponding proportional plant perturbation is

$$\Delta_P(s) = \frac{P(s) - P_0(s)}{P_0(s)} = \frac{\frac{g-g_0}{g_0} - s\theta}{1 + s\theta}. \quad (5.203)$$

At low frequencies the perturbation has approximate magnitude $|g - g_0|/g_0$. For performance robustness this number definitely needs to be less than the minimum of the bound on the right-hand side of (5.201), which is about $\frac{\varepsilon}{1+\varepsilon} 0.75$.

If $\varepsilon = 0.25$ then the latter number is about 0.15. Figure 5.37(a) includes a magnitude plot of Δ_P for $|g - g_0|/g_0 = 0.1$ (that is, $g = 0.9$ or $g = 1.1$) and $\theta = 0.1$. For this perturbation the performance robustness test is just satisfied. If g is made smaller than 0.9 or larger than 1.1 or θ is increased beyond 0.1 then the perturbation fails the test.

Figure 5.37(b) shows the magnitude plot of the sensitivity function S for $g = 1.1$ and $\theta = 0.1$. Comparison with the bound $|(1 + \varepsilon)S_0|$ shows that performance is robust, as expected. \square

5.8. Appendix: Proofs

In this Appendix to Chapter 5 the proofs of certain results in are sketched.

Proof of Bialas' test. We split the proof into two parts. In the *first* part we show a technical result that says that the Hurwitz matrix H_p of a polynomial p is nonsingular if p is stable, and that H_p is singular if p has imaginary zeros. This we use in the second part to prove Bialas' test.

Part 1. We show that H_p is singular if and only if $p(s)$ and $p(-s)$ have zeros in common. In particular this implies what we need in the second part, namely that H_p is nonsingular for strictly Hurwitz p and that H_p is singular for p with imaginary zeros (as these zeros come in conjugate pairs $s = j\omega$ and $s = -j\omega$.)

First we establish that

$$p(s) \text{ and } p(-s) \text{ have zeros in common.} \\ \iff$$

There is a nonzero polynomial r of degree at most $\deg(p) - 1$ such that rp is odd.

This is readily checked: rp is odd means $r(s)p(s) = sh(s^2)$ so the zeros of rp are then symmetric with respect to the imaginary axis. If $p(s)$ and $p(-s)$ have no zeros in common then obviously such r (of degree at most $\deg(p) - 1$) do not exist. If, on the other hand, $p(s)$ and $p(-s)$ do have zeros in common, then a polynomial r is readily found: If $p(0) = 0$ then $r(s) := p(-s)/s$ will do. If $p(s)$ and $p(-s)$ have a common zero at $\alpha \neq 0$, then $r(s) := sp(-s)/(s^2 - \alpha^2)$ will do.

The condition that rp is odd is the same as saying that the even part $[rp]_{\text{even}} = 0$, and that is a linear equation in the coefficients of r . To see this more clearly we define t depending on r and p as $t := rp$ and we write $t = rp$ out in terms of its coefficients

$$\begin{bmatrix} r_0 & r_1 & \cdots & r_{n-1} \end{bmatrix} \begin{bmatrix} p_0 & p_1 & p_2 & p_3 & \cdots & \cdots & 0 \\ 0 & p_0 & p_1 & p_2 & \cdots & \cdots & 0 \\ 0 & 0 & p_0 & p_1 & \cdots & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & 0 & 0 & 0 & p_0 & \cdots & p_n \end{bmatrix} = \begin{bmatrix} t_0 & t_1 & \cdots & t_{2n-1} \end{bmatrix}. \quad (5.204)$$

Here the coefficients are indexed as in $r(s) = r_0s^{n-1} + r_1s^{n-2} + \dots$ and $t(s) = t_0s^{2n-1} + t_1s^{2n-2} + \dots$, and the coefficients of p are denoted by p_i . $[rp]_{\text{even}}$ is zero iff the odd-indexed coefficients t_1, t_3, t_5, \dots are all zero. That is, iff removal of the 1st, 3rd, 5th etcetera columns of (5.204) results in the zero vector,

$$\begin{bmatrix} r_0 & r_1 & \cdots & r_{n-1} \end{bmatrix} \begin{bmatrix} p_1 & p_3 & \cdots & \cdots \\ p_0 & p_2 & \cdots & \cdots \\ 0 & p_1 & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & \cdots & \cdots \end{bmatrix} = \begin{bmatrix} 0 & 0 & \cdots & 0 \end{bmatrix}.$$

The big matrix in the middle may be recognized as the $n \times n$ Hurwitz matrix H_p of p , and, therefore, there exist nonzero polynomials r of degree at most $\deg(p) - 1$ such that $[rp]_{\text{even}} = 0$ iff the Hurwitz matrix of p is singular. That is what we intended to prove.

Part 2. We examine the stability properties of the convex combinations

$$p_\lambda := \lambda p + (1 - \lambda)q$$

of two polynomials p and q . Now if p is stable but for some $\lambda \in [0, 1]$ the polynomial $\lambda p + (1 - \lambda)q$ is not stable, then by continuity of the zeros of polynomials, there exists a $\lambda_* \in [0, 1]$ for which $\lambda_* p + (1 - \lambda_*)q$ has imaginary zeros. By linearity the corresponding Hurwitz matrix $H_{p_{\lambda_*}}$ equals

$$H_{p_{\lambda_*}} = \lambda_* H_p + (1 - \lambda_*) H_q.$$

Note that H_p is nonsingular if p is stable. Therefore, in that case H_{p_λ} may also be written as

$$H_{p_\lambda} = \lambda H_p \left[1 + \frac{1 - \lambda}{\lambda} H_p^{-1} H_q \right].$$

The value of $(1 - \lambda)/\lambda$ ranges over all nonnegative values as λ ranges over all $\lambda \in [0, 1]$, and this shows that H_{p_λ} is nonsingular for all such λ iff all *real valued* eigenvalues of $H_p^{-1} H_q$ are strictly positive. ■

Proof of Lemma 5.3.1. Introducing internal inputs and outputs as in Fig. 5.8(b), we obtain the signal balance equations

$$w_1 = v_1 + \Delta_H(v_2 + H w_1), \quad w_2 = v_2 + H(v_1 + \Delta_H w_1). \quad (5.205)$$

Rearrangement yields

$$(I - \Delta_H H)w_1 = v_1 + \Delta_H v_2, \quad (I - H \Delta_H)w_2 = H v_1 + v_2. \quad (5.206)$$

For internal stability we need the transfer matrices from v_1, v_2 to w_1, w_2 to exist and be stable. These transfer matrices are

$$\begin{aligned} w_1 &= (I - \Delta_H H)^{-1} v_1 + (I - \Delta_H H)^{-1} \Delta_H v_2, \\ w_2 &= (I - H \Delta_H)^{-1} H v_1 + (I - H \Delta_H)^{-1} v_2. \end{aligned}$$

By assumption both H and Δ_H are stable. Hence the above transfer matrices exist and are stable precisely if $(I - H \Delta_H)^{-1}$ and $(I - \Delta_H H)^{-1}$ exist and are stable. Now $\det(I - H \Delta_H) = \det(I - \Delta_H H)$ so existence and stability of $(I - H \Delta_H)^{-1}$ is equivalent to that of $(I - \Delta_H H)^{-1}$. ■

5.8.1. Small gain theorem

We prove Summary 5.4.1.

Proof of Theorem 5.4.1 (sketch). If $\|L\| < 1$ then L is stable. Hence the closed loop is internally stable iff $(I - L)^{-1}$ exists and is stable. The claim is that

$$(I - L)^{-1} = I + L + L^2 + L^3 + \dots$$

and that the right-hand side converges because $\|L\| < 1$. Define

$$Q_N = \sum_{k=0}^N L^k$$

Then $(I - L)Q_N = \sum_{k=0}^N (I - L)L^k = (I - L) + (L - L^2) + \dots + (L^N - L^{N+1}) = I - L^{N+1}$. Now by the submultiplicative property of induced norms we have that $\|L^k\| = \|L \cdot L \cdots L\| \leq \|L\| \cdot \|L\| \cdots \|L\| = \|L\|^k$. This allows to show that Q_N is bounded (independent of N),

$$\|Q_N\| \leq \|I + L + \dots + L^N\| \leq \|I\| + \|L\| + \dots + \|L\|^N = \sum_{k=0}^N \|L\|^k = \frac{1 - \|L\|^{N+1}}{1 - \|L\|} \leq \frac{1}{1 - \|L\|}$$

and that

$$\lim_{N \rightarrow \infty} \|I - (I + L)Q_N\| = \lim_{N \rightarrow \infty} \|I - (I - L^{N+1})\| = \lim_{N \rightarrow \infty} \|L^{N+1}\| \leq \lim_{N \rightarrow \infty} \|L\|^{N+1} = 0$$

It is standard result in functional analysis that the set of bounded operators form a Banach space, which for our case means that $Q_\infty := \lim_{N \rightarrow \infty} Q_N$ exists and is a bounded operator from \mathcal{U} to \mathcal{U} . ■

Proof of Theorem 5.4.4. By the small gain theorem the BPM is guaranteed to be internally stable if $\|H\Delta_H\|_\infty < 1$. Both conditions 1 and 2 imply $\|H\Delta_H\|_\infty < 1$. Indeed, by the submultiplicative property Condition 1 implies

$$\|H\Delta_H\|_\infty = \sup_{\omega} \bar{\sigma}(H(j\omega)\Delta_H(j\omega)) \leq \sup_{\omega} \bar{\sigma}(H(j\omega))\bar{\sigma}(\Delta_H(j\omega)) < 1$$

and Condition 2 implies

$$\|H\Delta_H\|_\infty \leq \|H\|_\infty \|\Delta_H\|_\infty < 1.$$

Proof of Theorem 5.4.5. We only prove 1 and only for $\gamma = 1$ (the case 2 is similar, and if $\gamma \neq 1$ a simple scaling will make it equal to 1). We know that $\|\Delta_H\|_\infty \leq 1$ and $\|H\|_\infty < 1$ imply internal stability. To prove the converse, suppose that $\|H\|_\infty \geq 1$. Then there is an $\omega_* \in \mathbb{R} \cup \{\infty\}$ for which $\bar{\sigma}(H(j\omega_*)) \geq 1$. Now define Δ to be the constant uncertainty

$$\Delta = \frac{1}{\bar{\sigma}(H(j\omega_*))^2} H(j\omega_*)^*.$$

Then

$$\bar{\sigma}(\Delta) = \frac{1}{\bar{\sigma}(H(j\omega_*))} \leq 1$$

and it is destabilizing, because

$$\det(I - \Delta H(j\omega_*)) = \det(I - \frac{1}{\bar{\sigma}(H(j\omega_*))^2} (H(j\omega_*))^* H(j\omega_*)) = 0.$$

■

5.8.2. Structured singular value of a dyadic matrix

We prove Summary 5.6.6. In fact we prove the generalization that for rank 1 matrices

$$M = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \end{bmatrix} \begin{bmatrix} b_1 & b_2 & \dots \end{bmatrix}, \quad a_i, b_j \in \mathbb{C} \quad (5.207)$$

and perturbation structure

$$\Delta = \begin{bmatrix} \Delta_1 & 0 & \dots \\ 0 & \Delta_2 & \dots \\ \dots & \dots & \dots \end{bmatrix}, \quad \Delta_i \in \mathbb{C} \quad (5.208)$$

there holds that $\mu(M) = \sum_i |a_i b_i|$.

Proof of structured singular value of dyadic matrix. Given M as the product of a column and row vector $M = ab$ we have

$$\det(I - M\Delta) = \det(I - ab\Delta) = \det(1 - b\Delta a) = 1 - \sum_i a_i b_i \Delta_i. \quad (5.209)$$

This gives a lower bound for the real part of $\det(I - M\Delta)$,

$$\det(I - M\Delta) = 1 - \sum_i a_i b_i \Delta_i \geq 1 - \sum_i |a_i b_i| \max_i |\Delta_i| = 1 - \sum_i |a_i b_i| \bar{\sigma}(\Delta). \quad (5.210)$$

For $\det(I - M\Delta)$ to be zero we need at least that the lower bound is not positive, that is, that $\bar{\sigma}(\Delta) \geq 1 / \sum_i |a_i b_i|$. Now take Δ equal to

$$\Delta = \frac{1}{\sum_i |a_i b_i|} \begin{bmatrix} \text{sgn}(a_1 b_1) & 0 & \dots \\ 0 & \text{sgn}(a_2 b_2) & \dots \\ \dots & \dots & \ddots \end{bmatrix}. \quad (5.211)$$

Then $\bar{\sigma}(\Delta) = 1 / \sum_i |a_i b_i|$ and it is such that

$$\det(I - M\Delta) = 1 - \sum_j a_j b_j \Delta_j = 1 - \sum_j a_j b_j \frac{\text{sgn } a_j b_j}{\sum_i |a_i b_i|} = 0. \quad (5.212)$$

Hence a smallest Δ (in norm) that makes $I - M\Delta$ singular has norm $1 / \sum_i |a_i b_i|$. Therefore $\mu(M) = \sum_i |a_i b_i|$.

It may be shown that the D -scaling upper bound for rank 1 matrices equals $\mu(M)$ in this case. ■

5.9. Exercises

5.1 *Stability margins.* The stability of the feedback system of Example 5.2.4 is quite robust with respect to variations in the parameters θ and g . Inspect the Nyquist plot of the nominal loop gain to determine the various stability margins of § 1.4.2 and Exercise 1.4.9(b) of the closed-loop system.

5.2 *Application of Bialas' test.* Test the stability of the system on the remaining three edges as given by (5.46).

5.3 *Parametric uncertainty.* Consider the family of polynomials

$$\chi_\alpha(s) = (3 + \alpha) + (2 + \alpha)s + (4 + 3\alpha)s^2 + \frac{2}{5}s^3$$

with $\alpha \in \mathbb{R}$ an uncertain constant in the interval $[-1, 1]$.

- Is each member of this family of polynomials strictly Hurwitz? (You may want to use MATLAB.)
- What can be deduced from Kharitonov's theorem?
- What can be deduced from the edge theorem?

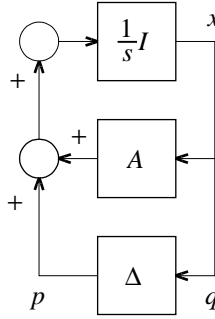


Figure 5.38: Uncertainty of the state space system

5.4 *Stability radius, (Hinrichsen and Pritchard, 1986).* Investigate what uncertainties of the matrix A make the system described by the state differential equation

$$\dot{x}(t) = Ax(t), \quad t \in \mathbb{R}, \quad (5.213)$$

unstable. Assume that the nominal system $\dot{x}(t) = Ax(t)$ is stable. The number

$$r(A) = \inf_{\substack{A + \Delta \text{ has at least one eigenvalue} \\ \text{in the closed right-half plane}}} \|\Delta\| \quad (5.214)$$

is called the *stability radius* of the matrix A . The matrix norm used is the spectral norm. If only real-valued uncertainties Δ are considered then $r(A)$ is the *real stability radius*, denoted $r_{\mathbb{R}}(A)$. If the elements of Δ may also assume complex values then $r(A)$ is the *complex stability radius*, denoted $r_{\mathbb{C}}(A)$.

- Prove that

$$r_{\mathbb{R}}(A) \geq r_{\mathbb{C}}(A) \geq \frac{1}{\|F\|_{\infty}}, \quad (5.215)$$

with F the rational matrix given by $F(s) = (sI - A)^{-1}$. *Hint:* Represent the system by the block diagram of Fig. 5.38.

b) A more structured uncertainty model results by assuming that A is perturbed as

$$A \longrightarrow A + B\Delta C, \quad (5.216)$$

with B and C matrices of suitable dimensions, and Δ the uncertainty. Adapt the block diagram of Fig. 5.38 and prove that the associated complex stability radius $r_{\mathbb{C}}(A, B, C)$ satisfies $r_{\mathbb{C}}(A, B, C) \geq 1/\|H\|_{\infty}$, with $H(s) = C(sI - A)^{-1}B$.

Hinrichsen and Pritchard (1986) prove that $r_{\mathbb{C}}(A) = 1/\|F\|_{\infty}$ and $r_{\mathbb{C}}(A, B, C) = 1/\|H\|_{\infty}$. Thus, explicit formulas are available for the complex stability radius. The real stability radius is more difficult to determine (Qiu et al., 1995).

5.5 Consider the unit feedback loop of Fig. 5.9 and suppose the loop gain L equals

$$L(s) = e^{-\varepsilon s} \frac{1}{s+1},$$

for some uncertain delay $\varepsilon \geq 0$. Nominally the loop gain has no delay ($\varepsilon = 0$).

- Model the uncertainty as an additive uncertainty and then use Inequality (5.70) to determine (numerically?) the largest delay ε_* below which the loop is internally stable.
- Model the uncertainty as a proportional (multiplicative) uncertainty and then use Inequality (5.70) to determine (numerically?) the largest delay ε_* below which the loop is internally stable.

5.6 Prove that k defined in (5.93) indeed satisfies (5.90).

5.7 *Structured singular value.* Fill in the details of the calculation of Summary 5.6.5.

5.8 *A MIMO feedback system.* Consider the MIMO system with the 2×2 loop gain transfer matrix

$$L(s) = \begin{bmatrix} 1/2 & 0 \\ ks/(s+2) & 0 \end{bmatrix}.$$

The loop gain depends on a parameter $k \in \mathbb{R}$.

- Find the ∞ -norm of L (the norm still depends on k).
- For which values of k does the small-gain theorem guarantee that the closed loop of Fig. 5.25 (page 214) is internally stable?
- Determine all k for which the closed loop of Fig. 5.25 is internally stable.
- Compute $\mu(L(j\omega))$ with respect to structure

$$\Delta = \begin{bmatrix} \Delta_{11} & \Delta_{12} \\ \Delta_{21} & \Delta_{22} \end{bmatrix}, \quad \Delta_{ij} \in \mathbb{C}.$$

5.9 *Numerator-denominator plant perturbation.* Table 5.1 lists several perturbation models.

- Verify the result for perturbed plant $P + V\Delta_P W$
- Verify the result for perturbed plant $(I + V\Delta_P W)^{-1}P$
- Verify the result for perturbed plant $(D + V\Delta_D W_1)^{-1}(N + V\Delta_N W_2)$

- 5.10 *Alternative characterization of the structured singular value.* Prove (Doyle, 1982) that if all uncertainties are complex then

$$\mu(M) = \max_{\Delta \in \mathcal{D}: \overline{\sigma}(\Delta) \leq 1} \rho(M\Delta), \quad (5.217)$$

with ρ denoting the spectral radius — that is, the magnitude of the largest eigenvalue (in magnitude). Show that if on the other hand some of the uncertainties are real then

$$\mu(M) = \max_{\Delta \in \mathcal{D}: \overline{\sigma}(\Delta) \leq 1} \rho_{\mathbb{R}}(M\Delta). \quad (5.218)$$

Here $\rho_{\mathbb{R}}(A)$ is largest of the magnitudes of the real eigenvalues of the complex matrix A . If A has no real eigenvalues then $\rho_{\mathbb{R}}(A) = 0$.

- 5.11 *Proof of the principal properties of the structured singular value* Fill in the details of the proof of Summary 5.6.5.

- 5.12 *Computation of structured singular values with MATLAB.* Reproduce the plots of Fig. 5.29 using the appropriate numerical routines from the μ -Toolbox or the Robust Control Toolbox for the computation of structured singular values.

- 5.13 *Structured singular values.* Let M be the 2×2 -matrix

$$M = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

- a) Determine $\overline{\sigma}(M)$
b) Determine $\mu(M)$ with respect to the structure

$$\Delta = \begin{bmatrix} \delta_1 & 0 \\ 0 & \delta_2 \end{bmatrix}, \quad \delta_1, \delta_2 \in \mathbb{C}.$$

- c) Determine

$$\inf_D \overline{\sigma}(DM D^{-1})$$

where the minimization is with respect to the diagonal matrices D of the form

$$D = \begin{bmatrix} d_1 & 0 \\ 0 & 1 \end{bmatrix}, \quad d_1 > 0.$$

- 5.14 *Multiple uncertain parameters in a state space description.* Suppose that $\dot{x} = Ax$ but that $A \in \mathbb{R}^{n \times n}$ is uncertain,

$$A = A_0 + \delta_1 A_1 + \delta_2 A_2 + \cdots + \delta_m A_m,$$

with A_i known and $\delta_i \in [-1, 1]$ unknown. The system $\dot{x} = Ax$ is *Robustly stable* if it is stable for every $\delta_i \in [-1, 1]$. Rewrite this robust stability problem into the problem of internal stability of a BPM.

- 5.15 *μ -analysis.* Consider the feedback loop of Fig. 6.1 and assume the plant is uncertain,

$$P(s) = \frac{N_0 + \sum_{i=1}^n \delta_{N_i} N_i}{D_0 + \sum_{i=1}^m \delta_{D_i} D_i}$$

where N_i and D_i are given polynomials, and $\delta_{N_i}, \delta_{D_i}$ uncertain real parameters in $[-1, 1]$.

- a) Given a controller C , determine
 - i. an interconnection matrix H ,
 - ii. an uncertainty Δ (expressed in terms of the δ_{Ni}, δ_{Di})
 such that the loop of Fig. 6.1 is internally stable for all possible $\delta_{Ni}, \delta_{Di} \in (-1, 1)$ iff $\sup_{\omega} \mu(H(j\omega)) \leq 1$.
- b) What is the rank of the interconnection matrix H as determined above?

5.16 *Bound on performance robustness.* In Example 5.7.4 we used the performance specification (in the form of W_1) to find bounds on the maximally allowable scaled uncertainties δ_P . Conversely, we may use the uncertainty specification (in the form of W_2) to estimate the largest possible performance variations.

- a) Suppose that g varies between 0.5 and 1.5 and θ between 0 and 0.2. Determine a function W_2 that tightly bounds the uncertainty Δ_P of (5.203).
- b) Use (5.201) to determine the smallest value of ε for which robust performance is guaranteed.
- c) Compute the sensitivity function S for a number of values of the parameters g and θ within the uncertainty region. Check how tight the bound on performance is that follows from the value of ε obtained in (5.16b).

6. \mathcal{H}_∞ -Optimization and μ -Synthesis

Overview – Design by \mathcal{H}_∞ -optimization involves the minimization of the peak magnitude of a suitable closed-loop system function. It is very well suited to frequency response shaping. Moreover, robustness against plant uncertainty may be handled more directly than with \mathcal{H}_2 optimization.

Design by μ -synthesis aims at reducing the peak value of the structured singular value. It accomplishes joint robustness and performance optimization.

6.1. Introduction

In this chapter we introduce what is known as \mathcal{H}_∞ -optimization as a design tool for linear multivariable control systems. \mathcal{H}_∞ -optimization amounts to the minimization of the ∞ -norm of a relevant frequency response function. The name derives from the fact that mathematically the problem may be set in the space \mathcal{H}_∞ (named after the British mathematician G. H. Hardy), which consists of all bounded functions that are analytic in the right-half complex plane. We do not go to this length, however.

\mathcal{H}_∞ -optimization resembles \mathcal{H}_2 -optimization, where the criterion is the 2-norm. Because the 2- and ∞ -norms have different properties the results naturally are not quite the same. An important aspect of \mathcal{H}_∞ optimization is that it allows to include robustness constraints explicitly in the criterion.

In § 6.2 (p. 240) we discuss the *mixed sensitivity problem*. This special \mathcal{H}_∞ problem is an important design tool. We show how this problem may be used to achieve the frequency response shaping targets enumerated in § 1.5 (p. 27).

In § 6.3 (p. 248) we introduce the *standard problem* of \mathcal{H}_∞ -optimization, which is the most general version. The mixed sensitivity problem is a special case. Several other special cases of the standard problem are exhibited.

The next section is devoted to *suboptimal* solutions of the standard \mathcal{H}_∞ -optimization problem. An example demonstrates the difficulties that can occur and they illustrate the type of assumptions needed for the solution of the \mathcal{H}_∞ -optimization problem.

In § 6.5 (p. 253) we review state space formulae for the suboptimal solutions and establish a lower bound for the ∞ -norm. Two algebraic Riccati equations are needed for the solution.

In § 6.6 (p. 255) we discuss optimal (as opposed to *suboptimal*) solutions and highlight some of their peculiarities. Section 6.7 (p. 258) explains how integral control and high-frequency roll-off

may be handled.

Section 6.8 (p. 266) is devoted to an introductory exposition of μ -synthesis. This approximate technique for joint robustness and performance optimization uses \mathcal{H}_∞ optimization to reduce the peak value of the structured singular value. Section 6.9 (p. 270) is given over to a rather elaborate description of an application of μ -synthesis.

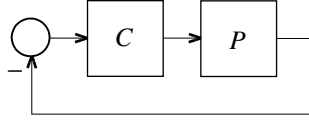


Figure 6.1: Feedback loop

6.2. The mixed sensitivity problem

In § 5.5.5 (p. 211) we studied the stability robustness of the feedback configuration of Fig. 6.1. We considered fractional perturbations of the type

$$L = ND^{-1} \longrightarrow (I + V\delta_N W_2)ND^{-1}(I + V\delta_D W_1)^{-1}. \quad (6.1)$$

The frequency dependent matrices V , W_1 , and W_2 are so chosen that the scaled perturbation $\delta_p = [-\delta_D \ \delta_N]$ satisfies $\|\delta_p\|_\infty \leq 1$. Stability robustness is guaranteed if

$$\|H\|_\infty < 1, \quad (6.2)$$

where

$$H = \begin{bmatrix} W_1 S V \\ -W_2 T V \end{bmatrix}. \quad (6.3)$$

$S = (I + L)^{-1}$ and $T = L(I + L)^{-1}$ are the sensitivity matrix and the complementary sensitivity matrix of the closed-loop system, respectively, with L the loop gain $L = PC$.

Given a feedback system with controller C and corresponding loop gain $L = PC$ that does not satisfy the inequality (6.2) one may look for a different controller that does achieve inequality. An effective way of doing this is to consider the problem of *minimizing* $\|H\|_\infty$ with respect to all controllers C that stabilize the system. If the minimal value γ of $\|H\|_\infty$ is greater than 1 then no controller exists that stabilizes the systems for all perturbations such that $\|\delta_p\|_\infty \leq 1$. In this case, stability robustness is only obtained for perturbations satisfying $\|\delta_p\|_\infty \leq 1/\gamma$.

The problem of minimizing

$$\left\| \begin{bmatrix} W_1 S V \\ -W_2 T V \end{bmatrix} \right\|_\infty \quad (6.4)$$

(Kwakernaak, 1983, 1985) is a version of what is known as the *mixed sensitivity problem* (Verma and Jonckheere, 1984). The name derives from the fact that the optimization involves both the sensitivity and the complementary sensitivity function.

In what follows we explain that the mixed sensitivity problem cannot only be used to verify stability robustness for a class of perturbations, but also to achieve a number of other important *design targets* for the one-degree-of-freedom feedback configuration of Fig. 6.1.

Before starting on this, however, we introduce a useful modification of the problem. We may write $W_2 T V = W_2 L(I + L)^{-1} V = W_2 P C(I + L)^{-1} V = W_2 P U V$, where

$$U = C(I + P C)^{-1} \quad (6.5)$$

is the input sensitivity matrix introduced in § 1.5 (p. 27). For a fixed plant we may absorb the plant transfer matrix P (and the minus sign) into the weighting matrix W_2 , and consider the modified problem of minimizing

$$\left\| \begin{bmatrix} W_1 S V \\ W_2 U V \end{bmatrix} \right\|_{\infty} \quad (6.6)$$

with respect to all stabilizing controllers. We redefine the problem of minimizing (6.6) as the mixed sensitivity problem.

We mainly consider the SISO mixed sensitivity problem. The criterion (6.6) then reduces to the square root of the scalar quantity

$$\sup_{\omega \in \mathbb{R}} (|W_1(j\omega)S(j\omega)V(j\omega)|^2 + |W_2(j\omega)U(j\omega)V(j\omega)|^2). \quad (6.7)$$

Many of the conclusions also hold for the MIMO case, although their application may be more involved.

6.2.1. Frequency response shaping

The mixed sensitivity problem may be used for simultaneously shaping the sensitivity and input sensitivity functions. The reason is that the solution of the mixed sensitivity problem often has the *equalizing property* (see § 6.6, p. 255). This property implies that the frequency dependent function

$$|W_1(j\omega)S(j\omega)V(j\omega)|^2 + |W_2(j\omega)U(j\omega)V(j\omega)|^2, \quad (6.8)$$

whose peak value is minimized, actually is a *constant* (Kwakernaak, 1985). If we denote the constant as γ^2 , with γ nonnegative, then it immediately follows from

$$|W_1(j\omega)S(j\omega)V(j\omega)|^2 + |W_2(j\omega)U(j\omega)V(j\omega)|^2 = \gamma^2 \quad (6.9)$$

that for the optimal solution

$$\begin{aligned} |W_1(j\omega)S(j\omega)V(j\omega)|^2 &\leq \gamma^2, & \omega \in \mathbb{R}, \\ |W_2(j\omega)U(j\omega)V(j\omega)|^2 &\leq \gamma^2, & \omega \in \mathbb{R}. \end{aligned} \quad (6.10)$$

Hence,

$$|S(j\omega)| \leq \frac{\gamma}{|W_1(j\omega)V(j\omega)|}, \quad \omega \in \mathbb{R}, \quad (6.11)$$

$$|U(j\omega)| \leq \frac{\gamma}{|W_2(j\omega)V(j\omega)|}, \quad \omega \in \mathbb{R}. \quad (6.12)$$

By choosing the functions W_1 , W_2 , and V correctly the functions S and U may be made small in appropriate frequency regions. This is also true if the optimal solution does not have the equalizing property.

If the weighting functions are suitably chosen (in particular, with $W_1 V$ large at low frequencies and $W_2 V$ large at high frequencies), then often the solution of the mixed sensitivity problem has the property that the first term of the criterion dominates at low frequencies and the second at high frequencies:

$$\underbrace{|W_1(j\omega)S(j\omega)V(j\omega)|^2}_{\text{dominates at low frequencies}} + \underbrace{|W_2(j\omega)U(j\omega)V(j\omega)|^2}_{\text{dominates at high frequencies}} = \gamma^2. \quad (6.13)$$

As a result,

$$|S(j\omega)| \approx \frac{\gamma}{|W_1(j\omega)V(j\omega)|} \quad \text{for } \omega \text{ small}, \quad (6.14)$$

$$|U(j\omega)| \approx \frac{\gamma}{|W_2(j\omega)V(j\omega)|} \quad \text{for } \omega \text{ large}. \quad (6.15)$$

This result allows quite effective control over the shape of the sensitivity and input sensitivity functions, and, hence, over the performance of the feedback system.

Because the ∞ -norm involves the supremum frequency response shaping based on minimization of the ∞ -norm is more direct than for the \mathcal{H}_2 optimization methods of § 4.5 (p. 155).

Note, however, that the limits of performance discussed in § 1.6 (p. 40) can never be violated. Hence, the weighting functions must be chosen with the respect due to these limits.

6.2.2. Type k control and high-frequency roll-off

In (6.14–6.15), equality may often be achieved *asymptotically*.

Type k control. Suppose that $|W_1(j\omega)V(j\omega)|$ behaves as $1/\omega^k$ as $\omega \rightarrow 0$, with k a nonnegative integer. This is the case if $W_1(s)V(s)$ includes a factor s^k in the denominator. Then $|S(j\omega)|$ behaves as ω^k as $\omega \rightarrow 0$, which implies a type k control system, with excellent low-frequency disturbance attenuation if $k \geq 1$. If $k = 1$ then the system has integrating action.

High-frequency roll-off. Likewise, suppose that $|W_2(j\omega)V(j\omega)|$ behaves as ω^m as $\omega \rightarrow \infty$. This is the case if $W_2 V$ is nonproper, that is, if the degree of the numerator of $W_2 V$ exceeds that of the denominator (by m). Then $|U(j\omega)|$ behaves as ω^{-m} as $\omega \rightarrow \infty$. From $U = C/(1 + PC)$ it follows that $C = U/(1 + UP)$. Hence, if P is strictly proper and $m \geq 0$ then also C behaves as ω^{-m} , and $T = PC/(1 + PC)$ behaves as $\omega^{-(m+e)}$, with e the pole excess¹ of P .

Hence, by choosing m we pre-assign the *high-frequency roll-off* of the controller transfer function, and the roll-offs of the complementary and input sensitivity functions. This is important for robustness against high-frequency unstructured plant perturbations.

Similar techniques to obtain type k control and high-frequency roll-off are used in § 4.5.4 (p. 158) and § 4.5.5 (p. 160), respectively, for \mathcal{H}_2 optimization.

6.2.3. Partial pole placement

There is a further important property of the solution of the mixed sensitivity problem that needs to be discussed before considering an example. This involves a pole cancellation phenomenon that is sometimes misunderstood. The equalizing property of § 6.2.1 (p. 241) implies that

$$W_1(s)W_1(-s)S(s)S(-s)V(s)V(-s) + W_2(s)W_2(-s)U(s)U(-s)V(s)V(-s) = \gamma^2 \quad (6.16)$$

¹The pole excess of a rational transfer function P is the difference between the number of poles and the number of zeros. This number is also known as the *relative degree* of P .

for all s in the complex plane. We write the transfer function P and the weighting functions W_1 , W_2 , and V in rational form as

$$P = \frac{N}{D}, \quad W_1 = \frac{A_1}{B_1}, \quad W_2 = \frac{A_2}{B_2}, \quad V = \frac{M}{E}, \quad (6.17)$$

with all numerators and denominators polynomials. If also the controller transfer function is represented in rational form as

$$C = \frac{Y}{X} \quad (6.18)$$

then it easily follows that

$$S = \frac{DX}{DX + NY}, \quad U = \frac{DY}{DX + NY}. \quad (6.19)$$

The denominator

$$D_{cl} = DX + NY \quad (6.20)$$

is the closed-loop characteristic polynomial of the feedback system. Substituting S and U as given by (6.19) into (6.16) we easily obtain

$$\frac{D \sim D \cdot M \sim M \cdot (A_1 \sim A_1 B_2 \sim B_2 X \sim X + A_2 \sim A_2 B_1 \sim B_1 Y \sim Y)}{E \sim E \cdot B_1 \sim B_1 \cdot B_2 \sim B_2 \cdot D_{cl} \sim D_{cl}} = \gamma^2. \quad (6.21)$$

If A is any rational or polynomial function then $A \sim$ is defined by $A \sim(s) = A(-s)$.

Since the right-hand side of (6.21) is a constant, all polynomial factors in the numerator of the rational function on the left cancel against corresponding factors in the denominator. In particular, the factor $D \sim D$ cancels. If there are no cancellations between $D \sim D$ and $E \sim E B_1 \sim B_1 B_2 \sim B_2$ then the closed-loop characteristic polynomial D_{cl} (which by stability has left-half plane roots only) necessarily has among its roots those roots of D that lie in the left-half plane, and the mirror images with respect to the imaginary axis of those roots of D that lie in the right-half plane. This means that the open-loop poles (the roots of D), possibly after having been mirrored into the left-half plane, *reappear* as closed-loop poles.

This phenomenon, which is not propitious for good feedback system design, may be prevented by choosing the denominator polynomial E of V equal to the plant denominator polynomial D , so that $V = M/D$. With this special choice of the denominator of V , the polynomial E cancels against D in the left-hand side of (6.21), so that the open-loop poles do *not* reappear as closed-loop poles.

Further inspection of (6.21) shows that if there are no cancellations between $M \sim M$ and $E \sim E B_1 \sim B_1 B_2 \sim B_2$, and we assume without loss of generality that M has left-half plane roots only, then the polynomial M cancels against a corresponding factor in D_{cl} . If we take V *proper* (which ensures $V(j\omega)$ to be finite at high frequencies) then the polynomial M has the same degree as D , and, hence, has the same number of roots as D .

All this means that by letting

$$V = \frac{M}{D}, \quad (6.22)$$

where the polynomial M has the same degree as the denominator polynomial D of the plant, the open-loop poles (the roots of D) are reassigned to the locations of the roots of M . By suitably

choosing the remaining weighting functions W_1 and W_2 these roots may often be arranged to be the *dominant* poles.

This technique, known as *partial pole placement* (Kwakernaak, 1986; Postlethwaite et al., 1990) allows further control over the design. It is very useful in designing for a specified bandwidth and good time response properties.

In the design examples in this chapter it is illustrated how the ideas of partial pole placement and frequency shaping are combined.

A discussion of the root loci properties of the mixed sensitivity problem may be found in Choi and Johnson (1996).

6.2.4. Example: Double integrator

We apply the mixed sensitivity problem to the same example as in § 4.6.2 (p. 161), where \mathcal{H}_2 optimization is illustrated². Consider a SISO plant with nominal transfer function

$$P_0(s) = \frac{1}{s^2}. \quad (6.23)$$

The actual, perturbed plant has the transfer function

$$P(s) = \frac{g}{s^2(1 + s\theta)}, \quad (6.24)$$

where g is nominally 1 and the nonnegative parasitic time constant θ is nominally 0.

Perturbation analysis. We start with a preliminary robustness analysis. The variations in the parasitic time constant θ mainly cause high-frequency perturbations, while the low-frequency perturbations are primarily the effect of the variations in the gain g . Accordingly, we model the effect of the parasitic time constant as a *numerator* perturbation, and the gain variations as *denominator* perturbations, and write

$$P(s) = \frac{N(s)}{D(s)} = \frac{\frac{1}{1+s\theta}}{\frac{s^2}{g}}. \quad (6.25)$$

Correspondingly, the relative perturbations of the denominator and the numerator are

$$\frac{D(s) - D_0(s)}{D_0(s)} = \frac{1}{g} - 1, \quad \frac{N(s) - N_0(s)}{N_0(s)} = \frac{-s\theta}{1 + s\theta}. \quad (6.26)$$

The relative perturbation of the denominator is constant over all frequencies, also in the crossover region. Because the plant is minimum-phase trouble-free crossover may be achieved (that is, without undue peaking of the sensitivity and complementary sensitivity functions). Hence, we expect that—in the absence of other perturbations—values of $|1/g - 1|$ up to almost 1 may be tolerated.

The size of the relative perturbation of the numerator is less than 1 for frequencies below $1/\theta$, and equal to 1 for high frequencies. To prevent destabilization it is advisable to make the complementary sensitivity small for frequencies greater than $1/\theta$. As the complementary sensitivity starts to decrease at the closed-loop bandwidth, the largest possible value of θ dictates

²Much of the text of this subsection has been taken from Kwakernaak (1993).

the bandwidth. Assuming that performance requirements specify the system to have a closed-loop bandwidth of 1, we expect that — in the absence of other perturbations — values of the parasitic time constant θ up to 1 do not destabilize the system.

Thus, both for robustness and for performance, we aim at a closed-loop bandwidth of 1 with small sensitivity at low frequencies and a sufficiently fast decrease of the complementary sensitivity at high frequencies with a smooth transition in the crossover region.

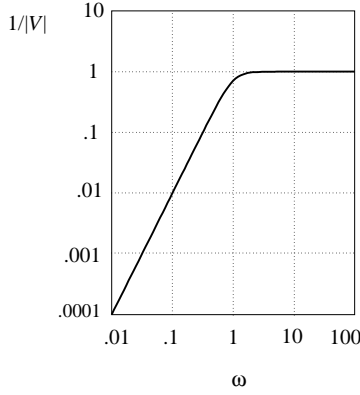


Figure 6.2: Bode magnitude plot of $1/V$

Choice of the weighting functions. To accomplish this with a mixed sensitivity design, we successively consider the choice of the functions $V = M/D$ (that is, of the polynomial M), W_1 and W_2 .

To obtain a good time response corresponding to the bandwidth 1, which does not suffer from sluggishness or excessive overshoot, we assign two dominant poles to the locations $\frac{1}{2}\sqrt{2}(-1 \pm j)$. This is achieved by choosing the polynomial M as

$$M(s) = [s - \frac{1}{2}\sqrt{2}(-1 + j)][s - \frac{1}{2}\sqrt{2}(-1 - j)] = s^2 + s\sqrt{2} + 1, \quad (6.27)$$

so that

$$V(s) = \frac{s^2 + s\sqrt{2} + 1}{s^2}. \quad (6.28)$$

We tentatively choose the weighting function W_1 equal to 1. Then if the first of the two terms of the mixed sensitivity criterion dominates at low frequencies from we have from (6.14) that

$$|S(j\omega)| \approx \frac{\gamma}{|V(j\omega)|} = \gamma \left| \frac{(j\omega)^2}{(j\omega)^2 + j\omega\sqrt{2} + 1} \right| \quad \text{at low frequencies.} \quad (6.29)$$

Figure 6.2 shows the magnitude plot of the factor $1/V$. The plot implies a very good low-frequency behavior of the sensitivity function. Owing to the presence of the double open-loop pole at the origin the feedback system is of type 2. There is no need to correct this low frequency behavior by choosing W_1 different from 1.

Next contemplate the high-frequency behavior. For high frequencies V is constant and equal to 1. Consider choosing W_2 as

$$W_2(s) = c(1 + rs), \quad (6.30)$$

with c and r nonnegative constants such that $c \neq 0$. Then for high frequencies the magnitude of $W_2(j\omega)$ asymptotically behaves as c if $r = 0$, and as $cr\omega$ if $r \neq 0$.

Hence, if $r = 0$ then the high-frequency roll-off of the input sensitivity function U and the controller transfer function C is 0 and that of the complementary sensitivity T is 2 decades/decade (40 dB/decade).

If $r \neq 0$ then U and C roll off at 1 decade/decade (20 dB/decade) and T rolls off at 3 decades/decade (60 dB/decade).

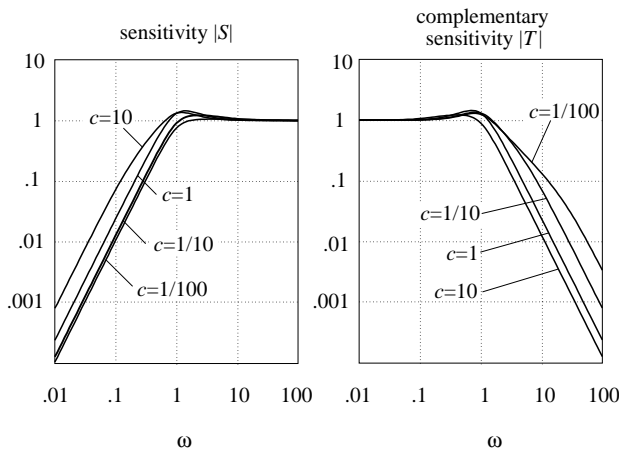


Figure 6.3: Bode magnitude plots of S and T for $r = 0$

Solution of the mixed sensitivity problem. We first study the case $r = 0$, which results in a proper but not strictly proper controller transfer function C , and a high-frequency roll-off of T of 2 decades/decade. Figure 6.3 shows³ the optimal sensitivity function S and the corresponding complementary sensitivity function T for $c = 1/100$, $c = 1/10$, $c = 1$, and $c = 10$. Inspection shows that as c increases, $|T|$ decreases and $|S|$ increases, which conforms to expectation. The smaller c is, the closer the shape of $|S|$ is to that of the plot of Fig. 6.2.

We choose $c = 1/10$. This makes the sensitivity small with little peaking at the cut-off frequency. The corresponding optimal controller has the transfer function

$$C(s) = 1.2586 \frac{s + 0.61967}{1 + 0.15563s}, \quad (6.31)$$

and results in the closed-loop poles $\frac{1}{2}\sqrt{2}(-1 \pm j)$ and -5.0114 . The two former poles dominate the latter pole, as planned. The minimal ∞ -norm is $\|H\|_{\infty} = 1.2861$.

Robustness against high-frequency perturbations may be improved by making the complementary sensitivity function T decrease faster at high frequencies. This is accomplished by taking the constant r nonzero. Inspection of W_2 as given by (6.30) shows that by choosing $r = 1$ the resulting extra roll-off of U , C , and T sets in at the frequency 1. For $r = 1/10$ the break point is shifted to the frequency 10. Figure 6.4 shows the resulting magnitude plots. For $r = 1/10$ the sensitivity function has little extra peaking while starting at the frequency 10 the complementary

³The actual computation of the controller is discussed in Example 6.6.1 (p. 256).

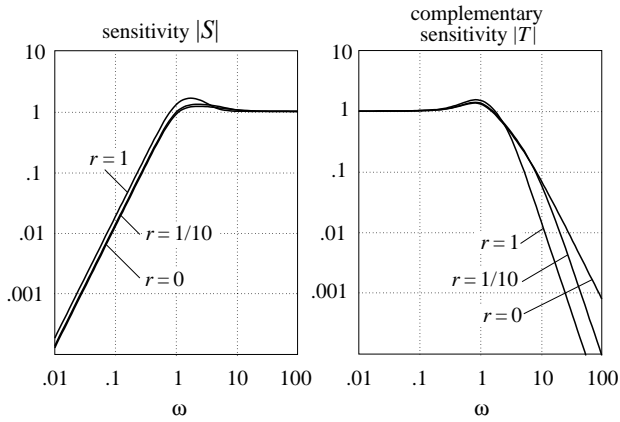


Figure 6.4: Bode magnitude plots of S and T for $c = 1/10$

sensitivity function rolls off at a rate of 3 decades/decade. The corresponding optimal controller transfer function is

$$C(s) = 1.2107 \frac{s + 0.5987}{1 + 0.20355s + 0.01267s^2}, \quad (6.32)$$

which results in the closed-loop poles $\frac{1}{2}\sqrt{2}(-1 \pm j)$ and $-7.3281 \pm j1.8765$. Again the former two poles dominate the latter. The minimal ∞ -norm is $\|H\|_{\infty} = 1.3833$.

Inspection of the two controllers (6.31) and (6.32) shows that both basically are lead controllers with high-frequency roll-off.

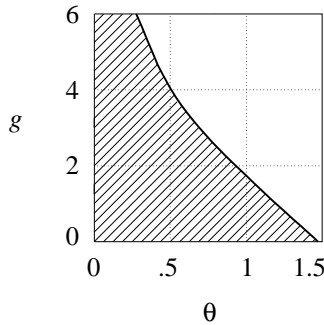


Figure 6.5: Stability region

Robustness analysis. We conclude this example with a brief analysis to check whether our expectations about robustness have come true. Given the controller $C = Y/X$ the closed-loop characteristic polynomial of the perturbed plant is $D_{cl}(s) = D(s)X(s) + N(s)Y(s) = (1 + s\theta)s^2X(s) + gY(s)$. By straightforward computation, which involves fixing one of the two parameters g and θ and varying the other, the stability region of Fig. 6.5 may be established for the controller (6.31). That for the other controller is similar. The diagram shows that for $\theta = 0$ the closed-loop system is stable for all $g > 0$, that is, for all $-1 < \frac{1}{g} - 1 < \infty$. This stability

interval is larger than predicted. For $g = 1$ the system is stable for $0 \leq \theta < 1.179$, which also is a somewhat larger interval than expected.

The controller (6.31) is similar to the modified PD controller that is obtained in Example 5.2.2 (p. 185) by root locus design. Likewise, the stability region of Fig. 6.5 is similar to that of Fig. 5.5.

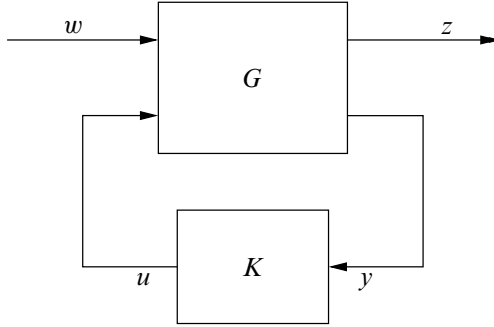


Figure 6.6: The standard \mathcal{H}_∞ problem configuration

6.3. The standard \mathcal{H}_∞ problem

The mixed sensitivity problem is a special case of the so-called *standard \mathcal{H}_∞ problem* (Doyle, 1984). This standard problem is defined by the configuration of Fig. 6.6. The “plant” is a given system with two sets of inputs and two sets of outputs. It is often referred to as the *generalized plant*. The signal w is an *external input*, and represents driving signals that generate disturbances, measurement noise, and reference inputs. The signal u is the *control input*. The output z has the meaning of *control error*, and ideally should be zero. The output y , finally, is the *observed output*, and is available for feedback. The plant has an open-loop transfer matrix G such that

$$\begin{bmatrix} z \\ y \end{bmatrix} = G \begin{bmatrix} w \\ u \end{bmatrix} = \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix} \begin{bmatrix} w \\ u \end{bmatrix}. \quad (6.33)$$

By connecting the feedback controller

$$u = Ky \quad (6.34)$$

we obtain from $y = G_{21}w + G_{22}u$ the closed-loop signal balance equation $y = G_{21}w + G_{22}Ky$, so that $y = (I - G_{22}K)^{-1}G_{21}w$. From $z = G_{11}w + G_{12}u = G_{11}w + G_{12}Ky$ we then have

$$z = \underbrace{[G_{11} + G_{12}K(I - G_{22}K)^{-1}G_{21}]}_H w. \quad (6.35)$$

Hence, the closed-loop transfer matrix H is

$$H = G_{11} + G_{12}K(I - G_{22}K)^{-1}G_{21}. \quad (6.36)$$

The standard \mathcal{H}_∞ -optimal regulation problem is the problem of determining a controller with transfer matrix K that

1. internally stabilizes the closed-loop system (as defined in § 1.3.2 (p. 12)), and

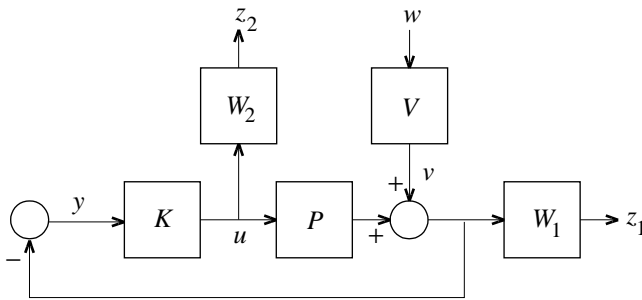


Figure 6.7: The mixed sensitivity problem

2. minimizes the ∞ -norm $\|H\|_\infty$ of the closed-loop transfer matrix H from the external input w to the control error z .

To explain why the mixed sensitivity problem is a special case of the standard problem we consider the block diagram of Fig. 6.7, where the controller transfer function now is denoted K rather than C . In this diagram, the external signal w generates the disturbance v after passing through a shaping filter with transfer matrix V . The “control error” z has two components, z_1 and z_2 . The first component z_1 is the control system output after passing through a weighting filter with transfer matrix W_1 . The second component z_2 is the plant input u after passing through a weighting filter with transfer matrix W_2 .

It is easy to verify that for the closed-loop system

$$z = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \underbrace{\begin{bmatrix} W_1 S V \\ -W_2 U V \end{bmatrix}}_H w, \quad (6.37)$$

so that minimization of the ∞ -norm of the closed-loop transfer matrix H amounts to minimization of

$$\left\| \begin{bmatrix} W_1 S V \\ W_2 U V \end{bmatrix} \right\|_\infty. \quad (6.38)$$

In the block diagram of Fig. 6.7 we denote the input to the controller as y , as in the standard problem of Fig. 6.6. We read off from the block diagram that

$$\begin{aligned} z_1 &= W_1 V w + W_1 P u, \\ z_2 &= W_2 u, \\ y &= -V w - P u. \end{aligned} \quad (6.39)$$

Hence, the open-loop transfer matrix G for the standard problem is

$$G = \left[\begin{array}{c|c} W_1 V & W_1 P \\ \hline 0 & W_2 \\ \hline -V & -P \end{array} \right] \quad (6.40)$$

Other well-known special cases of the standard problem \mathcal{H}_∞ problem are the *filtering problem* and the *minimum sensitivity problem* (see Exercises 6.6 and 6.3).

The \mathcal{H}_∞ -optimal regulation problem is treated in detail by [Green and Limebeer \(1995\)](#) and [Zhou et al. \(1996\)](#).

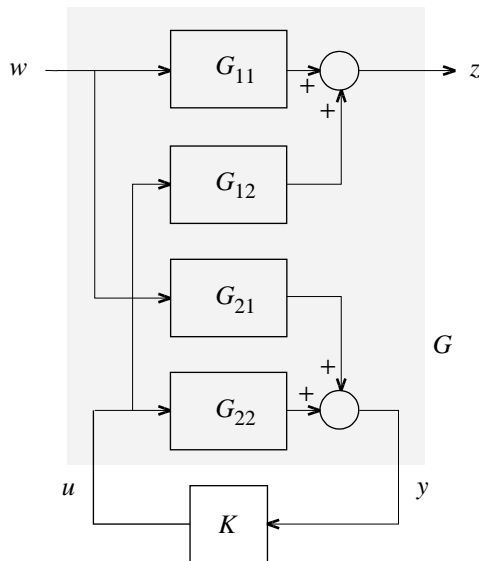


Figure 6.8: Feedback arrangement for the standard problem

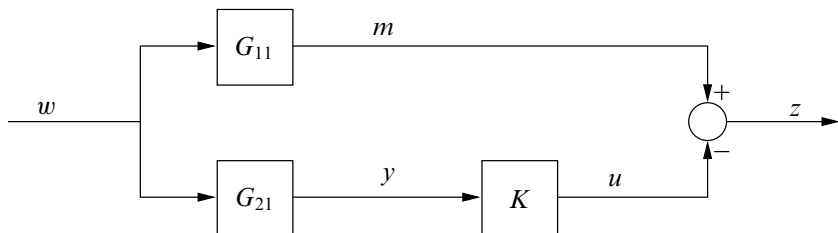


Figure 6.9: A filtering problem

6.4. Suboptimal solutions and and example

Determining the optimal controller K of the standard \mathcal{H}_∞ problem is not an easy task. To get a feel for the problems that can arise and for the assumptions that are required—assumptions also required by software—we examine in this section a problem that is easy enough to allow for an explicit solution.

Consider the configuration of Fig. 6.9. It is a special case of the standard H_∞ control configuration of Fig. 6.6 (see Exercise 6.6) but it is not a control problem as there is no feedback. It depicts a *filtering problem*. The idea being that the system K (called filter in this respect) should try to filter the information of the ‘message’ m out of the limited information of y . Ideally, then, the output of K equals m , rendering $z \equiv 0$, but generally this is not possible. It then makes sense to choose K so as to minimize z in some norm, for instance using the \mathcal{H}_2 - or \mathcal{H}_∞ -norm.

In this section we develop optimal and suboptimal solutions of the filtering \mathcal{H}_∞ -problem for the case that

$$G_{11}(s) = 1, \quad G_{21} = \frac{s-2}{s+2}.$$

Since both G_{11} and G_{21} are stable it will be clear that the loop of Fig. 6.9 is internally stable if

and only if K is stable. The transfer function H from w to z whose ∞ -norm we aim to minimize, is

$$H(s) = G_{11}(s) - G_{21}(s)K(s) = 1 - \frac{s-2}{s+2}K(s).$$

As we do not yet know what $K(s)$ is we also do not know what $H(s)$ is, except at $s = 2$ because for that value of s the term $\frac{s-2}{s+2}K(s)$ reduces to zero. Indeed, $K(s)$ is not allowed to have a pole at $s = 2$ because K needs to be stable. So, whatever stable K we take, we always have that

$$H(2) = 1.$$

At this point we recall a useful result, a result which is a consequence of the classical *Maximum Modulus Principle* from complex functions theory; our special version, though, may also be proved using the small gain theorem (see § 6.10 and also Summary 1.7.4):

Lemma 6.4.1. If $H(s)$ is a stable rational function then $\|H\|_\infty \geq |H(z_0)|$ for any z_0 in the right-half complex plane. \square

For our example this implies

$$\|H\|_\infty \geq 1$$

for any stable K . But, this norm can be achieved: simply take $K = 0$. It results in the constant $H = 1$ which has norm $\|H\|_\infty = 1$.

Next we derive so-called *suboptimal* solution. Suboptimal solutions are stabilizing K that achieve

$$\|H\|_\infty < \gamma, \tag{6.41}$$

with γ a given nonnegative number. The reason for determining suboptimal solutions is that optimal solutions in most cases cannot be found directly. Once suboptimal solutions may be obtained, optimal solutions may be approximated by searching on the number γ .

We first need the following result. Recall that $\|H\|_\infty$ is finite if and only if H is proper and stable and that then

$$\|H\|_\infty = \|H\|_{\mathcal{L}_\infty} := \sup_{\omega \in \mathbb{R}} \bar{\sigma}(H(j\omega)). \tag{6.42}$$

Summary 6.4.2 (Inequality for \mathcal{L}_∞ -norm). Let γ be a nonnegative number and H a rational transfer matrix. Then $\|H\|_{\mathcal{L}_\infty} \leq \gamma$ is equivalent to either of the following statements:

1. $H^\sim H \leq \gamma^2 I$ on the imaginary axis;
2. $HH^\sim \leq \gamma^2 I$ on the imaginary axis.

\square

Here H^\sim is defined by $H^\sim(s) = H^T(-s)$. H^\sim is called the *adjoint* of H . If A and B are two complex-valued matrices then $A \leq B$ means that $B - A$ is a nonnegative-definite matrix. “On the imaginary axis” means for all $s = j\omega$, $\omega \in \mathbb{R}$. The proof of Summary 6.4.2 may be found in § 6.10 (p. 281).

For our filtering problem we have as optimal value the bound

$$\gamma_{\text{opt}} = 1.$$

So suboptimal solutions exist iff $\gamma > 1$. Application of Summary 6.4.2 to our problem gives

$$\begin{aligned}
\|H\|_{\mathcal{L}_\infty} \leq \gamma &\iff (1 - \frac{s-2}{s+2}K)^\sim (1 - \frac{s-2}{s+2}K) \leq \gamma^2 \\
&\iff (1 - \frac{s+2}{s-2}K^\sim) (1 - \frac{s-2}{s+2}K) \leq \gamma^2 \\
&\iff 1 - \gamma^2 - \frac{s+2}{s-2}K^\sim - \frac{s-2}{s+2}K + K^\sim K \leq 0 \\
&\iff \begin{bmatrix} K^\sim & 1 \end{bmatrix} \begin{bmatrix} 1 & -\frac{s+2}{s-2} \\ -\frac{s-2}{s+2} & 1 - \gamma^2 \end{bmatrix} \begin{bmatrix} K \\ 1 \end{bmatrix} \leq 0.
\end{aligned}$$

Interestingly the matrix in the middle of the last inequality can also be written as

$$\begin{bmatrix} 1 & -\frac{s+2}{s-2} \\ -\frac{s-2}{s+2} & 1 - \gamma^2 \end{bmatrix} = W^\sim(s) \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} W(s)$$

with

$$W(s) = \begin{bmatrix} W_{11}(s) & W_{12}(s) \\ W_{21}(s) & W_{22}(s) \end{bmatrix} = \begin{bmatrix} \frac{\gamma^2 - \frac{s-2}{s+2}}{\gamma^2 - 1} & 1 \\ \frac{\gamma}{\gamma^2 - 1} (1 - \frac{s-2}{s+2}) & \gamma \end{bmatrix}. \quad (6.43)$$

Note that $W(s)$ is well defined for $\gamma > \gamma_{\text{opt}} = 1$. So we have

$$\|H\|_{\mathcal{L}_\infty} \leq \gamma \iff \begin{bmatrix} K^\sim & 1 \end{bmatrix} W^\sim \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} W \begin{bmatrix} K \\ 1 \end{bmatrix} \leq 0.$$

If we choose K as

$$K = -\frac{W_{21}}{W_{11}}$$

then by construction $W \begin{bmatrix} K \\ 1 \end{bmatrix}$ if of the form $\begin{bmatrix} 0 \\ A \end{bmatrix}$ so that

$$\begin{bmatrix} K^\sim & 1 \end{bmatrix} W^\sim \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} W \begin{bmatrix} K \\ 1 \end{bmatrix} = \begin{bmatrix} 0 & A^\sim \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 0 \\ A \end{bmatrix} = -A^\sim A \leq 0.$$

This implies that $\|H\|_{\mathcal{L}_\infty} \leq \gamma$. With the W of (6.43) the so constructed $K = -\frac{W_{21}}{W_{11}}$ equals

$$K(s) = \frac{s+2}{s + \frac{\gamma^2+2}{\gamma^2-1}} = \frac{(\gamma^2-1)(s+2)}{(\gamma^2-1)s + \gamma^2+2}. \quad (6.44)$$

This filter K has a pole at $s = -\frac{\gamma^2+2}{\gamma^2-1}$ so clearly for $\gamma > 1$ it is stable, hence it is a solution of the suboptimal \mathcal{H}_∞ problem with bound γ . (There is a theory that shows that stability of K is not a coincidence, but we do not need to bother about these details here.) As γ approaches $\gamma_{\text{opt}} = 1$ from above the pole of K goes to minus infinity, and at $\gamma = 1$ the pole disappears leaving the optimal filter $K = 0$.

This example illustrates a number of points, most of which also occur in more complicated problems.

- In this example it turned out to be possible to come up with an explicit formula for the *optimal* solution K . This generally is very complicated if not impossible.
- What is easier in the general case is to solve the *suboptimal* problem, i.e. the problem whether or not there exists a stabilizing K that achieve $\|H\|_\infty < \gamma$ for some given bound $\gamma > 0$.

- A line search in γ then can bring us as close as we want to the optimal controller K .
- Some coefficients of suboptimal controllers approach zero (or infinity) as γ approaches γ_{opt} . Only at $\gamma = \gamma_{\text{opt}}$ do these coefficients disappear: at $\gamma = \gamma_{\text{opt}}$ a cancellation of common factors occurs leaving an optimal controller of lower order. This is a common situation.

It is unfortunate that software usually does not perform the cancellation for us, even though it can be done. We have to do it manually and we can not discard it. Indeed, as the example shows the controller may have a pole that approaches minus infinity as γ approaches γ_{opt} . For γ close to the optimal value the controller hence has a very fast mode. This is to be avoided.

- If in this example G_{12} would have had imaginary zeros then K , which equals $K = \frac{G_{11}-H}{G_{12}}$, would normally have had imaginary poles (not allowed). This case complicates the standard theory and is typically ruled out by software.
- Similarly, if G_{12} would have been strictly proper then $K = \frac{G_{11}-H}{G_{12}}$ would normally have been non-proper. Also this case complicates the standard theory and it is ruled out by software. This situation is an indication that the problem is not well formulated. Indeed, if the optimal K tends to be nonproper then it may be wise to minimize not $\|H\|_{\infty}$ but, for instance, $\|[\frac{H}{W_2 K}]\|_{\infty}$ for some appropriate weight W_2 that ensures that (sub)optimal K 's to have sufficient high-frequency roll-off.

For the last two reasons one usually assumes that G_{12} in the filtering example has no zeros on the imaginary axis, including infinity.

6.5. State space solution of the standard \mathcal{H}_{∞} problem

Among the various solutions of the suboptimal standard \mathcal{H}_{∞} problem, the one based on state space realizations is the most popular, (Doyle et al., 1989; Stoorvogel, 1992; Trentelman and Stoorvogel, 1993). In these approaches it is assumed that the generalized plant G is proper. Hence it has a realization of the form

$$\dot{x} = Ax + B_1 w + B_2 u, \quad (6.45)$$

$$z = C_1 x + D_{11} w + D_{12} u, \quad (6.46)$$

$$y = C_2 x + D_{21} w + D_{22} u. \quad (6.47)$$

In the μ -TOOLS MATLAB toolbox (Balas et al., 1991) and the ROBUST CONTROL TOOLBOX (Chiang and Safonov, 1992) a solution of the corresponding \mathcal{H}_{∞} problem based on Riccati equations is implemented that requires the following conditions to be satisfied:

1. (A, B_2) is stabilizable and (C_2, A) is detectable.
2. $\begin{bmatrix} A - j\omega I & B_2 \\ C_1 & D_{12} \end{bmatrix}$ has full column rank for all $\omega \in \mathbb{R}$ (hence, D_{12} is tall⁴).
3. $\begin{bmatrix} A - j\omega I & B_1 \\ C_2 & D_{21} \end{bmatrix}$ has full row rank for all $\omega \in \mathbb{R}$ (hence, D_{21} is wide).
4. D_{12} and D_{21} have full rank.

⁴A matrix is tall if it has at least as many rows as columns. It is wide if it has at least as many columns as rows.

The first assumption is natural for otherwise the system can not be stabilized. The other assumptions ensure that G_{12} and G_{21} have full column rank and full row rank respectively on the imaginary axis, including infinity. With these assumptions the formulae for suboptimal controllers are rather technical but for a special case they are manageable:

Theorem 6.5.1 (The solution of the standard \mathcal{H}_∞ problem). Consider the configuration of Fig. 6.6 and assume the above four assumptions are satisfied, and for simplicity, that also

$$\begin{bmatrix} B_1 \\ D_{21} \end{bmatrix} D_{21}^T = \begin{bmatrix} 0 \\ I \end{bmatrix}, \quad D_{12}^T [C_1 \quad D_{12}] = [0 \quad I]. \quad (6.48)$$

Then there exists a stabilizing controller for which $\|H\|_\infty < \gamma$ iff the following three conditions hold.

1. $AQ + QA^T + Q(\frac{1}{\gamma^2}C_1^T C_1 - C_2^T C_2)Q + B_1 B_1^T = 0$ has a stabilizing solution $Q \geq 0$,
2. $PA + A^T P + P(\frac{1}{\gamma^2}B_1 B_1^T - B_2 B_2^T)P + C_1^T C_1 = 0$ has a stabilizing solution $P \geq 0$
3. All eigenvalues of QP have magnitude less than γ^2 .

If these three conditions are satisfied then one controller $u = Ky$ that stabilizes and achieves $\|H\|_\infty < \gamma$ is the controller with realization

$$\begin{cases} \dot{\hat{x}} &= (A + [\frac{1}{\gamma^2}B_1 B_1^T - B_2 B_2^T]P)\hat{x} + (I - \frac{1}{\gamma^2}QP)^{-1}QC_2^T(y - C_2\hat{x}) \\ u &= -B_2^T\hat{x} \end{cases} \quad (6.49)$$

The formulae for K are rather cumbersome if the assumptions (6.48) do not hold, computationally it makes no difference. The solution, as we see, involves two algebraic Riccati equations whose solutions define an observer *cum* state feedback law. The full solution is documented in a paper by Glover and Doyle (1988). More extensive treatments may be found in a celebrated paper by Doyle et al. (1989) and in Glover and Doyle (1989). Stoorvogel (1992) discusses a number of further refinements of the problem. The problem can also be solved using linear matrix inequalities (LMIs). LMIs are convex programs; an important topic but one that is not covered in this course.

6.5.1. Characteristics of the state space solution

We list a few properties of the state space solution.

1. For the two Riccati equations to have a solution it is required that the associated Hamiltonian matrices

$$\begin{bmatrix} A & B_1 B_1^T \\ -\frac{1}{\gamma^2}C_1^T C_1 + C_2^T C_2 & -A^T \end{bmatrix}, \quad \begin{bmatrix} A & \frac{1}{\gamma^2}B_1 B_1^T - B_2^T B_2 \\ -C_1^T C_1 & -A^T \end{bmatrix}$$

have no imaginary eigenvalues (see § 4.7.4). Stated differently, if γ_0 is the *largest* value of γ for which one or both of the above two Hamiltonian matrices has an imaginary eigenvalue, then

$$\gamma_{\text{opt}} \geq \gamma_0.$$

2. The controller (6.49) is stabilizing iff

$$Q \geq 0, \quad P \geq 0, \quad \lambda_{\max}(QP) < \gamma^2, \quad (6.50)$$

where λ_{\max} denotes the largest eigenvalue. This is a convenient way to test whether stabilizing controllers exist.

3. The controller (6.49) is of the same order as the generalized plant (6.45–6.47).

4. The transfer matrix of the controller (6.49) is strictly proper.

6.6. Optimal solutions to the \mathcal{H}_∞ problem

Finding optimal controllers as opposed to suboptimal controllers involves a search over the parameter γ . As the search advances the optimal controller is approached more and more closely. There are two possibilities for the optimal solution:

Type A solution. The suboptimal controller (6.49) is stabilizing for all $\gamma \geq \gamma_0$, with γ_0 the lower bound discussed in 6.5.1. In this case, the optimal solution is obtained for $\gamma = \gamma_0$.

Type B solution. As γ varies, the central controller becomes destabilizing as γ decreases below some number γ_{opt} with $\gamma_{\text{opt}} \geq \gamma_0$.

In type B solutions a somewhat disconcerting phenomenon occurs. In the example of § 6.4 about filtering several of the coefficients of central controller (filter) grow without bound as γ approaches $\gamma_{\text{opt}} = 1$. In the state space solution the phenomenon is manifested by the fact that as γ approaches γ_{opt} either of the two following eventualities occurs (Glover and Doyle, 1989):

(B1.) The matrix $I - \frac{1}{\gamma^2}QP$ becomes singular.

(B2.) The solutions P and/or Q grow without bound, and at $\gamma = \gamma_{\text{opt}}$ they do not exist.

In both cases large coefficients occur in the equations that define the central controller.

As illustrated in the example of § 6.4, however, the controller transfer matrix approaches a well-defined limit as $\gamma \downarrow \gamma_{\text{opt}}$, corresponding to the *optimal* controller. The type B optimal controller is of lower order than the suboptimal central controller. Also this is generally true. It is possible to characterize and compute *all optimal* solutions of type B (Glover et al., 1991; Kwakernaak, 1991).

An important characteristic of optimal solutions of type B is that the largest singular value $\bar{\sigma}(H(j\omega))$ of the optimal closed-loop frequency response matrix H is *constant* as a function of the frequency $\omega \in \mathbb{R}$. This is known as the *equalizing property* of optimal solutions.

Straightforward implementation of the two-Riccati equation algorithm leads to numerical difficulties for type B problems. As the solution approaches the optimum several or all of the coefficients of the controller become very large. Because eventually the numbers become too large the optimum cannot be approached too closely. This is something we have to keep in mind when we compute such controllers.

6.6.1. Numerical examples

We present the numerical solution of two examples of the standard problem.

Example 6.6.1 (Mixed sensitivity problem for the double integrator). We consider the mixed sensitivity problem discussed in § 6.2.4 (p. 244). With P , V , W_1 and W_2 as in 6.3, by (6.39) the standard plant transfer matrix is

$$G(s) = \left[\begin{array}{c|c} \frac{M(s)}{s^2} & \frac{1}{s^2} \\ \hline 0 & c(1 + rs) \\ \hline -\frac{M(s)}{s^2} & -\frac{1}{s^2} \end{array} \right], \quad (6.51)$$

where $M(s) = s^2 + s\sqrt{2} + 1$. We consider the case $r = 0$ and $c = 0.1$. It is easy to check that for $r = 0$ a state space representation of the plant is

$$\dot{x} = \underbrace{\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}}_A x + \underbrace{\begin{bmatrix} 1 \\ \sqrt{2} \end{bmatrix}}_{B_1} w + \underbrace{\begin{bmatrix} 1 \\ 0 \end{bmatrix}}_{B_2} u, \quad (6.52)$$

$$z = \underbrace{\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}}_{C_1} x + \underbrace{\begin{bmatrix} 1 \\ 0 \end{bmatrix}}_{D_{11}} w + \underbrace{\begin{bmatrix} 0 \\ c \end{bmatrix}}_{D_{12}} u, \quad (6.53)$$

$$y = \underbrace{\begin{bmatrix} 0 & -1 \end{bmatrix}}_{C_2} x + \underbrace{\begin{bmatrix} -1 \end{bmatrix}}_{D_{21}} w. \quad (6.54)$$

Note that we constructed a joint minimal realization of the blocks P and V in the diagram of Fig. 6.7. This is necessary to satisfy the stabilizability condition of § 6.5 (p. 253).

A numerical solution may be obtained with the help of the μ -TOOLS MATLAB toolbox (Balas et al., 1991). The search procedure for \mathcal{H}_∞ state space problems implemented in μ -TOOLS terminates at $\gamma = 1.2861$ (for a specified tolerance in the γ -search of 10^{-8}). The state space representation of the corresponding controller is

$$\dot{\hat{x}} = \begin{bmatrix} -0.3422 \times 10^9 & -1.7147 \times 10^9 \\ 1 & -1.4142 \end{bmatrix} \hat{x} + \begin{bmatrix} -0.3015 \\ -0.4264 \end{bmatrix} y, \quad (6.55)$$

$$u = [-1.1348 \times 10^9 \quad -5.6871 \times 10^9] \hat{x}. \quad (6.56)$$

Numerical computation of the transfer function of the controller results in

$$K(s) = \frac{2.7671 \times 10^9 s + 1.7147 \times 10^9}{s^2 + 0.3422 \times 10^9 s + 2.1986 \times 10^9}. \quad (6.57)$$

The solution is of type B as indicated by the large coefficients. By discarding the term s^2 in the denominator we may reduce the controller to

$$K(s) = 1.2586 \frac{s + 0.6197}{1 + 0.1556s}, \quad (6.58)$$

which is the result stated in (6.31)(p. 246). It may be checked that the optimal controller possesses the equalizing property: the mixed sensitivity criterion does not depend on frequency.

The case $r \neq 0$ is dealt with in Exercise 6.10 (p. 284). □

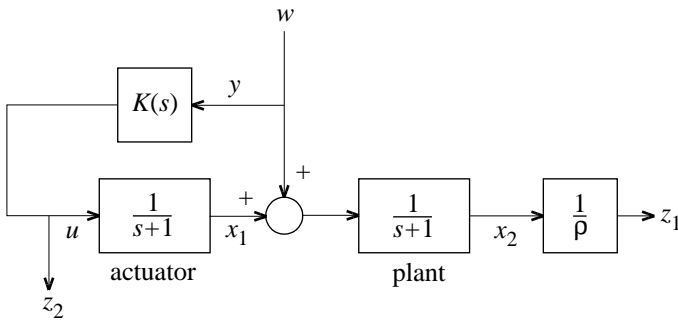


Figure 6.10: Feedforward configuration

Example 6.6.2 (Disturbance feedforward). Hagander and Bernhardsson (1992) discuss the example of Fig. 6.10. A disturbance w acts on a plant with transfer function $1/(s + 1)$. The plant is controlled by an input u through an actuator that also has the transfer function $1/(s + 1)$. The disturbance w may be measured directly, and it is desired to control the plant output by feedforward control from the disturbance. Thus, the observed output y is w . The “control error” has as components the weighted control system output $z_1 = x_2/\rho$, with ρ a positive coefficient, and the control input $z_2 = u$. The latter is included to prevent overly large inputs.

In state space form the standard plant is given by

$$\dot{x} = \underbrace{\begin{bmatrix} -1 & 0 \\ 1 & -1 \end{bmatrix}}_A x + \underbrace{\begin{bmatrix} 0 \\ 1 \end{bmatrix}}_{B_1} w + \underbrace{\begin{bmatrix} 1 \\ 0 \end{bmatrix}}_{B_2} u, \quad (6.59)$$

$$z = \underbrace{\begin{bmatrix} 0 & \frac{1}{\rho} \\ 0 & 0 \end{bmatrix}}_{C_1} x + \underbrace{\begin{bmatrix} 0 \\ 0 \end{bmatrix}}_{D_{11}} w + \underbrace{\begin{bmatrix} 0 \\ 1 \end{bmatrix}}_{D_{12}} u, \quad (6.60)$$

$$y = \underbrace{\begin{bmatrix} 0 & 0 \end{bmatrix}}_{C_2} x + \underbrace{\begin{bmatrix} 1 \end{bmatrix}}_{D_{21}} w. \quad (6.61)$$

The lower bound of § 6.5.1 (p. 254) may be found to be given by

$$\gamma_0 = \frac{1}{\sqrt{1 + \rho^2}}. \quad (6.62)$$

Define the number

$$\rho_c = \sqrt{\frac{\sqrt{5} + 1}{2}} = 1.270. \quad (6.63)$$

Hagander and Bernhardsson prove this (see Exercise 6.8):

1. For $\rho > \rho_c$ the optimal solution is of type A.
2. For $\rho < \rho_c$ the optimal solution is of type B.

For $\rho = 2$, for instance, numerical computation using μ -Tools with a tolerance 10^{-8} results in

$$\gamma_{\text{opt}} = \gamma_0 = \frac{1}{\sqrt{5}} = 0.4472, \quad (6.64)$$

with the central controller

$$\dot{\hat{x}} = \begin{bmatrix} -1.6229 & -0.4056 \\ 1 & -1 \end{bmatrix} \hat{x} + \begin{bmatrix} 0 \\ 0.7071 \end{bmatrix} y, \quad (6.65)$$

$$u = [-0.8809 \quad -0.5736] \hat{x}. \quad (6.66)$$

The coefficients of the state representation of the controller are of the order of unity, and the controller transfer function is

$$K(s) = \frac{-0.4056s - 0.4056}{s^2 + 2.6229s + 2.0285}. \quad (6.67)$$

There is no question of order reduction. The optimal solution is of type A.

For $\rho = 1$ numerical computation leads to

$$\gamma_{\text{opt}} = 0.7167 > \gamma_0 = 0.7071, \quad (6.68)$$

with the central controller

$$\dot{\hat{x}} = \begin{bmatrix} -6.9574 \times 10^7 & -4.9862 \times 10^7 \\ 1 & -1 \end{bmatrix} \hat{x} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} y, \quad (6.69)$$

$$u = [-6.9574 \times 10^7 \quad -4.9862 \times 10^7] \hat{x}. \quad (6.70)$$

The controller transfer function is

$$K(s) = \frac{-4.9862 \times 10^7 s - 4.9862 \times 10^7}{s^2 + 0.6957 \times 10^8 s + 1.1944 \times 10^8}. \quad (6.71)$$

By discarding the term s^2 in the denominator this reduces to the first-order controller

$$K(s) = -0.7167 \frac{s + 1}{s + 1.7167}. \quad (6.72)$$

The optimal solution now is of type B. □

6.7. Integral control and high-frequency roll-off

In this section we discuss the application of \mathcal{H}_∞ optimization, in particular the mixed sensitivity problem, to achieve two specific design targets: integral control and high-frequency roll-off. Integral control is dealt with in § 6.7.1. In § 6.7.2 we explain how to design for high-frequency roll-off. Subsection 6.7.3 is devoted to an example.

The methods to obtain integral control and high frequency roll-off discussed in this section for \mathcal{H}_∞ design may also be used with \mathcal{H}_2 optimization. This is illustrated for SISO systems in § 4.5.4 (p. 158) and § 4.5.5 (p. 160).

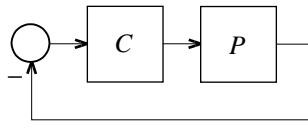


Figure 6.11: One-degree-of-freedom feedback system

6.7.1. Integral control

In § 2.3 (p. 64) it is explained that integral control is a powerful and important technique. By making sure that the control loop contains “integrating action” robust rejection of constant disturbances may be obtained, as well as excellent low-frequency disturbance attenuation and command signal tracking. In § 6.2.2 (p. 242) it is claimed that the mixed sensitivity problem allows to design for integrating action. Consider the SISO mixed sensitivity problem for the one-degree-of-freedom configuration of Fig. 6.11. The mixed sensitivity problem amounts to the minimization of the peak value of the function

$$|V(j\omega)W_1(j\omega)S(j\omega)|^2 + |V(j\omega)W_2(j\omega)U(j\omega)|^2, \quad \omega \in \mathbb{R}. \quad (6.73)$$

S and U are the sensitivity function and input sensitivity function

$$S = \frac{1}{1 + PC}, \quad U = \frac{C}{1 + PC}, \quad (6.74)$$

respectively, and V , W_1 and W_2 suitable frequency dependent weighting functions.

If the plant P has “natural” integrating action, that is, P has a pole at 0, then no special provisions are needed. In the absence of natural integrating action we may introduce integrating action by letting the product VW_1 have a pole at 0. This forces the sensitivity function S to have a zero at 0, because otherwise (6.73) is unbounded at $\omega = 0$.

There are two ways to introduce such a pole into VW_1 :

1. Let V have a pole at 0, that is, take

$$V(s) = \frac{V_0(s)}{s}, \quad (6.75)$$

with $V_0(0) \neq 0$. We call this the *constant disturbance model* method.

2. Let W_1 have a pole at 0, that is, take

$$W_1(s) = \frac{W_{1o}(s)}{s}, \quad (6.76)$$

with $W_{1o}(0) \neq 0$. This is the *constant error suppression* method.

We discuss these two possibilities.

Constant disturbance model. We first consider letting V have a pole at 0. Although VW_1 needs to have a pole at 0, the weighting function VW_2 cannot have such a pole. If VW_2 has a pole at 0 then U would be forced to be zero at 0. S and U cannot vanish simultaneously at 0. Hence, if V has a pole at 0 then W_2 should have a zero at 0.

This makes sense. Including a pole at 0 in V means that V contains a model for constant disturbances. Constant disturbances can only be rejected by constant inputs to the plant. Hence,

zero frequency inputs should not be penalized by W_2 . Therefore, if V is of the form (6.75) then we need

$$W_2(s) = sW_{2o}(s), \quad (6.77)$$

with $W_{2o}(0) \neq \infty$.

Figure 6.12(a) defines the interconnection matrix G for the resulting standard problem. Inspection shows that owing to the pole at 0 in the block $V_0(s)/s$ outside the feedback loop this standard problem does not satisfy the stabilizability condition of § 6.5 (p. 253).

The first step towards resolving this difficulty is to modify the diagram to that of Fig. 6.12(b), where the plant transfer matrix has been changed to $sP(s)/s$. The idea is to construct a simultaneous minimal state realization of the blocks $V_0(s)/s$ and $sP(s)/s$. This brings the unstable pole at 0 “inside the loop.”

The difficulty now is, of course, that owing to the cancellation in $sP(s)/s$ there is no minimal realization that is controllable from the plant input. Hence, we remove the factor s from the numerator of $sP(s)/s$ by modifying the diagram to that of Fig. 6.12(c). By a minimal joint realization of the blocks $V_0(s)/s$ and $P(s)/s$ the interconnection system may now be made stabilizable. This is illustrated in the example of § 6.7.3.

The block diagram of Fig. 6.12(c) defines a modified mixed sensitivity problem. Suppose that the controller K_0 solves this modified problem. Then the original problem is solved by controller

$$K(s) = \frac{K_0(s)}{s}. \quad (6.78)$$

Constant error suppression. We next consider choosing $W_1(s) = W_{1o}(s)/s$. This corresponds to penalizing constant errors in the output with infinite weight. Figure 6.13(a) shows the corresponding block diagram. Again straightforward realization results in violation of the stabilizability condition of § 6.5 (p. 253). The offending factor $1/s$ may be pulled inside the loop as in Fig. 6.13(b). Further block diagram substitution yields the arrangement of Fig. 6.13(c).

Integrator in the loop method. Contemplation of the block diagrams of Figs. 6.12(c) and 6.13(c) shows that both the constant disturbance model method and the constant error rejection method may be reduced to the *integrator in the loop* method. This method amounts to connecting an integrator $1/s$ in series with the plant P as in Fig. 6.14(a), and doing a mixed sensitivity design for this augmented system.

After a controller K_0 has been obtained for the augmented plant the actual controller that is implemented consists of the series connection $K(s) = K_0(s)/s$ of the controller K_0 and the integrator as in Fig. 6.14(b). To prevent the pole at 0 of the integrator from reappearing in the closed-loop system it is necessary to let V have a pole at 0. Replacing $V(s)$ with $V_0(s)/s$ in the diagram of Fig. 6.14(a) produces the block diagrams of Figs. 6.12(c) and 6.13(c).

6.7.2. High-frequency roll-off

As explained in § 6.2.2 (p. 242), high-frequency roll-off in the controller transfer function K and the input sensitivity function U may be pre-assigned by suitably choosing the high-frequency behavior of the weighting function W_2 . If $W_2(s)V(s)$ is of order s^m as s approaches ∞ then K and U have a high-frequency roll-off of m decades/decade. Normally V is chosen biproper⁵ such

⁵That is, both V and $1/V$ are proper. This means that $V(\infty)$ is finite and nonzero.

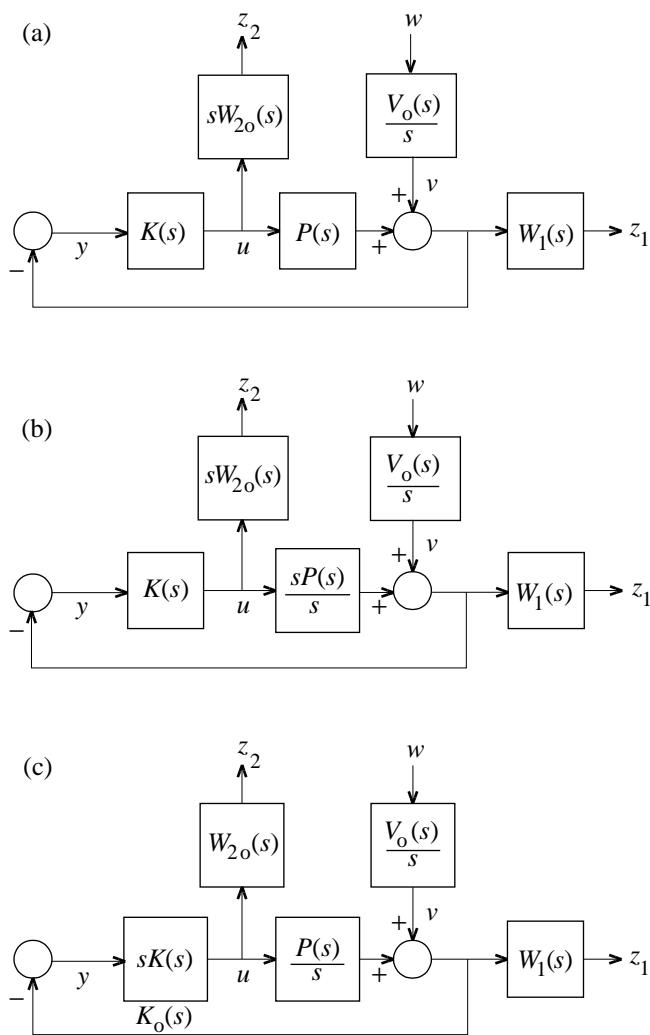


Figure 6.12: Constant disturbance model

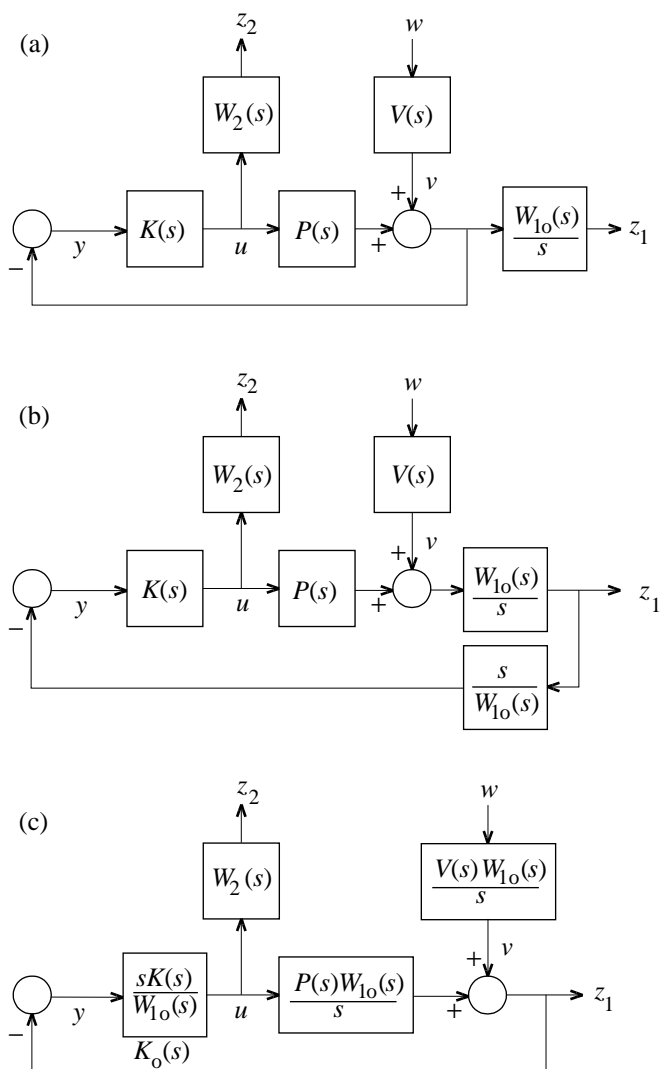


Figure 6.13: Constant error suppression

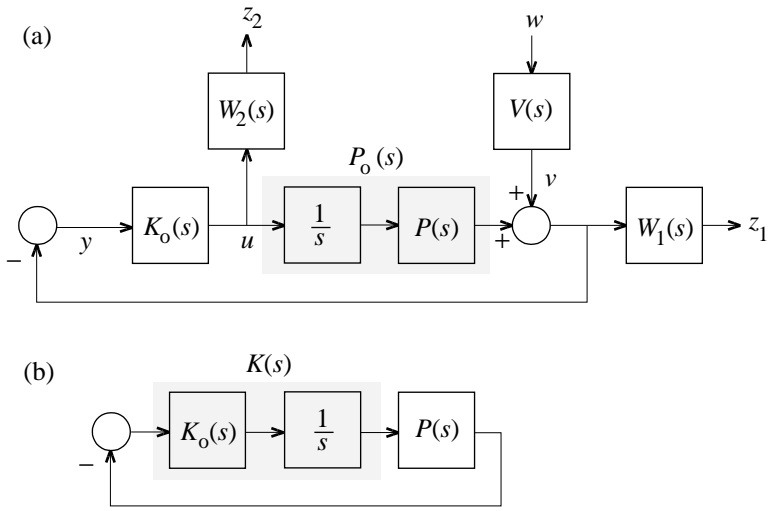


Figure 6.14: Integrator in the loop method

that $V(\infty) = 1$. Hence, to achieve a roll-off of m decades/decade it is necessary that W_2 behaves as s^m for large s .

If the desired roll-off m is greater than 0 then the weighting function W_2 needs to be nonproper. The resulting interconnection matrix

$$G = \left[\begin{array}{c|c} W_1 V & W_1 P \\ \hline 0 & W_2 \\ \hline -V & -P \end{array} \right] \quad (6.79)$$

is also nonproper and, hence, cannot be realized as a state system of the form needed for the state space solution of the \mathcal{H}_∞ -problem.

The makeshift solution usually seen in the literature is to cut off the roll-on at high frequency. Suppose by way of example that the desired form of W_2 is $W_2(s) = c(1 + rs)$, with c and r positive constants. Then W_2 may be made proper by modifying it to

$$W_2(s) = \frac{c(1 + rs)}{1 + s\tau}, \quad (6.80)$$

with $\tau \ll r$ a small positive time constant.

This contrivance may be avoided by the block diagram substitution of Fig. 6.15 (Krause, 1992). If W_2 is nonproper then, in the SISO case, W_2^{-1} is necessarily proper, and there is no difficulty in finding a state realization of the equivalent mixed sensitivity problem defined by Fig. 6.15 with the modified plant $P_0 = W_2^{-1}P$.

Suppose that the modified problem is solved by the controller K_0 . Then the original problem is solved by

$$K = K_0 W_2^{-1}. \quad (6.81)$$

Inspection suggests that the optimal controller K has the zeros of W_2 as its poles. Because this is in fact not the case the extra poles introduced into K cancel against corresponding zeros. This increase in complexity is the price of using the state space solution.

The example of § 6.7.3 illustrates the procedure.

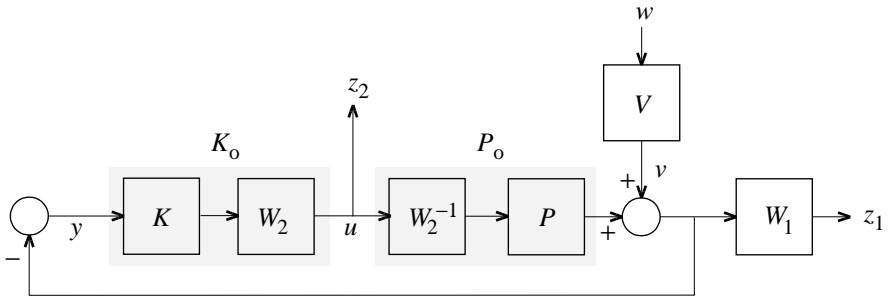


Figure 6.15: Block diagram substitution for nonproper W_2

6.7.3. Example

We present a simple example to illustrate the mechanics of designing for integrating action and high-frequency roll-off. Consider the first-order plant

$$P(s) = \frac{p}{s + p}, \quad (6.82)$$

with $p = 0.1$. These are the design specifications:

1. Constant disturbance rejection.
2. A closed-loop bandwidth of 1.
3. A suitable time response to step disturbances without undue sluggishness or overshoot.
4. High-frequency roll-off of the controller and input sensitivity at 1 decade/decade.

We use the integrator in the loop method of § 6.7.1 to design for integrating action. Accordingly, the plant is modified to

$$P_0(s) = \frac{p}{s(s + p)}. \quad (6.83)$$

Partial pole placement (see § 6.2.3, p. 242) is achieved by letting

$$V(s) = \frac{s^2 + as + b}{s(s + p)}. \quad (6.84)$$

The choice $a = \sqrt{2}$ and $b = 1$ places two closed-loop poles at the roots

$$\frac{1}{2}\sqrt{2}(-1 \pm j) \quad (6.85)$$

of the numerator of V . We plan these to be the dominant closed-loop poles in order to achieve a satisfactory time response and the required bandwidth.

To satisfy the high-frequency roll-off specification we let

$$W_2(s) = c(1 + rs), \quad (6.86)$$

with the positive constants c and r to be selected⁶. It does not look as if anything needs to be accomplished with W_1 so we simply let $W_1(s) = 1$.

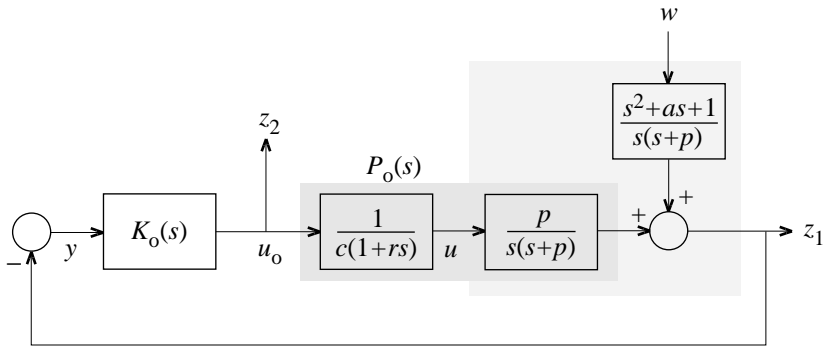


Figure 6.16: Modified block diagram for the example

Figure 6.16 shows the block diagram of the mixed sensitivity problem thus defined. It is obtained from Fig. 6.14(a) with the substitution of Fig. 6.15.

We can now see why in the eventual design the controller K has no pole at the zero $-1/r$ of the weighting factor W_2 . The diagram of Fig. 6.16 defines a standard mixed sensitivity problem with

$$P(s) = \frac{\frac{p}{rc}}{s(s + \frac{1}{r})(s + p)}, \quad (6.87)$$

$$V(s) = \frac{s^2 + as + b}{s(s + p)} = \frac{(s + \frac{1}{r})(s^2 + as + b)}{s(s + \frac{1}{r})(s + p)}, \quad (6.88)$$

and $W_1(s) = W_2(s) = 1$.

If V is in the form of the rightmost side of (6.88) then its numerator has a zero at $-1/r$. By the partial pole assignment argument of § 6.2.3 (p. 242) this zero reappears as a closed-loop pole. Because this zero is also an open-loop pole it is necessarily a zero of the controller K_0 that solves the mixed sensitivity problem of Fig. 6.16. This zero, in turn, cancels in the transfer function K of (6.81).

The system within the larger shaded block of Fig. 6.16 has the transfer matrix representation

$$z_1 = \begin{bmatrix} \frac{s^2 + as + b}{s(s + p)} & \frac{p}{s(s + p)} \end{bmatrix} \begin{bmatrix} w \\ u \end{bmatrix}. \quad (6.89)$$

This system has the minimal state realization

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 1 & -p \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} b \\ a - p \end{bmatrix} w + \begin{bmatrix} p \\ 0 \end{bmatrix} u, \quad (6.90)$$

$$z_1 = \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + w. \quad (6.91)$$

The block

$$u = \frac{1}{c(1 + rs)} u_0 \quad (6.92)$$

⁶See also Exercise 6.9 (p. 283).

may be represented in state form as

$$\dot{x}_3 = -\frac{1}{r}x_3 + \frac{1}{c}u_0, \quad (6.93)$$

$$u = \frac{1}{r}x_3. \quad (6.94)$$

Combining the state differential equations (6.90) and (6.93) and arranging the output equations we obtain the equations for the interconnection system that defines the standard problem as follows:

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & 0 & \frac{p}{r} \\ 1 & -p & 0 \\ 0 & 0 & -\frac{1}{r} \end{bmatrix}}_A \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \underbrace{\begin{bmatrix} b \\ a-p \\ 0 \end{bmatrix}}_{B_1} w + \underbrace{\begin{bmatrix} 0 \\ 0 \\ \frac{1}{c} \end{bmatrix}}_{B_2} u_0, \quad (6.95)$$

$$z = \underbrace{\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}}_{C_1} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \underbrace{\begin{bmatrix} 1 \\ 0 \end{bmatrix}}_{D_{11}} w + \underbrace{\begin{bmatrix} 0 \\ 1 \end{bmatrix}}_{D_{12}} u_0 \quad (6.96)$$

$$y = \underbrace{\begin{bmatrix} 0 & -1 & 0 \end{bmatrix}}_{C_2} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \underbrace{(-1)}_{D_{21}} w. \quad (6.97)$$

Let $c = 1$ and $r = 0.1$. Numerical computation of the optimal controller yields the type B solution

$$K_0(s) = \frac{1.377 \times 10^{10}s^2 + 14.055 \times 10^{10}s + 2.820 \times 10^{10}}{s^3 + 0.1142 \times 10^{10}s^2 + 1.3207 \times 10^{10}s + 1.9195 \times 10^{10}} \quad (6.98)$$

Neglecting the term s^3 in the denominator results in

$$K_0(s) = \frac{1.377s^2 + 14.055s + 2.820}{0.1142s^2 + 1.3207s + 1.91195} \quad (6.99)$$

$$= 12.056 \frac{(s+10)(s+0.2047)}{(s+9.8552)(s+1.7049)}. \quad (6.100)$$

Hence, the optimal controller for the original problem is

$$K(s) = \frac{K_0(s)}{sW_2(s)} = 12.056 \frac{(s+10)(s+0.2047)}{(s+9.8552)(s+1.7049)} \cdot \frac{10}{s(s+10)} \quad (6.101)$$

$$= \frac{120.56(s+0.2047)}{s(s+9.8552)(s+1.7049)}. \quad (6.102)$$

The pole at -10 cancels, as expected, and the controller has integrating action, as planned. The controller has a high-frequency roll-off of 2 decades/decade.

6.8. μ -Synthesis

In § 5.7 (p. 223) the use of the structured singular value for the analysis of stability and performance robustness is explained. We briefly review the conclusions.

In Fig. 6.17(a), the interconnection block H represents a control system, with structured perturbations represented by the block Δ . The signal w is the external input to the control system and z the control error. The input and output signals of the perturbation block are denoted p and q , respectively. The perturbations are assumed to have been scaled such that all possible structured perturbations are bounded by $\|\Delta\|_\infty \leq 1$.

Let H^Δ denote the transfer function from w to z (under perturbation Δ). Again, the transfer function is assumed to have been scaled in such a way that the control system performance is deemed satisfactory if $\|H^\Delta\|_\infty \leq 1$.

The central result established in § 5.7 is that the control system

1. is robustly stable, that is, is stable under all structured perturbations Δ such that $\|\Delta\|_\infty \leq 1$, and
2. has robust performance, that is, $\|H^\Delta\|_\infty \leq 1$ under all structured perturbations Δ such that $\|\Delta\|_\infty \leq 1$,

if and only if

$$\mu_H < 1. \quad (6.103)$$

The quantity μ_H , defined in (5.155), is the structured singular value of H with respect to perturbations as in Fig. 6.17(b), with Δ structured and Δ_0 unstructured.

Suppose that the control system performance and stability robustness may be modified by feedback through a controller with transfer matrix K , as in Fig. 6.18(a). The controller feeds the measured output y back to the control input u . Denote by μ_{H_K} the structured singular value of the closed-loop transfer matrix H_K of the system of Fig. 6.18(b) with respect to structured perturbations Δ and unstructured perturbations Δ_0 . Then μ -synthesis is any procedure to construct a controller K (if any exists) that stabilizes the nominal feedback system and makes

$$\mu_{H_K} < 1. \quad (6.104)$$

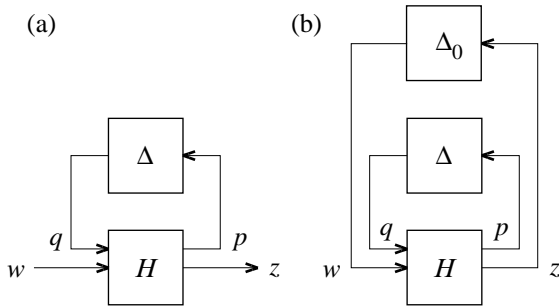


Figure 6.17: μ -Analysis for robust stability and performance.

6.8.1. Approximate μ -synthesis for SISO robust performance

In § 5.7 the SISO single-degree-of-freedom configuration of Fig. 6.19 is considered. We analyze its robustness with respect to proportional plant perturbations of the form

$$P \longrightarrow P(1 + \delta_p W_2) \quad (6.105)$$

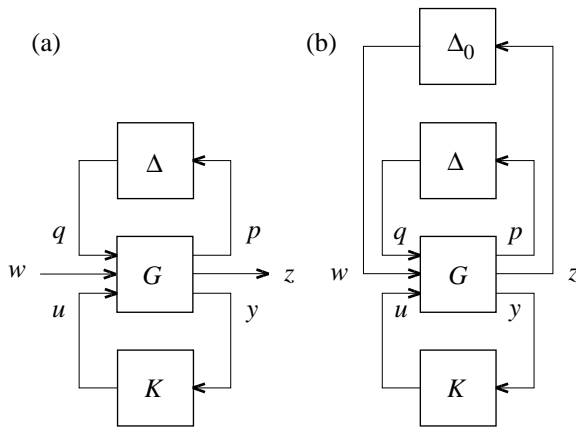


Figure 6.18: μ -Synthesis for robust stability and performance.

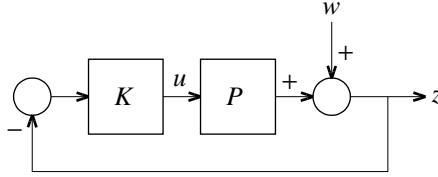


Figure 6.19: Single-degree- of-freedom configuration

with $\|\delta_P\|_\infty \leq 1$. The system is deemed to have satisfactory performance robustness if

$$\|W_1 S\|_\infty \leq 1 \quad (6.106)$$

for all perturbations, with S the sensitivity function of the closed-loop system. W_1 and W_2 are suitable frequency dependent functions.

By determining the structured singular value it follows in § 5.7 that the closed-loop system is robustly stable and has robust performance if and only if

$$\mu_{H_K} = \sup_{\omega \in \mathbb{R}} (|W_1(j\omega)S(j\omega)| + |W_2(j\omega)T(j\omega)|) < 1. \quad (6.107)$$

T is the complementary sensitivity function of the closed-loop system. Hence, the robust performance and stability problem consists of determining a feedback controller K (if any exists) that stabilizes the nominal system and satisfies $\mu_{H_K} < 1$.

One way of doing this is to *minimize* μ_{H_K} with respect to all stabilizing controllers K . Unfortunately, this is not a standard \mathcal{H}_∞ -optimal control problem. The problem cannot be solved by the techniques available for the standard \mathcal{H}_∞ problem. In fact, no solution is known to this problem.

By the well-known inequality $(a + b)^2 \leq 2(a^2 + b^2)$ for any two real numbers a and b it follows that

$$(|W_1 S| + |W_2 T|)^2 \leq 2(|W_1 S|^2 + |W_2 T|^2). \quad (6.108)$$

Hence,

$$\sup_{\omega \in \mathbb{R}} (|W_1(j\omega)S(j\omega)|^2 + |W_2(j\omega)T(j\omega)|^2) < \frac{1}{2} \quad (6.109)$$

implies $\mu_{H_K} < 1$. Therefore, if we can find a controller K such that

$$\sup_{\omega \in \mathbb{R}} (|W_1(j\omega)S(j\omega)|^2 + |W_2(j\omega)T(j\omega)|^2)^{1/2} = \left\| \begin{bmatrix} W_1 S \\ W_2 T \end{bmatrix} \right\|_{\infty} < \frac{1}{2}\sqrt{2} \quad (6.110)$$

then this controller achieves robust performance and stability. Such a controller, if any exists, may be obtained by minimizing

$$\left\| \begin{bmatrix} W_1 S \\ W_2 T \end{bmatrix} \right\|_{\infty}, \quad (6.111)$$

which is nothing but a mixed sensitivity problem.

Thus, the robust design problem has been reduced to a mixed sensitivity problem. This reduction is not necessarily successful in solving the robust design problem — see Exercise 6.13 (p. 284).

6.8.2. Approximate solution of the μ -synthesis problem

More complicated robust design problems than the simple SISO problem we discussed cannot be reduced to a tractable problem so easily. Various approximate solution methods to the μ -synthesis problem have been suggested. The best known of these (Doyle, 1984) relies on the property (see Summary 5.6.5(2–3))

$$\mu(M) \leq \bar{\sigma}[DM\bar{D}^{-1}], \quad (6.112)$$

with D and \bar{D} suitable diagonal matrices. The problem of minimizing

$$\mu_{H_K} = \sup_{\omega \in \mathbb{R}} \mu(H_K(j\omega)) \quad (6.113)$$

is now replaced with the problem of minimizing the bound

$$\sup_{\omega \in \mathbb{R}} \bar{\sigma}[D(j\omega)H_K(j\omega)\bar{D}^{-1}(j\omega)] = \|DH_K\bar{D}^{-1}\|_{\infty}, \quad (6.114)$$

where for each frequency ω the diagonal matrices $D(j\omega)$ and $\bar{D}(j\omega)$ are chosen so that the bound is the tightest possible. Minimizing (6.114) with respect to K is a standard \mathcal{H}_{∞} problem, provided D and \bar{D} are rational stable matrix functions.

Doyle's method is based on *D-K iteration*:

Summary 6.8.1 (*D-K iteration*).

1. Choose an initial controller K that stabilizes the closed-loop system, and compute the corresponding nominal closed-loop transfer matrix H_K .

One way of finding an initial controller is to minimize $\|H_K^0\|_{\infty}$ with respect to all stabilizing K , where H_K^0 is the closed-loop transfer matrix of the configuration of Fig. 6.18 from w to z with $\Delta = \Delta_0 = 0$.

2. Evaluate the upper bound

$$\min_{D(j\omega), \bar{D}(j\omega)} \bar{\sigma}[D(j\omega)H_K(j\omega)\bar{D}^{-1}(j\omega)], \quad (6.115)$$

with D and \bar{D} diagonal matrices as in Summary 5.6.5(c), for a number of values of ω on a suitable frequency grid. The maximum of this upper bound over the frequency grid is an estimate of μ_{H_K} .

If μ_{H_K} is small enough then stop. Otherwise, continue.

3. On the frequency grid, fit stable minimum-phase rational functions to the diagonal entries of D and \bar{D} . Because of the scaling property it is sufficient to fit their *magnitudes* only. The resulting extra freedom (in the phase) is used to improve the fit. Replace the original matrix functions D and \bar{D} with their rational approximations.
4. Given the rational approximations D and \bar{D} , minimize $\|DH_K\bar{D}^{-1}\|_\infty$ with respect to all stabilizing controllers K . Denote the minimizing controller as K and the corresponding closed-loop transfer matrix as H_K . Return to 2.

□

Any algorithm that solves the standard \mathcal{H}_∞ -problem exactly or approximately may be used in step (d). The procedure is continued until a satisfactory solution is obtained. Convergence is not guaranteed. The method may be implemented with routines provided in the μ -Tools toolbox (Balas et al., 1991). A lengthy example is discussed in § 6.9.

The method is essentially restricted to “complex” perturbations, that is, perturbations corresponding to dynamic uncertainties. “Real” perturbations, caused by uncertain parameters, need to be overbounded by dynamic uncertainties.

6.9. An application of μ -synthesis

To illustrate the application of μ -synthesis we consider the by now familiar SISO plant of Example 5.2.1 (p. 184) with transfer function

$$P(s) = \frac{g}{s^2(1 + s\theta)}. \quad (6.116)$$

Nominally, $g = g_0 = 1$ and $\theta = 0$, so that the nominal plant transfer function is

$$P_0(s) = \frac{g_0}{s^2}. \quad (6.117)$$

In Example 5.2.2 (p. 185) root locus techniques are used to design a modified PD controller with transfer function

$$C(s) = \frac{k + sT_d}{1 + sT_0}, \quad k = 1, \quad T_d = \sqrt{2} = 1.414, \quad T_0 = 0.1. \quad (6.118)$$

The corresponding closed-loop poles are $-0.7652 \pm j 0.7715$ and -8.4679 .

In this section we explore how D - K iteration may be used to obtain an improved design that achieves robust performance.

6.9.1. Performance specification

In Example 5.7.4 (p. 228) the robustness of the feedback system is studied with respect to the performance specification

$$|S(j\omega)| \leq \frac{1}{|V(j\omega)|}, \quad \omega \in \mathbb{R}. \quad (6.119)$$

The function V is chosen as

$$\frac{1}{V} = (1 + \varepsilon)S_0. \quad (6.120)$$

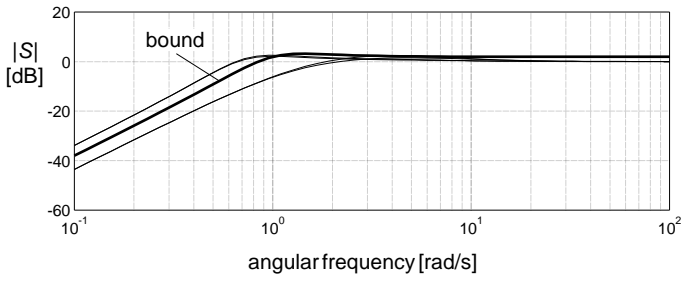


Figure 6.20: Magnitude plot of S for different parameter value combinations and the bound $1/V$

S_0 is the sensitivity function of the nominal closed-loop system and is given by

$$S_0(s) = \frac{s^2(1 + sT_0)}{\chi_0(s)}, \quad (6.121)$$

with $\chi_0(s) = T_0s^3 + s^2 + g_0T_d s + g_0k$ the nominal closed-loop characteristic polynomial. The number ε is a tolerance.

In Example 5.7.4 it is found that robust performance is guaranteed with a tolerance $\varepsilon = 0.25$ for parameter perturbations that satisfy

$$0.9 \leq g \leq 1.1, \quad 0 \leq \theta < 0.1. \quad (6.122)$$

The robustness test used in this example is based on a proportional loop perturbation model.

In the present example we wish to redesign the system such that performance robustness is achieved for a modified range of parameter perturbations, namely for

$$0.5 \leq g \leq 1.5, \quad 0 \leq \theta < 0.05. \quad (6.123)$$

For the purposes of this example we moreover change the performance specification model. Instead of relating performance to the nominal performance of a more or less arbitrary design we choose the weighting function V such that

$$\frac{1}{V(s)} = (1 + \varepsilon) \frac{s^2}{s^2 + 2\zeta\omega_0 s + \omega_0^2}. \quad (6.124)$$

Again, ε is a tolerance. The constant ω_0 specifies the minimum bandwidth and ζ determines the maximally allowable peaking. Numerically we choose $\varepsilon = 0.25$, $\omega_0 = 1$, and $\zeta = \frac{1}{2}$.

The design (6.118) does *not* meet the specification (6.119) with V given by (6.124) for the range of perturbations (6.123). Figure 6.20 shows plots of the bound $1/|V|$ together with plots of the perturbed sensitivity functions S for the four extreme combinations of parameter values. In particular, the bound is violated when the value of the gain g drops to too small values.

6.9.2. Setting up the problem

To set up the μ -synthesis problem we first define the perturbation model. As seen in Example 5.7.4 the loop gain has proportional perturbations

$$\Delta_P(s) = \frac{P(s) - P_0(s)}{P_0(s)} = \frac{\frac{g-g_0}{g_0} - s\theta}{1 + s\theta}. \quad (6.125)$$

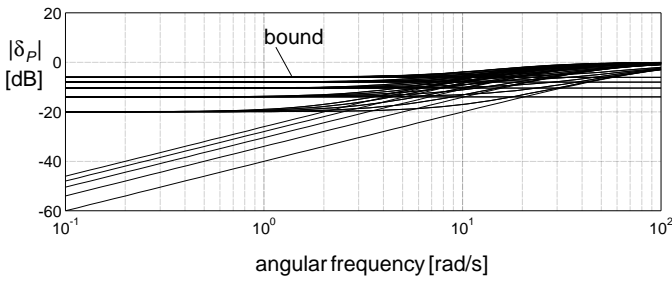


Figure 6.21: Magnitude plots of the perturbations δ_P for different parameter values and of the bound W_0

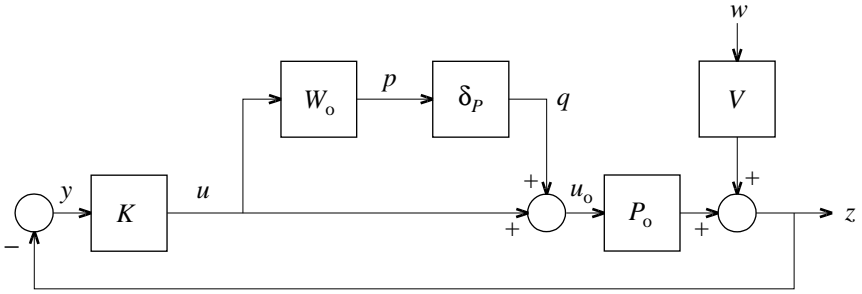


Figure 6.22: Perturbation and performance model

We bound the perturbations as $|\Delta_P(j\omega)| \leq |W_0(j\omega)|$ for all $\omega \in \mathbb{R}$, with

$$W_0(s) = \frac{\alpha + s\theta_0}{1 + s\theta_0}. \quad (6.126)$$

The number α , with $\alpha < 1$, is a bound $|g - g_0|/g_0 \leq \alpha$ for the relative perturbations in g , and θ_0 is the largest possible value for the parasitic time constant θ . Numerically we let $\alpha = 0.5$ and $\theta_0 = 0.05$. Figure 6.21 shows plots of the perturbation $|\delta_P|$ for various combinations of values of the uncertain parameters together with a magnitude plot of W_0 . Figure 6.22 shows the corresponding scaled perturbation model. It also includes the scaling function V used to normalize performance. The next step in preparing for D - K iteration is to compute the interconnection matrix G as in Fig. 6.18 from the interconnection equations

$$\begin{bmatrix} p \\ z \\ y \end{bmatrix} = G \begin{bmatrix} q \\ w \\ u \end{bmatrix}. \quad (6.127)$$

Inspection of the block diagram of Fig. 6.22 shows that

$$\begin{aligned} p &= W_0 u, \\ z &= V w + P_0(q + u), \\ y &= -V w - P_0(q + u), \end{aligned} \quad (6.128)$$

so that

$$\begin{bmatrix} p \\ z \\ y \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & 0 & W_0 \\ P_0 & W_1 & P_0 \\ -P_0 & -W_1 & -P_0 \end{bmatrix}}_G \begin{bmatrix} q \\ w \\ u \end{bmatrix}. \quad (6.129)$$

To apply the state space algorithm for the solution of the \mathcal{H}_∞ problem we need a state space representation of G . For the output z we have from the block diagram of Fig. 6.22

$$z = Vw + P_0 u_0 = \frac{1}{1+\varepsilon} \frac{s^2 + 2\zeta\omega_0 s + \omega_0^2}{s^2} w + \frac{g_0}{s^2} u_0. \quad (6.130)$$

This transfer function representation may be realized by the two-dimensional state space system

$$\begin{aligned} \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} &= \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} \frac{2\zeta\omega_0}{1+\varepsilon} & 0 \\ \frac{\omega_0^2}{1+\varepsilon} & g_0 \end{bmatrix} \begin{bmatrix} w \\ u_0 \end{bmatrix}, \\ z &= \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \frac{1}{1+\varepsilon} w. \end{aligned} \quad (6.131)$$

The weighting filter W_0 may be realized as

$$\begin{aligned} \dot{x}_3 &= -\frac{1}{\theta_0} x_3 + \frac{\alpha - 1}{\theta_0} u, \\ p &= u + x_3. \end{aligned} \quad (6.132)$$

Using the interconnection equation $u_0 = q + u$ we obtain from (6.131–6.132) the overall state differential and output equations

$$\begin{aligned} \dot{x} &= \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -\frac{1}{\theta_0} \end{bmatrix} x + \begin{bmatrix} 0 & \frac{2\zeta\omega_0}{1+\varepsilon} & 0 \\ g_0 & \frac{\omega_0^2}{1+\varepsilon} & g_0 \\ 0 & 0 & \frac{\alpha-1}{\theta_0} \end{bmatrix} \begin{bmatrix} q \\ w \\ u \end{bmatrix}, \\ \begin{bmatrix} p \\ z \\ y \end{bmatrix} &= \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix} x + \begin{bmatrix} 0 & 0 & 1 \\ 0 & \frac{1}{1+\varepsilon} & 0 \\ 0 & -\frac{1}{1+\varepsilon} & 0 \end{bmatrix} \begin{bmatrix} q \\ w \\ u \end{bmatrix}. \end{aligned} \quad (6.133)$$

To initialize the D – K iteration we select the controller that was obtained in Example 5.2.2 (p. 185) with transfer function

$$K_0(s) = \frac{k + sT_d}{1 + sT_0}. \quad (6.134)$$

This controller has the state space representation

$$\begin{aligned} \dot{x} &= -\frac{1}{T_0} x + \frac{1}{T_0} y, \\ u &= (k - \frac{T_d}{T_0}) x + \frac{T_d}{T_0} y. \end{aligned} \quad (6.135)$$

6.9.3. D – K iteration

The D – K iteration procedure is initialized by defining a frequency axis, calculating the frequency response of the initial closed-loop system H_0 on this axis, and computing and plotting the lower and upper bounds of the structured singular value of H_0 on this frequency axis⁷. The calculation of the upper bound also yields the “ D -scales.”

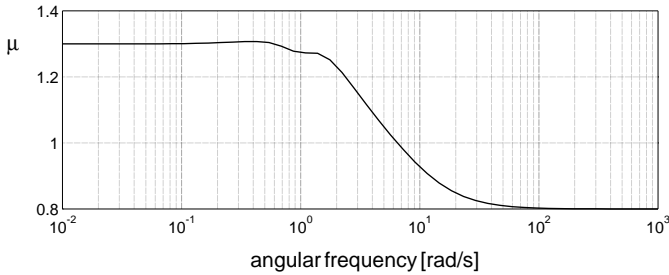


Figure 6.23: The structured singular value of the initial design

The plot of Fig. 6.23 shows that the structured singular value peaks to a value of about 1.3. Note that the computed lower and upper bounds coincide to the extent that they are indistinguishable in the plot. Since the peak value exceeds 1, the closed-loop system does not have robust performance, as we already know.

The next step is to fit rational functions to the D -scales. A quite good-looking fit may be obtained on the first diagonal entry of D with a transfer function of order 2. The second diagonal entry is normalized to be equal to 1, and, hence, does not need to be fitted. Figure 6.24 shows the calculated and fitted scales for the first diagonal entry. Because we have 1×1 perturbation blocks only, the left and right scales are equal.

The following step in the μ -synthesis procedure is to perform an \mathcal{H}_∞ optimization on the scaled system $G_1 = DGD^{-1}$. The result is a sub-optimal solution corresponding to the bound $\gamma_1 = 1.01$.

In the search procedure for the \mathcal{H}_∞ -optimal solution the search is terminated when the optimal value of the search parameter γ has been reached within a prespecified tolerance. This tolerance

⁷The calculations were done with the MATLAB μ -tools toolbox. The manual (Balas et al., 1991) offers a step-by-step explanation of the iterative process.

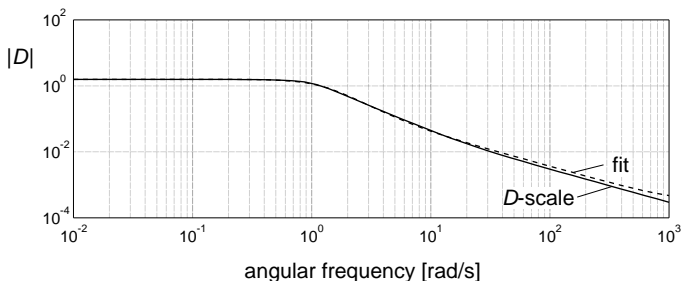


Figure 6.24: Calculated and fitted D -scales

should be chosen with care. Taking it too small not only prolongs the computation but — more importantly — results in a solution that is close to but not equal to the actual \mathcal{H}_∞ optimal solution. As seen in § 6.6 (p. 205), solutions that are close to optimal often have large coefficients. These large coefficients make the subsequent calculations, in particular that of the frequency response of the optimal closed-loop system, ill-conditioned, and easily lead to erroneous results. This difficulty is a weak point of algorithms for the solution of the \mathcal{H}_∞ problem that do not provide the actual optimal solution.

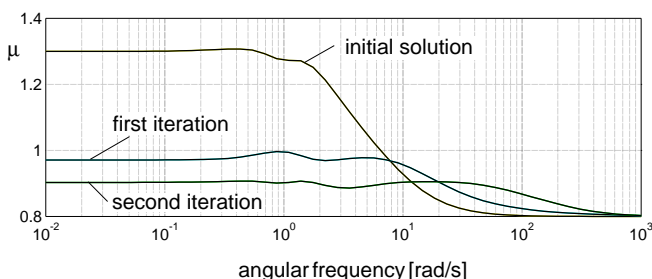


Figure 6.25: Structured singular values for the three successive designs

Figure 6.25 shows the structured singular value of the closed-loop system that corresponds to the controller that has been obtained after the first iteration. The peak value of the structured singular value is about 0.97. This means that the design achieves robust stability.

To increase the robustness margin another D – K iteration may be performed. Again the D -scale may be fitted with a second-order rational function. The \mathcal{H}_∞ optimization results in a suboptimal solution with level $\gamma_2 = 0.91$.

Figure 6.25 shows the structured singular value plots of the three designs we now have. For the third design the structured value has a peak value of about 0.9, well below the critical value 1.

The plot for the structured singular value for the final design is quite flat. This appears to be typical for minimum- μ designs (Lin et al., 1993).

6.9.4. Assessment of the solution

The controller K_2 achieves a peak structured singular value of 0.9. It therefore has robust performance with a good margin. The margin may be improved by further D – K iterations. We pause to analyze the solution that has been obtained.

Figure 6.26 shows plots of the sensitivity function of the closed-loop system with the controller K_2 for the four extreme combinations of the values of the uncertain parameters g and θ . The plots confirm that the system has robust performance.

Figure 6.27 gives plots of the nominal system functions S and U . The input sensitivity function U increases to quite a large value, and has no high-frequency roll-off, at least not in the frequency region shown. The plot shows that robust performance is obtained at the cost of a large controller gain-bandwidth product. The reason for this is that the design procedure has no explicit or implicit provision that restrains the bandwidth.

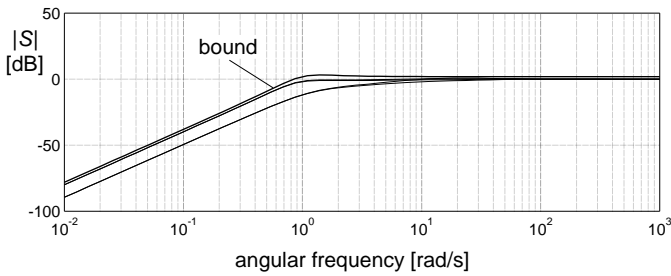


Figure 6.26: Perturbed sensitivity functions for K_2 and the bound $1/V$

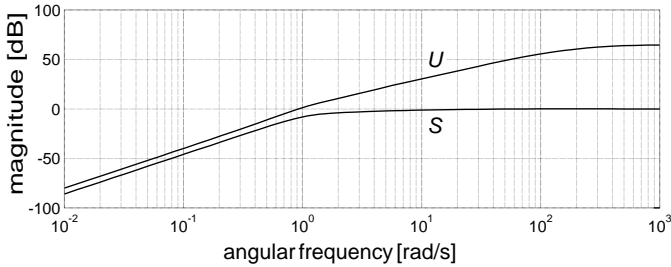


Figure 6.27: Nominal sensitivity S and input sensitivity U for the controller K_2

6.9.5. Bandwidth limitation

We revise the problem formulation to limit the bandwidth explicitly. One way of doing this is to impose bounds on the high-frequency behavior of the input sensitivity function U . This, in turn, may be done by bounding the high-frequency behavior of the weighted input sensitivity UV . Figure 6.28 shows a magnitude plot of UV for the controller K_2 . We wish to modify the design so that UV is bounded as

$$|U(j\omega)V(j\omega)| \leq \frac{1}{|W_2(j\omega)|}, \quad \omega \in \mathbb{R}, \quad (6.136)$$

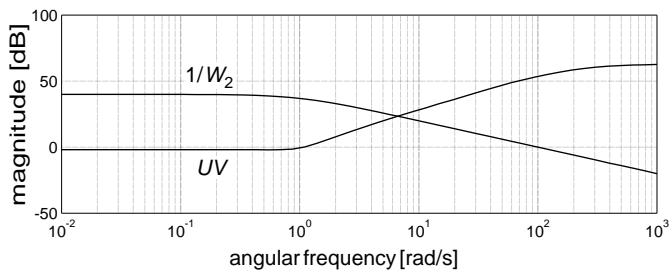


Figure 6.28: Magnitude plots of UV and the bound $1/W_2$

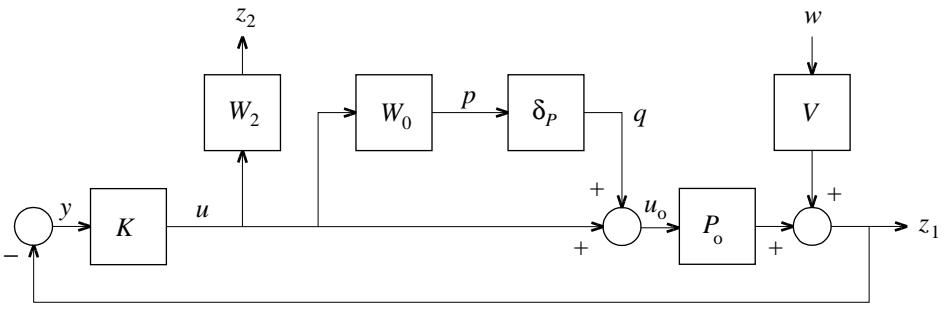


Figure 6.29: Modified block diagram for extra performance specification

where

$$\frac{1}{W_2(s)} = \frac{\beta}{s + \omega_1}. \quad (6.137)$$

Numerically, we tentatively choose $\beta = 100$ and $\omega_1 = 1$. Figure 6.28 includes the corresponding bound $1/|W_2|$ on $|U|$.

To include this extra performance specification we modify the block diagram of Fig. 6.22 to that of Fig. 6.29. The diagram includes an additional output z_2 while z has been renamed to z_1 . The closed-loop transfer function from w to z_2 is $-W_2UV$. To handle the performance specifications $\|SV\|_\infty \leq 1$ and $\|W_2UV\|_\infty \leq 1$ jointly we impose the performance specification

$$\left\| \begin{bmatrix} W_1SV \\ W_2UV \end{bmatrix} \right\|_\infty \leq 1, \quad (6.138)$$

where $W_1 = 1$. This is the familiar performance specification of the mixed sensitivity problem.

By inspection of Fig. 6.29 we see that the structured perturbation model used to design for robust performance now is

$$\begin{bmatrix} q \\ w \end{bmatrix} = \underbrace{\begin{bmatrix} \delta_P & 0 & 0 \\ 0 & \delta_1 & \delta_2 \end{bmatrix}}_{\Delta} \begin{bmatrix} p \\ z_1 \\ z_2 \end{bmatrix}, \quad (6.139)$$

with δ_P , δ_1 and δ_2 independent scalar complex perturbations such that $\|\Delta\|_\infty \leq 1$.

Because W_2 is a nonproper rational function we modify the diagram to the equivalent diagram of Fig. 6.30 (compare § 6.7, p. 258). The controller K and the filter W_2 are combined to a new controller $\tilde{K} = KW_2$, and any nonproper transfer functions now have been eliminated. It is easy

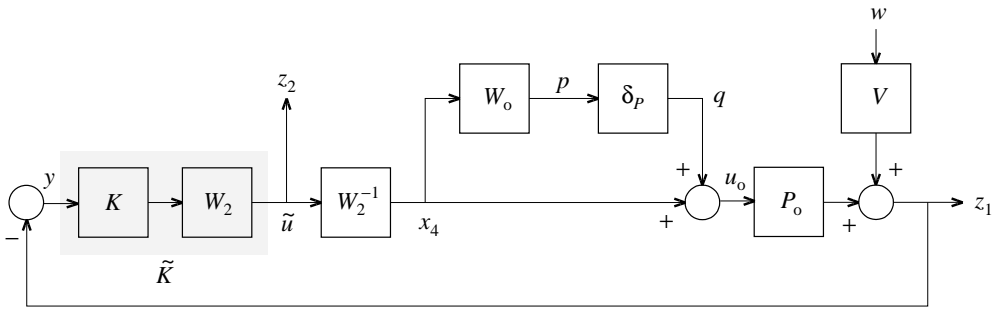


Figure 6.30: Revised modified block diagram

to check that the open-loop interconnection system according to Fig. 6.30 may be represented as

$$\dot{x} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & g_0 \\ 0 & 0 & -\frac{1}{\theta_0} & \frac{\alpha-1}{\theta_0} \\ 0 & 0 & 0 & -\omega_1 \end{bmatrix} x + \begin{bmatrix} 0 & \frac{2\zeta\omega_0}{1+\varepsilon} & 0 \\ g_0 & \frac{\omega_0^2}{1+\varepsilon} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \beta \end{bmatrix} \begin{bmatrix} q \\ w \\ u \end{bmatrix}, \quad (6.140)$$

$$\begin{bmatrix} p \\ z_1 \\ z_2 \\ y \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{bmatrix} x + \begin{bmatrix} 0 & 0 & 0 \\ 0 & \frac{1}{1+\varepsilon} & 0 \\ 0 & 0 & 1 \\ 0 & -\frac{1}{1+\varepsilon} & 0 \end{bmatrix} \begin{bmatrix} q \\ w \\ \tilde{u} \end{bmatrix}.$$

To let the structured perturbation Δ have square diagonal blocks we rename the external input w to w_1 and expand the interconnection system with a void additional external input w_2 :

$$\dot{x} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & g_0 \\ 0 & 0 & -\frac{1}{\theta_0} & \frac{\alpha-1}{\theta_0} \\ 0 & 0 & 0 & -\omega_1 \end{bmatrix} x + \begin{bmatrix} 0 & \frac{2\zeta\omega_0}{1+\varepsilon} & 0 & 0 \\ g_0 & \frac{\omega_0^2}{1+\varepsilon} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \beta \end{bmatrix} \begin{bmatrix} q \\ w_1 \\ w_2 \\ \tilde{u} \end{bmatrix}, \quad (6.141)$$

$$\begin{bmatrix} p \\ z_1 \\ z_2 \\ y \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{bmatrix} x + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{1}{1+\varepsilon} & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & -\frac{1}{1+\varepsilon} & 0 & 0 \end{bmatrix} \begin{bmatrix} q \\ w_1 \\ w_2 \\ \tilde{u} \end{bmatrix}.$$

The perturbation model now is

$$\begin{bmatrix} q \\ w \end{bmatrix} = \begin{bmatrix} \delta_P & 0 \\ 0 & \Delta_0 \end{bmatrix} \begin{bmatrix} p \\ z \end{bmatrix}. \quad (6.142)$$

with δ_P a scalar block and Δ_0 a full 2×2 block.

To obtain an initial design we do an \mathcal{H}_∞ -optimal design on the system with q and p removed, that is, on the nominal, unperturbed system. This amounts to the solution of a mixed sensitivity

problem. Nine iterations are needed to find a suboptimal solution according to the level $\gamma_0 = 0.97$. The latter number is less than 1, which means that nominal performance is achieved. The peak value of the structured singular value turns out to be 1.78, however, so that we do not have robust performance.

One D - K iteration leads to a design with a reduced peak value of the structured singular value of 1.32 (see Fig. 6.31). Further reduction does not seem possible. This means that the bandwidth constraint is too severe.

To relax the bandwidth constraint we change the value of β in the bound $1/W_2$ as given by (6.137) from 100 to 1000. Starting with \mathcal{H}_∞ -optimal initial design one D - K iteration leads to a design with a peak structured singular value of 1.1. Again, robust performance is not feasible.

Finally, after choosing $\beta = 10000$ the same procedure results after one D - K iteration in a controller with a peak structured singular value of 1.02. This controller very nearly has robust performance with the required bandwidth.

Figure 6.31 shows the structured singular value plots for the three controllers that are successively obtained for $\beta = 100$, $\beta = 1000$ and $\beta = 10000$.

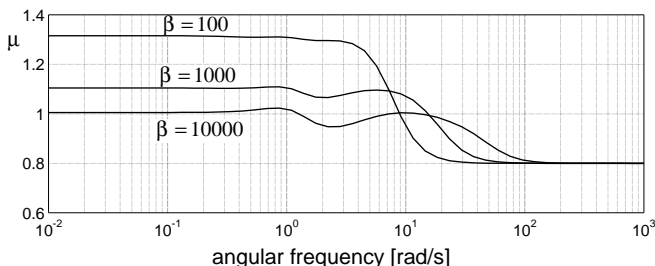


Figure 6.31: Structured singular value plots for three successive designs

6.9.6. Order reduction of the final design

The limited bandwidth robust controller has order 9. This number may be explained as follows. The generalized plant (6.141) has order 4. In the (single) D - K iteration that yields the controller the “ D -scale” is fitted by a rational function of order 2. Premultiplying the plant by D and postmultiplying by its inverse increases the plant order to $4 + 2 + 2 = 8$. Central suboptimal controllers \tilde{K} for this plant hence also have order 8. This means that the corresponding controllers $K = \tilde{K}W_2^{-1}$ for the configuration of Fig. 6.29 have order 9.

It is typical for the D - K algorithm that controllers of high order are obtained. This is caused by the process of fitting rational scales. Often the order of the controller may considerably be decreased without affecting performance or robustness.

Figure 6.32(a) shows the Bode magnitude and phase plots of the limited bandwidth robust controller. The controller is minimum-phase and has pole excess 2. Its Bode plot has break points near the frequencies 1, 60 and 1000 rad/s. We construct a simplified controller in two steps:

1. The break point at 1000 rad/s is caused by a large pole. We remove this pole by omitting the leading term s^9 from the denominator of the controller. This large pole corresponds to a pole at ∞ of the \mathcal{H}_∞ -optimal controller. Figure 6.32(b) is the Bode plot of the resulting simplified controller.

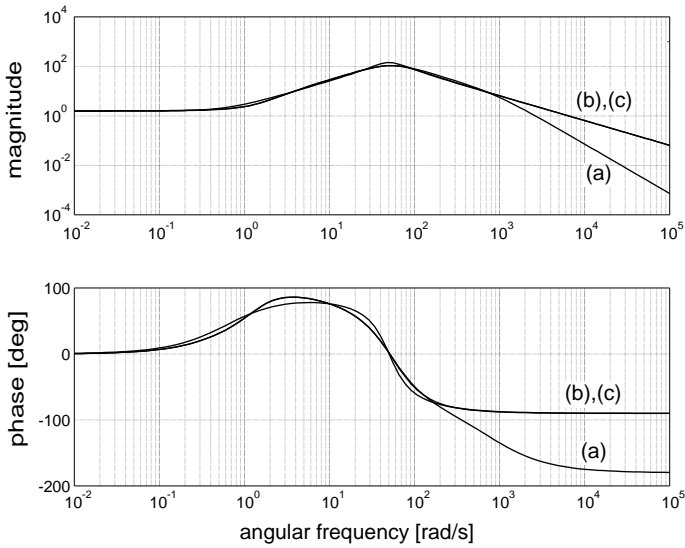


Figure 6.32: Exact (a) and approximate (b, c) frequency responses of the controller

2. The Bode plot that is now found shows that the controller has pole excess 1. It has break points at 1 rad/s (corresponding to a single zero) and at 60 rad/s (corresponding to a double pole). We therefore attempt to approximate the controller transfer function by a rational function with a single zero and two poles. A suitable numerical routine from MATLAB yields the approximation

$$K(s) = 6420 \frac{s + 0.6234}{(s + 22.43)^2 + 45.31^2}. \quad (6.143)$$

Figure 6.32(c) shows that the approximation is quite good.

Figure 6.33 gives the magnitudes of the perturbed sensitivity functions corresponding to the four extreme combinations of parameter values that are obtained for this reduced-order controller. The plots confirm that performance is very nearly robust.

Figure 6.34 displays the nominal performance. Comparison of the plot of the input sensitivity U with that of Fig. 6.27 confirms that the bandwidth has been reduced.

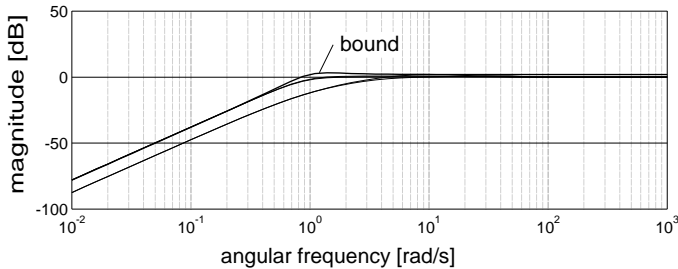


Figure 6.33: Perturbed sensitivity functions of the reduced-order limited-bandwidth design

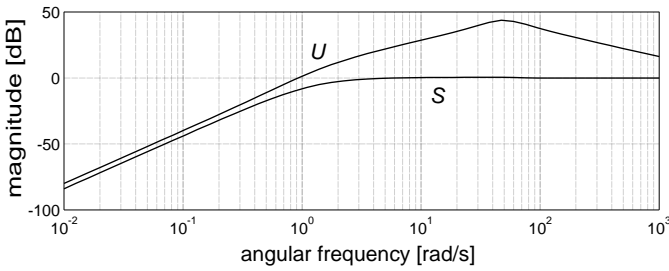


Figure 6.34: Nominal sensitivity S and input sensitivity U of the reduced-order limited- bandwidth design

6.9.7. Conclusion

The example illustrates that μ -synthesis makes it possible to design for robust performance while explicitly and quantitatively accounting for plant uncertainty. The method cannot be used naïvely, though, and considerable insight is needed.

A severe shortcoming is that the design method inherently leads to controllers of high order, often unnecessarily high. At this time only *ad hoc* methods are available to decrease the order. The usual approach is to apply an order reduction algorithm to the controller.

6.10. Appendix: Proofs

Proof of Lemma 6.4.1. By contradiction: if $\|H\|_\infty < |H(z_0)|$ then $\|H(s)/H(z_0)\|_\infty < 1$ hence by the small gain theorem $(I - H(s)/H(z_0))^{-1}$ is stable, but $I - H(s)/H(z_0)$ has a zero at $s = z_0$ so $(I - H(s)/H(z_0))^{-1}$ is not stable. Contradiction, hence $\|H\|_\infty \geq |H(z_0)|$. ■

Proof of Summary 6.4.2. By the definition of the ∞ -norm the inequality $\|H\|_\infty \leq \gamma$ is equivalent to the condition that $\bar{\sigma}(H^*(j\omega)H(j\omega)) \leq \gamma$ for $\omega \in \mathbb{R}$, with $\bar{\sigma}$ the largest singular value. This in turn is the same as the condition that $\lambda_i(H^*H) \leq \gamma^2$ on the imaginary axis for all i , with λ_i the i th largest eigenvalue. This finally amounts to the condition that $H^*H \leq \gamma^2 I$ on the imaginary axis, which proves Summary 6.4.2. ■

6.11. Exercises

6.1 *Closed-loop transfer matrix.* Verify (6.37).

6.2 *Minimum sensitivity problem.* A special case of the mixed sensitivity problem is the minimization of the infinity norm $\|WSV\|_\infty$ of the weighted sensitivity matrix S for the closed-loop system of Fig. 6.1. Show that this minimum sensitivity problem is a standard problem with open-loop plant

$$G = \begin{bmatrix} WV & WP \\ -V & -P \end{bmatrix}. \quad (6.144)$$

The SISO version of this problem historically is the first \mathcal{H}_∞ optimization problem that was studied. It was solved in a famous paper by Zames (1981).

6.3 *The model matching problem.* The model matching problem consists of finding a stable transfer matrix K that minimizes

$$\|P - K\|_{\mathcal{L}_\infty} := \sup_{\omega} \bar{\sigma}(P(j\omega) - K(j\omega)) \quad (6.145)$$

with P a given (unstable) transfer matrix. Is this a standard \mathcal{H}_∞ problem? (The problem is known as the (generalized) *Nehari* problem).

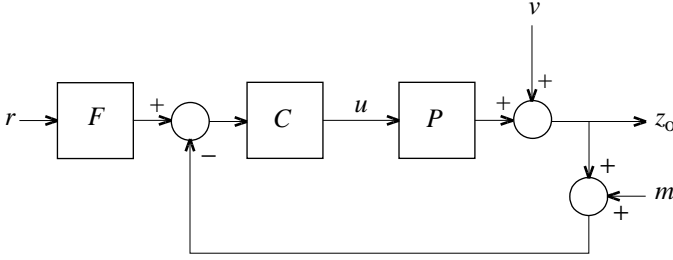


Figure 6.35: Two-degree-of-freedom feedback configuration

6.4 *Two-degree-of-freedom feedback system.* As a further application consider the two-degree-of-freedom configuration of Fig. 6.35. In Fig. 6.36 the diagram is expanded with a shaping filter V_1 that generates the disturbance v from the driving signal w_1 , a shaping filter V_2 that generates the measurement noise $m = V_2 w_2$ from the driving signal w_2 , a shaping filter V_0 that generates the reference signal r from the driving signal w_3 , a weighting filter W_1 that produces the weighted tracking error $z_1 = W_1(z_0 - r)$, and a weighting filter W_2 that produces the weighted plant input $z_2 = W_2 u$. Moreover, the controller C and the prefilter F are combined into a single block K . Define the “control error” z with components z_1 and

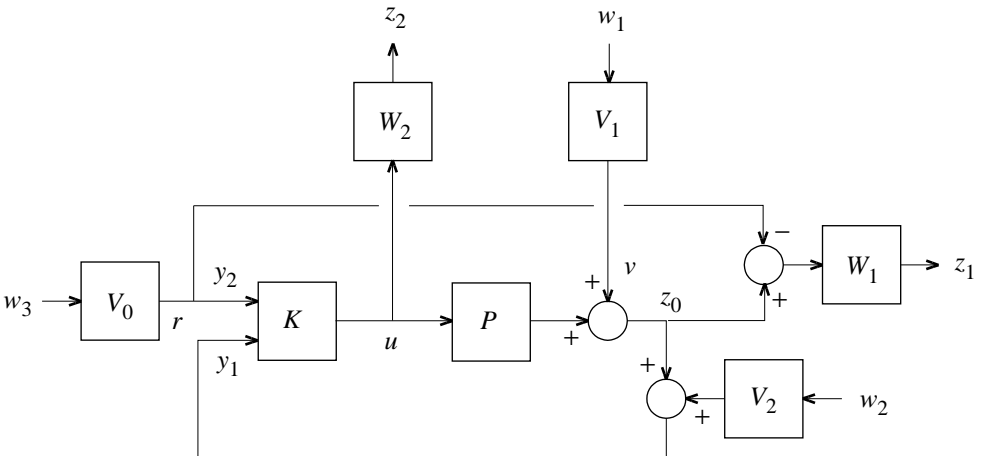


Figure 6.36: Block diagram for a standard problem

z_2 , the observed output y with components y_1 and y_2 as indicated in the block diagram,

the external input w with components w_1 , w_2 , and w_3 , and the control input u . Determine the open-loop generalized plant transfer matrix G and the closed-loop transfer matrix H .

6.5 *Mixed sensitivity problem.* Show that for the mixed sensitivity problem “stability of the loop around G_{22} ” is equivalent to stability of the feedback loop. Discuss what additional requirements are needed to ensure that the entire closed-loop system be stable.

6.6 *Filtering problem.* Show that the filtering problem discussed in § 6.4 is a standard \mathcal{H}_∞ problem.

6.7 *Mixed sensitivity problem.* Consider the unstable plant with transfer function

$$P(s) = \frac{1 - \frac{1}{4}s}{(1 - 10s)^2}.$$

- a) We aim for a closed loop bandwidth of about 1, and we want $S(0) = 0$ for constant disturbance rejection. Further we want sufficient high frequency roll-off of the controller, and that $|S|$ and $|T|$ do not peak much.

Use the mixed sensitivity approach of § 6.2 to design a suitable controller (see also § 6.7). Explain your choices of W_1 and W_2 and M and plot $|S|$ and $|T|$ and step response for the more successful choices.

- b) S or T designed above peak a bit. Is this inherent to the problem or a limitation of the mixed sensitivity design method? (Explain.)

6.8 *Disturbance feedforward.*

- a) Verify that the lower bound γ_0 is given by (6.62).
- b) Prove that ρ_c as given by (6.63) separates the solutions of type A and type B.
- c) Prove that if $\rho < \rho_c$ then the minimal ∞ -norm γ_{opt} is the positive solution of the equation $\gamma^4 + 2\gamma^3 = 1/\rho^2$ (Hagander and Bernhardsson, 1992).
- d) Check the frequency dependence of the largest singular value of the closed-loop frequency response matrix for the two types of solutions.
- e) What do you have to say about the optimal solution for $\rho = \rho_c$?

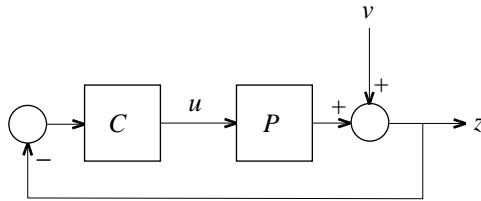


Figure 6.37: One-degree-of- freedom system with disturbance v

6.9 *Completion of the design.* Complete the design of this example.

- (a) For the controller (6.102), compute and plot the sensitivity function S , the input sensitivity function U , and the complementary sensitivity function T .

- (b) Check whether these functions behave satisfactorily, in particular, if S and T do not peak too much. See if these functions may be improved by changing the values of c and r by trial and error.
- (c) The roll-off of 2 decades/decade is more than called for by the design specifications. The desired roll-off of 1 decade/decade may be obtained with the constant weighting function $W_2(s) = c$. Recompute the optimal controller for this weighting function with $c = 1$.
- (d) Repeat (a) and (b) for the design of (c).
- (e) Compute and plot the time responses of the output z and the plant input u to a unit step disturbance v in the configuration of Fig. 6.37 for the designs selected in (b) and (d). Discuss the results.
- 6.10 *Roll-off in the double integrator example.* In Example 6.6.1 (p. 256) the numerical solution of the double integrator example of § 6.2.4 (p. 244) is presented for $r = 0$. Extend this to the case $r \neq 0$, and verify (6.32).
- 6.11 *Proof of inequality.* In § 6.8.1 we needed the inequality $(a + b)^2 \leq 2(a^2 + b^2)$ for any two real numbers a and b . Prove this.
- 6.12 *Minimization of upper bound.* Show that the upper bound on the right-hand side of (6.108) may also be obtained from Summary 5.6.5(2) (p. 220).
- 6.13 *Reduction to mixed sensitivity problem.* In § 6.8.1 (p. 267) it is argued that robust performance such that $\|SV\|_\infty < 1$ under proportional plant perturbations bounded by $\|\Delta_P\|_\infty \leq \|W_0\|_\infty$ is achieved iff

$$\sup_{\omega \in \mathbb{R}} (|S(j\omega)V(j\omega)| + |T(j\omega)W_0(j\omega)|) < 1. \quad (6.146)$$

A sufficient condition for this is that

$$\sup_{\omega \in \mathbb{R}} (|S(j\omega)V(j\omega)|^2 + |T(j\omega)W_0(j\omega)|^2) < \frac{1}{2}\sqrt{2}. \quad (6.147)$$

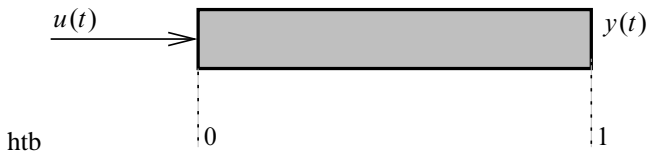
To see whether the latter condition may be satisfied we consider the mixed sensitivity problem consisting of the minimization of $\|H\|_\infty$, with

$$H = \begin{bmatrix} SV \\ -TW_0 \end{bmatrix}. \quad (6.148)$$

- (a) Show that the mixed sensitivity problem at hand may be solved as a standard \mathcal{H}_∞ problem with generalized plant

$$G = \left[\begin{array}{c|c} V & P_0 \\ 0 & W_0 V^{-1} P_0 \\ \hline -V & -P_0 \end{array} \right]. \quad (6.149)$$

- (b) Find a (minimal) state representation of G . (*Hint:* Compare (6.131) and (6.132).) Check that Assumption 4 of § 6.5 (p. 253) needed for the state space solution of the \mathcal{H}_∞ problem is not satisfied. To remedy this, modify the problem by adding a third component $z_1 = \rho u$ to the error signal z , with ρ small.



- (c) Solve the \mathcal{H}_∞ -optimization problem for a small value of ρ , say, $\rho = 10^{-6}$ or even smaller. Use the numerical values introduced earlier: $g_0 = 1$, $\varepsilon = 0.25$, $\omega_0 = 1$, $\zeta = 0.5$, $\alpha = 0.5$, and $\theta_0 = 0.05$. Check that $\|H\|_\infty$ cannot be made less than about 0.8, and, hence, the condition (6.147) cannot be met.
- (d) Verify that the solution of the mixed sensitivity problem does not satisfy (6.146). Also check that the solution does not achieve robust performance for the real parameter variations (6.123).

The fact that a robust design cannot be obtained this way does not mean that a robust design does not exist.

- 6.14 *State space representation.* Derive or prove the state space representations (6.131) and (6.132).
- 6.15 *Peak in the structured singular value plot.* Figure 6.31 shows that the singular value plot for the limited-bandwidth design has a small peak of magnitude slightly greater than 1 near the frequency 1 rad/s. Inspection of Fig. 6.33, however, reveals no violation of the performance specification on S near this frequency. What is the explanation for this peak?
- 6.16 *Comparison with other designs.* The controller (6.143) is not very different from the controllers found in Example 5.2.2 (p. 185) by the root locus method, by \mathcal{H}_2 optimization in § 4.6.2 (p. 161), and by solution of a mixed sensitivity problem in § 6.2.4 (p. 244).
 - a) Compare the four designs by computing their closed-loop poles and plotting the sensitivity and complementary sensitivity functions in one graph.
 - b) Compute the stability regions for the four designs and plot them in one graph as in Fig. 6.5.
 - c) Discuss the comparative robustness and performance of the four designs.
- 6.17 *μ -synthesis for an infinite dimensional system.* We are given a metal beam of one meter of which we can control the temperature u at the left end, and of which we can measure the temperature y at the right end (see the figure below). Assuming perfect isolation, it may be shown that the transfer function from u to y is

$$P(s) = \frac{1}{\cosh(\sqrt{s/a})} = \frac{1}{1 + \frac{1}{2!}(s/a) + \frac{1}{4!}(s/a)^2 + \frac{1}{6!}(s/a)^3 + \dots},$$

where $a > 0$ is the diffusion constant of the metal. It is a stable system. For copper, a equals $a = 1.16 \cdot 10^{-4}$ so the time constant $1/a$ is large, and heating hence may take a long time.

- a) Make the Bode magnitude plot of this (nonrational) P .

- b) The problem is to bring the temperature y of the right end of the beam to a given desired constant temperature T_{desired} by choice of u .

Formulate this problem as a μ -synthesis problem. (*You do not need to solve it with Matlab*). Your solution method should at least address the following.

- How to model the nonrational P as a perturbation of an appropriate ‘nominal’ rational P_0 ;
- How to ensure convergence of $y(t)$ to T_{desired} , and that the convergence is “reasonably” fast;
- What are the perturbations Δ , Δ_0 and generalized plant G as in Fig. 6.18 (page 270);
- What to do if D - K iteration does not result in $\mu_H := \sup_{\omega} \mu(H(j\omega)) < 1$?

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 i th row and j th 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 i th 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*

$$\chi_A(\lambda) = \det(\lambda I_n - A), \quad (\lambda \in \mathbb{C},)$$

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

$$A = VDV^{-1}, \quad \text{where } V \text{ and } D \text{ are square and } D \text{ is diagonal.}$$

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*, $\text{tr}(A)$ of a square matrix $A \in \mathbb{C}^{n \times n}$ is defined as $\text{tr}(A) = \sum_{i=1}^n A_{ii}$. It may be shown that

$$\text{tr}(A) = \sum_{i=1}^n \lambda_i(A).$$

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 $A^T A$ which in turn equals the number of nonzero eigenvalues of $A^T A$.

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 i th row and j th 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 $A^T = A$. Here A^T is the *transpose* of A is defined elementwise as $(A^T)_{ij} = A_{ji}$, $(i, j = 1, \dots, n)$.

A matrix $A \in \mathbb{C}^{n \times n}$ is *Hermitian* if $A^H = A$. Here A^H is the complex conjugate transpose of A defined as $(A^H)_{ij} = \overline{A_{ji}}$ $(i, j = 1, \dots, n)$. Overbars $\overline{x + jy}$ of a complex number $x + jy$ denote the complex conjugate: $\overline{x + jy} = x - jy$.

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: $V^H V = I$.

A symmetric or Hermitian matrix A is said to be *nonnegative definite* or *positive semi-definite* if $x^H A x \geq 0$ 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 $x^H A x > 0$ 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 (Nonnegative definite matrices). Let $A \in \mathbb{C}^{n \times n}$ be a Hermitian matrix. Then

1. All eigenvalues of A are real valued,
2. $A \geq 0 \iff \lambda_i(A) \geq 0 \quad (\forall i = 1, \dots, n)$,
3. $A > 0 \iff \lambda_i(A) > 0 \quad (\forall i = 1, \dots, n)$,
4. If T is nonsingular then $A \geq 0$ if and only $T^H A T \geq 0$.

□

A.2. Three matrix lemmas

Lemma A.2.1 (Eigenvalues of matrix products). Suppose A and B^H are matrices of the same dimension $n \times m$. Then for any $\lambda \in \mathbb{C}$ there holds

$$\det(\lambda I_n - AB) = \lambda^{n-m} \det(\lambda I_m - BA). \quad (\text{A.1})$$

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. $\text{tr}(AB) = \sum_i \lambda_i(AB) = \sum_j \lambda_j(BA) = \text{tr}(BA)$.

Lemma A.2.2 (Sherman-Morrison-Woodbury & 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,

$$(A + uv^H)^{-1} = A^{-1} - \underbrace{\frac{1}{1 + v^H A^{-1}u}}_{\text{rank-one}} (A^{-1}u)(v^H A^{-1}).$$

□

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).

□

B. Norms of signals and systems

For a SISO system with transfer function H the interpretation of $|H(j\omega)|$ as a “gain” from input to output is clear, but how may we define “gain” for a MIMO system with a transfer *matrix* H ? One way to do this is to use norms $\|H(j\omega)\|$ instead of absolute values. There are many different norms and in this appendix we review the most common norms for both signals and systems. The theory of norms leads to a re-assessment of stability.

B.1. Norms of vector-valued signals

We consider continuous-time signals z defined on the time axis \mathbb{R} . These signals may be scalar-valued (with values in \mathbb{R} or \mathbb{C}) or vector-valued (with values in \mathbb{R}^n or \mathbb{C}^n). They may be added and multiplied by real or complex numbers and, hence, are elements of what is known as a *vector space*.

Given such a signal z , its *norm* is a mathematically well-defined notion that is a measure for the “size” of the signal.

Definition B.1.1 (Norms). Let X be a vector space over the real or complex numbers. Then a function

$$\|\cdot\| : X \rightarrow \mathbb{R} \quad (\text{B.1})$$

that maps X into the real numbers \mathbb{R} is a *norm* if it satisfies the following properties:

1. $\|x\| \geq 0$ for all $x \in X$ (*nonnegativity*),
2. $\|x\| = 0$ if and only if $x = 0$ (*positive-definiteness*),
3. $\|\lambda x\| = |\lambda| \cdot \|x\|$ for every scalar λ and all $x \in X$ (*homogeneity with respect to scaling*),
4. $\|x + y\| \leq \|x\| + \|y\|$ for all $x \in X$ and $y \in X$ (*triangle inequality*).

The pair $(X, \|\cdot\|)$ is called a *normed* vector space. If it is clear which norm is used, X by itself is often called a normed vector space. \square

A well-known norm is the p -norm of vectors in \mathbb{C}^n .

Example B.1.2 (Norms of vectors in \mathbb{C}^n). Suppose that $x = (x_1, x_2, \dots, x_n)$ is an n -dimensional complex-valued vector, that is, x is an element of \mathbb{C}^n . Then for $1 \leq p \leq \infty$ the p -norm of x is defined as

$$\|x\|_p = \begin{cases} (\sum_{i=1}^n |x_i|^p)^{1/p} & \text{for } 1 \leq p < \infty, \\ \max_{i=1,2,\dots,n} |x_i| & \text{for } p = \infty. \end{cases} \quad (\text{B.2})$$

Well-known special cases are the norms

$$\|x\|_1 = \sum_{i=1}^n |x_i|, \quad \|x\|_2 = \left(\sum_{i=1}^n |x_i|^2 \right)^{1/2}, \quad \|x\|_\infty = \max_{i=1,2,\dots,n} |x_i|. \quad (\text{B.3})$$

$\|x\|_2$ is the familiar *Euclidean norm*. \square

B.2. Singular values of vectors and matrices

We review the notion of singular values of a matrix¹. If A is an $n \times m$ complex-valued matrix then the matrices $A^H A$ and $A A^H$ are both nonnegative-definite Hermitian. As a result, the eigenvalues of both $A^H A$ and $A A^H$ are all real and nonnegative. The $\min(n, m)$ largest eigenvalues

$$\lambda_i(A^H A), \quad \lambda_i(A A^H), \quad i = 1, 2, \dots, \min(n, m), \quad (\text{B.4})$$

of $A^H A$ and $A A^H$, respectively, ordered in decreasing magnitude, are equal. The remaining eigenvalues, if any, are zero. The square roots of these $\min(n, m)$ eigenvalues are called the *singular values* of the matrix A , and are denoted

$$\sigma_i(A) = \lambda_i^{1/2}(A^H A) = \lambda_i^{1/2}(A A^H), \quad i = 1, 2, \dots, \min(n, m). \quad (\text{B.5})$$

Obviously, they are nonnegative real numbers. The number of *nonzero* singular values equals the rank of A .

Summary B.2.1 (Singular value decomposition). Given $n \times m$ matrix A let Σ be the diagonal $n \times m$ matrix whose diagonal elements are $\sigma_i(A)$, $i = 1, 2, \dots, \min(n, m)$. Then there exist square unitary matrices U and V such that

$$A = U \Sigma V^H. \quad (\text{B.6})$$

A (complex-valued) matrix U is unitary if $U^H U = U U^H = I$, with I a unit matrix of correct dimensions. The representation (B.6) is known as the *singular value decomposition (SVD)* of the matrix A .

The largest and smallest singular value $\sigma_1(A)$ and $\sigma_{\min(n, m)}(A)$ respectively are commonly denoted by an overbar and an underbar

$$\overline{\sigma}(A) = \sigma_1(A), \quad \underline{\sigma}(A) = \sigma_{\min(n, m)}(A).$$

Moreover, the largest singular value $\overline{\sigma}(A)$ is a norm of A . It is known as the *spectral norm*. \square

There exist numerically reliable methods to compute the matrices U and V and the singular values. These methods are numerically more stable than any available for the computation of eigenvalues.

Example B.2.2 (Singular value decomposition). The singular value decomposition of the 3×1 matrix A given by

$$A = \begin{bmatrix} 0 \\ 3 \\ 4 \end{bmatrix} \quad (\text{B.7})$$

is $A = U \Sigma V^H$, where

$$U = \begin{bmatrix} 0 & -0.6 & -0.8 \\ 0.6 & 0.64 & -0.48 \\ 0.8 & -0.48 & 0.36 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 5 \\ 0 \\ 0 \end{bmatrix}, \quad V = 1. \quad (\text{B.8})$$

The matrix A has a single nonzero singular value (because it has rank 1), equal to 5. Hence, the spectral norm of A is $\overline{\sigma}_1(A) = 5$. \square

¹See for instance Section 10.8 of Noble (1969).

B.3. Norms of signals

The p -norm for constant vectors may easily be generalized to vector-valued signals.

Definition B.3.1 (\mathcal{L}_p -norm of a signal). For any $1 \leq p \leq \infty$ the p -norm or \mathcal{L}_p -norm $\|z\|_{\mathcal{L}_p}$ of a continuous-time scalar-valued signal z , is defined by

$$\|z\|_{\mathcal{L}_p} = \begin{cases} \left(\int_{-\infty}^{\infty} |z(t)|^p dt \right)^{1/p} & \text{for } 1 \leq p < \infty, \\ \sup_{t \in \mathbb{R}} |z(t)| & \text{for } p = \infty. \end{cases} \quad (\text{B.9})$$

If $z(t)$ is vector-valued with values in \mathbb{R}^n or \mathbb{C}^n , this definition is generalized to

$$\|z\|_{\mathcal{L}_p} = \begin{cases} \left(\int_{-\infty}^{\infty} \|z(t)\|^p dt \right)^{1/p} & \text{for } 1 \leq p < \infty, \\ \sup_{t \in \mathbb{R}} \|z(t)\| & \text{for } p = \infty, \end{cases} \quad (\text{B.10})$$

where $\|\cdot\|$ is any norm on the n -dimensional space \mathbb{R}^n or \mathbb{C}^n . □

The signal norms that we mostly need are the \mathcal{L}_2 -norm, defined as

$$\|z\|_{\mathcal{L}_2} = \left(\int_{-\infty}^{\infty} \|z(t)\|_2^2 dt \right)^{1/2} \quad (\text{B.11})$$

and the \mathcal{L}_∞ -norm, defined as

$$\|z\|_{\mathcal{L}_\infty} = \sup_{t \in \mathbb{R}} \|z(t)\|_\infty. \quad (\text{B.12})$$

The square $\|z\|_{\mathcal{L}_2}^2$ of the \mathcal{L}_2 -norm is often called the *energy* of the signal z , and the \mathcal{L}_∞ -norm $\|z\|_{\mathcal{L}_\infty}$ its *amplitude* or *peak value*.

Example B.3.2 (\mathcal{L}_2 and \mathcal{L}_∞ -norm). Consider the signal z with two entries

$$z(t) = \begin{bmatrix} e^{-t} \mathbb{1}(t) \\ 2e^{-3t} \mathbb{1}(t) \end{bmatrix}.$$

Here $\mathbb{1}(t)$ denotes the unit step, so $z(t)$ is zero for negative time. From $t = 0$ onwards both entries of $z(t)$ decay to zero as t increases. Therefore

$$\|z\|_{\mathcal{L}_\infty} = \sup_{t \geq 0} \max(e^{-t}, 2e^{-3t}) = 2.$$

The square of the \mathcal{L}_2 -norm follows as

$$\|z\|_{\mathcal{L}_2}^2 = \int_0^\infty e^{-2t} + 4e^{-6t} dt = \frac{e^{-2t}}{-2} + 4 \frac{e^{-6t}}{-6} \Big|_{t=0}^{t=\infty} = \frac{1}{2} + \frac{4}{6} = \frac{7}{6}$$

The energy of z equals $7/6$, its \mathcal{L}_2 -norm is $\sqrt{7/6}$. □

B.4. Norms of linear operators and systems

On the basis of normed signal spaces we may define norms of operators that act on these signals.

Definition B.4.1 (Induced norm of a linear operator). Suppose that ϕ is a linear map $\phi : \mathcal{U} \rightarrow \mathcal{Y}$ from a normed space \mathcal{U} with norm $\|\cdot\|_{\mathcal{U}}$ to a normed space \mathcal{Y} with norm $\|\cdot\|_{\mathcal{Y}}$. Then the norm of the operator ϕ induced by the norms $\|\cdot\|_{\mathcal{U}}$ and $\|\cdot\|_{\mathcal{Y}}$ is defined by

$$\|\phi\| = \sup_{\|u\|_{\mathcal{U}} \neq 0} \frac{\|\phi u\|_{\mathcal{Y}}}{\|u\|_{\mathcal{U}}}. \quad (\text{B.13})$$

□

Constant matrices represent linear maps, so that (B.13) may be used to define norms of matrices. For instance the spectral norm $\bar{\sigma}(M)$ is the norm induced by the 2-norm (see Exercise B.2(B.2a)).

Lemma B.4.2 (The spectral norm is an induced norm). Let $M \in \mathbb{C}^{m \times n}$. Then

$$\bar{\sigma}(M) = \sup_{x \in \mathbb{C}^n, x \neq 0} \frac{\|Mx\|_2}{\|x\|_2}.$$

□

B.4.1. Norms of linear systems

We next turn to a discussion of the norm of a system. Consider a system as in Fig. B.1, which maps the input signal u into an output signal y . Given an input u , we denote the output of the system as $y = \phi u$. If ϕ is a linear operator the system is said to be linear. The norm of the system is now defined as the norm of this operator.

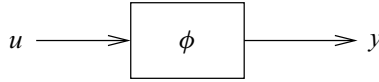


Figure B.1: Input-output mapping system

We establish formulas for the norms of a linear time-invariant system induced by the \mathcal{L}_2 - and \mathcal{L}_∞ -norms of the input and output signals.

Summary B.4.3 (Norms of linear time-invariant systems). Consider a MIMO convolution system with impulse response matrix h ,

$$y(t) = \int_{-\infty}^{\infty} h(\tau) u(t - \tau) d\tau, \quad t \in \mathbb{R}. \quad (\text{B.14})$$

Moreover let H denote the transfer matrix, i.e., H is the Laplace transform of h .

1. \mathcal{L}_∞ -induced norm. The norm of the system induced by the \mathcal{L}_∞ -norm (B.11) is given by

$$\max_{i=1,2,\dots,m} \int_{-\infty}^{\infty} \sum_{j=1}^k |h_{ij}(t)| dt, \quad (\text{B.15})$$

where h_{ij} is the (i, j) entry of the $m \times k$ impulse response matrix h .

2. \mathcal{L}_2 -induced norm. Suppose H is a rational matrix. The norm of the system induced by the \mathcal{L}_2 -norm exists if and only if H is proper and has no poles in the closed right-half plane. In that case the \mathcal{L}_2 -induced norm (B.13) equals the \mathcal{H}_∞ -norm of the transfer matrix defined as

$$\|H\|_{\mathcal{H}_\infty} = \sup_{\omega \in \mathbb{R}} \bar{\sigma}(H(j\omega)). \quad (\text{B.16})$$

□

A sketch of the proof is found in § B.6. For SISO systems the expressions for the two norms obtained in Summary B.4.3 simplify considerably.

Summary B.4.4 (Norms for SISO systems). The norms of a SISO system with (scalar) impulse response h and transfer function H induced by the \mathcal{L}_∞ - and \mathcal{L}_2 -norms are successively given by *action* of the impulse response

$$\|h\|_{\mathcal{L}_1} = \int_{-\infty}^{\infty} |h(t)| dt, \quad (\text{B.17})$$

and the peak value on the Bode plot if H has no unstable poles

$$\|H\|_{\mathcal{H}_\infty} = \sup_{\omega \in \mathbb{R}} |H(j\omega)|. \quad (\text{B.18})$$

If H has unstable poles then the induced norms do not exist.

□

Example B.4.5 (Norms of a simple system). As a simple example, consider a SISO first-order system with transfer function

$$H(s) = \frac{1}{1 + s\theta}, \quad (\text{B.19})$$

with θ a positive constant. The corresponding impulse response is

$$h(t) = \begin{cases} \frac{1}{\theta} e^{-t/\theta} & \text{for } t \geq 0, \\ 0 & \text{for } t < 0. \end{cases} \quad (\text{B.20})$$

It is easily found that the norm of the system induced by the \mathcal{L}_∞ -norm is

$$\|h\|_1 = \int_0^{\infty} \frac{1}{\theta} e^{-t/\theta} dt = 1. \quad (\text{B.21})$$

The norm induced by the \mathcal{L}_2 -norm follows as

$$\|H\|_{\mathcal{H}_\infty} = \sup_{\omega \in \mathbb{R}} \frac{1}{|1 + j\omega\theta|} = \sup_{\omega \in \mathbb{R}} \frac{1}{\sqrt{1 + \omega^2\theta^2}} = 1. \quad (\text{B.22})$$

For this example the two system norms are equal. Usually they are not.

□

Remark. In the robust control literature the \mathcal{H}_∞ -norm of a transfer matrix H is commonly denoted by $\|H\|_\infty$ and not by $\|H\|_{\mathcal{H}_\infty}$ as we have done here.

Summary B.4.6 (State-space computation of common system norms). Suppose a system has proper transfer matrix H and let $H(s) = C(sI_n - A)^{-1}B + D$ be a minimal realization of H . Then

1. $\|H\|_{\mathcal{H}_2}$ is finite if and only if $D = 0$ and A is a *stability matrix*². In that case

$$\|H\|_{\mathcal{H}_2} = \sqrt{\text{tr } B^T Y B} = \sqrt{\text{tr } C X C^T}$$

where X and Y are the solutions of the linear equations

$$AX + XA^T = -BB^T, \quad A^T Y + YA = -C^T C.$$

The solutions X and Y are unique and are respectively known as the *controllability gramian* and *observability gramian*.

2. The Hankel norm $\|H\|_{\text{H}}$ is finite only if A is a stability matrix. In that case

$$\|H\|_{\text{H}} = \sqrt{\lambda_{\max}(XY)}$$

where X and Y are the controllability gramian and observability gramian. The matrix XY has real nonnegative eigenvalues only, and $\lambda_{\max}(XY)$ denotes the largest of them.

3. The \mathcal{H}_{∞} -norm $\|H\|_{\mathcal{H}_{\infty}}$ is finite if and only if A is a stability matrix. Then $\gamma \in \mathbb{R}$ is an upper bound

$$\|H\|_{\mathcal{H}_{\infty}} < \gamma$$

if and only if $\overline{\sigma}(D) < \gamma$ and the $2n \times 2n$ matrix

$$\begin{bmatrix} A & 0 \\ -C^T C & -A^T \end{bmatrix} - \begin{bmatrix} -B \\ C^T D \end{bmatrix} (\gamma^2 I - D^T D)^{-1} \begin{bmatrix} D^T C & B^T \end{bmatrix} \quad (\text{B.23})$$

has no imaginary eigenvalues. This result shows that computation of the \mathcal{H}_{∞} -norm can be done with iteration (on γ) and computation of eigenvalues. There are various other approaches.

□

Proofs are listed in Appendix B.6.

B.5. BIBO and internal stability

Norms shed a new light on the notions of stability and internal stability. Consider the MIMO input-output mapping system of Fig. B.1 with linear input-output map ϕ . In § 1.3.2 we defined the system to be BIBO stable if any bounded input u results in a bounded output $y = \phi u$. Now bounded means having finite norm, and so different norms may yield different versions of BIBO stability. Normally the \mathcal{L}_{∞} - or \mathcal{L}_2 -norms of u and y are used.

We review a few fairly obvious facts.

Summary B.5.1 (BIBO stability).

1. If $\|\phi\|$ is finite, with $\|\cdot\|$ the norm induced by the norms of the input and output signals, then the system is said to be BIBO stable (with respect the

²A constant matrix A is a *stability matrix* if it is square and all its eigenvalues have negative real part.

2. Suppose that the system is linear time-invariant with rational transfer matrix H . Then the system is BIBO stable (both when the \mathcal{L}_∞ - and the \mathcal{L}_2 -norms of u and y are used) if and only if H is proper and has all its poles in the open left-half complex plane. □

Given the above it will be no surprise that we say that a transfer matrix is *stable* if it is proper and has all its poles in the open left-half complex plane. In § 1.3.2 we already defined the notion of *internal stability* for MIMO interconnected systems: An interconnected system is internally stable if the system obtained by adding external input and output signals in every exposed interconnection is BIBO stable. Here we specialize this for the MIMO system of Fig. B.2.

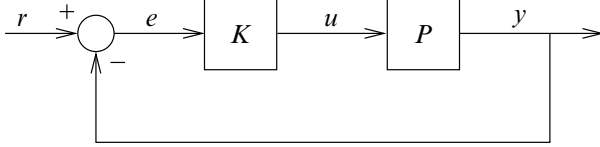


Figure B.2: Multivariable feedback structure

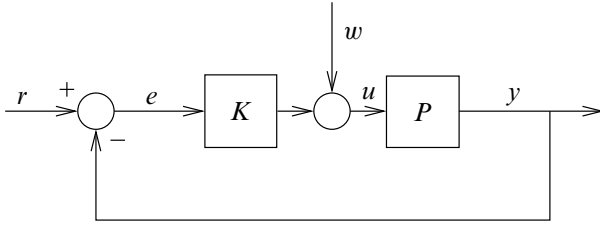


Figure B.3: Inputs r, w and outputs e, u defining internal stability

Assume that feedback system shown in Fig. B.2 is *multivariable* i.e., u or y have more than one entry. The closed loop of Fig. B.2 is by definition *internally stable* if in the extended closed loop of Fig. B.3 the maps from (w, r) to (e, u) are BIBO stable. In terms of the transfer matrices, these maps are

$$\begin{bmatrix} e \\ u \end{bmatrix} = \begin{bmatrix} I & P \\ -K & I \end{bmatrix}^{-1} \begin{bmatrix} r \\ w \end{bmatrix} \quad (\text{B.24})$$

$$= \begin{bmatrix} I - P(I + KP)^{-1}K & -P(I + KP)^{-1} \\ (I + KP)^{-1}K & (I + KP)^{-1} \end{bmatrix} \begin{bmatrix} r \\ w \end{bmatrix} \quad (\text{B.25})$$

Necessary for internal stability is that

$$\det(I + P(\infty)K(\infty)) \neq 0.$$

Feedback systems that satisfy this nonsingularity condition are said to be *well-posed*, Hsu and Chen (1968). The following is a variation of Eqn. (1.44).

Lemma B.5.2 (Internal stability). Suppose P and K are proper, having minimal realizations (A_P, B_P, C_P, D_P) and (A_K, B_K, C_K, D_K) . Then the feedback system of Fig. B.2 is internally stable if and only if $\det(I + D_P D_K) \neq 0$ and the polynomial

$$\det(I + P(s)K(s)) \det(sI - A_P) \det(sI - A_K)$$

has all its zeros in the open left-half plane. □

If the $n_y \times n_u$ system P is not square, then the dimension $n_y \times n_y$ of $I + PK$ and dimension $n_u \times n_u$ of $I + KP$ differ. The fact that $\det(I + PK) = \det(I + KP)$ allows to evaluate the stability of the system at the loop location having the lowest dimension.

Example B.5.3 (Internal stability of closed-loop system). Consider the loop gain

$$L(s) = \begin{bmatrix} \frac{1}{s+1} & \frac{1}{s-1} \\ 0 & \frac{1}{s+1} \end{bmatrix}.$$

The unity feedback around L gives as closed-loop characteristic polynomial:

$$\det(I + L(s)) \det(sI - A_L) = \left(\frac{s+2}{s+1} \right)^2 (s+1)^2 (s-1) = (s+2)^2 (s-1). \quad (\text{B.26})$$

The closed loop system hence is not internally stable. Evaluation of the sensitivity matrix

$$(I + L(s))^{-1} = \begin{bmatrix} \frac{s+1}{s+2} & -\frac{s+1}{(s+2)^2} \\ 0 & \frac{s+1}{s+2} \end{bmatrix}$$

shows that this transfer matrix in the closed-loop system is stable, but the complementary sensitivity matrix

$$(I + L(s))^{-1} L(s) = \begin{bmatrix} \frac{1}{s+2} & \frac{s^2+2s+3}{(s-1)(s+2)^2} \\ 0 & \frac{1}{s+2} \end{bmatrix}$$

is unstable, confirming that the closed-loop system indeed is not internally stable. □

Example B.5.4 (Internal stability and pole-zero cancellation). Consider

$$P(s) = \begin{bmatrix} \frac{1}{s} & \frac{1}{s} \\ \frac{2}{s+1} & \frac{1}{s} \end{bmatrix}, \quad K(s) = \begin{bmatrix} 1 & \frac{2s}{s-1} \\ 0 & -\frac{2s}{s-1} \end{bmatrix}.$$

The unstable pole $s = 1$ of the controller does not appear in the product PK ,

$$P(s)K(s) = \begin{bmatrix} \frac{1}{s} & 0 \\ \frac{2}{s+1} & \frac{2}{s+1} \end{bmatrix}.$$

This may suggest as in the SISO case that the map from r to u will be unstable. Indeed the lower-left entry of this map is unstable,

$$K(I + PK)^{-1} = \begin{bmatrix} * & * \\ -\frac{4s^2}{(s-1)(s+1)(s+3)} & * \end{bmatrix}.$$

Instability may also be verified using Lemma B.5.2; it is readily verified that

$$\begin{aligned} \det(I + P(s)K(s)) \det(sI - A_P) \det(sI - A_K) \\ = \det \begin{bmatrix} \frac{s+1}{s} & 0 \\ \frac{2}{s+1} & \frac{s+3}{s+1} \end{bmatrix} s^2 (s+1)(s-1) = s^2 (s-1)(s+1)(s+3) \end{aligned}$$

and this has a zero at $s = 1$. □

B.6. Appendix: Proofs

Proof of Summary B.4.6.

1. If D is not zero then the impulse response matrix $h(t) = Ce^{At}B\mathbb{1}(t) + \delta(t)D$ contains Dirac delta functions and as a result the \mathcal{H}_2 -norm can not be finite. So $D = 0$ and

$$h(t) = Ce^{At}B\mathbb{1}(t).$$

If A is not a stability matrix then $h(t)$ is unbounded, hence the \mathcal{H}_2 -norm is infinite. So $D = 0$ and A is a stability matrix. Then

$$\begin{aligned}\|H\|_{\mathcal{H}_2}^2 &= \text{tr} \int_{-\infty}^{\infty} h(t)^T h(t) dt \\ &= \text{tr} \int_0^{\infty} B^T e^{A^T t} C^T C e^{At} B dt \\ &= \text{tr} \left(B^T \underbrace{\int_0^{\infty} e^{A^T t} C^T C e^{At} dt}_Y B \right).\end{aligned}$$

The so defined matrix Y satisfies

$$A^T Y + Y A = \int_0^{\infty} A^T e^{A^T t} C^T C e^{At} + e^{A^T t} C^T C e^{At} A dt = e^{A^T t} C^T C e^{At} \Big|_{t=0}^{t=\infty} = -C^T C. \quad (\text{B.27})$$

That is, Y satisfies the Lyapunov equation $A^T Y + Y A = -C^T C$, and as A is a stability matrix the solution Y of (B.27) is well known to be unique.

2. Let X and Y be the controllability and observability gramians. By the property of state, the output y for positive time is a function of the state x_0 at $t = 0$. Then

$$\int_0^{\infty} y(t)^T y(t) dt = \int_0^{\infty} x_0^T e^{A^T t} C^T C e^{At} x_0 dt = x_0^T Y x_0.$$

If X satisfies $AX + XA^T = -BB^T$ then its inverse $Z := X^{-1}$ satisfies $ZA + A^T Z = -ZBB^T Z$. By completion of the square we may write

$$\begin{aligned}\frac{d}{dt} x^T Z x &= 2x^T Z \dot{x} = 2x^T (ZAx + ZBu) = x^T (ZA + A^T Z)x + 2x^T ZBu \\ &= x^T (-ZBB^T Z)x + 2x^T ZBu \\ &= u^T u - (u - B^T Zx)^T (u - B^T Zx).\end{aligned}$$

Therefore, assuming $x(-\infty) = 0$,

$$\int_{-\infty}^0 \|u(t)\|_2^2 - \|u(t) - B^T Zx(t)\|_2^2 dt = x^T(t)Zx(t) \Big|_{t=-\infty}^{t=0} = x_0^T Zx_0 = x_0^T X^{-1}x_0.$$

From this expression it follows that the smallest u (in norm) that steers x from $x(-\infty) = 0$ to $x(0) = x_0$ is $u(t) = B^T Zx(t)$ (verify this). This then shows that

$$\sup_u \frac{\int_0^{\infty} y^T(t)y(t) dt}{\int_{-\infty}^0 u^T(t)u(t) dt} = \frac{x_0^T Y x_0}{x_0^T X^{-1} x_0}.$$

Now this supremum is less than some γ^2 for some x_0 if and only if

$$x_0^T Y x_0 < \gamma^2 \cdot x_0^T X^{-1} x_0.$$

There exist such x_0 iff $Y < \gamma^2 X^{-1}$, which in turn is equivalent to that $\lambda_i(YX) = \lambda_i(X^{1/2} Y X^{1/2}) < \gamma^2$. This proves the result.

3. $\|H\|_{\mathcal{H}_\infty} < \gamma$ holds if and only if H is stable and $\gamma^2 I - H^* H$ is positive definite on $j\mathbb{R} \cup \infty$. This is the case iff it is positive definite at infinity (i.e., $\overline{\sigma}(D) < \gamma$, i.e. $\gamma^2 I - D^T D > 0$) and nowhere on the imaginary axis $\gamma^2 I - H^* H$ is singular. We will show that $\gamma^2 I - H^* H$ has imaginary zeros iff (B.23) has imaginary eigenvalues. It is readily verified that a realization of $\gamma^2 I - H^* H$ is

$$\left[\begin{array}{cc|c} A - sI & 0 & -B \\ -C^T C & -A^T - sI & C^T D \\ \hline D^T C & B^T & \gamma^2 I - D^T D \end{array} \right]$$

By the Schur complement results applied to this matrix we see that

$$\det \begin{bmatrix} A - sI & 0 \\ -C^T C & -A^T - sI \end{bmatrix} \cdot \det(\gamma^2 I - H^* H) = \det(\gamma^2 I - D^T D) \cdot \det(A_{\text{ham}} - sI).$$

where A_{ham} is the Hamiltonian matrix (B.23). As A has no imaginary eigenvalues we have that $\gamma^2 I - H^* H$ has no zeros on the imaginary axis iff A_{ham} has no imaginary eigenvalues. ■

Proof of Norms of linear time-invariant systems. We indicate how the formulas for the system norms as given in Summary B.4.3 are obtained.

1. \mathcal{L}_∞ -induced norm. In terms of explicit sums, the i th component y_i of the output $y = h * u$ of the convolution system may be bounded as

$$\begin{aligned} |y_i(t)| &= \left| \int_{-\infty}^{\infty} \sum_j h_{ij}(\tau) u_j(t - \tau) d\tau \right| \leq \int_{-\infty}^{\infty} \sum_j |h_{ij}(\tau)| \cdot |u_j(t - \tau)| d\tau \\ &\leq \left(\int_{-\infty}^{\infty} \sum_j |h_{ij}(\tau)| d\tau \right) \cdot \|u\|_{\mathcal{L}_\infty}, \quad t \in \mathbb{R}, \end{aligned} \quad (\text{B.28})$$

so that

$$\|y\|_{\mathcal{L}_\infty} \leq \max_i \left(\int_{-\infty}^{\infty} \sum_j |h_{ij}(\tau)| d\tau \right) \cdot \|u\|_{\mathcal{L}_\infty}. \quad (\text{B.29})$$

This shows that

$$\|\phi\| \leq \max_i \int_{-\infty}^{\infty} \sum_j |h_{ij}(\tau)| d\tau. \quad (\text{B.30})$$

The proof that the inequality may be replaced by equality follows by showing that there exists an input u such that (B.29) is achieved with *equality*.

2. \mathcal{L}_2 -induced norm. By Parseval's theorem (see e.g. Kwakernaak and Sivan (1991)) we have that

$$\begin{aligned} \frac{\|y\|_{\mathcal{L}_2}^2}{\|u\|_{\mathcal{L}_2}^2} &= \frac{\int_{-\infty}^{\infty} y^H(t)y(t) dt}{\int_{-\infty}^{\infty} u^H(t)u(t) dt} \\ &= \frac{\frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{y}^H(j\omega)\hat{y}(j\omega) d\omega}{\frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{u}^H(j\omega)\hat{u}(j\omega) d\omega} \end{aligned} \quad (\text{B.31})$$

$$= \frac{\int_{-\infty}^{\infty} \hat{u}^H(j\omega) H^H(j\omega) H(j\omega) \hat{u}(j\omega) d\omega}{\int_{-\infty}^{\infty} \hat{u}^H(j\omega)\hat{u}(j\omega) d\omega} \quad (\text{B.32})$$

where \hat{y} is the Laplace transform of y and \hat{u} that of u . For any fixed frequency ω we have that

$$\sup_{u(j\omega)} \frac{\hat{u}^H(j\omega) H^H(j\omega) H(j\omega) \hat{u}(j\omega)}{\hat{u}^H(j\omega)\hat{u}(j\omega)} = \bar{\sigma}^2(H(j\omega)).$$

Therefore

$$\sup_u \frac{\|y\|_{\mathcal{L}_2}^2}{\|u\|_{\mathcal{L}_2}^2} \leq \sup_{\omega} \bar{\sigma}^2(H(j\omega)). \quad (\text{B.33})$$

The right hand side of this inequality is by definition $\|H\|_{\mathcal{H}_\infty}^2$. The proof that the inequality may be replaced by equality follows by showing that there exists an input u for which (B.33) is achieved with equality within an arbitrarily small positive margin ε . ■

B.7. Problems

B.1 *Singular value decomposition.* Let $A = U \Sigma V^H$ be the singular value decomposition of the $n \times m$ matrix A , with singular values σ_i , $i = 1, 2, \dots, \min(n, m)$. Denote the columns of the $n \times n$ unitary matrix U as u_i , $i = 1, 2, \dots, n$, and those of the $m \times m$ unitary matrix V as v_i , $i = 1, 2, \dots, m$. Prove the following statements:

- For $i = 1, 2, \dots, \min(n, m)$ the column vector u_i is an eigenvector of AA^H corresponding to the eigenvalue σ_i^2 . Any remaining columns are eigenvectors corresponding to the eigenvalue 0.
- Similarly, for $i = 1, 2, \dots, \min(n, m)$ the column vector v_i is an eigenvector of $A^H A$ corresponding to the eigenvalue σ_i^2 . Any remaining columns are eigenvectors corresponding to the eigenvalue 0.
- For $i = 1, 2, \dots, \min(n, m)$ the vectors u_i and v_i satisfy

$$Av_i = \sigma_i u_i, \quad A^H u_i = \sigma_i v_i. \quad (\text{B.34})$$

B.2 *Singular values.* Given is a square $n \times n$ matrix A . Prove the following properties (compare p. R-5 of Chiang and Safonov (1988)):

- $\bar{\sigma}(A) = \max_{x \in \mathbb{C}^n, x \neq 0} \frac{\|Ax\|_2}{\|x\|_2}$.

b) $\underline{\sigma}(A) = \min_{x \in \mathbb{C}^n, x \neq 0} \frac{\|Ax\|_2}{\|x\|_2}$.

c) $\underline{\sigma}(A) \leq |\lambda_i(A)| \leq \overline{\sigma}(A)$, with $\lambda_i(A)$ the i th eigenvalue of A .

d) If A^{-1} exists then $\underline{\sigma}(A) = 1/\overline{\sigma}(A^{-1})$ and $\overline{\sigma}(A) = 1/\underline{\sigma}(A^{-1})$.

e) $\overline{\sigma}(\alpha A) = |\alpha| \overline{\sigma}(A)$, with α any complex number.

f) $\overline{\sigma}(A + B) \leq \overline{\sigma}(A) + \overline{\sigma}(B)$.

g) $\overline{\sigma}(AB) \leq \overline{\sigma}(A) \overline{\sigma}(B)$.

h) $\underline{\sigma}(A) - \overline{\sigma}(B) \leq \underline{\sigma}(A + B) \leq \underline{\sigma}(A) + \overline{\sigma}(B)$.

i) $\max(\overline{\sigma}(A), \overline{\sigma}(B)) \leq \overline{\sigma}([A \ B]) \leq \sqrt{2} \max(\overline{\sigma}(A), \overline{\sigma}(B))$.

j) $\max_{i,j} |A_{ij}| \leq \overline{\sigma}(A) \leq n \max_{i,j} |A_{ij}|$, with A_{ij} the (i, j) element of A .

k) $\sum_{i=1}^n \sigma_i^2(A) = \text{tr}(A^H A)$.

B.3 Induced norms of linear operators. The space of linear operators from a vector space \mathcal{U} to a vector space \mathcal{V} is itself a vector space.

- a) Show that the induced norm $\|\cdot\|$ as defined by (B.13) is indeed a norm on this space satisfying the properties of Definition B.1.1.

Prove that this norm has the following additional properties:

b) If $y = \phi u$, then $\|y\|_{\mathcal{V}} \leq \|\phi\| \cdot \|u\|_{\mathcal{U}}$ for any $u \in \mathcal{U}$.

- c) *Submultiplicative property:* Let $\phi_1 : \mathcal{U} \rightarrow \mathcal{V}$ and $\phi_2 : \mathcal{V} \rightarrow \mathcal{Y}$ be two linear operators, with \mathcal{U} , \mathcal{V} and \mathcal{Y} normed spaces. Then $\|\phi_2 \phi_1\| \leq \|\phi_2\| \cdot \|\phi_1\|$, with all the norms induced.

B.4 Matrix norms induced by vector norms. Consider the $m \times k$ complex-valued matrix M as a linear map $\mathbb{C}^k \rightarrow \mathbb{C}^m$ defined by $y = Mu$. Then depending on the norms defined on \mathbb{C}^k and \mathbb{C}^m we obtain different matrix norms.

- a) *Matrix norm induced by the 1-norm.* Prove that the norm of the matrix M induced by the 1-norm (both for \mathcal{U} and \mathcal{Y}) is the *maximum absolute column sum*

$$\|M\| = \max_j \sum_i |M_{ij}|, \quad (\text{B.35})$$

with M_{ij} the (i, j) entry of M .

- b) *Matrix norm induced by the ∞ -norm.* Prove that the norm of the matrix M induced by the ∞ -norm (both for \mathcal{U} and \mathcal{Y}) is the *maximum absolute row sum*

$$\|M\| = \max_i \sum_j |M_{ij}|, \quad (\text{B.36})$$

with M_{ij} the (i, j) entry of M .

Prove these statements.

B.5 \mathcal{H}_2 -norm and Hankel norm of MIMO systems. Two further norms of linear time-invariant systems are commonly encountered. Consider a stable MIMO system with transfer matrix H and impulse response matrix h .

a) \mathcal{H}_2 -norm. Show that the \mathcal{H}_2 -norm defined as

$$\|H\|_{\mathcal{H}_2} = \sqrt{\text{tr} \left(\int_{-\infty}^{\infty} H^T(-j2\pi f) H(j2\pi f) df \right)} = \sqrt{\text{tr} \left(\int_{-\infty}^{\infty} h^T(t) h(t) dt \right)} \quad (\text{B.37})$$

is a norm. The notation tr indicates the trace of a matrix.

b) *Hankel norm*. The impulse response matrix h defines an operator

$$y(t) = \int_{-\infty}^0 h(t - \tau) u(\tau) d\tau, \quad t \geq 0. \quad (\text{B.38})$$

which maps continuous-time signals u defined on the time axis $(-\infty, 0]$ to continuous-time signals y defined on the time axis $[0, \infty)$. The Hankel norm $\|H\|_{\text{H}}$ of the system with impulse response matrix h is defined as the norm of the map given by (B.38) induced by the \mathcal{L}_2 -norms of $u : (-\infty, 0] \rightarrow \mathbb{R}^{n_u}$ and $y : [0, \infty) \rightarrow \mathbb{R}^{n_y}$. Prove that this is indeed a norm.

c) Compute the \mathcal{H}_2 -norm and the Hankel norm of the SISO system of Example B.4.5. *Hint:* To compute the Hankel norm first show that if y satisfies (B.38) and h is given by (B.20) then $\|y\|_2^2 = \frac{1}{2\theta} \left(\int_{-\infty}^0 e^{\tau/\theta} u(\tau) d\tau \right)^2$. From this, prove that $\|H\|_{\text{H}} = \frac{1}{2}$.

B.6 Proof. Consider Summary B.5.1.

- Prove the statements 1 and 2.
- Show by a counterexample that the converse of 1 is not true, that is, for certain systems and norms the system may be BIBO stable in the sense as defined while $\|\phi\|$ is not necessarily finite.
- In the literature the following better though more complicated definition of BIBO stability is found: The system is BIBO stable if for every positive real constant N there exists a positive real constant M such that $\|u\| \leq N$ implies $\|y\| \leq M$. With this definition the system is BIBO stable if and only if $\|\phi\| < \infty$. Prove this.

C. Bibliography

- M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover Publications, Inc., New York, 1965.
- B. D. O. Anderson and E. I. Jury. A note on the Youla-Bongiorno-Lu condition. *Automatica*, 12: 387–388, 1976.
- B. D. O. Anderson and J. B. Moore. *Linear Optimal Control*. Prentice-Hall, Englewood Cliffs, NJ, 1971.
- B. D. O. Anderson and J. B. Moore. *Optimal Control: Linear Quadratic Methods*. Prentice Hall, Englewood Cliffs, NJ, 1990.
- J. D. Aplevich. *Implicit Linear Systems*, volume 152 of *Lecture Notes in Control and Information Sciences*. Springer-Verlag, Berlin, 1991. ISBN 0-387-53537-3.
- A. Bagchi. *Optimal control of stochastic systems*. International Series in Systems and Control Engineering. Prentice Hall International, Hempstead, UK, 1993.
- G. J. Balas, J. C. Doyle, K. Glover, A. Packard, and R. Smith. *User's Guide, μ -Analysis and Synthesis Toolbox*. The MathWorks, Natick, Mass., 1991.
- B. R. Barmish and H. I. Kang. A survey of extreme point results for robustness of control systems. *Automatica*, 29:13–35, 1993.
- A. C. Bartlett, C. V. Hollot, and Huang Lin. Root locations of an entire polytope of polynomials: It suffices to check the edges. *Math. Control Signals Systems*, 1:61–71, 1988.
- G. Basile and G. Marro. *Controlled and Conditioned Invariance in Linear System Theory*. Prentice Hall, Inc., Englewood Cliffs, NJ, 1992.
- G. Berman and R. G. Stanton. The asymptotes of the root locus. *SIAM Review*, 5:209–218, 1963.
- S. Białas. A necessary and sufficient condition for the stability of convex combinations of polynomials or matrices. *Bulletin Polish Acad. of Sciences*, 33:473–480, 1985.
- H. S. Black. Stabilized feedback amplifiers. *Bell System Technical Journal*, 13:1–18, 1934.
- V. Blondel. *Simultaneous Stabilization of Linear Systems*, volume 191 of *Lecture Notes in Control and Information Sciences*. Springer-Verlag, Berlin etc., 1994.
- H. W. Bode. Relations between attenuation and phase in feedback amplifier design. *Bell System Technical Journal*, 19:421–454, 1940.
- H. W. Bode. *Network Analysis and Feedback Amplifier Design*. Van Nostrand, New York, 1945.
- S. P. Boyd and C. H. Barratt. *Linear Controller Design: Limits of Performance*. Prentice Hall, Englewood Cliffs, NJ, 1991.

- E. H. Bristol. On a new measure of interaction for multivariable process control. *IEEE Transactions on Automatic Control*, 11:133–134, 1966.
- R. W. Brockett and M. D. Mesarovic. The reproducibility of multivariable systems. *Journal of Mathematical Analysis and Applications*, 11:548–563, 1965.
- P. J. Campo and M. Morari. Achievable closed-loop properties of systems under decentralized control: conditions involving the steady-state gain. *IEEE Transactions on Automatic Control*, 39:932–943, 1994.
- C.-T. Chen. *Introduction to Linear Systems Theory*. Holt, Rinehart and Winston, New York, 1970.
- R. Y. Chiang and M. G. Safonov. *User's Guide, Robust-Control Toolbox*. The MathWorks, South Natick, Mass., 1988.
- R. Y. Chiang and M. G. Safonov. *User's Guide, Robust Control Toolbox*. The MathWorks, Natick, Mass., USA, 1992.
- S.-G. Choi and M. A. Johnson. The root loci of H_∞ -optimal control: A polynomial approach. *Optimal Control Applications and Methods*, 17:79–105, 1996.
- Control Toolbox. *User's Guide*. The MathWorks, Natick, Mass., USA, 1990.
- E. J. Davison. Robust servomechanism problem. In *The Control Handbook*, pages 731–747. CRC press, 1996.
- E. J. Davison and S. H. Wang. Properties and calculation of transmission zeros of linear multivariable systems. *Automatica*, 10:643–658, 1974.
- E. J. Davison and S. H. Wang. 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, 1978.
- J. J. D'Azzo and C. H. Houpis. *Linear Control System Analysis and Design: Conventional and Modern*. McGraw-Hill Book Co., New York, NY, 1988. Third Edition.
- C. A. Desoer and J. D. Schulman. Zeros and poles of matrix transfer functions and their dynamical interpretations. *IEEE Transactions on Circuits and Systems*, 21:3–8, 1974.
- C. A. Desoer and M. Vidyasagar. *Feedback Systems: Input-Output Properties*. Academic Press, New York, 1975.
- C. A. Desoer and Y. T. Wang. Linear time-invariant robust servomechanism problem: a self-contained exposition. In C. T. Leondes, editor, *Control and Dynamic Systems*, volume 16, pages 81–129. Academic Press, New York, NY, 1980.
- R. C. Dorf. *Modern Control Systems*. Addison-Wesley Publishing Co., Reading, MA, 1992. Sixth Edition.
- J. C. Doyle. Robustness of multiloop linear feedback systems. In *Proc. 17th IEEE Conf. Decision & Control*, pages 12–18, 1979.
- J. C. Doyle. Analysis of feedback systems with structured uncertainties. *IEE Proc., Part D*, 129: 242–250, 1982.

- J. C. Doyle. Lecture Notes, ONR/Honeywell Workshop on Advances in Multivariable Control, Minneapolis, Minn., 1984.
- J. C. Doyle, B. A. Francis, and A. R. Tannenbaum. *Feedback Control Theory*. Macmillan, New York, 1992.
- J. C. Doyle, K. Glover, P. P. Khargonekar, and B. A. Francis. State-space solutions to standard \mathcal{H}_2 and \mathcal{H}_∞ control problems. *IEEE Trans. Aut. Control*, 34:831–847, 1989.
- J. C. Doyle and G. Stein. Multivariable feedback design: concepts for a classical/modern synthesis. *IEEE Transactions on Automatic Control*, 26:4–16, 1981.
- S. Engell. *Optimale lineare Regelung: Grenzen der erreichbaren Regelgüte in linearen zeitinvarianten Regelkreisen*, volume 18 of *Fachberichte Messen–Steuern–Regeln*. Springer-Verlag, Berlin, etc., 1988.
- W. R. Evans. Graphical analysis of control systems. *Trans. Amer. Institute of Electrical Engineers*, 67:547–551, 1948.
- W. R. Evans. Control system synthesis by root locus method. *Trans. Amer. Institute of Electrical Engineers*, 69:66–69, 1950.
- W. R. Evans. *Control System Dynamics*. McGraw-Hill Book Co., New York, NY, 1954.
- O. Föllinger. Über die Bestimmung der Wurzelortskurve. *Regelungstechnik*, 6:442–446, 1958.
- B. A. Francis and W. M. Wonham. The role of transmission zeros in linear multivariable regulators. *International Journal of Control*, 22:657–681, 1975.
- G. F. Franklin, J. D. Powell, and A. Emami-Naeini. *Feedback Control of Dynamic Systems*. Addison-Wesley Publ. Co., Reading, MA, 1986.
- G. F. Franklin, J. D. Powell, and A. Emami-Naeini. *Feedback Control of Dynamic Systems*. Addison-Wesley Publ. Co., Reading, MA, 1991. Second Edition.
- J. S. Freudenberg and D. P. Looze. Right half plane poles and zeros and design tradeoffs in feedback systems. *IEEE Trans. Automatic Control*, 30:555–565, 1985.
- J. S. Freudenberg and D. P. Looze. *Frequency Domain Properties of Scalar and Multivariable Feedback Systems*, volume 104 of *Lecture Notes in Control and Information Sciences*. Springer-Verlag, Berlin, etc., 1988.
- F. R. Gantmacher. *The Theory of Matrices*. Chelsea Publishing Company, New York, NY, 1964. Reprinted.
- K. Glover and J. C. Doyle. State-space formulae for all stabilizing controllers that satisfy an H_∞ -norm bound and relations to risk sensitivity. *Systems & Control Letters*, 11:167–172, 1988.
- K. Glover and J. C. Doyle. A state space approach to H_∞ optimal control. In H. Nijmeijer and J. M. Schumacher, editors, *Three Decades of Mathematical System Theory*, volume 135 of *Lecture Notes in Control and Information Sciences*. Springer-Verlag, Heidelberg, etc., 1989.
- K. Glover, D. J. N. Limebeer, J. C. Doyle, E. M. Kasenally, and M. G. Safonov. A characterization of all solutions to the four block general distance problem. *SIAM J. Control and Optimization*, 29:283–324, 1991.

- I. Gohberg, P. Lancaster, and L. Rodman. *Matrix polynomials*. Academic Press, New York, NY, 1982.
- G. M. Golub and C. Van Loan. *Matrix Computations*. The Johns Hopkins University Press, Baltimore, Maryland, 1983.
- D. Graham and R. C. Lathrop. The synthesis of ‘optimum’ transient response: Criteria and standard forms. *Trans. AIEE Part II (Applications and Industry)*, 72:273–288, 1953.
- M. Green and D. J. N. Limebeer. *Linear Robust Control*. Prentice Hall, Englewood Cliffs, 1995.
- M. J. Grimble. Two and a half degrees of freedom LQG controller and application to wind turbines. *IEEE Trans. Aut. Control*, 39(1):122–127, 1994.
- P. Grosdidier and M. Morari. A computer aided methodology for the design of decentralized controllers. *Computers and Chemical Engineering*, 11:423–433, 1987.
- P. Grosdidier, M. Morari, and B. R. Holt. Closed-loop properties from steady-state gain information. *Industrial and Engineering Chemistry. Fundamentals*, 24:221–235, 1985.
- P. Hagander and B. Bernhardsson. A simple test example for H_∞ optimal control. In *Preprints, 1992 American Control Conference, Chicago, Illinois*, 1992.
- P. Henrici. *Applied and Computational Complex Analysis, Vol. I*. Wiley-Interscience, New York, 1974.
- D. Hinrichsen and A. J. Pritchard. Stability radii of linear systems. *Systems & Control Letters*, 7:1–10, 1986.
- I. Horowitz. Quantitative feedback theory. *IEE Proc. Pt. D*, 129:215–226, November 1982.
- I. Horowitz and A. Gera. Blending of uncertain nonminimum-phase plants for elimination or reduction of nonminimum-phase property. *International Journal of Systems Science*, 10:1007–1024, 1979.
- I. Horowitz and M. Sidi. Synthesis of feedback systems with large plant ignorance for prescribed time domain tolerances. *Int. J. Control*, 126:287–309, 1972.
- I. M. Horowitz. *Synthesis of of Feedback Systems*. Academic Press, New York, 1963.
- M. Hovd and S. Skogestad. Simple frequency-dependent tools for control system analysis, structure selection and design. *Automatica*, 28:989–996, 1992.
- M. Hovd and S. Skogestad. Pairing criteria for decentralized control of unstable plants. *Industrial and Engineering Chemistry Research*, 33:2134–2139, 1994.
- C. H. Hsu and C. T. Chen. A proof of the stability of multivariable feedback systems. *Proceedings of the IEEE*, 56:2061–2062, 1968.
- H. M. James, N. B. Nichols, and R. S. Philips. *Theory of Servomechanisms*. McGraw-Hill, New York, 1947.
- T. Kailath. *Linear Systems*. Prentice Hall, Englewood Cliffs, N. J., 1980.
- R. E. Kalman and R. S. Bucy. New results in linear filtering and prediction theory. *J. Basic Engineering, Trans. ASME Series D*, 83:95–108, 1961.

- N. Karcanias and C. Giannakopoulos. Necessary and sufficient conditions for zero assignment by constant squaring down. *Linear Algebra and its Applications*, 122/123/124:415–446, 1989.
- V. L. Kharitonov. Asymptotic stability of an equilibrium position of a family of systems of linear differential equations. *Differentsial'nye Uraveniya*, 14:1483–1485, 1978a.
- V. L. Kharitonov. On a generalization of a stability criterion. *Akademii Nauk Kahskoi SSR, Fiziko-matematicheskaya*, 1:53–57, 1978b.
- U. Korn and H. H. Wilfert. *Mehrgrößenregelungen. Moderne Entwurfsprinzipien im Zeit- und Frequenzbereich*. Verlag Technik, Berlin, 1982.
- A. M. Krall. An extension and proof of the root-locus method. *Journal Soc. Industr. Applied Mathematics*, 9:644–653, 1961.
- A. M. Krall. A closed expression for the root locus method. *Journal Soc. Industr. Applied Mathematics*, 11:700–704, 1963.
- A. M. Krall. The root-locus method: A survey. *SIAM Review*, 12:64–72, 1970.
- A. M. Krall and R. Fornaro. An algorithm for generating root locus diagrams. *Communications of the ACM*, 10:186–188, 1967.
- J. M. Krause. Comments on Grimbale's comments on Stein's comments on rolloff of H^∞ optimal controllers. *IEEE Trans. Auto. Control*, 37:702, 1992.
- H. Kwakernaak. Asymptotic root loci of optimal linear regulators. *IEEE Trans. Aut. Control*, 21(3):378–382, 1976.
- H. Kwakernaak. Robustness optimization of linear feedback systems. In *Preprints, 22nd IEEE Conference of Decision and Control, San Antonio, Texas, USA*, 1983.
- H. Kwakernaak. Minimax frequency domain performance and robustness optimization of linear feedback systems. *IEEE Trans. Auto. Control*, 30:994–1004, 1985.
- H. Kwakernaak. A polynomial approach to minimax frequency domain optimization of multi-variable systems. *Int. J. Control*, 44:117–156, 1986.
- H. Kwakernaak. The polynomial approach to \mathcal{H}_∞ -optimal regulation. In E. Mosca and L. Pandolfi, editors, H_∞ -Control Theory, volume 1496 of *Lecture Notes in Mathematics*, pages 141–221. Springer-Verlag, Heidelberg, etc., 1991.
- H. Kwakernaak. Robust control and H^∞ -optimization. *Automatica*, 29:255–273, 1993.
- H. Kwakernaak. Symmetries in control system design. In A. Isidori, editor, *Trends in Control — A European Perspective*, pages 17–51. Springer, Heidelberg, etc., 1995.
- H. Kwakernaak and R. Sivan. *Linear Optimal Control Systems*. Wiley, New York, 1972.
- H. Kwakernaak and R. Sivan. *Modern Signals and Systems*. Prentice Hall, Englewood Cliffs, NJ, 1991.
- H. Kwakernaak and J. H. Westdijk. Regulability of a multiple inverted pendulum system. *Control — Theory and Advanced Technology*, 1:1–9, 1985.

- I. D. Landau, F. Rolland, C. Cyrot, and A. Voda. Régulation numérique robuste: Le placement de pôles avec calibrage de la fonction de sensibilité. In A. Oustaloup, editor, *Summer School on Automatic Control of Grenoble*. Hermes, 1993. To be published.
- A. J. Laub and B. C. Moore. Calculation of transmission zeros using QZ techniques. *Automatica*, 14:557–566, 1978.
- D. K. Le, O. D. I. Nwokah, and A. E. Frazho. Multivariable decentralized integral controllability. *International Journal of Control*, 54:481–496, 1991.
- V. X. Le and M. G. Safonov. Rational matrix GCD's and the design of squaring-down compensators- a state-space theory. *IEEE Transactions on Automatic Control*, 37:384–392, 1992.
- J.-L. Lin, I. Postlethwaite, and D.-W. Gu. μ - K Iteration: A new algorithm for μ -synthesis. *Automatica*, 29:219–224, 1993.
- J. Lunze. Determination of robust multivariable I-controllers by means of experiments and simulation. *Systems Analysis Modelling Simulation*, 2:227–249, 1985.
- J. Lunze. *Robust Multivariable Feedback Control*. Prentice Hall International Ltd, London, UK, 1988.
- J. Lunze. *Robust Multivariable Feedback Control*. International Series in Systems and Control Engineering. Prentice Hall International, Hempsstead, UK, 1989.
- C. C. MacDuffee. *The Theory of Matrices*. Chelsea Publishing Company, New York, N.Y., USA, 1956.
- A. G. J. MacFarlane and N. Karcnias. Poles and zeros of linear multivariable systems: a survey of the algebraic, geometric and complex-variable theory. *International Journal of Control*, 24: 33–74, 1976.
- J. M. Maciejowski. *Multivariable Feedback Design*. Addison-Wesley Publishing Company, Wokingham, UK, 1989.
- T. J. McAvoy. *Interaction Analysis. Principles and Applications*. Instrument Soc. Amer., Research Triangle Park, NC, 1983.
- D. C. McFarlane and K. Glover. *Robust Controller Design Using Normalized Coprime Factor Plant Descriptions*, volume 146. Springer-Verlag, Berlin, etc., 1990.
- R. H. Middleton. Trade-offs in linear control system design. *Automatica*, 27:281–292, 1991.
- R. J. Minnichelli, J. J. Anagnost, and C. A. Desoer. An elementary proof of Kharitonov's stability theorem with extensions. *IEEE Trans. Aut. Control*, 34:995–1001, 1989.
- M. Morari. Robust stability of systems with integral control. *IEEE Transactions on Automatic Control*, 30:574–577, 1985.
- M. Morari and E. Zafiriou. *Robust Process Control*. Prentice Hall International Ltd, London, UK, 1989.
- C. N. Nett and V. Manousiouthakis. Euclidean condition and block relative gain: connections, conjectures, and clarifications. *IEEE Transactions on Automatic Control*, 32:405–407, 1987.

- B. Noble. *Applied Linear Algebra*. Prentice Hall, Englewood Cliffs, NJ, 1969.
- O. D. I. Nwokah, A. E. Frazho, and D. K. Le. A note on decentralized integral controllability. *International Journal of Control*, 57:485–494, 1993.
- H. Nyquist. Regeneration theory. *Bell System Technical Journal*, 11:126–147, 1932.
- K. Ogata. *Modern Control Engineering*. Prentice Hall, Englewood Cliffs, NJ, 1970.
- A. Packard and J. C. Doyle. The complex structured singular value. *Automatica*, 29:71–109, 1993.
- I. Postlethwaite, J. M. Edmunds, and A. G. J. MacFarlane. Principal gains and principal phases in the analysis of linear multivariable feedback systems. *IEEE Transactions on Automatic Control*, 26:32–46, 1981.
- I. Postlethwaite and A. G. J. MacFarlane. *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., 1979.
- I. Postlethwaite, M. C. Tsai, and D.-W. Gu. Weighting function selection in H^∞ design. In *Proceedings, 11th IFAC World Congress, Tallinn*. Pergamon Press, Oxford, UK, 1990.
- L. Qiu, B. Bernhardsson, A. Rantzer, E. J. Davison, P. M. Young, and J. C. Doyle. A formula for computation of the real stability radius. *Automatica*, 31:879–890, 1995.
- J. Raisch. *Mehrgrößenregelung im Frequenzbereich*. R. Oldenbourg Verlag, München, BRD, 1993.
- H. H. Rosenbrock. *State-space and Multivariable Theory*. Nelson, London, 1970.
- H. H. Rosenbrock. The zeros of a system. *International Journal of Control*, 18:297–299, 1973.
- H. H. Rosenbrock. *Computer-aided Control System Design*. Academic Press, London, 1974a.
- H. H. Rosenbrock. Correction on "The zeros of a system". *International Journal of Control*, 20: 525–527, 1974b.
- A. Saberi, B. M. Chen, and P. Sannuti. *Loop transfer recovery: Analysis and design*. Prentice Hall, Englewood Cliffs, N. J., 1993.
- A. Saberi, P. Sannuti, and B. M. Chen. \mathcal{H}_2 optimal control. Prentice Hall, Englewood Cliffs, N. J., 1995. ISBN 0-13-489782-X.
- M. G. Safonov and M. Athans. A multiloop generalization of the circle criterion for stability margin analysis. *IEEE Trans. Automatic Control*, 26:415–422, 1981.
- M. K. Sain and C. B. Schrader. The role of zeros in the performance of multiinput, multioutput feedback systems. *IEEE Transactions on Education*, 33:244–257, 1990.
- C. J. Savant. *Basic Feedback Control System Design*. McGraw-Hill Book Co., New York, NY, 1958.
- C. B. Schrader and M. K. Sain. Research on system zeros: A survey. *International Journal of Control*, 50:1407–1433, 1989.

- O. A. Sebakhy, M. ElSingaby, and I. F. ElArabawy. Zero placement and squaring problem: a state space approach. *International Journal of Systems Science*, 17:1741–1750, 1986.
- U. Shaked. Design techniques for high feedback gain stability. *International Journal of Control*, 24:137–144, 1976.
- L. M. Silverman. Decoupling with state feedback and precompensation. *IEEE Trans. Automatic Control*, AC-15:487–489, 1970.
- V. Sima. *Algorithms for linear-quadratic optimization*. Marcel Dekker, New York-Basel, 1996. ISBN 0-8247-9612-8.
- S. Skogestad, M. Hovd, and P. Lundström. Simple frequency-dependent tools for analysis of inherent control limitations. *Modeling Identification and Control*, 12:159–177, 1991.
- S. Skogestad and M. Morari. Implications of large RGA elements on control performance. *Industrial and Engineering Chemistry Research*, 26:2323–2330, 1987.
- S. Skogestad and M. Morari. Variable selection for decentralized control. *Modeling Identification and Control*, 13:113–125, 1992.
- S. Skogestad and I. Postlethwaite. *Multivariable Feedback Control. Analysis and Design*. John Wiley and Sons Ltd., Chichester, Sussex, UK, 1995.
- A. A. Stoorvogel. *The H-Infinity Control Problem: A State Space Approach*. Prentice Hall, Englewood Cliffs, NJ, 1992.
- A. A. Stoorvogel and J. H. A. Ludlage. Squaring down and the problems of almost zeros for continuous time systems. *Systems and Control Letters*, 23:381–388, 1994.
- F. Svaricek. *Zuverlässige numerische Analyse linearer Regelungssysteme*. B.G.Teubner Verlagsgesellschaft, Stuttgart, BRD, 1995.
- G. J. Thaler. *Design of Feedback Systems*. Dowden, Hutchinson, and Ross, Stroudsburg, PA, 1973.
- H. Tolle. *Mehrgrößen-Regelkreissynthese. Band I. Grundlagen und Frequenzbereichsverfahren*. R.Oldenbourg Verlag, München, BRD, 1983.
- H. Tolle. *Mehrgrößen-Regelkreissynthese. Band II: Entwurf im Zustandsraum*. R.Oldenbourg Verlag, München, BRD, 1985.
- H. L. Trentelman and A. A. Stoorvogel. Control theory for linear systems. Vakgroep Wiskunde, RUG, November 1993. Course for the Dutch Graduate Network on Systems and Control, Winter Term 1993–1994.
- J. G. Truxal. *Automatic Feedback Control System Synthesis*. McGraw-Hill Book Co., New York, NY, 1955.
- J. Van de Vegte. *Feedback control systems*. Prentice-Hall, Englewood Cliffs, N. J., 1990. Second Edition.
- A. I. G. Vardulakis. *Linear Multivariable Control: Algebraic Analysis and Synthesis Methods*. John Wiley and Sons Ltd., Chichester, Sussex, UK, 1991.

- M. Verma and E. Jonckheere. L_∞ -compensation with mixed sensitivity as a broad-band matching problem. *Systems & Control Letters*, 14:295–306, 1984.
- M. Vidyasagar, H. Schneider, and B. A. Francis. Algebraic and topological aspects of feedback stabilization. *IEEE Trans. Aut. Control*, 27:880–894, 1982.
- M. Vidyasagar. *Control Systems Synthesis—A Factorization Approach*. MIT Press, Cambridge, MA, 1985.
- T. Williams and P. J. Antsaklis. Decoupling. In *The Control Handbook*, pages 795–804. CRC press, 1996.
- T. W. C. Williams and P. J. Antsaklis. A unifying approach to the decoupling of linear multivariable systems. *Internat. J. Control*, 44(1):181–201, 1986.
- G. T. Wilson. A convergence theorem for spectral factorization. *J. Multivariate Anal.*, 8:222–232, 1978.
- W. A. Wolovich. *Linear Multivariable Systems*. Springer Verlag, New York, NY, 1974.
- E. Wong. *Introduction to random processes*. Springer-Verlag, New York, 1983.
- W. M. Wonham. *Linear Multivariable Control: A Geometric Approach*. Springer-Verlag, New York, etc, 1979.
- D. C. Youla, J. J. Bongiorno, and C. N. Lu. Single-loop feedback-stabilization of linear multivariable dynamical plants. *Automatica*, 10:159–173, 1974.
- G. Zames. Feedback and optimal sensitivity: Model reference transformations, multiplicative seminorms, and approximate inverses. *IEEE Trans. Aut. Control*, 26:301–320, 1981.
- K. Zhou, J. C. Doyle, and K. Glover. *Robust and Optimal Control*. Prentice Hall, Englewood Cliffs, 1996. ISBN 0-13-456567-3.
- J. G. Ziegler and N. B. Nichols. Optimum settings for automatic controllers. *Trans. ASME, J. Dyn. Meas. and Control*, 64:759–768, 1942.

Index

- ~, 152
- $1\frac{1}{2}$ -degrees-of-freedom, 51
- 2-degrees-of-freedom, 11
- acceleration constant, 63
- action, 295
- additive uncertainty model, 196
- adjoint
 - of rational matrix, 251
- algebraic Riccati equation, 137
- amplitude, 293
- ARE, 137
 - solution, 144
- asymptotic stability, 12
- Bézout equation, 16
- bandwidth, 28, 80
 - improvement, 8
- basic perturbation model, 195
- BIBO stability, 13
- Blaschke product, 41
- Bode
 - diagram, 70
 - gain-phase relationship, 36
 - magnitude plot, 70
 - phase plot, 70
 - sensitivity integral, 38
- Butterworth polynomial, 91
- certainty equivalence, 150
- characteristic polynomial, 14, 287
- classical control theory, 59
- closed-loop
 - characteristic polynomial, 15
 - system matrix, 139
 - transfer matrix, 33
- complementary
 - sensitivity function, 23
 - sensitivity matrix, 33
- complex conjugate transpose, 288
- constant disturbance model, 259
- constant disturbance rejection, 65
- constant error suppression, 259
- control
 - decentralized-, 124
- controllability gramian, 296
- controller
 - stabilizing-, 14
- crossover region, 32
- decentralized
 - control, 105, 124
 - integral action, 127
- decoupling
 - control, 106
 - indices, 118
- delay
 - margin, 22
 - time, 81
- derivative time, 66
- determinant, 287
- DIC, 127
- disturbance condition number, 124
- D - K iteration, 269
- D -scaling upper bound, 222
- edge theorem, 192
- eigenvalue, 287
 - decomposition, 287
- eigenvector, 287
- elementary column operation, 109
- elementary row operation, 109
- energy, 293
- equalizing property, 255
- error covariance matrix, 148
- error signal, 5
- Euclidean norm, 291
- feedback equation, 6
- filtering, 250
- forward compensator, 5
- Freudenberg-looze equality, 40

- full uncertainty, 203
- functional
 - controllability, 117
 - reproducibility, 117
- FVE, 81
- gain, 88
 - at a frequency, 69
 - crossover frequency, 80
 - margin, 21, 80
- generalized Nyquist criterion, 19
- generalized plant, 154, 248
- gridding, 189
- Guillemin-Truxal, 90
- \mathcal{H}_2 -norm, 153, 303
- \mathcal{H}_2 -optimization, 153
- Hadamard product, 125
- Hankel norm, 303
- Hermitian, 288
- high-gain, 6
- \mathcal{H}_∞ -norm, 295
- \mathcal{H}_∞ -problem, 248
 - equalizing property, 255
- Hurwitz
 - matrix, 193
 - polynomial, 186
 - strictly-, 119
 - tableau, 187
- induced norm, 294
- inferential control, 156
- ∞ -norm
 - of transfer matrix, 295
- ∞ -norm
 - of transfer matrix, 198
 - of vector, 291
- input-output pairing, 124
- input sensitivity
 - function, 30
 - matrix, 33
- integral control, 66
- integrating action, 64
- integrator in the loop, 160
- intensity matrix, 146
- interaction, 105
- interconnection matrix, 195
- internal model principle, 65
- internal stability, 13, 297
- ITAE, 91
- Kalman-Yakubovič-Popov equality, 140
- Kalman filter, 148
- Kharitonov's theorem, 190
- KYP
 - equality, 140
 - inequality, 140
- \mathcal{L}_2 -norm, 293
- lag compensation, 84
- leading coefficient, 16
- lead compensation, 82
- linearity improvement, 8
- linear regulator problem
 - stochastic, 146
- \mathcal{L}_∞ -norm, 293
- loop
 - gain, 18
 - gain matrix, 18
 - IO map, 6, 10
 - shaping, 34, 82
 - transfer matrix, 15
 - transfer recovery, 152
- LQ, 136
- LQG, 145
- LTR, 152
- Lyapunov equation, 148
- Lyapunov matrix differential equation, 170
- margin
 - delay-, 22
 - gain-, 21, 80
 - modulus-, 21
 - multivariable robustness-, 218
 - phase-, 21, 80
- matrix, 287
 - closed loop system-, 139
 - complementary sensitivity-, 33
 - determinant, 287
 - error covariance-, 148
 - Hamiltonian, 145
 - Hermitian, 288
 - input sensitivity-, 33
 - intensity-, 146
 - loop transfer-, 15
 - nonnegative definite-, 288
 - nonsingular, 287
 - observer gain-, 147

- positive definite-, 288
- positive semi-definite-, 288
- proper rational-, 18
- rank, 287
- sensitivity-, 33
- singular, 287
- state feedback gain-, 137
- symmetric, 288
- system-, 14
- trace, 287
- unitary, 288
- M -circle, 75
- measurement noise, 31
- MIMO, 11, 33, 103
- minimal realization, 113
- minimum phase, 38
- mixed sensitivity problem, 240
 - type k control, 242
- modulus margin, 21
- monic, 141
- μ , 219
- μ -synthesis, 267
- μ_H , 220
- multiloop control, 124
- multiplicative uncertainty model, 196
- multivariable dynamic perturbation, 217
- multivariable robustness margin, 218
- N -circle, 75
- Nehari problem, 282
- Nichols
 - chart, 78
 - plot, 78
- nominal, 33
- nonnegative definite, 288
- nonsingular matrix, 287
- norm, 291
 - Euclidean, 291
 - Hankel, 303
 - induced, 294
 - ∞ -, 295
 - ∞ -, 198
 - \mathcal{L}_p -, 293
 - matrix, 302
 - on \mathbb{C}^n , 291
 - p -, 291
 - spectral-, 292
 - submultiplicative property, 302
- normal rank
 - polynomial matrix, 109
 - rational matrix, 110
- notch filter, 84
- Nyquist
 - plot, 18, 74
 - stability criterion, 18
- observability gramian, 296
- observer, 147
 - gain matrix, 147
 - poles, 151
- open-loop
 - characteristic polynomial, 15
 - pole, 88
 - stability, 20
 - zero, 88
- optimal linear regulator problem, 136
- order
 - of realization, 113
- output
 - controllability, 117
 - functional reproducible, 117
 - principal directions, 122
- overshoot, 81
- parametric perturbation, 184
- parasitic effect, 32
- parity interlacing property, 20
- partial pole placement, 244
- peak value, 293
- perturbation
 - multivariable dynamic, 217
 - parametric, 184
 - repeated real scalar, 217
 - repeated scalar dynamic, 217
- phase
 - crossover frequency, 80
 - margin, 21, 80
 - minimum-, 38
 - shift, 69
- PI control, 66
- PID control, 66
- plant capacity, 31
- p -norm, 291
 - of signal, 293
 - of vector, 291
- pole
 - assignment, 16
 - excess, 242

- of MIMO system, 111
- placement, 16
- pole-zero excess, 74
- polynomial
 - characteristic-, 14
 - invariant-, 109
 - matrix, Smith form, 109
 - matrix fraction representation, 211
 - monic, 141
 - unimodular matrix, 108
- position error constant, 62
- positive
 - definite, 288
 - semi definite, 288
- principal direction
 - input, 122
 - output, 122
- principal gain, 122
- proper, 18
 - rational function, 15
 - strictly-, 16
- proportional
 - control, 66
 - feedback, 7
 - uncertainty model, 196
- pure integral control, 66
- QFT, 93
- quantitative feedback theory, 93
- rank, 287
- rational matrix
 - fractional representation, 211
 - Smith McMillan form, 110
- realization
 - minimal, 113
 - order of, 113
- regulator
 - poles, 151
 - system, 2
- relative degree, 74, 242
- relative gain array, 125
- repeated real scalar perturbation, 217
- repeated scalar dynamic perturbation, 217
- reset time, 66
- resonance
 - frequency, 80
 - peak, 80
- return compensator, 5

- return difference, 139
 - equality, 140, 171
 - inequality, 140
- RGA, 125
- rise time, 81
- robustness, 7
- robust performance, 226
- roll-off, 32, 87
 - LQG, 160
 - mixed sensitivity, 242
- root loci, 89
- Schur complement, 180, 289
- sector bounded function, 205
- sensitivity
 - complementary-, 23
 - function, 23
 - input-, 30
 - matrix, 33
 - matrix, complementary-, 33
 - matrix, input-, 33
- separation principle, 150
- servocompensator, 128
- servomechanism, 128
- servo system, 2
- settling time, 81
- shaping
 - loop-, 82
- Sherman-Morrison-Woodbury, 289
- $\bar{\sigma}$, $\underline{\sigma}$, 292
- signal
 - amplitude, 293
 - energy, 293
 - \mathcal{L}_2 -norm, 293
 - \mathcal{L}_∞ -norm, 198
 - \mathcal{L}_∞ -norm, 293
 - peak value, 293
- single-degree-of-freedom, 12
- singular matrix, 287
- singular value, 292
 - decomposition, 292
- SISO, 11
- small gain theorem, 198
- Smith-McMillan form, 110
- Smith form, 109
- spectral norm, 292
- square down, 111
- stability, 11
 - asymptotic-, 12

- BIBO-, 13
- internal, 13, 297
- matrix, 296
- Nyquist (generalized), 19
- Nyquist criterion, 18
- of polynomial, 192
- of transfer matrix, 297
- open-loop, 20
- radius, 234
- radius, complex, 234
- radius, real, 234
- region, 187
- stabilizing, 14
- standard \mathcal{H}_∞ problem, 248
- standard \mathcal{H}_2 problem, 154
- state
 - estimation error, 147
 - feedback gain matrix, 137
- state-space realization, 113
- static output feedback, 114
- steady-state, 60
 - error, 60
- steady-state mean square error, 150
- step function
 - unit, 60
- strictly proper, 16
- structured singular value, 219
 - D -scaling upper bound, 222
 - lower bound, 221, 222
 - of transfer matrix, 220
 - scaling property, 220
 - upper bound, 221
- submultiplicative property, 302
- suboptimal solution (of \mathcal{H}_∞ problem), 251
- SVD, 292
- Sylvester equation, 16
- symmetric, 288
- system matrix, 14
- trace, 149, 287
- tracking system, 2
- transmission zero
 - rational matrix, 110
- triangle inequality, 291
- two-degrees-of-freedom, 11
- type k control
 - mixed sensitivity, 242
- type k system, 61
- uncertainty
 - full, 203
 - unstructured, 195
- uncertainty model
 - additive, 196
 - multiplicative, 196
 - proportional, 196
- unimodular, 108
- unit
 - feedback, 5
 - step function, 60
- unitary, 288
- unstructured uncertainty, 195
- vector space
 - normed, 291
- velocity constant, 62
- well-posedness, 297
- zero
 - open loop, 88
 - polynomial matrix, 109