# Feedback Systems

An Introduction for Scientists and Engineers

SECOND EDITION

Karl Johan Åström
Richard M. Murray

Version v3.0j (2019-08-18)

This is the electronic edition of *Feedback Systems* and is available from http://fbsbook.org. Hardcover editions may be purchased from Princeton University Press, http://press.princeton.edu/titles/8701.html.

This manuscript is for personal use only and may not be reproduced, in whole or in part, without written consent from the publisher (see http://press.princeton.edu/permissions.html).

# Chapter Fourteen

## Fundamental Limits

*Many people have seen theoretical advantages in the facts that front-drive rear-steering recumbent bicycles would have simpler transmissions than rear-drive recumbents and could have the center of mass nearer to the front wheel than the rear. The U.S. Department of Transportation commissioned the construction of a safe motorcycle with this configuration. It turned out to be safe in an unexpected way: No one could ride it.*

F. R. Whitt and D. G. Wilson, *Bicycling Science* [WW97]

In this chapter we will discuss properties that limit performance and robustness of control systems. Non-minimum phase dynamics, due to time delays and right half-plane poles and zeros impose severe limits. There are also nonlinear behaviors that appear at large and small signal levels. Large signal limits are caused by limited rate and power of actuators, or by constraints required to protect the process. Small signal limits are caused by measurement noise, friction, and quantization in converters. We will also discuss consequences of the limits for loop shaping, and give rules pole placement design.

### 14.1 SYSTEM DESIGN CONSIDERATIONS

The initial design of a system can have a significant impact on the ability to use feedback to provide robustness and performance improvements. It is particularly important to recognize fundamental limits in the performance of feedback systems early in the design process. For example, we may expect that a system with time delays cannot admit fast control because control actions are delayed. Similarly it seems reasonable that unstable systems will require fast controllers, which will depend on the bandwidth of sensors and actuators. These limits are caused by properties of the system dynamics and can often be captured by conditions on the poles and zeros of the process.

The freedom for the control designer depends very much on the situation. The designer can be faced with a process with given actuators and actuators and his or her task is to design a suitable controller. The designer then has limited freedom. In other cases he or she may be able to choose sensors, and in yet other cases the location and characteristics of sensors, actuators, and controller are designed simultaneously. The designer then has significant freedom. The typical case is somewhere in between these extremes.

In any case, design engineers should be aware of the fundamental limits of feedback systems and be able to deal with them as early as possible in the design process. Awareness of the limits and co-design of the process and the controller are good ways to avoid potential difficulties both for system and control designers.

The limits alluded to in the chapter quote are due to process dynamics and limits on actuation power and actuation rate. The dynamics limitations can be captured by time delays and poles and zeros in the right half-plane. It seems intuitively clear that a time delay in the process limits the achievable response speed. A less obvious case is that a process with a right half-plane pole/zero pair cannot be controlled robustly if the pole is close to the zero. Restriction in actuation can be captured by actuation power and actuation rates. These are all examples of fundamental limits whose potential impacts should be taken into account during initial system design.

**Stabilizability and Strong Stabilizability**

One of the most fundamental properties of a control system is the ability to design the dynamics of the (closed loop) system to meet a set of performance specifications. Often this can be captured by the location of poles and zeros in the relevant transfer functions, such as the Gang of Four. In Section 7.2 we found that a system must be reachable in order to find a state feedback that places closed-loop eigenvalues in arbitrary positions. The corresponding condition for output feedback is that the system is be both reachable and observable (Section 8.3). There are also trade-offs that are captured by the stability margin, bandwidth, peak values and locations of sensitivity functions, and many other features that we have encountered in the previous chapters.

One question of particular interest for system whose natural dynamics are unstable is when a system can be stabilized using feedback and whether it can be stabilized using a *stable* controller (a condition known as *strong stabilizability*). The question of stabilizability is slightly different than reachability since it may turn out that there are stable eigenvalues that cannot be modified by feedback, but we can still modify all unstable eigenvalues. Strong stabilizability is important for system-level design since we may not want to implement an unstable controller unless it is necessary to do so. (Note that a just having the controller be unstable does not mean that the closed loop system is unstable.)

A linear system with state feedback is always stabilizable if it is reachable. If a linear system is not reachable, it follows from Kalman's decomposition theorem (Section 8.3) that the system dynamics can be written as

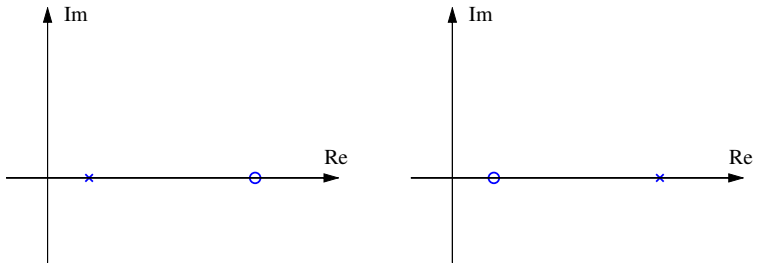
$$\frac{dx}{dt} = \frac{d}{dt}\begin{pmatrix} x_{\mathrm{r}} \\ x_{\bar{\mathrm{r}}} \end{pmatrix} = \begin{pmatrix} A_{\mathrm{r}} & 0 \\ * & A_{\bar{\mathrm{r}}} \end{pmatrix} \begin{pmatrix} x_{\mathrm{r}} \\ x_{\bar{\mathrm{r}}} \end{pmatrix} + \begin{pmatrix} B_{\mathrm{r}} \\ 0 \end{pmatrix} u, \tag{14.1}$$

where the states have been decomposed into two parts: the reachable states $x_{\mathrm{r}}$ and the unreachable states $x_{\bar{\mathrm{r}}}$. The dynamics in the invariant subspace represented $x_{\mathrm{r}}$ are reachable and it follows that we can always find a state feedback $K_{\mathrm{r}}$ such that $A_{\mathrm{r}} - K_{\mathrm{r}} B_{\mathrm{r}}$ has arbitrary eigenvalues. The system (14.1) is then stabilizable if and only if the eigenvalues of $A_{\bar{\mathrm{r}}}$ are in the left half-plane. A system with state feedback is thus stabilizable if the unreachable part of the system is stable.

Reachability and stabilizability for systems with state feedback can also be stated as a rank condition. A system with dynamics and control matrices $A$, $B$ having $n$ state variables is reachable if and only if

$$\operatorname{rank} \begin{pmatrix} A - sI & B \end{pmatrix} = n \tag{14.2}$$

for all values of $\lambda \in \mathbb{C}$. It is stabilizable if the condition holds for all $\lambda$ in the right




(a) Strongly stabilizable (b) Not strongly stabilizable

**Figure 14.1:** Pole zero diagrams for strongly stabilizable and non-strongly stabilizable systems. The system in (a) can be stabilized with a stable controller, but stabilization of the system in (b) requires a controller with a pole in the right half-plane.

half-plane $\mathrm{Re}\, s \geq 0$ ($\mathrm{Re}\, s > 0$ for strict stabilizability). Stabilizabilty for systems with state feedback is thus a weaker condition than reachability.

For a linear system with output feedback a controller can be constructed using an estimator and linear feedback on the state estimate, and the resulting controller has the transfer function given in equation (8.16), repeated here:

$$\frac{d\hat{x}}{dt} = A\hat{x} + Bu + L(y - C\hat{x}), \qquad u = -K\hat{x}. \tag{14.3}$$

The controller poles are the eigenvalues of the matrix $A - BK - LC$, and the controller zeros are the values of $s$ where the matrix

$$\begin{pmatrix} A - BK - LC - sI & L \\ K & 0] \end{pmatrix} \tag{14.4}$$

loses rank. If a system is stabilizable and observable, it is always possible to construct a stabilizing controller. However, the question about whether this controller is stable (corresponding to *strong* stabilizability is more subtle. Strong stabilizability can be expressed as conditions on the transfer function, as described in the following theorem.

**Theorem 14.1** (Strong stabilizability)**.** *Consider a linear system with the rational transfer function $P(s) = n(s)/d(s)$, where the polynomials $n(s)$ and $d(s)$ do not have a factor in common. The system can be strongly stabilized if and only if all $d(z_k)$ have the same sign for all $z_k$ such that $n(z_k) = 0$.*

This theorem is proven in Vidyasagar [Vid85, Theorem 3.1 and Corollary 3.3]) (see also Youla [YJJBL74]). For a system with a single pole/zero pair, this result implies that a process with $p > z$ requires a controller with a pole in the right half-plane, hence an *unstable* controller. This situation is illustrated in Figure 14.1. An example is given in Exercise 14.21. The root locus method gives significant insight into these cases (Exercise 14.15).

Another characterization of strong stabilizability is given in Doyle, Frances, and Tannenbaum [DFT92, Theorem 3, Chapter 5]:

**Theorem 14.2.** *A linear system $P$ is strongly stabilizable if and only if it has an*

*even number of real poles between every pair of real zeros in* $\Re\, s \geq 0$.

These two results show that strong stabilizability depends on the *patterns* of poles and zeros, which are often determined in the early stages of system design. Note that this does not imply that unstable systems should always be avoided, because instability may actually have advantages; typical examples are when high maneuverability is desired, as in high-performance aircraft.

### Right Half-Plane Zeros and Time Delays

In addition to questions related to stabilizability, we will see throughout this chapter that there can significant limitations on closed loop performance when a system has zeros in the right half-plane or time delays in the loop transfer function. A natural question to ask is whether these features can be avoided at the time of system design.

The poles of a system depend on the intrinsic dynamics of the system; they represent the modes of the system and they are given by the eigenvalues of the dynamics matrix $A$ of the linearized model. Sensors and actuators have no effect on the poles; the only way to change poles is by feedback or by redesign of the process. However, the zeros of a system depend on how the sensors and actuators are connected to the process. Zeros can thus be changed by moving or adding sensors and actuators, which is often simpler than redesigning the process dynamics.

The following example illustrates how the location of zeros can be determine through placement of sensors.

**Example 14.1 Vehicle steering**
Consider the vehicle steering system introduced in Example 3.11. The linearized (but non-normalized) model of the dynamics of the system relating lateral velocity to steering angle were given in Example 10.11 and have the form

$$P(s) = \frac{av_0 s + v_0^2}{bs},$$

where $v_0$ is the velocity of the vehicle, $a$ is the offset to the reference point for the vehicle position, and $b$ is the wheelbase. We observed that the system has a right half-plane zero when the velocity of the vehicle is negative and this can lead to limits in the closed loop performance of the system, such as those described in Example 10.11.

The existence of the right half-plane zero can be removed if we choose to measure the location of the vehicle by the position of the center of the rear wheels instead of the center of mass. This gives $a = 0$ and our dynamics become

$$G(s) = \frac{v_0^2}{bs},$$

which no longer has a right half-plane zero. Choosing this output can simplify the design constraints and is easily implemented by calibrating the position sensor for the vehicle so that it returns the position of the center of the rear wheels.

We note that this choice of "sensor" is subject to calibration errors and this can lead to a zero of the process transfer function at $v_0/\epsilon$, where $\epsilon$ represents the calibration error and the sign of the zero depends on the sign of the calibration error

and the direction of travel. We will see later in the chapter that this corresponds that what we call a "fast" zero and its impact on fundamental limits is relatively minor. It can thus it can be advantageous to choose the system output to be at a different point in order to simplify the feedback controller design. $\nabla$

Another source of limitations is due to time delays, which can add significant phase lag to the loop transfer function, making it difficult to maintain sufficient phase margin. Time delays may appear in the process, in communication channels, and in computations. Time delays have effects similar to right half-plane zeros. One way to see this is to consider the Padé approximation for a time delay. The first and second order Padé approximations are given by

$$G_1(s) = \frac{1 - s\tau/2}{1 + s\tau/2}, \qquad G_(s) = \frac{1 - \tau s/2 + (\tau s)^2/12}{1 + \tau s/2 + (\tau s)^2/12}.$$

The first-order Padé approximation has a right half plane zero at $2/\tau$ and the second order Padé approximation has a complex conjugate pair of right half-plane zeros at $s = (3 \pm i\sqrt{3})/\tau$.

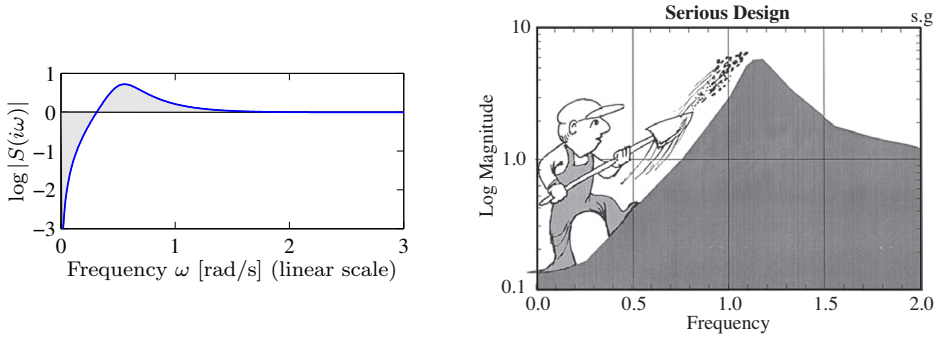
Unlike zeros, time delays cannot generally be avoided by choice of sensor or actuator location, and hence they should be avoided by proper design of the system's computing and communications architecture. Minimizing time delays whenever possible is usually a good design guideline for feedback control systems.

## 14.2 BODE'S INTEGRAL FORMULA

One of the most important limits in feedback control design was obtained by Bode, who showed that it was not possible to uniformly improve the performance of certain closed loop performance characteristics. Bode's result makes use of the the sensitivity function $S$ introduced in Section 12.1, which gives an overview of performance and robustness of a closed loop system. Specifically, it describes how disturbances are attenuated by feedback and allows comparison of disturbance attenuation of open and closed loop systems. We recall that disturbances with frequency $\omega$ are attenuated by feedback if $|S(i\omega)| < 1$, and disturbances with frequencies such that $|S(i\omega)| > 1$ are amplified. The maximum sensitivity $M_{\rm s} = \max_\omega |S(i\omega)|$ gives the largest amplification and is also a robustness measure, since $1/M_{\rm s}$ is equal to the stability margin $s_{\rm m}$ (see Figure 10.12).

A key observation is that the sensitivity function cannot be made small over a wide frequency range. There is an invariant (conserved quantity) called *Bode's integral formula* that implies that reducing the sensitivity at one frequency increases it at another, and the situation is worse if the process has right half-plane poles. Control design is thus always a compromise. The following theorem captures limits of performance under feedback.

**Theorem 14.3** (Bode's integral formula)**.** *Let $S(s)$ be the sensitivity function of an internally stable closed loop system with loop transfer function $L(s)$. Assume that the loop transfer function $L(s)$ is such that $sL(s)$ goes to zero as $s \to \infty$. Then*




(a) Bode integral formula (b) Control design process

**Figure 14.2:** Interpretation of the *waterbed effect*. The function $\log|S(i\omega)|$ is plotted versus $\omega$ in linear scales in (a). According to Bode's integral formula (14.5), the area of $\log|S(i\omega)|$ above zero must be equal to the area below zero. Gunter Stein's interpretation of design as a trade-off of sensitivities at different frequencies is shown in (b) (from [Ste03]).

*the sensitivity function has the property*

$$\int_0^\infty \log|S(i\omega)|\,d\omega = \int_0^\infty \log\frac{1}{|1+L(i\omega)|}\,d\omega = \pi\sum p_k. \tag{14.5}$$

*where the sum is over the right half-plane poles $p_k$ of $L(s)$.*

Equation (14.5) implies that if we design a controller that decreases the effect of disturbances for some frequencies it will increase the effect for other frequencies because the integral of $\log|S(i\omega)|$ remains constant. This property is sometimes referred to as the *waterbed effect*. It also follows that systems with open loop poles in the right half-plane have larger overall sensitivity than stable systems.

Equation (14.5) can be regarded as a *conservation law*: if the loop transfer function has no poles in the right half-plane, the equation simplifies to

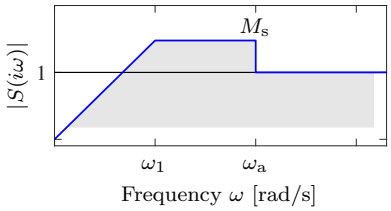
$$\int_0^\infty \log|S(i\omega)|d\omega = 0.$$

This formula can be given a nice geometric interpretation as illustrated in Figure 14.2, which shows $\log|S(i\omega)|$ as a function of $\omega$. The area over the horizontal axis must be equal to the area under the axis when the frequency is plotted on a *linear* scale. Thus if we wish to make the sensitivity smaller up to some frequency $\omega_{sc}$, we must balance this by increased sensitivity above $\omega_{sc}$. Control system design can be viewed as trading the disturbance attenuation at some frequencies for disturbance amplification at other frequencies. Notice that the assumption $\lim_{s\to\infty} sL(s) = 0$ is essential. Exercise 14.3 shows that without this assumption the sensitivity can be made arbitrarily small. A modification that covers $\lim_{s\to\infty} sL(s) = k$ is given in Exercise 14.4.

An equation similar to equation (14.5) holds for the complementary sensitivity function:

$$\int_0^\infty \frac{\log|T(i\omega)|}{\omega^2}\,d\omega = \pi\sum\frac{1}{z_i}, \quad T(s) = \frac{L(s)}{1+L(s)}, \tag{14.6}$$

(a) X-29 aircraft




(b) Sensitivity analysis

**Figure 14.3:** X-29 flight control system. The aircraft makes use of forward swept wings and a set of canards on the fuselage to achieve high maneuverability (a). The desired sensitivity for the closed loop system is shown in (b). We seek to use our control authority to shape the sensitivity curve so that we have low sensitivity (good performance) up to frequency $\omega_1$ by creating higher sensitivity up to our actuator bandwidth $\omega_a$.

where the summation is over all right half-plane zeros of the loop transfer function $L(s) = P(s)C(s)$ (Exercise 14.5). It follows from equation (14.6) that slow right half-plane zeros are worse than fast ones, just as equation (14.5) implies that fast right half-plane poles are worse than slow ones.

**Example 14.2 The X-29 aircraft**

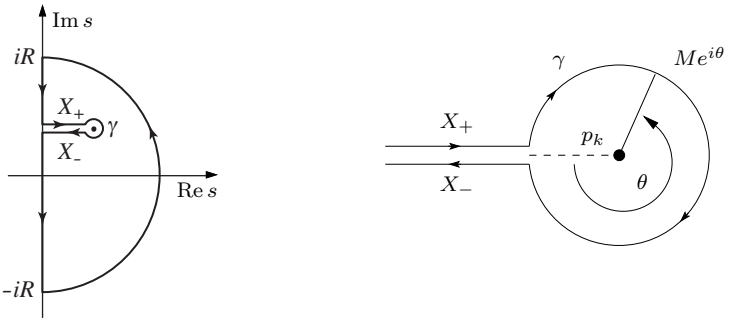
As an illustration of Bode's integral formula, we present an analysis of the control system for the X-29 aircraft (see Figure 14.3a), which has an unusual configuration of aerodynamic surfaces that is designed to enhance its maneuverability. This analysis was originally carried out by Gunter Stein in his inaugural IEEE Bode lecture "Respect the Unstable" [Ste03].

To analyze the system, we make use of a small set of parameters that describe the key properties of the system. A typical robustness requirement in aerospace systems is that the phase margins should be at least $\varphi_m = 45^\circ$. The X-29 has longitudinal dynamics that are similar to inverted pendulum dynamics (Exercise 9.5). It has a right half-plane pole at approximately $p = 6$ rad/s and a right half-plane zero at $z = 26$ rad/s. The actuators that stabilize the pitch have a bandwidth of $\omega_a = 40$ rad/s and the desired bandwidth of the pitch control loop is $\omega_1 = 3$ rad/s.

To evaluate the achievable performance, we search for a control law such that the sensitivity function is small up to the desired bandwidth and not greater than $M_s$ beyond that frequency. Because of Bode's integral formula, we know that $M_s$ must be greater than 1 at high frequencies to balance the small sensitivity at low frequency. We thus ask if we can find a controller that has the shape shown in Figure 14.3b with the smallest possible value of $M_s$. Note that the sensitivity above the frequency $\omega_a$ is 1 since we have no actuator authority above those frequencies. Thus, we desire to design a closed loop system that has low sensitivity at frequencies below $\omega_1$ and sensitivity that is not too large between $\omega_1$ and $\omega_a$.

From Bode's integral formula, we know that whatever controller we choose, equation (14.5) must hold. We will assume that the sensitivity function is given by

$$|S(i\omega)| = \begin{cases} \frac{\omega}{\omega_1} M_s & \text{if } \omega < \omega_1, \\ M_s & \text{if } \omega_1 \le \omega < \omega_a, \\ 1 & \text{if } \omega_a \le \omega < \infty, \end{cases}$$




(a) Closed contour (b) Open loop pole

**Figure 14.4:** Contour used to prove Bode's theorem. For each right half-plane pole $p_k$ of the loop transfer function $L(s)$, which is also a singularity of $\log S(s)$, we create a path from the imaginary axis that encircles the pole. To avoid clutter only one of the paths is shown.

corresponding to Figure 14.3b. Bode's integral becomes

$$\begin{aligned}\int_0^\infty \log|S(i\omega)|\,d\omega &= \int_0^{\omega_a} \log|S(i\omega)|\,d\omega \\ &= \int_0^{\omega_1} \log\frac{\omega M_s}{\omega_1}\,d\omega + (\omega_a - \omega_1)\log M_s = \pi p.\end{aligned}$$

Integration by parts gives, after some calculation, $-\omega_1 + \omega_a \log M_s = \pi p$ or

$$M_s = e^{(\pi p + \omega_1)/\omega_a}.$$

This formula tells us what the achievable value of $M_s$ will be for the given control specifications. In particular, using $p = 6$ rad/s, $\omega_1 = 3$ rad/s and $\omega_a = 40$ rad/s, we find that $M_s = 1.75$, which means that in the range of frequencies between $\omega_1$ and $\omega_a$, disturbances at the input to the process dynamics (such as wind) will be amplified by a factor of 1.75 in terms of their effect on the aircraft. With $M_s = 1.75$ we can also obtain an estimate of the phase margin as $\varphi_m \geq 2\arcsin 1/(2M_s) = 33^\circ$ (equation (10.7)), which indicates that the requirement $\varphi_m = 45^\circ$ may not be achievable. ∇

### Derivation of Bode's Formula

Bode's integral formula (Theorem 14.3) can be derived by contour integration. We assume that the loop transfer function has distinct poles at $s = p_k$ in the right half-plane and that $L(s)$ goes to zero faster than $1/s$ for large values of $s$. Consider the integral of the logarithm of the sensitivity function $S(s) = 1/(1 + L(s))$ along the contour $\Gamma$ shown in Figure 14.4. The contour encloses the right half-plane except for the points $s = p_k$ where the loop transfer function $L(s) = P(s)C(s)$ has poles and the sensitivity function $S(s)$ therefore has singularities (only one $p_k$ is shown in the figure). The direction of the contour is counterclockwise. The integral of the

logarithm of the sensitivity function around the contour $\Gamma$ is given by

$$I = \int_\Gamma \log(S(s))\,ds = I_1 + I_2 + I_3 = 0.$$

The integral $I$ is zero because the function $\log S(s)$ is analytic with no poles or zeros inside the contour. The term $I_1$ is the integral along the imaginary axis, the term $I_2$ is the integral along a large semicircle to the right with a radius $R$ that we will make infinitely large. The term $I_3$ is the integral along two parallel horizontal lines and a small circle enclosing $p_k$ as shown in Figure 14.4.

We now compute each of the terms in the contour integration. We have

$$I_1 = -i\int_{-R}^{R} \log(S(i\omega))d\omega = -2i\int_0^R \log(|S(i\omega)|)d\omega$$

because the real part of $\log S(i\omega)$ is an even function and the imaginary part is an odd function. Furthermore we have

$$I_2 = \int_\supset \log(S(s))\,ds = -\int_\supset \log(1 + L(s))\,ds \approx -\int_\supset L(s)\,ds,$$

where $\supset$ represents the semicircular portion of $\Gamma$ at radius $R$. Since $L(s)$ goes to zero faster than $1/s$ for large $s$, the integral goes to zero when the radius of the semicircle goes to infinity.

Next we consider the integral $I_3$. We split the contour into three parts: $X_+$, $\gamma$, and $X_-$, where $X_+$ and $X_-$ are horizontal lines from the imaginary axis to $p_k$, and $\gamma$ is a small circle with radius $r$ around the point $p_k$ (see Figure 14.4b). We can write the contour integral as

$$I_3 = \int_{X_+} \log S(s)\,ds + \int_\gamma \log S(s)\,ds + \int_{X-} \log S(s)\,ds$$

The point $p_k$ is a pole of $L(s)$ and hence a zero of $S(s)$, which causes $\log S(s)$ to become singular at $p_k$. The contour $\gamma$ for the middle integral is a small circle with radius $r$ around $p_k$. The magnitude of the integrand is of the order $\log r$ and the length of the path is $2\pi r$, and it can be shown that the magnitude of the integral goes to zero as the radius $r$ goes to zero. At the same time, $S(s) \approx k(s - p_k)$ near $p_k$, so the argument of $\log S(s)$ decreases by $2\pi$ as the contour encircles $p_k$ (in the clockwise direction). On the contours $X_+$ and $X_-$ we thus have

$$|S_{X_+}| = |S_{X_-}|, \qquad \arg S_{X_-} = \arg S_{X_+} - 2\pi.$$

Hence

$$\log(S_{X_+}) - \log(S_{X_-}) = \log(|S_{X_+}|) + i\,\arg(S_{X_+}) - \log(|S_{X_-}|) - i\,\arg(S_{X_-}) = 2\pi i.$$

Using the fact that the path $X_+$ is traversed in the opposite direction from $X_-$, the first and third terms can be combined to give

$$\int_{X_+} \log S(s)\,ds + \int_{X_-} \log S(s)\,ds = \int_{X_+} \big(\log S_{X_+}(s) - \log S_{X_-}(s)\big)\,ds.$$

The length of the path from the imaginary axis to $p_k$ is $\operatorname{Re} p_k$ and we get

$$\int_{X_+} \log S(s)\, ds + \int_{X_-} \log S(s)\, ds = 2\pi i \cdot \operatorname{Re} p_k.$$

Repeating the argument for all $p_k$ in the right half-plane, and letting the small circles go to zero gives

$$I_1 + I_2 + I_3 = -2i \int_0^\infty \log |S(i\omega)|\, d\omega + i \sum_k 2\pi \operatorname{Re} p_k = 0.$$

Since the $p_k$'s appear as complex conjugate pairs, we have $\sum_k \operatorname{Re} p_k = \sum_k p_k$, which gives Bode's formula (14.5).

## 14.3 GAIN CROSSOVER FREQUENCY INEQUALITY

We will now investigate the effect of non-minimum phase process dynamics for loop shaping design. The key idea of loop shaping design is to shape the loop transfer function $L(i\omega) = P(i\omega)C(i\omega)$ so that the closed loop system has good performance and robustness. Good performance is obtained by making $|L(i\omega)|$ large for frequencies where we want disturbance attenuation and small for high frequencies where measurement noise dominates. Recall from Figure 12.8 that good robustness is obtained by shaping the loop transfer function around the gain crossover frequency $\omega_{\mathrm{gc}}$. The performance limits show up very clearly in the design.

To explore the limits due to right half-plane poles and zeros, we factor the process transfer function as

$$P(s) = P_{\mathrm{mp}}(s) P_{\mathrm{ap}}(s), \tag{14.7}$$

where $P_{\mathrm{mp}}$ is the minimum phase factor and $P_{\mathrm{ap}}$ is the non-minimum phase factor, We do the factorization so that $P_{\mathrm{mp}}$ has all its poles and zeros in the open left half-plane. The factorization is normalized so that $|P_{\mathrm{ap}}(i\omega)| = 1$, and the sign is chosen so that $P_{\mathrm{ap}}$ has negative phase. The transfer function $P_{\mathrm{ap}}$ is called an *all-pass system* because it has unit gain for all frequencies. For example

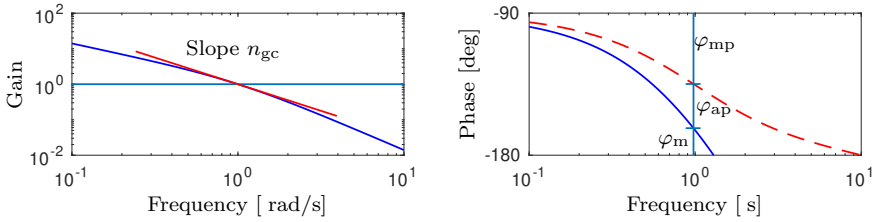
$$P(s) = \frac{s-2}{(s+1)(s-1)} = \frac{s+2}{(s+1)^2} \cdot \frac{(s-2)(s+1)}{(s+2)(s-1)} = P_{\mathrm{mp}}(s) \cdot P_{\mathrm{ap}}(s). \tag{14.8}$$

Since $|P_{\mathrm{ap}}(i\omega)| = 1$, the transfer functions $P(s)$ and $P_{\mathrm{mp}}(s)$ have the same gain curves but the transfer function $P(s)$ has larger phase lag than $P_{\mathrm{mp}}(s)$.

Consider the closed loop system obtained with a controller with the transfer function $C(s)$. Requiring that the phase margin be $\varphi_{\mathrm{m}}$, we get the inequality

$$\arg L(i\omega_{\mathrm{gc}}) = \arg P_{\mathrm{ap}}(i\omega_{\mathrm{gc}}) + \arg P_{\mathrm{mp}}(i\omega_{\mathrm{gc}}) + \arg C(i\omega_{\mathrm{gc}}) \geq -\pi + \varphi_{\mathrm{m}}, \tag{14.9}$$

where $\omega_{\mathrm{gc}}$ is the gain crossover frequency. Let $n_{\mathrm{gc}}$ be the slope of the gain curve of the loop transfer function $L(s) = P(s)C(s)$ at the crossover frequency. Since




(a) Gain curve (b) Phase curve

**Figure 14.5:** Illustration of the gain crossover frequency inequality. (a) Gain curve of the transfer function, with the slope of the curve at the gain crossover frequency $n_{\rm gc}$ marked. (b) Phase of the transfer function (solid) and its minimum phase component (dashed). The phase margin $\varphi_{\rm m}$, the phase lags $\varphi_{\rm mp}$ and $\varphi_{\rm ap}$ of the minimum phase component, and the all-pass component are shown in the figure.

$|P_{\rm ap}(i\omega)| = 1$ it follows that

$$n_{\rm gc} = \frac{d\log|L(i\omega)|}{d\log\omega}\bigg|_{\omega=\omega_{\rm gc}} = \frac{d\log|P_{\rm mp}(i\omega)C(i\omega)|}{d\log\omega}\bigg|_{\omega=\omega_{\rm gc}}.$$

Assuming that the controller $C(s)$ has neither poles nor zeros in the right half-plane, it then follows from Bode's relations (equation (10.9) on page 10-19) that

$$\arg P_{\rm mp}(i\omega) + \arg C(i\omega) \approx n_{\rm gc}\frac{\pi}{2}.$$

Combining this with equation (14.9) gives the following inequality for the allowable phase lag of the all-pass part at the gain crossover frequency, which we state as a theorem.

**Theorem 14.4** (Gain crossover frequency inequality)**.** *Let $P(s) = P_{mp}(s)P_{ap}(s)$ where $P_{ap}$ is an all-pass transfer function containing the non-minimum phase portion of $P(s)$. If $C(s)$ is a stabilizing compensator for the closed loop system with no right half-plane poles and zeros and with phase margin $\varphi_m$, gain crossover frequency $\omega_{gc}$, and gain crossover slope $n_{gc}$, then the allowable phase lag for the all-pass transfer function must satisfy the inequality*

$$\varphi_{ap} := -\arg P_{ap}(i\omega_{gc}) \le \pi - \varphi_m + n_{gc}\frac{\pi}{2} := \overline{\varphi}_{ap}. \tag{14.10}$$

The gain crossover frequency inequality is illustrated in Figure 14.5. The condition (14.10) requires that the gain crossover frequency must be chosen so that the phase lag of the all-pass factor is not too large. For systems with high robustness requirements we may choose a phase margin of 60° ($\varphi_{\rm m} = \pi/3$). To have a reasonable flexibility in choosing the gain crossover frequency we choose $n_{\rm gc} = -1$, which gives an admissible phase lag $\overline{\varphi}_{\rm ap} = \pi/6 = 0.52$ rad (30°) for the all-pass component. For systems where we can accept a lower robustness we might choose a phase margin of 45° ($\varphi_{\rm m} = \pi/4$) and the slope $n_{\rm gc} = -1/2$, which gives an admissible phase lag $\overline{\varphi}_{\rm ap} = \pi/2 = 1.57$ rad (90°).

The gain crossover frequency inequality (14.10) shows that non-minimum phase

components impose severe restrictions on possible crossover frequencies and that there are systems that cannot be controlled with sufficient stability margins. We illustrate the limits in a number of commonly encountered situations.

**Example 14.3 Crossover frequency limits for a process with a zero in the right half-plane**
The non-minimum phase part of the process transfer function for a system with a right half-plane zero is

$$P_{\text{ap}}(s) = \frac{z - s}{z + s},$$

where $z > 0$. Notice that we have $z - s$ in the numerator instead of $s - z$ to satisfy the condition that $P_{\text{ap}}$ should have negative phase. The phase lag of the all-pass factor is

$$\varphi_{\text{ap}} = -\arg P_{\text{ap}}(i\omega) = 2 \arctan \frac{\omega}{z}.$$

Let the admissible phase lag of the all-pass factor be $\overline{\varphi}_{\text{ap}}$. The inequality (14.10) then gives the following bound on the crossover frequency:

$$\omega_{\text{gc}} \leq z \, \tan(\overline{\varphi}_{\text{ap}}/2). \tag{14.11}$$

With $\overline{\varphi}_{\text{ap}} = \pi/3$ we get $\omega_{\text{gc}} < 0.6\, z$. We can thus conclude that a right half-plane zero limits the achievable gain crossover frequency $\omega_{\text{gc}}$, and slow right half-plane zeros ($z$ small) give lower crossover frequency than fast right half-plane zeros. $\nabla$

Processes with zeros in the right half-plane are not uncommon, and are they are inherent consequences of the physics, as in Exercise 14.6, which models hydroelectric power generation. Another example is the shrink and swell phenomenon in drum level control discussed in Example 3.14. In that example the zero in the right half-plane is associated with the inverse response characteristic, where the step response initially moves in the wrong direction. The effect also appears in product development projects where the cost initially increases during the development phase and then hopefully decreases to give profit when the product appears on the market.

We next consider the case of right half-plane poles.

**Example 14.4 Crossover frequency limits for a process with a pole in the right half-plane**
The non-minimum phase part of the transfer function for a system with a pole in the right half-plane is

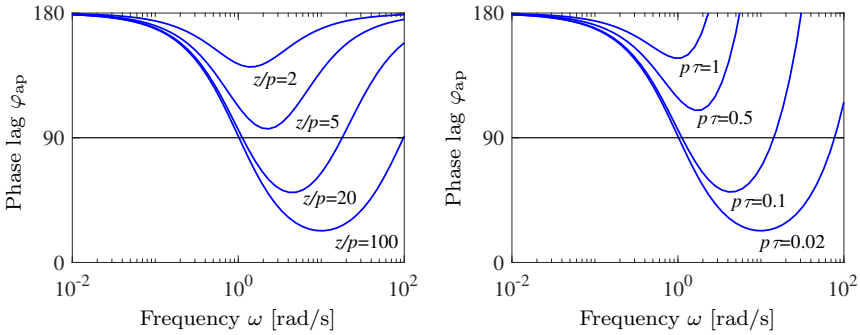
$$P_{\text{ap}}(s) = \frac{s + p}{s - p},$$

where $p > 0$. The sign of $P_{\text{ap}}$ is dictated by the condition that it should have negative phase. The phase lag of the non-minimum phase part is

$$\varphi_{\text{ap}} = -\arg P_{\text{ap}}(i\omega) = 2 \arctan \frac{p}{\omega},$$

and the inequality (14.10) gives the following bound on the crossover frequency:

$$\omega_{\text{gc}} \geq \frac{p}{\tan(\overline{\varphi}_{\text{ap}}/2)}, \tag{14.12}$$




(a) Right half-plane pole/zero pair (b) Right half-plane pole and time delay

**Figure 14.6:** Illustration of the gain crossover frequency inequality for systems with a zero and a pole in the right half-plane (a) and systems with a time delay and a right half-plane pole (b). The figures show the phase lag $\varphi_{ap}$ of the all-pass factor $P_{ap}$ as a function of frequency for the systems using equations (14.14) and (14.15). All systems have a right half-plane pole at $p = 1$. The systems in (a) have zeros at $z = 2$, 5, 20, and 100, and the systems in (b) have time delays $\tau = 0.02$, 0.1, 0.5, and 1.

where $\bar{\varphi}_{ap}$ is the maximum admissible phase lag of the all-pass factor $P_{ap}$. Right half-plane poles thus require that the closed loop system has a sufficiently high gain crossover frequency. With $\bar{\varphi}_{ap} = \pi/3$ we get $\omega_{gc} > 1.7p$. Fast right half-plane poles ($p$ large) therefore require a larger gain crossover frequency than slower right half-plane poles. Robust control of unstable systems thus requires that the bandwidths of the process the actuators and the sensors are sufficiently high. ∇

**Example 14.5 Phase lag for processes with a right half-plane pole/zero pair**

Consider a system with a right half-plane zero $z$ and a right half-plane pole $p$. The transfer function of the process and its all-pass factor are given by

$$P(s) = \frac{a - z}{s - p}, \qquad P_{ap}(s) = \frac{(z - s)(s + p)}{(z + s)(s - p)}. \tag{14.13}$$

The all-pass factor has the phase lag

$$\varphi_{ap} = -\arg P_{ap}(i\omega) = 2\arctan(\omega/z) + 2\arctan(p/\omega), \tag{14.14}$$

which is plotted in Figure 14.6a for $z/p = 1, 2, 5, 10, 20, 100$. ∇

We will illustrate with some numerical values. If we require that the phase lag $\varphi_{ap}$ of the non-minimum phase factor be less than 90°, we must require that the ratio $z/p$ be larger than 6 (from Figure 14.6). The pole and the zero must thus be sufficiently separated (Exercise 14.7). The values of the gain crossover frequency $\omega_{gc}$ are also quite restricted.

Notice that we cannot apply Theorem 14.4 if $p > z$ because a stabilizing controller must then have a pole in the right half plane (see Figure 14.1).

Time delays also impose limits similar to those given by zeros in the right half-plane.

For a process with time delay, $P_{\mathrm{ap}}(s) = e^{\tau s}$. Using the gain crossover frequency inequality (14.10) we get $\omega_{\mathrm{gc}}\tau \leq \overline{\varphi}_{\mathrm{ap}}$, where $\tau$ is the time delay. Time delays are thus similar to right half-plane zeros because they require that the bandwidth and the crossover frequencies be sufficiently small.

**Example 14.6 Phase lag for processes with a right half-plane pole and time delay**
Consider a system with all-pass factor and phase lag given by

$$P_{\mathrm{ap}}(s) = \frac{s+p}{s-p}e^{-\tau s}, \qquad \varphi_{\mathrm{ap}} = -\arg P_{p\tau}(i\omega) = \omega\tau + 2\arctan(p/\omega). \tag{14.15}$$

A plot of the phase lag of the all-pass factor is given in Figure 14.6b. The figure shows that the behavior is similar to a system with a right half-plane pole/zero pair. The phase lag $\varphi_{\mathrm{ap}}$ has a minimum $\sqrt{\tau(2-p\tau)} + 2\arctan\sqrt{p\tau/(2-p\tau)}$ for $\omega\tau = \sqrt{p\tau(2-p\tau)}$ (Exercise 14.8). It follows from equation (14.9) that a system with a right half-plane pole $p$ and a time delay $\tau$ cannot be stabilized by a controller with no poles and zeros in the right half-plane if $p\tau \geq 2$. ∇

Systems with a pole/zero pair in the right half-plane are not common. In Example 14.2 we encountered the X-29 aircraft (Exercise 14.11). The next example is another illustration.

**Example 14.7 Balance system**
As an example of a system with both right half-plane poles and zeros, consider the balance system with zero damping introduced in Example 3.2. The transfer functions from force $F$ to output angle $\theta$ and position $q$ were derived in Example 9.11:

$$H_{\theta F}(s) = \frac{ml}{(M_{\mathrm{t}}J_{\mathrm{t}} - m^2l^2)s^2 - mglM_{\mathrm{t}}},$$
$$H_{qF}(s) = \frac{J_{\mathrm{t}}s^2 - mgl}{s^2\big((M_{\mathrm{t}}J_{\mathrm{t}} - m^2l^2)s^2 - mglM_{\mathrm{t}}\big)}.$$

Assume that we want to stabilize the pendulum by using the cart position as the measured signal. The transfer function $H_{qF}$ from the input force $F$ to the cart position $q$ has poles $\{0, 0, \pm\sqrt{mglM_{\mathrm{t}}/(M_{\mathrm{t}}J_{\mathrm{t}} - m^2l^2)}\}$ and zeros $\{\pm\sqrt{mgl/J_{\mathrm{t}}}\}$. Using the parameters in Example 7.7, the right half-plane pole is at $p = 2.68$ and the zero is at $z = 2.09$. With the best choice of the gain crossover frequency, it follows from equation (14.14) that the phase lag of the all-pass component $P_{\mathrm{ap}}$ is 166°, which implies that it impossible to obtain a reasonable phase margin. The pole/zero ratio is 1.28, which is far from the value 6 required to control the system robustly. Using Figure 14.6, we see that the amount of achievable phase margin for the system is very small if we desire a bandwidth in the range of 2–4 rad/s.

The right half-plane zero of the system can be eliminated by changing the output of the system. For example, if we choose the output to correspond to a position at a distance $r$ along the pendulum, we have $y = q - r\sin\theta$ and the transfer function for the linearized output becomes

$$H_{yF}(s) = H_{qF}(s) - rH_{\theta F}(s) = \frac{(J_{\mathrm{t}} - mlr)s^2 - mgl^2}{s^2\big((M_{\mathrm{t}}J_{\mathrm{t}} - m^2l^2)s^2 - mglM_{\mathrm{t}}\big)}.$$

If we choose $r$ sufficiently large, then $mlr - J_\mathrm{t} > 0$ and we eliminate the right half-plane zero, obtaining instead two pure imaginary zeros. The gain crossover frequency is determined by the right half-plane pole $p = \sqrt{mglM_\mathrm{t}/(M_\mathrm{t}J_\mathrm{t} - m^2l^2)}$ (Example 14.4). If our admissible phase lag for the non-minimum phase part is $\varphi_l = 45^\circ$, then our gain crossover must satisfy

$$\omega_\mathrm{gc} \geq \frac{p}{\tan(\varphi_l/2)} = 6.48 \text{ rad/s}.$$

If the actuators have sufficiently high bandwidth, e.g., a factor of 10 above $\omega_\mathrm{gc}$ or roughly 10 Hz, then we can provide robust tracking up to this frequency. $\nabla$

## 14.4 THE MAXIMUM MODULUS PRINCIPLE

Significant insight into the fundamental limits imposed by poles and zeros in the right half-plane and time delays can be obtained with simple calculations by using the *maximum modulus principle*.

**Theorem 14.5** (Maximum modulus principle)**.** *Let $\Omega \subset \mathbb{C}$ be a nonempty, bounded, open and connected set in the complex plane and let $G : \overline{\Omega} \to \mathbb{C}$ be continuous on the closure of $\Omega$ and analytic on $\Omega$. Then*

$$\sup_{s\in\overline{\Omega}} |G(s)| = \sup_{s\in\partial\Omega} |G(s)|$$

This theorem can be used to give bounds on transfer functions, such as the sensitivity functions, by using the Nyquist contour as the boundary of the open right half-plane. We state this result as a corollary.

**Corollary.** *Let $G(s)$ be an analytic transfer function on the closed, right half-plane. Then $|G(s)|$ assumes its largest value on the imaginary axis:*

$$\max_{\omega\in\mathbb{R}} |G(i\omega)| \geq \max_{\mathrm{Re}\, s\geq 0} |G(s)|.$$

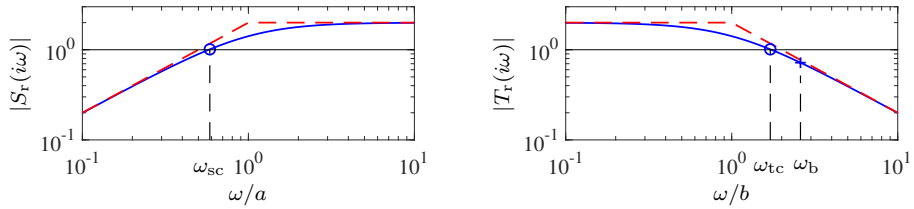
To see how this result can be applied, consider the transfer functions

$$S(s) = \frac{1}{1 + P(s)C(s)}, \qquad T(s) = \frac{P(s)C(s)}{1 + P(s)C(s)},$$

and note that $S(s) + T(s) = 1$. The zeros of the sensitivity function $S(s)$ are the poles of the process and the controller, and the zeros of the complementary sensitivity function are the zeros of the process and the controller. We find from the above equation that $S(z) = 1$ for zeros $z$ of the process or the controller. Similarly we have $T(p) = 1$ for poles $p$ of the poles of the process or the controller.

We can use the maximum modulus principle to obtain requirements on disturbance attenuation and robustness, formulated as conditions on the sensitivity functions. We will use the following nominal transfer functions to capture our desired sensitivity requirements:

$$S_\mathrm{r}(s) = \frac{M_\mathrm{s}\, s}{s + a}, \qquad T_\mathrm{r}(s) = \frac{M_\mathrm{t}\, b}{s + b}. \tag{14.16}$$




(a) Requirements for sensitivity (b) Requirements for complementary sensitivity

**Figure 14.7:** Gain curves for the transfer functions (a) $S_r(s) = M_s\ s/(s+a)$ and (b) $T_r(s) = M_t\ b/(s+b)$, which give requirements for sensitivity and complementary sensitivity. The plots are drawn for $M_s = M_t = 2$, the gain crossover frequencies are denoted by $\circ$, and the bandwidth defined by $T(\omega_b) = 1/\sqrt{2}$ by $+$.

Bode plots of the gain curves of the transfer functions $S_r(s)$ and $T_r(s)$ are shown in Figure 14.7a. We will consider requirements defined by

$$|S(i\omega)| \leq |S_r(i\omega)|, \qquad |T(i\omega)| \leq |T_r(i\omega)|, \tag{14.17}$$

which guarantee that the maximum sensitivities are less than $M_s$ or $M_t$. The sensitivity crossover frequencies of the transfer functions (14.16) and the bandwidth are given by

$$\omega_{sc} = \frac{a}{\sqrt{M_s^2 - 1}}, \qquad \omega_{tc} = b\sqrt{M_t^2 - 1}, \qquad \omega_b = b\sqrt{2M_t^2 - 1}. \tag{14.18}$$

We will now use the maximum modulus principle to investigate the effects of poles and zeros in the right half-plane, and to establish limits on achievable performance.

**Example 14.8 Sensitivity limits for a system with a zero in the right half-plane**

Assume that the process $P(s)$ has a zero $s = z$ in the right half-plane and no other poles and zeros in the right half-plane. The sensitivity function is analytic in the right half-plane for all controllers that stabilize the system, and equation (14.17) implies that

$$\max_\omega \left| \frac{S(i\omega)}{S_r(i\omega)} \right| \leq 1. \tag{14.19}$$

The function $S(s)/S_r(s)$ is analytic in the right half-plane and on the imaginary axis. If the process has a zero $s = z$ in the right half-plane the sensitivity function has the property that $S(z) = 1$. Applying the maximum modulus principle to the function $S(s)/S_r(s)$ then gives

$$\max_\omega \left| \frac{S(i\omega)}{S_r(i\omega)} \right| \geq \left| \frac{S(z)}{S_r(z)} \right| = S(z)\,\frac{z+a}{M_s z} = \frac{z+a}{M_s z}.$$

This inequality is compatible with equation (14.19) only if $z + a \leq M_s z$, hence

$$a \leq z\,(M_s - 1), \qquad \omega_{sc} \leq z\sqrt{\frac{M_s - 1}{M_s + 1}}, \tag{14.20}$$

where the bound on $\omega_{sc}$ follows after some algebra. We see that a right half-plane zero $z$ limits the sensitivity crossover frequency $\omega_{sc}$ of the closed loop system and thus also the range of frequencies over which the sensitivity can be kept small (compare with Example 14.3). $\nabla$

If we make the calculations for a system with complex zeros $s = z_{re} \pm i\, z_{im}$, we obtain the following conditions (Exercise 14.12):

$$\begin{aligned} a &\leq \sqrt{M_s^2 z_{re}^2 + (M_s^2 - 1) z_{im}^2} - z_{re}, \\ \omega_{sc} &= \frac{a}{\sqrt{M_s^2 - 1}} \leq \frac{\sqrt{M_s^2 z_{re}^2 + (M_s^2 - 1) z_{im}^2} - z_{re}}{\sqrt{M_s^2 - 1}}, \end{aligned} \tag{14.21}$$

which are equal to equation (14.20) for $z_{im} = 0$. Robust control of a process with right half-plane zeros therefore requires that the sensitivity crossover frequency $\omega_{sc}$ is not too high (equations (14.20) and (14.21)). If there are several right half-plane zeros the limit is given by the smallest bound.

A similar analysis based the complementary sensitivity function gives the consequences of right half-plane poles (see Exercise 14.13). We conclude that robust control in the presence of right half-plane poles requires that the complementary sensitivity crossover frequency $\omega_{tc}$ and the bandwidth $\omega_b$ are sufficiently large.

Next we will consider the effect of both poles and zeros in the right half-plane. Since robust control of a process with a right half-plane zero $z$ requires that the sensitivity crossover frequency (or the bandwidth) is sufficiently low and a right half-plane pole requires that the sensitivity crossover frequency is sufficiently high, we may expect that systems with a right half-plane pole/zero pair cannot be controlled robustly if the poles and zeros are close and we may expect that a system cannot be controlled at all if $p > z$. Indeed, it can be shown (Exercise 12.15) that a process cannot be stabilized by a stable controller if $p > z$. We will analyze the situation in the next example.

**Example 14.9 Sensitivity limits for processes with poles and zeros in the right half-plane**
Consider a process $P(s)$ with the right half-plane zeros $z_k$ and the right half-plane poles be $p_k$. Introduce the polynomial $n(s)$ with zeros $s = z_k$ and the polynomial $d(s)$ with zeros $s = p_k$. The process transfer function can then be written as

$$P(s) = \frac{n(s)}{d(s)} \tilde{P}(s), \tag{14.22}$$

where $\tilde{P}(s)$ has no poles or zeros in the right half plane. Furthermore we consider controllers that stabilize the process. The sensitivity function

$$S(s) = \frac{1}{1 + P(s)C(s)} = \frac{d(s)}{d(s) + n(s)\tilde{P}(s)C(s)},$$

has the zeros $s = p_k$ in the right half-plane, and we have $S(z_k) = 1$ for all zeros $z_k$ of the polynomial $n(s)$. Introduce the weighting function

$$W_p(s) = \frac{d(-s)}{d(s)}.$$

The poles and zeros of this function are symmetric with respect to the imaginary axis, which implies that $|W_{\mathrm{p}}(i\omega)| = 1$. The function $W_{\mathrm{p}}(s)S(s)$ is analytic in the right half-plane, since the polynomial $d(s)$ is canceled and $d(-s)$ has all its roots in the left half-plane. Since $S(z_k) = 1$, it follows from the maximum modulus principle that

$$M_{\mathrm{s}} = \max_{\omega} |S(i\omega)| = \max_{\omega} |W_{\mathrm{p}}(i\omega)S(i\omega)| \geq |W_{\mathrm{p}}(z_k)S(z_k)| = \Big|\frac{d(-z_k)}{d(z_k)}\Big|, \quad (14.23)$$

which implies

$$M_{\mathrm{s}} \geq \max_{k} \Big|\frac{d(-z_k)}{d(z_k)}\Big|. \quad (14.24)$$

For a system with a pole/zero pair in the right half-plane we have $n(s) = s - z$ and $d(s) = s - p$. Since there is only one zero the above equation becomes

$$M_{\mathrm{s}} \geq \Big|\frac{z+p}{z-p}\Big|, \quad (14.25)$$

which implies that

$$\frac{z}{p} \geq \frac{M_{\mathrm{s}}+1}{M_{\mathrm{s}}-1} \quad \text{if } z > p \qquad \text{or} \qquad \frac{z}{p} \leq \frac{M_{\mathrm{s}}-1}{M_{\mathrm{s}}+1} \quad \text{if } z > p. \quad (14.26)$$

$\nabla$

To find controllers with a maximum sensitivity less than $M_{\mathrm{s}}$ for a process with a right half-plane pole/zero pair, it follows from equation (14.26) that the pole and zero must be sufficiently separated. The zero/pole ratio must either be smaller than $(M_{\mathrm{s}}-1)/(M_{\mathrm{s}}+1)$ or larger than $(M_{\mathrm{s}}+1)/(M_{\mathrm{s}}-1)$. For $M_{\mathrm{s}} = 2$ the critical ratios are 0.5 and 2 and for $M_{\mathrm{s}} = 1.4$ they are 1/6 and 6.

A calculation similar to the one in Example 14.9 for the complementary sensitivity gives (Exercise 14.14)

$$M_{\mathrm{t}} \geq \max_{k} \Big|\frac{n(-p_k)}{n(p_k)}\Big|. \quad (14.27)$$

In the special case of a single pole/zero pair the condition becomes

$$M_{\mathrm{t}} \geq \Big|\frac{z+p}{z-p}\Big| \qquad \Longrightarrow \qquad \frac{z}{p} \geq \frac{M_{\mathrm{t}}+1}{M_{\mathrm{t}}-1} \quad \text{or} \quad \frac{z}{p} \leq \frac{M_{\mathrm{t}}-1}{M_{\mathrm{t}}+1}. \quad (14.28)$$

We illustrate the results with an example.

**Example 14.10 Bicycle with rear-wheel steering**
Figure 14.8 shows two bicycles with rear wheel steering. Bicycle dynamics were discussed in Section 4.2, where the following model was obtained:

$$J\frac{d^2\varphi}{dt^2} - \frac{Dv_0}{b}\frac{d\delta}{dt} = mgh\sin\varphi + \frac{mv_0^2 h}{b}\delta.$$

The wheelbase is $b$, the mass of the bicycle and the driver is $m$, and the distance from the center of mass to ground is $h$. Furthermore, $J$ is the moment of inertia with respect to the line through the contact points of the wheels with the ground

(a) Unrideable bicycle (b) Rideable bicycle

**Figure 14.8:** Two bicycles with rear wheel steering (a) is unrideable and (b) is rideable. Figures courtesy of Richard Klein [Kle89].

and $D$ is the inertia product. We have $J \approx mh^2$ and $D \approx mah$, where $a$ the distance between the projection of the center of mass on the ground and the contact point of the driving wheel. The model for a bicycle with rear wheel steering is obtained simply by reversing the sign of the velocity and we get

$$mh^2 \frac{d^2\varphi}{dt^2} + \frac{mhav_0}{b}\frac{d\delta}{dt} = mgh\sin\varphi + \frac{mv_0^2 h}{b}\delta.$$

The transfer function from steering angle $\delta$ to tilt angle $\varphi$ is

$$P_{\varphi\delta} = \frac{-av_0 s + v_0^2}{b(hs^2 - g)} = \frac{av_0}{bh}\,\frac{-s + v_0/a}{s^2 - g/h}.$$

The transfer function has a right half-plane pole $p = \sqrt{g/h}$ and a right half-plane zero at $z = v_0/a$. The condition (14.26) then gives

$$\frac{z}{p} = \frac{v_0}{a}\sqrt{\frac{h}{g}} \geq \frac{M_\mathrm{s}+1}{M_\mathrm{s}-1} \qquad \Longrightarrow \qquad v_0 \geq a\sqrt{\frac{g}{h}}\,\frac{M_\mathrm{s}+1}{M_\mathrm{s}-1}.$$

The unstable pole $p = \sqrt{g/h}$ does not depend on the velocity but the right half-plane zero $z = v_0/a$ is proportional to the velocity. To ride the bicycle comfortably the velocity must therefore be sufficiently large. Evaluating the parameters for the bicycles in Figure 14.8 with $M_\mathrm{s} = 2$ we find $v_0 \geq 9.4$ m/s (34 km/h) for the bicycle in Figure 14.8a and $v_0 \geq 1.2$ m/s (3.8 km/h) for the bicycle in Figure 14.8b. The bicycle in Figure 14.8a has indeed proven to be unrideable, while the bicycle in Figure 14.8b is rideable [Kle89]. ∇

In view of the robustness results for systems with a single right half-plane pole or single right half-plane zero, it is perhaps surprising that processes with $p > z$ can actually be controlled robustly. This is in fact possible, though it requires more clever design techniques. A detailed discussion of stabilizability is given by Youla [YJJBL74], where it is proven that a system with right half-plane poles and zeros can be stabilized with a stable controller if and only if the number of poles

**Table 14.1:** Summary of limits by time delays and right half-plane (RHP) poles and zeros, $\omega_{sc}$ and $\omega_{tc}$ are the crossover frequencies for the sensitivity function and the complementary sensitivity function.

| Process feature | Limits |
|---|---|
| Real RHP zero $z$ | $\omega_{sc} \leq z\sqrt{\frac{M_s - 1}{M_s + 1}}$ |
| Complex RHP zeros $z = z_{re} \pm iz_{im}$ | $\omega_{sc} \leq \frac{\sqrt{M_s^2 z_{re}^2 + (M_s^2 - 1)z_{im}^2} - z_{re}}{\sqrt{M_s^2 - 1}}$ |
| Real RHP pole $p$ | $\omega_{tc} \geq p\sqrt{\frac{M_t + 1}{M_t - 1}}$ |
| Complex RHP poles $p = p_{re} \pm ip_{im}$ | $\omega_{tc} \geq \frac{\sqrt{M_t^2 p_{re}^2 + (M_t^2 - 1)p_{im}^2} + p_{re}}{\sqrt{M_t^2 - 1}}$ |
| RHP pole/zero pair $p, z$ | $M_s \geq \left\lvert\frac{p+z}{p-z}\right\rvert,\ M_t \geq \left\lvert\frac{p+z}{p-z}\right\rvert$ |
| RHP poles and zeros $d(s), n(s)$ | $M_s \geq \max_k \left\lvert\frac{d(-z_k)}{d(z_k)}\right\rvert,\ M_t \geq \max_k \left\lvert\frac{n(-p_k)}{n(p_k)}\right\rvert$ |
| RHP pole $p$ and time delay $\tau$ | $M_t \geq e^{p\tau},\ M_s \geq e^{p\tau} - 1$ |

between every pair of right half-plane zeros is even (see also Vidyasagar [Vid85, Theorem 3.1]). For a system with a single pole/zero pair, the result implies that a process with $p > z$ requires a controller with a pole in the right half-plane, hence an *unstable* controller. An example is given in Exercise 14.21. The root locus method gives significant insight into these cases (Exercise 14.15).

We have focused here on the effects of right half-plane poles and zeros. Another common source of limits is the existence of time delays. The limits imposed by a time delay and a right half-plane pole are similar to the limits by a right half-plane pole/zero pair. A list of various limits are summarized in Table 14.1.

## 14.5 ROBUST POLE PLACEMENT

When using any design method that does not include requirements on robustness it is necessary to check the robustness of the design. In Section 7.2 we used state feedback to assign the eigenvalues of the closed loop system and showed that if a system is reachable then the eigenvalues of the closed loop system can be set to arbitrary values. This design technique is also called "pole placement" and in this section we will show that the insights into the roles of poles and zeros can give us a deeper understanding of how to design such controllers. In particular we will show that it is necessary to take the process zeros into account when choosing the desired closed loop poles. We will first analyze examples where seemingly reasonable designs lead to closed loop systems that are not robust. We will then we then present design rules for pole (eigenvalue) placement that guarantee that the closed loop systems is robust.

### Fast Stable Process Poles

A pole stable if it is in the left half-plane and unstable if it is in the right half-plane. We call it "fast" if its magnitude is larger than the intended closed loop bandwidth. We will explore the effects of fast stable process poles on pole placement design through a simple example that illustrates the basic design rule.

**Example 14.11 Robust pole placement for fast process poles**
Consider a PI controller for a first-order system, where the process and the controller have the transfer functions $P(s) = b/(s+a)$, with $a > 0$, and $C(s) = k_{\mathrm{p}} + k_{\mathrm{i}}/s$. The loop transfer function is

$$L(s) = \frac{b(k_{\mathrm{p}}s + k_{\mathrm{i}})}{s(s+a)},$$

and the closed loop characteristic polynomial is

$$s(s+a) + b(k_{\mathrm{p}}s + k_{\mathrm{i}}) = s^2 + (a + bk_{\mathrm{p}})s + k_{\mathrm{i}}b.$$

If we specify that the desired closed loop poles should be $-p_1$ and $-p_2$, we find that the controller parameters are given by

$$k_{\mathrm{p}} = \frac{p_1 + p_2 - a}{b}, \qquad k_{\mathrm{i}} = \frac{p_1 p_2}{b}.$$

The sensitivity functions are then

$$S(s) = \frac{s(s+a)}{(s+p_1)(s+p_2)}, \qquad T(s) = \frac{(p_1 + p_2 - a)s + p_1 p_2}{(s+p_1)(s+p_2)}.$$

Assume that process pole $a$ is faster than the closed loop poles $p_1 < p_2 < a$. The proportional gain $k_{\mathrm{p}}$ is then negative and the controller has a zero in the right half-plane, an indication that the system may have bad properties. Consider the gain $|S(i\omega)|$ of the sensitivity function plotted in Figure 14.9a for $a = b = 1$, $p_1 = 0.05$, and $p_2 = 0.2$. We have $S(i\omega) \approx 1$ for high frequencies. Moving backwards in frequency we find that the sensitivity increases around $\omega = a$ corresponding to the fast process pole. The sensitivity continues to increase with decreasing frequency and it does not decrease until the frequency is below the closed loop pole $p_2$. The net effect is a large sensitivity peak, approximately $\omega = a/\sqrt{p_1 p_2} \approx 10$.

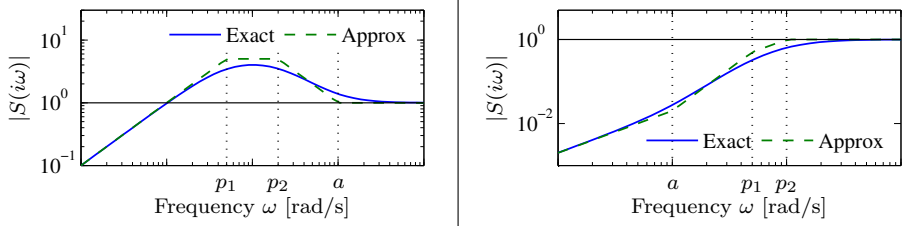
The problem with poor robustness can be avoided by choosing one closed loop pole equal to the process pole, i.e., $p_2 = a$. The controller gains then become

$$k_{\mathrm{p}} = \frac{p_1}{b}, \qquad k_{\mathrm{i}} = \frac{ap_1}{b},$$

which means that the fast process pole is canceled by a controller zero at $s = -a$. The loop transfer function and the sensitivity functions are

$$L(s) = \frac{bk_{\mathrm{p}}}{s}, \qquad S(s) = \frac{s}{s + bk_{\mathrm{p}}}, \qquad T(s) = \frac{bk_{\mathrm{p}}}{s + bk_{\mathrm{p}}}.$$

Figure 14.9b shows the gain curve of the sensitivity function for the case when the closed loop poles ($p_1 = 5$, $p_2 = 20$) are faster than the process pole ($a = 1$). There is no peak of the sensitivity function in this case. ∇




(a) $p_1 < p_2 < a$ (b) $a < p_1 < p_2$

**Figure 14.9:** Gain curves of the sensitivity function $S$ for designs in Example 14.11. The solid lines are the true sensitivities, and the dashed lines are the asymptotes. Notice the high peak of the sensitivity function in (a) and that there is no peak in (b).

### Slow Stable Process Zeros

We call a zero "stable" if it is in the left half-plane and "unstable" if it is in the right half-plane. Furthermore a zero is said to be "slow" if its magnitude is smaller than the intended closed loop bandwidth. We will explore the effects of slow stable process zeros in pole placement design, and we begin with a simple example.

**Example 14.12 Vehicle steering**
Consider the model for vehicle steering in Example 9.10, where the transfer function from steering angle to lateral position is

$$P(s) = \frac{\gamma s + 1}{s^2} = \gamma \, \frac{s + 1/\gamma}{s^2}.$$

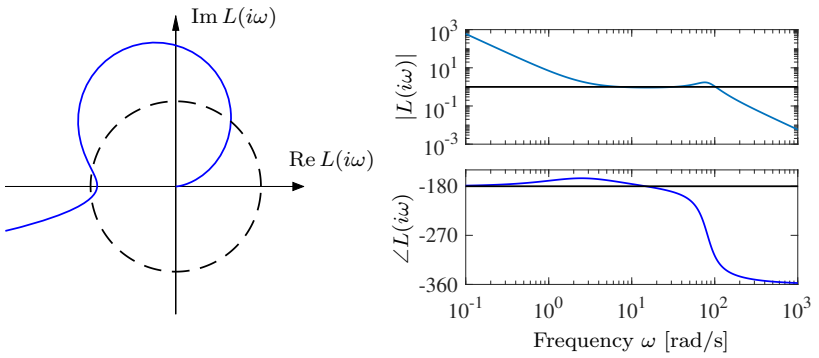
A controller based on state feedback was designed in Example 7.4, and state feedback was combined with an observer in Example 8.4. The system simulated in Figure 8.8 has closed loop poles specified by $\omega_c = 0.7$, $\zeta_c = 0.707$, $\omega_o = 1$, and $\zeta_o = 0.707$. Assume that we want a faster closed loop system and choose $\omega_c = 10$, $\zeta_c = 0.707$, $\omega_o = 20$, and $\zeta_o = 0.707$. Using the state representation in Example 8.3, a pole placement design gives state feedback gains $k_1 = 100$ and $k_2 = -35.86$ and observer gains $l_1 = 28.28$ and $l_2 = 400$. The controller transfer function is

$$C(s) = \frac{-11516s + 40000}{s^2 + 42.4s + 6657.9}. \tag{14.29}$$

Figure 14.10 shows Nyquist and Bode plots of the loop transfer function.

The Nyquist plot indicates that the robustness is poor since the loop transfer function is very close to the critical point $-1$. The phase margin is 7° and the gain margin is $g_m = 1.08$, which means that the system becomes unstable if the gain is increased by 8%. The poor robustness also shows up in the Bode plot, where the gain curve hovers around the value 1 while the phase curve is close to $-180°$ for a wide frequency range (3-40 rad/s). Additional insight is obtained by analyzing the sensitivity functions, shown as dashed lines in Figure 14.11. The maximum sensitivities are $M_s = 13$ and $M_t = 12$.

It is surprising that the closed loop is so sensitive to process variations when we have designed a controller so that the closed loop system has well-damped closed




(a) Nyquist plot of $L(s) = P(s)C(s)$ (b) Bode plot of $L(s) = P(s)C(s)$

**Figure 14.10:** Observer-based control of vehicle steering. Nyquist and Bode plots of the loop transfer function for vehicle steering with a controller based on state feedback and an observer. The controller provides stable operation, but with very poor robustness.

loop poles. We have an indication that something is unusual because the design gives a controller that has a zero in the right half-plane at $s = 3.5$, while the observer and controller have complex poles with $\omega_c = 10$ and $\omega_o = 20$. Recall the results from Example 14.3, which indicate that robust control of a process with a zero at $s = 3.5$ cannot have a gain crossover frequency larger than $\omega_{gc} = 2$.

To understand what happens, we will investigate the reason for the peaks of the sensitivity functions. Let the transfer functions of the process and the controller be

$$P(s) = \frac{n_p(s)}{d_p(s)}, \qquad C(s) = \frac{n_c(s)}{d_c(s)}, \tag{14.30}$$

where $n_p(s)$, $n_c(s)$, $d_p(s)$ and $d_c(s)$ are the numerator and denominator polynomi-

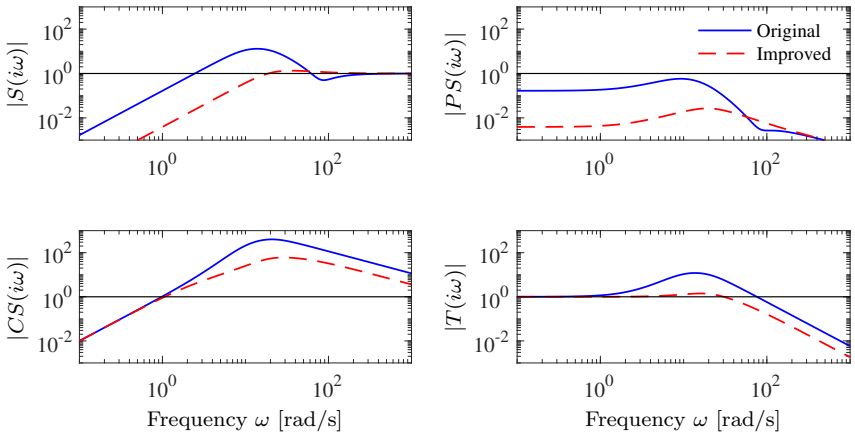



**Figure 14.11:** Gain curves of the sensitivity functions for systems with observer-based control of vehicle steering. The plots for the original controller with $\omega_c = 10$, $\zeta_c = 0.707$, $\omega_o = 20$, $\zeta_o = 0.707$ is shown in solid lines and the improved controller with $\omega_c = 10$, $\zeta_c = 2.6$ is shown in dashed lines.

als. The complementary sensitivity function is

$$T(s) = \frac{P(s)C(s)}{1 + P(s)C(s)} = \frac{n_{\rm p}(s)n_{\rm c}(s)}{d_{\rm p}(s)d_{\rm c}(s) + n_{\rm p}(s)n_{\rm c}(s)}.$$

The poles of $T(s)$ are the poles of the closed loop system and the zeros of $T(s)$ are the zeros of the process and the controller transfer functions. A plot of the gain curve of $T(s)$ for the original controller is shown as the solid line in the lower right plot in Figure 14.11. We have $T(0) = 1$, because $L(0) = P(0)C(0) = \infty$ due to the double integrator of $P$. The gain $|T(i\omega)|$ increases for increasing $\omega$ due to the process zero at $\omega = 2$. It increases further at the controller zero at $\omega = 3.5$, and it does not start to decrease until the closed loop poles appear at $\omega = 10$ and $\omega = 20$. The net result is a high peak of the gain of the complementary sensitivity function.

The peak in the complementary sensitivity function can be avoided by assigning a closed loop pole at the slow process zero or close to it. We can achieve this by choosing $\omega_{\rm c} = 10$ and $\zeta_{\rm c} = 2.6$, which gives closed loop poles at $s = -2$ and $s = -50$. The controller transfer function then becomes

$$C(s) = \frac{3628s + 40000}{s^2 + 80.28s + 156.56} = 3628\,\frac{s + 11.02}{(s + 2)(s + 78.28)}. \tag{14.31}$$

Notice that the new controller has a pole at $s = -2$ that cancels the process zero. Also notice the large differences in the zero frequency gains of the controllers $C(0) = 6.0$ for the controller (14.29) and $C(0) = 255$ for the controller (14.31). Cancellation of the slow zero gives a dramatic increase of the low-frequency gain of the controller. The gain curves for the sensitivity function of the improved controller are shown with dashed lines in Figure 14.11. The closed loop system has the maximum sensitivities $M_{\rm s} = 1.34$ and $M_{\rm t} = 1.41$, which indicate good robustness.

This example shows that a robust design can be obtained by first canceling the slow stable process zero, designing the controller for the system without the zero, and then adding the pole to the controller. Notice that the plot of $|PS(i\omega)|$ shows that the improved system, represented by the dashed lines, has much better disturbance attenuation and the plot of $|CS(i\omega)|$ shows that is it not as sensitive to measurement noise. The large differences in low-frequency gains of the controllers are clearly visible in the gain curves for $S$ and $PS$. ∇

We can learn several things from this example. First, that it is essential to evaluate the closed loop system carefully, for example by plotting the gain curves of the Gang of Four. We have also seen that seemingly reasonable design methods do not necessarily give robust closed loop systems. For designs based on pole placement it is necessary to consider the open loop poles and zeros when specifying the desired closed loop dynamics, and in particular robustness requires that there must be closed loop poles that are equal to or close to slow stable process zeros. Another lesson is that slow unstable process zeros impose limits on the achievable bandwidth, as already noted in Section 14.4.

One potential issue with the choice of controller poles and zeros that exactly cancel the open loop poles and zeros is that they may lead to undesirable dynamics or lack of robustness (if there are model uncertainties). We address this important issue in more detail below.

### Design Rules for Robust Pole Placement

Based on the insight gained from the previous examples, we can now formulate design rules that give controllers with good robustness for pole placement design. Consider the expression (13.12) for maximum complementary sensitivity, repeated here:

$$M_{\mathrm{t}} = \sup_{\omega} |T(i\omega)| = \left\| \frac{PC}{1+PC} \right\|_{\infty}.$$

Let $\omega_{\mathrm{gc}}$ be the desired gain crossover frequency. Assume that the process has zeros that are slower than $\omega_{\mathrm{gc}}$. The complementary sensitivity function is 1 for low frequencies, and it increases for frequencies close to the process zeros unless there is a closed loop pole in the neighborhood (as seen, for instance, in Figure 14.11 of the previous example). To avoid large values of the complementary sensitivity function we find that the closed loop system should therefore have poles close to or equal to the slow stable zeros. This means that slow stable zeros should be canceled by controller poles. Since unstable zeros cannot be canceled, the presence of slow *unstable* zeros means that achievable gain crossover frequency must be smaller than the slowest unstable process zero.

Now consider process poles that are faster than the desired gain crossover frequency. Consider the expression for the maximum of the sensitivity function:

$$M_{\mathrm{s}} = \sup_{\omega} |S(i\omega)| = \left\| \frac{1}{1+PC} \right\|_{\infty}.$$

The sensitivity function is 1 for high frequencies. Moving from high to low frequencies, the sensitivity function increases at the fast process poles. The sensitivity function will have large peaks unless there are closed loop poles that are close to the fast process poles. To avoid large peaks in the sensitivity the closed loop system should therefore have poles close the fast process poles. One way to achieve this is to have controller zeros close to the fast process pole. Since unstable modes cannot be canceled, the presence of a fast unstable pole implies that the gain crossover frequency must be sufficiently large, as was discussed in Section 14.3 (Example 14.4).

To summarize, we obtain the following simple rule for choosing closed loop poles: slow stable process zeros should be matched by slow closed loop poles, and fast stable process poles should be matched by fast closed loop poles. Slow unstable process zeros and fast unstable process poles impose severe limits.

## 14.6 NONLINEAR EFFECTS

Although we focus primarily on linear systems in this chapter, there are some nonlinearities that must be considered when designing a control system. Limits on actuation power set bounds on response speed. Nonlinearities due to friction, round-off error in A/D and D/A converters, and numerical representations in computation bound the precision that can be obtained in regulation and tracking. We briefly describe some of the effects of these limits here, illustrated primarily through examples.

### Actuation Limits

Many limits are associated with constraints on how large signals and variables can be. Motors have limited torque, amplifiers have limits on currents, and pumps have limited flow. There are also limits due to equipment protection: the temperature of a component must not be too high and compressor stall must be avoided, for example. Limits may appear as restrictions on the amplitude and the rate of change of the control signal. There may also be restrictions on internal process variables and their rates.

A real-world example of the consequences of actuator limits is the grounding of a passenger ferry in 2004. The ferry was grounded while entering the port of Umeå due to high winds (20 m/s). The incident analysis revealed that the wind forces of 600 kN and higher were much larger than the forces generated by the ship's propellers and rudder, and even assistance from a tugboat capable of applying 260 kN of thrust could not have helped. In the setting of control systems, this example illustrates a situation where actuators do not have the sufficient power to counteract the load disturbances.

The following simple analytical example demonstrates how these types of considerations can be taken into account in the design stage of a project.

**Example 14.13 Current limits in servo systems**
Response time is a common requirement for motor drives. The achievable response time depends critically on actuation power and physical limits of the process. To determine the response time we can compute the minimum time to make transitions from one state to the other, subject to the physical constraints on the process and the actuator.

Consider a simple servo system where the actuator is a current-driven voice coil. The system can be modeled by

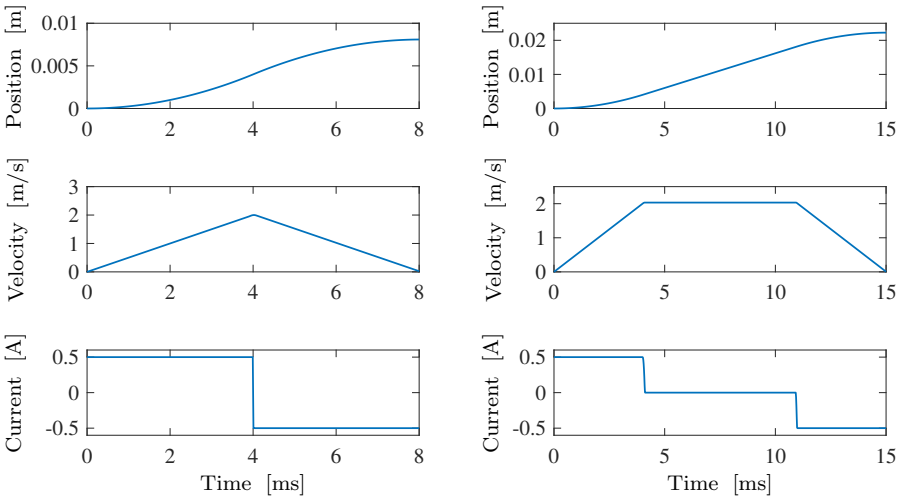
$$m\frac{d^2x}{dt^2} = F = k_I I, \tag{14.32}$$

where $m$ is the mass of the system, $x$ is the position of the mass, $F$ is the force, $I$ is the current through the voice coil, and $k_I$ is the motor constant. The maximum acceleration $a_{\max} = F_{\max}/m = k_I I_{\max}/m$ is given by the maximum current $I_{\max}$. There is also a limit on the maximum velocity: for a voice coil drive the maximum velocity is $v_{\max} = V_{\max}/k_I$, where $V_{\max}$ is the largest supply voltage.

If there is no limit on the velocity, the problem of moving the mass from one position to another in minimum time is simply to apply maximum acceleration until the mid position is reached and then apply maximum deceleration, so-called "bang-bang" control. If there is a velocity limit, the maximum acceleration is only applied until the maximum velocity is reached. The minimum time solution is illustrated in Figure 14.12. When the acceleration $a$ is constant, the velocity increases as $v(t) = at$ and the position is $x(t) = at^2/2 = v^2(t)/(2a)$. A straightforward calculation shows that the minimum time for a transition over a distance $\ell$ with zero velocity at start and end is

$$t = \begin{cases} 2\sqrt{\ell/a_{\max}} & \text{if } \ell \le v_{\max}^2/a_{\max}, \\ \ell/v_{\max} + v_{\max}/a_{\max} & \text{if } \ell > v_{\max}^2/a_{\max}. \end{cases} \tag{14.33}$$

We can derive requirements on the actuator from this equation. ∇




(a) No velocity saturation (b) Velocity saturation

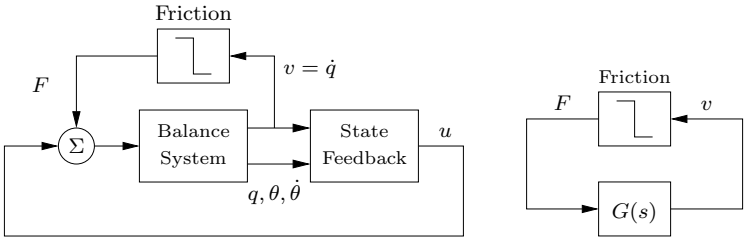
**Figure 14.12:** Minimum time transition for a servo system. (a) The case of short movements when the velocity does not reach the saturation limit. The control is of the "bang-bang" type where maximum current is applied to accelerate or brake. (b) Illustration of what happens for large motions. Full acceleration $a_{\max} = 500$ m/s$^2$ is applied until $t = 5$ ms when maximum velocity $v_{\max} = 2$ m/s is reached and the drive circuit saturates. The current is then zero until time $t = 10$ ms when full braking current is applied. The parameter values are $m = 2.5 \times 10^{-3}$ kg, $k_I = 2.5$ N/A $= 2.5$ Vs/m, $I_{\max} = 0.5$ A, and $V_{\max} = 5$ V.

This simple example can be solved analytically. Software for computing minimum time control is readily available for more complex systems.

Saturation limits can also affect the stability of a feedback system. We saw in Section 10.5 two different methods for reasoning about the effects of (static) nonlinearities in a feedback system: the circle criterion and describing functions. Both of these techniques use the Nyquist plot as a means of analyzing the effects of the nonlinearity on closed loop stability. In the particular case of actuation limits, the circle criterion allows the saturation to be modeled as a sector-bounded nonlinearity with $k_{\text{low}} = 0$ and $k_{\text{high}} = 1$, which implies that the system is stable if the Nyquist curve for the linear dynamics has $\text{Re}\, H(s) > -1$. The describing function method is slightly less constraining, since the image of the describing function for a saturation nonlinearity is given by the negative real axis from $-\infty$ to $-1$, and hence the Nyquist curve for $H(s)$ should not cross the negative real axis at at a gain greater than one. (Note that the describing function method is only an approximation, although it is often a very useful for preliminary design.)

### Measurement Noise and Friction

There are many sources of measurement noise: the physics of the sensor, the electronics, the transmission equipment, and the A/D and D/A converters. The controller in a closed loop system feeds measurement noise into the system, creating fluctuations in all variables. Fluctuations in the output limits regulation and tracking performance. Fluctuations in the control signal causes wear or even saturation




(a) Balance system with state feedback (b) Transformed diagram

**Figure 14.13:** Block diagrams of a balance system with state feedback and friction. (a) Detailed block diagram showing the balance system with inputs $u$ and $F$ and outputs $q$, $\theta$, $v = \dot{q}$, $\dot{\theta}$. (b) Block diagram obtained after transformations. It has two blocks: the nonlinear friction block a linear block with the transfer function $G(s)$ from friction force $F$ to velocity $v$.

of the actuator, and cannot be permitted to be too large. Since measurement noise is typically dominated by high frequencies, it limits the high-frequency gain of the controller, the bandwidth, and thus the response time of the closed loop system.

The effects of measurement noise and quantization can be estimated using linear methods by calculating the transfer function from the noise sources to the control signal and the process variables, and they can be alleviated by filtering and a controller with high-frequency roll-off. Quantization can be approximated as noise with a variance of $\delta^2/12$, where $\delta$ is the quantization level.

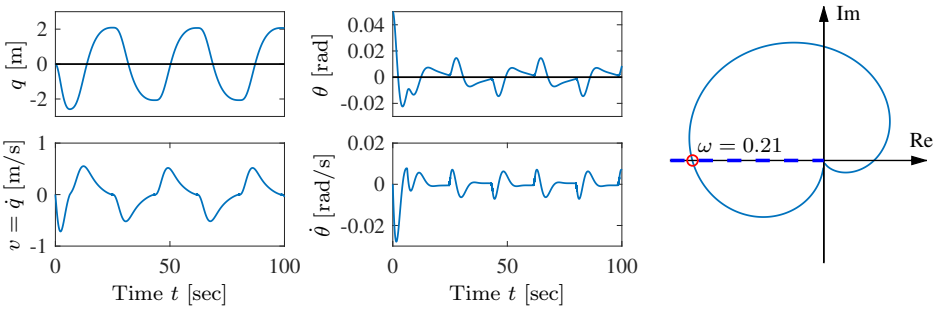
Friction typically generates oscillations that limit regulation and tracking performance. Similar oscillations can be caused by quantization. Oscillations can be reduced by nonlinear friction compensation. Friction is inherently a nonlinear phenomenon, and accurate analysis requires nonlinear methods. Some insight can be obtained using the describing function method discussed in Section 10.5. We illustrate with an example.

**Example 14.14 Effect of friction in a cart–pendulum system**
The cart-pendulum or balance system was introduced in Example 3.2 and we designed a state feedback for it in Example 7.7. Experiments with cart–pendulum systems have shown that friction on the cart creates oscillations. To explore this we will investigate the effects of friction by simulation and analysis.

A block diagram of a balance system with friction is shown in Figure 14.13a. To simulate the system we use Coulomb's model for friction, where the friction force is $F$ is given by

$$F = -\mu_{\mathrm{f}} M_{\mathrm{t}} g \,\mathrm{sgn}(v), \tag{14.34}$$

where $\mu_{\mathrm{f}} = 0.001$ is the coefficient for rolling friction, $M_{\mathrm{t}}$ is the total mass, $g$ is the acceleration of gravity, and $v$ is the cart velocity. We use the parameter values from Example 3.2, and the controller is the state space feedback in Example 7.7 with the slower closed loop poles. Results of a simulation of the system are shown in Figure 14.14a. The upper plots in the figure show the cart position $q$ (left) and the pendulum angle $\theta$ (right), and the lower plots show the cart velocity $v = \dot{q}$ (left) and the angular velocity of the pendulum $\dot{\theta}$ (right). The plots show clearly that there are oscillations with period $T_{\mathrm{p}} = 37$ s. The oscillation of the cart velocity has amplitude $A \approx 0.52$ m/s. The waveforms of the oscillations are far from sinusoidal, as can be seen in the plots on the right in Figure 14.14a.




(a) Pendulum angle and cart velocity (b) Nyquist plot of $G$

**Figure 14.14:** Time and frequency responses of the cart-pendulum system. (a) Time responses when the pendulum has an initial misalignment. (b) Frequency response of the transfer function $G(s)$ (solid line), given by (14.36), and the locus of the negative inverse $-1/N(a)$ (dashed line) of the describing function $N(a)$ for friction, which given by equation (14.35).

We can make a simple physical argument to understand how friction may cause oscillation. The pendulum is unstable and will start to fall for any perturbation. The control law then attempts to stabilize the system by applying a force to the cart, but the cart will remain stationary until the pendulum has fallen so much that the control signal is large enough to generate a force that is larger than the friction force. The cart then moves, causing the pendulum to move towards the upright position. The process will repeat itself creating an oscillation.

We will now use the describing function method, introduced in Section 10.5, to understand the behavior of the system. To do this we first use block diagram algebra to reduce Figure 14.13a to the two-block system in Figure 14.13b. One block represents the nonlinear friction model (14.34), which has the describing function

$$N(a) = \frac{4\mu_{\mathrm{f}} M_{\mathrm{t}} g}{a\pi}, \tag{14.35}$$

where $a$ is the amplitude of the input (cart velocity). The other block in Figure 14.13b represents the linear closed loop dynamics from friction force $F$ to cart velocity $v$, when friction is not present. The transfer function can be computed from the state space representation of the closed loop dynamics

$$\frac{d}{dt}x = (A - BK)x + BF, \qquad v = \begin{pmatrix} 0 & 0 & 1 & 0 \end{pmatrix} x,$$

where $x = (q, \theta, \dot{q}, \dot{\theta})$, $A$, $B$, and $K$ are given in Example 7.7. The resulting transfer function is given by

$$G(s) = \frac{0.01837s^3 - 0.08s}{s^4 + 1.046s^3 + 0.9109s^2 + 0.2552s + 0.03781}, \tag{14.36}$$

where the numerical values are based on the parameter values from Example 7.7.

Figure 14.14b shows a Nyquist plot of the transfer function $G$ (solid line) and the

negative inverse $-1/N(a)$ of the describing function (dashed line). Recall that the condition for oscillation is $G(i\omega)N(a) = -1$, which corresponds to an intersection of the solid and dashed lines in the figure. The intersection occurs for $\omega = 0.21$, and $1/N(a) = 0.39$. The describing function method then indicates that there may be an oscillation with period $T_\mathrm{p} = 2\pi/0.21 = 30$ s and amplitude $a = 4 \times 0.39\, \mu_\mathrm{f}\, M_\mathrm{t}\, g/\pi = 0.43$ m/s. Notice that the describing function method assumes that the velocity variation is sinusoidal, which explains the difference from the values $T = 37$ s and $a = 0.52$ m/s obtained by simulation. ∇

## 14.7 FURTHER READING

The effects of actuator limits are conveniently explored using optimal control theory [AF66, BH75], which permits solution of problems that are much more complicated than the one in Figure 14.12. The limitations caused by right half-plane poles and zeros were well known by Bode, who coined the term non-minimum phase to emphasize that such systems had much more phase lag than the equivalent minimum phase systems [Bod45]. The paper [Ste03], which is based on the inaugural IEEE Bode Lecture gives important insights into the effects of unstable poles and is strongly recommended. Horowitz [Hor63] also discussed the limits caused by poles and zeros in the right half-plane. The section on the maximum modulus theorem is based on [RÅ15]; more details are found in [GGS01, SP05]. The section on loop shaping design is based on [Åst00]. The design rules for pole placement are not widely known.

## EXERCISES

**14.1** (Large signals) Verify Figure 14.12 by hand calculation.

**14.2** (Noise limits bandwidth) Consider PI control of an integrator, where the transfer functions of the process and the controller are

$$P(s) = \frac{1}{s}, \qquad C(s) = k_\mathrm{p} + \frac{k_\mathrm{i}}{s},$$

and $k_\mathrm{p} = 2\zeta\omega_0$ and $k_\mathrm{i} = \omega_0^2$, with $\zeta = 0.707$. Assume that the inputs and outputs range from 0 to 10 V, that there is measurement noise with a standard deviation of 10 mV, and that the largest permissible variation in the control signal due to noise is 2 V. Show that the bandwidth, defined as $\omega_\mathrm{bw} = 2\omega_0$, cannot be larger than 283.

**14.3** (Effect of roll-off) Consider a closed loop system consisting of a first-order process and a proportional controller. Let the loop transfer function be

$$L(s) = P(s)C(s) = \frac{k}{s+1},$$

where parameter $k > 0$ is the controller gain. Show that the sensitivity function can be made arbitrarily small.

**14.4** (Bode's integral) In Theorem 14.3 it was assumed that $sL(s)$ goes to zero as $s \to \infty$. Assume instead that $\lim sL(s) = a$ and show that

$$\int_0^\infty \log |S(i\omega)|\, d\omega = \int_0^\infty \log \frac{1}{|1 + L(i\omega)|}\, d\omega = \pi \sum p_k - a\,\frac{\pi}{2},$$

where $p_k$ are the poles of the loop transfer function $L(s)$ in the right half-plane.

**14.5** (Integral formula for complementary sensitivity) Prove the formula (14.6) for the complementary sensitivity.

**14.6** (Water turbine dynamics) Consider the problem of power generation in an hydroelectric power station. Let the control signal be the opening area $a$ at the turbine entrance and $\ell$ be the length of the tube, which has area $A$. Formulate a mathematical model for the system, then linearize the model around a nominal valve opening $u_0 = a/A$ and a nominal power $P_0$. Show that the transfer function is

$$G(s) = \frac{P_0}{a_0}\,\frac{1 - 2u_0 s\tau}{1 + u_0 s\tau},$$

where $\tau = \ell/\sqrt{2gh}$ and $g$ is the acceleration due to gravity.

**14.7** (The pole/zero ratio) Consider a process with the loop transfer function

$$L(s) = k\frac{z - s}{s - p},$$

with positive $z$ and $p$. Show that the system is stable if $p/z < k < 1$ or $1 < k < p/z$ and that the largest stability margin is $s_m = |p - z|/(p + z)$, which is obtained for $k = 2p/(p + z)$. Determine the pole/zero ratios that give the stability margin $s_m = 2/3$.

**14.8** (Phase lag of systems with right half-plane pole/zero pair and delay and right half-plane pole) Consider the transfer functions for a process with a right half-plane pole and right half-plane zero in Example 14.5 and a right half-plane pole and a time delay in Example 14.6. The phase lags of their all-pass factors are given in equations (14.14) and (14.15). Show that the largest phase lags are

$$\begin{aligned}
\varphi_{\text{ap1}} &= -\arg P_{pz}(i\omega) \le 2\arctan\left(2\sqrt{pz}/|z - p|\right),\\
\varphi_{\text{ap2}} &= -\arg P_{p\tau}(i\omega) \le \sqrt{p\tau(2 - p\tau)} + 2\arctan\sqrt{p\tau/(2p - p\tau)}
\end{aligned}$$

and that they occur for $\omega_1 = \sqrt{pz}$ and $\omega_2 = \sqrt{2p/\tau - p^2}$ respectively.

**14.9** Consider a process with the transfer function

$$P(s) = \frac{(s + 3)(s + 200)}{(s + 1)(s^2 + 10s + 40)(s + 40)}.$$

Discuss suitable choices of closed loop poles for a design that gives dominant poles with undamped natural frequency 1 and 10.

**14.10** (Vehicle steering) Consider the Nyquist curve in Figure 14.10. Explain why part of the curve is approximately a circle. Derive a formula for the center and the radius and compare with the actual Nyquist curve.

**14.11** (X-29) A simplified model of the X-29 aircraft in a certain flight condition has a right-hand pole/zero pair with $p = 6$ rad/s and $z = 26$ rad/s. Estimate the achievable stability margins and compare with the results in Example 14.2.

**14.12** (Sensitivity inequalities) Prove the inequalities given by equation (14.21). (Hint: Use the maximum modulus theorem.)

**14.13** (Sensitivity limits due to poles in the right half-plane) Let $T_{\mathrm{r}} = M_{\mathrm{t}}\, b/(s+b)$ represent an upper bound on the desired sensitivity let $\omega_{\mathrm{tc}}$ represent the complementary sensitivity crossover frequency. Show that for a process $P(s)$ with a right half-plane pole $s = p$ but no other singularities in the right half-plane, the following inequalities hold:

$$b \geq \frac{p_{\mathrm{re}} + \sqrt{M_{\mathrm{t}}^2 p_{\mathrm{re}}^2 + (M_{\mathrm{t}}^2 - 1)p_{\mathrm{im}}^2}}{M_{\mathrm{t}}^2 - 1}, \qquad \omega_{\mathrm{tc}} \leq \frac{p_{\mathrm{re}} + \sqrt{M_{\mathrm{t}}^2 p_{\mathrm{re}}^2 + (M_{\mathrm{t}}^2 - 1)p_{\mathrm{im}}^2}}{\sqrt{M_{\mathrm{t}}^2 - 1}}. \tag{14.37}$$

**14.14** (Maximum complementary sensitivity for multiple right half-plane poles and zeros) Consider a process $P(s)$ with the right half-plane zeros $z_k$ and right half-plane poles $p_k$. Introduce the polynomial $n(s)$ with zeros $s = z_k$ and the polynomial $d(s)$ with zeros $s = p_k$. Show that the complementary sensitivity function has the property

$$M_{\mathrm{t}} \geq \max_k \left| \frac{n(-p_k)}{n(p_k)} \right|.$$

Also show that the equations (14.28) hold.

**14.15** (Right half-plane pole/zero pair lag control) Consider a process with the transfer functions

$$P(s) = \frac{s - z}{s - p}.$$

a) Show that the system can be controlled by a controller with the transfer function $C(s) = b/(s - a)$. Design such a controller that gives the closed loop poles $s = -\zeta\omega_0 \pm \omega_0\sqrt{1 - \zeta^2}$.

b) Calculate the maximum sensitivities of the closed loop system as a function of $\omega_{\mathrm{b}}$ and compare the bound imposed by the right half-plane poles and zeros of the system. Discuss the differences between the cases $z > p$ and $z < p$.

c) Plot the root locus of the process with the PI controller and describe qualitative the differences between the cases $z > p$ and $z < p$. Use the numerical values $\omega_0 = 2, \zeta = 1, p = 1, z = 5$ and $p = 5, z = 1$.

**14.16** (Time delay and a pole in the right half-plane) Consider a process with the transfer function

$$P(s) = \frac{e^{-s\tau}}{s - p}\bar{P}(s),$$

where $\bar{P}(s)$ has no poles and zeros in the right half-plane. Show that the sensitivity functions have the properties listed in Table 14.1:

$$M_{\mathrm{t}} \geq e^{p\tau}, \qquad M_{\mathrm{s}} \geq e^{p\tau} - 1.$$

**14.17** (Stabilization of an inverted pendulum with visual feedback) Consider stabilization of an inverted pendulum based on visual feedback using a video camera with a 50-Hz frame rate. Let the effective pendulum length be $l$. Assume that we want the loop transfer function to have a slope of $n_{\mathrm{gc}} = -1/2$ at the crossover frequency. Use the gain crossover frequency inequality to determine the minimum length of the pendulum that can be stabilized if we desire a phase margin of 45°.

**14.18** (Design with poor robustness) Consider a process with the transfer function

$$P(s) = \frac{1}{s+1}.$$

Design a PI controller that gives a closed loop system with the characteristic polynomial $s^2 + 0.2s + 0.01$. Plot the Nyquist curve of the loop transfer function and determine the gain, phase, and stability margins.

**14.19** (Rear-steered bicycle) Consider the simple model of a bicycle in equation (4.5), which has one pole in the right half-plane. The model is also valid for a bicycle with rear wheel steering, but the sign of the velocity is then reversed and the system also has a zero in the right half-plane. Use the results of Exercise 14.7 to give a condition on the physical parameters that admits a controller with the stability margin $s_{\mathrm{m}}$.

**14.20** (Right half-plane pole/zero pair PI control) Consider a process with the transfer function

$$P(s) = \frac{s-z}{s-p}.$$

a) Show that the system can be controlled by a PI controller and design a PI controller that gives a closed loop system with poles at $s = -\zeta\omega_0 \pm \omega_0\sqrt{1-\zeta^2}$.

b) Calculate the maximum sensitivity of the closed loop system as a function of $\omega_0$ and compare with the bound imposed by the the right half-plane poles and zeros of the system. Discuss the differences between the cases $z > p$ and $z < p$.

c) Plot the root locus of the process with the PI controller and qualitatively describe how it changes with the process pole and the process zero. Use the numerical values $\omega_0 = 1, \zeta = 1$, $p = 1$, $z = 5$ and $p = 5$, $z = 1$.

**14.21** (Right half-plane pole/zero pair lag control) A process with the transfer function

$$P(s) = \frac{s-0.1}{s-1}$$

cannot be controlled by a stable controller. Show that the controller

$$C = \frac{b}{s-2}$$

with $b = 6$ stabilizes the system. Plot the Nyquist curve of the closed loop system and determine the maximum sensitivities. Also plot the root locus with respect to the parameter $b$.