Technische Universität Kaiserslautern

Chapter 1

Introduction to robust control

Prof. Dr.-Ing. Steven Liu
Lehrstuhl für Regelungssysteme
Technische Universität Kaiserslautern

Uncertainty and feedback control (1)

A key reason for using feedback is to reduce the effects of uncertainty which may appear in different forms as disturbances, noises or as other imperfections in the models used for control. A robust control should be able to work well even uncertainties exist in the system to be controlled. Thus, robustness means insensitiveness or tolerance to system uncertainties in the sense: even if the mathematical model of the system is somewhat incorrect, the controlled system should be at least stable and perhaps also close to "optimal".

Model uncertainty and robustness have been a central theme in the development of feedback control. In fact, without uncertainty there is no need for feedback: if we have complete certainty already, what extra benefit would we obtain from control theory being the situation already under control? In the early stage of the development history of control, Black (feedback amplifier), Nyquist (stability criterion) or Bode (Bode's integrals and relations) have essentially contributed to the ideas and design methods of robust control using transfer function, frequency response and graphical techniques. Later, Horowitz generalized Bode's work on robust design and introduced the Qualitative Feedback Theory (QFT). In the 1960s the state-space theory represented a significant paradigm change. New insight, new concepts and new design methods based on rigorous, analytical descriptions and formulated as optimization problems (e.g. LQG, Kalman filter) were established. They gained large success in control of linear systems with deterministic or stochastic disturbances, benefiting from advances in numerical linear algebra and efficient software. On the

Uncertainty and feedback control (2)

other hand, however, it is not straightforward to capture model uncertainty in a state variable setting.

In 1980s, another paradigm shift came into the control design which brought robustness to the forefront. It started a new development that led to the so called $\mathcal{H}_\infty$ theory. The idea was to develop systematic design methods that were guaranteed to give closed-loop stability and some performance specification for systems with model uncertainty. The problem was successfully addressed both using transfer functions and in state-space formulation. Game theory is another approach to $\mathcal{H}_\infty$ theory. Newer results bring $\mathcal{H}_\infty$ even closer to the classical control design.

Another approach to deal with unknown dynamics and model uncertainties is the use of adaptive control. In this case the controller adapts itself to the changing conditions or uncertainties. Adaptive control is different from robust control in the sense that it does not need a priori information about the bounds on these uncertain or time-varying parameters; robust control guarantees that if the changes are within given bounds the control law need not be changed, while adaptive control is precisely concerned with control law changes. We'll not discuss adaptive control in this lecture.

How can we deal with uncertainty?

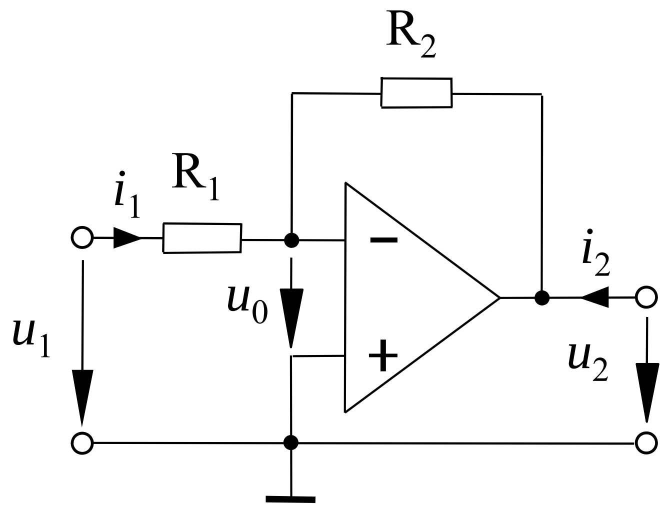
– Live with it: Robust control!

– Reduce it: Adaptive control!

Example 1.1: Black's feedback amplifier

$$u_1 = R_1 i_1 + u_0$$
$$u_2 = R_2 i_2 + u_0$$
$$u_2 = -V u_0$$
$$i_1 = -i_2$$

$$\Rightarrow \quad \frac{u_2}{u_1} = -\frac{R_2}{R_1} \frac{1}{1 + \frac{1}{V}\left(1 + \frac{R_2}{R_1}\right)}$$



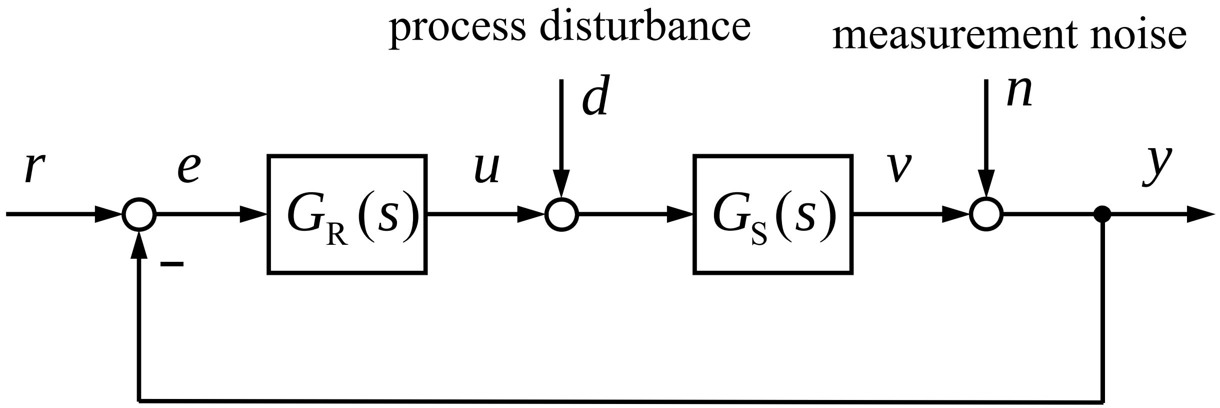
Let the open-loop gain of the amplifier be V. The input-output relation, as described by the gain relation u_2/u_1 above, is essentially given by the ratio R_2/R_1. If the raw amplifier gain V is large enough the gain of the feedback amplifier is virtually independent of V. Assume for example that $R_2/R_1 = 100$ and $V = 10^4$. A 10% change of V gives only a gain variation of 0.1%. Feedback thus has the amazing property of reducing the effects of uncertainty dramatically. Also the amplifier linearity is increased significantly.

The risk for instability is a drawback of feedback. This is the issue we always have to care about. A single feedback amplifier, however, is always stable.

Standard feedback control loop

Basic requirements:

- Stability
- Attenuation of process disturbances
- Reduction of effects of measurement noise
- Good tracking of reference signal



There are three inputs r, d, and n, and four interesting signals v, y, e, and u. They form in total 12 possible transfer functions

$$G_{vr} = \frac{G_R G_S}{1+G_R G_S} \qquad G_{yr} = G_{vr} \qquad G_{er} = 1 - G_{vr} = G_{yn} \qquad G_{ur} = \frac{G_R}{1+G_R G_S}$$

$$G_{vd} = \frac{G_S}{1+G_R G_S} \qquad G_{yd} = G_{vd} \qquad G_{ed} = -G_{vd} \qquad G_{ud} = -G_{vr}$$

$$G_{vn} = -G_{vr} \qquad G_{yn} = \frac{1}{1+G_R G_S} \qquad G_{en} = -G_{yn} \qquad G_{un} = -G_{ur}$$

Sensitivity and complementary sensitivity function

Only four transfer functions are really independent ("*The Gang of Four*"):

$$G_{yr}(s)=\frac{G_R G_S}{1+G_R G_S}=T(s) \qquad \text{complementary sensitivity function}$$

$$G_{yd}(s)=\frac{G_S}{1+G_R G_S} \qquad \text{process disturbance sensitivity function}$$

$$G_{yn}(s)=\frac{1}{1+G_R G_S}=S(s) \qquad \text{sensitivity function} \tag{1.1}$$

$$G_{ur}(s)=\frac{G_R}{1+G_R G_S} \qquad \text{noise sensitivity function}$$

The *sensitivity function* S and the *complementary sensitivity function* T are particularly important for robustness analysis of closed-loop control systems. Both S and T depend only on the loop transfer function $L=G_R G_S$. The term complementary sensitivity for T follows from the identity

$$S(s)+T(s)=1 \tag{1.2}$$

It is $\big|\,|S| - |T|\,\big| < |S+T| = 1$ at any frequency so the absolute values of S and T differ at most by 1.

Stability and stability margins

Many properties of the standard control loop can be derived from the loop transfer function.

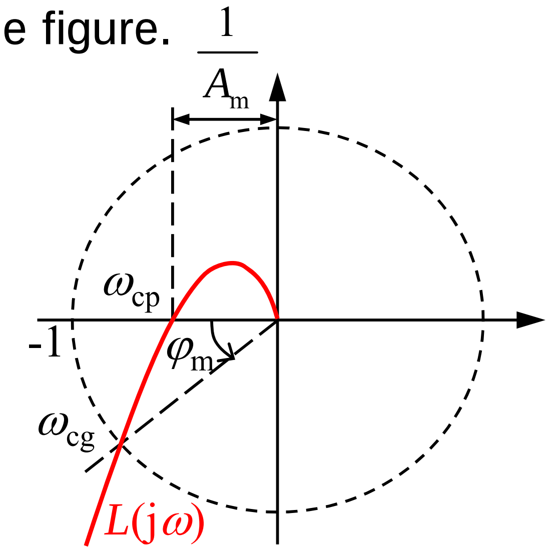
The stability of the system can be varified by Nyquist's stability criterion which says that the closed-loop system is stable if and only if the Nyquist plot of L for $-\infty < \omega < \infty$, encircles the point $-1+j0$ as many times anticlockwise as L has RHP poles. That is, using the term of argument variation,

$$\Delta \arg(1+L(j\omega)) = (2n_r + n_i)\pi \tag{1.3}$$

where $\Delta\arg$ is the argument variation when ω goes from $-\infty$ to $+\infty$ and n_r, n_i are the number of poles of L in the right half plane (RHP) and on the imaginary axis respectively.

Stability is normally investigated by analyzing the Nyquist plot, as shown in the figure.

To achieve stability the Nyquist plot must be sufficiently far away from the critical point -1. the distance from the critical point can also be used as a measure of the degree of stability, i.e., amplitude margin A_m and phase margin φ_m. An amplitude margin A_m implies that the gain can be increased with a factor less than A_m without making the system unstable. Similarly, for a system with a phase margin φ_m it is possible to increase the phase shift in the loop by a quantity less than φ_m without making the system unstable.



Small process variations

We now consider the effects of small variations in the process (plant).

As shown before, the signal transmission of the closed-loop system from reference r to output y is described by the complementary sensitivity function

$$T(s) = \frac{G_\mathrm{R} G_\mathrm{S}}{1+G_\mathrm{R} G_\mathrm{S}}$$

The variation of T can be calculated as

$$\frac{\mathrm{d}T}{T} = \frac{\mathrm{d}G_\mathrm{S}}{G_\mathrm{S}} - \frac{G_\mathrm{R}\mathrm{d}G_\mathrm{S}}{1+G_\mathrm{R} G_\mathrm{S}} = \frac{1}{1+G_\mathrm{R} G_\mathrm{S}} \cdot \frac{\mathrm{d}G_\mathrm{S}}{G_\mathrm{S}} = S\frac{\mathrm{d}G_\mathrm{S}}{G_\mathrm{S}} \tag{1.4}$$

Thus, the sensitivity function S tells us how the closed-loop properties are influenced by small variations in the process. As robustness measures the maximum sensitivities

$$M_S = \max\left|S(\mathrm{j}\omega)\right|, \qquad M_T = \max\left|T(\mathrm{j}\omega)\right| \tag{1.5}$$

are also used (we later introduce $M_S = \|S\|_\infty$ and $M_T = \|T\|_\infty$ in terms of $\mathcal{H}_\infty$ norm). The value $1/M_S$ can be seen as the *shortest distance* between the Nyquist plot and the critical point -1. Typically, it is required that M_S is less than about 2 (6 dB) and M_T is less than about 1.25 (2 dB). A large value of M_S or M_T (larger than about 4) indicates poor performance as well as poor robustness.

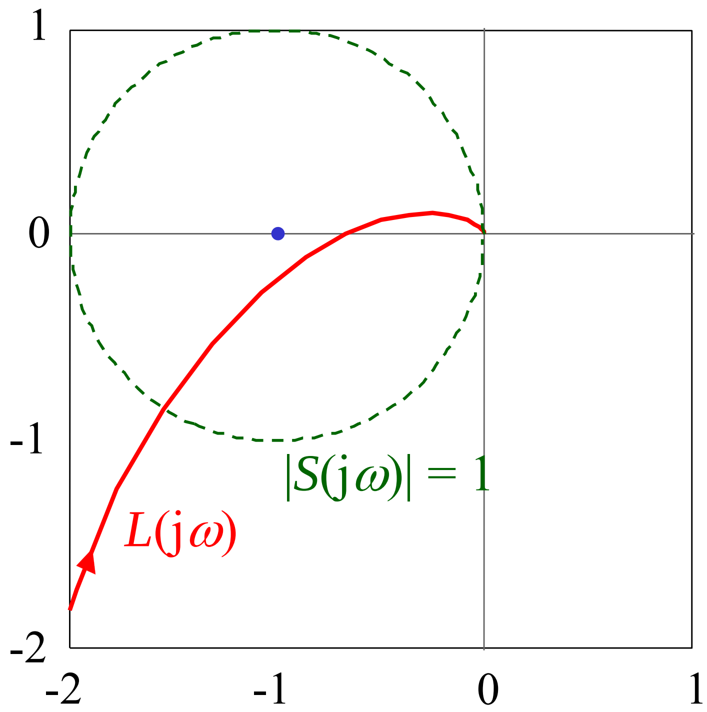
Influence of feedback on disturbance

The sensitivity function also has other physical interpretations. Consider the standard control loop. If there is no feedback the Laplace transform of the output is Y_{ol}. The output under closed-loop control is given by

$$Y_{cl} = \frac{1}{1+G_{\mathrm{R}}G_{\mathrm{S}}} Y_{ol}$$

It follows that

$$\frac{Y_{cl}}{Y_{ol}} = \frac{1}{1+G_{\mathrm{R}}G_{\mathrm{S}}} = S \qquad (1.6)$$



Thus, the sensitivity function tells us also how the disturbances are influenced by feedback. Disturbances with frequencies such that $|S(\mathrm{j}\omega)|$ is less than one are reduced by an amount equal to the distance to the critical point and those with frequencies such that $|S(\mathrm{j}\omega)|$ is larger than one are amplified by the feedback, as illustrated in the figure.

Maximum sensitivities and performance (1)

The importance of bounding the value of M_S and M_T can be also explained from different points of view. We first consider some simple relations between the maximum sensitivities and the stability margins.

At the phase crossover frequency we have

$$L(\mathrm{j}\omega_{cp}) = -\frac{1}{A_\mathrm{m}}$$

and therefore

$$T(\mathrm{j}\omega_{cp}) = \frac{L(\mathrm{j}\omega_{cp})}{1+L(\mathrm{j}\omega_{cp})} = \frac{-1}{A_\mathrm{m}-1} \le M_T \;;$$

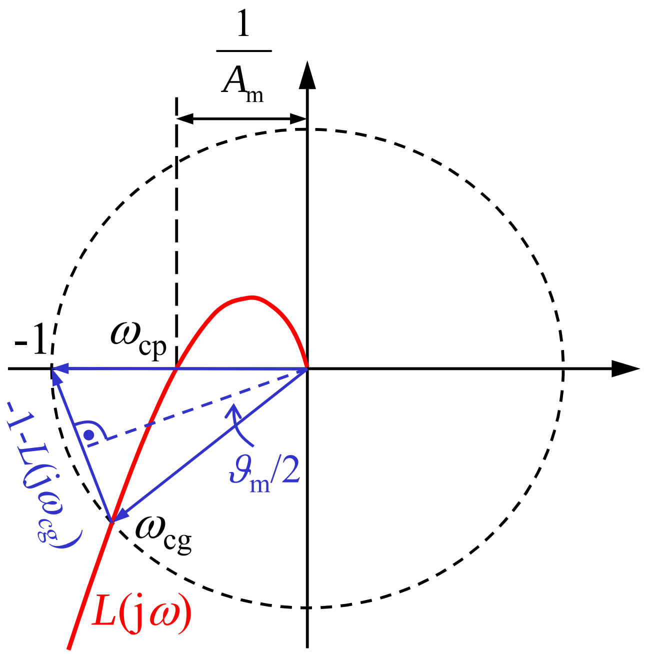
$$S(\mathrm{j}\omega_{cp}) = \frac{1}{1+L(\mathrm{j}\omega_{cp})} = \frac{A_\mathrm{m}}{A_\mathrm{m}-1} \le M_S$$

It follows then

$$A_\mathrm{m} \ge \frac{M_S}{M_S-1} \quad \text{and} \quad A_\mathrm{m} \ge 1-\frac{1}{M_T}$$

As well as

$$\varphi_\mathrm{m} \ge 2\arcsin\left(\frac{1}{2M_S}\right) \ge \frac{1}{M_S} \quad \text{and} \quad \varphi_\mathrm{m} \ge 2\arcsin\left(\frac{1}{2M_T}\right) \ge \frac{1}{M_T}$$



Maximum sensitivities and performance

Another index of control performance is the ability of disturbance rejection. Without control ($u = 0$) the error signal is

$$e = r - y = r - G_S d$$

and with feedback control

$$e = r - y = S(r - G_S d)$$

Thus, feedback control improves performance in terms of reducing $|e|$ at all frequencies where $|S| < 1$. Usually, $|S|$ is small at low frequencies: for example $|S(0)| = 0$ for systems with integral action. Because all real systems are strictly proper we must have $|L| \to 0$ or $|S| \to 1$ at high frequencies. At intermediate frequencies we can not avoid in practice a peak value, M_S, larger than 1 (explained later). Thus, in the frequency range near M_S the control performance is degraded, and the value of M_S is naturally a measure of the worst-case performance degradation.

In summary, both for stability and performance we want to have M_S close to 1.

Large process variations (1)

Eq. (1.4) gives the sensitivity for small perturbations of the process. To see how much the process can change (large process variations) without making the closed-loop system unstable we again use the Nyquist diagram. Consider a point on the Nyquist plot. The distance to the critical point is $|1+L|$. If the process changes by ΔG_S, the point changes by $G_R \Delta G_S$. The system will remain stable as long as

$$\left|G_R \Delta G_S\right| < \left|1+G_R G_S\right| \tag{1.7}$$

and the number of right hand poles and the poles on j-axis of $G_R G_S$ does not change. This implies that the perturbations must have the property that ΔG_S is a stable transfer function.

The closed-loop system will remain stable as long as

$$\left|\frac{\Delta G_S}{G_S}\right| < \left|\frac{1+G_R G_S}{G_R G_S}\right| = \left|\frac{1}{T}\right| \tag{1.8}$$

A conservative estimation for the allowable variation in process dynamics is then

$$\left|\Delta G_S\right| \le \frac{\left|G_S\right|}{M_T} \tag{1.9}$$

The largest value M_T of the complementary sensitivity function T is therefore a simple measure of robustness to process variations. It implies that the variations can be large for those frequencies where

Large process variations (2)

T is small and that smaller variations are allowed for frequencies where T is large.

A similar estimate based on the maximum sensitivity is that

$$\left|\Delta G_{\mathrm{S}}\right| < \left|\frac{1}{SG_{\mathrm{R}}}\right| \leq \frac{1}{M_S \left|G_{\mathrm{R}}\right|} \tag{1.10}$$

It follows from Eqs (1.9) and (1.10) that it would be highly desirable to make both the sensitivity functions S and T as small as possible. This is unfortunately not possible because of the complementary property (1.2). Besides, there are also other constraints on the sensitivities. For example, if p is a RHP pole of $L(s)$ then

$$T(p) = 1\,, \quad S(p) = 0 \tag{1.11}$$

Similarly, if z is a RHP zero of $L(s)$ then

$$T(z) = 0\,, \quad S(z) = 1 \tag{1.12}$$

These *interpolation constraints* also clearly restrict the allowable S and T. Further, more important constraints are discussed in the following.

The waterbed effect: sensitivity integrals

A typical sensitivity function S has a peak value greater than 1: this is unavoidable in practice. It was shown by Bode (*Bode's sensitivity integral*) that if the closed-loop system is stable then

$$\int_0^\infty \ln|S(\mathrm{j}\omega)|\mathrm{d}\omega = \frac{\pi}{2}\left(-\lim_{s\to\infty}(sL(s)) + 2\sum p_i\right) \tag{1.13}$$

where p_i are the possible RHP poles of L. If $L(s)$ has *at least a relative degree of two* Eq. (1.13) (*waterbed formula*) is further simplified to

$$\int_0^\infty \ln|S(\mathrm{j}\omega)|\mathrm{d}\omega = \pi\sum p_i \tag{1.14}$$

Middleton showed (1991) for the complementary sensitivity function the dual result

$$\int_0^\infty \ln|T(\mathrm{j}\omega)|\frac{\mathrm{d}\omega}{\omega^2} = \pi\sum \frac{1}{z_i} \tag{1.15}$$

where z_i are the right half plane zeros of L. The equations (1.13) - (1.15) indicate that if we push the sensitivities down at some frequencies then it will have to increase at others. This phenomena is sometimes called the *waterbed effect*. It also follows from the equations that the presence of poles in the RHP increases the sensitivity and that zeros in the right half plane increase the complementary sensitivity. A fast RHP pole gives higher sensitivity than a slow pole, and a slow RHP zero gives higher sensitivity than a fast zero.

Bode's gain-phase relation (1)

The magnitude and phase diagrams are also related, at least for *minimum phase systems*. It is not possible to achieve high phase advance without using high gains and it is not possible to obtain transfer functions that decrease rapidly without having large phase lags.

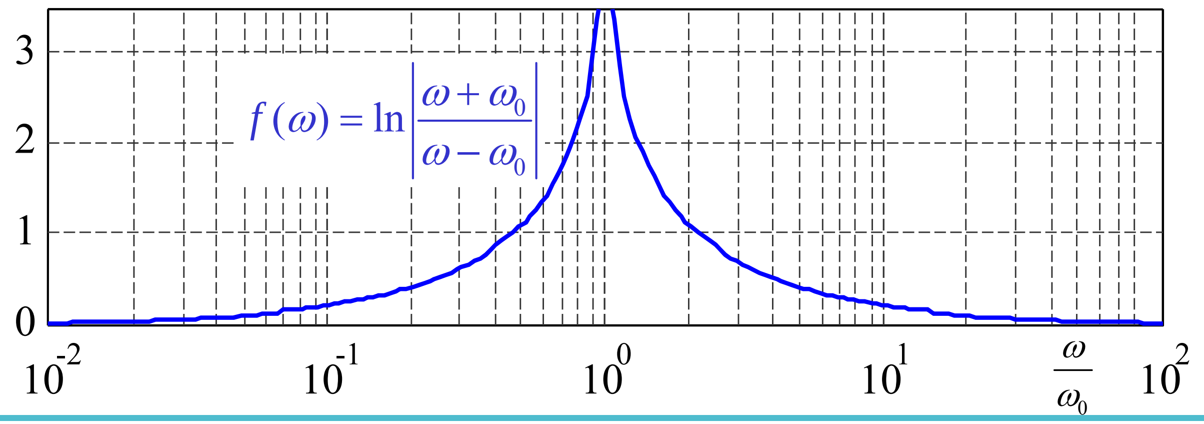
Consider a stable and minimum phase transfer function $G(s)$ with $G(\mathrm{j}\omega) = A(\omega)\mathrm{e}^{\mathrm{j}\Phi(\omega)}$ and $G(s) > 0$. It was shown by Bode that

$$\Phi(\omega_0) = \frac{2\omega_0}{\pi}\int_0^\infty \frac{\ln A(\omega) - \ln A(\omega_0)}{\omega^2 - \omega_0^2}\mathrm{d}\omega = \frac{1}{\pi}\int_0^\infty \frac{\mathrm{d}(\ln A(\omega))}{\mathrm{d}(\ln\omega)} \ln\left|\frac{\omega+\omega_0}{\omega-\omega_0}\right|\frac{\mathrm{d}\omega}{\omega}$$

$$= \frac{1}{\pi}\int_0^\infty c(\omega) f(\omega)\frac{\mathrm{d}\omega}{\omega} \qquad (1.16)$$

where the ratio $c(\omega) = \mathrm{d}(\ln|G(\mathrm{j}\omega)|)/\mathrm{d}(\ln\omega)$ is the slop of the Bode plot and the following relation holds for the weighting kernel $f(\omega)$:

$$\int_0^\infty \ln\left|\frac{\omega+\omega_0}{\omega-\omega_0}\right|\frac{\mathrm{d}\omega}{\omega} = \frac{\pi^2}{2} \qquad (1.17)$$



Bode's gain-phase relation (2)

An approximate version of (1.16) for $c(\omega) \approx \text{const.}$ based on (1.17) is that

$$\Phi(\omega) \approx \frac{\pi}{2} c(\omega) \tag{1.18}$$

This means that if the slope of the magnitude diagram is constant the phase is roughly $c\pi/2$!

Bode's gain-phase relations impose fundamental limitations on the performance that can be achieved by a feedback control. A simple observation is that even if it is desirable that the loop gain decreases rapidly at the crossover frequency, it is not possible to have a steeper slope than -2 without violating stability constraints.

Bode suggested an ideal shape of the loop transfer function which has the form

$$L(s) = \left(\frac{s}{\omega_{cg}}\right)^c \quad c < 0 \tag{1.19}$$

where ω_{cg} is the gain crossover frequency. The Nyquist curve for this loop transfer function is simply just a straight line through the origin with $\arg\{L(j\omega)\} = c\pi/2$. The relation (1.19) is also called *Bode's ideal loop transfer function*.

One special reason why Bode made this particular choice of $L(s)$ is that it gives a close-loop system

Bode's gain-phase relation (3)

that is insensitive to gain changes. Variations in the process gain will change the crossover frequency while the phase margin is $\varphi_m = \pi(1+c/2)$ for all values of the gain. The gain margin is infinite. The slopes $c = -1.333$, -1.5, and -1.667 correspond to phase margins of $60°$, $45°$, and $30°$. Bode's idea to use loop shaping to design controller that are insensitive to gain variations of the plant were later generalized by Horowitz to systems that are insensitive also to other variations of the plant (QFT method).

The transfer function given by (1.19) is an irrational transfer function for non-integer c. It can be approximated arbitrarily close by rational transfer functions. It is usually sufficient to approximate L over a frequency range around the desired crossover frequency ω_{cg}. Assume for example that the gain of the process varies between $k_{\min}$ and $k_{\max}$ and that it is desired to have a loop transfer function that is close to (1.19) in the frequency range $(\omega_{\min}, \omega_{\max})$. It follows from (1.19) that

$$\frac{\omega_{\max}}{\omega_{\min}} = \left(\frac{k_{\min}}{k_{\max}}\right)^{1/c} \tag{1.20}$$

For $c = -5/3$ and a gain ratio of 100 we obtain a frequency ratio of about 16 and for $c = -4/3$ we get a frequency ratio of 32. To obtain a gain variation range as large as possible it is thus useful to have c as small as possible. There is, however, a compromise since the phase margin decreases with decreasing c and the system becomes unstable for $c = -2$!

Loop shaping

Classical controller design in the frequency domain is usually based on the loop shaping method, that is, the open-loop transfer function $L(\mathrm{j}\omega)$ is shaped to obtain desired frequency-domain specifications such as crossover frequencies, phase margin, slopes, etc. Essentially, to get the benefits of feedback control we want the loop gain, $|L(\mathrm{j}\omega)|$, to be as large as possible in the bandwidth region. However, due to time delays, RHP zeros, high-frequency noises and limitations on the allowed manipulated inputs, the loop gain has to drop below 1 at and above some frequency (crossover frequency). Thus, disregarding stability for the moment, it is desirable that $|L(\mathrm{j}\omega)|$ falls sharply with frequency (roll-off).

The shape of $L(\mathrm{j}\omega)$ is most crucial and difficult in the crossover region between ω_{cg} (where $|L| = 1$) and ω_{cp} (where $\angle L = -180°$). For stability, we at least need the loop gain to be less than 1 at ω_{cp}. Thus, to get a high bandwidth (fast response) we want ω_{cg} and therefore ω_{cp} large, that is, we want the phase lag in L to be small. Unfortunately, this is not consistent with the desire that $|L|$ should fall sharply. In addition, if the slope is made steeper at higher frequencies, then this will add unwanted phase lag at intermediate frequencies, thus degrading the stability. The situation becomes even worse for cases with delays or RHP zeros in $L(s)$ which add undesirable phase lag to $L(\mathrm{j}\omega)$ without contributing to a desirable negative slope in $L(\mathrm{j}\omega)$.

Typically, a desired loop shape for $L(\mathrm{j}\omega)$ shows a slope of about -1 in the crossover range, and a slope of -2 or higher beyond this region. At low frequencies, the desired shape of $|L(\mathrm{j}\omega)|$ depends on what disturbances and references we are designing for and includes usually at least one integrator.

Example 1.2: Loop shaping (1)

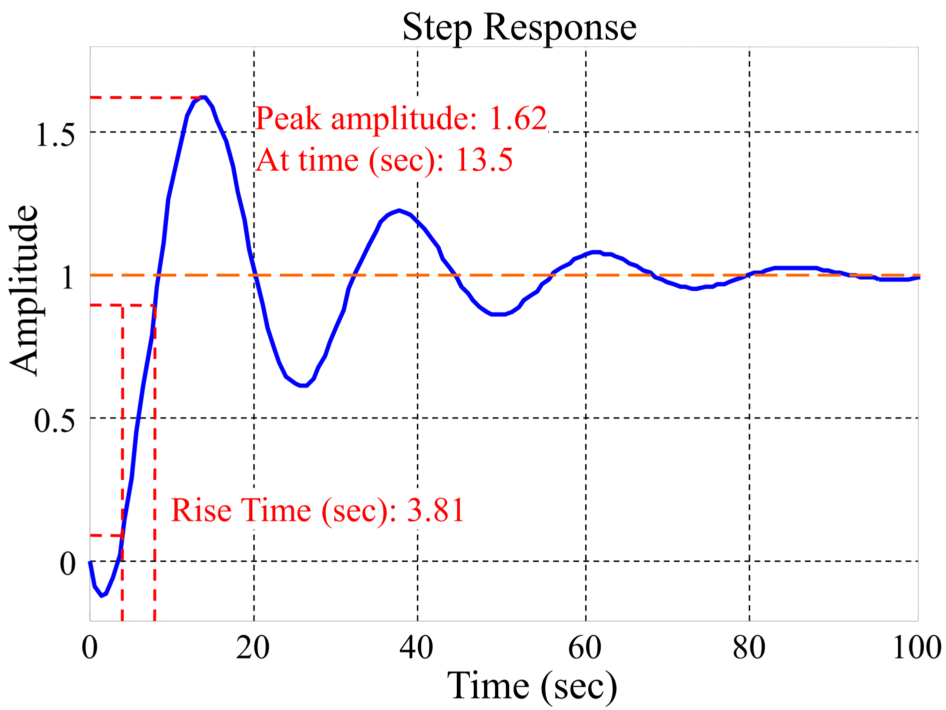
Exp. 1.2: We consider the control of a dynamic process described by the plant transfer function

$$G_S(s) = \frac{3(-2s+1)}{(10s+1)(5s+1)}$$

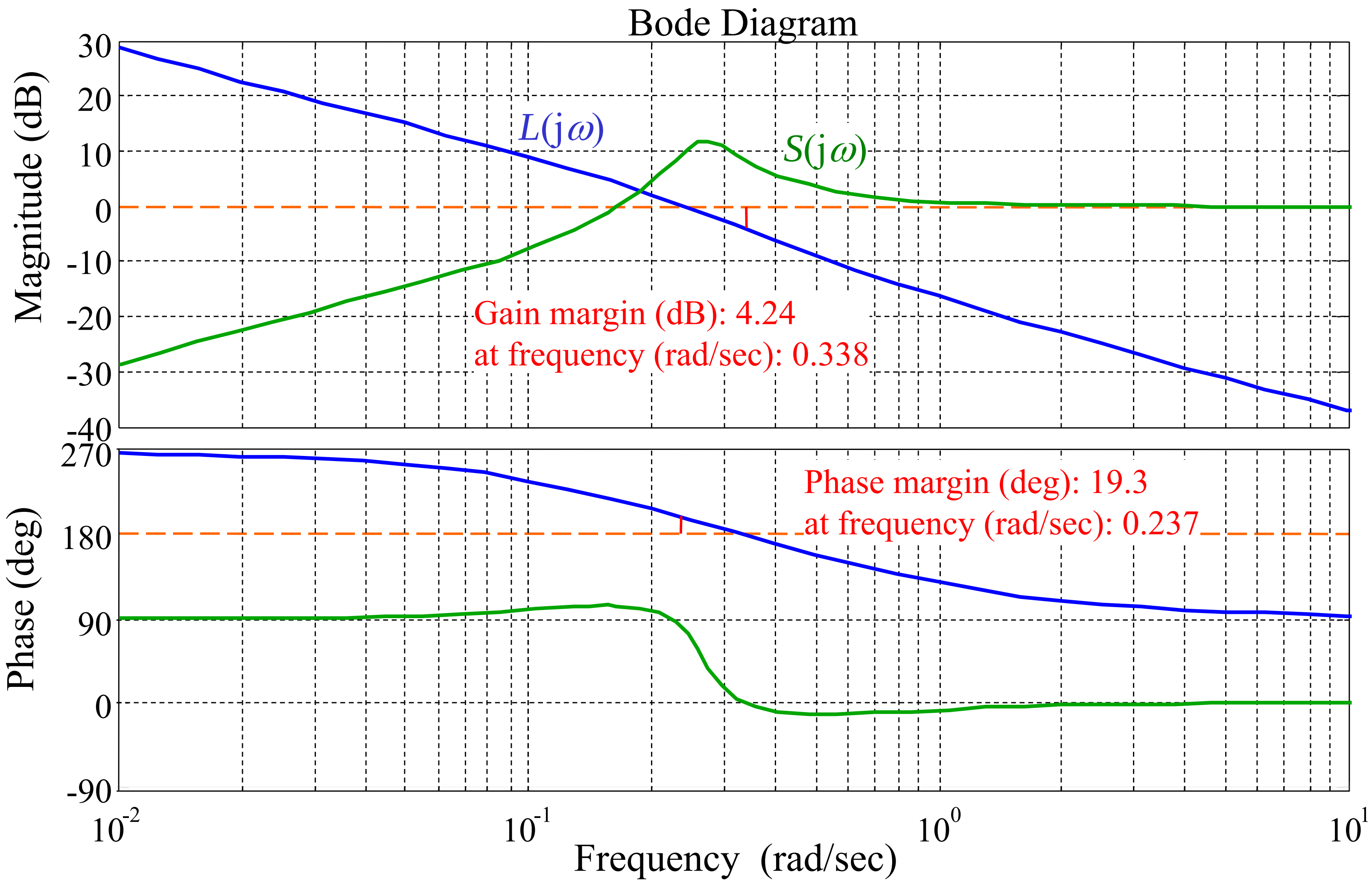
We first choose a PI type controller which eliminates the steady-state control error. If the classical tuning rules of Ziegler and Nichols are used for determining the controller parameters, then we have

$$G_R(s) = 1.14 \cdot (1 + \frac{1}{12.7s})$$

The closed-loop response to a step change in reference, shown in the figure below, is quite oscillatory and the overshoot is about 62% which is much larger than one would normally like for reference tracking. The stability margins are $A_m \approx 4.24$ dB and $\varphi_m \approx 19.3$ deg, and we can see from the loop frequency response $L(j\omega)$ that the sensitivity is pretty poor. The maximum value of the complementary sensitivity is $M_S \approx 3.94 \approx 11.9$ dB.



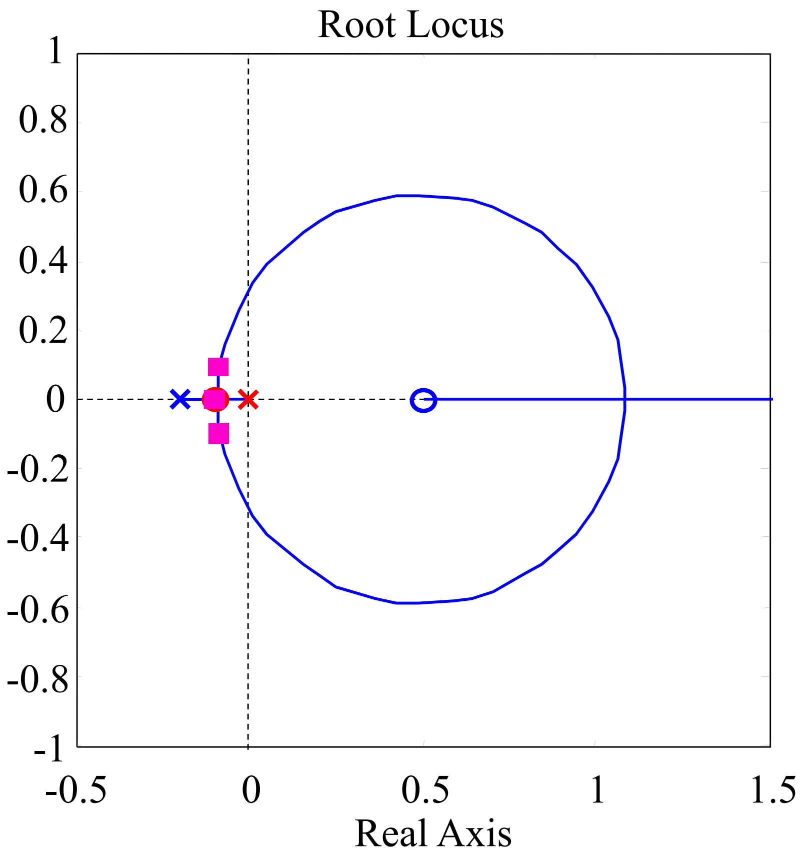
Example 1.2: Loop shaping (2)

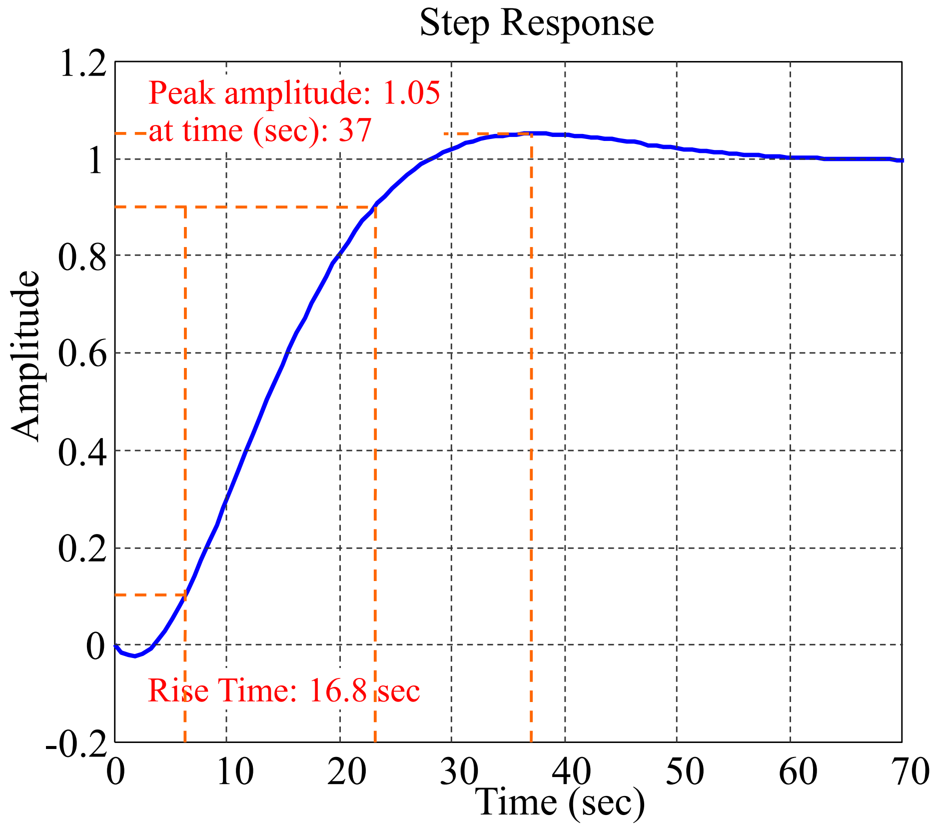


Example 1.2: Loop shaping (3)

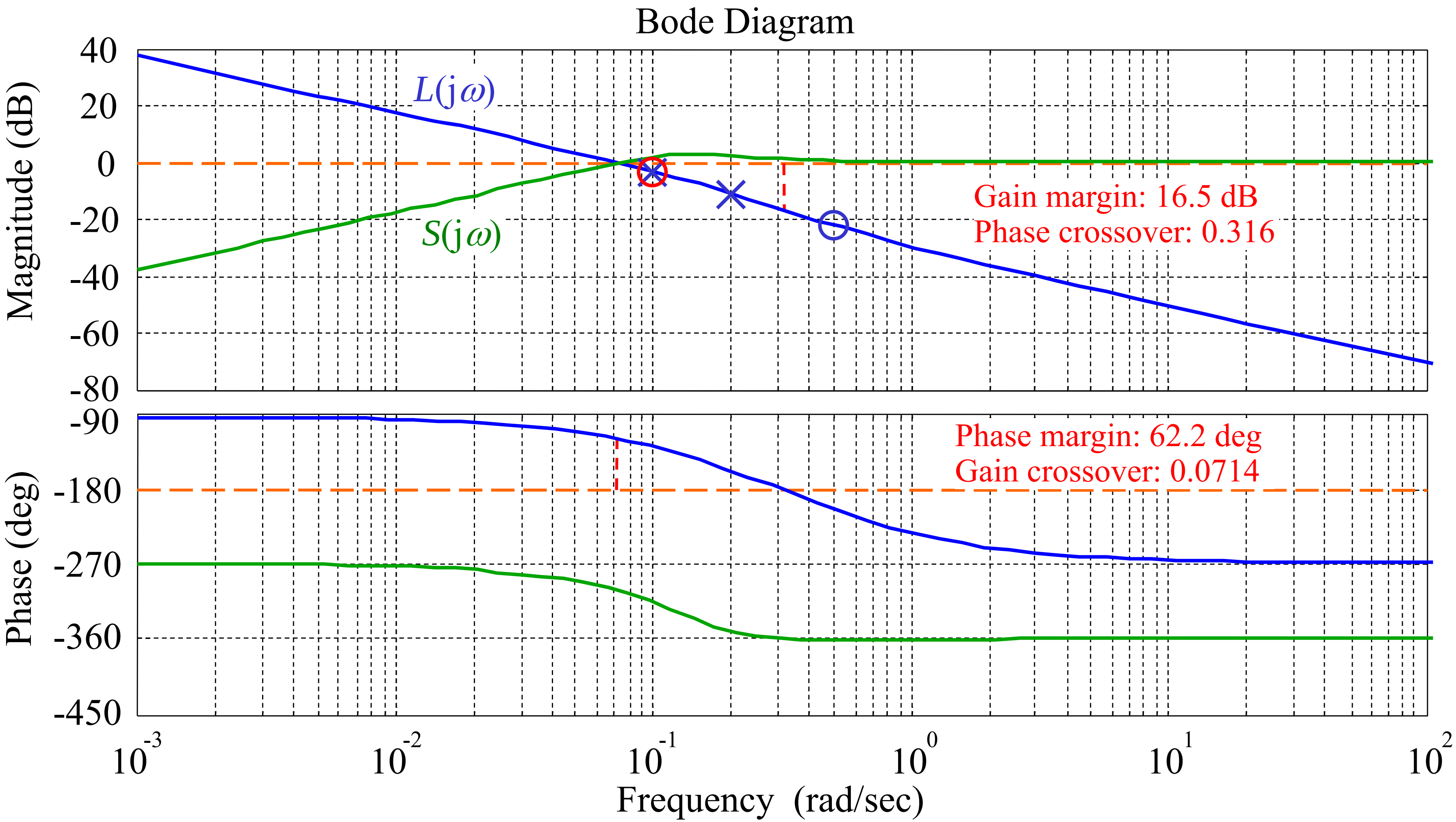
We now design a PI controller using root locus method. For convenience the slower plant pole is cancelled by the controller zero. From the root locus we choose

$$G_R(s) = 0.25 \cdot \left(1 + \frac{1}{10s}\right)$$





Example 1.2: Loop shaping (4)



Example 1.2: Loop shaping (5)

We now design a controller using loop shaping method. Since there is a RHP zero at $s = 0.5$ which cannot be cancelled by the controller, the loop function L must therefore contain this zero of G_S. This zero limits the achievable bandwidth and so the crossover region will be about 0.5 rad/sec. We require the controller to have one integrator and thus a reasonable approach is to shape the loop transfer function having a slope of -1 at low frequencies, and then rolling off with a higher slope at frequencies beyond 0.5 rad/sec. A possible choice of the loop transfer function is then

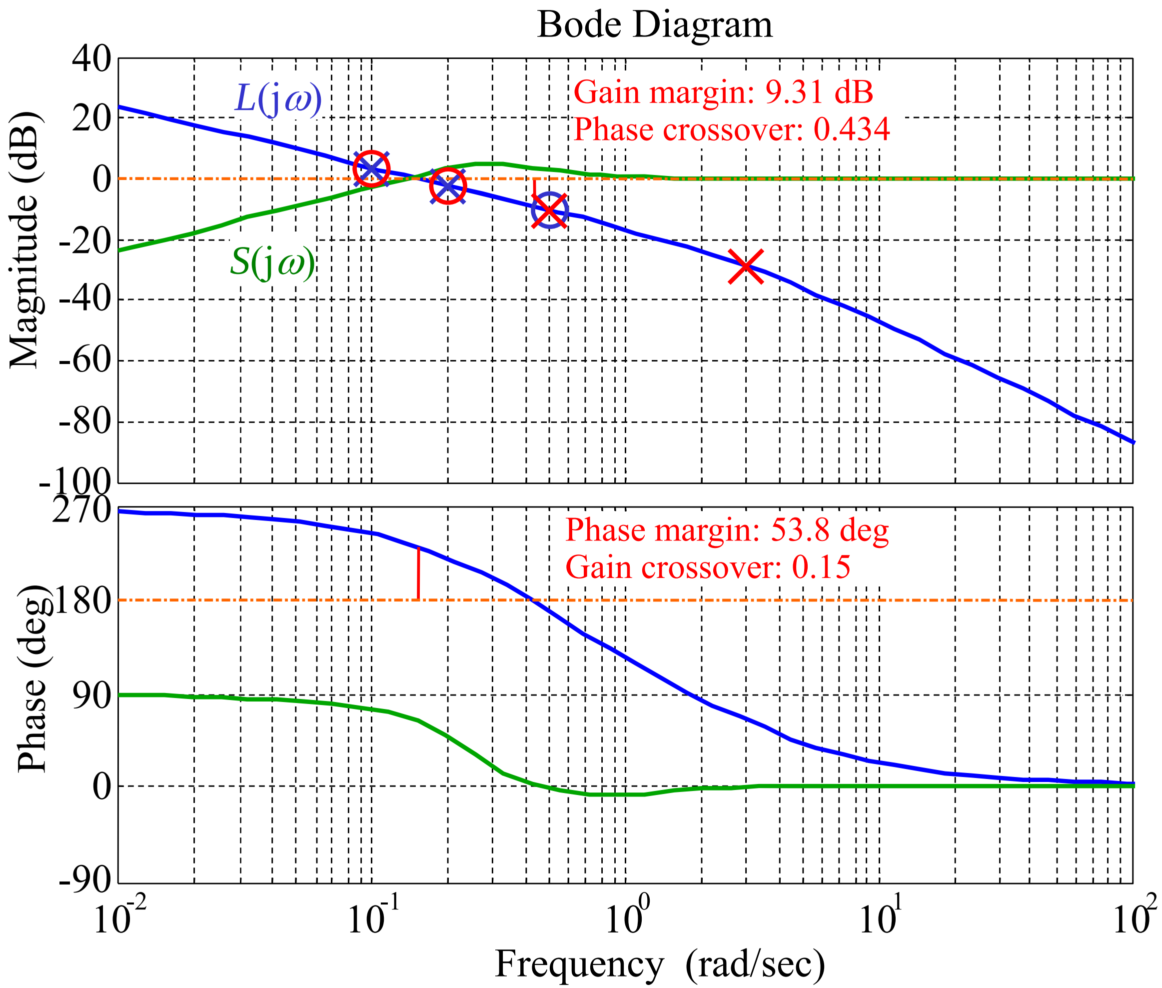
$$L(s) = \frac{K_L(-2s+1)}{s(2s+1)(s/3+1)}$$

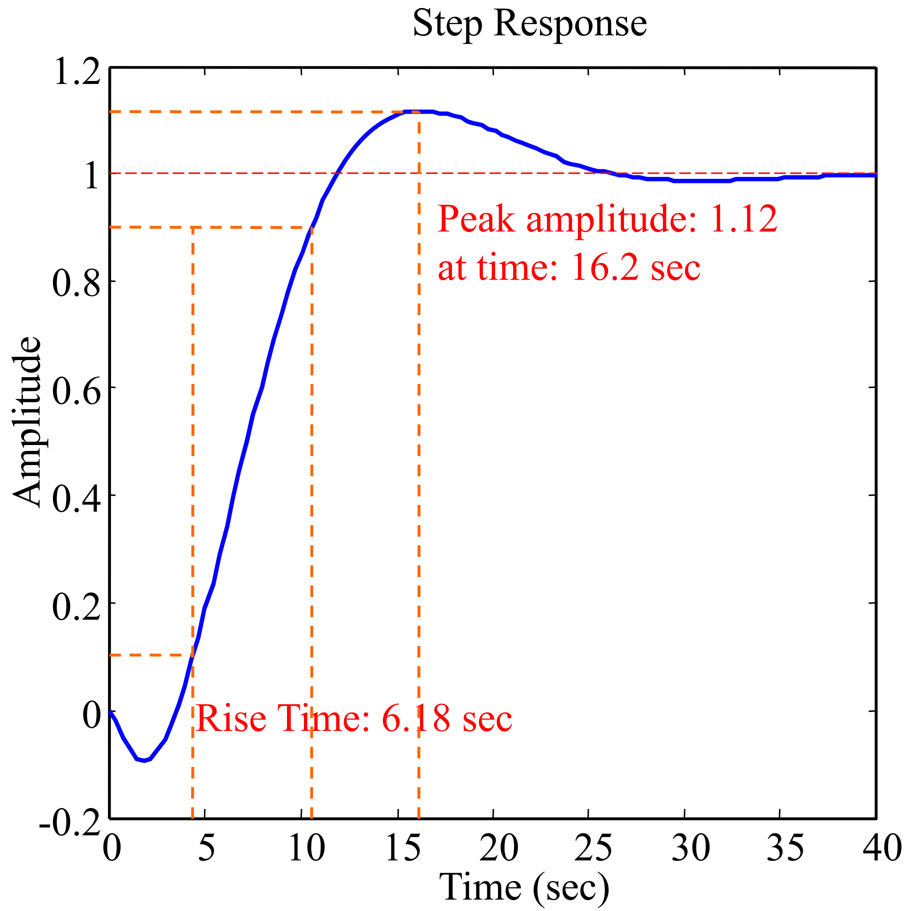
With $K_L = 0.15$ we obtain the following controller

$$G_R(s) = 0.05\frac{(10s+1)(5s+1)}{s(2s+1)(s/3+1)}$$

The controller zeros cancel the plant poles since we don't want the slope to drop before crossover. The all-pass term $(-2s+1)/(2s+1)$ is introduced to compensate for the amplitude effect of the plant RHP zero as much as possible. It causes, however, a relatively strong additional phase lag. Thus, the controller gain must be chosen so that the crossover is somewhat below 0.5 rad/sec to provide enough phase margin. The third controller pole, located beyond the crossover region, leads to a stronger roll off of L. The maximum value of the sensitivity function is $M_S \approx 1.72 \approx 4.71$ dB.

Example 1.2: Loop shaping (6)





Example 1.2: Loop shaping (7)

More phase margin can be obtained if we move the second controller pole towards higher frequency even though the influence of the RHP zero will somehow increase in the time response. A possible choice is for example

$$L(s) = \frac{0.18(-2s+1)}{s(s+1)(s/3+1)}$$

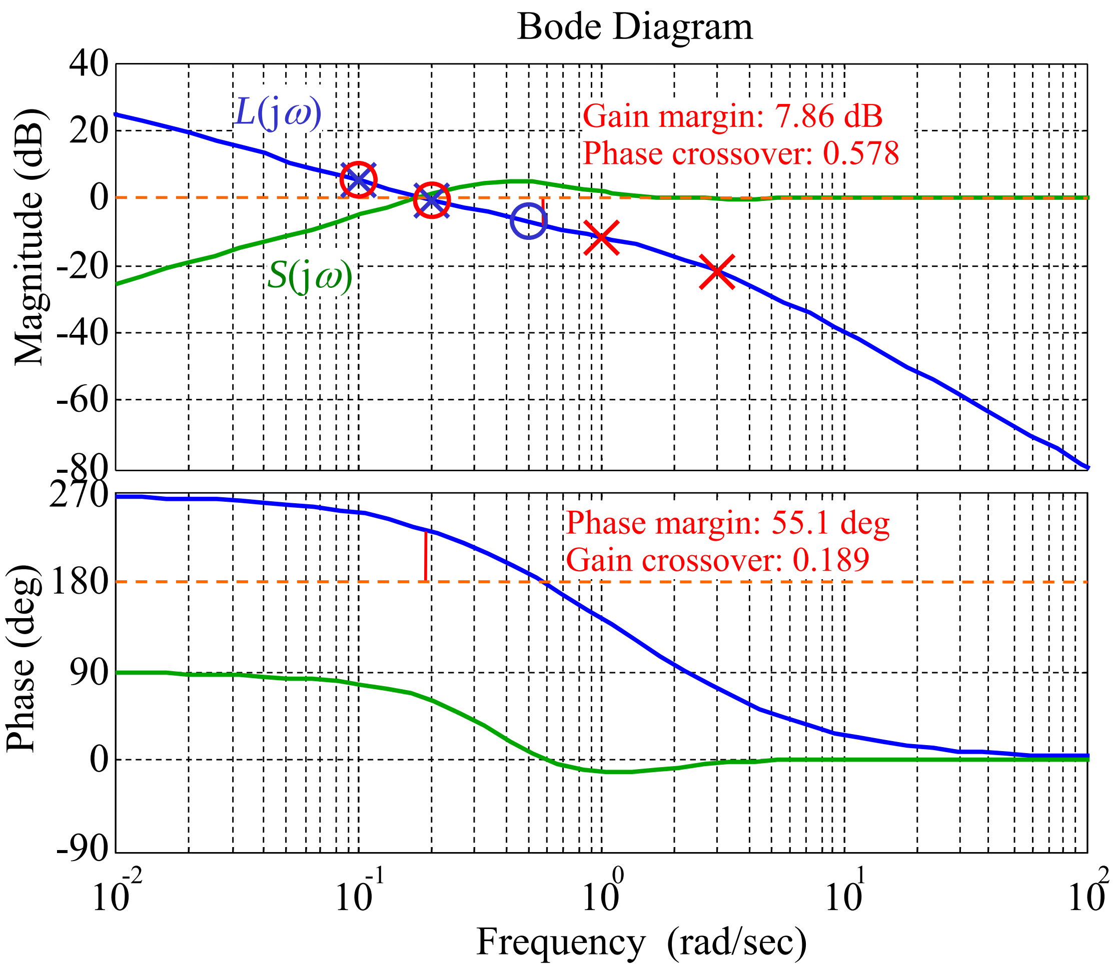
The corresponding controller is then

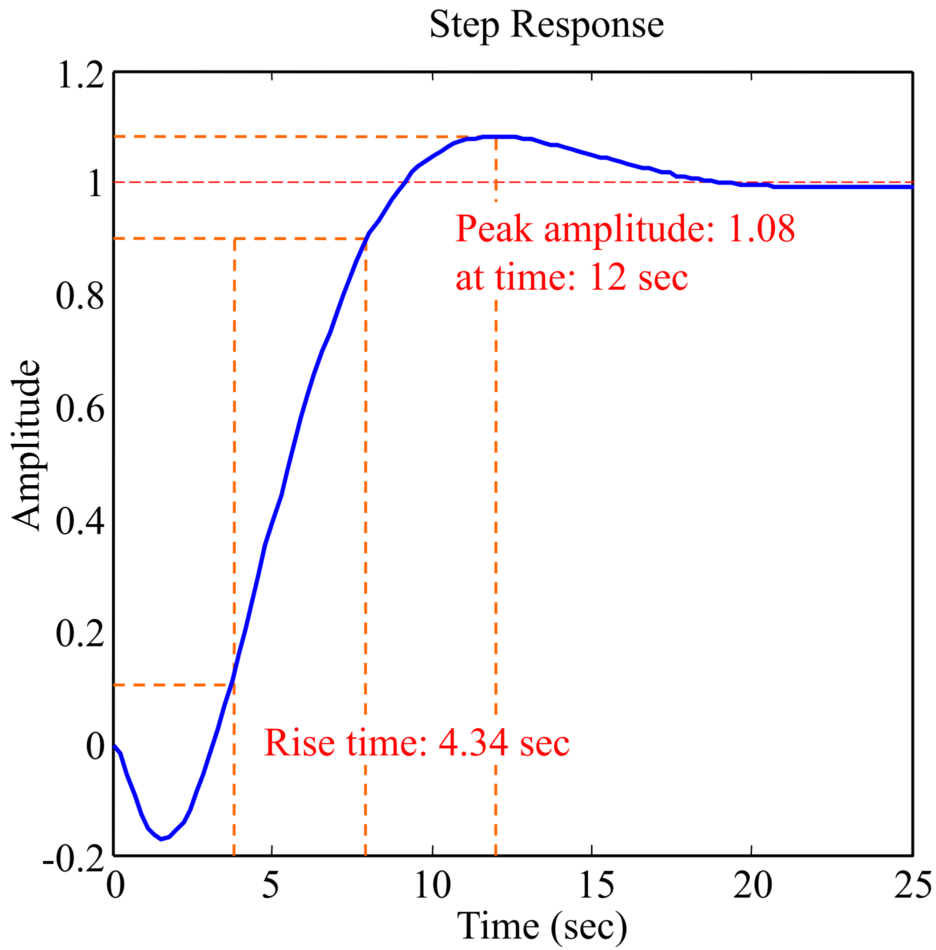
$$G_{\mathrm{R}}(s) = 0.06\frac{(10s+1)(5s+1)}{s(s+1)(s/3+1)}$$

Both the stability margins and dynamic parameters are improved. However, the influence of the RHP zero is now stronger.

The maximum value of the sensitivity function in this case is $M_S \approx 1.84 \approx 5.28\ \mathrm{dB}$.

Example 1.2: Loop shaping (8)





State-space theory and system robustness

The state-space theory represented a paradigm shift with many useful system concepts. For linear time-invariant systems the standard state-space model is given by

$$\begin{aligned} \dot{\boldsymbol{x}} &= \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{u} + \boldsymbol{w} \\ \boldsymbol{y} &= \boldsymbol{C}\boldsymbol{x} + \boldsymbol{v} \end{aligned} \qquad (1.21)$$

where $\boldsymbol{u}$ is the input, $\boldsymbol{y}$ the output and $\boldsymbol{x}$ is the state variable. The uncertainty is represented by the disturbances $\boldsymbol{w}$ and $\boldsymbol{v}$ as well as by variations in the elements of the matrices $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$. the disturbances $\boldsymbol{w}$ and $\boldsymbol{v}$ are typically described as stochastic processes.

The control problem is usually formulated either as quadratic optimal control minimizing a given cost function or as state-space tracking based on pole placement technique. In many cases a state observer is needed to estimate the state vector components which cannot be measured directly. Generally, controllability and observability are key conditions for solving the control problem.

In the model (1.21) it is natural to describe model uncertainties as variations in the elements of the matrices $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$. This is, however, only a very restricted class of perturbations and does not cover neglected dynamics or small time delays which are easier to be described in the frequency domain. This is one major drawback of the state-space control methods.

The only formal requirements on the system to be controlled in state-space theory is that it is controllable and observable. There are no consideration of RHP poles and zeros or time delays. Thus, one has to pay special attention to the system design to achieve good robustness.

Example 1.3: a fast system with a low bandwidth (1)

Exp. 1.3: Consider a system that is described by

$$\dot{\boldsymbol{x}} = \begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix} \boldsymbol{x} + \begin{pmatrix} -10 \\ 1 \end{pmatrix} u$$

$$y = \begin{pmatrix} 1 & 0 \end{pmatrix} \boldsymbol{x}$$

It is easy to proof that the system is both controllable and observable. The system is of second order and only one state variable is supposed to be measurable. Thus, a state controller with an observer of first order is designed, which leads to a closed-loop system of third order. Now, assume that the characteristic polynomial of the closed-loop system is

$$\Delta(s) = (s+10)(s+5+\mathrm{j}5\sqrt{3})(s+5-\mathrm{j}5\sqrt{3}) = (s+10)(s^2+10s+100)$$

Obviously, the closed-loop poles are fast and well-damped. Since the plant transfer function is

$$G_{\mathrm{S}}(s) = \frac{-10s+1}{s(s+1)}$$

and the equivalent transfer function of the controller and observer has the general form

$$G_{\mathrm{R}}(s) = \frac{\beta_1 s + \beta_0}{s+\alpha}$$

Example 1.3: a fast system with a low bandwidth (2)

the controller parameters can be calculated based on coefficient comparison

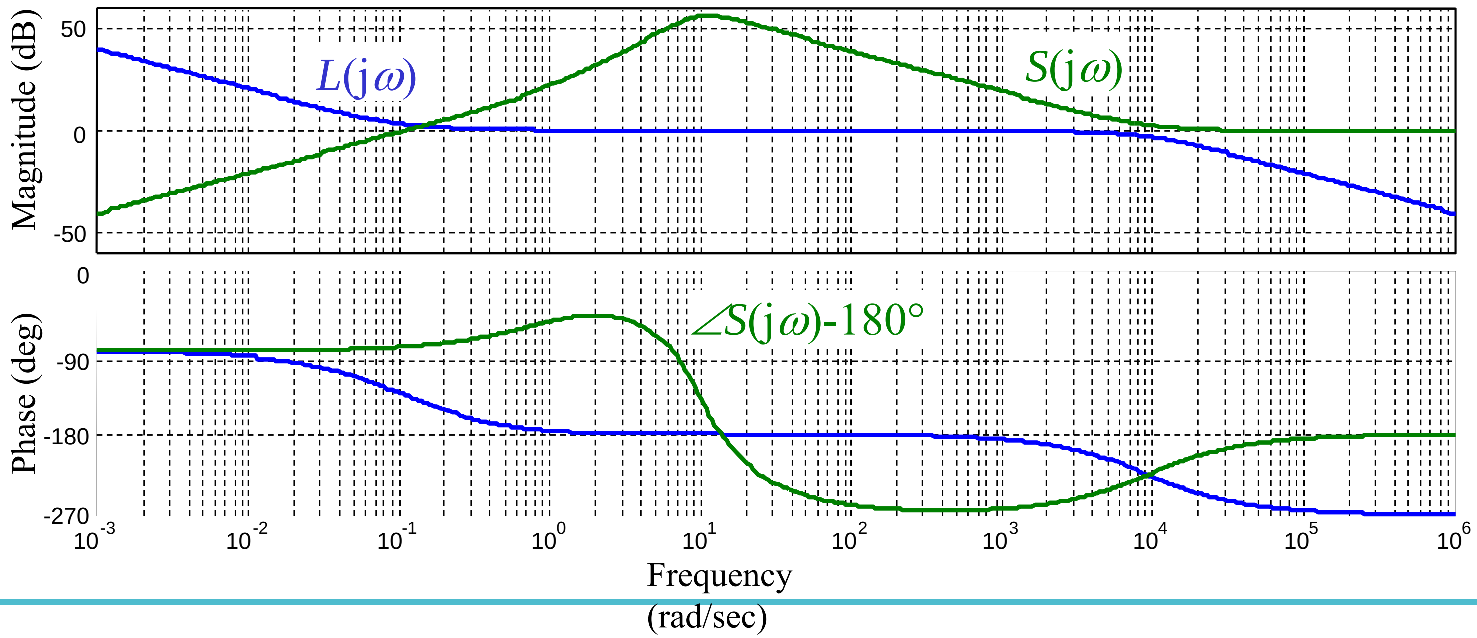
$$\beta_0 = 1000\,; \quad \beta_1 = 925.5\,; \quad \alpha = 9274.5$$

Thus, the loop function is

$$L(s) = \frac{(\beta_1 s + \beta_0)(1-10s)}{s(s+1)(s+\alpha)} = -9255\frac{(s-0.1)(s+1.0805)}{s(s+1)(s+9274)}$$

almost cancelled!

Now, we consider the Bode diagram of the loop transfer function shown in the figure.



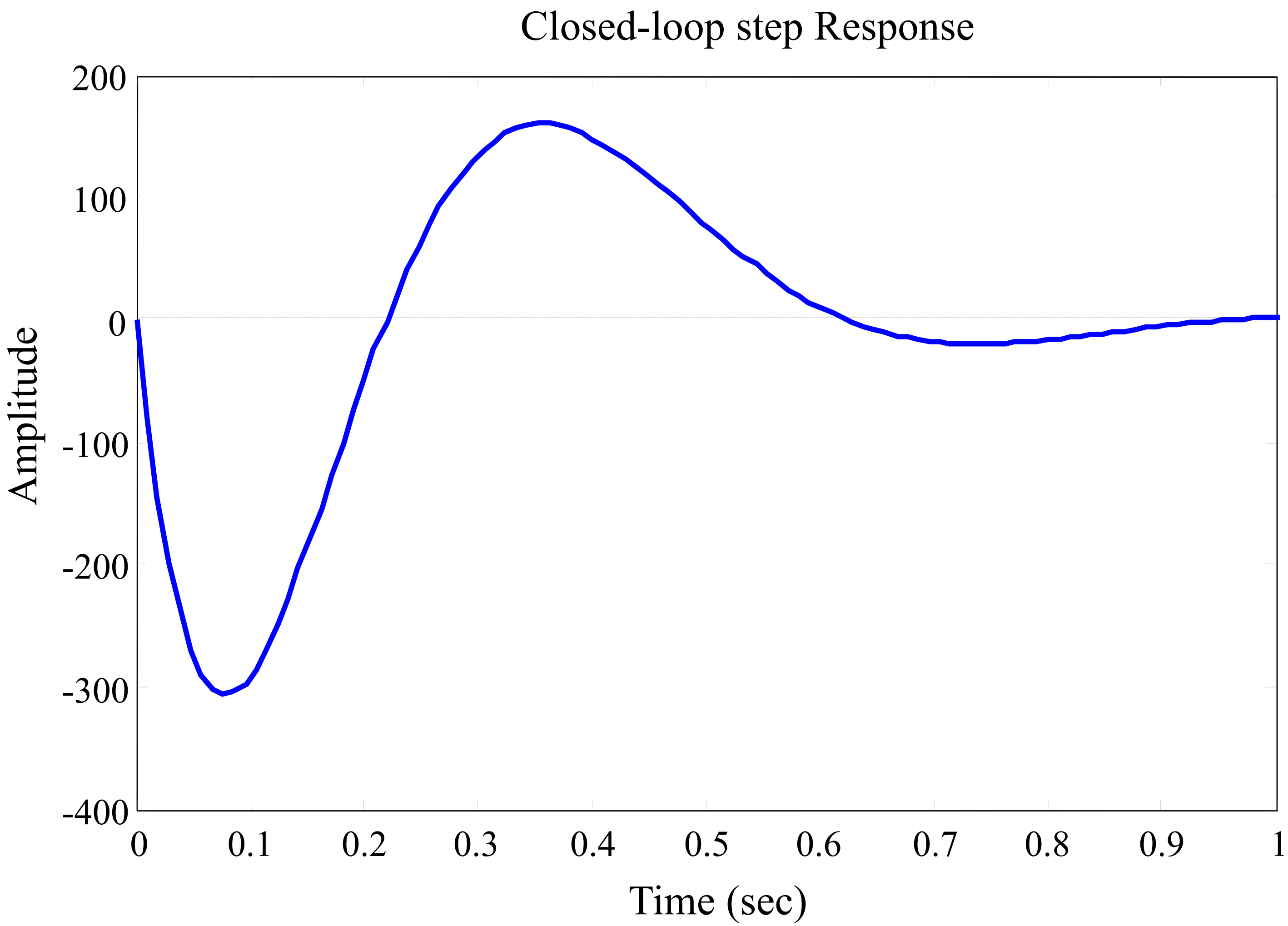
Example 1.3: a fast system with a low bandwidth (3)

The loop transfer function has a low frequency asymptote that intersects the 0 dB-line $\log|L(\mathrm{j}\omega)| = 0$ at ω = -0.1, i.e., at the slow non-minimum phase zero. The magnitude then becomes close to one and it remains so until the break point at around $\omega \approx 9275$, i.e. near the controller pole. The phase is very close to -$180°$ over that frequency range which means that the stability margin is very small. The gain crossover frequency is 6.58 and the phase margin $\varphi_{\mathrm{m}} = 0.15°$. The maximum sensitivities are $M_S = 678$ and $M_T = 677$ which also indicate that the system is extremely sensitive. The slope of the magnitude curve at crossover is also very small which is another indication of the poor robustness of the system.

That example shows clearly the danger of using a design method in a routine manner, and that it is not sufficient to check controllability, observability and closed-loop pole location. What is the reason for this very poor performance and robustness? For this particular example there are several limitations caused by the RHP zero. Violating these limitations by making the closed-loop system too fast we obtain a system with very poor stability margins even if the closed-loop poles are well-damped. Notice also that even if the gain crossover frequency is 6.58 rad/s the sensitivity S becomes larger than one for $\omega = 0.107$, which is close to the RHP zero. Feedback is thus not effective for disturbances having higher frequencies than about 0.107, since they are amplified by the feedback.

Respect the system limitations!

Example 1.3: a fast system with a low bandwidth (4)



Summary

- Model uncertainty is a key motivation for introducing feedback
- Process uncertainty can be easily described as a variation in the process transfer function
- Robustness can be measured using gain and phase margins and the sensitivity functions
- Transfer function, Nyquist's stability theory, Nyquist plot, Bode diagrams, Bode's integrals and Bode's ideal loop transfer function are important concepts and tools for robustness studies
- Fundamental limitations including RHP poles and zeros exist which one has to be aware of during control design
- Loop shaping is an effective method to design a proper controller taking robustness aspects into account
- The original state-space theory is an elegant approach to control design. The robustness, however, can not be handled properly. Special attention must be paid to the design procedure to avoid robustness difficulties. Evaluation of design robustness is needed to find a suitable compromise between different control requirements.

Chapter's references

[1.1] Åström, K.J. Limitations on control system performance
European Journal of Control, Vol. 6, 2000, pp. 1 - 19

[1.2] Black, H.S. Stabilized feedback amplifier
Bell System Technical Journal, Vol. 13, 1934, pp. 1 - 18

[1.3] Bode, H.W. Network analysis and feedback amplifier design
Van Nostrand, New York, 1945

[1.4] Freudenberg, J.S. /Looze, D.P. Right half plane pole and zeros and design trade-offs in feedback systems, *IEEE Transactions on Automatic Control*, Vol. AC-30, 1985, pp. 555 – 565

[1.5] Horowitz, I.M. Synthesis of feedback systems
Academic Press, New York, 1963

[1.6] Middleton, R.H. Trade-offs in linear control system design
Automatica, Vol. 27, 1991, pp. 281 – 292