EEEN60108
Control Fundamentals

William Heath

Minor revisions
September 2015

Contents

Part I
Preliminaries

1 Introduction

1.1 Feedback systems

Feedback mechanisms are ubiquitous in both nature and man-made systems. We are concerned with the use of negative feedback to control or regulate various system attributes to desired levels.

There are many examples in biology. In the eye the iris opens and closes to control the amount of light that falls on the retina. In bright light it closes while in dim light it opens. Similarly the form of the lens is modified using feedback to ensure correct focus of an object as its distance varies (so-called accommodation). The ability of biological systems to maintain attributes at a desired even level is known as "homeostasis." Almost all mammals are "warm-blooded"—that is to say they regulate their blood temperature.

There are also many examples at the level of cell biology. For example, most enzyme production is regulated by negative feedback. Too much of a product acts as an inhibitor on an earlier stage of production.

Negative feedback operates on a macro scale in ecological systems. For example feedback mechanisms maintain the proportion of oxygen in the atmosphere near 21%, and the proportion of carbon dioxide near 0.03%. This phenomenon gives rise to the "Gaia hypothesis" that the earth is alive. Mankind is currently testing both this hypothesis and the feedback mechanisms to their limit.

Some control mechanisms in nature are highly sophisticated. Consider a tennis player. He/she has highly developed actuators (the tennis racquet, limbs and muscles) and highly developed sensors (vision and feel). But to be a good player, he/she also needs considerable skill and finesse to co-ordinate the sensors and actuators into a winning combination. Nevertheless feedback is crucial. When catching a high ball, sportsmen/sportswomen do *not* calculate the trajectory of the ball. Rather they use feedback heuristics such as the following: *"fix your gaze on the ball, start running, and adjust your running speed so that the angle of the gaze remains constant."*[*]

1.2 History of feedback controllers

In this course we are concerned with the design of feedback controllers for engineering systems. Some common examples are extremely old. For example the ballcock in a lavatory prevents the cistern from overflowing. Note that as this involves a switching mechanism, it beyond the scope of this course.

Watt's governor (or the centrifugal governor) was one of the catalysts for the industrial revolution (Fig 1). When used as a governor for a steam engine, its function can be described heuristically as follows: as the flyballs spin faster, they move out and open a valve which releases steam; this in turn slows the speed at which the flyballs spin; similarly as the speed drops the flyballs move in, and close the valve. There is an interesting example of such a governor at Quarry Bank Mill in Styal, where the governor is used to regulate the amount

[*] G. Gigerenzer: Gut Feelings, Allen Lane, 2007.

of water falling onto a mill wheel. Until very recently the same principle was used for the control of Diesel engines.

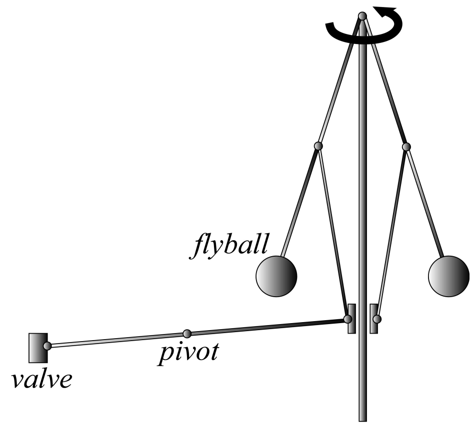


Figure 1: Watt's governor.

The invention of the feedback amplifier in 1927 ushered in long distance telecommunications. The feedback amplifier has stable linear characteristics despite uncertainties and nonlinearities in the high gain amplifier (originally a vacuum tube amplifier) around which it is built.

Traditionally control has been divided into the fields of:

- Process control, concerned with the product or operation of chemical engineering systems (for example, controlling the yield of a distillation column).
- Mechanical control, concerned with the operation of mechanical systems (for example the Wright brothers' main innovation was an aircraft that was open-loop unstable, but that could be stabilized and easily maneuvered using feedback).
- Electrical and electronic control, concerned with the regulation of electrical circuits, whether on the very small scale (e.g. nano technology) or very large scale (e.g. control of power networks).

Computer control breaks down the distinction between these fields. The report "The Impact of Control Technology: Overview, Success Stories, and Research Challenges" gives a fascinating account of both the state-of-the-art in control technology and future challenges that are being addressed. It is available on-line from:

`http://ieeecss.org/general/impact-control-technology`

1.3 Example: automotive control

A modern car contains many control loops. The most obvious example is the engine control, where the amount of fuel (as well as the fuel/air ratio in petrol engines) injected to each cylinder is computer controlled. The controller takes into account many measured variables, but the principle of feedback control is neatly illustrated by idle speed control (the state of idle is usually defined to occur when the foot is off the accelerator). Here the controller measures the engine speed (usually from the flywheel) and regulates the fuel injected to maintain a constant low speed. Diesel engines are unstable at idle in open-loop so some sort of feedback mechanism is necessary—without this they would either run away or stall. Historically Diesel engines had mechanical governors that regulated the injected fuel according to the speed. Nowadays electronic control units (ECUs) have replaced mechanical governors. The speed is measured (usually from the flywheel) and the ECU then determines the amount of fuel to inject. On a modern common rail system, where a small reservoir of fuel is kept at high pressure, this is achieved by determining the length of time a valve is opened allowing flow of fuel from the reservoir to the cylinder. The control loop is illustrated in Fig 2. For petrol engines a governor is not necessary but the use of ECUs has greatly enhanced performance.

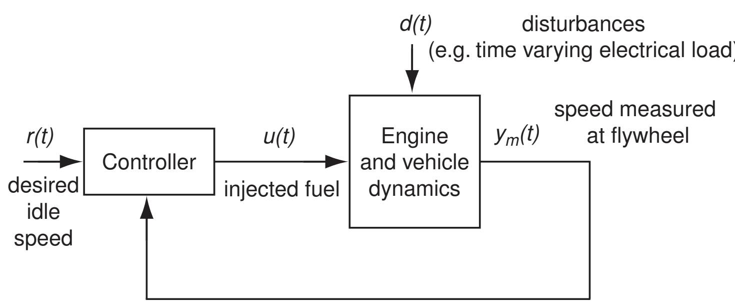


Figure 2: Control loop for a Diesel engine at idle.

Cruise control is another control loop. A constant engine speed is maintained (even in the presence of changes in road slope). Some high end cars now have active cruise control, where the speed and position of the car in front is detected (for example via radar) and a set distance between the car and the car in front is maintained. Some experimental cars also have automated steering where lanes are detected using video cameras, and the car's position in the centre of the lane is automatically maintained.

There are many other loops in modern cars. Some, such as traction control and anti-lock breaking, improve the dynamic performance of the vehicle. Others, such as air conditioning, simply enhance the passenger comfort. On a larger scale, navigation and traffic flow systems may involve feedback. The vehicle manufacturing process also has many control loops (for robotics, plant scheduling etc.).

1.4 Scope of the course

We will discuss two fundamental concepts in this course:

Feedback, the use of measured signals from a plant to help determine the actuator values.

- We will consider how feedback can be used to overcome uncertainty.
- We will consider how feedback can be used to achieve certain performance criteria.
- We will consider simple control structures such as PID controllers.

Dynamics, the modeling of a plant's time response using differential equations.

- We will consider dynamics in the time domain.
- We will consider dynamics in the frequency domain.
- We will characterise simple first and second order systems.

Much of the material on dynamics has already been covered in Signals and Systems. We will focus our attention on transfer function descriptions.

We will combine the two concepts by considering the dynamics of *closed-loop* systems.

We will then consider three analysis/design techniques:

Root locus, broadly associated with time domain dynamics.

The Nyquist criterion, a frequency response analysis tool.

Phase lead and lag compensation, a frequency response design technique.

These can be used to deepen our understanding of PID controllers and second order systems, but also to analyse more complicated control structures.

1.5 Limitations of the course

This is an introductory course, where we will outline the rudiments of classical control. This was mostly developed in the 1930s and 1940s, and is still used today. However, many subsequent (and important) developments are beyond its scope. For example, we will not consider:

- digital control,
- nonlinear control,
- multivariable control,
- switched control,
- networked control,

- state space models,
- optimal control,
- modern robust control.

Similarly we will pay little attention to the important topics of system identification and modelling.

One of the main differences between classical control (that we study in this course) and more modern approaches is that classical control relies on graphical techniques that are relatively easy to draw by hand. By contrast modern approaches tend to rely on the availability of computer software to find design and synthesis solutions. The classical approach results in intuitive solutions, which is why it continues to be used (and to be useful) today. Nevertheless we will find a computer useful to generate (or validate) solutions and graphs.

1.6 Further reading

The following is an excellent new introductory course to control and feedback systems:

> K. J. Åström and R. M. Murray. Feedback Systems; an introduction for scientists and engineers. Princetown University Press, 2008.

At the time of writing an electronic pdf version is available from:

`http://www.cds.caltech.edu/~murray/amwiki/Main_Page`

The following three books cover the same material as the course, and include many worked examples and practical applications:

> G. F. Franklin, J. D. Powell and A. Emami-Naeni. Feedback control of dynamic systems (6th edition). Pearson, 2010.

> R. C. Dorf and R. H. Bishop. Modern Control Systems (12th edition). Pearson, 2011.

> N. S. Nise. Control Systems Engineering (6th edition). Wiley, 2011.

Also recommended:

> B. Lurie and P. Enright. Classical feedback control: with Matlab and Simulink (2nd edition). CRC Press, 2011.

Both the IEEE Control Systems Society webpage and the IFAC (International Federation of Automatic Control) webpage have a wealth of information on current applications and developments:

`http://www.ieeecss.org`
`http://www.ifac-control.org/`

In particular the IEEE page has contents of the IEEE Control Systems Society Magazine. This has many accessible articles on both the theory and application of control systems. The full articles can be downloaded from:

`http://ieeexplore.ieee.org`

There is also useful information at the IET:

`http://kn.theiet.org/communities/controlauto/index.cfm`

2 Feedback

2.1 Negative feedback amplifier

One of the main uses of feedback is to overcome uncertainty. This is nicely illustrated by the negative feedback amplifier. Op-amps rely of feedback for their successful application. The op-amp itself is a high gain device (typically over 100dB at steady state). When configured in feedback the emergent properties are largely unaffected by noise, nonlinearities etc. Furthermore, with an ideal op amp there is zero output impedance—i.e. the output voltage does not vary with output current. This is crucial for modularity, and is a direct result of using negative feedback. Op-amps are considered in more detail as a Case Study towards the end of these notes.

2.2 Standard control loop

A typical control system is depicted in Fig 3. We define the **plant** as the system to be controlled. Plants may have more than one actuator, more than one sensor and more than one controller, but we will assume only one of each. Thus we assume we have some measured variable $y_m(t)$, measuring some plant attribute $y_p(t)$, and we would like it to follow some set point (or reference signal) $r(t)$. To that end the controller computes the manipulated variable $u(t)$ that is fed to the plant actuator.

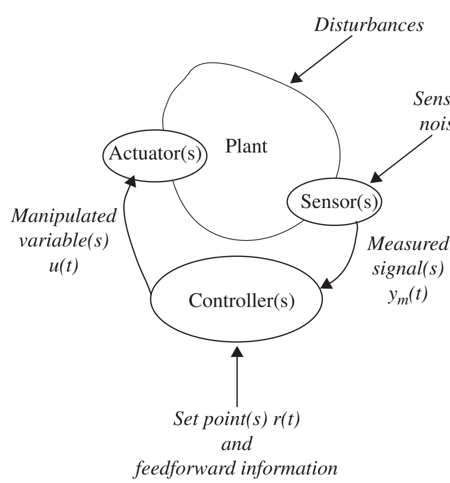


Figure 3: Control system.

We will assume the control loop can be simplified to the configuration shown in Fig 4. This is an over-simplification, but is suitable for much of the analysis in this course. In practice control structures usually include additional feedforward paths. The plant (with its actuator) is represented as the block G together with

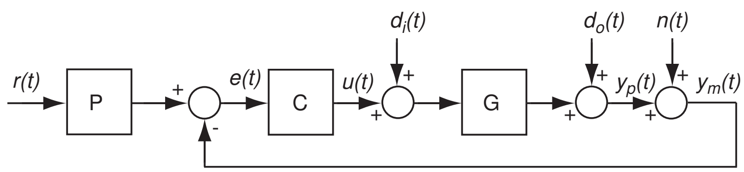


Figure 4: Closed loop system.

disturbance signals. Thus the plant output y_p is related to the plant input u as

$$y_p = G(u + d_i) + d_o.$$

The signals $d_i(t)$ and $d_o(t)$ represent input and output disturbances respectively. The sensor is represented simply by the addition of a noise signal $n(t)$, so that the measurement of the plant output is given by

$$y_m = y_p + n.$$

The aim of the controller is to determine the manipulated variable $u(t)$ so that the plant output $y_p(t)$ tracks the reference signal $r(t)$ as closely as possible. The controller takes a very specific structure

$$\begin{aligned} u &= C(e), \\ e &= P(r) - y_m. \end{aligned}$$

2.3 Standard control loop without dynamics

For the present discussion we will make the *further* simplification that G, C and P are all simple gains. The system equations can be written:

$$\begin{aligned} y_p(t) &= Gu(t) + Gd_i(t) + d_o(t), \\ y_m(t) &= y_p(t) + n(t), \\ e(t) &= Pr(t) - y_m(t), \\ u(t) &= Ce(t). \end{aligned}$$

Substituting for $e(t)$ we find

$$\begin{aligned} y_p(t) &= d_o(t) + Gu(t) + Gd_i(t), \\ u(t) &= CPr(t) - Cy_m(t), \\ y_m(t) &= y_p(t) + n(t). \end{aligned}$$

Substituting for $y_m(t)$ and $u(t)$ gives

$$y_p(t) = d_o(t) + Gd_i(t) + GCPr(t) - GCy_p(t) - GCn(t).$$

We may rearrange this to give

$$(1 + GC)y_p(t) = GCPr(t) + Gd_i(t) + d_o(t) - GCn(t),$$

and hence we obtain the closed-loop equation

$$y_p(t) = \frac{GC}{1+GC}Pr(t) + \frac{G}{1+GC}d_i(t) + \frac{1}{1+GC}d_o(t) - \frac{GC}{1+GC}n(t). \qquad (1)$$

Note that equation (1) gives an expression for $y_p(t)$ that is independent of $u(t)$.

If C is very large, we may say

$$\frac{GC}{1+GC} \approx 1, \;\; \frac{G}{1+GC} \approx 0 \text{ and } \frac{1}{1+GC} \approx 0.$$

Thus if $P = 1$ we find

$$y_p(t) \approx r(t) - n(t).$$

Thus the use of high gain feedback allows us virtually to eliminate the effects of input and output disturbances, although not the effects of sensor noise.

We will find the quantities $\frac{1}{1+GC}$ and $\frac{GC}{1+GC}$ recur often, so we give them special names and symbols:

- the sensitivity $S = \frac{1}{1+GC}$;
- the complementary sensitivity $T = \frac{GC}{1+GC}$.

Note that

$$S + T = 1.$$

The block represented by P is usually termed the pre-filter.

2.4 Stability and dynamics

In the previous section we suggested that the use of high gain control could render near perfect control, with the only limit on performance set by the size of the sensor noise signal $n(t)$. In practice such a high gain control would likely require undesirable and/or infeasible actuator energy. It would also be likely to send the closed-loop system unstable, or at least result in some highly undesirable dynamics. An obvious example of an unstable feedback loop is when callers to talkback radio leave their own radios on to listen to themselves.

Feedback can also be used to stabilise a plant that is unstable in open-loop. We have already seen this in the example of Diesel engine control.

To analyse such effects we will need a dynamic model of the relation between $y_p(t)$ and $u(t)$. This will be the subject of the next chapter. We will also find it useful to allow our controller C and pre-filter P to have dynamic characteristics. Our aim is to return to a figure such as Fig 4 but where G, C and P capture dynamic behaviour. Similarly we would like to model $y_p(t)$ in terms of $r(t)$, $d_i(t)$, $d_o(t)$ and $n(t)$ as in (1) but again capturing dynamic behaviour.

2.5 Key points

- Feedback is everywhere.
- Feedback has useful properties: in particular in the presence of uncertainty.
- Feedback can be bad - we need to consider dynamics carefully.
- There are inescapable design tradeoffs when designing a control system.

Part II
Review of Signals and Systems

3 Differential equations

When designing and analysing feedback systems we find it useful to think in terms of input/output response. We assume that given a model of the plant G and its input $u = u(t)$ we can determine its output $y = y(t)$ (see Fig 5; we are ignoring the effects of disturbances and noise for the moment). We use the notation $y = Gu$ (which should be interpreted as $y = G(u)$). Representing G as a static gain is usually too simplistic. It turns out that representing G as a linear differential equation is useful for a wide variety of plants (and also control structures).

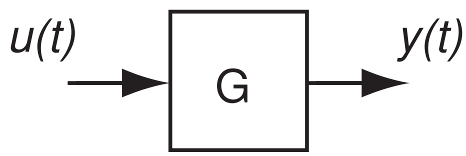


Figure 5: Open loop input-output response.

3.1 Linear differential equations

An n-th order linear differential equation takes the form

$$\begin{aligned} & a_n \frac{d^n y}{dt^n}(t) + a_{n-1}\frac{d^{n-1}y}{dt^{n-1}}(t) + \cdots + a_1 \frac{dy}{dt}(t) + a_0 y(t) \\ = \quad & b_m \frac{d^m u}{dt^m}(t) + b_{m-1}\frac{d^{m-1}u}{dt^{m-1}}(t) + \cdots + b_1 \frac{du}{dt}(t) + b_0 u(t). \end{aligned}$$

We must have $a_n \neq 0$; otherwise the order of the equation is less than n. There is also some redundancy in the choice of the parameters $a_n, a_{n-1}, \ldots, a_1, a_0$ and $b_m, b_{m-1}, \ldots, b_1, b_0$. The usual convention is to set $a_n = 1$ without loss of generality. So our n-th order linear differential equation takes the form

$$\begin{aligned} & \frac{d^n y}{dt^n}(t) + a_{n-1}\frac{d^{n-1}y}{dt^{n-1}}(t) + \cdots + a_1 \frac{dy}{dt}(t) + a_0 y(t) \\ = \quad & b_m \frac{d^m u}{dt^m}(t) + b_{m-1}\frac{d^{m-1}u}{dt^{m-1}}(t) + \cdots + b_1 \frac{du}{dt}(t) + b_0 u(t). \qquad (2) \end{aligned}$$

We say $y = G(u)$ if, given an input signal u, the output signal y satisfies (2).

Linearity says that if, given u_1 and u_2, we have

$$y_1 = G(u_1),$$

and

$$y_2 = G(u_2),$$

for some y_1 and y_2 then

$$\alpha_1 y_1 + \alpha_2 y_2 = G(\alpha_1 u_1 + \alpha_2 u_2),$$

for any α_1 and α_2.

3.2 Example: electrical circuit

Consider the simple RLC circuit depicted in Fig 6.

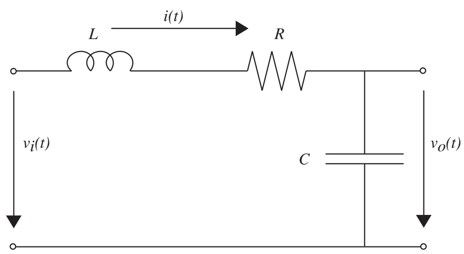


Figure 6: Simple RLC circuit.

We have

$$L\frac{di}{dt}(t) + Ri(t) + v_o(t) = v_i(t),$$
$$C\frac{dv_o}{dt}(t) = i(t).$$

Eliminating $i(t)$ yields

$$LC\frac{d^2v_o}{dt^2}(t) + RC\frac{dv_o}{dt} + v_o(t) = v_i(t),$$

or

$$\frac{d^2v_o}{dt^2}(t) + \frac{R}{L}\frac{dv_o}{dt} + \frac{1}{LC}v_o(t) = \frac{1}{LC}v_i(t).$$

It seems natural to associate $v_i(t)$ with an input $u(t)$ and $v_o(t)$ with an output $y(t)$. However this requires an additional assumption of causality—that is to say that $v_o(t)$ is determined by $v_i(t)$ while $v_i(t)$ is determined independently (or more simply, the assumption that $v_i(t)$ is the driving voltage). This requires external conditions which cannot be inferred from the equations alone. If such an assumption is justified, then the equation takes the standard form with $n = 2$, $m = 0$:

$$\frac{d^2y}{dt^2} + a_1\frac{dy}{dt} + a_0y = b_0u,$$

with $y = v_o$, $u = v_i$, $a_1 = \frac{R}{L}$, $a_0 = \frac{1}{LC}$ and $b_0 = \frac{1}{LC}$.

3.3 Example: mechanical system

Consider the simple one-dimensional mechanical system depicted in Fig 7. A mass M is connected to a spring with spring constant K and a damper with damping coefficient B. A force $f(t)$ acts on the mass and its distance $y(t)$

is measured (calibrated such that 0 corresponds to the equilibrium position). Newton's second law gives

$$f(t) - Ky(t) - B\frac{dy}{dt}(t) = M\frac{d^2y}{dt^2}(t),$$

which we can rearrange to give

$$\frac{d^2y}{dt^2}(t) + \frac{B}{M}\frac{dy}{dt}(t) + \frac{K}{M}y(t) = \frac{1}{M}f(t).$$

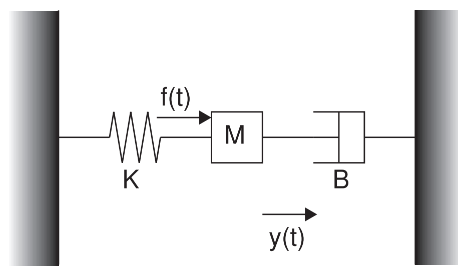


Figure 7: Simple mechanical system.

Thus if the force is determined independently from the system we may associate it with an input $u(t)$ and our equation takes the standard form with $n = 2$, $m = 0$:

$$\frac{d^2y}{dt^2} + a_1\frac{dy}{dt} + a_0 = b_0u,$$

with $u = f$, $a_1 = \frac{B}{M}$, $a_0 = \frac{K}{M}$ and $b_0 = \frac{1}{M}$.

3.4 Example: flow system

Consider the fluid tank depicted in Fig 8. The flow in is $q_i(t)$ and the flow out is $q_o(t)$ while the measured fluid height is $h(t)$. If the tank has constant area A then

$$A\frac{dh}{dt}(t) = q_i(t) - q_o(t).$$

We will assume that the flow in $q_i(t)$ is determined independently from conditions in the tank, whereas the flow out $q_o(t)$ is determined by the height of the fluid in the tank. Suppose we fix the input flow to $q_i(t) = \bar{q}$ for some constant $\bar{q}$. After a while the height of the tank will settle to some $\bar{h}$ and the out-flow will equal the input flow $q_o(t) = q_i(t) = \bar{q}$. Then for small variation in input flow we can say to good approximation

$$q_o(t) - \bar{q} = \lambda\left[h(t) - \bar{h}\right] \text{ for some constant } \lambda.$$

(The value of λ will depend on the cross sectional area of the outflow pipe and the settings of downstream valves as well as the properties of the fluid itself.)

Thus

$$A\frac{dh}{dt}(t) = q_i(t) - \bar{q} - \lambda\left[h(t) - \bar{h}\right]$$

If we set our measured variable $y(t)$ to be the deviation of height from this steady state value $y(t) = h(t) - \bar{h}$ and similarly set the manipulated variable $u(t)$ to be the deviation of input flow from the fixed value $u(t) = q_i(t) - \bar{q}$ then we obtain

$$A\frac{dy}{dt}(t) + \lambda y(t) = u(t),$$

or equivalently

$$\frac{dy}{dt} + a_0 y = b_0 u,$$

with $a_0 = \frac{\lambda}{A}$ and $b_0 = \frac{1}{A}$.

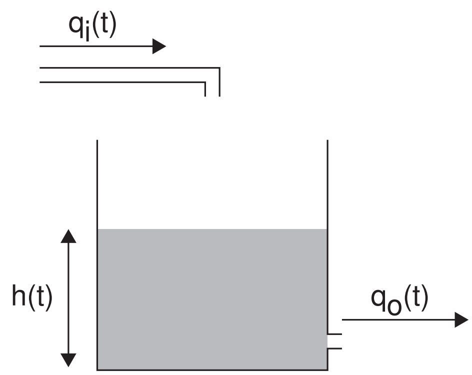


Figure 8: Flow system.

3.5 Discussion

We have seen that three different systems can be described by similar dynamic equations. Both the electrical circuit and the mechanical system are well modelled using second order linear differential equations. The flow system is approximately modelled using a simpler first order differential equation. Worked Example 5 concerns a pair of tanks which can be approximately modelled using a second order differential equation. More generally a description in terms of linear differential equations is often appropriate from the point of view of control system design

Historically the systems are prototypes for electrical, mechanical and chemical engineering. As a discipline control engineering has spanned all three branches of engineering. Often a system has elements from each (for example both electric motors and dvd players have both electrical and mechanical components). Most controllers are nowadays computer based; although the practicalities of such implementation are beyond the scope of this course, it is entirely appropriate that all control engineering should now be seen as a sub-discipline of electrical engineering.

3.6 Key point

- Linear differential equations can be used to describe the input-output properties of many simple systems.

4 Laplace transforms and transfer functions

In Section 2 we saw that it would be desirable to model plants using an input-output description

$$y = Gu. \tag{3}$$

In Section 3 we found a generic description of the form

$$\frac{d^n y}{dt^n}(t) + \cdots + a_0 y(t) = b_m \frac{d^m u}{dt^u}(t) + \cdots + b_0 u(t). \tag{4}$$

We used operator notation to write this as $y = G(u)$, but this is not in the form required (3). However we can express (4) in the form of (3) by use of the Laplace transform.

If $U(s)$ and $Y(s)$ are the respective Laplace transforms of $u(t)$ and $y(t)$, and if u and y are related by the differential equation (4) then

$$Y(s) = G(s)U(s),$$

with

$$G(s) = \frac{b_m s^m + \cdots + b_0}{s^n + \cdots + a_0}, \tag{5}$$

provided certain initial conditions are satisfied.

In this section we will consider the Laplace transform of signals—i.e. the objects $Y(s)$ and $U(s)$. We will also consider transfer functions such as the object $G(s)$.

4.1 Laplace transform definition

The Laplace transform of a signal $x(t)$ is defined as

$$X(s) = \mathcal{L}\left[x(t)\right] = \int_0^\infty x(t)e^{-st}dt.$$

Note that if we have two signals $x_1(t)$ and $x_2(t)$ that are equal for all $t \geq 0$ then their Laplace transforms are equal. To avoid ambiguity we will usually consider signals that are zero for $t < 0$.

(For mathematicians only: there is a distinction between $\int_{0^-}^\infty$ and $\int_{0^+}^\infty$. Such subtleties are beyond the scope of this course, but we are in fact using $\int_{0^-}^\infty$.)

4.2 Laplace transform examples

4.2.1 Unit step

We define the unit step $h(t)$ to be

$$h(t) = \begin{cases} 1 \text{ for } t \geq 0, \\ 0 \text{ for } t < 0. \end{cases}$$

Its Laplace transform $H(s)$ can be computed as

$$\begin{aligned} H(s) &= \int_0^\infty 1e^{-st}dt \\ &= \left[-\frac{1}{s}e^{-st}\right]_0^\infty \\ &= \frac{1}{s}. \end{aligned}$$

4.2.2 Exponential function

Suppose $x(t)$ is an exponential function

$$x(t) = e^{-at},$$

for some a. Its Laplace transform is

$$\begin{aligned} X(s) &= \int_0^\infty e^{-at}e^{-st}dt \\ &= \int_0^\infty e^{-(s+a)t}dt \\ &= \left[-\frac{1}{s+a}e^{-(s+a)t}\right]_0^\infty \\ &= \frac{1}{s+a}. \end{aligned}$$

4.2.3 Sinusoid

Suppose

$$x(t) = \sin\omega t.$$

We may use sine formulae to express $x(t)$ as

$$x(t) = \frac{1}{2j}\left(e^{j\omega t} - e^{-j\omega t}\right).$$

It follows from the Laplace transform of an exponential that

$$\begin{aligned} X(s) &= \frac{1}{2j}\left(\frac{1}{s-j\omega} - \frac{1}{s+j\omega}\right) \\ &= \frac{1}{2j} \times \frac{2j\omega}{s^2+\omega^2} \\ &= \frac{\omega}{s^2+\omega^2}. \end{aligned}$$

4.2.4 Dirac delta function

The Dirac delta function $\delta(t)$ is an impulse of infinite magnitude, infinitesimal duration and unit area. It has the fundamental property

$$\int_{-\infty}^{\infty} f(t)\delta(t)dt = f(0) \text{ for any } f(t).$$

Thus its Laplace transform is

$$\begin{aligned}\mathcal{L}[\delta(t)] &= \int_0^{\infty} \delta(t)e^{-st}dt \\ &= e^{-s\times 0} \\ &= 1.\end{aligned}$$

4.3 Laplace transform properties

The Laplace transform has many useful properties. In the following we will assume $x(t)$ has Laplace transform $X(s)$.

4.3.1 Linearity

The Laplace transform is linear. That is to say

$$\mathcal{L}\left[\alpha_1 x_1(t) + \alpha_2 x_2(t)\right] = \alpha_1\mathcal{L}\left[x_1(t)\right] + \alpha_2\mathcal{L}\left[x_2(t)\right].$$

4.3.2 Differentiation

$$\begin{aligned}\mathcal{L}\left[\frac{dx}{dt}(t)\right] &= \int_0^{\infty} \frac{dx}{dt}(t)e^{-st}dt \\ &= \left[x(t)e^{-st}\right]_0^{\infty} - \int_0^{\infty}(-s)x(t)e^{-st}dt \text{ by parts} \\ &= -x(0) + s\int_0^{\infty} x(t)e^{-st}dt \\ &= s\mathcal{L}[x(t)] - x(0) \\ &= sX(s) - x(0).\end{aligned}$$

Example: Suppose $u(t)$ and $y(t)$ are related by the first order differential equations

$$\frac{dy}{dt}(t) + a_0 y(t) = b_0 u(t).$$

Then taking Laplace transforms yields

$$sY(s) - y(0) + a_0 Y(s) = b_0 U(s).$$

Thus, provided $y(0) = 0$ we have

$$Y(s) = \frac{b_0}{s + a_0} U(s).$$

This verifies (5) for first order differential equations.

Similar relations follow for higher derivatives. For example, if follows immediately that the Laplace transform of the second derivative of $x(t)$ satisfies

$$\begin{aligned} \mathcal{L}\left[\frac{d^2x}{dt^2}(t)\right] &= s\mathcal{L}\left[\frac{dx}{dt}(t)\right] - \frac{dx}{dt}(0) \\ &= s^2 X(s) - sx(0) - \frac{dx}{dt}(0). \end{aligned}$$

4.3.3 Integration

Suppose

$$y(t) = \int_0^t u(\tau) d\tau.$$

Then

$$\frac{dy}{dt}(t) = u(t) \text{ and } y(0) = 0.$$

It follows that

$$Y(s) = \frac{1}{s} U(s).$$

4.3.4 Delay

Suppose

$$y(t) = u(t - \tau).$$

Then

$$\begin{aligned} Y(s) &= \mathcal{L}[u(t-\tau)] \\ &= \int_0^\infty u(t-\tau) e^{-st} dt \\ &= \int_{-\tau}^\infty u(t') e^{-s(t'+\tau)} dt' \\ &= \int_0^\infty u(t') e^{-s(t'+\tau)} dt' \\ &= e^{-s\tau} \int_0^\infty u(t') e^{-st'} dt' \\ &= e^{-s\tau} U(s). \end{aligned}$$

4.3.5 Final value theorem

Suppose $x(t)$ converges for large t to some steady state value x_{ss}. That is to say

$$\lim_{t\to\infty} x(t) = x_{ss} \text{ exists.}$$

We have the relation

$$sX(s) = x(0) + \int_0^\infty \frac{dx}{dt}(t)e^{-st}dt.$$

If we let $s \to 0$ we find

$$\begin{aligned}
\lim_{s\to 0} sX(s) &= x(0) + \lim_{s\to 0}\int_0^\infty \frac{dx}{dt}(t)e^{-st}dt \\
&= x(0) + \int_0^\infty \frac{dx}{dt}(t)dt \\
&= x(0) + [x(t)]_0^\infty \\
&= x_{ss}.
\end{aligned}$$

This is the final value theorem for the Laplace transform of signals.

4.4 Transfer functions

We have succeeded in our aim of an input-output description

$$Y(s) = G(s)U(s),$$

that captures linear dynamics. The object $G(s)$ is called the (linear) transfer function. It is a versatile tool for studying closed-loop systems (systems with controllers). We also need to consider what plant properties we can discern from (or model with) $G(s)$.

4.4.1 Definitions

We write

$$G(s) = \frac{b_m s^m + b_{m-1}s^{m-1} + \cdots b_0}{s^n + a_{n-1}s^{n-1} + \cdots + a_0}.$$

The order of G is n. The relative degree of G is $n - m$. A transfer function is said to be *proper* if $n \geq m$ and *improper* if $n < m$. We do not allow improper transfer functions. Proper transfer functions can be further classified as *strictly proper* when $n > m$ and *bi-proper* when $n = m$.

We may write

$$G(s) = \frac{n(s)}{d(s)},$$

where the polynomials $n(s)$ and $d(s)$ are respectively the numerator and denominator of the transfer function. The poles of G are the roots of $d(s)$ while the zeros of G are the roots of $n(s)$.

We say a transfer function is *stable* if all its poles are in the left half plane. It is *unstable* if it has any pole in the right half plane. If it has a pole on the imaginary axis (i.e. its real part is zero) but with no poles in the right half-plane then we say it is *marginally stable.*

Example: Suppose

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

Then G has three poles at $s = 0$, $s = -\frac{1}{2} + j\frac{\sqrt{3}}{2}$ and $s = -\frac{1}{2} - j\frac{\sqrt{3}}{2}$. It has one zero at $s = -1$. The order of G is 3 while the relative degree of G is 2. $G(s)$ is strictly proper (because its relative degree is greater than zero) and marginally stable (because it has one pole on the imaginary axis and the remainder in the left half plane).

4.4.2 Simulation

The notion of a transfer function being *proper* or *improper* is closely related to whether it is possible to *simulate* the system it represents. In particular, an input-output relation determined by a differential equation is usually simulated using integrators.

Example: Suppose

$$\frac{d}{dt}y(t) + ay(t) = bu(t).$$

Equivalently we may write

$$y(t) + a\int_0^t y\,dt = b\int_0^t u\,dt.$$

This is represented in the Laplace domain as

$$Y(s) + a\frac{1}{s}Y(s) = b\frac{1}{s}U(s).$$

This can be simulated as in Fig 9.

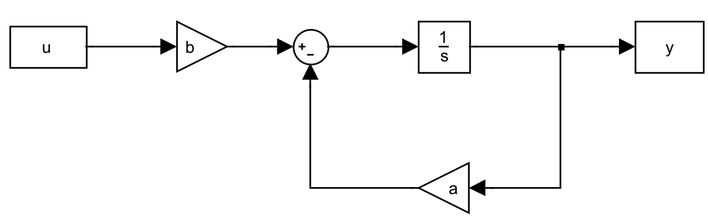


Figure 9: Simulation of a first order differential equation using Simulink.

A transfer function is only proper if the causal relation between the input and output can be realised using a network of integrators. The ideas are more formally expressed using *state space* methods, which are beyond the scope of this course.

4.4.3 Block diagrams and closed-loop response

Transfer functions can be combined in series and parallel. See Figs 10 and 11.

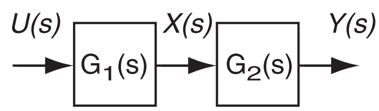


Figure 10: Transfer functions in series. We have $Y = G_2X$ and $X = G_1U$. Hence $Y = G_2G_1U$.

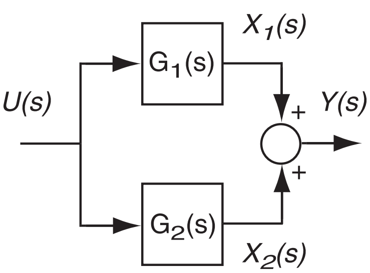


Figure 11: Transfer functions in parallel. We have $Y = X_1+X_2$ with $X_1 = G_1U$ and $X_2 = G_2U$. Hence $Y = (G_1 + G_2)\,U$.

But they can also be combined in feedback. See Fig 12. Thus our key innovation is that we can represent control systems, such as that depicted in Fig 3, using block diagrams, such as that depicted in Fig 4, with transfer functions that include a description of dynamic behaviour.

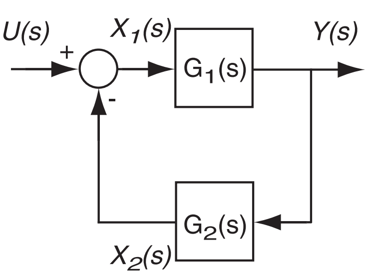


Figure 12: Transfer functions in feedback. We have $Y = G_1X_1$, $X_1 = U - X_2$ and $X_2 = G_2Y$. Substituting for X_1 and X_2 yields $Y = G_1U - G_1G_2Y$ and hence $Y = (1 + G_1G_2)^{-1}\,G_1U$.

Example: Suppose a plant can be represented by the transfer function $G(s)$ given by

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

The controller is a simple gain given by

$$C(s) = k_p.$$

We wish the plant output $y(t)$ to track some reference signal $r(t)$. For simplicity we assume there is no pre-filter $P(s)$, and no noise or disturbance signals. So the feedback loop can be depicted as in Fig 13.

We have the following relationships

$$\begin{aligned} Y(s) &= G(s)U(s), \\ U(s) &= C(s)E(s), \\ E(s) &= R(s) - Y(s). \end{aligned}$$

Solving gives

$$\begin{aligned} Y(s) &= \frac{GC}{1+GC}R(s) \\ &= \frac{k_p \frac{1}{s(s+1)}}{1 + k_p \frac{1}{s(s+1)}}R(s) \\ &= \frac{k_p}{s^2 + s + k_p}R(s). \end{aligned}$$

If $r(t)$ is a unit step

$$r(t) = \begin{cases} 0 \text{ for } t < 0, \\ 1 \text{ for } t \geq 1. \end{cases}$$

then $R(s) = 1/s$ and

$$Y(s) = \frac{k_p}{s^2 + s + k_p} \times \frac{1}{s}.$$

The evolution of the corresponding $y(t)$ is shown in Fig 14 for various values of k_p.

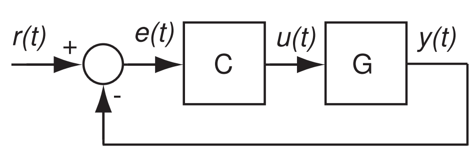


Figure 13: Simple closed loop system.

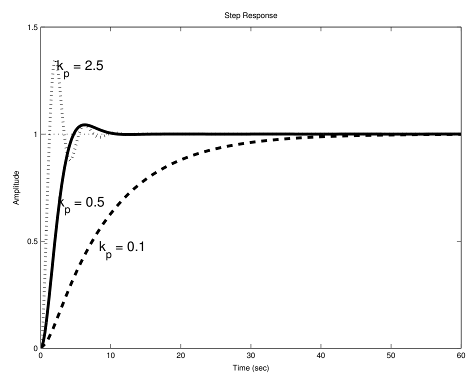


Figure 14: Closed-loop step responses for $k_p = 0.1$, 0.5 and 2.5. As we increase k_p the speed of response gets faster, but the response also becomes more oscillatory. For this example all three responses have the same steady state response $y_{ss} = 1$.

4.4.4 Sensitivity and complementary sensitivity

A more general model of a closed-loop system is given in Fig 4, repeated below as Fig 15, but with G, C and P (the pre-filter) generalised to transfer functions $G = G(s)$, $C = C(s)$ and $P = P(s)$. We may generalise our analysis provided we work in the Laplace domain.

Thus

$$\begin{aligned} Y_p(s) &= D_o(s) + G(s)U(s) + G(s)D_i(s), \\ U(s) &= C(s)P(s)R(s) - C(s)Y_m(s), \\ Y_m(s) &= Y_p(s) + N(s). \end{aligned}$$

Here $Y_p(s)$ is the Laplace transform of $y_p(t)$, the plant output. Similarly $Y_m(s)$ corresponds to the measured output, $U(s)$ to the input, $D_o(s)$ to the output disturbance, $D_i(s)$ to the input disturbance and $N(s)$ to the noise. See Fig 15.

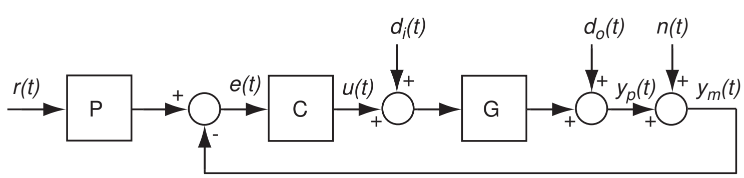


Figure 15: Closed loop system.

Substituting as before yields

$$\begin{aligned} Y_p(s) &= \frac{G(s)C(s)}{1+G(s)C(s)}P(s)R(s) - \frac{G(s)C(s)}{1+G(s)C(s)}N(s) \\ &\quad + \frac{G(s)}{1+G(s)C(s)}D_i(s) + \frac{1}{1+G(s)C(s)}D_o(s). \end{aligned}$$

Once again we define:

Sensitivity

$$S(s) = \frac{1}{1+G(s)C(s)}.$$

Complementary sensitivity

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

so that

$$Y_p(s) = T(s)P(s)R(s) - T(s)N(s) + G(s)S(s)D_i(s) + S(s)D_o(s).$$

As before we have the relations

$$T(s) = G(s)C(s)S(s),$$

and

$$S(s) + T(s) = 1.$$

Suppose

$$G(s) = \frac{n(s)}{d(s)} \text{ and } C(s) = \frac{n_c(s)}{d_c(s)},$$

with $n(s)$, $d(s)$, $n_c(s)$ and $d_c(s)$ all polynomials. Then

$$\begin{aligned} S(s) &= 1/\left(1 + \frac{n(s)}{d(s)}\frac{n_c(s)}{d_c(s)}\right) \\ &= \frac{d(s)d_c(s)}{d(s)d_c(s) + n(s)n_c(s)}, \end{aligned}$$

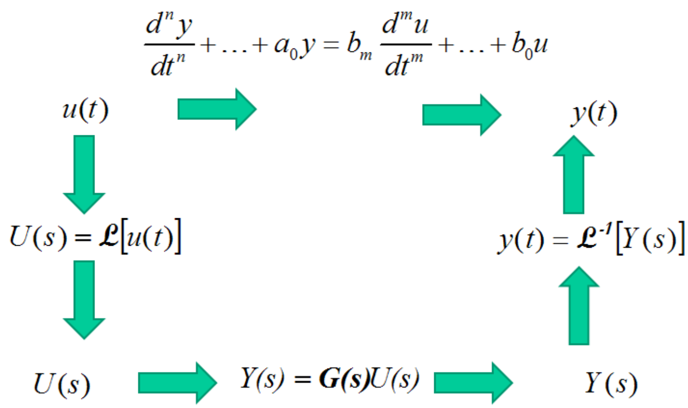
and

$$\begin{aligned} T(s) &= \left(\frac{n(s)}{d(s)}\frac{n_c(s)}{d_c(s)}\right) / \left(1 + \frac{n(s)}{d(s)}\frac{n_c(s)}{d_c(s)}\right) \\ &= \frac{n(s)n_c(s)}{d(s)d_c(s) + n(s)n_c(s)}. \end{aligned}$$

Note that $S(s)$ and $T(s)$ share the same denominator polynomial.

4.5 Big picture

- The relation between input-output signals in the time and Laplace domains are summarised as follows:



5 Step responses

The previous example illustrated that it is straightforward to compute *closed-loop* transfer functions. Given either a simulation package or the use of inverse Laplace transforms it is also straightforward to compute closed-loop time responses. However the relation between the response and the gain k_p is not necessarily straightforward. Before we analyse such relations, we will build up our intuition of open-loop responses. In this section we will study step-responses; in Section 6 we will study frequency response.

5.1 Steady state gain

Before we consider step responses, we can make a general observation about steady state behaviour. Suppose

$$Y(s) = G(s)U(s),$$

and $u(t)$ is such that

$$u_{ss} = \lim_{t\to\infty} u(t)$$

exists. The final value theorem states that

$$u_{ss} = \lim_{s\to 0} sU(s).$$

Thus, if it exists, we may say

$$\begin{aligned} y_{ss} &= \lim_{t\to\infty} y(t), \\ &= \lim_{s\to 0} sY(s), \\ &= \lim_{s\to 0} sG(s)U(s), \\ &= G(0)u_{ss}. \end{aligned}$$

If $G(0)$ exists we call it the *steady state gain*, although we require the further proviso that $G(s)$ be stable for it to be well defined.

Example 1 Suppose $G(s) = \frac{b}{s+a}$. Then the steady state gain is b/a. This is illustrated in Fig 16 with $b = 2$ and $a = 1$.

Example 2 Suppose

$$G(s) = e^{-s}\frac{1.4}{s^2 + 0.6s + 2} + e^{-2s}\frac{1.2}{s^2 + 0.4s + 4}.$$

The system is depicted in Fig 17.

Suppose further $u(t)$ is as illustrated in Fig 18. Clearly this is a complicated system with peculiar dynamics. Nevertheless $u_{ss} = 1$ and $G(0) = 1$. It follows that $y_{ss} = 1$ also, as demonstrated in Fig 19.

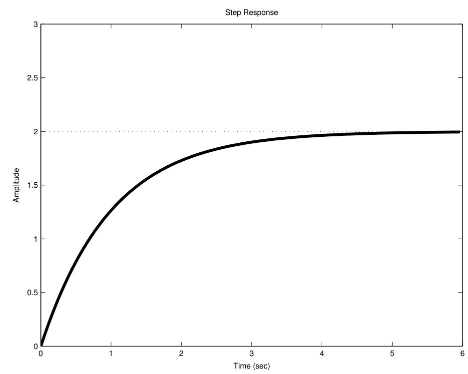


Figure 16: First order step response.

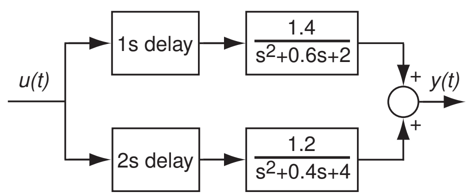


Figure 17: $G(s)$ for the example.

5.2 First order systems

A first order system has transfer function

$$G(s) = \frac{b}{s+a},$$

or equivalently, satisfies the first order differential equation

$$\frac{dy}{dt}(t) + ay(t) = bu(t).$$

Its time response to a unit step can be derived via Laplace transforms and using a partial fraction expansion:

$$\begin{aligned} Y(s) &= G(s)U(s), \\ &= \frac{b}{s+a} \times \frac{1}{s}, \\ &= \frac{b}{a}\left(\frac{1}{s} - \frac{1}{s+a}\right). \end{aligned}$$

So for $t \geq 0$ we have

$$y(t) = \frac{b}{a}\left(1 - e^{-at}\right).$$

This is illustrated in Fig 20.

We may observe the following:

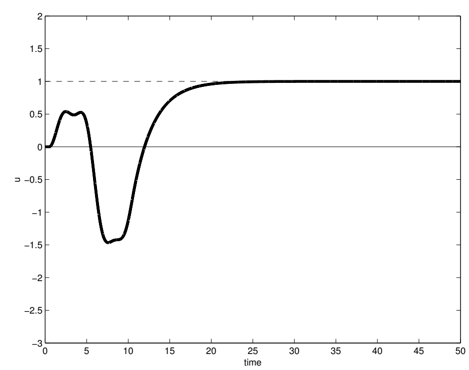


Figure 18: Input for the example.

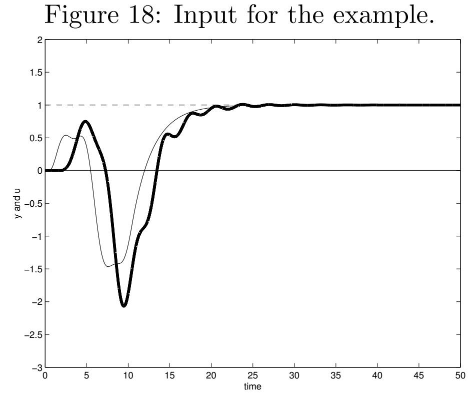


Figure 19: Output for the example.

- Provided a is positive, $y(t)$ settles to a steady state value for large t. As $t \to \infty$, we find $y(t) \to \frac{b}{a}$. This agrees with the final value theorem.
- The response reaches approximately 63.21% of its final value at $t = 1/a$. The value $1/a$ is the time constant. The larger the time constant (or equivalently, the smaller the value of a), the slower the system response.
- Similarly the response reaches 10% of its final value after $t = 0.1054/a$ and 90% of its final value after $t = 2.306/a$. Some authors define the rise time to be the difference between these two values (i.e. the time it takes to go from 10% to 90% of its final value), while some authors define it simply to be the time from 0 to 90% of its final value.
- The slope of the response at $t = 0$ is b.
- If a is negative then $y(t)$ blows up to infinity; in this case the response is said to be unstable.

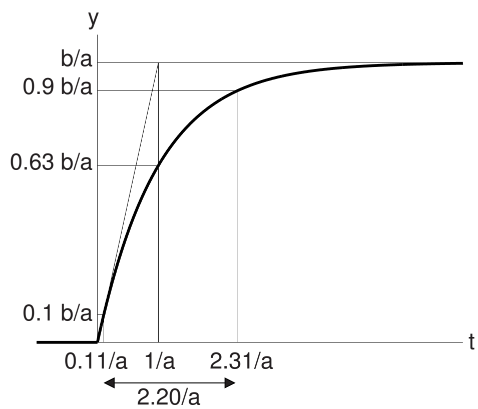


Figure 20: First order step response.

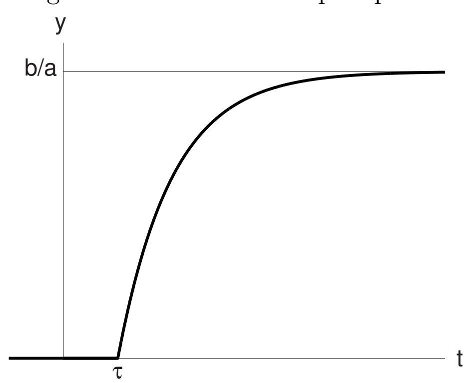


Figure 21: First order step response with delay of τ seconds.

Example: Consider a plant described by a first order transfer function

$$G(s) = \frac{b_o}{s + a_o},$$

in feedback with a proportional gain k_p. The closed-loop system is depicted in Fig 22. The closed-loop transfer function is

$$Y(s) = \frac{k_p b_o}{s + a_o + k_p b_o} R(s),$$

with steady state gain $k_p b_o/(a_o + k_p b_o)$ and time constant $1/(a_o + k_p b_o)$.

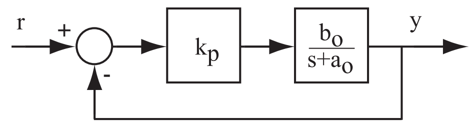


Figure 22: First order transfer function with proportional feedback

5.2.1 First order system with delay

Many systems can be adequately modeled as a first order system with a delay. If the delay is τ seconds, then the transfer function is

$$G(s) = e^{-s\tau}\frac{b}{s+a}.$$

The response is illustrated in Fig 21.

5.3 Second order systems

Second order systems take the general form

$$G(s) = \frac{b_1 s + b_0}{s^2 + a_1 s + a_0}$$

and show a much wider variety of dynamic response. We saw in Section 3 that second order systems are often adequate to describe simple electrical and mechanical systems. They may also be used to loosely approximate first order systems with delay. Specifically the transfer function

$$G(s) = e^{-s\tau}\frac{b}{s+a}$$

can be approximated as

$$G(s) = \frac{-\tau s/2 + 1}{\tau s/2 + 1} \times \frac{b}{s+a}.$$

This is known as a Padé approximation. See Fig 23.

We need to understand how systems with second order transfer functions will behave for different values of b_0, b_1, a_0 and a_1. To simplify matters we will begin by assuming $b_1 = 0$. We will also set $b_0 = ga_0$, and assume that $g = 1$. This means the steady state gain $G(0) = 1$. The response for other values of g can be found by scaling the following results.

Despite these simplifications, we will still need to consider separate cases. We will classify the behaviour according to the roots of the denominator, which are termed the *poles* of the system. They are either both real, or a complex conjugate pair.

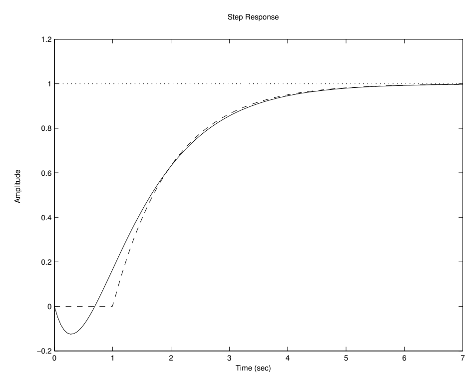


Figure 23: Step response of a first order system with delay $G(s) = e^{-s}\frac{1}{s+1}$ (dashed) and the step response of its approximation: a second order system $G(s) = \frac{-s/2+1}{s/2+1} \times \frac{1}{s+1}$ (solid).

5.3.1 The poles are separate and real

Suppose we may write

$$G(s) = \frac{p_1 p_2}{(s+p_1)(s+p_2)},$$

with both p_1 and p_2 real. The step response has Laplace transform

$$\begin{aligned} Y(s) &= \frac{p_1 p_2}{(s+p_1)(s+p_2)}\frac{1}{s}, \\ &= \frac{1}{s} + \left(\frac{p_2}{p_1 - p_2}\right)\frac{1}{s+p_1} - \left(\frac{p_1}{p_1 - p_2}\right)\frac{1}{s+p_2}. \end{aligned}$$

Hence the step response itself takes the form

$$y(t) = 1 + \left(\frac{p_2}{p_1 - p_2}\right)e^{-p_1 t} - \left(\frac{p_1}{p_1 - p_2}\right)e^{-p_2 t}.$$

We may observe the following (see Fig 24):

- If either p_1 or p_2 is negative (i.e. the corresponding pole is positive) then the step response blows up and we say the transfer function is unstable. If both p_1 and p_2 are positive (i.e. the corresponding poles are negative) then the transfer function is stable.

- If $p_1 >> p_2$ then the step response may be approximated as

$$y(t) \approx 1 - e^{-p_2 t}.$$

  The response is, to the eye, similar to that of a first order system with a pole at $-p_2$ and we say the smaller pole dominates. The main significant difference is that the gradient of the second order response at $t = 0$ is 0.

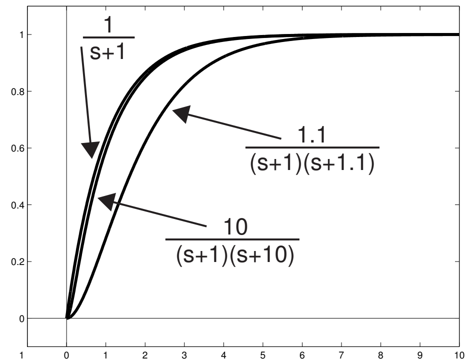


Figure 24: Step responses for second order transfer functions with real poles

- If p_1 and p_2 are similar, the response is significantly different from any first order response.

5.3.2 The poles are identical and real

Suppose both poles are identical and we may write

$$G(s) = \frac{p^2}{(s+p)^2}.$$

The step response has Laplace transform

$$\begin{aligned} Y(s) &= \frac{p^2}{(s+p)^2}\frac{1}{s}, \\ &= \frac{1}{s} - \frac{1}{s+p} - p\frac{1}{(s+p)^2}. \end{aligned}$$

Hence the step response itself takes the form

$$y(t) = 1 - e^{-pt} - pte^{-pt}.$$

Although mathematically this looks quite different from the case where the poles are separate, with reference to Fig 27 we see that the behaviour is in fact very similar. Once again we must have both poles negative for the response to be stable.

5.3.3 The poles are complex

If the poles form a complex conjugate pair, then it is standard to write the transfer function as

$$G(s) = \frac{g\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2}.$$

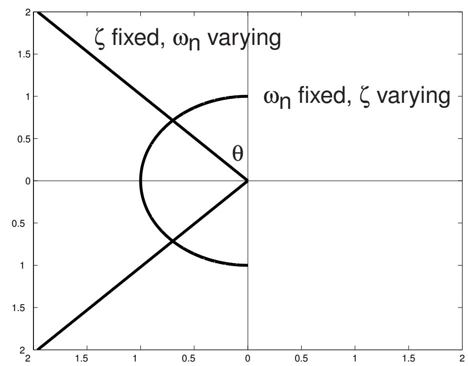


Figure 25: Locus of poles when the denominator of the transfer function is $s^2 + 2\zeta\omega_n s + \omega_n^2$.

Once again we assume $g = 1$. The poles lie at

$$p = -\omega_n \zeta \pm j\omega_n \sqrt{1-\zeta^2}.$$

Omega (ω_n) is the *natural frequency* while zeta (ζ) is the *damping ratio*. Zeta takes values $0 < \zeta < 1$. Fixing ω_n and varying ζ gives a locus of poles on a semi-circle, radius ω_n, which intersects the real axis at $-\omega_n$ when $\zeta = 1$ and the imaginary axis at $\pm\omega_n j$ when $\zeta = 0$. Fixing ζ and varying ω_n gives a locus of poles on a ray at angle θ satisfying

$$\tan\theta = \frac{\sqrt{1-\zeta^2}}{\zeta}.$$

These loci are illustrated in Fig 25.

A tedious calculation gives the step response as

$$y(t) = 1 - \frac{1}{\sqrt{1-\zeta^2}} e^{-\zeta\omega_n t} \sin\left(\omega_n \sqrt{1-\zeta^2}\, t + \theta\right) \tag{6}$$

Just as varying g changes only the magnitude of the response, so changing ω_n changes only the speed of the response. Fig 26 shows three step responses, all with ζ set arbitrarily to 0.3. Respectively they show step responses for:

$$\begin{aligned} G(s) &= \frac{1}{s^2+0.6s+1}, \text{ i.e. with } \omega_n = 1 \text{ and } g = 1 \\ G(s) &= \frac{10}{s^2+0.6s+1}, \text{ i.e. with } \omega_n = 1 \text{ and } g = 10 \\ G(s) &= \frac{100}{s^2+6s+100}, \text{ i.e. with } \omega_n = 10 \text{ and } g = 1 \end{aligned}$$

The shapes of the three plots are identical. Only the relative scalings, indicated by the x- and y-axes, differ.

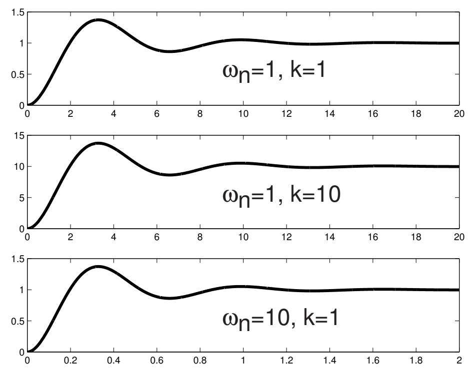


Figure 26: Step responses of second order systems with differing values of ω_n and g.

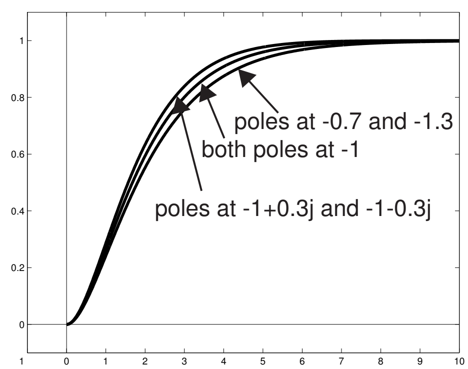


Figure 27: Step responses with two poles close together.

If ζ is near 1 then once again the response is similar to that with two identical poles on the real axis. Fig 27 shows the step response of the following three transfer functions:

$$\begin{aligned} G_1(s) &= \frac{0.7 \times 1.3}{(s+0.7)(s+1.3)} = \frac{0.91}{s^2+2s+0.91}, \\ G_2(s) &= \frac{1}{(s+1)^2} = \frac{1}{s^2+2s+1}, \\ G_3(s) &= \frac{1.09}{(s+1+j0.03)(s+1-j0.3)} = \frac{1.09}{s^2+2s+1.09}. \end{aligned}$$

Note that $G_3(s)$ has natural frequency $\omega_n = \sqrt{1.09} \approx 1.04$ and damping ratio $\zeta \approx 0.96$.

Now suppose we fix ω_n and g but allow ζ to vary. Various step responses are shown in Figs 28 and 29.

For $\zeta < 1$ there is some overshoot to the step response. As ζ decreases, the overshoot increases.

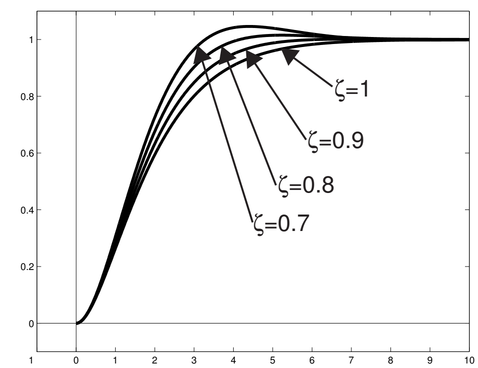


Figure 28: Locus of poles with high damping.

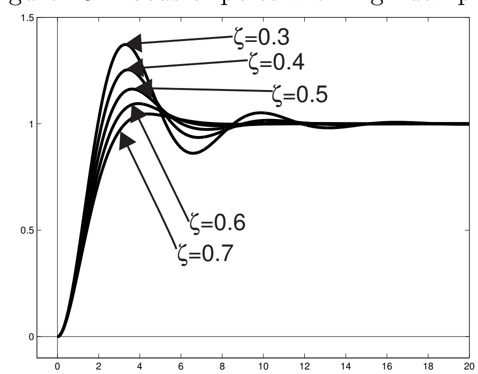


Figure 29: Locus of poles with low damping.

In the limit as $\zeta \to 0$, the response tends to

$$\lim_{\zeta\to 0} y(t) = 1 - \sin\left(\omega t + \frac{\pi}{2}\right).$$

The poles for this limit lie on the imaginary axis, so we say the response is marginally stable. For $\zeta < 0$ the poles lie in the right half plane (i.e. $G(s)$ is unstable) and the response blows up.

5.3.4 Time domain step response specifications

There are three standard time response specifications.

Peak overshoot M_p. For a stable response with a well-defined final value $y_{ss} = \lim_{t\to\infty} y(t)$ the peak value M_p is given by

$$M_p = \frac{y_{\max} - y_{ss}}{y_{ss}},$$

where $y_{\max}$ is the maximum value of $y(t)$.

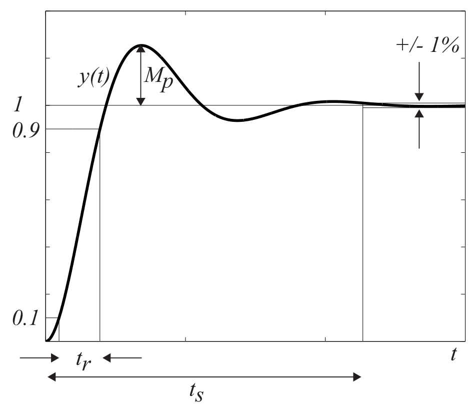


Figure 30: Time domain step response specifications.

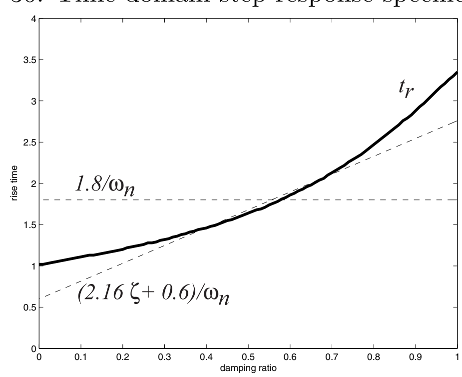


Figure 31: Rise time for second order system with no zero against damping ratio ζ with natural frequency $\omega_n = 1$.

For a second order response with $\zeta < 1$ and no zero we can find the peak value by evaluating where the derivative of (6) is zero. We find the peak occurs at

$$t = \frac{\pi}{\omega_n \sqrt{1-\zeta^2}},$$

and

$$M_p = e^{-\frac{\zeta}{\sqrt{1-\zeta^2}}\pi}.$$

Rise time. We defined the rise time for first order systems as the time for the response to go from 10% to 90% of the final value. The definition applies to more general responses, but is less useful when there is overshoot.

Figure 31 shows the rise time against damping ratio for a second order system with no zero when the natural frequency $\omega_n = 1$. Two commonly used approximations are also shown:

$$\begin{aligned} t_r &\approx \frac{2.16\zeta + 0.6}{\omega_n}, \\ t_r &\approx \frac{1.8}{\omega_n}. \end{aligned}$$

Settling time. The settling time is the time for which the response reaches to within $\pm 1\%$ of the final value and then stays there.

$$\left| \frac{y(t) - y_{ss}}{y_{ss}} \right| < 0.01 \text{ for all } t \geq t_s.$$

For a second order system we may approximate

$$t_s \approx \frac{4.6}{\zeta \omega_n}.$$

The approximation is surprisingly good for $\zeta < 1$, though it derives from the less good approxiation

$$y(t) \approx 1 - e^{-\omega_n \zeta t} \text{ for large } t.$$

5.3.5 Second order transfer functions with a zero

A zero in a transfer function can have a significant effect on the time response. Suppose G is given by the transfer function

$$G(s) = \frac{(bs+1) g \omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2}.$$

We will find it useful to write

$$G(s) = G_1(s) + G_2(s),$$

with

$$\begin{aligned} G_1(s) &= \frac{g\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2}, \\ G_2(s) &= \frac{bg\omega_n^2 s}{s^2 + 2\zeta\omega_n s + \omega_n^2}. \end{aligned}$$

If the corresponding time response is $y_1(t) + y_2(t)$ then y_1 takes the form we studied in the previous section. Meanwhile, by the properties of Laplace transforms,

$$y_2(t) = b \frac{dy_1}{dt}(t).$$

We will not analyse y_2 too closely, except to note

$$y_2(0) = \lim_{t \to \infty} y_2(t) = 0.$$

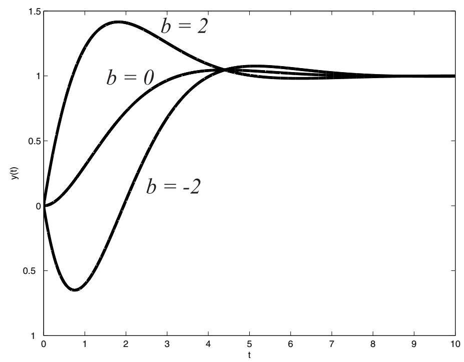


Figure 32: Step response for second order systems with a zero.

Consider the second order transfer function

$$G(s) = \frac{bs + 1}{s^2 + 1.4s + 1}.$$

This has a natural frequency $\omega_n = 1$ and a damping ratio $\zeta = 0.7$. Its steady state gain is 1 for all values of b. Figure 32 shows the time response for the values $b = -2$, $b = 0$ and $b = 2$. Allowing b to be non-zero affects the peak overshoot M_p, rise time t_r and settling time t_s; the approximations of the previous section are no longer valid.

Note in particular the case $b = -2$. The response dips below zero before rising to the steady state value. This so-called “inverse-response” is characteristic of transfer functions with zeros in the right half plane. Transfer functions with such zeros are termed “non-minimum phase” transfer functions, for reasons we will discuss in the next section.

5.4 Key points

- First order dynamics are relatively simple.
- Second order dynamics yield rich behaviour.
- Complex conjugate poles correspond to oscillatory behaviour.

6 Frequency response analysis

In the previous section we classified step responses for low order transfer functions. The classification is not straightforward, even when we write the transfer function in terms of natural frequency ω_n and damping ratio ζ.

An exception is the steady state response. Specifically, if $\lim_{t\to\infty} u(t)$ exists and if $G(s)$ is stable then

$$\lim_{t\to\infty} y(t) = G(0) \lim_{t\to\infty} u(t).$$

The relation is valid for $G(s)$ of any order.

Remarkably, the result may be generalised to the case where $u(t)$ is a sine wave. Specifically, suppose

$$\lim_{t\to\infty} [u(t) - \sin(\omega t)] = 0,$$

and $G(s)$ is stable. Then

$$\lim_{t\to\infty} [y(t) - g\sin(\omega t + \phi)] = 0,$$

for some g and ϕ. Furthermore, if we write

$$G(j\omega) = |G(j\omega)| \exp(j\angle G(j\omega)),$$

then

$$g = |G(j\omega)|,$$
$$\phi = \angle G(j\omega).$$

We usually say (loosely) that if

$$u(t) = \sin(\omega t),$$

and if $G(s)$ is stable then

$$y(t) = |G(j\omega)| \sin(\omega t + \angle G(j\omega)),$$

provided we ignore initial transient conditions.

We call $|G(j\omega)|$ the *gain* of G and $\angle G(j\omega)$ the *phase* of G, each at frequency ω.

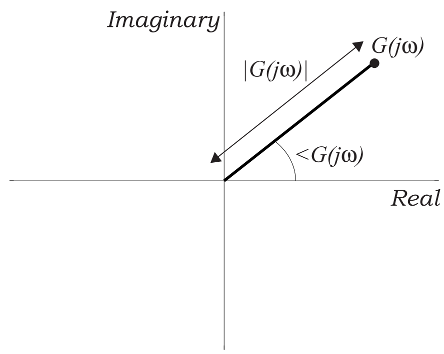


Figure 33: Gain and phase of G evaluated at a specific frequency $G(j\omega)$.

6.1 First order system

Suppose

$$G(s) = \frac{g}{s+a}.$$

Then

$$\begin{aligned}
G(j\omega) &= \frac{g}{a+j\omega}, \\
|G(j\omega)| &= \left|\frac{g}{a+j\omega}\right|, \\
&= g\left|\frac{1}{a+j\omega}\times\frac{a-j\omega}{a-j\omega}\right|, \\
&= g\left|\frac{a}{a^2+\omega^2} - j\frac{\omega}{a^2+\omega^2}\right|, \\
&= g\sqrt{\left(\frac{a}{a^2+\omega^2}\right)^2 + \left(\frac{\omega}{a^2+\omega^2}\right)^2}, \\
&= g\sqrt{\frac{a^2+\omega^2}{(a^2+\omega^2)^2}}, \\
&= \frac{g}{\sqrt{a^2+\omega^2}}, \\
\angle G(j\omega) &= \text{atan}\left[\left(-\frac{\omega}{a^2+\omega^2}\right) / \left(\frac{a}{a^2+\omega^2}\right)\right], \\
&= \text{atan}\left[-\omega/a\right], \\
&= -\text{atan}[\omega/a].
\end{aligned}$$

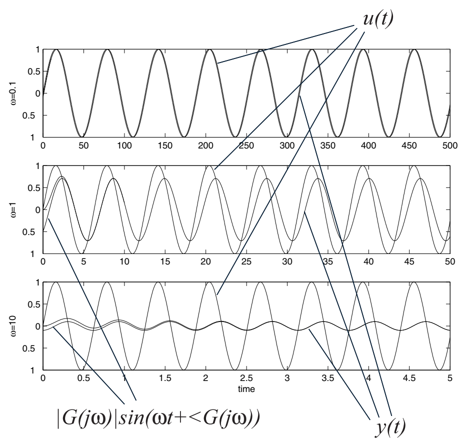


Figure 34: Sine wave response for the first order plant $G(s) = \frac{1}{s+1}$ at frequencies $\omega = 0.1$, $\omega = 1$ and $\omega = 10$. Note the time scales are different for each plot.

- When ω is small
$$|G(j\omega)| \approx \frac{g}{a} \text{ and } \angle G(j\omega) \approx 0^o.$$
- When $\omega = a$
$$|G(j\omega)| = \frac{1}{\sqrt{2}}\frac{g}{a} \text{ and } \angle G(j\omega) = -45^o.$$
- When ω is large
$$|G(j\omega)| \approx \frac{g}{\omega} \text{ and } \angle G(j\omega) \approx -90^o.$$

Fig 34 shows the response of $G(s) = \frac{1}{1+s}$ to sine waves at $\omega = 0.1$, $\omega = 1$ and $\omega = 10$.

6.2 Bode plot

The Bode plot shows gain and phase against frequency on separate graphs.

For practical reasons it is useful to show the log of the gain versus the log of the frequency. Suppose

$$\begin{aligned} G(j\omega) &= G_1(j\omega) \times G_2(j\omega)/G_3(j\omega), \\ &= \left(g_1 e^{j\phi_1}\right) \times \left(g_2 e^{j\phi_2}\right) \div \left(g_3 e^{j\phi_3}\right), \\ &= \frac{g_1 g_2}{g_3} e^{j(\phi_1+\phi_2-\phi_3)}. \end{aligned}$$

Then the gain of G is given by

$$|G(j\omega)| = \frac{g_1 g_2}{g_3},$$

and

$$\log|G(j\omega)| = \log g_1 + \log g_2 - \log g_3.$$

By tradition the gain is shown in decibels defined as

$$|G|_{dB} = 20\log_{10}|G|.$$

For a first order system $G(s) = \frac{g}{s+a}$ we find

- when ω is small
$$|G(j\omega)|_{dB} \approx 20\log_{10}(g/a);$$
in particular a steady state gain of 1 corresponds to 0_{dB};
- when ω is large
$$|G(j\omega)|_{dB} \approx 20\log_{10}(g/\omega);$$
if $\omega = 10a$,
$$|G(j10a)|_{dB} \approx 20\log_{10}(g/a) - 20,$$
and similarly if $\omega = 10^2 a$
$$|G(j10^2 a)|_{dB} \approx 20\log_{10}(g/a) - 40.$$

On a log-log scale the gain follows the horizontal line $20\log_{10}(g/a)$ for small frequencies and then drops at 20dB per decade for large frequencies. It deviates most from these two asymptotes where they meet at $\omega = a$. This frequency is known as the *break frequency*. Here

$$\begin{aligned} |G(j\omega)|_{dB} &= 20\log_{10}\left(\frac{g}{\sqrt{2a^2}}\right), \\ &= 20\log_{10}(g/a) - 20\log_{10}\sqrt{2}, \\ &\approx 20\log_{10} g/a - 3dB. \end{aligned}$$

Similarly the phase of G is given by

$$\angle G(j\omega) = \phi_1 + \phi_2 - \phi_3.$$

Therefore on a Bode plot the phase is drawn on a linear scale, but against frequency scaled in a logarithmic manner, as for the gain. At $\omega = a$ we have

$$\angle G(ja) = -\text{atan}[a/a] = -45^o,$$

and near $\omega = a$ we may approximate

$$\angle G(j\omega) \approx -\frac{\pi}{4} - \frac{1}{2}\log\frac{\omega}{a} \text{ (in radians)}.$$

This corresponds to a tangent that touches 0^o when $\omega/a = e^{-\pi/2} \approx 0.2$ and touches -90^o when $\omega/a = e^{\pi/2} \approx 5$.

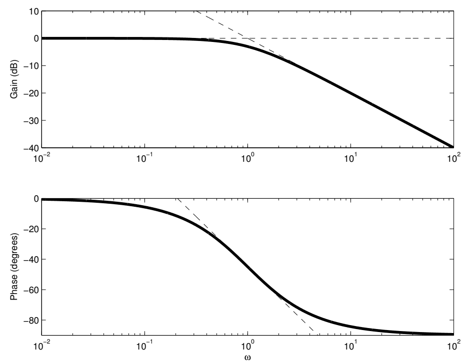


Figure 35: Bode plot of the first order plant $G(s) = \frac{1}{s+1}$.

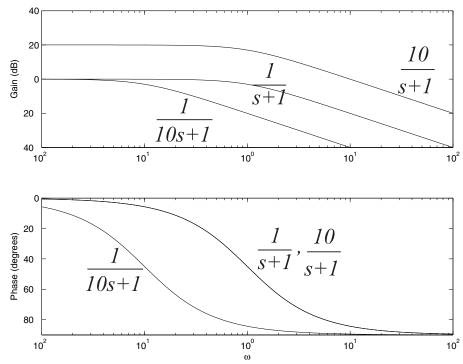


Figure 36: Bode plots of $G(s) = \frac{1}{s+1}$, $G(s) = \frac{10}{s+1}$ and $G(s) = \frac{1}{10s+1}$.

Scaling relations: the Bode plot of $\frac{g}{s+1}$ is the same as $\frac{1}{s+1}$ but moved up by g. The Bode plot of $\frac{1}{as+1}$ is the same as $\frac{1}{s+1}$ but with $-20dB$-per-decade asymptote intersecting $0dB$ at $\omega = 1/a$. See Fig 36.

Grid lines: The Bode plot shows two graphs: the first is a graph of log gain against log frequency, while the second is a graph of phase against log frequency. As the gain is usually expressed in decibels, the corresponding y-axis scale becomes linear. Therefore the standard convention for grid-lines is for them to be linear for both gain and phase, but logarithmic for frequency. See Fig 37.

6.3 Closed-loop response

Since $G(j\omega)$ is a special case of $G(s)$ (i.e. where the complex operator s is mapped onto the positive imaginary axis, $j\omega$ with $0 \leq \omega < \infty$), we may use the same description of closed-loop response we used with Laplace transforms.

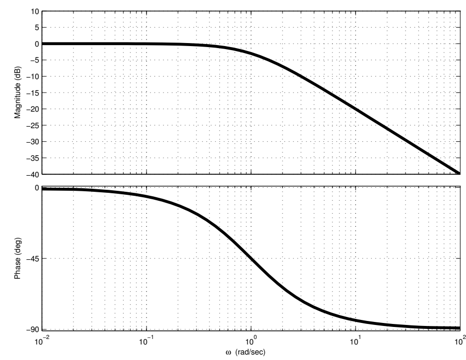


Figure 37: Bode plot of $G(s) = \frac{1}{s+1}$, showing the standard convention for grid lines.

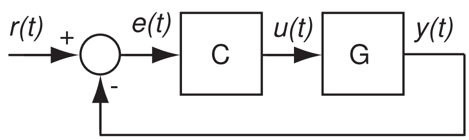


Figure 38: Simple closed loop system.

Example: Suppose (as before) a plant can be represented by the transfer function $G(s)$ given by

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

The controller is a simple gain given by

$$C(s) = k_p.$$

We wish the plant output $y(t)$ to track some reference signal $r(t)$. For simplicity we assume there is no pre-filter $P(s)$, and no noise or disturbance signals. So the feedback loop can be depicted as in Fig 38.

We saw before that

$$\begin{aligned} Y(s) &= \frac{GC}{1+GC}R(s), \\ &= \frac{k_p}{s^2+s+k_p}R(s), \end{aligned}$$

so we may also say

$$Y(j\omega) = \frac{k_p}{-\omega^2 + j\omega + k_p}R(j\omega).$$

Our definitions of sensitivity and complementary sensitivity carry over to the frequency domain:

Sensitivity

$$S(j\omega) = \frac{1}{1 + G(j\omega)C(j\omega)}.$$

Complementary sensitivity

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

As before we have the relations

$$T(j\omega) = G(j\omega)C(j\omega)S(j\omega),$$

and

$$S(j\omega) + T(j\omega) = 1,$$

at all frequencies ω.

6.4 First order elements

It is useful to be able to draw the Bode plots for the objects $s+1$, $s-1$, $s-1$, $\frac{1}{1-s}$, $\frac{1}{s-1}$ and $\frac{1-s}{1+s}$. Note that the first three of these are improper while the fourth and fifth are unstable.

- $G(s) = s + 1$,

$$\begin{aligned} |(1+j\omega)| &= \sqrt{1+\omega^2} \approx \begin{cases} 0dB \text{ for } \omega << 1 \\ +3dB \text{ for } \omega = 1 \\ 20\log_{10}\omega \text{ for } \omega >> 1, \end{cases} \\ \angle(1+j\omega) &= \begin{cases} 0^o \text{ for } \omega \to 0 \\ 45^o \text{ for } \omega = 1 \\ 90^o \text{ for } \omega \to \infty. \end{cases} \end{aligned}$$

  The phase may also be expressed as

$$\angle(1+j\omega) = \begin{cases} -360^o \text{ for } \omega \to 0 \\ -315^o \text{ for } \omega = 1 \\ -270^o \text{ for } \omega \to \infty. \end{cases}$$

- $G(s) = s - 1$.

$$\begin{aligned} |(-1+j\omega)| &= \sqrt{1+\omega^2} \approx \begin{cases} 0dB \text{ for } \omega << 1 \\ +3dB \text{ for } \omega = 1 \\ 20\log_{10}\omega \text{ for } \omega >> 1, \end{cases} \\ \angle(-1+j\omega) &= \begin{cases} -180^o \text{ for } \omega \to 0 \\ -225^o \text{ for } \omega = 1 \\ -270^o \text{ for } \omega \to \infty. \end{cases} \end{aligned}$$

- $G(s) = 1 - s$.

$$
\begin{aligned}
|(1-j\omega)| &= \sqrt{1+\omega^2} \approx \begin{cases} 0dB \text{ for } \omega << 1 \\ +3dB \text{ for } \omega = 1 \\ 20\log_{10}\omega \text{ for } \omega >> 1, \end{cases} \\
\angle(1-j\omega) &= \begin{cases} 0^o \text{ for } \omega \to 0 \\ -45^o \text{ for } \omega = 1 \\ -90^o \text{ for } \omega \to \infty. \end{cases}
\end{aligned}
$$

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

$$
\begin{aligned}
|\frac{1}{1-j\omega}| &= \frac{1}{\sqrt{1+\omega^2}} \approx \begin{cases} 0dB \text{ for } \omega << 1 \\ -3dB \text{ for } \omega = 1 \\ -20\log_{10}\omega \text{ for } \omega >> 1, \end{cases} \\
\angle\frac{1}{1-j\omega} &= \begin{cases} -360^o \text{ for } \omega \to 0 \\ -315^o \text{ for } \omega = 1 \\ -270^o \text{ for } \omega \to \infty. \end{cases}
\end{aligned}
$$

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

$$
\begin{aligned}
|\frac{1}{1-j\omega}| &= \frac{1}{\sqrt{1+\omega^2}} \approx \begin{cases} 0dB \text{ for } \omega << 1 \\ -3dB \text{ for } \omega = 1 \\ -20\log_{10}\omega \text{ for } \omega >> 1, \end{cases} \\
\angle\frac{1}{j\omega-1} &= \begin{cases} -180^o \text{ for } \omega \to 0 \\ -135^o \text{ for } \omega = 1 \\ -90^o \text{ for } \omega \to \infty. \end{cases}
\end{aligned}
$$

These are shown in Fig 39.

- $G(s) = \frac{1-s}{1+s}$.
  The gain is given by

$$|G(j\omega)|_{dB} = 20\log_{10}(1+\omega^2) - 20\log_{10}(1+\omega^2) = 0dB.$$

  The phase is

$$\angle G(j\omega) = \angle(1-j\omega) - \angle(1+j\omega) = -2\angle(1+j\omega).$$

- A delay of τ given by $G(s) = e^{-\tau s}$.
  The gain is

$$|G(j\omega))|_{dB} = 0dB,$$

  while the phase is

$$\angle G(j\omega) = \angle e^{-j\omega\tau} = -\omega\tau.$$

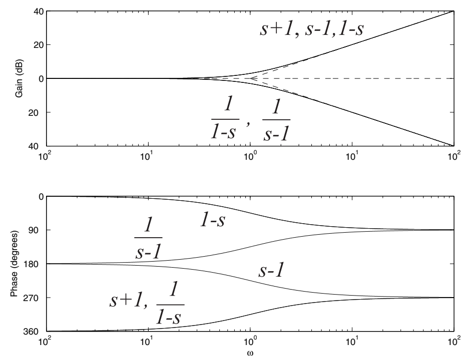


Figure 39: Superimposed Bode plots for $s+1$, $s-1$, $1-s$, $\frac{1}{1-s}$ and $\frac{1}{s-1}$.

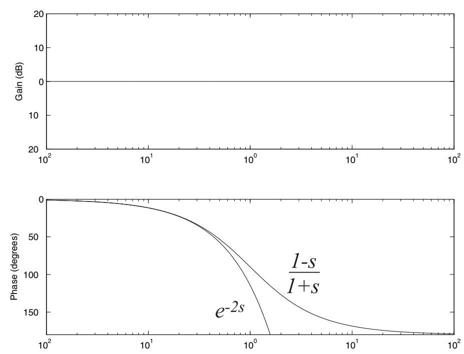


Figure 40: Superimposed Bode plots for $\frac{1-s}{1+s}$ and e^{-2s}.

Since the gain is everywhere 1 (i.e. $0dB$) these are known as all-pass filters. For any transfer function with a pole or zero in the right half-plane, or with delay, there is a corresponding transfer function with the same gain but reduced (negative) phase. Thus linear transfer functions with all poles and zeros in the left half plane (and with no delay) are termed minimum phase transfer functions. We have seen that transfer functions of the form

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

are often used to approximate the effects of delay.

6.5 Second order systems

6.5.1 Separate real poles

Consider the stable second order transfer function

$$G(s) = \frac{g}{(s+p_1)(s+p_2)},$$

with both p_1 and p_2 positive. Without loss of generality suppose $p_1 \leq p_2$.

This has steady state gain $g/(p_1 p_2)$ and two break frequencies $\omega = p_1$ and $\omega = p_2$. Hence we may say:

At low frequencies $\omega < p_1$. The gain can be approximated by a horizonal line at $g/(p_1 p_2)$. The phase tends to 0^o as $\omega \to 0$.

At mid-frequencies $p_1 < \omega < p_2$. The gain can be approximated by an asymptote dropping at -20dB/decade. The phase passes through -90^o at $\omega = \sqrt{p_1 p_2}$.

At high frequencies $\omega > p_2$. The gain can be approximated by an asymptote dropping at -40dB/decade. The phase tends to -180^o as $\omega \to \infty$.

At $\omega = p_1$. The gain is approximately -3dB below the steady state value. The phase is approximately -45^o. If p_2 is close in value to p_1 both these values will be lower.

At $\omega = p_2$. The gain is approximately -3dB below the intersection of the two sloping asymptotes. The phase is approximately -135^o. If p_2 is close in value to p_1 both these values will be higher.

Figs 41 and 42 show Bode plots of two such transfer functions, one with the poles far apart (in frequency) and one with the poles close (in frequency). Grid lines are included so that the intersections of the asymptotes can be seen.

6.5.2 Identical real poles

Suppose we have the stable second order transfer function

$$G(s) = \frac{g}{(s+p)^2}.$$

Its steady state gain is g/p^2. We can obtain the remaining characteristics from superimposing two first order transfer functions, each with a pole at $s = -p$. Hence:

At low frequencies $\omega < p$. The gain can be approximated by a horizonal line at g/p^2. The phase tends to 0^o as $\omega \to 0$.

At high frequencies $\omega > p$. The gain can be approximated by an asymptote dropping at -40dB/decade. The phase tends to -180^o as $\omega \to \infty$.

At $\omega = p$. The gain is approximately -6dB below the steady state value. The phase is approximately -90^o.

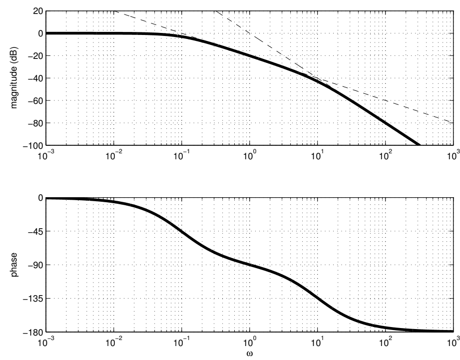


Figure 41: Bode plot of $G(s) = \frac{1}{(s+0.1)(s+10)}$. The two real poles are far apart.

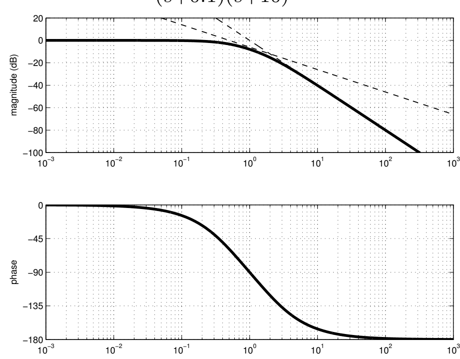


Figure 42: Bode plot of $G(s) = \frac{1}{(s+0.5)(s+2)}$. The two real poles are close together.

6.5.3 Complex conjugate poles

Consider the second order transfer function

$$G(s) = \frac{g\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2}.$$

Its frequency response at frequency ω is

$$\begin{aligned} G(j\omega) &= \frac{g\omega_n^2}{\omega_n^2 - \omega^2 + 2j\zeta\omega_n\omega}, \\ &= \frac{g}{1 - \omega^2/\omega_n^2 + 2j\zeta\omega/\omega_n}. \end{aligned}$$

When ω is small,

$$\begin{aligned} G(j\omega) &\approx g, \\ |G(j\omega)| &\approx 20\log_{10} g dB, \\ \angle G(j\omega) &\approx 0^o. \end{aligned}$$

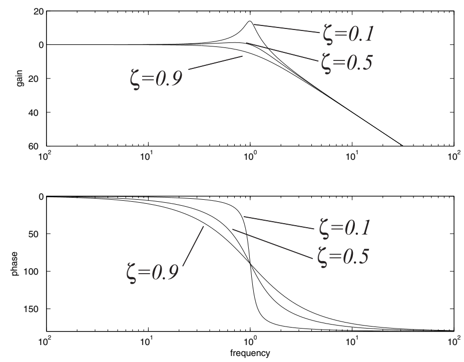


Figure 43: Bode plot of three second order systems with damping ratios $\zeta = 0.1$, $\zeta = 0.5$ and $\zeta = 0.9$.

When ω is large,

$$\begin{aligned} G(j\omega) &\approx -\frac{g\omega_n^2}{\omega^2}, \\ |G(j\omega)| &\approx (20\log_{10} g + 40\log_{10}\omega_n - 40\log_{10}\omega)dB, \\ \angle G(j\omega) &\approx -180^o. \end{aligned}$$

When $\omega = \omega_n$ for the more general case

$$\begin{aligned} G(j\omega) &= \frac{g}{2j\zeta}, \\ |G(j\omega)| &= (20\log_{10} g - 20\log_{10} 2\zeta)dB, \\ \angle G(j\omega) &= -90^o. \end{aligned}$$

If the damping ratio $\zeta < 1/\sqrt{2}$ there is a peak in the gain. Since we have

$$|G(j\omega)|^2 = \frac{(g\omega_n^2)^2}{(\omega_n^2 - \omega^2)^2 + (2\zeta\omega\omega_n)^2},$$

the peak occurs where

$$f(\omega^2) = (\omega_n^2 - \omega^2)^2 + (2\zeta\omega\omega_n)^2,$$

is minimum. Differentiating with respect to ω^2 gives

$$\frac{df(\omega^2)}{d(\omega^2)} = 2(\omega^2 - \omega_n^2) + 4\zeta^2\omega_n^2,$$

which is zero when

$$\omega^2 = \omega_n^2(1 - 2\zeta^2),$$

or equivalently

$$\omega = \omega_n\sqrt{1 - 2\zeta^2}.$$

At this frequency we find

$$\begin{aligned} G(j\omega) &= \frac{g}{2\zeta^2 + 2j\zeta\sqrt{1-2\zeta^2}}, \\ |G(j\omega)| &= \frac{g}{2\zeta\sqrt{1-\zeta^2}}. \end{aligned}$$

Bode plots of second order systems with $g = 1$ and $\omega_n = 1$ are shown in Fig 43 for three different damping ratios. Note that in the limit as $\zeta \to 0$ the peak gain tends to infinity and the phase transition from 0^o to -180^o tends to a step with respect to frequency.

6.6 Derivation of the main result

Suppose $G(s)$ is a stable rational transfer function with separate poles $-p_1$, $-p_2$, …, $-p_n$ (these may have complex values). Since $G(s)$ is stable we must have Real$[p_i] > 0$ for all i. We may write

$$G(s) = \frac{b(s)}{(s+p_1)(s+p_2)\cdots(s+p_n)},$$

for some numerator polynomial $b(s)$. Suppose the input to the plant is a sine wave

$$u(t) = \sin \omega t,$$

with Laplace transform

$$U(s) = \frac{w}{s^2 + \omega^2}.$$

The output is

$$\begin{aligned} Y(s) &= G(s)U(s), \\ &= \frac{b(s)}{(s+p_1)(s+p_2)\cdots(s+p_n)} \frac{w}{(s+j\omega)(s-j\omega)}, \\ &= \frac{A_1}{s+j\omega} + \frac{A_2}{s-j\omega} + \frac{B_1}{s+p_1} + \ldots + \frac{B_n}{s+p_n}, \end{aligned}$$

by partial fraction expansion for some $A_1, A_2, B_1, \ldots, B_n$. So

$$y(t) = A_1 e^{-j\omega t} + A_2 e^{j\omega t} + B_1 e^{-p_1 t} + \cdots + B_n e^{-p_n t}.$$

For large t the terms with coefficients B_i die out, leaving

$$y_{ss}(t) = A_1 e^{-j\omega t} + A_2 e^{j\omega t}.$$

From the partial fraction expansion we find

$$
\begin{aligned}
A_1 &= G(s)\frac{\omega}{s-j\omega}\bigg|_{s=-j\omega}, \\
&= \frac{j}{2}G(-j\omega), \\
A_2 &= G(s)\frac{\omega}{s+j\omega}\bigg|_{s=j\omega}, \\
&= \frac{-j}{2}G(j\omega).
\end{aligned}
$$

It is possible to show that

$$|G(-j\omega)| = |G(j\omega)| \text{ and } \angle G(-j\omega) = -\angle G(j\omega).$$

Hence

$$
\begin{aligned}
y_{ss}(t) &= \frac{j}{2}G(-j\omega)e^{-j\omega t} - \frac{j}{2}G(j\omega)e^{j\omega t}, \\
&= \frac{j}{2}|G(j\omega)|\left(e^{-j\angle G(j\omega)}e^{-j\omega t} - e^{j\angle G(j\omega)}e^{j\omega t}\right), \\
&= \frac{j}{2}|G(j\omega)|(-2j)\sin\left(\omega t + \angle G(j\omega)\right), \\
&= |G(j\omega)|\sin\left(\omega t + \angle G(j\omega)\right).
\end{aligned}
$$

The result is readily extendable to more general transfer functions.

6.7 Key points

- For a transfer function $G(s)$ the object $G(j\omega)$ characterises the response to injected sine waves.
- Bode plots depict the frequency response graphically. The first sub-plot shows gain (in dB) against frequency (log scale). The second sub-plot shows phase (linear scale) against frequency (log scale).
- We have built a catalogue of first and second order responses.

7 Review of analysis tools: first order response to a sine wave

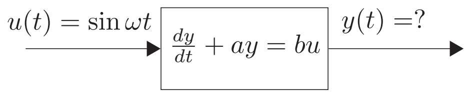


Figure 44: Response of a system with first order dynamics to a sine wave input.

Let us pause to compare the analytical tools we now have available. We will consider a simple dynamic system with a simple input. Suppose our input-output relation is

$$\frac{dy}{dt} + ay = bu,$$

and the input is a sine wave, $u(t) = \sin \omega t$ for some frequency ω. We will derive the solution for y in three ways:

1. Time domain analysis. This provides a complete solution, but is not straightforward.
2. Laplace domain analysis. Once again, this provides a complete solution. Finding the solution is simpler than with time domain analysis, though finding the parameters (via partial fraction expansion) can be tricky.
3. Frequency domain analysis. This is really easy, though we lose sight of the initial transients.

7.1 Time domain analysis

We can find a complete solution using time domain analysis.

The general solution is

$$y(t) = M \sin(\omega t + \phi) + K e^{-at},$$

for some $M = M(\omega)$, $\phi = \phi(\omega)$ and $K = K(\omega)$.

Substituting back into the differential equation we find

$$M\omega \cos(\omega t + \phi) + aM \sin(\omega t + \phi) = b \sin \omega t.$$

This becomes

$$M\omega \cos \omega t \cos \phi - M\omega \sin \omega t \sin \phi + aM \sin \omega t \cos \phi + aM \cos \omega t \sin \phi = b \sin \omega t.$$

Equating coefficients in $\cos \omega t$ gives

$$M\omega \cos \phi + aM \sin \phi = 0,$$

and hence

$$\tan\phi = -\frac{\omega}{a}.$$

Equating coefficients in $\sin\omega t$ gives

$$-M\omega\sin\phi + aM\cos\phi = b,$$

and hence

$$-M\omega\tan\phi + aM = b\sqrt{1+\tan^2\phi},$$

which reduces to

$$M = \frac{b}{\sqrt{\omega^2 + a^2}}.$$

Checking initial conditions (we assume both y and u are zero for $t \leq 0$) we find

$$M\sin\phi + K = 0.$$

Hence

$$K = \frac{b\omega}{\omega^2 + a^2}.$$

But since the term in e^{-at} decays over time (provided $a > 0$), we have the solution

$$y(t) \approx M\sin(\omega t + \phi) \text{ for large } t.$$

7.2 Laplace domain analysis

It is usually easier to find $y(t)$ in the Laplace domain. We have

$$U(s) = \frac{\omega}{s^2 + \omega^2},$$

and

$$Y(s) = G(s)U(s) \text{ with } G(s) = \frac{b}{s+a}.$$

Hence

$$\begin{aligned} Y(s) &= \frac{b}{s+a} \times \frac{\omega}{s^2+\omega^2}, \\ &= \frac{A}{s+a} + \frac{Bs+C}{s^2+\omega^2}, \end{aligned}$$

with

$$\begin{aligned} A + B &= 0, \\ aB + C &= 0, \\ \omega^2 A + aC &= b\omega. \end{aligned}$$

This has solution

$$A = \frac{b\omega}{\omega^2+a^2}, B = -\frac{b\omega}{\omega^2+a^2}, C = \frac{ab\omega}{\omega^2+a^2}.$$

So

$$y(t) = \frac{b\omega}{\omega^2 + a^2} e^{-at} - \frac{b\omega}{\omega^2 + a^2} \cos \omega t + \frac{ab}{\omega^2 + a^2} \sin \omega t,$$
$$= \frac{b\omega}{\omega^2 + a^2} e^{-at} + \frac{b}{\sqrt{\omega^2 + a^2}} \sin(\omega t + \phi),$$

where $\tan \phi = -\omega/a$.

7.3 Frequency domain analysis

It is even more straight forward to derive the frequency response directly. If we ignore initial transients, then

$$y(t) = |G(j\omega)| \sin(\omega t + \phi) \text{ with } \phi = \angle G(j\omega).$$

We have

$$G(s) = \frac{b}{s + a},$$

so

$$G(j\omega) = \frac{b}{j\omega + a} = \frac{b(a - j\omega)}{a^2 + \omega^2}.$$

Hence

$$|G(j\omega)| = \frac{b}{a^2 + \omega^2} \sqrt{a^2 + \omega^2} = \frac{b}{\sqrt{\omega^2 + a^2}},$$

and

$$\tan \angle G(j\omega) = -\frac{\omega}{a}.$$

Note that even these equations disguise the simple nature of the response which becomes clear when expressed in a Bode plot.

Part III
Classical control

8 PID Control

PID control is the standard form of control. Loops with PID controllers can be found in many diverse industrial applications. We will introduce the basic concepts, using our insight in time domain, Laplace domain and frequency domain properties.

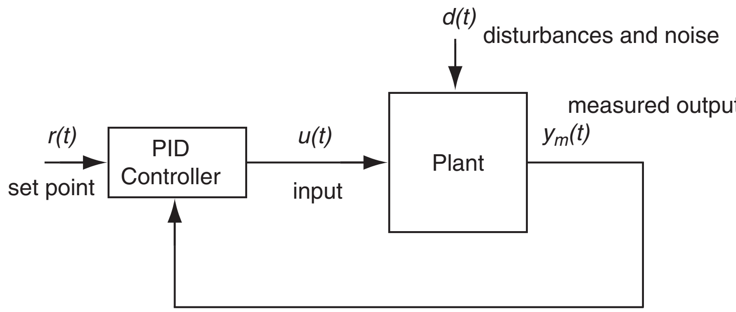


Figure 45: PID controller in a closed-loop system.

8.1 Proportional control

The most simple form of continuous feedback is proportional control, where

$$u(t) = k_p \left(r(t) - y(t) \right),$$

for some *proportional gain* $k_p > 0$. If the open-loop response is

$$Y(s) = G(s)U(s) + D(s),$$

then in closed loop we have

$$Y(s) = \frac{k_p G(s)}{1 + k_p G(s)} R(s) + \frac{1}{1 + k_p G(s)} D(s).$$

8.1.1 Stable first order plant

Suppose the plant is first order with

$$G(s) = \frac{b}{s+a},$$

with $a, b > 0$. Then the sensitivity is

$$S(s) = \frac{1}{1 + k_p G(s)} = \frac{s+a}{s+a+k_p b},$$

and similarly the complementary sensitivity is

$$T(s) = \frac{k_p G(s)}{1 + k_p G(s)} = \frac{k_p b}{s+a+k_p b}.$$

Thus in closed-loop we have

$$Y(s) = \frac{k_p b}{s + a + k_p b} R(s) + \frac{s + a}{s + a + k_p b} D(s).$$

We should observe the following:

- We have the usual relation $S(s) + T(s) = 1$.
- The closed-loop response is always stable.
- At steady state we find

$$y_{ss} = T(0) r_{ss} + S(0) d_{ss} = \frac{k_p b}{a + k_p b} r_{ss} + \frac{a}{a + k_p b} d_{ss}.$$

- $S(s)$ and $T(s)$ are both first order. They share the same denominator with break frequency $a + k_p b$.
- The closed-loop bandwidth is higher than the open-loop bandwidth, and increases with k_p.

As an example, consider the first order system

$$Y(s) = \frac{1}{1 + s} R(s) + D(s).$$

Its open loop time response to a square wave set point change and a step disturbance at time $t = 30$ is shown in Fig 46. Also shown is its closed-loop step response with proportional control $k_p = 10$. Fig 47 shows the open-loop gain $|G(j\omega)|$ and the closed-loop sensitivities $S(j\omega)$ and $T(j\omega)$.

8.1.2 Unstable first order plant

Suppose instead $G(s)$ is given by

$$G(s) = \frac{b}{s - a},$$

with $b > 0$ and $a > 0$. The plant is open-loop unstable, but the closed-loop response

$$Y(s) = \frac{k_p b}{s + k_p b - a} R(s) + \frac{s - a}{s + k_p b - a} D(s),$$

is stable provided

$$k_p b - a > 0.$$

This is an example of negative feedback stabilizing an open-loop unstable plant.

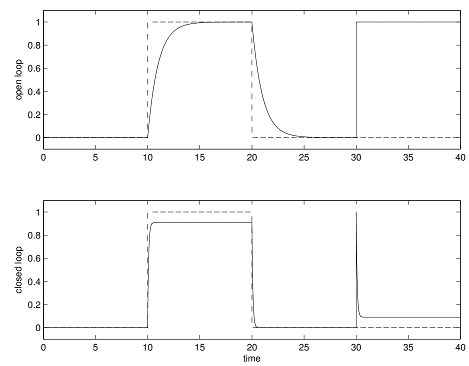


Figure 46: Open and closed-loop time responses for first order plant with proportional control.

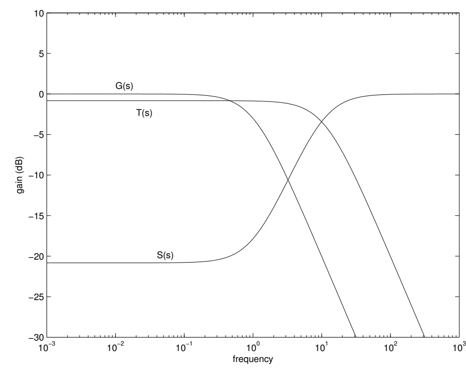


Figure 47: Open and closed-loop frequency responses for first order plant with proportional control.

8.1.3 Stable second order plant (with no zero)

Suppose the plant is

$$G(s) = \frac{b_0}{s^2 + a_1 s + a_0}.$$

The closed-loop response is determined by the sensitivities

$$\begin{aligned} S(s) &= \frac{1}{1 + k_p G(s)} = \frac{s^2 + a_1 s + a_0}{s^2 + a_1 s + a_0 + k_p b}, \\ T(s) &= \frac{k_p G(s)}{1 + k_p G(s)} = \frac{k_p b_0}{s^2 + a_1 s + a_0 + k_p b}. \end{aligned}$$

Focusing on the complementary sensitivity $T(s)$, we should observe:

- The steady state gain is

$$\frac{k_p b_0}{a_0 + k_p b_0},$$

and approaches 1 for large k_p.

- The natural frequency is

$$\omega_n = \sqrt{a_0 + k_p b_0},$$

and increases as k_p increases.

- The damping ratio is

$$\zeta = \frac{a_1}{2\sqrt{a_0 + k_p b_0}},$$

and approaches 0 for large k_p.

Thus we expect a faster but more oscillatory response as we increase k_p. This is illustrated in Figs 48 and 49 where we show the open and closed-loop responses of

$$G(s) = \frac{4}{s^2 + 5s + 4},$$

with the proportional controller $k_p = 10$.

8.1.4 Higher order plants and plants with delay

If the gain is sufficiently high then proportional control will destabilize any higher order plant with relative degree greater than two. Sufficiently high gain proportional control will also destabilize any plant with delay.

8.2 Integral action and PI control

We have seen that high gain control is desirable in that it reduces steady state error (to both set point changes and to disturbances). However it is often unachievable; it leads to low damping ratios with second order plants and leads to instability with higher order plants or when there is any delay.

For most plants it is possible to ensure zero steady state error (to step set point changes and to step disturbances) by introducing integral action. The most simple form is PI (proportional integral) control, which takes the form

$$u(t) = k_p e(t) + \frac{k_p}{T_i}\int^t e(\tau)\,d\tau.$$

Taking Laplace transforms

$$Y(s) = k_p\left(1 + \frac{1}{T_i s}\right)U(s) = \frac{k_p(T_i s + 1)}{T_i s}U(s).$$

So the transfer function for a PI controller is

$$C(s) = \frac{k_p(T_i s + 1)}{T_i s}.$$

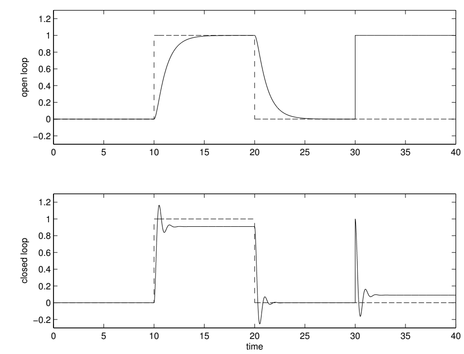


Figure 48: Open and closed-loop time responses for second order plant with proportional control.

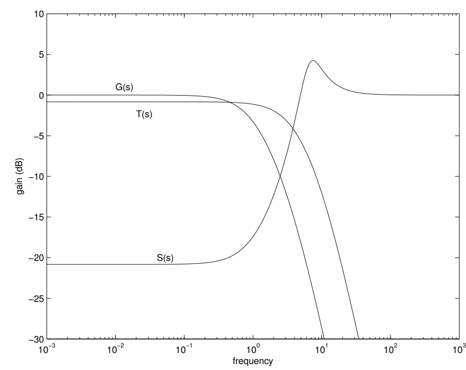


Figure 49: Open and closed-loop frequency responses for second order plant with proportional control.

The term T_i is called the *integral time*. High values of T_i reduce the integral action, while low values increase it. The Bode plot of a PI controller is illustrated in Fig 50

For a first order plant

$$G(s) = \frac{b}{s+a},$$

we find

$$\begin{aligned} S(s) &= \frac{s(s+a)}{s^2 + (a + k_p b)s + k_p b/T_i}, \\ T(s) &= \frac{k_p b(s + 1/T_i)}{s^2 + (a + k_p b)s + k_p b/T_i}. \end{aligned}$$

In particular

$$\begin{aligned} S(0) &= 0, \\ T(0) &= 1. \end{aligned}$$

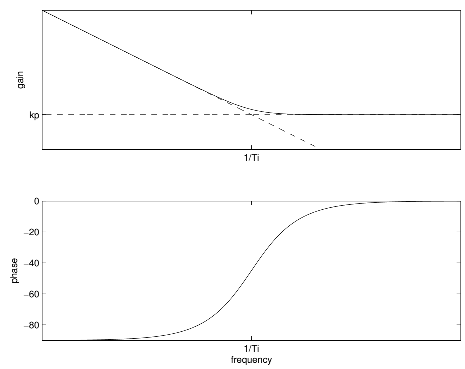


Figure 50: Bode plot of a PI controller.

Fig 51 illustrates the closed-loop response of the plant

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

with the controllers

$$C(s) = 4\left(1 + \frac{3}{4s}\right),\ C(s) = 4\left(1 + \frac{1}{s}\right) \text{ and } C(s) = 4\left(1 + \frac{5}{4s}\right).$$

Fig 52 illustrates the corresponding sensitivities and complementary sensitivities.

A similar phenomenon occurs with more general plants, and with more general controllers. Suppose the plant has numerator $n(s)$ and denominator $d(s)$ while the controller has numerator $n_c(s)$ and denominator $d_c(s)$. Thus

$$G(s) = \frac{n(s)}{d(s)},\ C(s) = \frac{n_c(s)}{d_c(s)},$$

and

$$\begin{aligned} T(s) &= \frac{n(s)n_c(s)}{n(s)n_c(s) + d(s)d_c(s)}, \\ S(s) &= \frac{d(s)d_c(s)}{n(s)n_c(s) + d(s)d_c(s)}. \end{aligned}$$

The controller has integral action if

$$|C(j\omega)| \to \infty \text{ as } \omega \to 0.$$

This occurs if $d_c(0) = 0$ but $n_c(0) \neq 0$. In this case we may write

$$d_c(s) = s\tilde{d}_c(s),$$

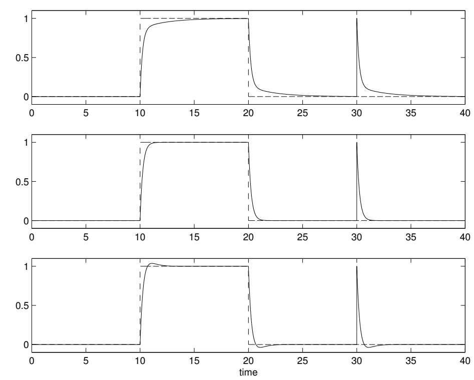


Figure 51: Closed loop responses of plant with three different PI controllers.

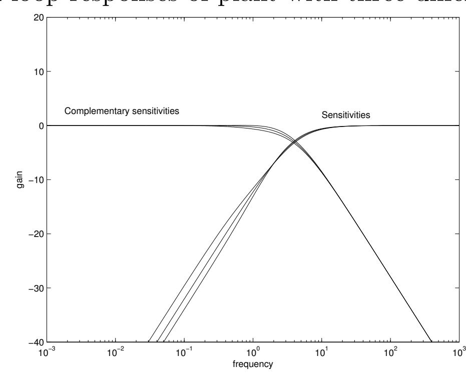


Figure 52: Closed-loop sensitivities of plant with three different PI controllers.

for some other polynomial $\tilde{d_c}(s)$. Suppose further that $n(0) \neq 0$. Then we have

$$\begin{aligned} T(0) &= \frac{n(0)n_c(0)}{n(0)n_c(0)} = 1, \\ S(0) &= \frac{0}{n(0)n_c(0)} = 0. \end{aligned}$$

8.3 Derivative action and PD control

8.3.1 PD and PDγ control

High gain control is also desirable in that it increases the bandwidth. But in the case where high gain proportional control is unachievable, integral action will not generally increase the bandwidth. In this case derivative action is useful, and the simplest form is PD (proportional derivative) control.

An idealised PD controller takes the form

$$u(t) = k_p e(t) + k_p T_d \frac{de(t)}{dt}.$$

T_d is called the *derivative time*. The transfer function of an idealised PD controller is

$$C_{PD}(s) = k_p(1 + T_d s).$$

This is improper (and hence unrealisable). In practice analog PD controllers are usually implemented with transfer function

$$C_{PD\gamma}(s) = k_p\left(1 + \frac{T_d s}{1 + \gamma T_d s}\right),$$

with γ small. This is sometimes called PDγ control.

8.3.2 Frequency response

Typical frequency responses of C_{PD} and $C_{PD\gamma}$ are shown in Fig 53. The gain of C_{PD} is k_p at low frequency, but increases with ω at high frequencies, with a break point at $1/T_d$. By contrast, the gain of $C_{PD\gamma}$ reaches a maximum of $k_p(1+\gamma)/\gamma$ with break points at $1/(T_d + T_d\gamma)$ and $1/(T_d\gamma)$. The phase of C_{PD} rises from 0^o to 90^o, while the phase of $C_{PD\gamma}$ returns to 0^o at high frequency.

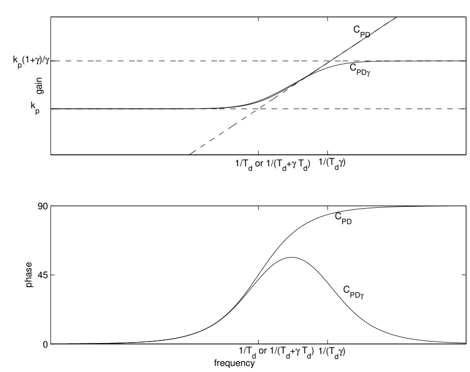


Figure 53: Typical Bode plots of PD and PDγ controllers.

8.3.3 Second order plant

Consider the second order system

$$G(s) = \frac{b_0}{s^2 + a_1 s + a_0},$$

together with an idealised PD controller C_{PD}. The closed-loop sensitivities are

$$\begin{aligned} T(s) &= \frac{G(s)C_{PD}(s)}{1+G(s)C_{PD}(s)}, \\ &= \frac{k_p b_0(1+T_d s)}{s^2+(a_1+k_pT_db_0)s+(a_0+k_pb_0)}, \\ S(s) &= \frac{1}{1+G(s)C_{PD}(s)}, \\ &= \frac{s^2+a_1s+a_0}{s^2+(a_1+k_pT_db_0)s+(a_0+k_pb_0)}. \end{aligned}$$

The derivative action gives us an extra degree of freedom, that allows us to choose the closed-loop natural frequency ω_n and the closed-loop damping ζ independently. Specifically we can choose

$$\begin{aligned} k_p &= \frac{\omega_n^2-a_0}{b_0}, \\ T_d &= \frac{2\zeta\omega_n-a_1}{k_pb_0}. \end{aligned}$$

Example: Consider the second order plant

$$G(s)=\frac{4}{s^2+5s+4}.$$

We saw that a proportional only controller with $k_p=10$ led to a step response with some brief oscillations. This gain gives a closed-loop natural frequency

$$\omega_n^2=b_0k_p+a_0=4\times 10+4=44.$$

With $T_d=0$ the damping ratio is

$$\zeta=\frac{a_1}{2\omega_n}=\frac{5}{2\sqrt{44}}\approx 0.38.$$

Suppose we require a damping ratio $\zeta=0.75$. Then we should set

$$T_d=\frac{2\zeta\omega_n-a_1}{k_pb_0}=\frac{2\times 0.75\times\sqrt{44}-5}{10\times 4}\approx 0.12.$$

A PDγ controller with $\gamma=0.1$ takes the form

$$C(s)=10\left(1+\frac{0.12s}{1+0.012s}\right).$$

Fig 54 shows the closed-loop response of the plant with such a controller, and should be compared with Fig 48 where only proportional action was used.

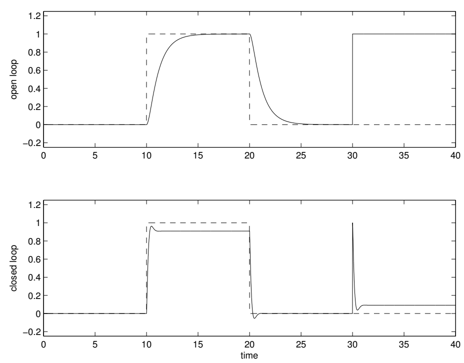


Figure 54: Closed-loop time response for a second order plant with PDγ control.

8.4 PID control and Ziegler-Nichols tuning

PID controllers include proportional, integral and derivative action. The standard form is:

$$U(s) = C_{PID}(s)E(s),$$

with

$$e(t) = r(t) - y_m(t),$$

and

$$C_{PID}(s) = k_p\left(1 + \frac{1}{T_i s} + T_d s\right).$$

As with the idealised PD controller this is unrealisable, and a PID controller usually takes the form

$$C_{PID\gamma}(s) = k_p\left(1 + \frac{1}{T_i s} + \frac{T_d s}{1 + \gamma T_d s}\right),$$

with γ small.

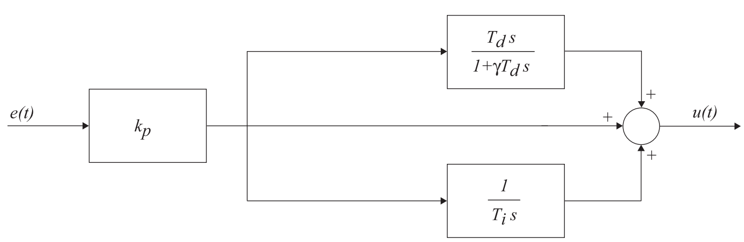


Figure 55: Idealised structure for PIDγ controller.

There are many tuning rules for choosing k_p, T_i and T_d in a PID controller. The Ziegler-Nichols rules are the most popular: there are two different sets of rules.

Ziegler-Nichols step response method

The Ziegler-Nichols step response tuning rules are appropriate for plants whose unit step response appears similar to that in Fig 56 (this is reasonable for many process control applications). In this case a tangent line is drawn at the steepest part of the slope, and the parameters a and L determined from where the tangent intersects the y- and a-axes respectively. Gains are then chosen according to Table 1. The Bode plot of a typical PID controller with Ziegler-Nichols tuning is shown in Fig 57.

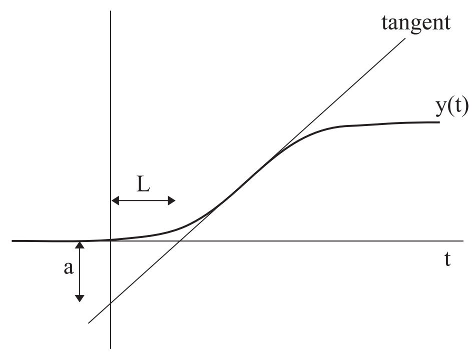


Figure 56: Measurements for Ziegler-Nichols step response method.

Controller	k_p	T_i	T_d
P	$1/a$		
PI	$0.9/a$	$3L$	
PID	$1.2/a$	$2L$	$L/2$

Table 1: Ziegler-Nichols step response method gains

Controller	k_p	T_i	T_d
P	$0.5K_u$		
PI	$0.4K_u$	$0.8T_u$	
PID	$0.6K_u$	$0.5T_u$	$0.125T_u$

Table 2: Ziegler-Nichols frequency response method gains.

Ziegler-Nichols frequency response method

For some plants it is inappropriate to perform open-loop step tests (for example, the plant may be marginally stable in open loop). In this case, it is sometimes possible to tune a PID controller using the Ziegler-Nichols frequency response method. Here the plant is kept in closed-loop, but with a proportional only controller. The gain of the controller is increased until the plant oscillates with steady amplitude (ie with a higher gain the

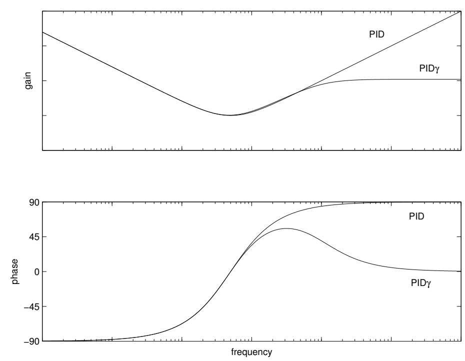


Figure 57: Typical Bode plots of PID and PIDγ controllers.

plant goes unstable and with a lower gain the closed-loop step response of the plant attenuates). The corresponding gain is called the *ultimate gain* K_u and the period of the corresponding oscillations is called the *ultimate period* T_u. PID gains should then be set according to Table 2.

As with any tuning rule, the Ziegler-Nichols tuning rules should be taken as first pass only. Often they lead to an underdamped closed-loop response, and the response can be improved by reducing the proportional gain. Where precise tuning is required the rules can usually be improved on by taking into account the particular control system design specifications.

8.5 PID implementation

PID controllers are often implemented using off-the-shelf software. Such controllers often have non-standard configuration and use non-standard terminology. In particular:

- Sometimes an integral gain k_i is used instead of integral time T_i. Increasing k_i increases integral action (as opposed to increasing T_i which reduces integral action). Similarly a derivative gain k_d may be used with different units (scaling) to T_d.
- So far we have discussed the so-called "standard form". There are two other forms:

  1. Parallel form where
  $$C(s) = k_p + \frac{k_i}{s} + k_d s.$$
  2. Interaction form where
  $$C(s) = k_p' \left(1 + \frac{1}{sT_i'}\right)\left(1 + sT_d'\right).$$

- Modern PID controllers are implemented digitally. PID gains (and structures) may be tailored to such implementation—this is beyond the scope of this course.

8.6 Set-point tracking

While derivative action is used to improve disturbance rejection, it is often overly severe for set point tracking. In particular step response demands may lead to sharp changes in actuator demands and output overshoot. This phenomenon is known as *derivative kick*. One solution is to include a pre-filter. Another is to allow the derivative action to act on the output only (and not the set point). The PID controller becomes

$$U(s) = k_p E(s) + \frac{k_p}{T_i s} E(s) - k_p \frac{T_d s}{1 + \gamma T_d s} Y(s).$$

Similarly it may be advantageous to limit the amount the proportional action acts on the set point, so the controller becomes

$$U(s) = k_p \lambda R(s) - k_p Y(s) + \frac{k_p}{T_i s} E(s) - k_p \frac{T_d s}{1 + \gamma T_d s} Y(s),$$

with λ a tuning parameter, usually between 0 and 1.

Note that the closed-loop poles are not affected by these changes. Also the integral action continues to act on the error.

As an example, consider the plant

$$G(s) = \frac{1}{s^3 + 3s^2 + 3s + 1}.$$

If the controller is a simple gain k_p this is stable in closed-loop for $k_p < 8$ and unstable for $k_p > 8$. When $k_p = 8$ the plant output oscillates with a period of $2\pi/\sqrt(8) \approx 3.6$. This is straightforward to verify either by simulation or by computing the closed-loop response. Thus the ultimate gain is 8 and the ultimate period is $2\pi/\sqrt(3) \approx 3.6$. The Ziegler-Nichols frequency response tuning rules suggest PID gains $k_p = 4.8$, $T_i = 1.8$ and $T_d = 0.45$. Fig 58 shows the outputs and inputs of the plant with a PIDγ controller subject to an output disturbance at $t = 20$, an output disturbance at $t = 40$ and a set point change at $t = 60$. The dashed lines show the signals when the derivative and proportional gains act on the error, while the solid lines show the signals when they act only on the output. The two disturbance responses are identical, but the set point responses are very different.

8.7 Anti-windup

So far we have assumed that the input signal (manipulated variable) has an infinite range. In practice there are always limits to actuator movement—for example a valve may cannot be more than 100% open. If actuators reach their

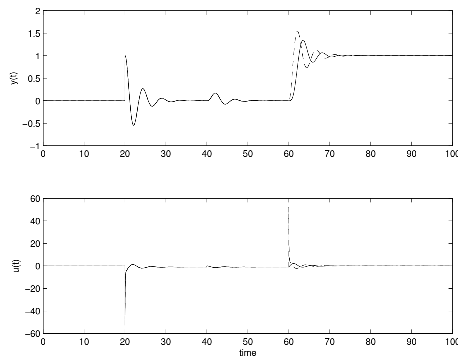


Figure 58: Input and output response of a third order plant with PIDγ control.

limits during normal operation, this causes degradation of performance. The worst case scenario is with a PI or PID controller where the set-point or disturbance is such that the actuator sits on its limit for a period of time without the output reaching its set-point. In this case the error will not go zero, and the I term will continue integrating. Under these circumstances it can take a long time for the control system to recover.

To remedy this, anti-windup structures are implemented. These entail some kind of integrator management (one possibility is simply to switch off the integration when an actuator is at its limit). Fig 59 shows one possible structure that is commonly used. Note that an additional tuning gain T_t is introduced. Fig 60 compares input and output behaviour for the previous example when the actuator is limited to lie between

$$-2 \leq u(t) \leq 2.$$

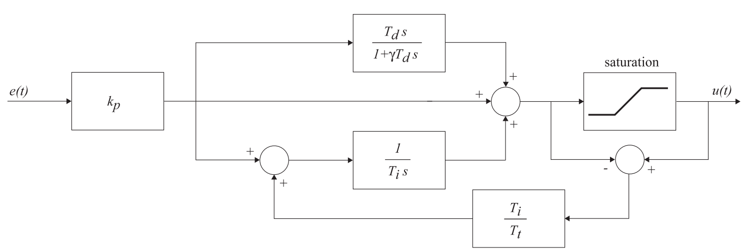


Figure 59: Anti-windup structure for a PID controller.

8.8 Internal model principle

We have seen that integral action guarantees zero steady state error to step changes in the set point (provided the closed-loop response is stable). The

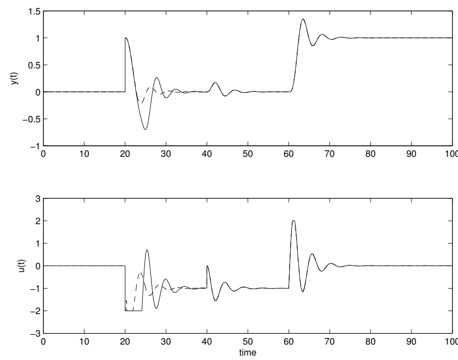


Figure 60: Input and output response of a third order plant with PIDγ. The input is limited to lie between -2 and 2. The solid line shows the response without anti-windup, while the dashed line shows the response when anti-windup is included.

integral action may occur in either the controller $C(s)$ or the plant $G(s)$. Define the *forward loop transfer function* $kL(s)$ as

$$kL(s) = G(s)C(s) = k\frac{n_L(s)}{sd_L(s)},$$

where $n_L(s)$ and $d_L(s)$ are polynomials, with $n_L(0)$ non-zero. (We will find it useful to separate the gain k in later sections.) Then the error response to a unit step is

$$E(s) = R(s) - Y(s) = \left[1 - \frac{kL(s)}{1 + kL(s)}\right]\frac{1}{s} = \frac{1}{1 + kL(s)}\frac{1}{s} = \frac{d_L(s)}{sd_L(s) + kn_L(s)}.$$

We find

$$\lim_{t\to\infty} e(t) = \lim_{s\to 0} sE(s) = \lim_{s\to 0}\frac{sd_L(s)}{sd_L(s) + kn_L(s)} = \frac{0}{kn_L(0)} = 0,$$

as required.

Similarly we find the response to an step output disturbance in steady state is zero: if all other signals entering the loop are zero then

$$Y(s) = \frac{1}{1 + kL(s)}D_o(s) = \frac{1}{1 + kL(s)}\frac{1}{s},$$

and

$$\lim_{t\to\infty} y(t) = \lim_{s\to 0} sY(s) = 0.$$

If we require zero steady state error to *both* input and output step disturbance then we require the integral action to be in the controller: recall that the response to an input disturbance is

$$Y(s) = \frac{G(s)}{1 + G(s)C(s)}D_i(s),$$

while the response to an output disturbance is

$$Y(s) = \frac{1}{1 + G(s)C(s)} D_o(s).$$

1. Suppose the integral action is in the controller, so that

$$C(s) = k\frac{n_c(s)}{sd_c(s)}, \text{ and } G(s) = \frac{n(s)}{d(s)},$$

with $n_c(0)$ non-zero. The response to an input disturbance is

$$Y(s) = \frac{sn(s)d_c(s)}{sd(s)d_c(s) + kn(s)n_c(s)} \frac{1}{s},$$

and

$$\lim_{t\to\infty} y(t) = \lim_{s\to 0} \frac{sn(s)d_c(s)}{sd(s)d_c(s) + kn(s)n_c(s)} = 0.$$

Similarly the response to an output disturbance is

$$Y(s) = \frac{sd(s)d_c(s)}{sd(s)d_c(s) + kn(s)n_c(s)} \frac{1}{s},$$

and

$$\lim_{t\to\infty} y(t) = \lim_{s\to 0} \frac{sd(s)d_c(s)}{sd(s)d_c(s) + kn(s)n_c(s)} = 0.$$

2. Suppose instead the plant is integrating but the controller has no integral action, so that

$$C(s) = k\frac{n_c(s)}{d_c(s)}, \text{ and } G(s) = \frac{n(s)}{sd(s)},$$

with $d_c(s)$ and $n(s)$ non-zero. Then the response to a step input is

$$Y(s) = \frac{n(s)d_c(s)}{sd(s)d_c(s) + kn(s)n_c(s)} \frac{1}{s},$$

and

$$\lim_{t\to\infty} y(t) = \lim_{s\to 0} \frac{n(s)d_c(s)}{sd(s)d_c(s) + kn(s)n_c(s)} \neq 0.$$

This is an example of the *internal model principle.* Suppose we wish to have zero steady state error to a setpoint with Laplace transform

$$R(s) = \frac{n_r(s)}{d_r(s)}.$$

Then $d_r(s)$ must be in the denominator of the forward loop transfer function. Similarly if we have an input disturbance with Laplace transform

$$D_i(s) = \frac{n_i(s)}{d_i(s)},$$

and we want the output to be zero in steady state, then we must have $d_i(s)$ in the denominator of the controller (unless $d_i(s)$ is in the numerator of the plant).

Examples:

1. Suppose we require zero steady state error to a ramp set point

$$r(t) = t,\ R(s) = \frac{1}{s^2}.$$

   Then we require there to be a double integrator in the forward loop transfer function.

2. Suppose a plant is subject to sine wave disturbances with known frequency ω at both input and output. If we want the steady state output to be zero, then we require the controller to take the form

$$C(s) = \frac{1}{s^2+\omega^2}\frac{n(s)}{d(s)},$$

   for some $n(s)$ and $d(s)$.

8.9 Digital PID

Most PID controllers are implemented digitally. The compensator itself can be implemented with just a few lines of software code. Although digital control is beyond the scope of this course, in this section we outline how the ideas of Section 8 translate to digital implementation. Much of the analysis for continuous time control loops carries over at low bandwidths with respect to the sampling frequency.

8.9.1 Digital P control

Digital P-only control can be written simply as

$$u_k = k_p e_k,$$

with

$$e_k = r_k - y_k,$$

and k an integer (the discrete time index). Here we assume that all signals (y, u, r etc.) are sampled with fixed sample time T_s. Thus

$$y_k = y(kT_s) \text{ etc.}$$

The proportional gain k_p is exactly the same as for continuous time P-only control, and has similar effect. High gain control gives a fast response and small steady state error, at the expense of high actuator energy, increased oscillations, noise amplification and a risk of instability. Low gain control is usually safer but gives a sluggish response.

However the forward loop dynamics from u_k to y_k do not necessarily correspond even closely to their continuous time counterparts. Attention must be paid to issues of sampling (extracting y_k from the continuous time signal), modulation (forming a continuous time manipulated variable $u(t)$ from u_k) and any internal filtering and amplification within the digital controller.

As for continuous time control, care must be taken with offset values. Suppose for example the control input is a voltage signal between 0V and 12V so that a sensible offset value is 6V. Then u_k represents deviation from this offset value and the implemented voltage is $(6 + u_k)$V.

8.9.2 Digital PD control

As with continuous control, derivative action can be useful for damping oscillations. But usually the derivative of a signal is no longer available (an exception is when both a signal and its derivative are sensed separately - for example with separate speed and acceleration measurements). It is standard to approximate the derivative with a difference. Recall the definition:

$$\frac{d}{dt}e(t) = \lim_{\delta t \to 0} \frac{e(t+\delta t) - e(t)}{\delta t}.$$

This definition involves future values $e(t+\delta t)$. But we can also write

$$\frac{d}{dt}e(t) = \lim_{\delta t \to 0} \frac{e(t) - e(t-\delta t)}{\delta t}.$$

Hence we can approximate:

$$\frac{d}{dt}e(t) \approx \frac{e(t) - e(t-T_s)}{T_s},$$

so that digital PD control takes the form:

$$u_k = k_p \left(e_k + \frac{T_d}{T_s}(e_k - e_{k-1}) \right).$$

This requires a remembered state e_{k-1}, representing the error one sample ago. Care must be taken to code this correctly.

Note that the contoller is causal (unlike its continuous counterpart), so no additional filtering is strictly necessary. However there remains high gain at high frequency which can amplify noise. Thus an additional low pass filter remains advisable. In some implementation this is in series with the compensator; with others it acts only on the derivative part.

As with continuous PD control, it is better that the derivative action acts only on the measured variable. Hence a better implementation is:

$$u_k = k_p e_k - k_p \frac{T_d}{T_s}(y_k - y_{k-1}).$$

8.9.3 Digital PI control

Integral action is useful for ensuring zero steady state error. This time we make the approximation:

$$\int_{-\infty}^{t} e(t)\, dt \approx \sum_{n=0}^{\infty} T_s e(t - nT_s).$$

Hence a digital PI controller takes the form:

$$u_k = k_p \left(e_k + \frac{T_s}{T_i} \sum_{n=0}^{\infty} e_{k-n} \right).$$

As with PD control we require a remembered state, this time corresponding to the sum of errors. A standard implementation is so-called velocuty form. We find

$$u_k - u_{k-1} = k_p \left(e_k - e_{k-1} + \frac{T_s}{T_i} e_k \right),$$

and hence the implementation

$$u_k = u_{k-1} + k_p \left(e_k - e_{k-1} + \frac{T_s}{T_i} e_k \right).$$

Note that there are two remembered states in this form, u_{k-1} and e_{k-1}. A simple form of anti-windup is to replace the calculated remembered state u_{k-1} with the value of u actually implemented one sample ago.

8.9.4 Digital PID control

As with continuous control, a digital PID controller can be formed by adding proportional, derivative and integral action in parallel. It is often written simply as:

$$u_k = k_p e_k + k_i \sum_{n=0}^{\infty} e_{k-n} + k_d(e_k - e_{k-1}),$$

or in velocity form:

$$u_k = u_{k-1} + k_p(e_k - e_{k-1}) + k_i e_k + k_d(e_k - 2e_{k-1} + e_{k-2}).$$

8.10 Key points

- PID control is an industry standard.
- Proportional control provides the grunt.
- Integral action ensures zero steady state error (under appropriate conditions).
- Derivative action can be used to remove transient oscillations.
- The Ziegler-Nichols rules are a reasonable starting point for tuning.

9 Root locus analysis

9.1 The root locus plot

Consider the closed-loop system depicted in Fig 61. The *root locus* is the map of the closed-loop poles as we change the control gain from zero to infinity (*root* refers to the roots of the closed loop denominator while *locus* is Latin for place). The map is in the complex s-plane.

Since we do not change the structure of the controller in the analysis, it is convenient to define the *forward loop transfer function* $kL(s)$ as

$$kL(s) = C(s)G(s).$$

The root locus maps the poles of the closed-loop transfer function for the system depicted in Fig 62 as we vary k.

We write

$$L(s) = \frac{n_L(s)}{d_L(s)},$$

where $n_L(s)$ and $d_L(s)$ are polynomials, so the closed-loop sensitivities are

$$\begin{aligned} S(s) &= \frac{1}{1+kL(s)} = \frac{d_L(s)}{d_L(s)+kn_L(s)}, \\ T(s) &= \frac{kL(s)}{1+kL(s)} = \frac{kn_L(s)}{d_L(s)+kn_L(s)}. \end{aligned}$$

We are seeking the roots of the polynomial $d_L(s) + kn_L(s)$ as we vary k. The order of $d_L(s) + kn_L(s)$ is the same as the order of $d_L(s)$, so the number of closed-loop poles is equal to the number of open-loop poles.

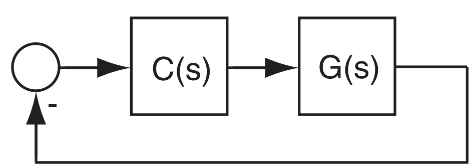


Figure 61: Closed loop system with transfer functions for controller $C(s)$ and plant $G(s)$.

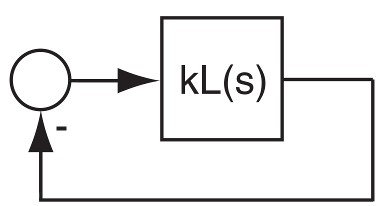


Figure 62: Equivalent closed loop system with forward loop transfer function $kL(s)$.

When k is small, $d_L(s) + kn_L(s) \approx d_L(s)$. We can interpret this as saying the closed-loop poles are very close to the open loop poles. In particular, if the open-loop system is stable, then the closed-loop system with small gain will also be stable. Similarly, if the open-loop system is unstable, then small gain control cannot stabilize the system.

When k is very large, the expression $d_L(s)+kn_L(s)$ is dominated by $kn_L(s)$. That is to say, there are closed-loop poles near the open-loop zeros when k is large. But if the relative degree of the plant $n_p - n_z$ is greater than zero (i.e. if n_p, the order of $d_L(s)$, is greater than n_z, the order of $n_L(s)$) then $d_L(s)+kn_L(s)$ has an additional $n_p - n_z$ poles. These additional poles tend to infinity (not necessarily on the real line) as k tends to infinity.

Traditional classical control textbooks devote considerable attention to drawing root locus diagrams by hand. However, they can easily be generated using computer software. For example in Matlab, given a forward loop transfer function L, the root locus plot can be generated by typing:

```
>>rlocus(L)
```

In this course we will consider the interpretation of such computer generated plots.

We begin my stating the following properties which will be seen in all the examples:

Property 1: the open-loop poles are part of the locus.

Property 2: the open-loop zeros are part of the locus.

Property 3: the location of the roots varies continuously with k. Hence the locus comprises n_p continuous curves (which may intersect) in the complex plane.

Property 4: as $k \to 0$, the curves of the root locus approach the open-loop poles.

Property 5: as $k \to \infty$ the curves of the root locus either approach the open-loop zeros or go to infinity.

9.2 Root locus examples

Example 1: $G(s) = \frac{1}{s-1}$, $C(s) = k$.

We have

$$L(s) = \frac{1}{s-1}.$$

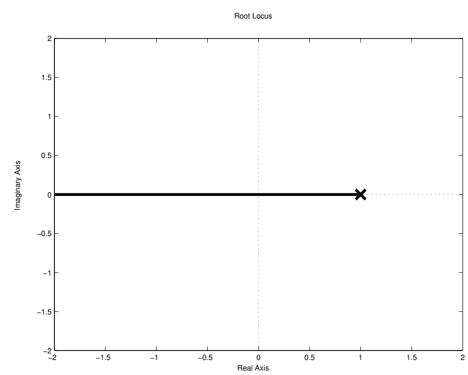


Figure 63: Root locus for example 1. The plot is obtained by using the "rlocus" command in Matlab.

Interpretation

In this case it is easy to calculate the closed-loop poles by hand. The complementary sensitivity is:

$$T(s) = \frac{kL(s)}{1 + kL(s)},$$
$$= \frac{k}{s + k - 1}.$$

The plant is open loop unstable with a pole at +1. The closed-loop pole is at $1 - k$. When $k \approx 0$ the pole is near +1 and so for small gain k the closed-loop system is also unstable. As we increase k the closed-loop pole moves to the left. When $k = 1$ the closed-loop pole is at 0 (i.e. marginally stable). When $k > 1$ the closed-loop system is stable. As the gain is further increased, the closed-loop system becomes faster.

Example 2: $G(s) = \dfrac{1}{s^2 + s - 2}$, $C(s) = k$.

We have

$$L(s) = \frac{1}{s^2 + s - 2} = \frac{1}{(s+2)(s-1)}.$$

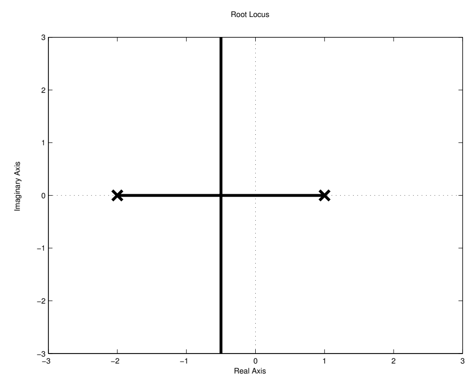


Figure 64: Root locus for example 2.

Interpretation

The complementary sensitivity is:

$$\begin{aligned} T(s) &= \frac{kL(s)}{1 + kL(s)}, \\ &= \frac{k}{s^2 + s + k - 2}. \end{aligned}$$

The plant is open loop unstable. For small gain k the closed-loop system is also unstable, with closed-loop poles close to the open-loop poles. When $k = 2$ the closed-loop denominator is $s^2 + s$ with poles at 0 and -1. When $k > 2$ the closed-loop system is stable. When $k = 2.25$ the closed-loop denominator is $(s + 0.5)^2$ so the two closed-loop poles coincide. For $k > 2.25$ the closed-loop poles form a complex conjugate pair, whose imaginary part increases as k increases. For high gain the closed-loop system becomes highly oscillatory.

Example 3: $G(s) = \dfrac{1}{s^2 + s + 1}$, $C(s) = k$.

We have

$$L(s) = \frac{1}{s^2 + s + 1} = \frac{1}{(s + 1/2 - j\sqrt(3)/2)(s + 1/2 + j\sqrt(3)/2)}.$$

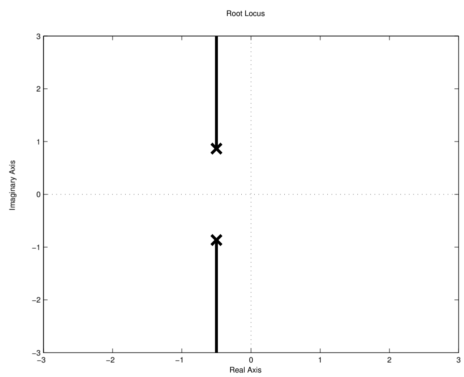


Figure 65: Root locus for example 3.

Interpretation

We see that the locus of Example 3 is a subset of the locus of Example 2. As the gain is increased, the closed-loop response becomes more oscillatory.

Example 4: $G(s) = \dfrac{1}{s+1}$, $C(s) = \dfrac{s+2}{s}$.

We have

$$L(s) = \frac{s+2}{s(s+1)}.$$

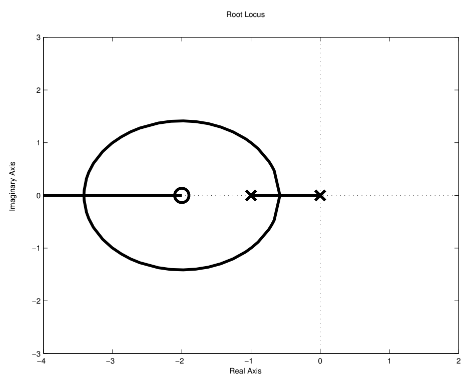


Figure 66: Root locus for example 4.

Interpretation

The open loop system has a pole at the origin and is marginally stable. For any value of $k > 0$ the closed-loop system is stable. As k is increased the closed-loop poles become a complex conjugate pair. However the presence of the zero stops the damping ratio from dropping too low. For high values of gain k both poles return to the real line; one tends to the zero at -2 while the other tends to $-\infty$.

Example 5: $G(s) = \dfrac{1}{s^2 + s - 2}$, $C(s) = (s + 1)$.

Note that the PD controller $C(s)$ is not proper for this example. We have

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

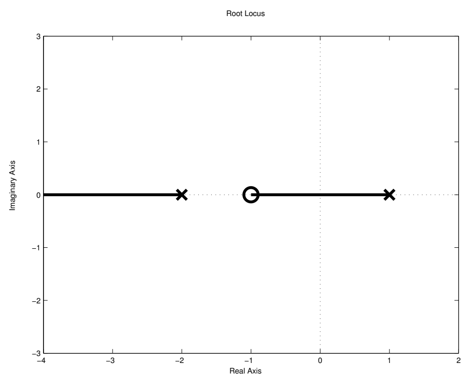


Figure 67: Root locus for example 5.

Interpretation

The plant is unstable in open loop. The closed-loop system is stable for $k > 2$. Both poles are real for all values of $k > 0$.

Example 6: $G(s) = \dfrac{1}{(s+1)(s^2+2s+2)}$, $C(s) = k$.

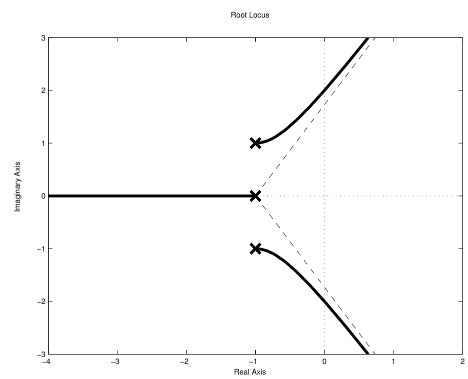


Figure 68: Root locus for example 6.

Interpretation

The relative degree is 3. All three poles are stable in open loop. As k is increased the real pole becomes faster (moves to the left). However the two complex poles move to the right in a direction that reduces the damping ratio. For k sufficiently large the poles move to the right half plane so the closed-loop system becomes unstable. In the next chapter we will see how to find the transition value of k via the Nyquist criterion. It is also possible to calculate this value of k by hand. The complementary sensitivity is:

$$\begin{aligned} T(s) &= \frac{kL(s)}{1+kL(s)}, \\ &= \frac{k}{(s+1)(s^2+2s+2)+k}, \\ &= \frac{k}{s^3+3s^2+4s+2+k}. \end{aligned}$$

The transition value of k is the value where the closed-loop denominator has poles on the imaginary axis—i.e. where there is a factor $(s^2+\omega^2)$ for some ω in the closed-loop denominator. Hence we need to find k such that, for some α and ω:

$$s^3+3s^2+4s+2+k = (s+\alpha)(s^2+\omega^2).$$

Equating coefficients gives:

$$\begin{aligned} \alpha &= 3, \\ \omega^2 &= 4, \\ \alpha\omega^2 &= 2 + k. \end{aligned}$$

Hence the transition value is $k = 10$. When $k > 10$ the closed-loop system becomes unstable.

Example 7: $G(s) = \dfrac{1}{(s-1)(s^2+4s+8)}$, $C(s) = k$.

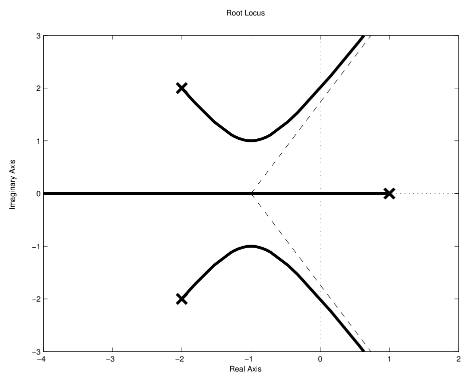


Figure 69: Root locus for example 7.

Interpretation

The relative degree is 3. The plant is unstable in open loop, so the closed-loop system is also unstable for small k. As k is increased the real pole moves to the left, and there is some value k for which it moves into the left half plane. However the two complex conjugate poles move to the right as k increases and there is also a transition value where they move into the right half plane.

The complementary sensitivity is:

$$\begin{aligned} T(s) &= \frac{kL(s)}{1+kL(s)}, \\ &= \frac{k}{(s-1)(s^2+4s+8)+k}, \\ &= \frac{k}{s^3+3s^2+4s+k-8}. \end{aligned}$$

It is easy to see the transition value for the real pole is $k = 8$. The transition value for the complex conjugate poles is where

$$s^3+3s^2+4s+k-8 = (s+\alpha)(s^2+\omega^2),$$

for some α and ω. Equating coefficients gives:

$$\begin{aligned} \alpha &= 3, \\ \omega^2 &= 4, \\ \alpha\omega^2 &= k-8. \end{aligned}$$

Hence the transition value is $k = 20$. We conclude that the closed-loop system is stable for $8 < k < 20$.

9.3 Grid lines

If grid lines are included in a root locus plot, it is conventional to show lines of constant frequency and damping ratio for second order systems corresponding to pairs of conjugate poles. Grid lines for Example 7 are shown in Fig 70. Note in particular for this example that for small gain the closed-loop damping ratio increases as we increase the gain. This effect is reversed for higher gains.

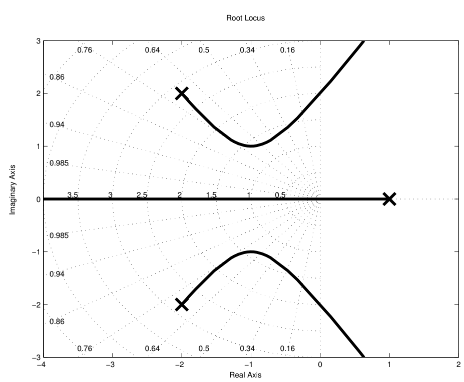


Figure 70: Root locus for example 7 with grid lines.

9.4 Key points

- The root locus plot gives a map of closed-loop poles as the feedback gain changes.
- At low gain the poles are near the open-loop poles.
- At high gain the poles are either near the open-loop zeros or near infinity.

10 The Nyquist criterion

10.1 Introduction

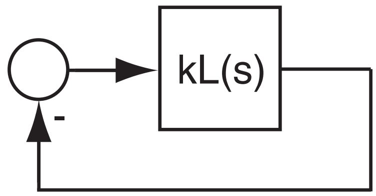


Figure 71: Closed loop system with forward loop transfer function $kL(s)$.

We have seen that the frequency response, encapsulated by the Bode diagram, provides useful information about a system's behaviour. For example, peaks in the gain of the sensitivity $S(j\omega)$ or complementary sensitivity $T(j\omega)$ are indicative of poor closed-loop behaviour. In this section we will examine how the closed-loop response can be *inferred* from the frequency response of the forward loop transfer function $kL(j\omega)$. Recall that we have the relations

$$S(j\omega) = \frac{1}{1 + kL(j\omega)} \text{ and } T(j\omega) = \frac{kL(j\omega)}{1 + kL(j\omega)}.$$

We will find it useful to represent the frequency response of $kL(j\omega)$ with the *Nyquist plot*. This plots the imaginary part of $kL(j\omega)$ against the real part of $kL(j\omega)$. We will then see how closed-loop performance and closed-loop stability can be inferred from this plot.

10.2 The Nyquist plot

As stated above, the Nyquist plot shows $\text{Im}\,[kL(j\omega)]$ against $\text{Re}\,[kL(j\omega)]$.

Example 1: $kL(s) = \dfrac{k}{s+1}$.

We have

$$\begin{aligned} kL(j\omega) &= \frac{k}{1 + j\omega}, \\ &= k\left(\frac{1}{1 + j\omega}\right)\left(\frac{1 - j\omega}{1 - j\omega}\right), \\ &= k\frac{1 - j\omega}{1 + \omega^2}. \end{aligned}$$

So

$$\begin{aligned} \text{Re}\,[kL(j\omega)] &= \frac{k}{1 + \omega^2}, \\ \text{Im}\,[kL(j\omega)] &= \frac{-k\omega}{1 + \omega^2}. \end{aligned}$$

We can quickly check the following values:

ω	$\mathrm{Re}\left[kL(j\omega)\right]$	$\mathrm{Im}\left[kL(j\omega)\right]$
0	k	0
1	$k/2$	$-k/2$
$\to \infty$	$\to 0$	$\to 0$

It would be straightforward to include more values in the table, either by calculation or by reading them from the Bode plot. In fact the Nyquist plot shows the same information as a Bode plot, except for the dependence on frequency. For this particular example, it is possible to find the curve analytically. We find

$$\begin{aligned}
&\left(\mathrm{Re}\left[kL(j\omega)\right] - \frac{k}{2}\right)^2 + \mathrm{Im}\left[kL(j\omega)\right]^2 \\
&\qquad = k^2\left(\frac{1}{1+\omega^2} - \frac{1}{2}\right)^2 + k^2\left(\frac{-\omega}{1+\omega^2}\right)^2, \\
&\qquad = k^2\frac{\left(2-\left(1+\omega^2\right)\right)^2 + \left(2\omega\right)^2}{4\left(1+\omega^2\right)^2}, \\
&\qquad = k^2\frac{1-2\omega^2+\omega^4+4\omega^2}{4\left(1+\omega^2\right)^2}, \\
&\qquad = k^2\frac{1+2\omega^2+\omega^4}{4\left(1+\omega^2\right)^2}, \\
&\qquad = \frac{k^2}{4}.
\end{aligned}$$

So the Nyquist plot lies on the circle radius $k/2$, centred on $+k/2$. We see from the tabulated values that with $\omega \geq 0$ it forms the semicircle whose imaginary part is less than or equal to zero. See Fig 72.

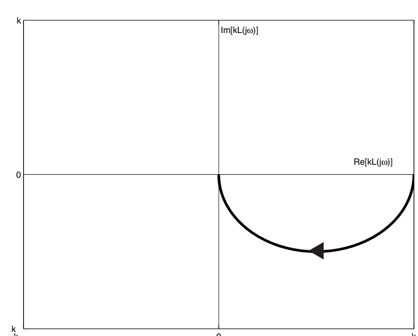


Figure 72: Nyquist plot for Example 1.

Example 2: $kL(s) = \dfrac{k}{s(s+1)}$.

We have

$$kL(j\omega) = \frac{k}{(j\omega)(1+j\omega)},$$
$$= \frac{-jk(1-j\omega)}{\omega(1+\omega^2)}.$$

So

$$\mathrm{Re}\,[kL(j\omega)] = \frac{-k}{1+\omega^2},$$
$$\mathrm{Im}\,[kL(j\omega)] = \frac{-k}{\omega(1+\omega^2)}.$$

Once again we can quickly check the following values:

ω	$\mathrm{Re}\,[kL(j\omega)]$	$\mathrm{Im}\,[kL(j\omega)]$
$\to 0$	$\to -k$	$\to -\infty$
1	$-k/2$	$-k/2$
$\to \infty$	$\to 0$	$\to 0$

The curve for positive values of ω is shown in Fig 73.

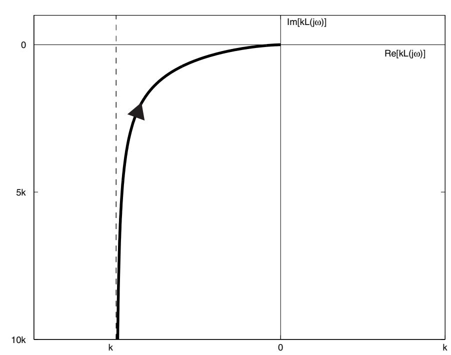


Figure 73: Nyquist plot for Example 2.

Example 3: $kL(s) = \frac{k}{(s+1)^3}$.

We have

$$kL(j\omega) = \frac{k}{-j\omega^3 - 3\omega^2 + 3j\omega + 1}.$$

So we can quickly check the following values:

ω	$kL(j\omega)$	$\mathrm{Re}\,[kL(j\omega)]$	$\mathrm{Im}\,[kL(j\omega)]$
0	k	k	0
1	k/(-2+2j)	$-k/4$	$-k/4$
$\to\infty$	0	$\to 0$	$\to 0$

We know from the Bode plot that the phase drops to -270^o for high frequency, so the curve must cross both the real and the imaginary axes. We can find these values by equating coefficients in our original expression for $kL(j\omega)$.

- $\mathrm{Im}\,[kL(j\omega)] = 0$ when $-\omega^3 + 3\omega = 0$. I.e. when $\omega^2 = 3$. At this frequency we find

$$kL(j\omega) = \frac{k}{-3\omega^2 + 1} = -\frac{k}{8}.$$

- $\mathrm{Re}\,[kL(j\omega)] = 0$ when $-3\omega^2 + 1 = 0$. I.e. when $\omega^2 = 1/3$. At this frequency we find

$$kL(j\omega) = \frac{k}{-j\omega^3 + 3j\omega} = -j\frac{k3\sqrt{3}}{8}.$$

The curve for positive values of ω is shown in Fig 74.

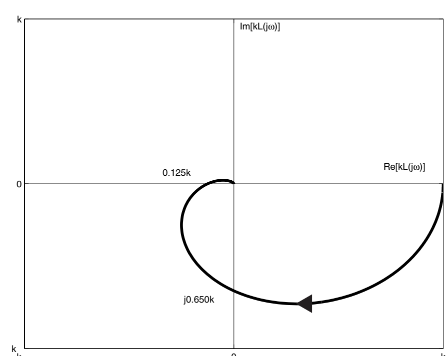


Figure 74: Nyquist plot for Example 3.

Example 4: $kL(s) = \dfrac{k}{(s-1)(s^2+4s+8)}$.

We have

$$\begin{aligned} kL(j\omega) &= \frac{k}{(j\omega-1)(-\omega^2+4j\omega+8)}, \\ &= \frac{k(-j\omega-1)(-\omega^2+8-4j\omega)}{(1+\omega^2)(64+\omega^4)}, \\ &= \frac{k(-8-3\omega^2)}{(1+\omega^2)(64+\omega^4)} + j\frac{k(-4\omega+\omega^3)}{(1+\omega^2)(64+\omega^4)}. \end{aligned}$$

The real part is always negative for $\omega > 0$, but tends to zero as $\omega \to \infty$. The imaginary part is zero when $\omega = 0$, $\omega = 2$ and as $\omega \to \infty$. Hence we can tabulate the following values:

ω	$\mathrm{Re}\,[kL(j\omega)]$	$\mathrm{Im}\,[kL(j\omega)]$
0	$-0.125k$	0
1	$-0.0846k$	$-0.2031k$
2	$-0.05k$	0
4	$-0.0103k$	$0.0088k$
$\to \infty$	$\to 0$	$\to 0$

The curve for positive values of ω is shown in Fig 75.

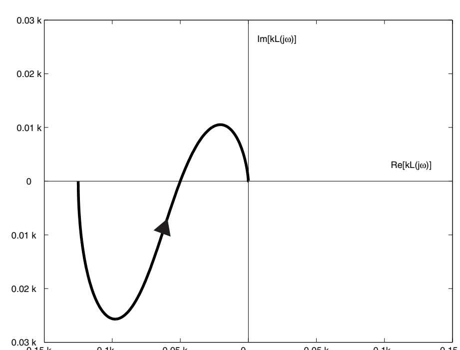


Figure 75: Nyquist plot for Example 4.

10.3 The −1 point

The reason we draw the Nyquist plot of the forward loop transfer function $kL(j\omega)$ is to infer the behaviour of the closed-loop system: in particular the sensitivity $S(j\omega)$ and complementary sensitivity $T(j\omega)$.

If the Nyquist plot of $kL(j\omega)$ passes through -1 for some frequency ω, then at that frequency we have

$$1 + kL(j\omega) = 0.$$

Hence the sensitivity and complementary sensitivity will be infinite at that frequency. Similarly if $kL(j\omega)$ passes near -1 then the sensitivity and complementary sensitivity will be large at the corresponding frequency. This is clearly undesirable.

Also, if the closed-loop sensitivities are infinite at a certain frequency ω, this corresponds to a closed-loop pole at $j\omega$. This in turn corresponds to the root-locus passing through the imaginary axis. Hence if the gain is varied, the value k is a candidate for the transition between closed-loop stability and instability.

Example: consider Example 4 from the previous section, which corresponds to Example 7 from the root locus section. The Nyquist curve crosses the real axis at 0 (as $\omega \to \infty$), at $-k/8$ (when $\omega = 0$) and at $-k/20$ (when $\omega = 2$). Thus if we vary k the Nyquist plot passes through -1 when $k = 8$ and $k = 20$. We saw in the root locus section that the example is stable for $8 < k < 20$.

10.4 The Nyquist criterion

The Nyquist criterion allows us to assess whether the closed-loop system is stable directly from the Nyquist diagram of the forward loop transfer function $kL(j\omega)$. It is easily stated, but first we need to extend the Nyquist diagram to include negative frequencies. We exploit the relations

$$\begin{aligned} L(-j\omega) &= L(j\omega)^* \text{ where * denotes complex conjugate,} \\ \mathrm{Re}\left[L(-j\omega)\right] &= \mathrm{Re}\left[L(j\omega)\right], \\ \mathrm{Im}\left[L(-j\omega)\right] &= -\mathrm{Im}\left[L(j\omega)\right]. \end{aligned}$$

Figures 76 to 79 show the extended Nyquist diagrams for Examples 1 to 4 with $k = 1$ plotted using the Matlab function `>>nyquist`. We have retained the automatic scaling from Matlab. It is often useful to re-scale the axes.

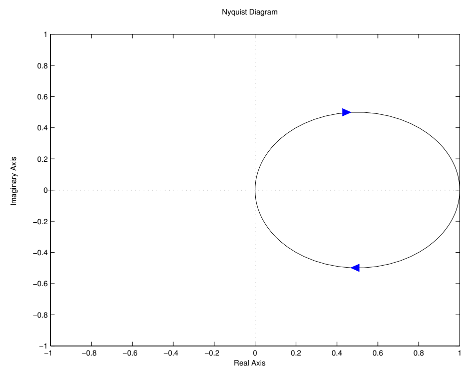


Figure 76: Extended Nyquist plot for Example 1 with $k = 1$

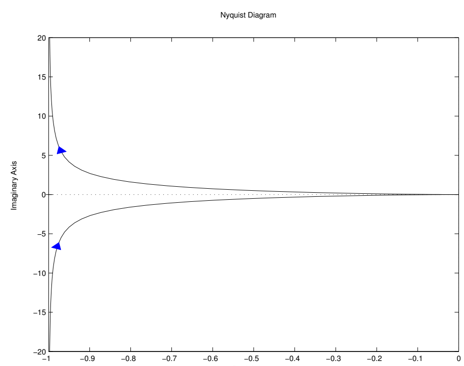


Figure 77: Extended Nyquist plot for Example 2 with $k = 1$.

In Examples 1, 3 and 4 the extended Nyquist curve forms a continuous loop. The Nyquist criterion can then be stated:

Let Z be the number of closed-loop unstable poles.
Let P be the number of open-loop unstable poles.
Let N be the number of times the Nyquist curve of kL encircles the -1 point clockwise.
Then
$$Z = N + P.$$

Example 1: $kL(s) = \dfrac{k}{s+1}$.

The forward loop transfer function $kL(s)$ has one stable pole and no unstable poles, so $P = 0$. From Fig 76 we see that the Nyquist curve never

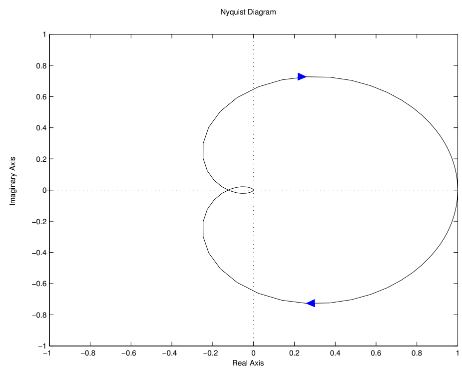


Figure 78: Extended Nyquist plot for Example 3 with $k = 1$.

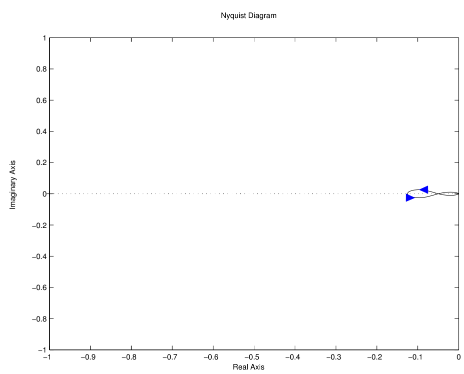


Figure 79: Extended Nyquist plot for Example 4 with $k = 1$.

encircles -1 for $k > 0$. Hence $N = 0$ for $k > 0$. The Nyquist criterion states that $Z = 0$ for $k > 0$—i.e. the system is stable in closed loop for any positive k.

Example 3: $kL(s) = \dfrac{k}{(s+1)^3}$.

The forward loop transfer function $kL(s)$ has three stable poles and no unstable poles, so $P = 0$. From Fig 78 we see that the Nyquist curve does not encircle -1 for k positive and $k < 8$. Hence $N = 0$ for $0 < k < 8$. In this case the Nyquist criterion states that $Z = 0$—i.e. the system is stable in closed loop for $0 < k < 8$.

But when $k = 8$ the curve passes exactly through the -1 point. Fig 80 shows the curve with $k = 20$. For $k > 8$ the curve encircles the -1 point twice clockwise, so $N = 2$. As P remains 0 we have $Z = 2$. So for $k > 8$ the system is unstable in closed-loop.

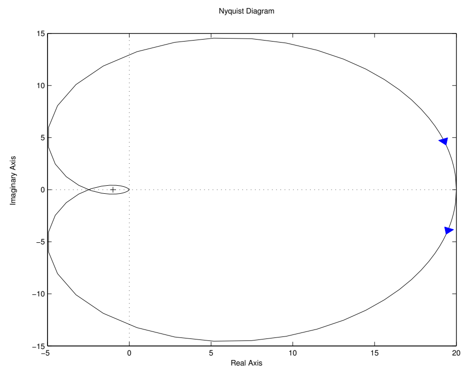


Figure 80: Extended Nyquist plot for Example 3 with $k = 20$.

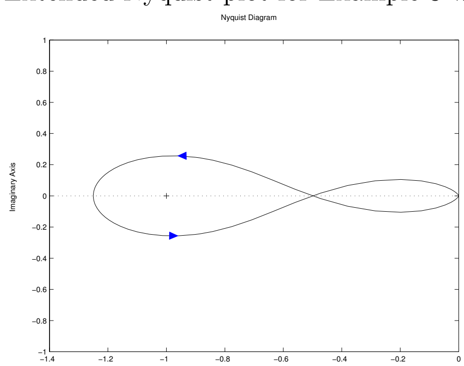


Figure 81: Extended Nyquist plot for Example 4 with $k = 10$.

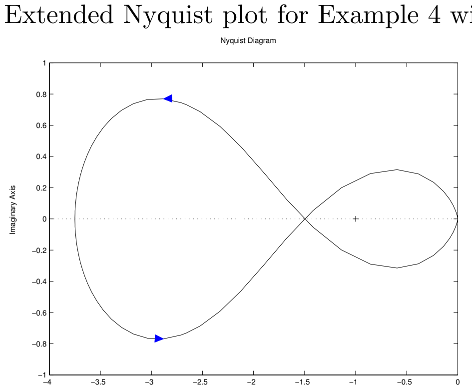


Figure 82: Extended Nyquist plot for Example 4 with $k = 30$.

Example 4: $kL(s) = \dfrac{k}{(s-1)(s^2+4s+8)}$.

The forward loop transfer function $kL(s)$ has two stable poles and one unstable pole so $P = 1$. We know the curve crosses the real axis at -1 when $k = 8$ and when $k = 20$.

When $k < 8$, we see from Fig 79 that the curve does not encircle the -1 point, so $N = 0$. Hence $Z = 1$ and the system is unstable in closed-loop.

Fig 81 shows the curve when $k = 10$. For $8 < k < 20$ the curve encircles -1 once anti-clockwise so $N = -1$. Hence $Z = 0$ and the system is stable in closed-loop.

Fig 82 shows the curve when $k = 30$. For $20 < k$ the curve encircles -1 once clockwise so $N = 1$. Hence $Z = 2$ and the system is unstable in closed-loop.

We have not yet considered Example 2. With reference to Fig 77 we see that the extended Nyquist curve does not form a closed loop. This is because there is an open-loop pole at 0. More generally, any open-loop pole on the imaginary axis will result in a discontinuity in the Nyquist curve.

We can still apply the Nyquist criterion, by *deeming* the pole at $s = 0$ to be in the left half plane. For this example there is one other stable pole, so P is *deemed* to be 0. Then, rather than trace the Nyquist curve along all frequencies from $-\infty$ to $+\infty$, we take a small detour to the right around the pole at $s = 0$, with a semicircle of infinitesimal radius ε (Fig 83).

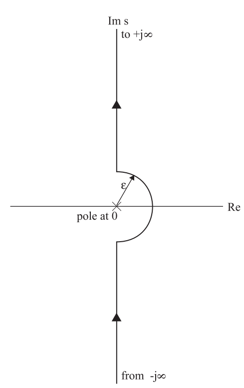


Figure 83: Path in the s-plane along which we evaluate $kL(s)$ for Example 2.

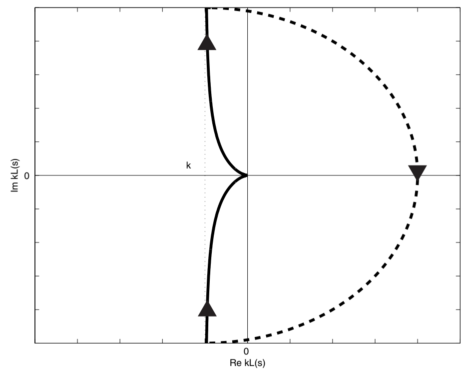


Figure 84: Values of $kL(s)$ for Example 2.

The corresponding Nyquist curve of $kL(s)$ then includes a semicircle of infinite radius from $j\infty$ to $-j\infty$. Since $kL(\varepsilon)$ is positive, the semicircle must be around the right half plane (Fig 84). Hence the Nyquist curve does not encircle -1, so $N = 0$. We conclude that $Z = 0$ for this case, and the closed-loop system is stable for all positive values of k.

10.5 Stability margins

If the Nyquist plot of the forward loop transfer function $kL(s)$ passes near the -1 point, then the closed-loop sensitivity $S(s)$ and complementary sensitivity $T(s)$ will be large at the corresponding frequency. For this reason, compensators $C(s)$ are *designed* so that $kL(s) = C(s)G(s)$ are an acceptable distance from -1. The measure of distance is called a *stability margin.*

In modern control, the usual measure is the minimum radius r of a circle centred at -1 touching $kL(j\omega)$ (see Fig 85). We find

$$\begin{aligned} r &= \min_{\omega} \left|(-1) - kL(j\omega)\right|, \\ &= \min_{\omega} \left|1 + kL(j\omega)\right|, \\ &= \min_{\omega} \left|\frac{1}{S(j\omega)}\right|, \\ &= \frac{1}{\max_{\omega} \left|S(j\omega)\right|}. \end{aligned}$$

The quantity $\max_{\omega} |S(j\omega)|$ is termed the *infinity norm* of $S(s)$, written $\|S\|_{\infty}$, so

$$r = \|S\|_{\infty}^{-1}.$$

In classical control there are two standard stability margins, the *gain margin* and the *phase margin.*

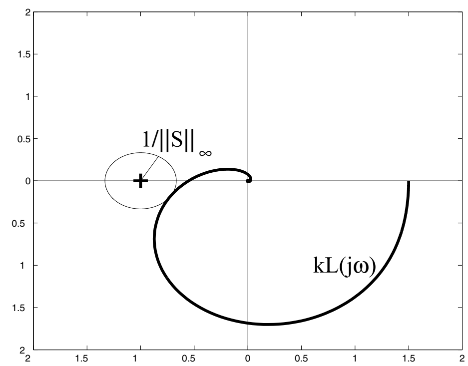


Figure 85: Infinity norm of the sensitivity as a stability margin.

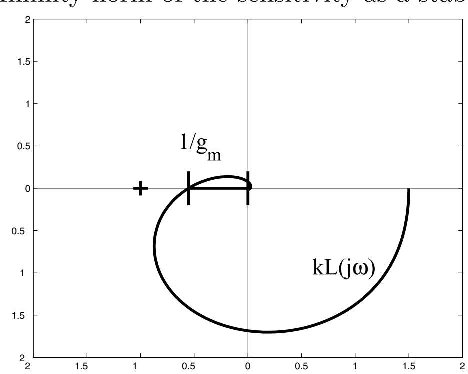


Figure 86: Gain margin.

Gain margin: If $kL(j\omega)$ crosses the negative real axis between 0 and -1 at $1/g_m$, then g_m is the gain margin. It is the amount of extra gain required for $g_m kL(k\omega)$ to pass through -1. It is usually measured in decibels. See Fig 86.

Phase margin: The phase margin ϕ_m is measured where $kL(j\omega)$ crosses a circle with unit radius centred at the origin. It is the angle between $L(j\omega)$ at this point and the negative real axis. It can be interpreted as the amount of phase change required for $kL(j\omega)$ to pass through -1. See Fig 87.

The gain and phase margins are popular because they can be read directly from the Bode plot of $kL(j\omega)$. Specifically, the gain margin is the gain below 0dB at the frequency where the phase drops to -180^o. Similarly the phase margin is the phase above -180^o at the frequency where the gain drops to 0dB. Gain and phase margins for the Example 3 with $k = 4$ are shown on a Nyquist plot in Fig 88 and on a Bode plot in Fig 89. The frequency at which the gain drops to 0dB is termed the *crossover frequency* ω_c. Note that 6dB approximates

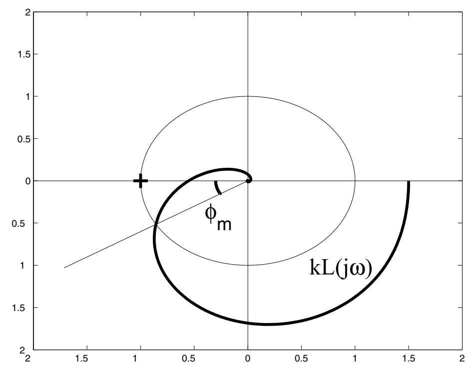


Figure 87: Phase margin.

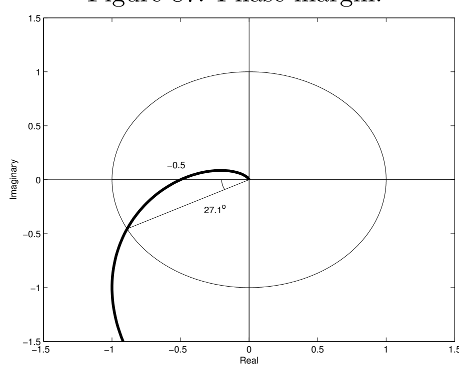


Figure 88: Gain and phase margins for Example 3 with $k = 4$ on a Nyquist plot of $kL(j\omega)$.

to a gain of 2 since

$$20 \log_{10} 2 \approx 6.02.$$

10.6 Grid lines

The Nyquist plot shows the frequency response corresponding to the open-loop forward loop transfer function. In the last section we saw that it is easy to infer the corresponding closed-loop sensitivity. Specifically, contours of constant sensitivity magnitude form circles centred on the -1 point (see Fig 85). It is useful to include contours of constant complementary sensitivity magnitude. In the jargon these are called *M-circles*, and are given as follows:

The M-circle corresponding to gain g is a circle centred at $g^2/(1-g^2)$ with radius $|g/(1-g^2)|$. There is a bifurcation when $g = 1$ (or 0dB) where the radius goes to infinity and the M-circle for 0dB is in fact a vertical straight line intersecting $-1/2$ on the real line.

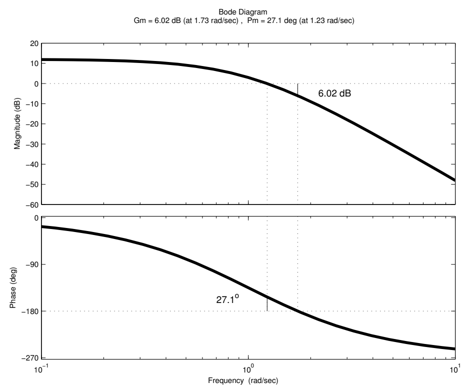


Figure 89: Gain and phase margins for Example 3 with $k = 4$ on a Bode plot of $kL(j\omega)$.

It is conventional to show M-circles as grid lines on a Nyquist plot. Fig 90 shows the Nyquist plot for Example 3 with gain $k = 4$ and with grid-lines shown. The figure is focused on the region near -1.

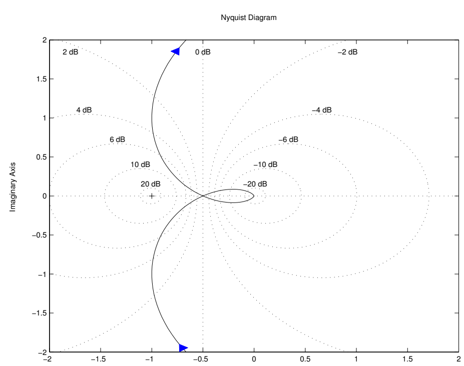


Figure 90: Nyquist plot for Example 3 with $k = 4$ and with M-circles shown.

10.7 Nichols charts

The Nyquist plot can be read (almost) directly from a Bode plot, and carries the advantage that the magnitude of closed-loop sensitivity and complementary sensitivity can be easily inferred from it. However it loses the neat scaling of a Bode plot whose gain is drawn on a logarithmic scale. In classical control the Nichols chart combines many of the advantages of Nyquist and Bode plots. The Nichols chart is a graph of log magnitude against phase. The standard convention is to draw M-circles and N-circles as grid lines. In the previous section we introduced M-circles as contours of constant magnitude of the complementary sensitivity. N-circles are contours of constant phase of the complementary sensitivity. Note that neither appear as actual circles on the Nichols cahrt. It is arguable that it would be more useful to show contours of constant magnitude of the sensitivity in place of N-circles. Fig 91 shows the Nichols chart for Example 2 (with $k = 1$) and Fig 92 shows the Nichols chart for Example 3 (with $k = 4$).

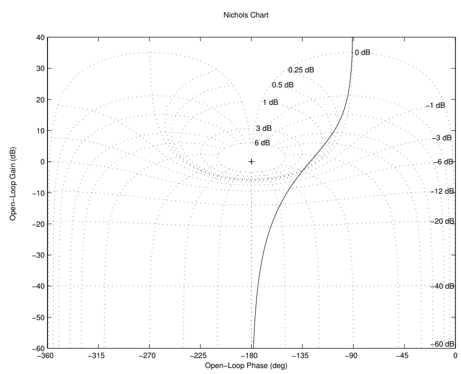


Figure 91: Nichols chart for Example 2 with $k = 1$.

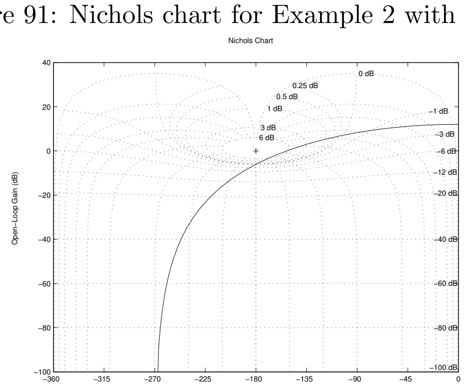


Figure 92: Nichols chart for Example 3 with $k = 4$.

10.8 Justification of the Nyquist criterion

The Nyquist criterion can be derived from *Cauchy's principle of the argument.* This concerns how the argument (angle) of a transfer function $H(s)$ changes around a simple closed curve P in the s-plane. Let $\Delta_P H$ denote the change in argument as we move around P in a clockwise direction.

Example: Suppose $H(s) = s$ and P does not enclose the origin. Then $\Delta_P H = 0$. But suppose P encloses the origin. Then $\Delta_P H = -360^o$ (NB phase is measured in an *anti*-clockwise direction). It is tautologous to say that $H(s)$ encircles the origin if and only if P also encircles the origin. See Figs 93 and 94.

Example: Suppose $H(s) = s - z$ for some zero z. Then $H(s)$ encircles the origin if and only if P encloses z (see Fig 95).

For more general transfer functions, we can exploit relations of the form

$$\angle\left(\frac{s-z}{(s-p_1)(s-p_2)}\right) = \angle(s-z) - \angle(s-p_1) - \angle(s-p_2).$$

Hence we may say:

Suppose a transfer function $H(s)$ has n_p poles and n_z zeros within a simple closed curve P. Then

$$\Delta_P H = -(n_z - n_p) \times 360^o,$$

and $H(s)$ encircles the origin $n_z - n_p$ times when evaluated around P.

For the Nyquist criterion, we let P be the *Nyquist path* which encloses the whole of the right half plane. Thus the Nyquist path consists of the imaginary axis together with a semicircular sweep to the right of the imaginary axis with infinite radius (see Fig 96). We now apply the principle of the argument to $H(s) = 1 + kL(s)$. The poles of $H(s)$ are also the poles of $kL(s)$, i.e. the open-loop poles. The zeros of $H(s)$ are the closed-loop poles. Thus P corresponds to the number of poles of $H(s)$ enclosed by P, while Z corresponds to the number of zeros of $H(s)$ enclosed by P. So $H(s)$ encircles the origin $P - Z$ times, clockwise. But this corresponds to the number of times $kL(s) = H(s)-1$ encircles the -1 point. Hence $N = Z - P$. This is the Nyquist criterion.

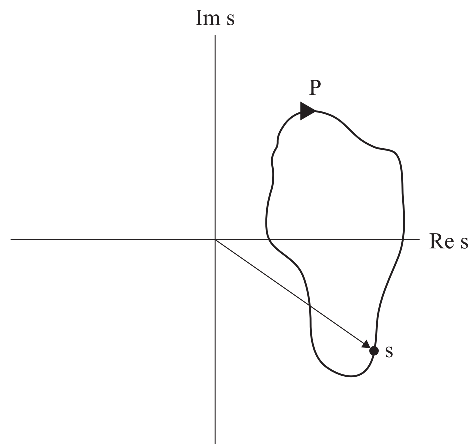


Figure 93: Path P and $H(s) = s$, neither encircling the origin.

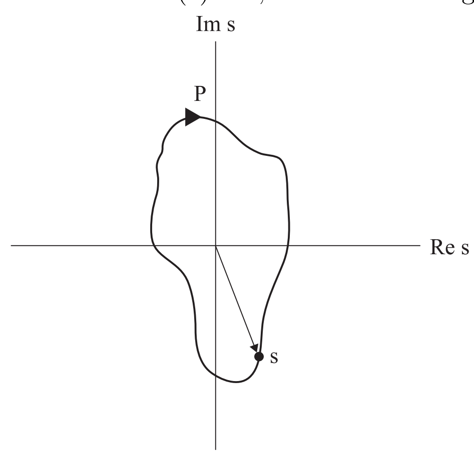


Figure 94: Path P and $H(s) = s$, both encircling the origin.

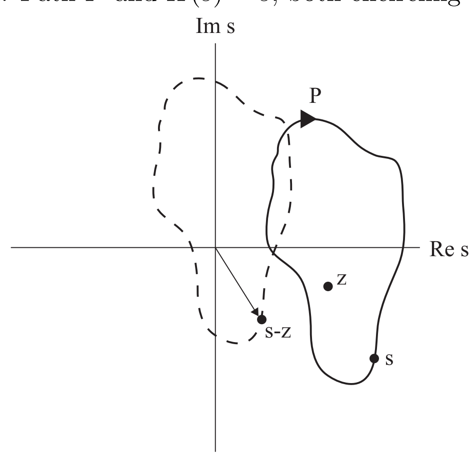


Figure 95: Path P and $H(s) = s - z$, with P encircling z.

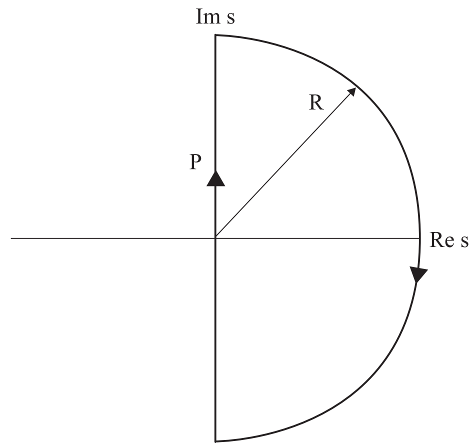


Figure 96: P corresponds to the Nyquist curve as $R \to \infty$.

10.9 Key points

- Closed-loop properties can be inferred from the Nyquist plot of the forward loop transfer function.
- The -1 point is crucial in the analysis.
- The Nyquist criterion determines the stability of the closed-loop.
- Gain and phase margins are measures of distance from the -1 point.

11 Design of compensators

11.1 Design considerations and constraints

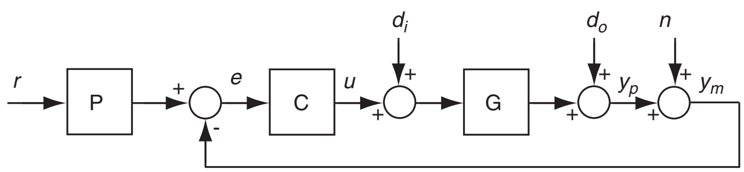


Figure 97: Standard closed loop system.

Recall our standard closed-loop configuration. The Laplace transform of the plant output is

$$Y_p(s) = T(s)P(s)R(s) - T(s)N(s) + G(s)S(s)D_i(s) + S(s)D_o(s),$$

where R represents the reference signal, N the sensor noise, D_i the input disturbance and D_o the output disturbance. Here $S(s)$ and $T(s)$ are respectively the closed-loop sensitivity and complementary sensitivity given, as before, by

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

with

$$kL(s) = G(s)C(s).$$

We have the following mutually contradictory requirements:

- Set point tracking: $T(s)P(s) \approx 1$.
- Noise suppression: $T(s) \approx 0$.
- Disturbance rejection: $S(s) \approx 0$.

In particular we are constrained by the identity

$$S(j\omega) + T(j\omega) = 1 \text{ at all frequencies},$$

so we cannot set both $S(j\omega)$ and $T(j\omega)$ to zero at the same frequency.

The usual design compromise is to choose

$$S(j\omega) \approx 0 \text{ and } T(j\omega) \approx 1 \text{ at low frequencies},$$
$$S(j\omega) \approx 1 \text{ and } T(j\omega) \approx 0 \text{ at high frequencies}.$$

Typical (good) values of S and T are shown in Fig 98. The closed-loop bandwidth ω_b is shown, usually defined as the frequency at which the complementary sensitivity drops below -3dB. Achieving such a closed-loop response

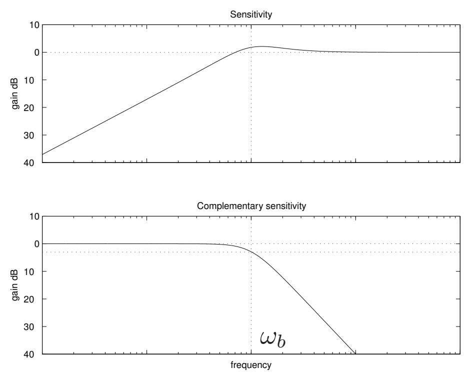


Figure 98: Well-designed closed-loop sensitivity and complementary sensitivity. The complementary sensitivity rolls off to zero at high frequency, while the sensitivity rolls off to zero at low frequency. Neither peaks badly at mid-frequencies.

corresponds to choosing $|kL(j\omega)|$ to be big at low frequencies and small at high frequencies. Care must be taken at the cross-over frequency ω_c. The cross-over frequency is usually near to, but a bit smaller than, the closed-loop bandwidth.

Varying k allows us to choose any cross-over frequency we like. However, varying k has no effect on the phase, so the phase margin at our chosen k may be arbitrarily small or even negative (leading to an unstable closed-loop system). We need to ensure the phase margin is sufficiently large. In terms of performance the following rule of thumb is sometimes useful. The closed-loop response may often be approximated as a second order transfer function with damping ratio

$$\zeta \approx \frac{p_m}{100}.$$

Thus for a closed-loop damping ratio of 0.5 or above, we require a phase margin of 50^o or above.

Unfortunately we cannot increase the phase independent of gain. Nor can we change the *shape* of the gain independent of phase.

11.2 Phase lead design

A phase lead compensator "buys" extra phase margin at the "expense" of increased gain. Its original popularity stemmed from it being implementable using passive electronic components. It takes the form

$$C(s) = k\frac{1 + T_L s}{1 + \alpha T_L s},$$

with $\alpha < 1$. Observe that it has the same structure as a PDγ controller. The Nyquist plot of a phase lead compensator is shown in Fig 99. The plot is a

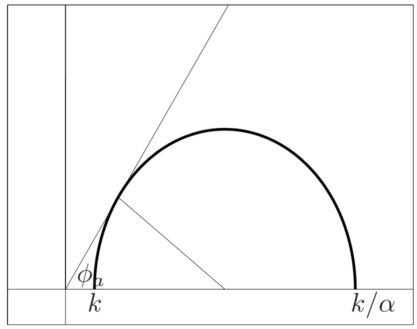


Figure 99: Nyquist plot of a phase lead compensator.

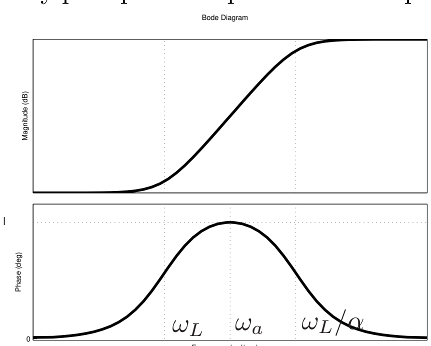


Figure 100: Bode plot of a phase lead compensator. The gain is $20\log_{10} k$ at low frequencies and rises to $20\log_{10}(k/\alpha)$ at high frequencies.

semicircle with centre at

$$c = \frac{1}{2}\left(k + \frac{k}{\alpha}\right) = \frac{k}{2\alpha}(1+\alpha),$$

and radius

$$r = c - k = \frac{k}{2\alpha}(1-\alpha).$$

It follows that the maximum phase advance ϕ_a occurs when

$$\sin\phi_a = \frac{r}{c} = \frac{1-\alpha}{1+\alpha}.$$

Equivalently, given a desired phase advance ϕ_a, we may specify

$$\alpha = \frac{1-\sin\phi_a}{1+\sin\phi_a}.$$

Put $\omega_L = 1/T_L$ and $\omega_a = \omega_L/\sqrt{\alpha}$. From the Bode plot (Fig 100) we see that the maximum phase advance occurs when $\omega = \omega_a$.

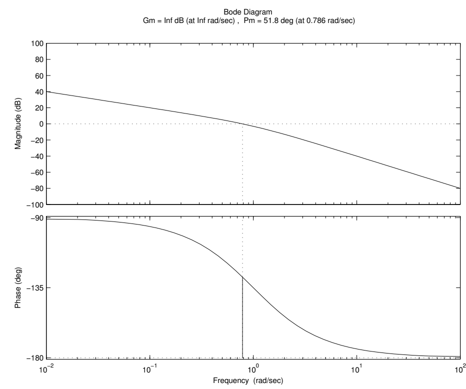


Figure 101: Phase margin of example with $k = 1$.

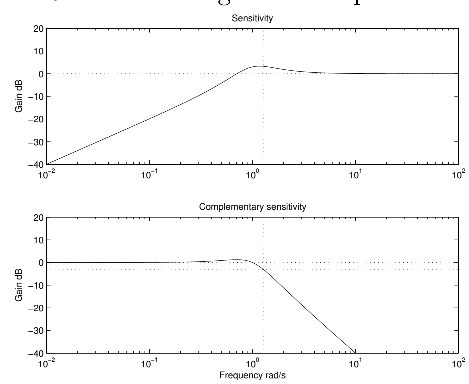


Figure 102: Closed-loop sensitivities with $k = 1$.

Example: $G(s) = \dfrac{1}{s(s+1)}$.

If $C(s) = 1$ so that $kL(s) = G(s)$, the gain margin is infinite and the phase margin is 51.8^o (Fig 101). The resulting sensitivities have reasonably small overshoot, with a closed-loop bandwidth of 1.27 rad/s (Fig 102).

Suppose instead $C(s) = 10$. This gives a cross-over frequency of around 3 rad/s and a closed-loop bandwidth of around 5 rad/s. The gain margin is still infinite, but the phase margin is only 18^o (Fig 103). The closed-loop sensitivities show some nasty peaks (Fig 104).

Suppose we have a design specification that the cross-over frequency should be 3 rad/s but the phase margin should be at least 50^o. We need a phase lead compensator with a maximum of 32^o phase advance at 3 rad/s.

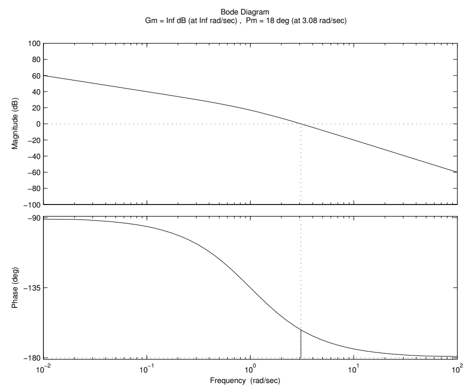


Figure 103: Phase margin of example with $k = 10$.

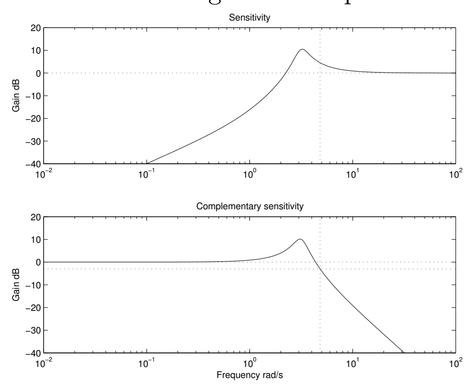


Figure 104: Closed-loop sensitivities with $k = 10$.

So put

$$
\begin{aligned}
\alpha &= \frac{1 - \sin 32^o}{1 + \sin 32^o} \approx 0.31, \\
\omega_a &= 3, \\
\omega_L &= \omega_a \sqrt{\alpha} \approx 1.66, \\
C(s) &= k \frac{\omega_L + s}{\omega_L + \alpha s} \approx k \frac{1.66 + s}{1.66 + 0.31 s}.
\end{aligned}
$$

The Bode plot of the compensator with $k = 1$ is shown in Fig 105. The Bode plot GC (with $k = 1$) together with that of G is shown in Fig 106.

To attain the cross-over frequency $\omega_c = 3$, we need to raise k to 5.25 (this can be read from Fig 106, where the gain of GC with $k = 1$ is -14dB at $\omega = 3$; we have $20 \log_{10} 5.25 \approx 14$). The phase margin is then 50^o as required (Fig 107). The resulting sensitivities are shown in Fig 108.

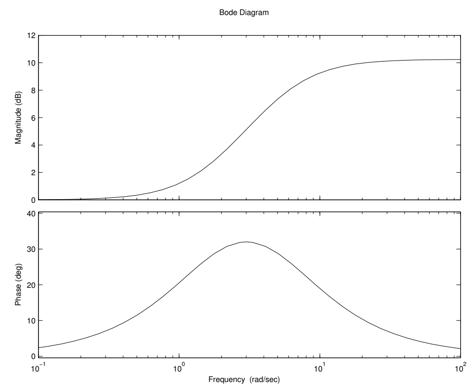


Figure 105: Phase lead compensator for example (with $k = 1$).

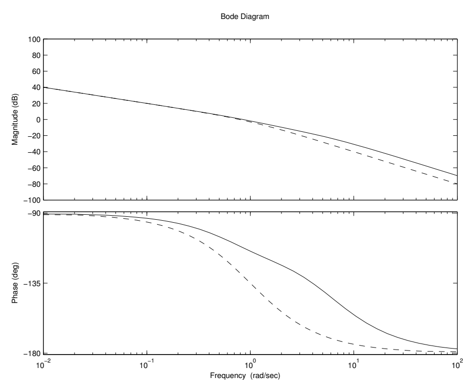


Figure 106: Comparison of compensated (solid) and uncompensated (dashed) forward loop transfer functions for example (with $k = 1$).

In the example we have specified the cross-over frequency and the phase margin. This allowed us to choose the phase lead compensator directly. Usually we will have further specifications (for example, we may require higher low frequency gain to improve disturbance rejection) in which case an iterative design procedure is more appropriate. Different textbooks recommend different procedures.

ω_c	cross-over frequency of the forward loop transfer function
ω_b	bandwidth of the complementary sensitivity
ω_a	frequency of compensator peak phase advance
ω_L	first break frequency of compensator

Table 1. Defined frequencies in phase-lead compensator design.

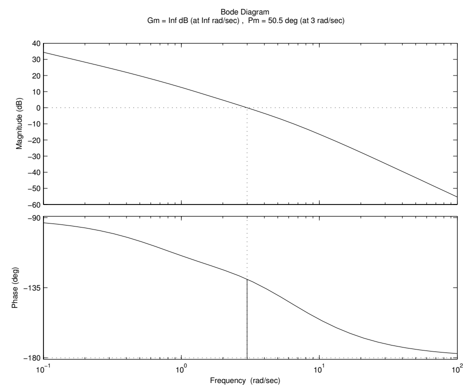


Figure 107: Phase margin of example with phase lead compensator ($k = 5.25$).

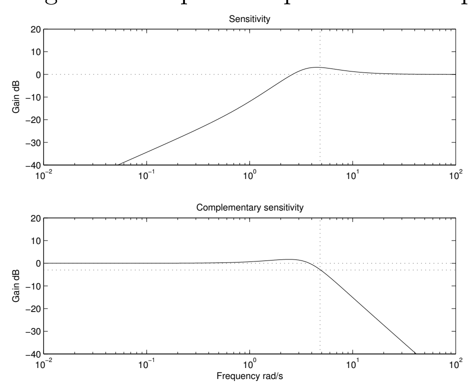


Figure 108: Closed-loop sensitivities with phase lead compensator ($k = 5.25$).

11.3 Phase lag design

A phase lag compensator has the same structure as a phase lead compensator

$$C(s) = k\frac{1 + T_L s}{1 + \alpha T_L s},$$

but with $\alpha > 1$. Its effect is opposite to that of a phase lead compensator; it "buys" extra gain at low frequencies, at the "expense" of increased phase. In this sense its role is similar to that of a PI controller. Once again it may be implemented using passive electronic components.

If we write $\omega_L = 1/T_L$ so that

$$C(s) = k\frac{s + \omega_L}{\alpha s + \omega_L},$$

then we should choose ω_L well below the cross-over frequency, so that the phase margin is not affected.

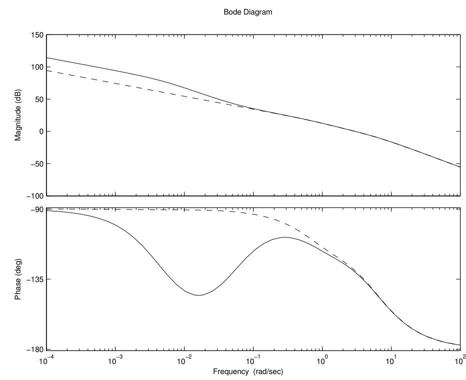


Figure 109: Bode plot of forward loop transfer function for the example with phase lead compensation (dashed) and both phase lead and phase lag compensation (solid).

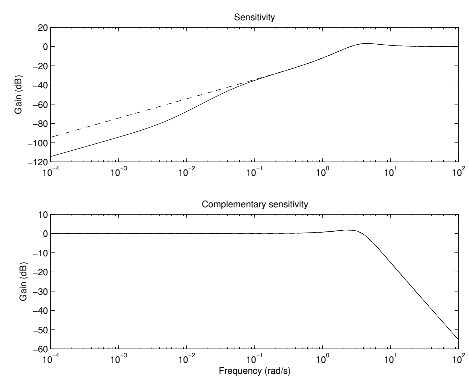


Figure 110: Closed-loop sensitivities for the example. Lead compensation is shown dashed while combined lead and lag compensation is shown solid.

Example: Consider our previous example $G(s) = \dfrac{1}{s(s+1)}$ with phase lead compensator

$$C_{\text{lead}}(s) = 5.25\frac{1.66 + s}{1.66 + 0.31s}.$$

We will include a phase lag compensator

$$C_{\text{lag}}(s) = \frac{s + 0.05}{s + 0.005},$$

so that the overall controller is given by

$$C(s) = C_{\text{lead}}(s)C_{\text{lag}}(s).$$

The Bode plots of both $G(s)C_{\text{lead}}(s)$ and $G(s)C_{\text{lead}}(s)C_{\text{lag}}(s)$ are shown in Fig 109.

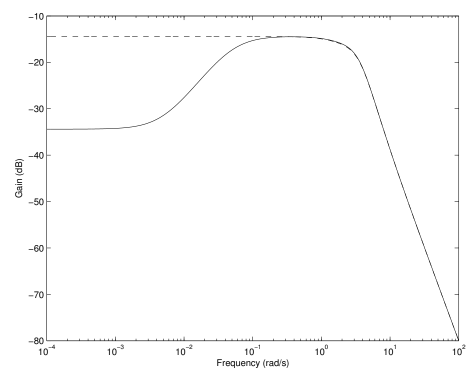


Figure 111: Gain of GS for the two cases. Lead compensation is shown dashed while combined lead and lag compensation is shown solid.

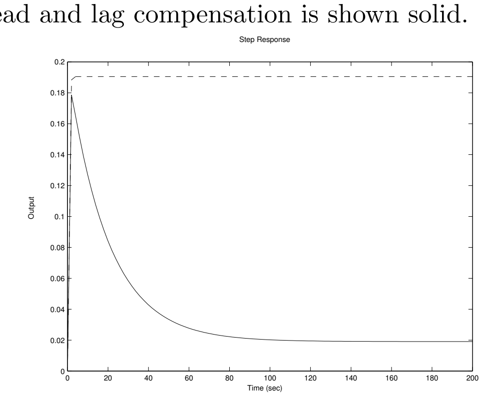


Figure 112: Responses to a unit step input disturbance. Lead compensation is shown dashed while combined lead and lag compensation is shown solid.

The closed loop sensitivities are shown in Fig 110. The sensitivity with a phase lag compensator has lower gain at very low frequency. The two complementary sensitivities look similar for this example.

For this example the phase lag compensator gives a significant improvement to the rejection of input disturbances. Fig 111 shows the gains of GS for each case. The gain with lead compensation is significantly smaller at low frequencies. The corresponding step responses to unit input disturbances are shown in Fig 112. The performance is considerably improved with the inclusion of a phase lag compensator.

11.4 Lead-lag compensators

In the previous section we saw that phase lead compensators and phase lag compensators can by usefully combined. The combined compensator has transfer function

$$\begin{aligned} C(s) &= C_{\text{lead}}(s)C_{\text{lag}}(s), \\ &= k\left(\frac{1+T_1 s}{1+\alpha_1 T_1 s}\right)\left(\frac{1+T_2 s}{1+\alpha_2 T_2 s}\right), \end{aligned}$$

with

$$\alpha_1 < 1 \text{ and } \alpha_2 > 1.$$

Any such compensator is termed a "lead-lag" compensator. However, historically it was common to implement such a compensator with

$$\alpha = \alpha_1 = \frac{1}{\alpha_2}.$$

In this case the compensator has transfer function

$$\begin{aligned} C_{\text{lead-lag}}(s) &= k\left(\frac{1+T_1 s}{1+\alpha T_1 s}\right)\left(\frac{1+T_2 s}{1+(T_2/\alpha)s}\right), \\ &= k\left(\frac{s+1/T_1}{s+1/(\alpha T_1)}\right)\left(\frac{s+1/T_2}{s+\alpha/T_2}\right). \end{aligned}$$

The advantage of such a restriction is that the compensator (without the gain also set to $k = 1$) can be implemented as a passive circuit with only two resistors and two capacitors. Both lead and lag compensators can also be implemented using such a passive circuit realization. However the structure of such circuits is beyond the scope of this course.

11.5 Notch filters

Notch filters provide an additional tool in classical control. They are usually used when the plant has a resonant response at high frequency: this may not be apparent in open loop, but effectively limits the closed-loop bandwidth without careful control design. As an illustrative example, consider a plant with transfer function

$$G(s) = e^{-s/10}\left[\left(\frac{1}{10s+1}\right) + \left(\frac{1}{s^2+s+100}\right)\right].$$

Its Bode plot, together with phase and gain margins, is shown in Fig 113. A suitable notch filter would have zeros near the two resonant poles. For example, $C(s)$ might be chosen as

$$C(s) = \frac{s^2+s+81}{s^2+18s+81}.$$

The Bode plot of C is shown in Fig 114 and the Bode plot of GC (together with gain and phase margins) is shown in Fig 115. It is apparent that the notch filter allows higher control gain. To illustrate this, Fig 116 shows the complementary sensitivities for the plant with and without a notch filter when the proportional gain is 5.

Note that a notch filter effectively ensures there is *no* control action near the problematic frequency. This means the technique will be ineffectual if input disturbances in this band region are problematic.

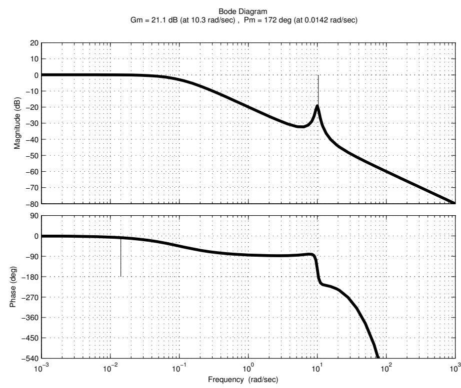


Figure 113: Bode plot of a plant with a high frequency resonance.

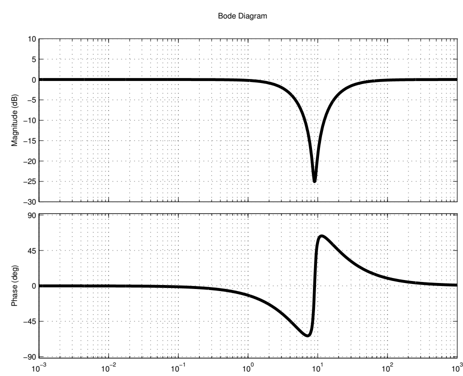


Figure 114: Bode plot of a notch filter.

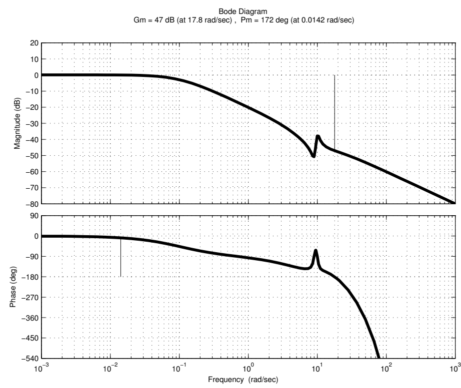


Figure 115: Bode plot of the forward loop transfer function with a notch filter.

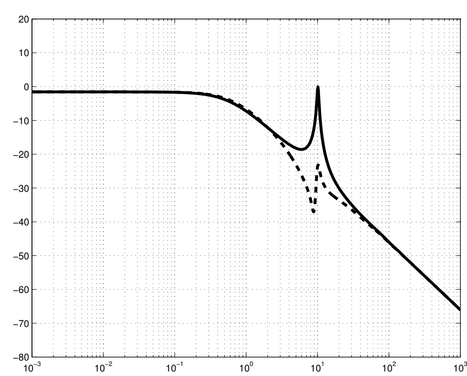


Figure 116: Complementary sensitivities with (dashed) and without (solid) a notch filter; the proportional gain is 5 in both cases.

11.6 Key points

- Control design tradeoffs are best expressed in the frequency domain.
- Phase lead compensators allow higher cross-over frequency and/or bigger phase margin.
- Phase lead compensators are structurally equivalent to PDγ control.
- Phase lag compensators allow increased gain at low bandwidth.

Part IV
Tutorial questions

Tutorial 1

Question 1.1

The voltage $u(t)$ shown in Figure T1.1 is an exponential function. Its tangent at time $t = 0$ is also shown. It is applied as the input of a plant with the transfer function

$$G(s) = \frac{3}{(s+1)(s+4)}$$

Propose an appropriate expression for the input $u(t)$ and hence compute the output signal $y(t)$. Sketch it for $0 \le t \le 10$.

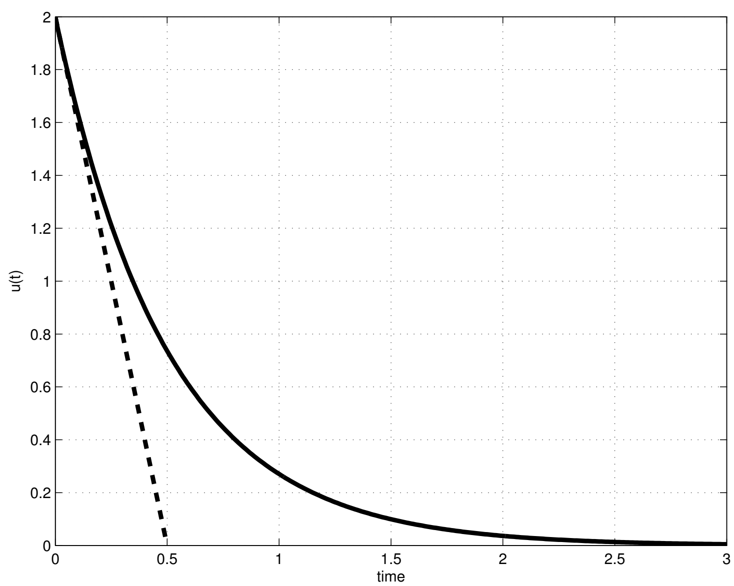


Figure T1.1: Input voltage $u(t)$.

Question 1.2

1. For the RLC network in Figure T1.2, find the transfer function

$$G(s) = \frac{Y(s)}{U(s)}$$

   Give an expression for the natural frequency ω_n and the damping ratio ζ in terms of R_1, R_2, C and L.

2. Suppose $C = 10\mu F$ and $L = 1mH$. What shape step response do you expect for the following values of R_1 and R_2?

   (a) $R_1 = 2\Omega$, $R_2 = 10\Omega$,

   (b) $R_1 = 10\Omega$, $R_2 = 2\Omega$,.

   You *may* find the following approximations useful:

$$M_p \approx 1 - \frac{\zeta}{0.6},\ t_r \approx \frac{1.7}{\omega_n},\ t_s \approx \frac{4.6}{\omega_n \zeta}$$

   Plot the responses using Matlab; how good are the approximations?

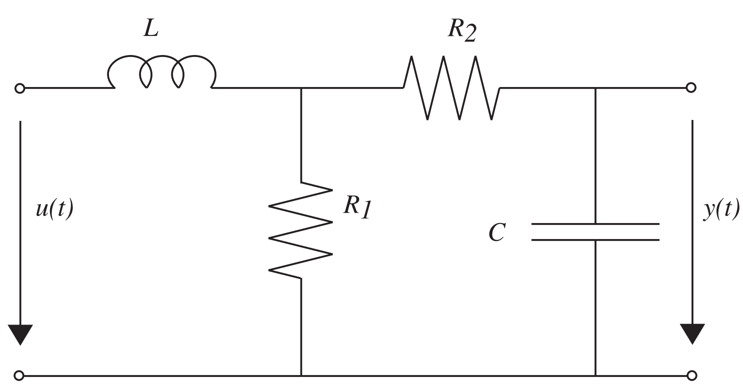


Figure T1.2: RLC network.

Tutorial 2

Question 2.1

Consider the speed control system shown in Fig T2.1. When $k_0 = 1$ and $\tau = 0.2$, what gain k_v is required to keep the steady state error $v_r - v$ to a step input less than 0.05 when $v_r = 1$?

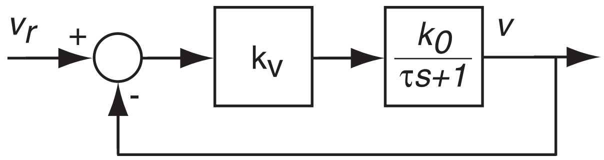


Figure T2.1: Speed control loop.

Question 2.2

The control loop in Problem 1 has been modified from speed control to control of the position $y = \int_{t_0}^{t} v\, dt$, as shown in Fig T2.2. Numerical values are $k_0 = 1$ and $\tau = 0.2$.

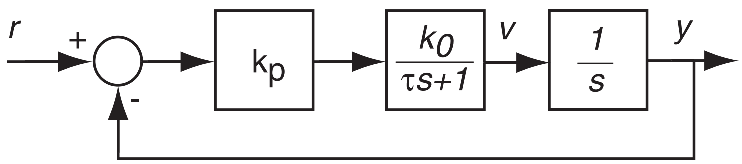


Figure T2.2: Position control loop.

1. What is the steady state error to a unit step position demand?
2. What is the fastest 1% settling time t_s that can be achieved with proportional gain?
3. Assume that both the position $y(t)$ and the speed $v(t)$ are measured and used for feedback control according to

$$u(t) = k_p\,[r(t) - y(t)] - k_v v(t)$$

   Redraw the block diagram and find controller gains k_p and k_v such that the damping ratio is not less than 0.7 and the settling time is no more than 0.5.

Question 2.3

Fig T2.3 shows the Bode plot of a plant. Estimate its transfer function $G(s)$.

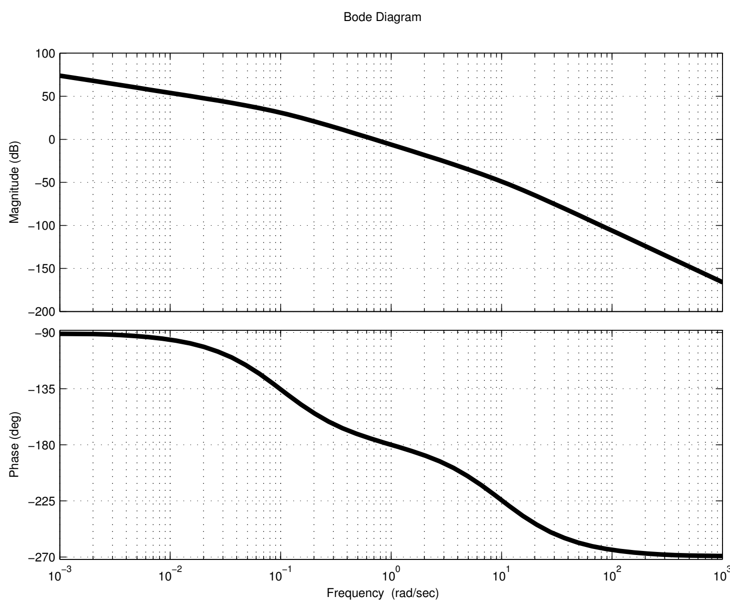


Figure T2.3: Bode plot of a plant.

Tutorial 3

Question 3.1

Consider the control system depicted in figure T3.1. The plant has transfer function

$$G(s) = \frac{18}{s(s+6)}$$

and is subject to step input disturbances $d_i(t)$. The plant output is required to track a reference signal $r(t)$ in closed-loop.

Suppose $C(s)$ is a proportional feedback controller

$$C(s) = k_p$$

a) Use Matlab to draw the root locus for the forward loop transfer function.

b) What gain k_p is required to ensure the peak overshoot in the response to a step change in $r(t)$ is less than 17%?

c) What gain k_p is required to ensure that a unit step change in $d_i(t)$ should change the steady state value of the output by less than 0.05?

d) Is it possible to satisfy both performance specifications at once?

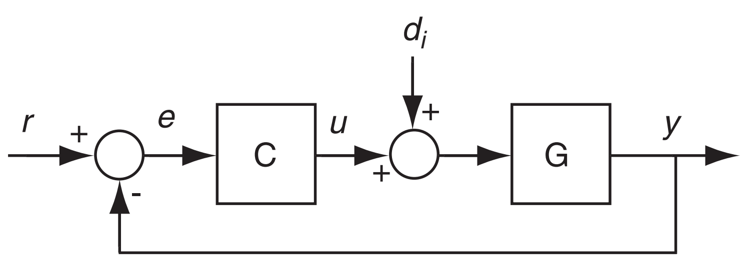


Figure T3.1: Feedback control system.

Question 3.2

Design a PD controller

$$C(s) = k_p(1 + T_D s)$$

for the plant that meets both performance specifications. Use Matlab to draw the root locus of the forward loop transfer function with an appropriate value of T_D.

Question 3.3

Suppose the controller is implemented as a PDγ controller

$$C(s) = k_p \left(1 + \frac{T_D s}{1 + \gamma T_D s}\right)$$

with $\gamma = 0.1$. For the same value of T_D, use Matlab to draw the root locus of the forward loop transfer function.

Tutorial 4

Question 4.1

Suppose a plant with transfer function

$$G(s) = \frac{5}{s(s+0.1)(s+10)}$$

is controlled using proportional feedback with gain k.

1. Draw the Nyquist plot of the forward loop transfer function when $k = 1$.
2. Use the Nyquist stability criterion to determine closed-loop stability for both large and small k.
3. Confirm your answer by drawing the root locus using Matlab.

Question 4.2

A plant has frequency response shown in Fig T4.1. It is to be controlled using proportional feedback with gain k.

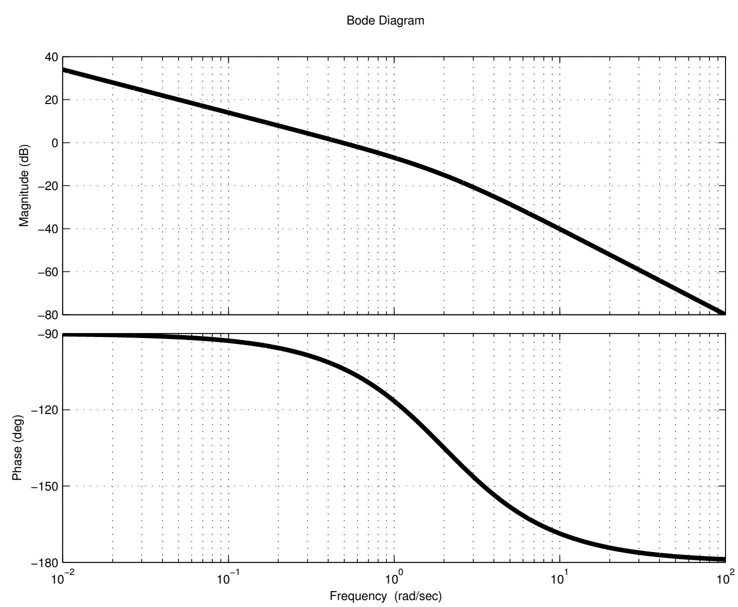


Figure T4.1: Bode plot.

1. Estimate the transfer function of the plant.
2. What is the minimum steady state error to a unit ramp set point if the phase margin should be at least 30^o?
3. What is maximum phase margin if the gain is sufficiently large for the steady state error to a unit ramp set point to be less than 0.01?

Question 4.3

For the same plant whose frequency response is shown in Fig T4.1, design a phase lead compensator $C(s)$ that ensures the following:

- the steady state error to a unit ramp set point is at most 0.01,
- the phase margin is at least 30^o.

Part V
Worked examples

Worked examples

Problem 1

The armature voltage v is used to control the shaft angle θ of a DC motor, whose schematic is shown in Fig WE1.1. Let J be the motor shaft inertia, T_m be the motor torque and T_l be the load torque. Also let R be the armature resistance, let i be the armature current and let e_b be the back emf. Neglecting inductance, shaft compliance and friction, the motor can be described by the equations

$$\begin{aligned} J\ddot{\theta} &= T_m + T_l \\ v &= Ri + e_b \end{aligned}$$

Let K_m be the motor constant. Then assuming that the field is constant, we have

$$\begin{aligned} T_m &= K_m i \\ e_b &= K_m \dot{\theta} \end{aligned}$$

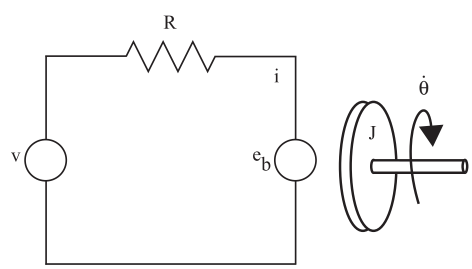


Figure WE1.1: Schematic of a DC motor.

1. Find the transfer function and draw a block diagram of the plant with inputs and v and T_l and with output θ.

2. Suppose v is used as a manipulated variable to control the shaft angle (with T_l considered as an input disturbance). Suppose further that proportional feedback with gain K_p is used, so that

$$v = K_p(\theta_r - \theta)$$

   for some angle demand signal θ_r. If $J = 0.1$, $R = 0.1$ and $K_m = 0.5$ then:

   (a) After a step change in θ_r, what is the 1% settling time t_s?

   (b) What condition must K_p satisfy for the peak overshoot to be less than 17%?

3. Suppose in addition that a tachometer is used to measure the motor speed $\dot{\theta}$, and that this measurement is used for velocity feedback of the form

$$v = K_p(\theta_r - \theta) - K_T\dot{\theta}$$

Draw the block diagram of the closed-loop system such that the plant output is $\dot{\theta}$, with θ obtained via an additional integrator. What is the closed-loop transfer function from θ_r to θ?

4. Find values of K_p and K_T such that the peak overshoot is no more than 17% and the 1% settling time is no more than 0.1.

Problem 2

The control system shown in Fig WE2.1 is required to meet the following specifications:

- The effect of a constant input disturbance d_i in steady state operation should be reduced to less than 10%.
- The damping ratio ζ should not be less than 0.7.

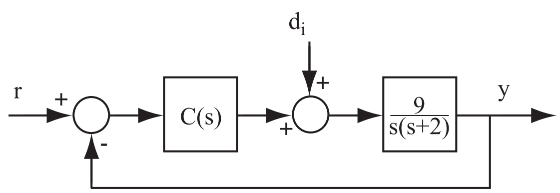


Figure WE2.1: Control system block diagram.

1. Suppose proportional feedback is used, i.e.

$$C(s) = K_P$$

   (a) Use Matlab to draw the root locus of the system.
   (b) Can the specifications be met? If so, find a suitable value for K_p.

2. Suppose an idealised PD control is used with,

$$C(s) = K_P(1 + T_D s)$$

   (a) Can the specification be met? If so, find suitable values for K_p and T_d.
   (b) Use Matlab to draw the root locus of the system with this value of T_d.

3. Propose a modification to the PD controller such that the design specifications can be met with the controller realizable.

Problem 3

A controller is to be designed for a plant with transfer function

$$G(s) = \frac{4}{s(s+2)}$$

such that the steady state error to a unit ramp set point is less than 0.05, and the peak overshoot of the step response is no more than 17%. The Bode plot of $G(s)$ is shown in Fig WE3.1.

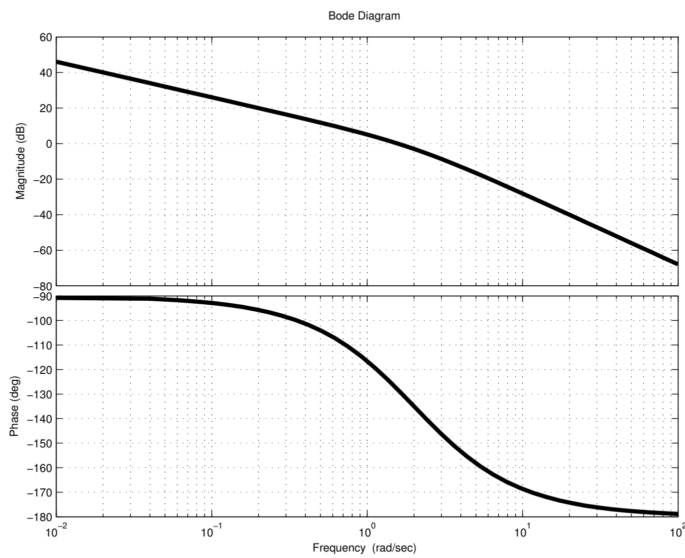


Figure WE3.1: Bode plot of $G(s)$.

1. When proportional feedback

   $$C(s) = K_p$$

   is used, what is the steady state error to a unit ramp set point, and under what condition on the gain K_P can the ramp set point specification be met?

2. With a controller gain that meets the condition on the ramp set point, what is the largest achievable phase margin?

3. What phase margin is required to achieve the peak overshoot specification?

4. Design a lead compensator $C(s)$ that meets both specifications.

Problem 4

Consider the control loop shown in Fig WE4.1. The transfer function of the plant is

$$G(s) = \frac{s+2}{s(s-5)}$$

Its Bode plot is shown in Fig WE4.2.

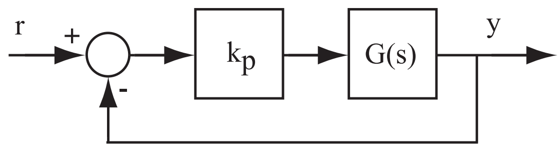


Figure WE4.1: Control system block diagram.

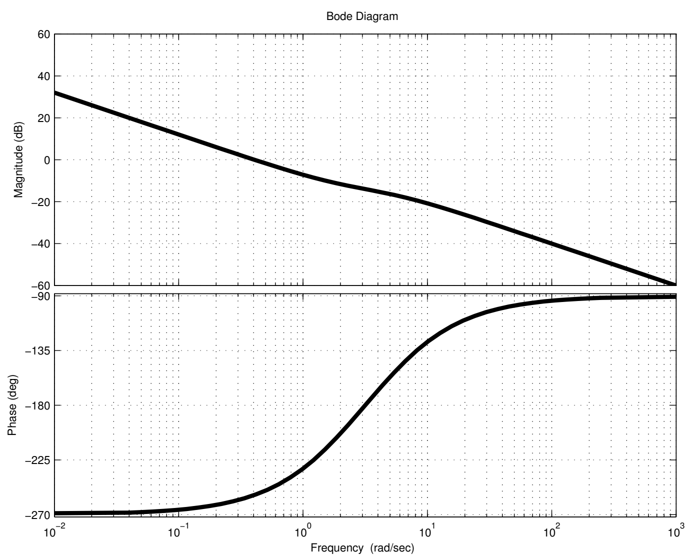


Figure WE4.2: Bode plot of $G(s)$.

1. When $G(s)$ is evaluated along the Nyquist path, what is the phase angle for $s = j\omega$ with frequencies $\omega \to -\infty$ and $\omega \to \infty$? If $\varepsilon > 0$ is a small constant, what is the phase angle for $s = j\omega$ at frequencies $\omega = -\varepsilon$ and $\omega = \varepsilon$? What are the properties of $G(s)$ take when $s = \varepsilon$?

2. Sketch the Nyquist plot of $G(s)$.

3. Discuss closed-loop stability when K_P is allowed to take values from 0 to ∞.

4. Use Matlab to draw the root locus for the system and hence confirm your statements on closed-loop stability.

Problem 5

The flow of a liquid through a system of two coupled tanks illustrated in Fig WE5.1 is to be controlled. We consider only the deviation of flow rates from the steady state value. The variables are:

- q_i - the inflow rate into tank 1,
- h_1 - the liquid level in tank 1,
- q_1 - the outflow rate of tank 1, which is the inflow rate of tank 2,
- h_2 - the liquid level in tank 2,
- q_2 - the outflow rate of tank 2.

The liquid level in tank 1 is governed by the equations

$$A_1\dot{h}_1 = q_i - q_1$$
$$R_1 = \frac{h_1 - h_2}{q_1}$$

where R_1 is the resistance at the outlet and A_1 is the cross-sectional area of the tank. Similarly the liquid level in tank 2 is governed by

$$A_2\dot{h}_2 = q_1 - q_2$$
$$R_2 = \frac{h_2}{q2}$$

Suppose we have the numerical values:

$$R_1 = 4,\ R_2 = 2,\ A_1 = 1,\ A_2 = 2$$

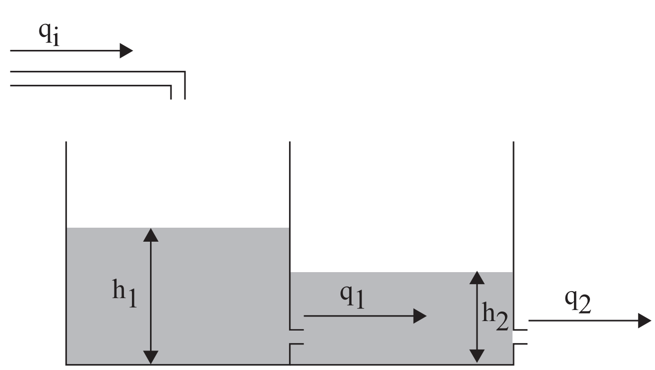


Figure WE5.1: Coupled tanks with flows and liquid levels.

1. Find the transfer function $G(s)$ from tank 1 inflow q_i to tank 2 outflow q_2.

2. With the inflow q_i as control input and the outflow q_2 as measured output, suppose a proportional controller is used. What gain K_P is required to achieve a rise time of approximately 1 after a unit step as reference input? Estimate the peak overshoot with this controller.

3. Suppose instead a PD controller is used. What derivative time T_D is required to keep the peak overshoot below 10% when the gain is chosen to achieve a rise time of 1?

4. What is the relative steady state error to a step input with the above controllers?

Problem 6

A plant has transfer function

$$G(s) = \frac{1}{(s+3)(s+6)}$$

1. When proportional feedback is used to control the output of the plant, what is the minimum value of the gain K_P required to achieve a steady state error of less than 10% to a step set point change?

2. Suppose the controller is given as pure integral action

$$C(s) = \frac{K_P}{s}$$

   Use Matlab to draw the root locus for the forward loop transfer function in this case. Explain why such a controller would lead to a poor transient response.

3. Suppose the controller is given as

$$C(s) = K_P \frac{s+K_I}{s}$$

   (a) If $K_I = 3$, express the closed-loop damping ratio as a function of K_P.

   (b) If $K_I = 6$, express the closed-loop damping ratio as a function of K_P.

   (c) Use Matlab to draw the root locus diagrams for the three cases $K_I = 3$, $K_I = 4.5$ and $K_I = 6$. Compare them briefly.

Problem 7

A plant with transfer function

$$G(s) = \frac{10}{(s-1)(s+100)}$$

is to be controlled with proportional feedback.

1. Sketch the Bode plot of the plant.
2. Roughly sketch the Nyquist plot of the plant.
3. Apply the Nyquist criterion to find the values of the gain K_P for which the closed-loop system is stable.

Problem 8

A plant with the frequency response shown in Fig WE8.1 is to be controlled.

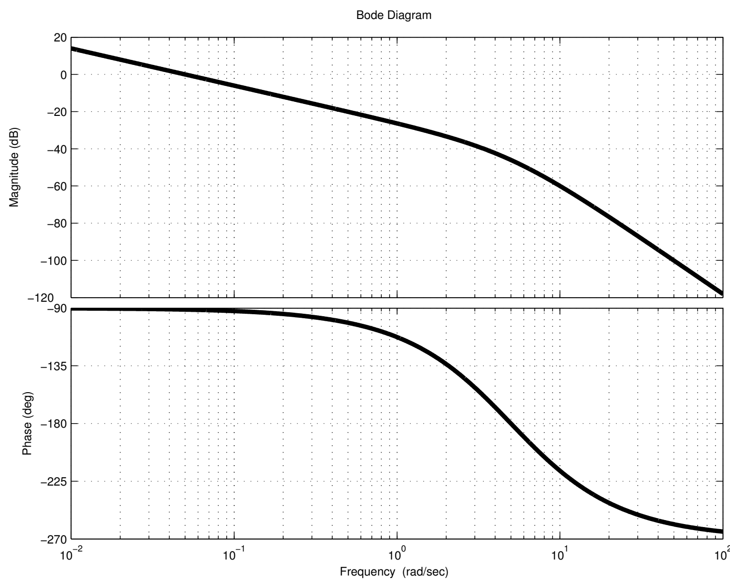


Figure WE8.1: Bode plot of plant.

1. Design a phase-lead compensator to meet the following specifications:
   - The closed-loop bandwidth should be at least 3 rad/sec.
   - The peak overshoot should be less than 10%.
2. How does the controller affect the sensitivity of the control system to high frequency noise?

Worked example solutions

Problem 1

1. Taking Laplace transforms we find

$$
\begin{aligned}
Js^2\Theta(s) &= T_m(s) + T_l(s) \\
V(s) &= RI(s) + E_b(s) \\
T_m(s) &= K_m I(s) \\
E_b(s) &= K_m s\Theta(s)
\end{aligned}
$$

   We need to substitute for $T_m(s)$, $I(s)$ and $E_b(s)$. First eliminate $T_m(s)$ and $E_b(s)$:

$$
\begin{aligned}
Js^2\Theta(s) &= K_m I(s) + T_l(s) \\
V(s) &= RI(s) + K_m s\Theta(s)
\end{aligned}
$$

   Now eliminate $I(s)$:

$$JRs^2\Theta(s) = K_m[V(s) - K_m s\Theta(s)] + RT_l(s)$$

   This simplifies to

$$\Theta(s) = \frac{1}{s(JRs + K_m^2)}\left[K_m V(s) + RT_l(s)\right]$$

   The block diagram of the open-loop system is shown in Fig WE1.2.

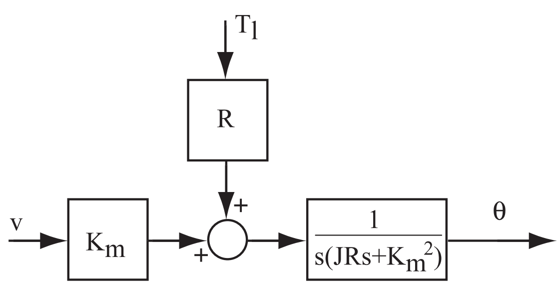


Figure WE1.2: Block diagram of DC motor.

2. (a) The transfer function from $V(s)$ to $\Theta(s)$ is

$$
\begin{aligned}
G(s) &= \frac{K_m}{JRs^2 + K_m^2 s} \\
&= \frac{0.5}{0.01s^2 + 0.25s} \\
&= \frac{50}{s^2 + 25s}
\end{aligned}
$$

If

$$v = K_p(\theta_r - \theta)$$

then the closed-loop transfer function from Θ_r to Θ is

$$\begin{aligned} H(s) &= \frac{K_p G(s)}{1 + K_p G(s)} \\ &= \frac{50K_p}{s^2 + 25s + 50K_p} \end{aligned}$$

If we write this as

$$H(s) = \frac{k\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2}$$

then

$$\begin{aligned} k &= 1 \\ \omega_n^2 &= 50K_p \\ 2\zeta\omega_n &= 25 \end{aligned}$$

The 1% settling time can be approximated as

$$\begin{aligned} t_s &\approx \frac{4.6}{\zeta\omega_n} \\ &= 0.368 \end{aligned}$$

(b) The peak overshoot is given by

$$M_p = \exp\left(-\frac{\pi\zeta}{\sqrt{1-\zeta^2}}\right)$$

So we require

$$-\frac{\pi\zeta}{\sqrt{1-\zeta^2}} \leq \log 0.17$$

Hence

$$\pi^2\zeta^2 \geq (\log 0.17)^2(1-\zeta^2)$$

Hence

$$\begin{aligned} \zeta &\geq \frac{|\log 0.17|}{\sqrt{\pi^2 + (\log 0.17)^2}} \\ &\approx 0.5 \end{aligned}$$

But

$$\begin{aligned} \zeta &= \frac{25}{2\omega_n} \\ &= \frac{25}{2\sqrt{50K_p}} \end{aligned}$$

So we require

$$\frac{25}{2\sqrt{50K_p}} \geq \frac{1}{2}$$

This gives

$$50K_p \leq 25^2$$

and hence

$$K_p \leq 12.5$$

3. The block diagram of the closed-loop system is shown in Fig WE1.3. Introduce

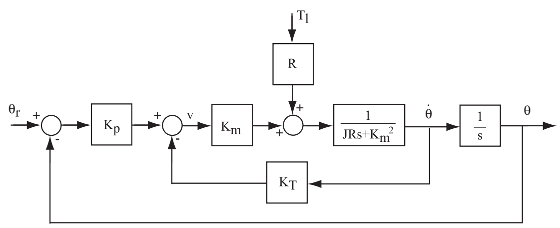


Figure WE1.3: Closed-loop system.

$$\phi = \dot{\theta}$$

Then the transfer function of the inner loop (ignoring T_l) is

$$\begin{aligned}\Phi(s) &= \frac{K_m/(JRs + K_m^2)}{1 + K_T K_m/(JRs + K_m^2)} K_p \left[\Theta_r(s) - \Theta(s)\right] \\ &= \frac{K_m K_p}{JRs + K_m^2 + K_T K_m} \left[\Theta_r(s) - \Theta(s)\right]\end{aligned}$$

Hence the transfer function from θ_r to θ is

$$\begin{aligned}\Theta(s) &= \frac{K_m K_p/(JRs^2 + (K_m^2 + K_T K_m)s)}{1 + K_m K_p/(JRs^2 + (K_m^2 + K_T K_m)s)} \Theta_r(s) \\ &= \frac{K_m K_p}{JRs^2 + (K_m^2 + K_T K_m)s + K_m K_p} \Theta_r(s) \\ &= \frac{50K_p}{s^2 + (25 + 50K_T)s + 50K_p} \Theta_r(s)\end{aligned}$$

4. We have

$$\begin{aligned}\omega_n^2 &= 50K_p \\ 2\zeta\omega_n &= 25 + 50K_T\end{aligned}$$

The two conditions are

$$\frac{4.6}{\zeta\omega_n} \leq 0.1$$
$$\zeta \geq 0.5$$

Hence we require

$$\frac{4.6}{(25 + 50K_T)/2} \leq 0.1$$
$$\frac{25 + 50K_T}{2\sqrt{50K_p}} \geq 0.5$$

The first gives

$$K_T \geq 1.34$$

If $K_T = 1.34$ then the second is satisfied with

$$K_p \geq 170$$

Problem 2

1. (a) In Matlab the root locus can be generated as follows:

```
>>L = tf(9,[1 2 0]);
>>rlocus(L);
```

The locus is shown in Fig WE2.2.

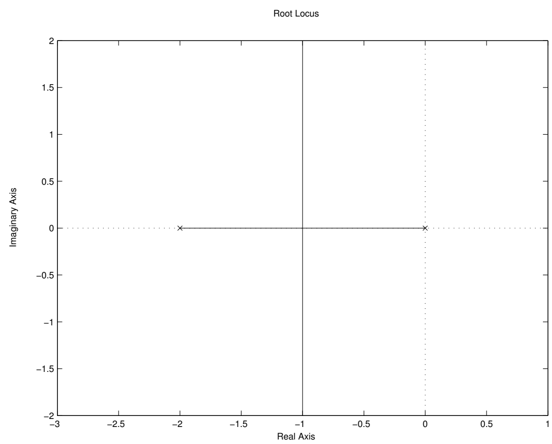


Figure WE2.2: Root locus of the system.

(b) The closed-loop system is given by

$$
\begin{aligned}
Y(s) &= \frac{G(s)}{1+G(s)C(s)}D_i(s) + \frac{G(s)C(s)}{1+G(s)C(s)}R(s) \\
&= \frac{9}{s^2+2s+9K_p}D_i(s) + \frac{9K_p}{s^2+2s+9K_p}R(s)
\end{aligned}
$$

To achieve the input disturbance specification we require

$$0.1 > \left. \frac{9}{s^2+2s+9K_p} \right|_{s=0} = \frac{1}{K_p}$$

and hence

$$K_p > 10$$

To achieve the damping ratio specification we require

$$\frac{2}{2\sqrt{9K_p}} > 0.7$$

and hence

$$K_p < \frac{2}{9}$$

The two conditions cannot be satisfied simultaneously.

2. (a) The closed-loop system is given by

$$Y(s) = \frac{9}{s^2 + (2 + 9K_pT_d)s + 9K_p} D_i(s) + \frac{9K_p(1 + T_ds)}{s^2 + (2 + 9K_pT_d)s + 9K_p} R(s)$$

The conditions are satisfied with $K_p = 10$ and

$$\frac{2 + 90T_d}{2\sqrt{90}} \geq 0.7$$

Which gives

$$T_d \geq 0.1254$$

For simplicity, choose

$$T_d = 0.2$$

(b) The forward loop transfer function is

$$kL(s) = K_p \frac{9(1 + 0.2s)}{s(s + 2)}$$

In Matlab the root locus can be generated as follows:

```
>>L = tf(9*[0.2 1],[1 2 0]);
>>rlocus(L);
```

The locus is shown in Fig WE2.3.

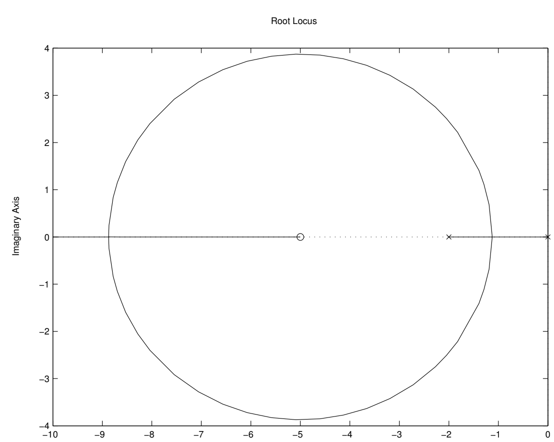


Figure WE2.3: Root locus of the system.

3. The PD controller should be replaced by a PDγ controller with structure

$$C(s) = K_p \left(1 + \frac{T_ds}{1 + \gamma T_ds}\right)$$

with γ small.

Problem 3

1. The closed loop transfer function from set point r to output y is

$$\begin{aligned} Y(s)/R(s) &= \frac{G(s)C(s)}{1+G(s)C(s)} \\ &= \frac{4K_p}{s^2+2s+4K_p} \end{aligned}$$

   The transfer function from r to error $e = r - y$ is

$$\begin{aligned} E(s)/R(s) &= 1 - Y(s)/R(s) \\ &= \frac{s^2+2s}{s^2+2s+4K_p} \end{aligned}$$

   The set point is a unit ramp, so $R(s) = 1/s^2$. The final value theorem gives

$$\begin{aligned} \lim_{t\to\infty} e(t) &= \lim_{s\to 0} sE(s) \\ &= \lim_{s\to 0} s\frac{s^2+2s}{s^2+2s+4K_p}\frac{1}{s^2} \\ &= \lim_{s\to 0} \frac{s+2}{s^2+2s+4K_p} \\ &= \frac{1}{2K_p} \end{aligned}$$

   So we require $Kp > 10$ for an error less than 0.05.

2. If $K_p = 10$, the gain in the Bode plot is raised by $20\log_{10} 10 = 20$dB. The Bode plot of $G(s)$ crosses the -20dB point at $\omega \approx 6$, where the phase is $\angle G(j\omega) \approx -162^o$. So the phase margin with $K_p = 10$ is $\phi_m = 18^o$. If the gain is further increased, the phase margin is reduced.

3. The rule of thumb $\zeta \approx \phi_m/100$ gives a closed-loop damping ratio of approximately 0.2. The overshoot is then

$$\begin{aligned} M_p &= e^{-\frac{\zeta}{\sqrt{1-\zeta^2}}\pi} \\ &\approx 0.56 \end{aligned}$$

4. To meet the overshoot specification we require the damping ratio $\zeta \geq 0.5$, which in turn corresponds to a phase margin $\phi_m \geq 50^o$. We also need to preserve the gain at low frequency, so a phase lead compensator will increase the cross-over frequency. Let us assume the cross-over frequency is doubled to $\omega_c \approx 12$. From the Bode plot, the phase is approximately -170^o at this frequency, so we require a phase advance of 40^o.

Put

$$\alpha = \frac{1 - \sin 40^o}{1 + \sin 40^o} \approx 0.22$$
$$\omega_l = \sqrt{\alpha} 12 \approx 5.6$$
$$C(s) = K_p \frac{\omega_l + s}{\omega_l + \alpha s} = 10 \frac{5.6 + s}{5.6 + 0.22s}$$

This gives a phase margin of 51^o with a cross over frequency 8 rad/s. For this particular case we have satisfied the specifications and it is not necessary to iterate further. See Fig WE3.2.

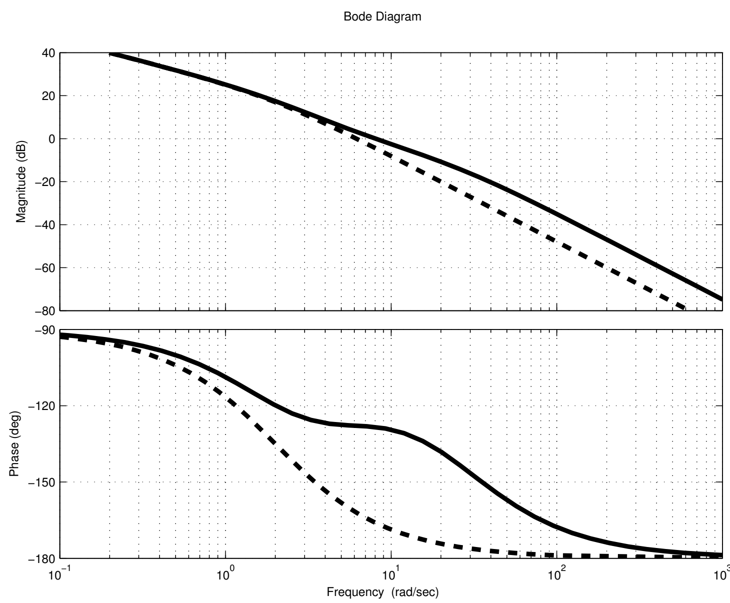


Figure WE3.2: Bode plot of $10G$ (dashed) and GC (solid) where C is the phase lead compensator.

Problem 4

1. For large frequencies $\omega \to \infty$ we can read from the Bode plot $\angle G(j\omega) \to -90^o$. Similarly for small frequencies $\omega \to 0^+$ we can read $\angle G(j\omega) \to -270^o$. By symmetry, we have $\angle G(j\omega) \to +90^o = -270^o$ as $\omega \to -\infty$ and $\angle G(j\omega) \to 270^o = -90^o$ as $\omega \to 0^-$. When $s = \varepsilon > 0$ with ε small, $G(\varepsilon)$ is large but negative.

2. See Fig WE4.3.

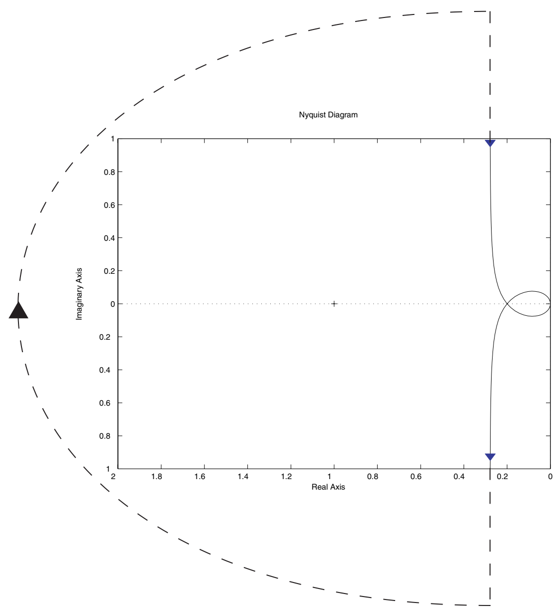


Figure WE4.3: Nyquist plot of $G(s)$.

3. We need to find when $G(j\omega)$ crosses the real axis.

$$\begin{aligned} G(j\omega) &= \frac{2 + j\omega}{j\omega(j\omega - 5)} \\ &= \frac{-j(2 + j\omega)(-5 - j\omega)}{\omega(25^2 + \omega^2)} \\ &= \frac{-7}{(25 + \omega^2)} + j\frac{10 - \omega^2}{\omega(25 + \omega^2)} \end{aligned}$$

The imaginary part is zero when $\omega^2 = 10$. At this frequency

$$\begin{aligned} G(j\sqrt{10}) &= \frac{-7}{25 + 10} \\ &= \frac{-1}{5} \end{aligned}$$

Applying the Nyquist criterion, the number of open-loop poles in the right half plane is $P = 1$. If $K_P < 5$ the Nyquist plot encircles -1 once clockwise, so $N = 1$ and the number of unstable closed-loop poles is $Z = N + P = 2$. But if $K_p > 5$ the Nyquist plot encircles -1 once anticlockwise, so $N = -1$ and so $Z = 0$. So the closed-loop system is stable for $K_p > 5$.

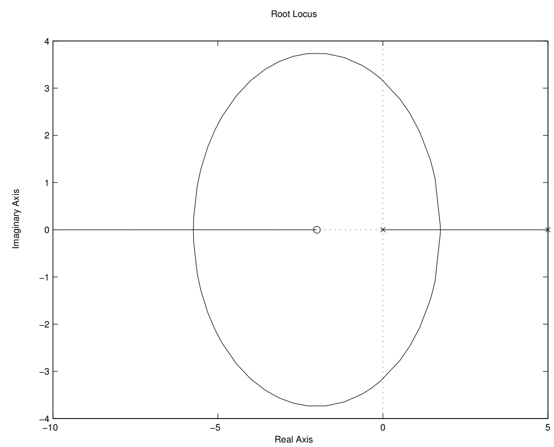


Figure WE4.4: Root locus plot of $G(s)$.

4. The root locus can be drawn in Matlab using the commands:

```
>>L = tf([1 2],[1 -5 0]);
>>rlocus(L);
```

This shows that the locus is unstable for low gains, but for k sufficiently large the closed-loop poles move into the left half plane. When $K_p = 5$ the closed-loop denominator is:

$$\begin{aligned} d(s) &= 5(s+2) + s(s-5) \\ &= s^2 + 10 \end{aligned}$$

With this value the closed-loop poles lie on the imaginary axis, confirming that it represents the transition from instability to stability.

Problem 5

1. Taking Laplace transforms:

$$\begin{aligned} sA_1H_1(s) &= Q_i(s) - Q_1(s) \\ R_1Q_1(s) &= H_1(s) - H_2(s) \\ sA_2H_2(s) &= Q_1(s) - Q_2(s) \\ R_2Q_2(s) &= H_2(s) \end{aligned}$$

   Substituting for H_1 and H_2 gives

$$\begin{aligned} sA_1\left(R_1Q_1(s) + R_2Q_2(s)\right) &= Q_i(s) - Q_1(s) \\ sA_2R_2Q_2(s) &= Q_1(s) - Q_2(s) \end{aligned}$$

   Substituting for $Q_1(s)$ gives

$$sA_1\left[R_1(1 + sA_2R_2)Q_2(s) + R_2Q_2(s)\right] = Q_i(s) - (1 + sA_2R_2)Q_2(s)$$

   This simplifies to

$$(1 + s(A_1R_1 + A_1R_2 + A_2R_2) + s^2A_1A_2R_1R_2)Q_2(s) = Q_i(s)$$

   Inserting the numerical values gives

$$Q_2(s) = \frac{1}{16s^2 + 10s + 1}Q_i(s)$$

2. The closed-loop response is

$$\frac{G(s)C(s)}{1 + G(s)C(s)} = \frac{K_p}{16s^2 + 10s + 1 + K_p}$$

   The rise time can be approximated as

$$t_r \approx \frac{1.8}{\omega_n}$$

   with

$$\omega_n^2 = \frac{1 + K_p}{16}$$

   So the specification is satisfied when

$$K_p \approx 16 \times 1.8^2 - 1 \approx 51$$

   The damping ratio is given by

$$\zeta = \frac{10/16}{2\omega_n} \approx 0.17$$

   for this value of K_p. The peak overshoot is

$$M_p = e^{-\frac{\zeta}{\sqrt{1-\zeta^2}}\pi} \approx 0.58$$

3. With a PD controller
$$C(s) = K_p(1 + T_d s)$$
the closed-loop response is
$$\frac{G(s)C(s)}{1 + G(s)C(s)} = \frac{K_p(1 + T_d)}{16s^2 + (10 + K_p T_d)s + 1 + K_p}$$
We still require $K_p = 51$. A peak overshoot of 10% corresponds to $\zeta \approx 0.6$. So we require
$$0.6 = \frac{(10 + 51T_d)/16}{2 \times 1.8}$$
which gives
$$T_d \approx 0.48$$
Note that this analysis disregards the effect of the zero that the PD controller introduces.

4. In each case the steady state closed-loop gain is
$$\frac{G(0)C(0)}{1 + G(0)C(0)} = \frac{K_p}{1 + K_p} \approx 0.98$$
so the percentage error is 2%.

Problem 6

1. The transfer function from set point r to error $e = r - y$ is

$$\frac{1}{1+G(s)C(s)} = \frac{(s+3)(s+6)}{(s+3)(s+6)+K_p}$$

which has steady sate gain $18/(18+K_p)$. So we require

$$K_p \geq 162$$

2. The forward loop transfer function is

$$K_pL(s) = \frac{K_p}{s(s+3)(s+6)}$$

Hence the root locus can be drawn in Matlab using the commands:

```
>>L = tf(1,[1 0]) * tf(1,[1 3]) * tf(1,[1 6]);
>>rlocus(L);
```

See Fig WE6.1. The pure integral action "pulls" the location of the closed-

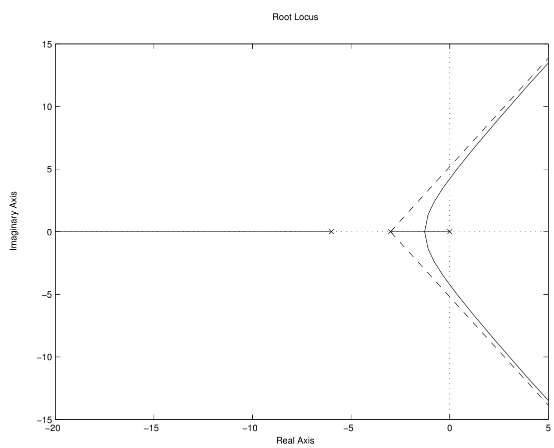


Figure WE6.1: Root locus plot of $L(s)$ with pure integral control.

loop poles to the right. For medium to high gain they will be oscillatory or unstable.

3. For a PI controller there is an extra zero at $-K_I$. If $K_I = 3$ or $K_I = 6$ it cancels one of the open-loop poles, so the closed-loop transfer function is second order:

   (a) $K_I = 3$.

$$G(s)C(s) = \frac{K_P}{s(s+6)}$$

$$\frac{GC}{1+GC} = \frac{K_P}{s^2+6s+K_P}$$

The natural frequency is $\omega_n = \sqrt{K_P}$. The damping ratio is given by $2\zeta\omega_n = 6$. Hence

$$\zeta = \frac{3}{\sqrt{K_P}}$$

(b) $K_I = 6$.

$$G(s)C(s) = \frac{K_P}{s(s+3)}$$

$$\frac{GC}{1+GC} = \frac{K_P}{s^2+3s+K_P}$$

The natural frequency is $\omega_n = \sqrt{K_P}$ as before. The damping ratio is given by $2\zeta\omega_n = 3$. Hence

$$\zeta = \frac{1.5}{\sqrt{K_P}}$$

(c) The root locus for the case (say) $K_I = 4.5$ can be drawn as follows:

```
>>G = tf(1,conv([1 3],[1 6]));
>>C = tf([1 4.5],[1 0]);
>>rlocus(G*C);
```

The root locus diagrams for the three cases are shown in Figs WE6.2 to WE6.4.

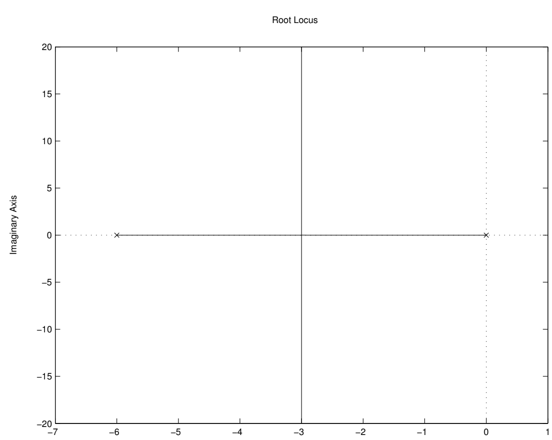


Figure WE6.2: Root locus plot with PI control, $K_I = 3$.

In each case there is a pair of complex conjugate poles when K_P is sufficiently high. When $K_I = 3$ or $K_I = 6$ the locus of these poles is a straight vertical line. When $K_I = 4.5$ the locus tends to a vertical asymptote. The vertical lines move to the right as we increase K_I.

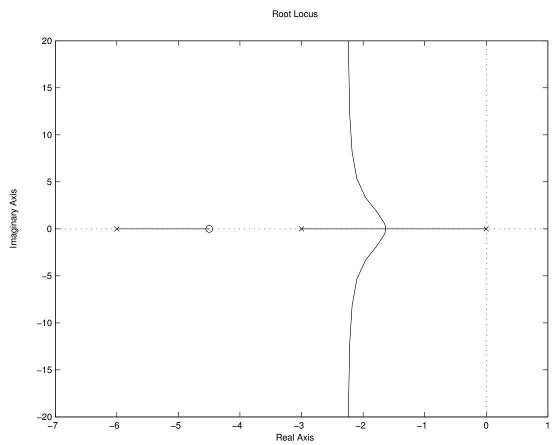


Figure WE6.3: Root locus plot with PI control, $K_I = 4.5$.

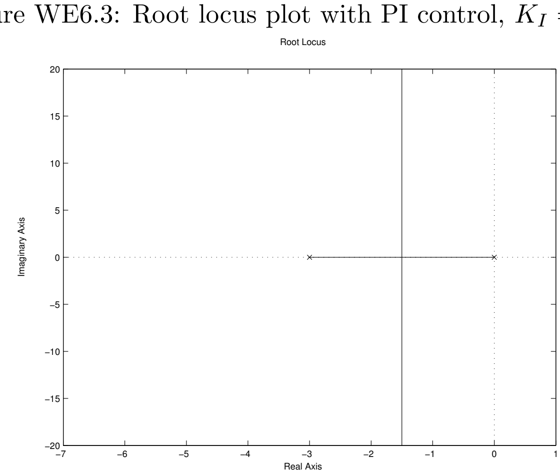


Figure WE6.4: Root locus plot with PI control, $K_I = 6$.

Problem 7

1. The two break-frequencies are $\omega = 1$ and $\omega = 100$. The frequency response is

$$G(j\omega) = \frac{10}{-\omega^2 + 99j\omega - 100}$$

We create the following table:

Frequency	Approximate transfer function	Gain (dB)	Phase
$\omega << 1$	$\frac{10}{-100}$	-20	-180^o
$\omega = 1$	$\frac{10}{(j-1)100}$	$-20-3$	$-180^o + 45^o$
$\omega = 10$	$\frac{10}{10j \times 100}$	-40	-90^o (approx)
$\omega = 100$	$\frac{10}{100j \times (100+100j)}$	$-60-3$	$-90^o - 45^o$
$\omega >> 100$	$\frac{10}{-\omega^2}$	$20 - 40\log_{10}\omega$	-180^o

The Bode plot is shown in Fig WE7.1

2. The Nyquist plot is shown in Fig WE7.2. Note that there is a small cusp near the origin—see Fig WE7.3.

3. There is one unstable open-loop pole so $P = 1$. If $K_p < 10$ the Nyquist plot does not encircle the -1, so $N = 0$ and hence the number of closed-loop unstable poles is $Z = N + P = 1$. If $K_p > 10$ then there is one anticlockwise encirclement, so $N = -1$ and $Z = 1 - 1 = 0$. Hence for $K_p > 0$ the system is stable in closed-loop.

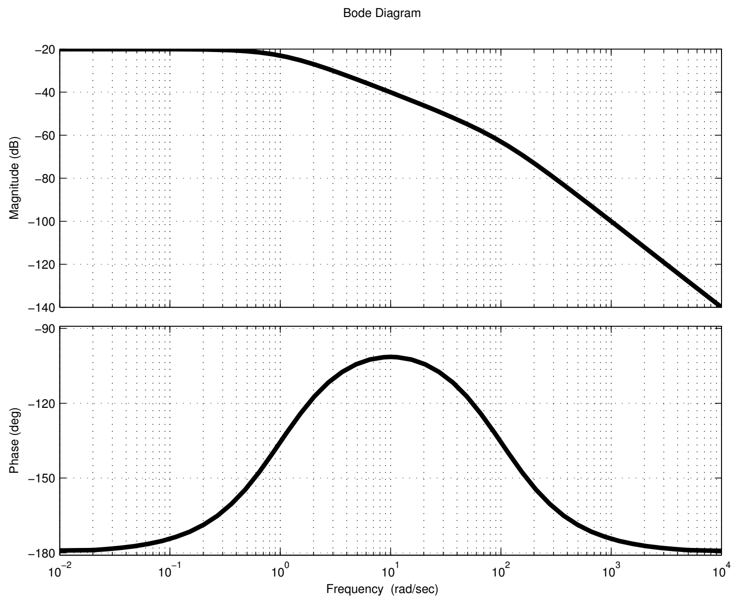


Figure WE7.1: Bode plot of plant.

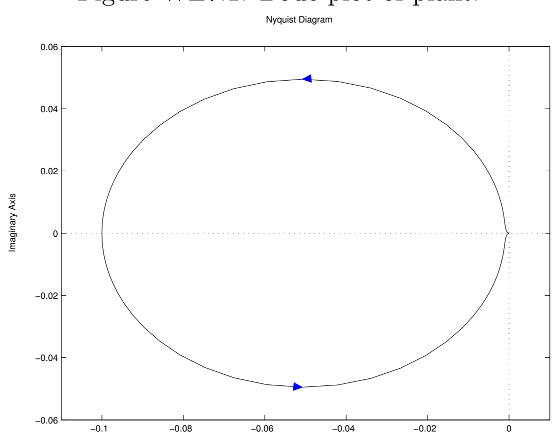


Figure WE7.2: Nyquist plot of plant.

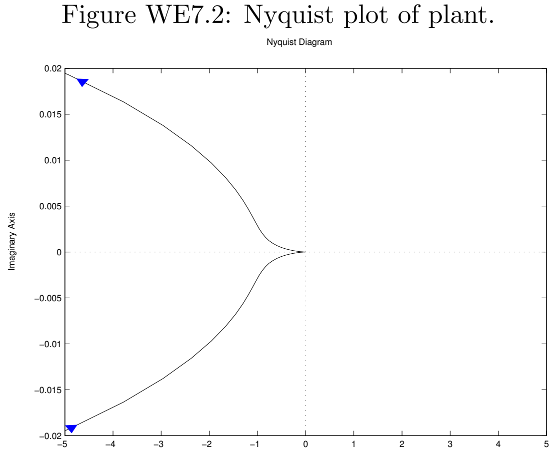


Figure WE7.3: Nyquist plot of plant near the origin.

Problem 8

1. The formula
$$M_p = e^{-\frac{\zeta}{\sqrt{1-\zeta^2}}\pi}$$
gives a peak overshoot less than 10% for a damping ratio $\zeta \geq 0.6$. This in turn corresponds to a phase margin $\phi_m = 60^o$ using the approximation
$$\zeta \approx \frac{\phi_m}{100}$$
We also require a closed-loop bandwidth of at least 3 rad/s. It should be sufficient to choose the cross-over frequency to be 3 rad/s.

From the Bode plot, the phase at 3 rad/s is approximately -150^o, so we need a phase advance of 30^o. Choose
$$\begin{aligned}\alpha &= \frac{1-\sin 30^o}{1+\sin 30^o}\\ &= 0.33\\ \omega_l &= \omega_c\sqrt{\alpha} = 3 \times \sqrt{0.33} = 1.73\\ C(s) &= k\frac{s+\omega_l}{\alpha s+\omega_l}\\ &= k\frac{s+1.73}{0.33s+1.73}\end{aligned}$$
for some k.

The gain of G at $\omega_c = 3$ is approximately $-40dB$. The gain of C with $k = 1$ at $\omega_c = 3$ is $20\log_1 0(1/0.33)/2 = 4.8$dB.

So we require k to be $40 - 4.8 = 35.2$dB. This corresponds to
$$k = 10^{35.2/20} = 57.5$$

2. A high gain controller will make the system more sensitive to high frequency noise.

Part VI
Case studies

Op-amps

Op-amps rely of feedback for their successful application. The op-amp itself is a high gain device (typically over 100dB at steady state). When configured in feedback the emergent properties are largely unaffected by noise, nonlinearities etc.

Consider the non-inverting op-amp configuration shown in Fig CS1. If the op-amp has open-loop gain A then

$$v_o(t) = A\left(v_s(t) - v_f(t)\right).$$

The feedback mechanism is a voltage divider so the feedback voltage is given as

$$v_f(t) = \frac{R_1}{R_1 + R_2} v_o(t).$$

It follows that in closed-loop $v_o(t)$ can be expressed as

$$v_o(t) = \frac{A}{1 + \beta A} v_s(t) \approx \frac{1}{\beta} v_s(t) \text{ with } \beta = \frac{R_1}{R_1 + R_2}.$$

The feedback loop is illustrated with an equivalent block diagram in Fig CS2.

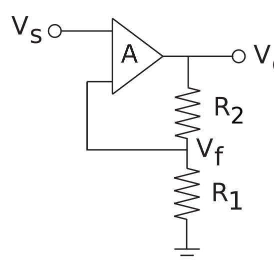


Figure CS1: Op-amp with non-inverting feedback.

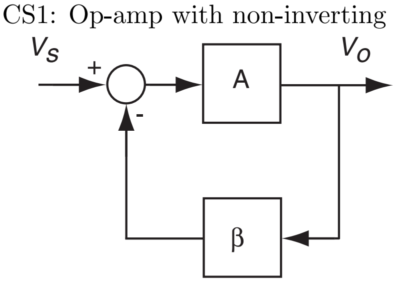


Figure CS2: Equivalent block diagram.

This is *not* a standard control system as illustrated in Fig 3 of Chapter 2. In particular there is no "controller" block distinct from the "plant". In addition, if $v_s(t)$ is our reference signal, then we are not trying to track it directly (although we *are* trying to "track" $v_s(t)/\beta$). Nevertheless, the underlying principles still apply - in fact much of classical control theory was pioneered with op-amps as the application.

Suppose, for example, the op-amp is noisy in open-loop, so the output voltage is given by

$$v_o(t) = A\left(v_s(t) + v_f(t)\right) + d(t).$$

We have labeled the noise as $d(t)$ since its effect is that of a disturbance signal we want to attenuate at the output: see Fig CS3. In closed-loop $v_o(t)$ can now be expressed as

$$v_o(t) = \frac{A}{1+\beta A}v_s(t) + \frac{1}{1+\beta A}d(t).$$

But for large A,

$$\frac{1}{1+\beta A}d(t) \approx 0,$$

so once again

$$v_o(t) \approx \frac{1}{\beta}v_s(t).$$

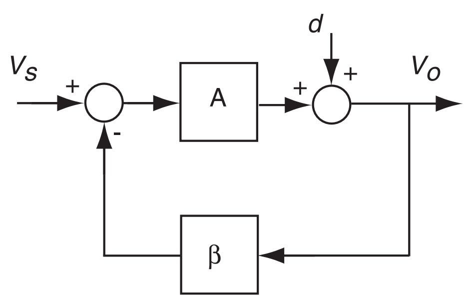


Figure CS3: Equivalent block diagram with disturbance signal representing amplifier noise.

In practice the op-amp device will have some associated dynamics, so that the open-loop gain is frequency dependent. Care must then be taken to ensure the closed-loop response is stable for all possible values of β in the range $0 < \beta < 1$.

Suppose the op-amp can be characterised with a first order transfer function

$$A(s) = \frac{M}{\tau s + 1}.$$

For large M this has phase margin $PM \approx 90^o$, with cross-over frequency $\omega_c \approx M/\tau$. See Fig CS4. In the Laplace domain the closed-loop behaviour can be described as

$$V_o(s) = \frac{M}{\tau s + 1 + \beta M}V_s(s).$$

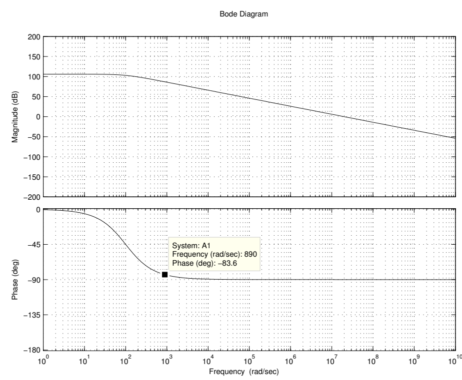


Figure CS4: Bode plot of op-amp with first order dynamics.

But if the op-amp's transfer function is second-or-more order and the dominant poles are close together, then the closed-loop will at best have a high resonance, if not be unstable. As an extreme example, suppose the op-amp can be characterised by the second order transfer function

$$A(s) = \frac{M}{(\tau s + 1)^2}.$$

For large M this has phase margin $PM \approx 0^o$, with cross-over frequency $\omega_c \approx \sqrt{M}/\tau$. The closed-loop behaviour is

$$V_o(s) = \frac{M}{(\tau s + 1)^2 + \beta M} V_s(s),$$

which has very low damping ratio.

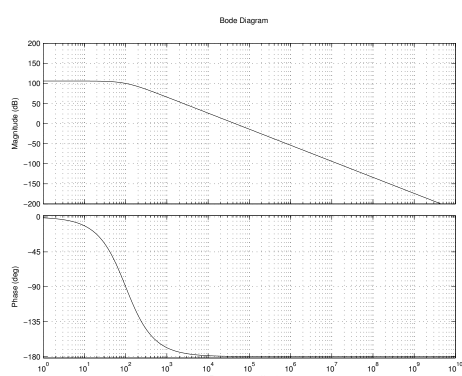


Figure CS5: Bode plot of op-amp with second order dynamics and poles identical.

Therefore op-amps are usually designed to have a dominant low frequency pole, with all other dynamics occurring at much higher frequency. Inserting appropriate internal capacitors results in so-called *pole splitting* via the *Miller effect*. A typical design might have the second pole at (minus) double the cross-over frequency so that the dynamics can be approximated by the second-order transfer function

$$V_o(s) = \frac{M}{(\tau s + 1)(\tau s/2M + 1)} V_s(s),$$

giving a phase margin $PM > 60^o$.

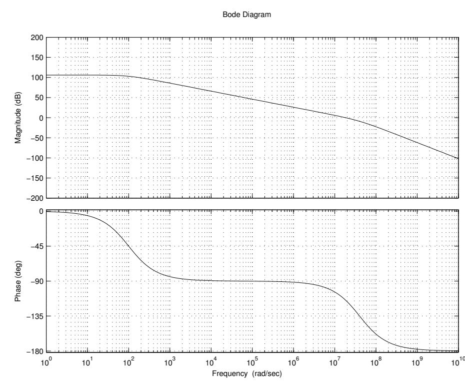


Figure CS6: Bode plot of op-amp with second order dynamics and poles far apart.

The closed-loop step responses for the first order case and for the second order case with poles far apart are both illustrated in Fig CS7. The constant β is chosen as 0.5. The closed-loop step response for the second order case where the poles are identical is illustrated in Fig CS8, again with $\beta = 0.5$ The ringing due to the low damping is clear; note the different time scale for this graph.

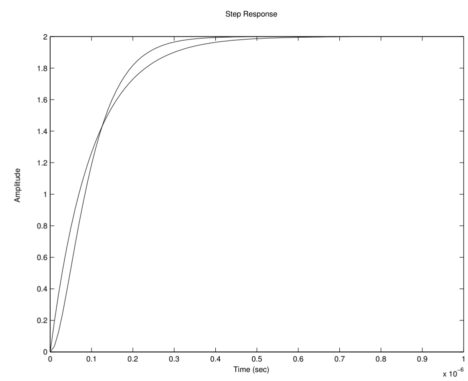


Figure CS7: Closed-loop step responses for the op-amp with first order dynamics and for the op-amp with two poles far apart. In both cases $\beta = 0.5$.

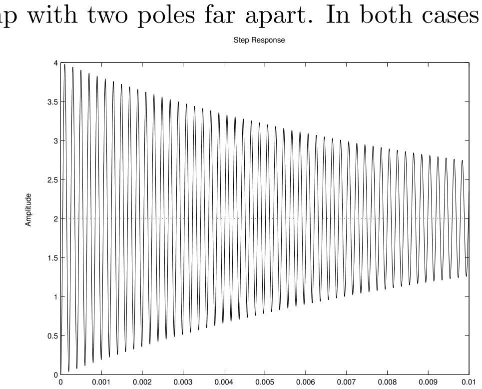


Figure CS8: Closed-loop step response for the op-amp with two identical poles and $\beta = 0.5$.

Edible oil refining

Fig CS9 shows a process plant where feedback control can be used to reduce variation*. Crude edible oil flows through two separators producing refined edible oil. Flow is measured at *Fl1* while pressure in the separators are measured at *Pl1* and *Pl2*. Originally there was no feedback around the two separators, although the flow was controlled by manipulating the automatic valve *AV1*. Typical flows and pressures are illustrated in Fig CS11. Then feedback was installed, taking measurements from *Pl1* and *Pl2* and feeding them back to the automatic valves *AV2* and *AV3*. The control loop is shown in Fig CS10—note that schematically it is identical to Fig 2 from the Introduction. After designing and tuning a controller, the resultant flows and pressures are shown in Fig CS12. We see considerably less variation in the two pressures. This in turn allows the operators to run the plant much closer to the operating limit, allowing greater yield without impinging on product quality. In a more general sense, the effect of limiting plant variation on the quality/yield tradeoff is illustrated in Fig CS13.

It was also possible to use the controller to specify the operating pressure, as illustrated in Fig CS14. In this case the controller was an advanced strategy known as MPC (model predictive control) and was implemented via a PLC (programmable logic controller), as illustrated in Fig CS15. Plant conditions are monitored via a SCADA (supervisory control and data acquisition).

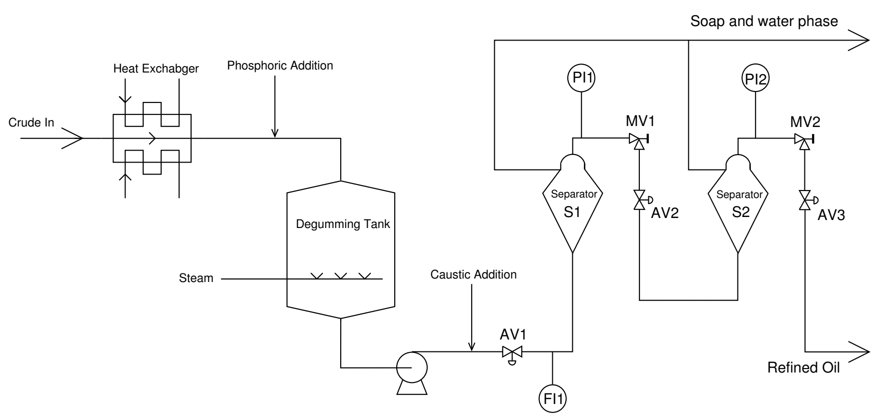


Figure CS9: Schematic of edible oil refining.

*More details are reported in Wills and Heath: Application of barrier function based model predictive control to an edible oil refining process, Journal of Process Control, 2005.

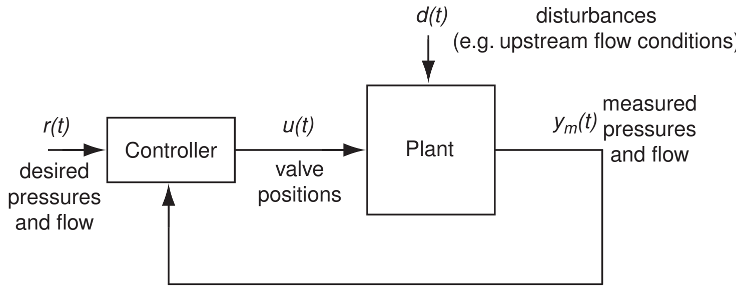


Figure CS10: Control loop for edible oil refining.

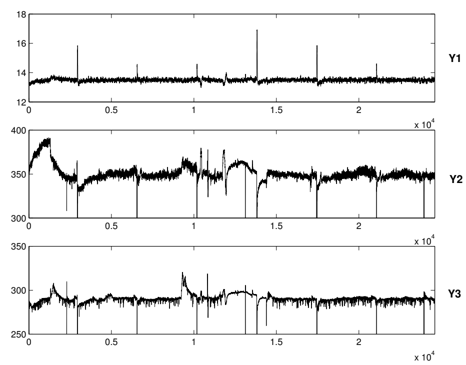


Figure CS11: Flow and two pressure measurements against time. Only the flow is controlled with feedback.

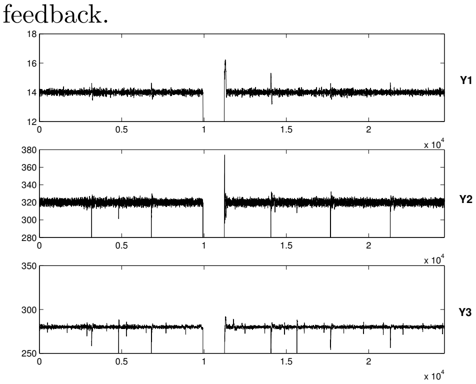


Figure CS12: Flow and two pressure measurements against time. All three variables are controlled using feedback.

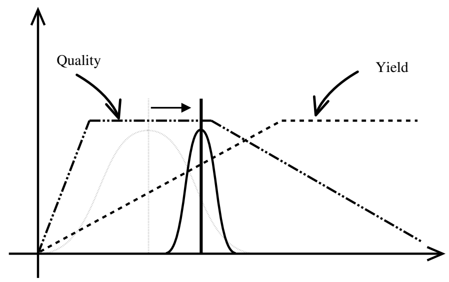


Figure CS13: Reducing variation allows an improvement in yield without affecting quality.

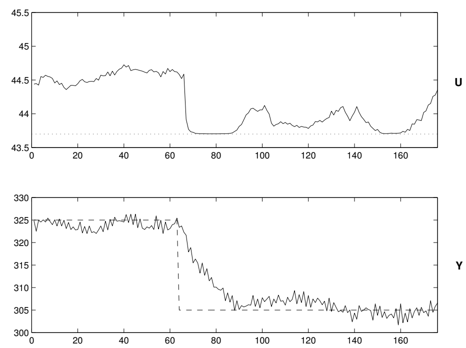


Figure CS14: Step response in pressure. The top figure shows the valve position (manipulated variable) against time while the bottom figure shows pressure (measured variable) against time.

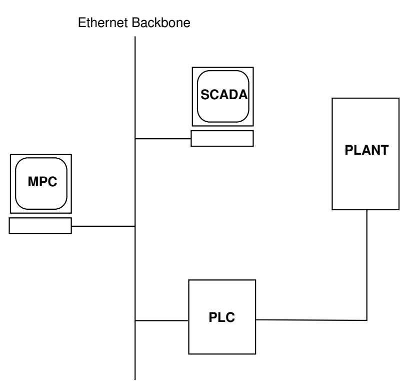


Figure CS15: Hardware configuration for the controller.

Excitation systems

This case study is based on an example in "Power System Control and Stability (2nd ed.)" by P. M. Anderson and A. A. Fouad, IEEE Press, 2003, pp327–332.

A requirement for power generation is to match the generator to the load network. It is standard to model a turbine providing power to the grid as a machine connected to an infinite bus via a transmission line. An infinite bus is a source of invariable frequency and voltage. It is therefore necessary to match the machine's speed and voltage to those of the infinite bus; this is achieved by feedback control. There are two loops to consider:

- a speed governor ensures the turbine provides the correct frequency ω_R, assumed to be $60 \times 2\pi$ rad/s here;
- the excitation system ensures the transmitted voltage is matched to that of the infinite bus.

In this case study we will consider the excitation system.

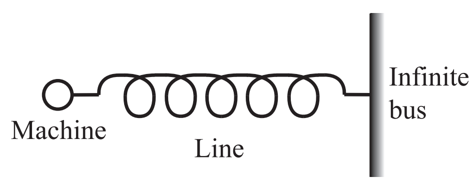


Figure CS16: Infinite bus.

Wikipedia defines excitation thus: "An electric generator or electric motor consists of a rotor spinning in a magnetic field. The magnetic field may be produced by permanent magnets or by field coils. In the case of a machine with field coils, a current must flow in the coils to generate the field, otherwise no power is transferred to or from the rotor. The process of generating a magnetic field by means of an electric current is called excitation." A schematic of a self-excited generator is shown in Fig CS17. It follows that varying the excitation voltage can be used as an actuation signal in order to vary the generated voltage. A block diagram of the feedback loop is shown in Fig CS18.

Linearised models for both the machine and the excitation system are derived by Anderson and Fouad (2003). Block diagrams of each linearised model are shown in Figs CS19 and CS20 respectively. Details of the model for the machine are beyond the scope of this case study. The main excitation dynamics are modelled as a first order unstable response

$$\frac{1}{K_E + T_E s} \text{ with } K_E \text{ negative.}$$

In addition there is a regulator amplifier with first order dynamics

$$\frac{K_A}{1 + T_A s}.$$

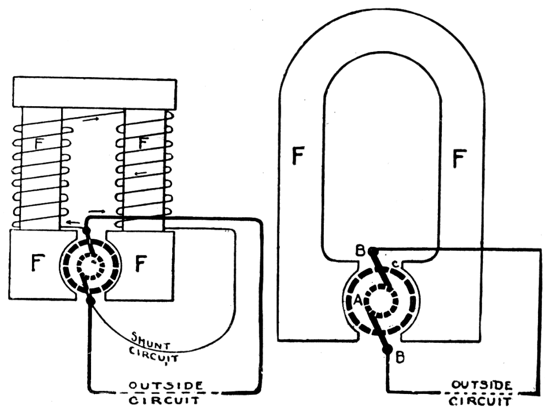


Figure CS17: From Wikipedia: "A self-excited shunt-wound DC generator is shown on the left, and a magneto DC generator with permanent field magnets is shown on the right. The shunt-wound generator output varies with the current draw, while the magneto output is steady regardless of load variations." The figure originally appeared in Hawkins Electrical Guide, Volume 1, 1917 by Theo. Audel & Co.

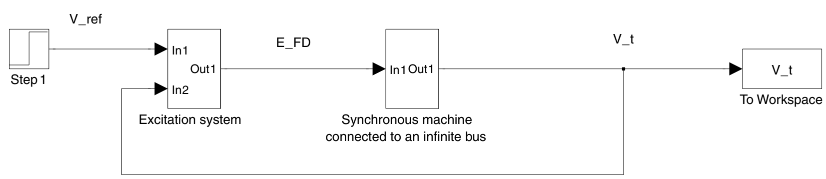


Figure CS18: Excitation system control loop. The infinite bus provides a reference voltage V_{ref} that the machine's generated voltage V_t should track as closely as possible. The excitation voltage is given by E_{FD}.

There is also the possibility of including rate feedback (i.e. pseudo derivative action) with dynamics

$$\frac{K_F s}{1 + T_F s}.$$

In Anderson and Fouad (2003) the various parameters are given the following values:

$$\begin{array}{lll} K_A = 40, & K_E = -0.17, & \\ T_A = 0.05, & T_E = 0.95, & \\ 2H = 4.74, & D = 2.0, & T'_{do} = 5.9, \\ K_1 = 1.4479, & K_2 = 1.3174, & K_3 = 0.3072, \\ K_4 = 1.8052, & K_5 = 0.0294, & K_6 = 0.5257. \end{array}$$

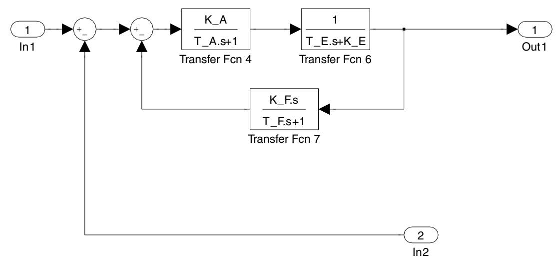


Figure CS19: Linearised model of the excitation system in block diagram form.

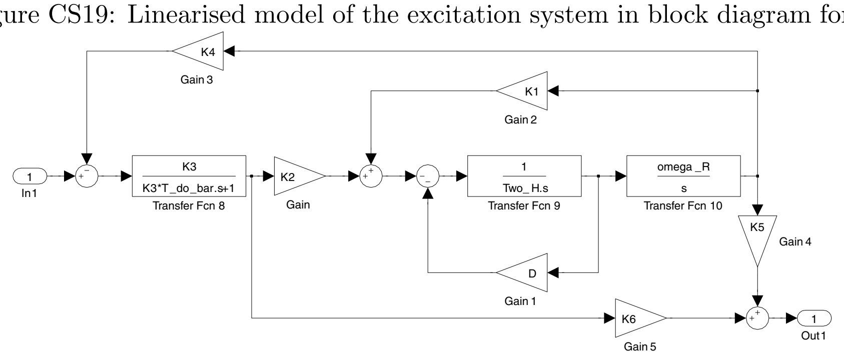


Figure CS20: Linearised model of the machine in block diagram form.

If there is no rate feedback (i.e. if $K_F = 0$) then the closed-loop system is unstable. But if rate feedback is included with $K_F = 0.04$ and $T_F = 1$ then the closed-loop system performs well. This is analysed by Anderson and Fouad (2003) as follows.

The closed-loop system may be transformed by block diagram manipulation to the configuration shown in Fig CS21. We write

$$\begin{aligned} N(s) &= \frac{N_{num}(s)}{N_{den}(s)} \\ &= \frac{K_3K_6(2Hs^2 + Ds + K_1\omega_R) - \omega_R K_2 K_3 K_5}{(K_3T'_{do}s + 1)(2Hs^2 + Ds + K_1\omega_R) - \omega_R K_2 K_3 K_4}. \end{aligned}$$

If there is no rate feedback then the loop is equivalent to unit feedback applied to the forward loop transfer function

$$L(s) = N(s) \times \frac{1}{K_E + T_E s} \times \frac{K_A}{1 + T_A s}.$$

This has zeros at $-0.21 \pm j10.45$ and poles at -20, $-0.35 \pm j10.73$, -0.2732 and $+0.1789$. The root locus is shown in Fig CS22 and in close-up in Fig CS23. There is only a small range of feedback gain for which the closed-loop system is stable, between $k = 0.0014$ and $k = 0.0509$. In particular we are interested

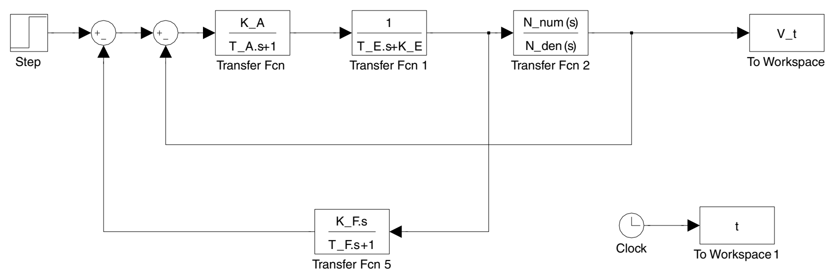


Figure CS21: Equivalent block diagram representation of the closed-loop system.

in the configuration with unit gain feedback, i.e. $k = 1$—it follows that this is unstable.

With rate control $L(s)$ becomes*

$$L(s) = \left(N(s) + \frac{K_F s}{1 + T_F s} \right) \times \frac{1}{K_E + T_E s} \times \frac{K_A}{1 + T_A s}.$$

This change in structure adds one additional pole, moves the two zeros and adds an additional two zeros. With the choice $K_F = 0.04$ and $T_F = 1$ the zeros are at $-0.40 \pm j10.69$ and $-1.20 \pm j0.83$ while the additional pole is at -1. The root locus is shown in Fig CS24 and in close-up in Fig CS25. The additional zeros ensure the loop is stable for high gain (in fact for any gain above 0.0014) and in particular the loop is stable for unity feedback.

Step responses and bode plots for the two cases may be generated via Matlab and Simulink files available on Blackboard.

*Strictly speaking $L(s)$ is no longer the forward loop transfer function but the return ratio, since there are dynamics in the feedback path. See the associated Matlab files for computation of the appropriate closed-loop transfer functions.

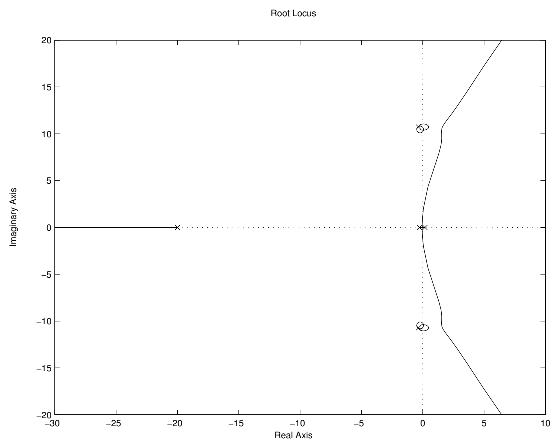


Figure CS22: Root locus diagram of the system without rate feedback.

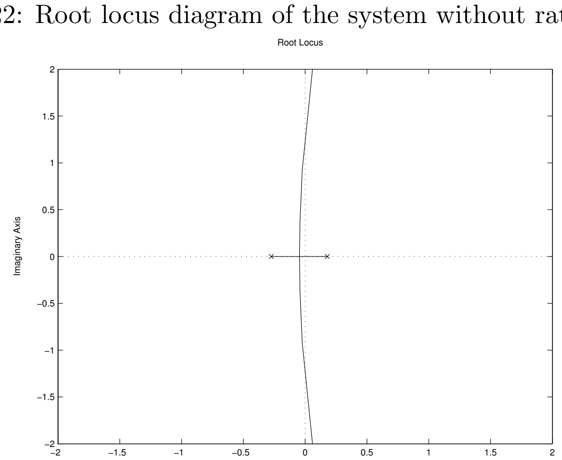


Figure CS23: Close-up of the root locus diagram of the system without rate feedback.

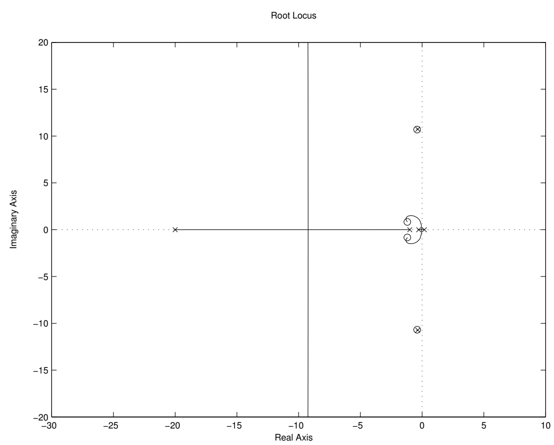


Figure CS24: Root locus diagram of the system with rate feedback.

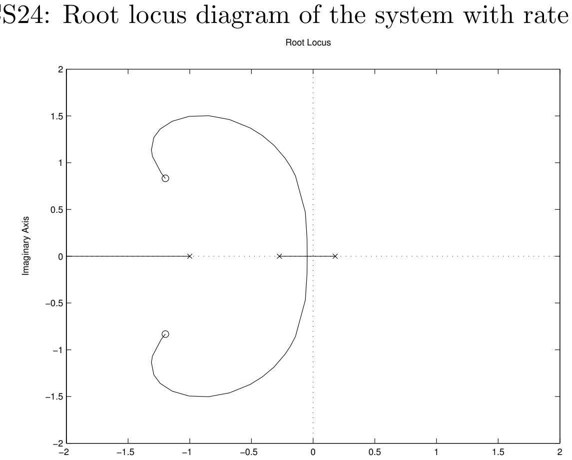


Figure CS25: Close-up of the root locus diagram of the system with rate feedback.