Feedback Systems

An Introduction for Scientists and Engineers
SECOND EDITION

Karl Johan Åström
Richard M. Murray

Version v3.0c (10 January 2014)

PRINCETON UNIVERSITY PRESS
PRINCETON AND OXFORD

Chapter Three
Feedback Principles

Feedback - it is the fundamental principle that underlies all self-regulating systems, not only machines but also the processes of life and the tides of human affairs.

A. Tustin, "Feedback", *Scientific American*, 1952, [Tus52].

This chapter presents examples that illustrate fundamental properties of feedback: disturbance attenuation, command signal following, robustness and shaping of behavior. Simple methods for analysis and design of low order systems are introduced. After reading this chapter, readers should have some insight into the power of feedback, they should know about transfer functions and block diagrams and be able to design simple feedback systems.

3.1 Mathematical Models

The fundamental properties of feedback will be illustrated using a collection of examples. We need a modest set of concepts and tools to analyze simple feedback systems: linear differential equations, transfer functions, block diagrams and block diagram algebra. In addition we need a simulation tool. In this section we will introduce some of these tools, refining them in further chapters.

Linear Differential Equations and Transfer Functions

In many practical situations, the input/output behavior of a system can be modeled by a linear differential equation of the form

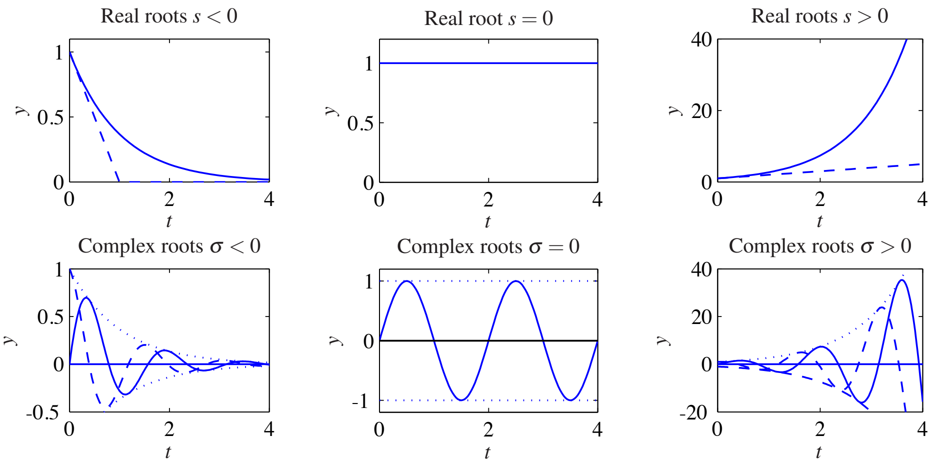
$$\frac{d^n y}{dt^n} + a_1 \frac{d^{n-1} y}{dt^{n-1}} + \cdots + a_n y = b_1 \frac{d^{n-1} u}{dt^{n-1}} + \cdots + b_n u. \tag{3.1}$$

where the coefficients a_k and b_k are real numbers. The model (3.1) is more general than the model given by equation (2.7) in Section 2.2 because the right hand side has terms with derivatives of the input. The differential equation (3.1) is characterized by two polynomials

$$a(s) = s^n + a_1 s^{n-1} + \cdots + a_n, \qquad b(s) = b_1 s^{n-1} + b_2 s^{n-2} + \cdots + b_n, \tag{3.2}$$

where $a(s)$ is the *characteristic polynomial* of the differential equation (3.1).

The solution to equation (3.1) is the sum of two terms: the *general solution to the homogeneous equation*, which does not depend on the input, and a *particular solution*, which depends on the input. The particular solution is not unique unless initial conditions or other conditions are imposed.



(a) $\mathrm{Re}\, s < 0$ (b) $\mathrm{Re}\, s = 0$ (c) $\mathrm{Re}\, s > 0$

Figure 3.1: The exponential function $y(t) = e^{st}$. The top row shows the function for real s, the bottom row shows the function for complex $s = \sigma + i\omega$. The left column shows $\mathrm{Re}\, s < 0$, the center column $\mathrm{Re}\, s = 0$ and the right column $\mathrm{Re}\, s > 0$.

The *homogeneous equation* associated with equation (3.1) is

$$\frac{d^n y}{dt^n} + a_1 \frac{d^{n-1} y}{dt^{n-1}} + \ldots + a_n y = 0, \tag{3.3}$$

and its general solution is a sum of exponentials, where the exponents are the roots s_k of the *characteristic equation* $a(s) = 0$. If there are no multiple roots s_k the solution is

$$y(t) = \sum_{k=1}^{n} C_k e^{s_k t}, \tag{3.4}$$

where C_k are arbitrary constants. The solution has n free parameters $C_1, \ldots, C_n$.

Since the coefficients a_k are real, the roots of the characteristic equation are real or complex conjugated pairs. A real root s_k of the characteristic equation corresponds to the exponential function $e^{s_k t}$. This function decreases over time if s_k is negative, it is constant if $s_k = 0$, and it increases if s_k is positive, as shown in the top row of Figure 3.1. For real roots s_k the parameter $T = 1/s_k$ is the *time constant*.

A complex root $s_k = \sigma \pm i\omega$ corresponds to the time functions

$$e^{\sigma t} \sin(\omega t), \qquad e^{\sigma t} \cos(\omega t),$$

which have oscillatory behavior, as illustrated in the bottom row of Figure 3.1. The damped sine is shown in full lines and the damped cosine in dashed lines. The dotted lines shows the envelopes, which correspond to the exponential function $\pm e^{\sigma t}$. The distance between zero crossings is π/ω, and the ratio of successive peaks is $e^{2\sigma\pi/\omega}$.

When the characteristic equation (3.4) has multiple roots, the solutions to the homogeneous equation (3.3) are

$$y(t) = \sum_{k=1}^{n} C_k(t) e^{s_k t}, \tag{3.5}$$

where $C_k(t)$ is a polynomial with degree less than the multiplicity of the root s_k. The solution (3.5) has n free parameters which can be determined from initial conditions.

Having explored the solution to the homogeneous equation, we now turn to the input-dependent part of the solution. The solution to equation (3.1) for an exponential input is of particular interest. We set $u(t) = e^{st}$ and investigate if there is a unique particular solution of the form $y(t) = G(s)e^{st}$. Assuming this to be the case, we find

$$\begin{aligned} &\frac{du}{dt} = se^{st}, && \frac{d^2u}{dt^2} = s^2 e^{st}, && \cdots && \frac{d^n u}{dt^n} = s^n e^{st} \\ &\frac{dy}{dt} = sG(s)e^{st}, && \frac{d^2y}{dt^2} = s^2 G(s) e^{st}, && \cdots && \frac{d^n y}{dt^n} = s^n G(s) e^{st}. \end{aligned} \tag{3.6}$$

The differential equation (3.1) then becomes

$$(s^n + a_1 s^{n-1} + \cdots + a_n)G(s)e^{st} = (b_1 s^{n-1} + b_2 s^{n-2} + \cdots + b_n)e^{st},$$

and hence

$$G(s) = \frac{b_1 s^{n-1} + b_2 s^{n-2} + \cdots + b_n}{s^n + a_1 s^{n-1} + \cdots + a_n} = \frac{b(s)}{a(s)}. \tag{3.7}$$

This function is called the *transfer function* of the system. It describes a particular solution to the differential equation for the input $e^{s_k t}$ and it is a convenient way to characterize the system described by the differential equation.

To further show the relation between the transfer function and the differential equation, introduce the differential operator $p = d/dt$. We have $p^2 = d^2/dt^2$ and the differential equation (3.1) can be written as

$$p^n y + a_1 p^{n-1} y + \cdots a_n y = b_1 p^{n-1} u + b_2 p^{n-1} u + \cdots + b_n u,$$

or

$$(p^n + a_1 p^{n-1} + \cdots + a_n)y = (b_1 p^{n-1} + b_2 p^{n-2} + \cdots + b_n)u.$$

The relation between the transfer function (3.7) and the differential equation (3.1) is clear: the transfer function (3.7) can be obtained by inspection from the differential equation (3.1), and conversely the differential equation can be obtained from the transfer function. The transfer function can thus be regarded as a shorthand notation for the differential equation (3.1).

To deal with oscillatory signals, like those shown in Figure 3.1, it is convenient to allow s to be a complex number. The transfer function is a function $G : \mathbb{C} \to \mathbb{C}$ that maps complex numbers to complex numbers. The roots of the characteristic equation $a(s) = 0$ are called *poles* of the transfer function. A pole s_k appears as exponent in the general solution to the homogeneous equation (3.4). The roots of the polynomial $b(s)$ are called *zeros* of the transfer function. The reason is that if

$b(s_k) = 0$ it follows that $G(s_k) = 0$, and the particular solution for the input $e^{s_k t}$ is zero. A system theoretic interpretation is that the transmission of the exponential signal $e^{s_k t}$ is blocked by the zero $s = s_k$.

The particular solution for a constant input $u(t) = 1$ is $y(t) = G(0)$. The quantity $G(0)$ is called the *zero frequency gain* or the *static gain*. The particular solution for a sinusoidal input $u = \cos(\omega t) = \operatorname{Re} e^{i\omega t}$ is

$$\begin{aligned} y(t) &= \operatorname{Re}\left(G(i\omega)\, e^{i\omega t}\right) = \operatorname{Re}\left(|G(i\omega)|\, e^{i \arg G(i\omega)}\, e^{i\omega t}\right) \\ &= |G(i\omega)| \operatorname{Re} e^{i(\arg G(i\omega) + \omega t)} = |G(i\omega)| \cos(\omega t + \arg G(i\omega)). \end{aligned}$$

The input is thus amplified by the $|G(i\omega)|$ and the phase shift between input and output is $\arg G(i\omega)$, where arg denotes the angle of a complex variable. The functions $G(i\omega)$, $|G(i\omega)|$ and $\arg G(i\omega)$ are called the *frequency response*, *gain* and *phase*. The gain and the phase are also called *magnitude* and *angle*.

The actual response to a sine or a cosine function is the sum of a particular solution and the general solution to the homogeneous equation (3.4) or (3.5). The coefficients in the general solution can be determined from the initial conditions. If all roots of the characteristic equation have negative real parts, all solutions to the homogeneous equation go to zero and the general solution converges to the particular solution as time increases.

The transfer function is a useful representation of a linear time-invariant system. It has many physical interpretations that can be exploited for analysis and design. The transfer function makes it possible to apply algebra to manipulate dynamical systems and to get insight into their behavior. The transfer function can also convey a great deal of intuition: $G(0)$ is the steady state gain for constant inputs and frequency response $G(i\omega)$ captures the steady state response to sinusoidal functions. The frequency response can be determined experimentally by exploring the response of a system to sinusoidal signals. The approximations of $G(s)$ for small and large s captures the propagation of slow and fast signals respectively. Consider for example the spring-mass system in equation (2.14), with input u and output q, which has the transfer function

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

For small s we have $G(s) \approx 1/k$. The corresponding input-output relation is $q = (1/k)u$ and the system behaves like a spring for low frequency input. For large s we have $G(s) \approx 1/(ms^2)$. The corresponding differential equation is $m\ddot{q} = u$, and the system behaves like mass (a double integrator) for high frequency inputs. Approximations of transfer functions will be discussed more in Section 8.4.

More detailed discussions of transfer functions and the frequency response will be given in later chapters, particularly in Chapter 8.

Transfer Functions and Laplace Transforms

We have defined transfer functions as a particular solution for the exponential input e^{st}. Transfer functions can also be conveniently defined using Laplace transforms.

Let $u(t)$ be the input to the system (3.1) and let $y(t)$ be the corresponding output when the initial conditions are zero. Furthermore let $U(s)$ and $Y(s)$ be the Laplace transforms of the input and the output

$$U(s) = \int_0^\infty e^{-st} u(t) dt, \qquad Y(s) = \int_0^\infty e^{-st} y(t) dt.$$

The transfer function of the system is then simply $G(s) = \dfrac{Y(s)}{U(s)}$.

Stability: The Routh-Hurwitz Criterion

When using feedback there is always the danger that the system may become unstable. It is therefore important to have a stability criterion. The differential equation (3.1) is called *stable* if all solutions of the homogeneous equation (3.3) go to zero after a perturbation. It follows from equation (3.5) that this requires that all the roots of the characteristic equation

$$a(s) = s^n + a_1 s^{n-1} + \cdots + a_n = 0,$$

have negative real parts. The *Routh-Hurwitz criterion* is a stability criterion that does not require calculation of the roots, because it gives conditions in terms of the coefficients of the characteristic polynomial..

A first order differential equation is stable if the coefficient a_1 of the characteristic polynomial is positive, since the zero of the characteristic polynomial will be $s = -a_1 < 0$. A second order polynomial is stable if the coefficients a_1 and a_2 are all positive. Since the roots are

$$s = \frac{1}{2}\left(-a_1 \pm \sqrt{a_1^2 - 4a_2}\right),$$

it is easy to verify that the real parts are negative if and only if $a_1 > 0$ and $a_2 > 0$. A third order differential equation is more complicated, but the roots can be shown to have negative real parts if the coefficients a_1, a_2 and a_3 are all positive and if

$$a_1 a_2 > a_3. \tag{3.8}$$

A fourth order differential equation is stable if all coefficients are positive and if

$$a_1 a_2 > a_3, \qquad a_1 a_2 a_3 > a_1^2 a_4 + a_3^2. \tag{3.9}$$

The Routh-Hurwitz criterion [Gan60] gives similar conditions for arbitrarily high order polynomials. Stability of a linear differential equation can thus be investigated just by analyzing the signs of various combinations of the coefficients of the characteristic polynomial.

Block Diagrams and Transfer Functions

Figure 3.2 shows a block diagram of a typical control system. If each block is modeled as a high order differential equation (3.1), we need to find the differential equation that relates the signals in the complete system. A block can be considered

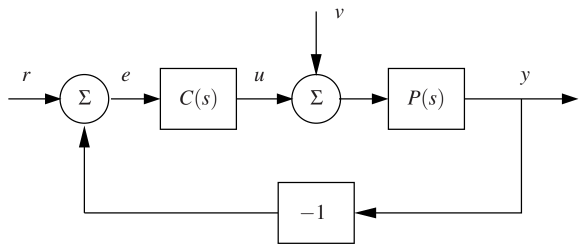


Figure 3.2: Block diagram of simple feedback system. The controller transfer function is $C(s)$ and the process transfer function is $P(s)$. The process output is y, the external signals are the reference r and the load disturbance v.

as a filter that generates the output from the input and the block is characterized by its transfer function, which is a nice shorthand notation for the differential equation describing the input-output relation.

Assume that the disturbance v is zero and that we want to find the differential equation that describes how the output y is influenced by the reference signal r. Let the transfer functions of the controller and the process be characterized by the polynomials $b_c(s)$, $a_c(s)$, $b_p(s)$ and $a_p(s)$, so that

$$C(s) = \frac{b_c(s)}{a_c(s)}, \qquad P(s) = \frac{b_p(s)}{a_p(s)}. \tag{3.10}$$

The corresponding differential equations are

$$a_c(p)\,u(t) = b_c(p)\,(r(t) - y(t)), \qquad a_p(p)\,y(t) = b_p(p)\,u(t),$$

where we have introduced $p = \frac{d}{dt}$ to simplify the notation. Multiplying the first equation by $a_p(p)$ and the second with $a_c(p)$ we find that

$$a_c(p)\,a_p(p)\,y(t) = a_c(p)\,b_p(p)\,u(t) = b_p(p)\,b_c(p)\,(r(t) - y(t)).$$

Solving for $y(t)$ gives

$$\big(a_c(p)\,a_p(p) + b_p(p)\,b_c(p)\big)\,y(t) = b_p(p)\,b_c(p)\,r(t), \tag{3.11}$$

which is the differential equation that relates the output to the reference. We see that the polynomial notation makes it easy to manipulate differential equations. Forming linear combinations of differential equations and their derivatives corresponds to polynomial multiplication.

The differential equation (3.11) corresponds to the transfer function

$$G_{yr} = \frac{b_p(s)b_c(s)}{a_c(s)a_p(s) + b_p(s)b_c(s)} = \frac{P(s)C(s)}{1 + P(s)C(s)} \tag{3.12}$$

Proceeding in the same way we obtain the following transfer functions

$$G_{ur} = \frac{C(s)}{1 + P(s)C(s)}, \qquad G_{yv} = \frac{P(s)}{1 + P(s)C(s)}, \qquad G_{uv} = \frac{-P(s)C(s)}{1 + P(s)C(s)}. \tag{3.13}$$

By using polynomials and transfer functions the relations between signals in a feedback system can be obtained by algebra. The transfer functions relating two signals can be obtained from the block diagram by inspection. The denominator is always $1+P(s)C(s)$ and the numerator is a product of the transfer functions between the signals, for example, the transfer functions from disturbance v to control u in Figure 3.2 are $P(s)$, -1 and $C(s)$.

3.2 Using Feedback to Improve Disturbance Attenuation

Reducing the effects of disturbances is a primary use of feedback. It was used by James Watt to make steam engines run at constant speed in spite of varying load and by electrical engineers to make generators driven by water turbines deliver electricity with constant frequency and voltage. Feedback is commonly used in process control, in machine tool control, in power generation, and for engine and cruise control in cars. In humans the pupillary reflex is used to make sure that the light intensity of the retina is reasonably constant in spite of large variations in the ambient light. The human body exploits feedback to keep body temperature, blood pressure and other important variables constant. Keeping variables close to a desired, constant reference values in spite of disturbances is called a *regulation problem*.

Disturbance attenuation will be illustrated by control of a process whose dynamics can be approximated by a first order system. A block diagram of the system is shown in Figure 3.2. Since we will focus on the effects of a load disturbance v we will assume that the reference r is zero. The transfer functions G_{yv} and G_{uv} relating the output y and the control u to the load disturbance are given by equation (3.13). For simplicity we will assume that the process is modeled by the first order differential equation

$$\frac{dy}{dt}+ay=bu, \qquad a\geq 0, \quad b>0.$$

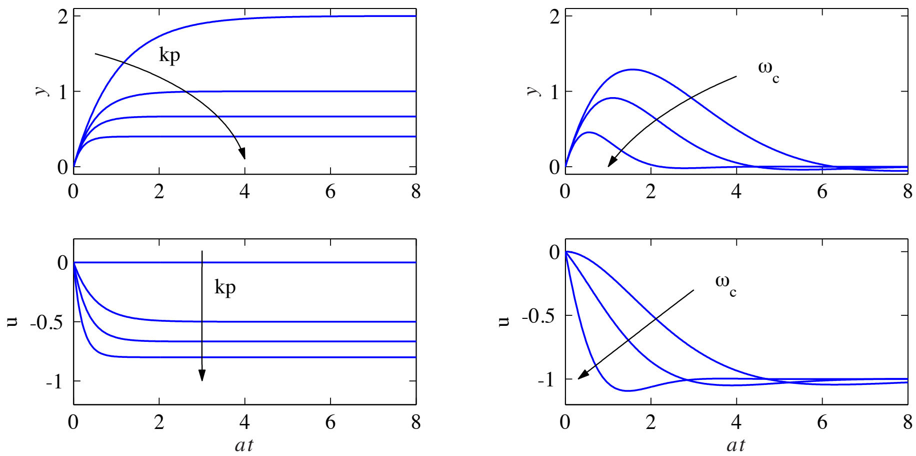
A straightforward calculation gives the transfer function

$$P(s)=\frac{b}{s+a}. \tag{3.14}$$

A first order system is a reasonable model of a physical system if the storage of mass, momentum or energy can be captured by one state variable. Typical examples are the velocity of a car on a road, the angular velocity of rotating system and the level of a tank.

Proportional Control

We will first investigate the case of proportional (P) control, when the control signal is proportional to the output error: $u=k_pe$, see Section 1.4. The controller transfer function is then $C(s)=k_p$. The effect of the disturbance on the output is



(a) Proportional control (P)

(b) Proportional-Integral control (PI)

Figure 3.3: Responses of open and closed loop system with proportional control (a) and PI control (b). The process transfer function is $P = 2/(s+1)$. The controller gains for proportional control are $k_p = 0$, 0.5, 1 and 2. The PI controller is designed using equation (3.20) with $\zeta = 0.707$ and $\omega_c = 0.707$, 1 and 2, which gives the controller parameters $k_p = 0$, 0.207, 0.914 and $k_i = 0.25$, 0.50 and 2.

described by the transfer function

$$G_{yv}(s) = \frac{b}{s + a + bk_p}.$$

The relation between the disturbance v and the output y is thus given by the differential equation

$$\frac{dy}{dt} + (a + bk_p)y = bv.$$

The closed loop system is stable if $a + bk_p > 0$. A constant disturbance $v = v_0$ then gives an output that approaches the steady state value

$$y_0 = G_{yv}(0) = \frac{v_0}{a + bk_p} v_0,$$

exponentially with the time constant $T = 1/(a + bk_p)$. Without feedback $k_p = 0$ and a constant disturbance v_0 thus gives the steady state error v_0/a. The error decreases when using feedback if $k_p > 0$.

We have thus shown that a constant disturbance gives an error that can be reduced by feedback using a proportional controller. The error decreases with increasing controller gain. Figure 3.3 shows the responses for a few values of controller gain k_p.

Proportional-Integral (PI) Control

The PI controller, introduced in Section 1.4, is described by

$$u(t) = k_p e(t) + k_i \int_0^t e(\tau) d\tau. \qquad (3.15)$$

To determine the transfer function of the controller we differentiate, hence

$$\frac{du}{dt} = k_p \frac{de}{dt} + k_i e$$

and we find by inspection that the transfer function is $C(s) = k_p + k_i/s$. To investigate the effect of the disturbance v on the output we use the block diagram in Figure 3.2 and we find by inspection that the transfer function from v to y is

$$G_{yv}(s) = \frac{P(s)}{1+P(s)C(s)} = \frac{s}{s^2+(a+bk_p)s+bk_i}. \qquad (3.16)$$

The relation between the load disturbance and the output is thus given by the differential equation

$$\frac{d^2y}{dt^2} + (a+bk_p)\frac{dy}{dt} + bk_i y = \frac{dv}{dt}. \qquad (3.17)$$

Notice that, since the disturbance enters as a derivative in the right hand side, a constant disturbance gives no steady state error. The same conclusion can be drawn from the observation that $G_{yv}(0) = 0$. Compare with the discussion of integral action and steady state error in Section 1.4.

To find suitable values of the controller parameters k_p and k_i we consider the characteristic polynomial of the differential equation (3.17),

$$a_{cl}(s) = s^2 + (a+bk_p)s + bk_i. \qquad (3.18)$$

We can assign arbitrary roots to the characteristic polynomial by choosing the controller gains k_p and k_i, and we choose controller parameters that give the characteristic polynomial

$$(s+\sigma+i\,\omega)(s+\sigma-i\,\omega) = s^2+2\sigma s+\sigma^2+\omega^2. \qquad (3.19)$$

This polynomial has roots at $s = -\sigma \pm i\,\omega$. The general solution to the homogeneous equation is then a linear combination of the terms

$$e^{-\sigma t}\sin(\omega t), \qquad e^{-\sigma t}\sin(\omega t),$$

which are damped sine and cosine functions, as shown in the lower left plot in Figure 3.1. The coefficient σ determines the decay rate and the parameter ω gives the frequency of the decaying oscillation. Identifying coefficients of equal powers of s in the polynomials (3.18) and (3.19) gives

$$k_p = \frac{2\sigma - a}{b}, \qquad k_i = \frac{\sigma^2+\omega^2}{b}. \qquad (3.20)$$

Instead of parameterizing the closed loop system in terms of σ and ω it is common practice to use the *undamped natural frequency* $\omega_c = \sqrt{\sigma^2+\omega^2}$ and the *damping ratio* $\zeta = \sigma/\omega_c$. The closed loop characteristic polynomial is then

$$a_{cl}(s) = s^2 + 2\sigma s + \sigma^2 + \omega^2 = s^2 + 2\zeta\omega_c s + \omega_c^2.$$

This parameterization has the advantage that ζ determines the shape of the response and that ω_c gives the response speed.

Figure 3.3 shows the output y and the control signal u for $\zeta = 0.707$ and different values of ω_c. Proportional control gives a steady-state error which decreases with increasing controller gain. With PI control the steady-state error is zero. Both the decay rate and the peak error decrease when the design parameter ω_c is increased. Larger controller gains give smaller errors and control signals that react faster to the disturbance. To achieve the benefits of large control gains the model must be accurate over wide frequency ranges.

In summary, we find that the analysis gives a simple way to find the parameters of PI controllers for processes whose dynamics can be approximated by a first order system. The technique can be generalized to more complicated systems but the controller will be more complex.

Unmodeled Dynamics

The design we have made indicate that there are no limits to the performance that can be achieved. Figure 3.3 shows that arbitrarily fast response can be obtained simply by making ω_c sufficiently large. In reality there are of course limitations to what can be achieved. One reason is that the controller gains increase with ω_c, the integral gain is $k_i = \omega_c^2/b$. A large value of ω_c thus gives large controller gains and the actuator may saturate. Another reason is that the model (3.14) is a simplification, it is only valid in a given frequency range. If the model is instead

$$P(s) = \frac{b}{(s+a)(1+sT)}, \tag{3.21}$$

where the term $1+sT$ represents dynamics in sensors or actuators or other dynamics that was neglected when deriving (3.14), so-called *unmodeled dynamics*, the closed loop characteristic polynomial for the closed loop system becomes

$$a_{cl} = s(s+a)(1+sT) + k_p s + k_i = s^3 T + s^2(1+aT) + 2\zeta\omega_c s + \omega_c^2.$$

It follows from the Routh-Hurwitz criterion (3.8) that the closed loop system is stable if $\omega_c^2 T < 2\zeta\omega_c(1+aT)$ or if

$$\omega_c < \frac{2\zeta(1+aT)}{T}.$$

The frequency ω_c and the achievable response time are thus limited by the unmodeled dynamics represented by T.

3.3 Using Feedback to Follow Command Signals

Another major application of feedback is to make a system output follow a command signal. It is called the *servo problem*. Cruise control and steering of a car, tracking a satellite with an antenna or a star with a telescope are some examples. Other examples are high performance audio amplifiers, machine tools and industrial robots.

To illustrate command signal following we will consider the system in Figure 3.2 where the process is a first order system and the controller is a PI controller. The transfer functions of the process and the controller are

$$P(s) = \frac{b}{s+a}, \qquad C(s) = \frac{k_p s + k_i}{s}. \tag{3.22}$$

Since we will focus on command signal following we will neglect the load disturbance, $v = 0$. It follows from equation (3.12) that the transfer function from the command signal r to the output y is

$$G_{yr}(s) = \frac{P(s)C(s)}{1+P(s)C(s)} = \frac{bk_p s + bk_i}{s + (a+bk_p)s + bk_i}. \tag{3.23}$$

Since $G_{yr}(0) = 1$ it follows that $r = y$ when r and y are constant, independent of the parameters a and b. The output is thus equal to the reference in steady state, a useful property of controllers with integral action.

To determine suitable values of the controller parameters k_p and k_i we proceed as in Section 3.2 by choosing controller parameters that makes the closed-loop characteristic polynomial

$$a_{cl}(s) = s^2 + (a+bk_p)s + bk_i \tag{3.24}$$

equal to $s^2 + 2\zeta\omega_c s + \omega_c^2$. Identifying coefficients of equal powers of s in these polynomials give

$$k_p = \frac{2\zeta\omega_c - a}{b}, \qquad k_i = \frac{\omega_c^2}{b}. \tag{3.25}$$

Notice that integral gain increases with the square of ω_c. Figure 3.4 shows the output signal y and the control signal u for different values of the design parameters ζ and ω_c. The response time decreases with increasing ω_c and the initial value of the control signal also increases because it takes more effort to move rapidly. The overshoot decreases with increasing ζ. For $\omega_c = 2$, the design choice $\zeta = 1$ gives a short settling time and a response without overshoot.

It is desirable that the output y will track the reference r for time-varying references. This means that we would like the transfer function $G_{yr}(s)$ to be close to 1 for large frequency ranges.With the chosen design we have

$$G_{yr}(s) = \frac{(2\zeta\omega_c - a)s + \omega_c^2}{s^2 + 2\zeta\omega_c s + \omega_c^2}.$$

It is thus desirable to have a large ω_c to be able to track fast changes in the reference signal. Plotting the frequency response of G_{yr} gives a quantitative representation

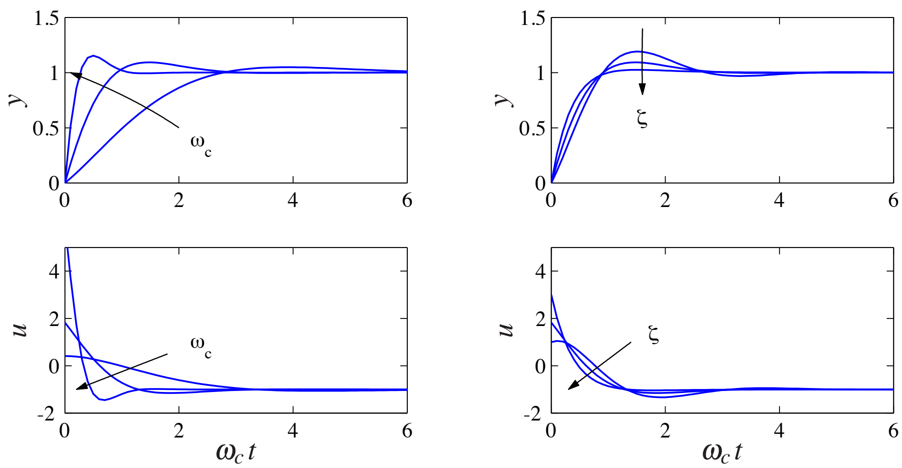


Figure 3.4: Responses to a step change in the command signal for different values of the design parameters. The left figure shows responses for fixed $\zeta = 0.707$ and $\omega_c = 1$, 2 and 5. The right figure shows responses for $\omega_c = 2$ and $\zeta = 0.5$, 0.707, and 1. The process parameters are $a = b = 1$.

of the tracking abilities.

Controllers with Two Degrees of Freedom

The control law in Figure 3.2 has *error feedback* because the control signal u is generated from the error $e = r - y$. With proportional control, a step in the reference signal gives an immediate step change in the control signal. This rapid reaction can be an advantageous, but it may give a large overshoot, which can be avoided by a replacing the PI controller in equation (3.15) with

$$u(t) = k_p\big(\beta r(t) - y(t)\big) + k_i \int_0^t (r(\tau) - y(\tau))d\tau, \tag{3.26}$$

In this modified PI algorithm, the proportional action only acts on the fraction β of the reference signal. The transfer functions from reference r to u and from output y to u are

$$C_{ur}(s) = \beta k_p + \frac{k_i}{s}, \qquad C_{uy}(s) = k_p + \frac{k_i}{s} = C(s). \tag{3.27}$$

The controller (3.26) is is called a controller with *two degrees of freedom* since the transfer functions $C_{ur}(s)$ and $C_{uy}(s)$ are different.

A block diagram of a closed loop system with a PI controller having two degrees of freedom is shown in Figure 3.5. Let the process transfer function be $P(s) = b/(s+a)$. The transfer functions from reference r and disturbance v to output y are

$$G_{yr}(s) = \frac{b\beta k_p s + bk_i}{s^2 + (a + bk_p)s + bk_i}, \qquad G_{yv}(s) = \frac{s}{s^2 + (a + bk_p)s + bk_i}. \tag{3.28}$$

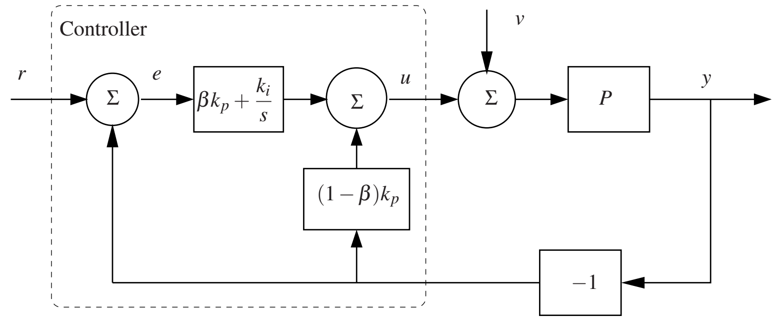


Figure 3.5: Block diagram of a closed-loop system with a PI controller having two degrees of freedom.

Comparing with the corresponding transfer function for a controller with error feedback in equations (3.16) and (3.23) we find that the responses to the load disturbances are the same but the response to reference values are different.

A simulation of the closed loop system for $a = 0$ and $b = 1$ is shown in Figure 3.6. The figure shows that the parameter β has a significant effect on the responses. Comparing the system with error feedback ($\beta = 1$) with the system with smaller values of β we find that using a system with two degrees of freedom gives the same settling time with less overshoot and gentler control actions.

The example shows that command signal response can be improved by using a controller architecture having two degrees of freedom. In Section 3.6 we will show that the responses to command signals and disturbances can be completely separated by using a more general system architecture. To use a system with two degrees of freedom both the reference signal and the command signals must be available. There are situations where only the error signal can be measured, typical examples are DVD players, optical memories and atomic force microscopes.

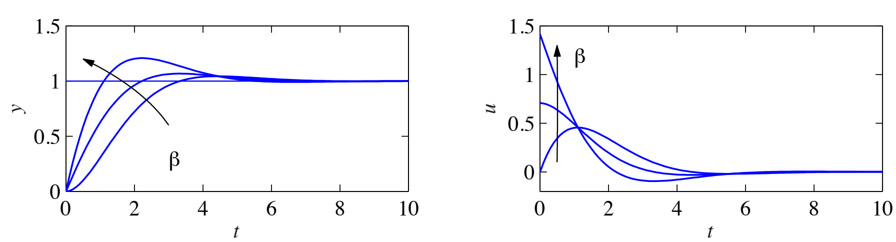


Figure 3.6: Response to a step change in the command signal for a system with a PI controller having two degrees of freedom. The process transfer function is $P(s) = 1/s$ and the controller gains are $k_p = 1.414$, $k_i = 1$ and $\beta = 0$, 0.5 and 1.

3.4 Using Feedback to Provide Robustness

Feedback can be used to make good systems from poor components. The development of the electric feedback amplifier for transmission of telephone signals is an early example [Ben93]. Design of amplifiers with constant linear gain was a major problem. The basic component in the amplifier was the vacuum tube, which was nonlinear and time varying. A major accomplishment was the invention of the feedback amplifier. The idea is to close a feedback loop by arranging a feedback loop around the vacuum tube, which gives a closed loop system with a linear input/output relation with constant gain.

The idea to use feedback to linearize input/output characteristics and to make it insensitive to process variations is common. The recipe is to localize the source of the variations and to close feedback loops around them. This idea is used extensively to obtain linear amplifiers and actuators, and to reduce effects of friction in mechanical systems. We will illustrate with a simple model of an electronic amplifier.

A Nonlinear Amplifier

Consider an amplifier with a static, nonlinear input/output relation with considerable variability as illustrated in Figure 3.7a. The nominal input/output characteristics is shown in heavy dashed line and examples of variations in thin lines. The nonlinearity in the figure is actually

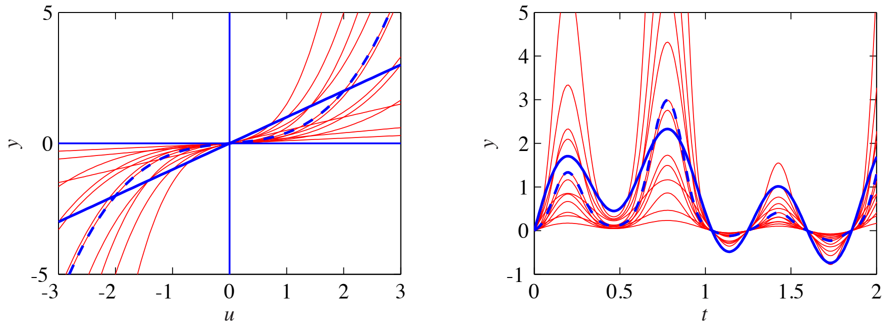
$$y = f(u) = \alpha(u + \beta u^3), \qquad -3 \le u \le 3. \tag{3.29}$$

The nominal values corresponding to the dashed line are $\alpha = 0.2$ and $\beta = 1$. The variations of the parameters α and β are in the ranges $0.1 \le \alpha \le 0.5$, $0 \le \beta \le 2$. The responses of the system to the input

$$r(t) = \sin(t) + \sin(\pi t) + \sin(\pi^2 t). \tag{3.30}$$

are shown in Figure 3.7b. The desired response $y = u$ is shown in heavy full lines and responses for a range of parameters are shown in thin lines. The nominal response of the nonlinear system is shown in heavy dashed lines. It is distorted due to the nonlinearity. Notice in particular the heavy distortion both for small and large signal amplitudes. The behavior of the system is clearly not satisfactory.

The behavior of the system can be improved significantly by introducing feedback. A block diagram of a system with a simple integral controller is shown in Figure 3.8. Figure 3.9 shows the behavior of the closed loop system with the same parameter variations as in Figure 3.7. The input/output plot in Figure 3.9a is a scatter plot of the inputs and the outputs of the feedback system. The input output relation is practically linear and close to the desired response. There is some variability because the feedback introduces dynamics. Figure 3.9b shows the responses to the reference signal, notice the dramatic improvement compared with Figure 3.7b. Figure 3.9c shows the tracking error.



(a) Input/output relationship (b) Output response

Figure 3.7: Response of a nonlinear system. The left figure shows the input/output relation of the open-loop system and the right figure shows responses to the input signal. The nominal response of the nonlinear system is shown in dashed thick lines and the variations of the response due to parameter variations is shown in thin lines. The ideal response is shown in full thick lines.

Analysis

Analysis of the closed loop system is difficult because it is nonlinear. We can however obtain significant insight by using approximations. We first observe that the system is linear when $\beta = 0$. In other situations we will approximate the nonlinear function by a straight line around an operating point $u = u_0$. The slope of the nonlinear function at $u = u_0$ is $f'(u_0)$ and we will approximate the process with a linear system with the gain $f'(u_0)$. The transfer functions of the process and the controller are

$$P(s) = b, \qquad b = \alpha(1 + 3\beta u_0^2), \qquad C(s) = \frac{k_i}{s}, \tag{3.31}$$

where u_0 denotes the operating condition. The process gain $b = \alpha(1 + 3\beta u_0^2)$ is in the range 0.1–27.5 depending on the values of α, β and u_0. It follows from equation (3.13) that the transfer functions relating the output y and the error e to the reference signal are

$$G_{yr}(s) = \frac{bk_i}{s + bk_i}, \qquad G_{er}(s) = 1 - G_{yr} = \frac{s}{s + bk_i}. \tag{3.32}$$

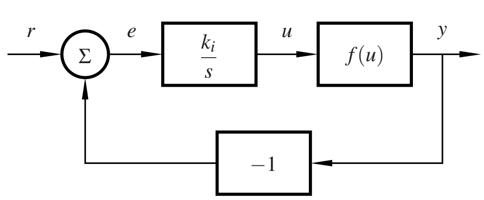


Figure 3.8: Block diagram of a nonlinear system with integral feedback.

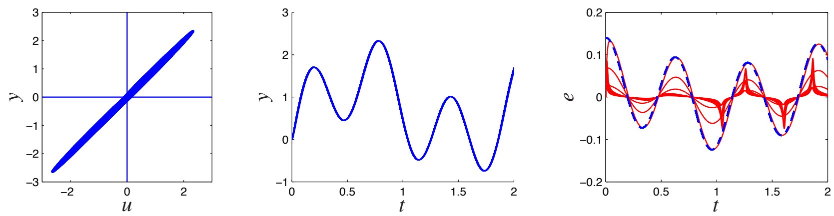


Figure 3.9: Responses of the system with integral feedback. The left plot, is a scatter plot of inputs and outputs. The center plot shows the response of the closed loop system to the input signal r, and the right plot shows the control error. The parameter variations are the same as in Figure 3.7. Notice the dramatic improvement compared to Figure 3.7b

The closed loop system is a first-order system with the pole $s = -bk_i$ and the time constant $T = 1/(bk_i)$. The integral gain is chosen as $k_i = 1000$. The closed loop pole ranges from 100 rad/s to 2.75×10^4 rad/s, which is fast compared to the high frequency component 9.9 rad/s of the input signal.

The error for the approximated system is described by the differential equation

$$\frac{de}{dt} = -bk_i e + \frac{dr}{dt}, \qquad \frac{dr}{dt} = \cos(t) + \pi\cos(\pi t) + \pi^2\cos(\pi^2 t). \tag{3.33}$$

The fast frequency component of the input (3.30) has the frequency $\pi^2 = 9.8$; it is slower than the process dynamics for all parameter variations and we have

$$e \approx \frac{1}{bk_i}\frac{dr}{dt}. \tag{3.34}$$

This estimate is shown as the dashed line in Figure 3.9c. The peak error is approximately $\pi^2/(bk_i) = 0.1$ when $bk_i = 100$. The error deviates significantly from the estimate (3.34) because the system is nonlinear. It follows from (3.34) that the error is smaller when $bk_i > 50$, which explains why the dashed line in Figure 3.9c is an upper bound.

This analysis has given a simple procedure to design an integrating controller for a system whose dynamics can be approximated by a static model. Design is essentially the choice of a single parameter: the integral gain of the controller. The closed loop transfer function from reference to output is given by equation (3.32) where integral gain is $k_i = 1/(bT_{cl})$ where T_{cl} is the desired time constant of the closed loop system. The integral gain is inversely proportional to T_{cl} and the largest integral gain is limited by unmodeled dynamics.

The example illustrates that feedback can be used to design an amplifier that has practically linear input/output relation even if the basic amplifier is nonlinear with strongly varying characteristics.

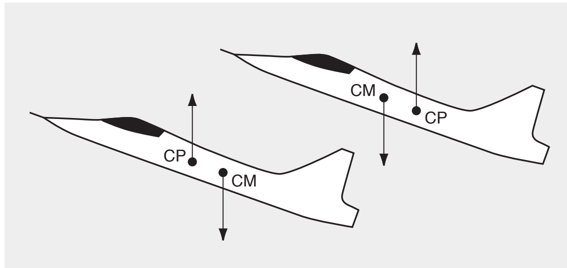


Figure 3.10: Schematic digram of two aircrafts. The aircraft at the top is stable because it has the center of pressure CP behind the center of mass CM. The aircraft at the bottom is unstable because the positions of center of mass and center of pressure are reversed.

3.5 Using Feedback to Shape Behavior

The regulation and servo problems discussed in Sections 3.2 and 3.3 are classical applications of feedback. In Section 3.4 it was shown that feedback can be used to obtain stable linear input/output behavior for a nonlinear system with strong variability. In this section we will show how feedback can be used to shape the dynamic behavior of a system.

Collision avoidance is a useful behavior of moving robots. Feedback is used in automobiles to create behaviors that avoid locking brakes, skids and collision with pedestrians. Feedback is used to make the dynamic behavior of airplanes invariant to operating conditions. Feedback is also an essential element of human balancing and locomotion.

Bacteria use simple feedback mechanisms to search for areas where there is high concentration of food or light. The principle is to sense a variable and to make exploratory moves to see if the concentration increases. A similar mechanism can be used to avoid harmful substances. Optimization is also used in computer systems to maintain maximum throughput of servers.

Stabilization

Stabilizing an unstable system is a typical example of how feedback can be used to change behavior. Many systems are naturally unstable. The ability to stand upright, walk and run has given humans many advantages but it requires stabilization. Stability and maneuverability are conflicting goals in vehicle design. The ship designer Minorsky realized that there was a trade-off between maneuverable and stability and he emphasized that a stable ship is difficult to steer. The Wright Flyer, which was maneuverable but unstable, inspired Sperry to design an autopilot. Feedback has been used extensively in aircraft, from simple systems for stability augmentation to systems that provide full autonomy.

Military airplanes gain significant competitive advantage by making them unstable. Schematic pictures of two airplanes are shown in Figure 3.10. The positions of the center of mass CM and the center of pressure CP are key elements. To be stable the center of pressure must be behind of the center of mass. The center of pressure of an aircraft shifts backwards when a plane goes supersonic. If the plane is stable at subsonic speeds it becomes even more stable at supersonic speeds because of the long distance between CM and CP. Large forces and large control surfaces are then required to maneuver the airplane and the plane will be more sluggish. A more balanced design is obtained by placing the center of pressure in front of the center of mass at subsonic speeds. Such a plane will have superior performance, but it is unstable at subsonic speeds, i.e. at takeoff and landing. When the control system is mission critical there are strong demands on the robustness and reliability of the control system.

Stabilization of an inverted pendulum is a prototype example. Consider the cart–pendulum discussed in Examples 2.1 and 2.2. Neglecting damping, assuming that the cart is much heavier than the pendulum and assuming that the tilt angle θ is small, equation (2.10) can be approximated by the differential equation

$$J_t\ddot{\theta} - mgl\theta = u. \tag{3.35}$$

The transfer function of the open loop system is

$$G_{\theta u} = \frac{1}{J_t s^2 - mgl}, \qquad a_{cl}(s) = J_t s^2 - mgl.$$

The system is unstable because it has a pole $s = \sqrt{mgl/J_t} = \omega_0$ in the right half plane. It can be stabilized with a *proportional-derivative (PD) controller* that has the transfer function

$$C(s) = -k_d s - k_p. \tag{3.36}$$

The closed-loop characteristic polynomial is

$$a_{cl}(s) = J_t s^2 + k_d s + (k_p - mgl),$$

and all of its roots are in the left half plane if $k_p > mgl$.

One way to find controller parameters is to choose the controller gains so that the characteristic polynomial has natural frequency ω_c and damping ratio ζ, hence

$$k_d = 2\zeta\omega_c J_t, \qquad k_p = J_t\omega_c^2 + mgl.$$

Choosing $\omega_c^2 = \omega_0$ moves the open loop poles from $\pm\omega_0$ to $-\zeta\omega_0 \pm i\sqrt{1-\zeta^2}\omega_0$. The controller gains are then $k_p = 2mgl$ and $k_d = 2\zeta\sqrt{mglJ_t}$. The control law (3.36) stabilizes the pendulum but is does not stabilize the motion of the cart. To do this it is necessary to introduce feedback from cart position and cart velocity.

The Segway

The Segway discussed in Example 2.1 is essentially a pendulum on a cart and can be modeled by equation (2.9) with an added torque τ on the pendulum that is exerted by the person leaning on the platform. Hence

$$M_t\ddot{p} - ml\ddot{\theta} = u, \qquad -ml\ddot{p} + J_t\ddot{\theta} - mgl\theta = -\tau,$$

where u is the force generated by the motor and τ is the torque generated by the lean of the rider. Since the Segway is similar to the inverted pendulum on a cart, we will explore if the feedback (3.36) can be used to stabilize the system. The closed loop system is described by

$$M_t\ddot{p} - ml\ddot{\theta} = -k_d\dot{\theta} - k_p\theta, \qquad -ml\ddot{p} + J_t\ddot{\theta} - mgl\theta = -\tau.$$

Elimination of $\ddot{p}$ gives

$$(M_tJ_t - m^2l^2)\ddot{\theta} + mlk_d\dot{\theta} + ml(k_p - M_tg)\theta = -M_t\tau$$

Since $M_tJ_t > m^2l^2$, the differential equation is stable if $k_p > M_tg$. To find out how the tilt influences the forward motion we eliminate θ and its derivatives and we find that the transfer function relating forward acceleration to τ is

$$G_{\ddot{p}\tau} = -\frac{-m^2l^2s^2 + mlk_ds + mlk_p}{(M_tJ_t - m^2l^2)s^2 + mlk_ds + ml(k_p - M_tg)} \approx \frac{k_p}{k_p - M_tg}$$

where the approximation is valid for small s. The feedback (3.36), which stabilizes the Segway, thus creates a behavior where the acceleration is proportional to the torque τ. Stabilizing the tilt angle thus gives a mechanism where the forward acceleration is proportional to the forward tilt torque.

Impedance Control and Haptics

Changing behavior of a mechanical system is common in robotics and haptics. Position control is not sufficient when industrial robots are used for grinding, polishing and assembly. The robot can be brought into proximity with the workspace by position control but to carry out the operations it is desirable to shape how the force depends on the distance between the tool and the workspace. A spring-like behavior is an example. The general problem is to create a behavior specified by a given differential equation between force and motion, a procedure called *impedance control*. Similar situations occur in teleoperation in hazardous environment or in telesurgery. In this situation the workpiece is operated remotely using a joystick. It is useful for the operator to have some indication of the forces between the tool and the workpiece. This can be accomplished by generating a force on the operators joystick that mimics the force on the workpiece.

Figure 3.11 shows two haptic input devices. The systems are pen-like with levers or gimbals containing angle sensors and force actuation. By sensing position and orientation, and generating a force depending on position and velocity, it is possible to create a behavior that simulates touching real or virtual objects. Forces that simulate friction and surface structure can also be generated.

We illustrate the principle with a joystick having a low friction joint. Let J be the moment of inertia, and let the actuation torque and the external torque from the operator be T and T_a, respectively. The equation of motion is

$$J\frac{d^2\theta}{dt^2} = T + T_a.$$

Figure 3.11: Haptic devices, the left figure shows the PHANTOMTM and the right a system is developed by Quanser.

By measuring the angle θ and its first two derivatives we can create the feedback

$$T_a = k_p(\theta_r - \theta) - k_d \frac{d\theta}{dt} - k_a \frac{d^2\theta}{dt^2}.$$

The closed loop system is then

$$(J + k_a)\frac{d^2\theta}{dt^2} + k_d \frac{d\theta}{dt} + k_p(\theta - \theta_r) = T.$$

The feedback has thus provided virtual inertia k_a, virtual damping k_d and virtual spring action k_p. If no torque is applied the joystick will assume the orientation given by the reference signal θ_r. If a the user applies a torque the joystick will behave like a damped spring-mass system.

3.6 Feedback and Feedforward

Feedback and feedforward have complementary properties. Feedback only acts when there are deviations between the actual and the desired behavior, feedforward acts on planned behavior. Some of the properties are summarized in Table 3.1. In economics feedback represents a market economy and feedforward a plan economy. Feedback and feedforward can be combined to improve response to command signals and to reduce the effect of disturbances that can be measured.

Table 3.1: Properties of feedback and feedforward

Feedback	Feedforward
Closed loop	Open loop
Acts on deviations	Acts on plans
Robust to model uncertainty	Sensitive to model uncertainty
Risk for instability	No risk for instability
Sensitive to measurement noise	Insensitive to measurement noise

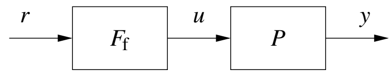


Figure 3.12: Obtaining a system with the desired transfer function $G_m(s)$ by pure feedforward. The process transfer function is $P(s)$, the feedforward compensator has the transfer function $G_{ff}(s) = P^{-1}(s)G_m(s)$

Feedforward and System Inversion

To explore feedforward control we will first investigate command signal following. Consider the system modeled by the differential equation (3.1):

$$\frac{d^n y}{dt^n} + a_1 \frac{d^{n-1}y}{dt^{n-1}} + \ldots + a_n y = b_1 \frac{d^{n-1}u}{dt^{n-1}} + \ldots + b_n u.$$

Assume that we want to find a control signal u that gives the response y_r. It follows from (3.1) that the control signal is given by

$$b_1 \frac{d^{n-1}u}{dt^{n-1}} + \ldots + b_n u = \frac{d^n y_r}{dt^n} + a_1 \frac{d^{n-1}y_r}{dt^{n-1}} + \ldots + a_n y_r, \tag{3.37}$$

This equation is called the *inverse* of equation (3.1) because it is obtained by exchanging inputs and outputs. If the transfer function of the original system is $P(s)$, the transfer function of the inverse system is simply $P^{-1}(s)$.

Consider a system with the transfer function $P(s)$, and assume that we want to find a feedforward controller so that the response to command signals is given by the transfer function $F_m(s)$, as shown in Figure 3.12. The feedforward compensator is then

$$G_{ff}(s) = P^{-1}(s)G_m(s) \tag{3.38}$$

because $P(s)G_{ff}(s) = G_m(s)$. Design of a feedforward compensator is thus closely related to system inversion.

There are problems with system inversion since the inverse may require differentiations and it may be unstable. If $b_1 \neq 0$ we have $G^{-1}(s) \approx s/b_1$ for large s, which implies that to obtain a bounded control signal we must require that the reference signal has a smooth first derivative. If $b_1 = 0$ we must similarly require that the reference signal has a smooth second derivative.

Difficulties with Feedforward Compensation

Let the process and the desired response have the transfer functions

$$P(s) = \frac{1}{(s+1)^2}, \qquad F_m(s) = \frac{\omega_c^2}{s^2 + 2\zeta\omega_c s + \omega_c^2}.$$

The feedforward transfer function is then given by equation (3.38), hence

$$F_f = P^{-1}(s)F_m(s) = \frac{\omega_c^2(s+1)}{s^2 + 2\zeta\omega_c s + \omega_c^2}.$$

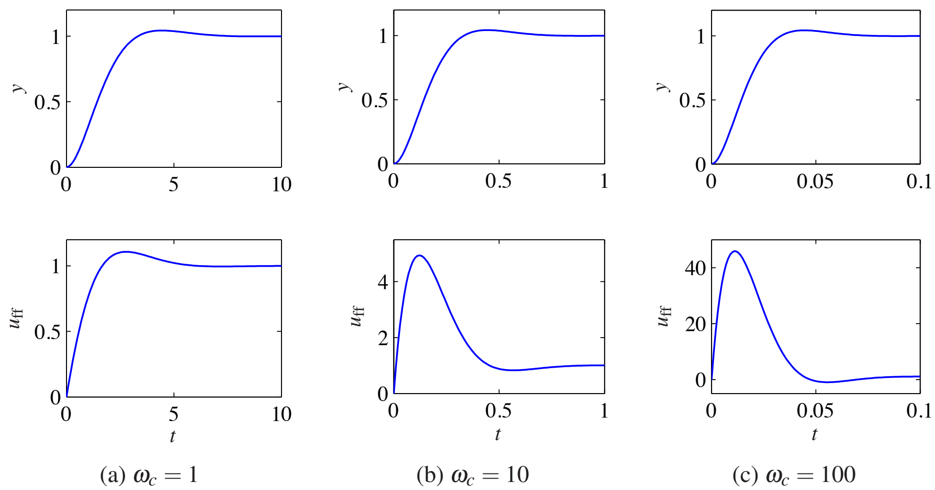


Figure 3.13: Outputs y (top plots) and feedforward signals (lower plots) for $\omega_c = 1$ (left) 10 (center) and 100 right. The outputs are identical apart from the time scale, but the control signals required to generate the output differs significantly. Notice that the largest value increases significantly with increasing ω_c.

Figure 3.13 shows the outputs y and the feedforward signals u_{ff} for different values of ω_c. Notice that large control signals required naturally are required to obtain fast responses. Achievable performance is thus limited by the size of admissible control signals.

Another difficulty with feedforward is that the inverse process dynamics may be unstable, and the feedforward signal may then be infinitely large as time increases. To have a bounded feedforward signal it follows from equation (3.38) that the transfer function G_{yr} must have the same right half-plane zero as the process. Right-half plane process zeros thus limit what can be achieved with feedforward.

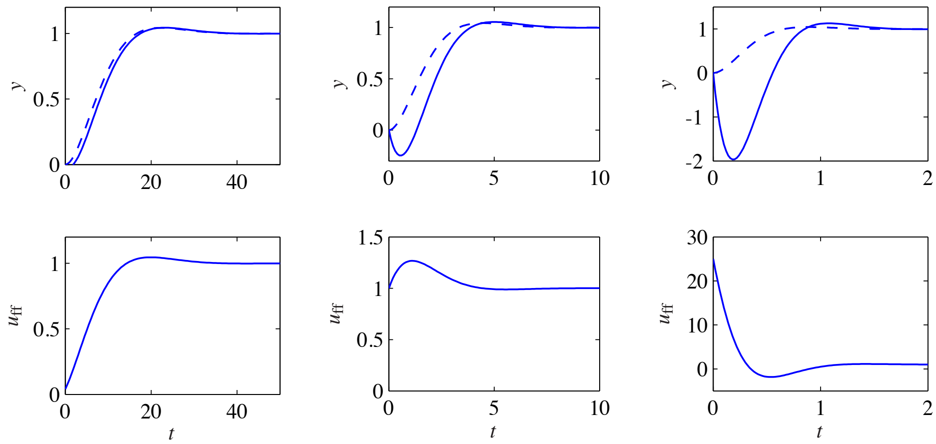
Let the process and the desired response be characterized by the transfer functions

$$P(s) = \frac{1-s}{(s+1)^2}, \qquad F_{\text{m}}(s) = \frac{\omega_c^2(1-s)}{s^2 + 2\zeta\omega_c s + \omega_c^2}.$$

Since the process has a right half plane zero at $s = 1$ the inverse model is unstable and it follows from equation (3.38) that we must require that the transfer function of the desired response has the same zero. Equation (3.38) gives the feedforward transfer function

$$F_{ff}(s) = \frac{\omega_c^2(s+1)^2}{s^2 + 2\zeta\omega_c s + \omega_c^2}. \tag{3.39}$$

Figure 3.14 shows the outputs y and the feedforward signals for different values of ω_c. The response to the command signal always goes in the wrong direction initially because of the right half plane zero at $s = 1$. This effect, called *inverse response*, is barely noticeable if the response is slow ($\omega_c = 1$) but increases with increasing response speed. For $\omega_c = 5$ the undershoot is more than 200%. The



(a) $\omega_c = 0.2$ (b) $\omega_c = 1$ (c) $\omega_c = 5$

Figure 3.14: Outputs y (top plots) and feedforward signals (lower plots) for $\omega_c = 1$ (left) 10 (center) and 100 (right) for a unit step command in the reference signal. The dashed curve shows the response that could be achieved if the process did not have the right half plane zero.

right half plane zero thus severely limits the response time. The behavior of the control signal changes qualitatively with ω_c. To understand what happens we note that the zero frequency gain of the transfer function (3.39) is $F_f(0) = 1$ and that the high frequency gain is $F_f(\infty) = \omega_c^2$. The initial value of the control signal is thus $u_{ff}(0) = F_{ff}(\infty)\omega_c^2 r$ and the final value is $u_{ff}(\infty) = F_f(0)r$. For $\omega_c = 0.2$ the control signal grows from 0.04 to the final value 1 with a small overshoot. For $\omega_c = 1$ the control signal starts from 1 has an overshoot and settles on the final value. For $\omega_c = 5$ the control signals starts at 25 and decays towards the final value 1 with an undershoot.

Sensitivity to Process Variations

The transfer function from reference r to output y of a system with pure feedforward control is

$$G_{yr}(s) = P(s)F_f(s). \tag{3.40}$$

To find the sensitivity of G_{yr} to variations in the process transfer function p we take logarithm of equation (3.40) and differentiate to obtain

$$\frac{dG_{yr}(s)}{G_{yr}(s)} = \frac{dP(s)}{P(s)}. \tag{3.41}$$

The relative variations in the system with feedforward is the same as those in the process and is thus sensitive to process variations.

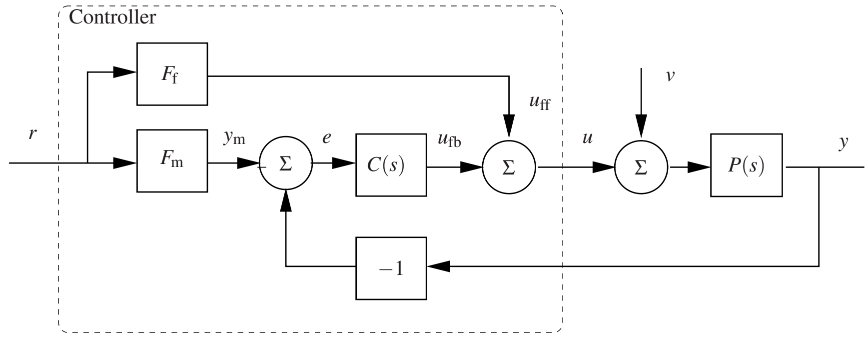


Figure 3.15: Block diagram of a closed loop system with a controller having an architecture with two degrees of freedom. The signals y_m and u_{ff} are generated by feedforward from the reference r. The feedback controller $C(s)$ acts on the control error $e = y_m - y$ and generates the feedback control signal u_{fb}.

Combining Feedforward with Feedback

Since feedback can give systems that are robust to model uncertainties it seems natural to combine feedforward with feedback. The architecture of such a controller is shown in the block diagram of Figure 3.15. The controller has three blocks representing the feedback transfer function $C(s)$ and the feedforward transfer functions F_m and F_f. The desired response to command signals is $y_m = F_m r$, which is the reference signal to the controller $C(s)$. The feedback signal u_{fb} is generated by the feedback controller $C(s)$ that acts on the error $e = y_m - y$. The feedforward signal $u_{ff} = F_f r$ is designed to make the process give a response that is close to the desired output y_m. The control signal is the sum of the feedforward u_{ff} and the feedback signals u_{fb}. The controller architecture in Figure 3.15 is highly intuitive. The feedforward signal u_{ff} generates the ideal output $y = y_m$, the error is then zero and the feedback signal u_{fb} is zero. All control is thus handled by the feedforward action. If there are modeling errors, the error e will not be zero and the feedback controller $C(s)$ will make corrections.

The controller in Figure 3.15 is a generalization of the controller with two degrees of freedom introduced in Section 3.3 (see Figure 3.5). A nice property is that it gives a separation of command signal following, robustness and disturbance attenuation. Command signal following is dealt with by design of the feedforward transfer functions F_m and F_f. Robustness and disturbance attenuation is dealt with by design of the feedback transfer function $C(s)$.

The transfer function from r to y for the system in Figure 3.15 is

$$G_{yr} = \frac{P(F_f + CF_m)}{1+PC} = F_m + \frac{PF_f - F_m}{1+PC}. \tag{3.42}$$

The transfer function G_{yr} is equal to G_m if F_m and F_f are chosen so that

$$F_m(s) = P(s)F_f(s). \tag{3.43}$$

The process transfer function imposes limitations on the choice of the transfer

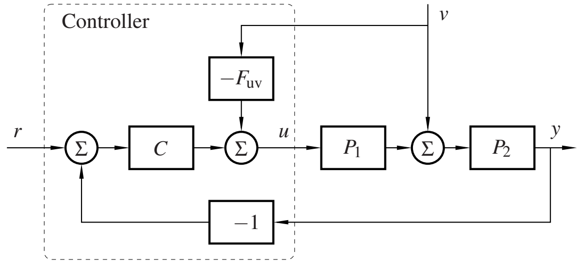


Figure 3.16: Using Feedback and feedforward to reduce the effect of a disturbance v that can be measured.

function F_m. The transfer functions relating the output to disturbances are

$$G_{yv}(s) = \frac{P(s)}{1+P(s)C(s)}, \qquad G_{uw}(s) = -\frac{C(s)}{1+P(s)C(s)}. \tag{3.44}$$

These transfer functions do not depend on the feedforward feedforward transfer functions. The system controller shown in Figure 3.15 admits a decoupling of the response to command signals to response to disturbances. The feedback controller $C(s)$ is designed to give robustness to process variations and attenuation of load disturbances. The desired response to command signals is obtained by design of the feedforward signal generator.

To investigate the effect of process uncertainty we consider the case of small variations. Taking the logarithm of G_{yr} in equation (3.44) gives

$$\log G_{yr} = \log P + \log\left(F_f + CF_m\right) - \log\left(1+PC\right).$$

Differentiating with respect to P gives the following expression for the sensitivity

$$\frac{dG_{yr}(s)}{G_{yr}(s)} = \frac{dP(s)}{P(s)} - \frac{C(s)dP(s)}{1+P(s)C(s)} = \frac{1}{1+P(s)C(s)}\frac{dP(s)}{P(s)}. \tag{3.45}$$

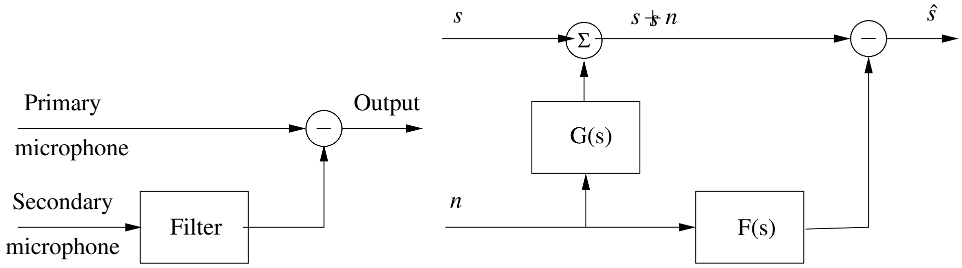
The relative error in the closed loop transfer function $G_{yr}(s)$ can thus be smaller than the relative error in the process transfer function $P(s)$ for frequencies where $P(s)C(s)$ is large. Compare with the corresponding expression (3.41) for pure feedforward. It is thus useful to combine feedback and feedforward.

Using Feedforward to Attenuate of Measured Disturbances

Feedforward can also be used to mitigate the effect of disturbances that can be measured. Such a scheme is shown in Figure 3.16. The process transfer function P is composed of two factors, $P = P_1P_2$. A measured disturbance v enters at the input of process section P_2. The measured disturbance is fed to the process input via the feedforward transfer function F_{uv}.

The transfer function from the disturbance v to process output y is

$$G_{yv}(s) = \frac{P_2(1-P_1F_{uv})}{1+PC}. \tag{3.46}$$



(a) Schematic diagram (b) Block diagram

Figure 3.17: Schematic and block diagrams for noise cancellation.

This equation shows that there are two ways of reducing the disturbance. The transfer function $1 - P_1 F_{uv}$ can be made small by a proper choice of the feedforward transfer function F_{uv}. In feedback compensation the effect of the disturbance is instead reduced by making the loop transfer function PC large. Feedforward to makes the error small by subtraction. Feedback instead makes the error small by dividing with $1 + PC$. An immediate consequence is that feedforward is more sensitive than feedback since we are trying to match two terms. Feedback gives better robustness but there is a risk of instability. Feedback and feedforward are therefore complementary, and we again can see that it is useful to combine them.

Feedforward is most effective when the disturbance v enters early in the process. This occurs when most of the dynamics are in process section P_2. When $P_1 = 1$, and therefore $P_2 = P$, the ideal feedforward compensator is realizable, and the effects of the disturbance can be eliminated from the process output y. On the other hand, when the dynamics enter late in the process, so that $P_1 \approx P$, the effects of the disturbance are seen in the process output y at the same time as they appear in the feedforward signal. In this case, there is no advantage of using feedforward compared to feedback.

Noise cancellation is a common example of use of feedforward to cancel effects of disturbances. Consider, for example, a pilot that has to communicate in a noisy cabin. The environmental noise will seriously deteriorate the communication because the pilots microphone will pick up ambient noise. The noise can be reduced significantly by using two microphones as illustrated in Figure 3.17. The primary microphone is directed towards the pilot. It picks up the pilots voice and ambient noise. The second microphone is directed away from the pilot and it picks up the ambient noise. The effect of the noise can be reduced by filtering the signal from the secondary microphone and subtracting it from the signal from the primary microphone. A block diagram of the system is shown in Figure 3.17b. The transfer function $G(s)$ represents the dynamics of the acoustic transmission from the secondary microphone to the first microphone. The transfer function $F(s)$ is the transfer function of the filter. To cancel the effect of the noise the transfer function $F(s)$ should be close to $G(s)$. Since the noise transmission depends on the situation, for example how the pilot turns his head, it is common to let the filter be adaptive so that it can adjust, as described later in Example 4.16. Noise cancellation has many applications, in headphones, to create noise-free spaces by active noise control, or to measure electrocardiogram and heartbeat of mother and fetus.

3.7 Further Reading

The books by Bennett [Ben79, Ben93] and Mindel [Min02, Min08] give interesting perspective on the development of control. Much of the material touched upon in this chapter is classical control see [JNP47], [CM51] and [Tru55].The notion of controllers with two degrees of freedom was introduced by Horowitz [Hor63]. The analysis will be elaborated in the rest of the book. Transfer functions and other descriptions of dynamics are discussed in Chapters 5 and 8, methods to investigate stability in Chapter 9. The simple method to find parameters of controllers based on matching of coefficients of the closed loop characteristic polynomial is developed further in Chapters 6, 7 and 12. Feedforward control is discussed in Section 7.5.

Exercises

3.1 Let $y \in \mathbb{R}$ and $u \in \mathbb{R}$. Solve the differential equations

$$\frac{dy}{dt} + ay = bu, \qquad \frac{d^2y}{dt^2} + 2\frac{dy}{dt} + y = 2\frac{du}{dt} + u.$$

Determine the responses to a unit step $u(t) = 1$ and the exponential signal $u(t) = e^{st}$ when the initial condition is zero. Derive the transfer functions of the systems.

3.2 Let $y_0(t)$ be the response of a system with the transfer function $G_0(s)$ to a given input. The transfer function $G(s) = (1 + sT)G_0(s)$ has the same zero frequency gain but it has an additional zero at $z = -1/T$. Let $y(t)$ be the response of the system with the transfer function $G(s)$, show that

$$y(t) = y_0(t) + T\frac{dy_0}{dt}, \tag{3.47}$$

Then cosider the system with the transfer function

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

which has unit zero-frequency-gain ($G(0) = 1$). Use the result in equation (3.47) to explore the effect of the zero $s = -a$ on the step response of a system.

3.3 Consider a process and a controller modeled by

$$\frac{dy}{dt} + ay = bu, \qquad u = k_p(r - y) + k_i\int_0^t (r(\tau) - y(\tau))d\tau,$$

where r is the reference, u the control variable and y the process output. Derive a differential equation relating the output y to the difference by direct manipulation of the equations. Draw a block diagram of the system. Derive the transfer functions of the process and the controller. Compute the transfer function from reference r to output y of the closed loop system. Make the derivations both by direct manipulation of the system equations and by polynomial algebra. Compare the results with a direct determination of the transfer functions by inspection of the block diagram.

3.4 The dynamics of the pupillary reflex is approximated by a linear system with the transfer function

$$P(s) = \frac{0.2(1-0.1s)}{(1+0.1s)^3}.$$

Assume that the nerve system that control the pupil opening is be modeled as a proportional controller with the gain k. Use Routh-Hurwitz theorem to determine the largest gain that gives a stable closed loop system.

3.5 A simple model for the relation between speed v and throttle u for a car is given by the transfer function

$$G_{vu} = \frac{b}{s+a}$$

where $b = 1\ \mathrm{m/s^2}$ and $a = 0.025\ \mathrm{rad/s}$, see Appendix A.3. The control signal is normalized to the range $0 \le u \le 1$. Design a PI controller for the system that gives a closed loop system with the characteristic polynomial

$$a_{cl}(s) = s^2 + 2\zeta\omega_c s + \omega_c^2.$$

What are the consequences of choosing different values of the design parameters ζ and ω_c. Use you judgment to find suitable values. Hint: Investigate maximum acceleration and maximum velocity for step changes in the velocity reference.

3.6 Consider the feedback system in Figure 3.2. Let the disturbance $v = 0$, $P(s) = 1$ and $C(s) = k_i/s$. Determine the transfer function G_{yr} from reference r to output y. Also determine how much G_{yr} is changed when the process gain changes by 10%.

3.7 The calculations in Section 3.2 can be interpreted as a design method for a PI controller for a first order system. A similar calculation can be made for PID control of the second order system. Let the transfer functions of the process and the controller be

$$P(s) = \frac{b}{s^2 + a_1 s + a_2}, \qquad C(s) = k_p + \frac{k_i}{s} + k_d s.$$

Show that the controller parameters

$$k_p = \frac{(1+2\alpha\zeta)\omega_c^2 - a_2}{b}, \qquad k_i = \frac{\alpha\omega_c^3}{b}, \qquad k_d = \frac{(\alpha+2\zeta)\omega_c - a_1}{b}.$$

give a closed loop system with the characteristic polynomial

$$(s^2 + 2\zeta\omega_c s + \omega_c^2)(s + \alpha\omega_c).$$

3.8 Consider an open loop system with the nonlinear input-output relation $y = f(u)$. Assume that the system is closed with the proportional controller $u = k(r - y)$. Show that the input-output relation of the closed loop system is

$$y + \frac{1}{k} f^{-1}(y) = r.$$

Estimate the largest deviation from ideal linear response $y = r$. Illustrate by plotting the input output responses for a) $f(u) = \sqrt{u}$ and b) $f(u) = u^2$ with $0 \le u \le 1$ and $k = 5, 10$ and 100.

3.9 Consider the system in Section 3.2 where the controller was designed to give a closed loop system characterized by $\omega_c = 1$ and $\zeta = 0.707$. The transfer functions of the process and the controller are

$$P(s) = \frac{2}{s+1}, \qquad C(s) = \frac{0.207s + 0.5}{s}.$$

The response of the closed loop system to command signals has a settling time (time required to stay within 2% of the final value, see Figure 5.9) of $4/\zeta\omega_c \approx 5.66$ Assume that the attenuation of the load disturbances is satisfactory but that we want a closed loop system system that responds five times faster to command signals without overshoot. Determine the transfer functions of a controller with the architecture in Figure 3.15 that gives a response to command signals with a first order dynamics. Simulate the system in the nominal case of a perfect model and explore the effects of modeling errors by simulation.

3.10 Consider a queuing system modeled by

$$\frac{dx}{dt} = \lambda - \mu_{\max} \frac{x}{x+1}.$$

The model is nonlinear and the dynamics of the systems changes significantly with the queuing length, see Example 2.11. Investigate the situation when a PI controller is used for admission control. The arrival intensity λ is then given by

$$\lambda = k_p(r - x) + k_i \int^t (r(t) - x(t))dt.$$

The controller parameters are determined from the approximate model

$$\frac{dx}{dt} = \lambda.$$

Find controller parameters that give the closed loop characteristic polynomial $s^2 + 2s + 1$ for the approximate model. Investigate the behavior of the control strategy for the nonlinear model by simulation for the input $r = 5 + 4\sin(0.1t)$.