

---

# Feedback Systems

---

*An Introduction for Scientists and Engineers*  
*SECOND EDITION*

Karl Johan Åström  
Richard M. Murray

Version v3.0f (30 Aug 2015)

This is the electronic edition of *Feedback Systems* and is available from <http://www.cds.caltech.edu/~murray/FBS>. Hardcover editions may be purchased from Princeton University Press, <http://press.princeton.edu/titles/8701.html>.

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

PRINCETON UNIVERSITY PRESS  
PRINCETON AND OXFORD

---

## Chapter Two

### 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 some 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. The basic concepts described in this chapter are explained in more detail in the remainder of the text, and this chapter can be skipped for readers who prefer to move directly to the more detailed analysis and design techniques.

#### 2.1 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.

##### Linear Differential Equations and Transfer Functions

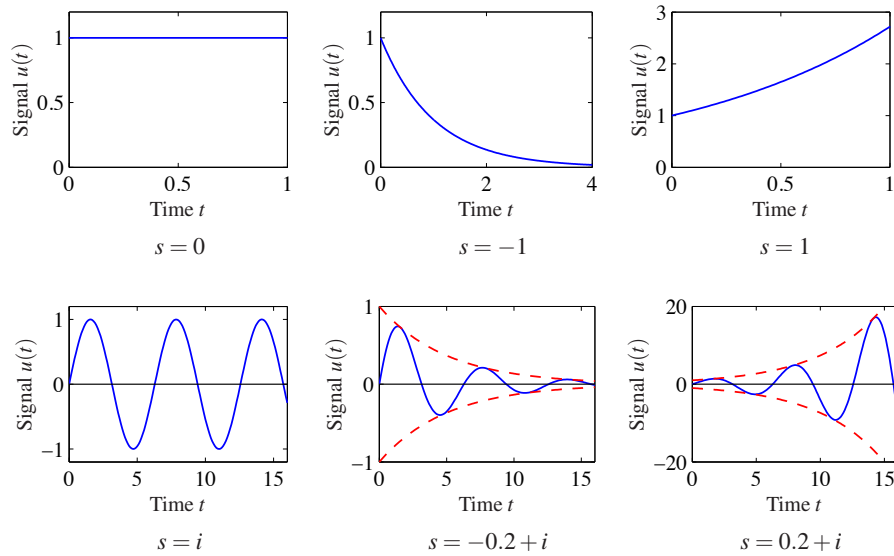
In many practical situations, the input/output behavior of a system can be modeled by a linear differential equation of the form

$$\frac{d^n y}{dt^n} + a_1 \frac{d^{n-1} y}{dt^{n-1}} + \cdots + a_n y = b_1 \frac{d^{n-1} u}{dt^{n-1}} + \cdots + b_n u. \quad (2.1)$$

where  $u$  is the input,  $y$  is the output and the coefficients  $a_k$  and  $b_k$  are real numbers. The model (2.1) is more general than the model given by equation (3.7) in Section 3.2 because the right hand side has terms with derivatives of the input  $u$ . The differential equation (2.1) is characterized by two polynomials

$$a(s) = s^n + a_1 s^{n-1} + \cdots + a_n, \quad b(s) = b_1 s^{n-1} + b_2 s^{n-2} + \cdots + b_n, \quad (2.2)$$

where  $a(s)$  is the *characteristic polynomial* of the differential equation (2.1).



**Figure 2.1:** Examples of exponential signals. The top row corresponds to exponential signals with a real exponent, and the bottom row corresponds to those with complex exponents. The dashed line in the last two cases denotes the bounding envelope for the oscillatory signals. In each case, if the real part of the exponent is negative then the signal decays, while if the real part is positive then it grows.

The solution to equation (2.1) 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 *homogeneous equation* associated with equation (2.1) is

$$\frac{d^n y}{dt^n} + a_1 \frac{d^{n-1} y}{dt^{n-1}} + \dots + a_n y = 0. \quad (2.3)$$

Letting  $s_k$  represent the roots of the *characteristic equation*  $a(s) = 0$ , the solution to equation (2.3) is

$$y(t) = \sum_{k=1}^n C_k e^{s_k t} \quad (2.4)$$

if the characteristic does not have multiple roots  $s_k$ . The parameters  $C_1, \dots, C_n$  are constants that can be determined from the initial conditions.

Since the coefficients  $a_k$  are real, the roots of the characteristic equation are either real-valued or occur in complex conjugate 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, is constant if  $s_k = 0$  and increases if  $s_k$  is positive, as shown in the top row of Figure 2.1. 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

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

which have oscillatory behavior, as illustrated in the bottom row of Figure 2.1. The sine terms are shown as solid lines and the cosine terms as dashed lines; they have zero crossings with the spacing  $\pi/\omega$ . The dotted lines show the envelopes, which correspond to the exponential function  $\pm e^{\sigma t}$ .

When the characteristic equation (2.4) has multiple roots, the solutions to the homogeneous equation (2.3) take the form

$$y(t) = \sum_{k=1}^m C_k(t) e^{s_k t}, \quad (2.5)$$

where  $C_k(t)$  is a polynomial with degree less than the multiplicity of the root  $s_k$ . The solution (2.5) has  $\sum_{k=1}^m (\deg C_k + 1) = n$  free parameters.

Having explored the solution to the homogeneous equation, we now turn to the input-dependent part of the solution. The solution to equation (2.1) 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

$$\begin{aligned} \frac{du}{dt} &= s e^{st}, & \frac{d^2 u}{dt^2} &= s^2 e^{st}, & \dots & \frac{d^n u}{dt^n} = s^n e^{st} \\ \frac{dy}{dt} &= s G(s) e^{st}, & \frac{d^2 y}{dt^2} &= s^2 G(s) e^{st}, & \dots & \frac{d^n y}{dt^n} = s^n G(s) e^{st}. \end{aligned} \quad (2.6)$$

Inserting these expressions into the differential equation (2.1) gives

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

and hence

$$G(s) = \frac{b_1 s^{n-1} + b_2 s^{n-2} + \dots + b_n}{s^n + a_1 s^{n-1} + \dots + a_n} = \frac{b(s)}{a(s)}. \quad (2.7)$$

This function is called the *transfer function* of the system. It describes a particular solution to the differential equation for the input  $e^{st}$  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  $p = \frac{d}{dt}$  and the notation  $p^k = \frac{d^k}{dt^k}$ . The differential equation (2.1) can be written as

$$p^n y + a_1 p^{n-1} y + \dots + a_n y = b_1 p^{n-1} u + b_2 p^{n-2} u + \dots + b_n u,$$

or

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

The relation between the transfer function (2.7) and the differential equation (2.1) is clear: the transfer function (2.7) can be obtained by inspection from the differential equation (2.1), 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 (2.1).

To deal with oscillatory signals, like those shown in Figure 2.1, it is convenient to allow  $s$  to be a complex number. The transfer function is a function  $G: \mathbb{C} \rightarrow \mathbb{C}$



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 (2.4). 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  $y(t) = G(0) = b_n/a_n$ . The quantity  $G(0)$  is called the *zero frequency gain* or the *static gain*. Letting  $\arg$  represent the argument (phase) of a complex number and  $|\cdot|$  the magnitude, the particular solution for the input  $u = \cos(\omega t) = \operatorname{Re} e^{i\omega t}$  is

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

The input is thus amplified by  $|G(i\omega)|$  and the phase shift between input and output is  $\arg G(i\omega)$ . The functions  $G(i\omega)$ ,  $|G(i\omega)|$  and  $\arg G(i\omega)$  are called the *frequency response*, *gain* and *phase* and the transfer function  $G$ . 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 (2.4) or (2.5). 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 as  $t \rightarrow \infty$  and the general solution converges to the particular solution as time increases.

The transfer function  $G(s)$  is a useful representation of the differential equation (2.1) and of the system modeled by the differential equation. The transfer function has many physical interpretations that can be exploited for analysis and design. The transfer function makes it possible to apply algebra to determine relations between signals in a complex system. The transfer function can also convey a great deal of intuition:  $G(0)$  is the steady state gain for constant inputs and the frequency response  $G(i\omega)$  capture the steady state response to sinusoidal functions. The frequency response of a stable system can be determined experimentally by exploring the response of a system to sinusoidal signals. The approximations of  $G(s)$  for small and large  $s$  capture the propagation of slow and fast signals respectively. Consider for example the spring-mass system in equation (3.16), 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  $q = (1/k)u$  which implies that for low frequency inputs, the system behaves like a spring driven by a force. For large  $s$  we have  $G(s) \approx 1/(ms^2)$ . The corresponding differential equation is  $m\ddot{q} = u$ , and for high frequency inputs the system behaves like mass driven by a force (a double integrator). A more elaborate treatment of

transfer functions and the frequency response will be given in later chapters, particularly in Chapter 9.

### Stability: The Routh-Hurwitz Criterion

When using feedback there is always the danger that the system may become unstable, and it is therefore important to have a stability criterion. The differential equation (2.1) is called *stable* if all solutions of the homogeneous equation (2.3) go to zero for any initial condition. It follows from equation (2.5) 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 explicit calculation of the roots, because it gives conditions in terms of the coefficients of the characteristic polynomial.

We illustrate the Routh-Hurwitz criterion by describing it for first, second and third order differential equations. 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 and only 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 and only if

$$a_1, a_2, a_3 > 0, \quad \text{and} \quad a_1 a_2 > a_3. \quad (2.8)$$

The corresponding conditions for a fourth order differential equation are

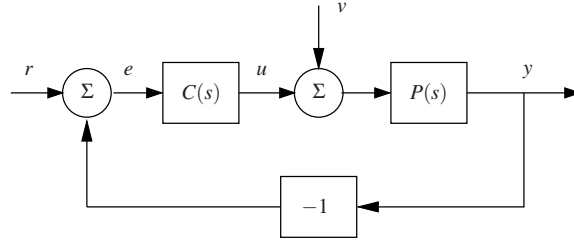
$$a_1, a_2, a_3 > 0, \quad a_1 a_2 > a_3, \quad \text{and} \quad a_1 a_2 a_3 > a_1^2 a_4 + a_3^2. \quad (2.9)$$

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

As described in Chapter 3, control systems are often described using block diagrams. Figure 2.2 shows a block diagram of a typical control system. The control system takes a reference signal  $r$  and compares it to the system output  $y$  (assumed to be in the same units). The resulting error is fed to a controller, whose output  $u$  drives the process. We also model process disturbances using the signal  $v$ , which we assume enters in the same location as the controller input. If each block is modeled as a linear differential equation (2.1), we need to find the differential equation that relates the signals in the complete system. A block can be considered as a



**Figure 2.2:** 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  $v$ .

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  $v$  in Figure 2.2 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

$$C(s) = \frac{b_c(s)}{a_c(s)}, \quad P(s) = \frac{b_p(s)}{a_p(s)}. \quad (2.10)$$

The corresponding differential equations are

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

where we recall that  $p^k = \frac{d^k}{dt^k}$ . 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

$$(a_c(p)a_p(p) + b_p(p)b_c(p))y(t) = b_p(p)b_c(p)r(t), \quad (2.11)$$

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 (2.11) corresponds to the transfer function

$$G_{yr} = \frac{b_p(s)b_c(s)}{a_c(s)a_p(s) + b_p(s)b_c(s)} = \frac{P(s)C(s)}{1 + P(s)C(s)}, \quad (2.12)$$

where we use the notation  $G_{yr}$  for the transfer function from  $r$  to  $y$ . Proceeding in the same way we obtain the following transfer functions for other input/output

pairs:

$$G_{ur} = \frac{C(s)}{1 + P(s)C(s)}, \quad G_{yv} = \frac{P(s)}{1 + P(s)C(s)}, \quad G_{uv} = \frac{-P(s)C(s)}{1 + P(s)C(s)}. \quad (2.13)$$

Instead of manipulating the differential equations we can also determine transfer functions by tracing signals around the loop. Assume, for example that we would like to determine the transfer function  $G_{yv}(s)$ , then let the input be an exponential signal  $v(t) = e^{st}$ . The output is then  $y(t) = G_{yv}(s)e^{st}$ . Tracing signals around the loop in the block diagram in Figure 2.2 gives

$$y(t) = G_{yv}(s)e^{st} = P(s)e^{st} - P(s)C(s)y(t) = P(s)e^{st} - P(s)C(s)G_{yv}(s)e^{st}$$

and hence

$$G_{yv}(s)e^{st} = P(s)e^{st} - P(s)C(s)G_{yv}(s)e^{st}$$

and solving for  $G_{yv}(s)$  gives the expression (2.13).

By using polynomials and transfer functions the relations between signals in a feedback system can thus be obtained by algebra. In fact, 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 2.2 are  $P(s)$ ,  $-1$  and  $C(s)$  and so the numerator for  $G_{uv}$  is  $-P(s)C(s)$ .

## 2.2 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 to alleviate disturbances in the process industry, for machine tools 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 2.2. 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 (2.13). For simplicity we will assume that the process is modeled by the first order differential equation

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

The corresponding transfer function is

$$P(s) = \frac{b}{s+a}. \quad (2.14)$$

A first order system is a reasonable model of a physical system if the storage of mass, momentum or energy can be captured by a single state variable. Typical examples are the velocity of a car on a road, the angular velocity of a 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:  $u = k_p e$ , see Section 1.5. The controller transfer function is then  $C(s) = k_p$ . The process transfer function is given by equation (2.14) and the effect of the disturbance on the output is then described by the transfer function (2.13):

$$G_{yv}(s) = \frac{P(s)}{1 + P(s)C(s)} = \frac{\frac{b}{s+a}}{1 + \frac{bk_p}{s+a}} = \frac{b}{s+a+bk_p}.$$

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

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

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 steady state error thus 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 2.3a shows the responses for a few values of controller gain  $k_p$ .

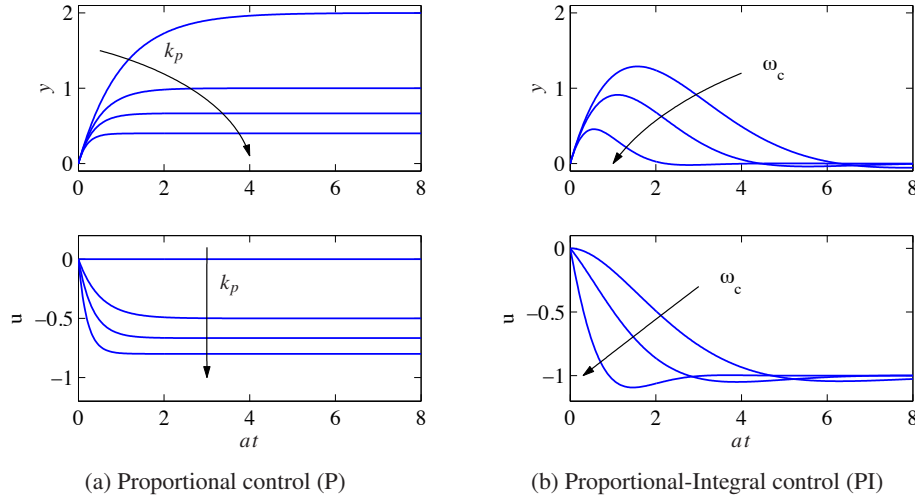
### Proportional-Integral (PI) Control

The PI controller, introduced in Section 1.5, is described by

$$u(t) = k_p e(t) + k_i \int_0^t e(\tau) d\tau. \quad (2.15)$$

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

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



**Figure 2.3:** 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 (2.20) with  $\zeta_c = 0.707$  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  $2$ .

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 2.2 and we find by inspection that the transfer function from  $v$  to  $y$  is

$$G_{yv}(s) = \frac{P(s)}{1 + P(s)C(s)} = \frac{bs}{s^2 + (a + bk_p)s + bk_i}. \quad (2.16)$$

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

$$\frac{d^2y}{dt^2} + (a + bk_p)\frac{dy}{dt} + bk_iy = b\frac{dv}{dt}. \quad (2.17)$$

Notice that, since the disturbance enters as a derivative on 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 1.5.

To find suitable values of the controller parameters  $k_p$  and  $k_i$ , we consider the characteristic polynomial of the differential equation (2.17),

$$a_{cl}(s) = s^2 + (a + bk_p)s + bk_i. \quad (2.18)$$

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

$$(s + \sigma_d + i\omega_d)(s + \sigma_d - i\omega_d) = s^2 + 2\sigma_d s + \sigma_d^2 + \omega_d^2. \quad (2.19)$$

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

$$e^{-\sigma_d t} \sin(\omega_d t), \quad e^{-\sigma_d t} \cos(\omega_d t),$$

which are damped sine and cosine functions, as shown in the lower left plot in Figure 2.1. The coefficient  $\sigma_d$  determines the decay rate and the parameter  $\omega_d$  gives the frequency of the decaying oscillation. Identifying coefficients of equal powers of  $s$  in the polynomials (2.18) and (2.19) gives

$$k_p = \frac{2\sigma_d - a}{b}, \quad k_i = \frac{\sigma_d^2 + \omega_d^2}{b}. \quad (2.20)$$

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

$$a_{cl}(s) = s^2 + 2\sigma_d s + \sigma_d^2 + \omega_d^2 = s^2 + 2\zeta_c \omega_c s + \omega_c^2.$$

This parameterization has the advantage that  $\zeta_c$ , which is in the range  $[-1, 1]$ , determines the shape of the response and  $\omega_c$  gives the response speed.

Figure 2.3b shows the output  $y$  and the control signal  $u$  for  $\zeta_c = 1/\sqrt{2} = 0.707$  and different values of  $\omega_c$ . Proportional control gives a steady-state error which decreases with increasing controller gain  $k_p$ . With PI control the steady-state error is zero. Both the decay rate and the peak error decrease when the design parameter  $\omega_c$  is increased. Larger controller gain give smaller errors and control signals that react faster to the disturbance.

With the controller parameters (2.20) the transfer function (2.16) from disturbance  $v$  to process output becomes

$$G_{yv}(s) = \frac{P(s)}{1 + P(s)C(s)} = \frac{bs}{s^2 + 2\zeta_c \omega_c s + \omega_c^2}.$$

For efficient attenuation of disturbances is desirable that  $|G_{yv}(i\omega)|$  is small for all  $\omega$ . For small values of  $\omega$  we have  $|G_{yv}(i\omega)| \approx b\omega/\omega_c$ , for large  $\omega$  we have  $|G_{yv}(i\omega)| \approx b/\omega$ . The largest value of  $|G_{yv}(i\omega)|$  is  $b/(2\zeta_c \omega_c)$  for  $\omega = \omega_c$ . It thus follows that a large value of  $\omega_c$  gives good load disturbance attenuation.

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. To achieve the benefits of large control gains the model must be accurate over wide frequency ranges, as will be discussed next.

### Unmodeled Dynamics

The analysis we have made so far indicates that there are no limits to the performance that can be achieved. Figure 2.3b shows that arbitrarily fast response can be obtained simply by making  $\omega_c$  sufficiently large. In reality there are of course limitations on what is achievable. One reason is that the controller gains increase with

$\omega_c$ , the proportional gain is  $k_p = (2\zeta_c\omega_c - a)/b$  and the integral gain is  $k_i = \omega_c^2/b$ . A large value of  $\omega_c$  thus gives large controller gains and the control signal may be so large that actuator saturates. Another reason is that the model (2.14) is a simplification: it is only valid in a given frequency range. If the model is instead

$$P(s) = \frac{b}{(s+a)(1+sT)}, \quad (2.21)$$

where the term  $1+sT$  represents the dynamics of sensors, actuators or other dynamics that was neglected when deriving equation (2.14), so-called *unmodeled dynamics*, the closed loop characteristic polynomial for the closed loop system becomes

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

It follows from the Routh-Hurwitz criterion (2.8) that the closed loop system is stable if  $\omega_c^2T < 2\zeta_c\omega_c(1+aT)$  or if

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

The frequency  $\omega_c$  and the achievable response time are thus limited by the unmodeled dynamics represented by  $T$ . When models are developed for control it is therefore important to also consider the unmodeled dynamics.

### 2.3 Using Feedback to Follow Command Signals

Another major application of feedback is to make a system output follow a command signal. This 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 appear in high performance audio amplifiers, machine tools and industrial robots.

To illustrate command signal following we will consider the system in Figure 2.2 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

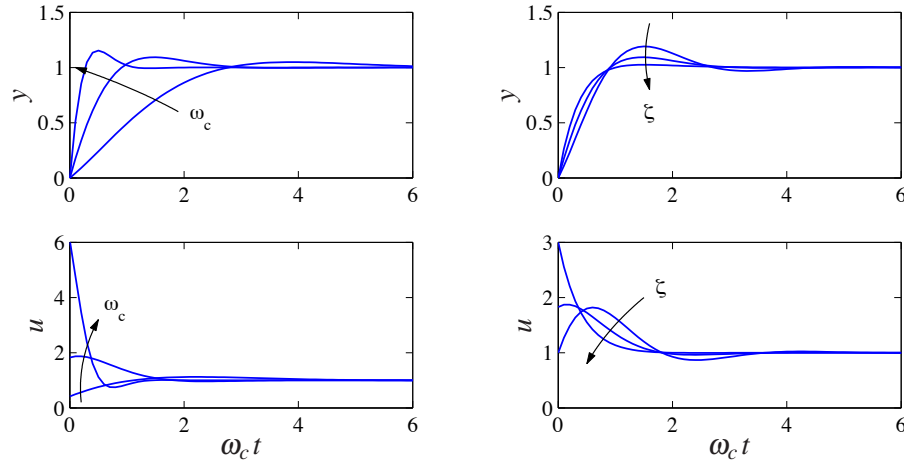
$$P(s) = \frac{b}{s+a}, \quad C(s) = \frac{k_ps + k_i}{s}. \quad (2.22)$$

Since we will focus on following the command signal  $r$  we will neglect the load disturbance,  $v = 0$ . It follows from equation (2.12) that the transfer function from the command signal  $r$  to the output  $y$  is

$$G_{yr}(s) = \frac{P(s)C(s)}{1 + P(s)C(s)} = \frac{bk_ps + bk_i}{s^2 + (a + bk_p)s + bk_i}. \quad (2.23)$$

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





**Figure 2.4:** Responses to a unit step change in the command signal for different values of the design parameters  $\omega_c$  and  $\zeta_c$ . The left figure shows responses for fixed  $\zeta_c = 0.707$  and  $\omega_c = 1, 2$  and  $5$ . The right figure shows responses for  $\omega_c = 2$  and  $\zeta_c = 0.5, 0.707$ , and  $1$ . The process parameters are  $a = b = 1$ . The initial value of the control signal is  $k_p$ .

To determine suitable values of the controller parameters  $k_p$  and  $k_i$  we proceed as in Section 2.2 by choosing controller parameters that makes the closed-loop characteristic polynomial

$$a_{cl}(s) = s^2 + (a + bk_p)s + bk_i \quad (2.24)$$

equal to  $s^2 + 2\zeta_c\omega_c s + \omega_c^2$  with  $\zeta_c > 0$  and  $\omega_c > 0$ . Identifying coefficients of equal powers of  $s$  in these polynomials give

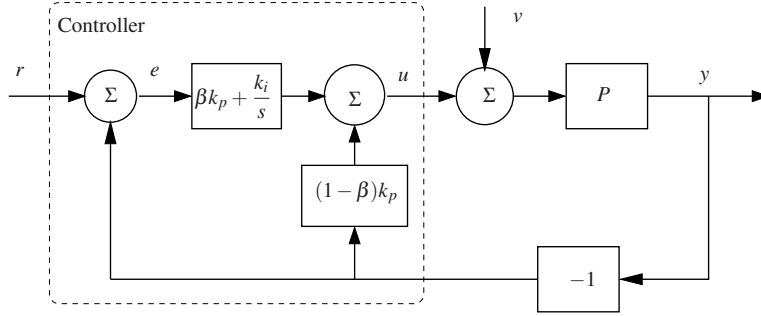
$$k_p = \frac{2\zeta_c\omega_c - a}{b}, \quad k_i = \frac{\omega_c^2}{b}, \quad (2.25)$$

which is equivalent to equation (2.20). Notice that integral gain increases with the square of  $\omega_c$ . Figure 2.4 shows the output signal  $y$  and the control signal  $u$  for different values of the design parameters  $\zeta_c$  and  $\omega_c$ . The response time decreases with increasing  $\omega_c$  and the initial value of the control signal also increases because it takes more effort to move rapidly. The overshoot decreases with increasing  $\zeta_c$ . For  $\omega_c = 2$ , the design choice  $\zeta_c = 1$  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 controller parameters (2.25) it follows from equation (2.23) that

$$G_{yr}(s) = \frac{P(s)C(s)}{1 + P(s)C(s)} = \frac{(2\zeta_c\omega_c - a)s + \omega_c^2}{s^2 + 2\zeta_c\omega_c s + \omega_c^2}.$$

The magnitude will be close to 1 when  $\omega_c$  is large compared to  $s$ . It is thus de-



**Figure 2.5:** Block diagram of a closed-loop system with a PI controller having an architecture with two degrees of freedom.

sirable to have a large  $\omega_c$  to be able to track fast changes in the reference signal. The frequency response of  $G_{yr}$  gives a quantitative representation of the tracking abilities.

### Controllers with Two Degrees of Freedom

The control law in Figure 2.2 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  $r$  gives an immediate step change in the control signal  $u$ . This rapid reaction can be advantageous, but it may give large overshoot, which can be avoided by a replacing the PI controller in equation (2.15) with a controller of the form

$$u(t) = k_p(\beta r(t) - y(t)) + k_i \int_0^t (r(\tau) - y(\tau)) d\tau. \quad (2.26)$$

In this modified PI algorithm, the proportional action only acts on the fraction  $\beta$  of the reference signal. The signal transmission from reference  $r$  to  $u$  and from output  $y$  to  $u$  can be represented by the transfer functions

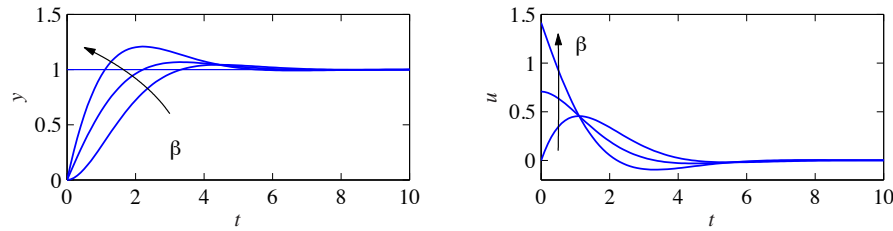
$$C_{ur}(s) = \beta k_p + \frac{k_i}{s}, \quad C_{uy}(s) = k_p + \frac{k_i}{s} = C(s). \quad (2.27)$$

The controller (2.26) 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 2.5. Let the process transfer function be  $P(s) = b/(s + a)$ . The transfer functions from reference  $r$  and disturbance  $v$  to output  $y$  are

$$G_{yr}(s) = \frac{b\beta k_p s + b k_i}{s^2 + (a + b k_p)s + b k_i}, \quad G_{yv}(s) = \frac{s}{s^2 + (a + b k_p)s + b k_i}. \quad (2.28)$$

Comparing with the corresponding transfer function for a controller with error feedback in equations (2.16) and (2.23) we find that the responses to the load disturbances are the same but the response to reference values are different.



**Figure 2.6:** 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  $1$ .

A simulation of the closed loop system for  $a = 0$  and  $b = 1$  is shown in Figure 2.6. The figure shows that the parameter  $\beta$  has a significant effect on the responses. Comparing the system with error feedback ( $\beta = 1$ ) to the system with smaller values of  $\beta$  we find that using a system with two degrees of freedom gives the same settling time with less overshoot and gentler control actions.

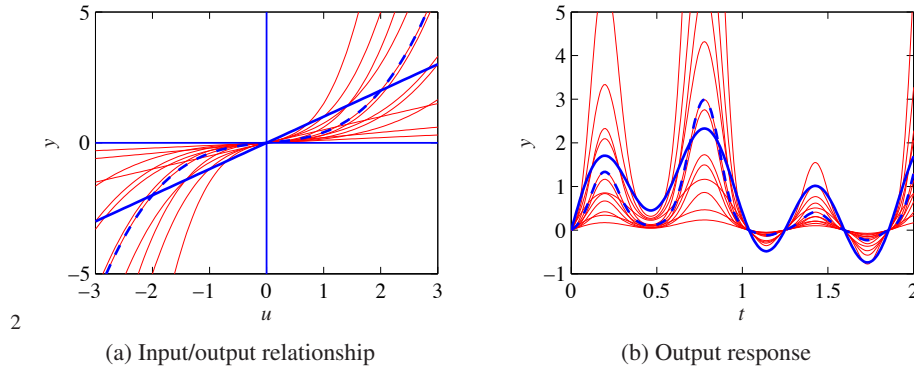
The example shows that command signal response can be improved by using a controller architecture having two degrees of freedom. In Section 2.7 we will show that the responses to command signals and disturbances can be completely separated by using a more general system architecture. To use a system with two degrees of freedom both the reference signal  $r$  and the output signal  $y$  must be measured. There are situations where only the error signal  $e = r - y$  can be measured, typical examples are DVD players, optical memories and atomic force microscopes.

## 2.4 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 having 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.





**Figure 2.7:** Responses of a static 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 (2.30). The ideal response is shown solid bold lines. The nominal response of the nonlinear system is shown using dashed bold lines and the responses for different parameter values are shown using thin lines.

### A Nonlinear Amplifier

Consider an amplifier with a static, nonlinear input/output relation with considerable variability as illustrated in Figure 2.7a. The nominal input/output characteristics is shown as a dashed bold line and examples of variations as thin lines. The nonlinearity in the figure is actually

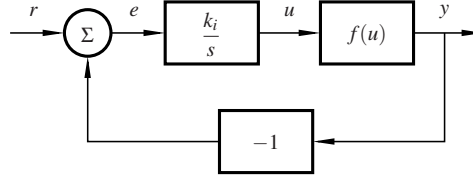
$$y = f(u) = \alpha(u + \beta u^3), \quad -3 \leq u \leq 3. \quad (2.29)$$

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

$$r(t) = \sin(t) + \sin(\pi t) + \sin(\pi^2 t). \quad (2.30)$$

are shown in Figure 2.7b. The desired response  $y = u$  is shown as a solid bold line and responses for a range of parameters are shown with thin lines. The nominal response of the nonlinear system is shown as a dashed bold line. 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 2.8. Figure 2.9 shows the behavior of the closed loop system with the same parameter variations as in Figure 2.7. The input/output plot in Figure 2.9a 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 of the dynamics introduced by the feedback. Figure 2.9b shows the responses to the reference signal; notice the dramatic improvement compared with Figure 2.7b. The tracking error is shown in Figure 2.9c.



**Figure 2.8:** Block diagram of a nonlinear system with integral feedback.



### 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

$$P(s) = b = f'(u_0) = \alpha(1 + 3\beta u_0^2), \quad C(s) = \frac{k_i}{s}, \quad (2.31)$$

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  $\alpha, \beta$  and  $u_0$ . It follows from equation (2.13) that the transfer functions relating the output  $y$  and the error  $e$  to the reference signal are

$$G_{yr}(s) = \frac{bk_i}{s + bk_i}, \quad G_{er}(s) = 1 - G_{yr} = \frac{s}{s + bk_i}. \quad (2.32)$$

The closed loop system is a first-order system with the pole  $s = -bk_i$  and the time constant  $T_{cl} = 1/(bk_i)$ . The integral gain is chosen as  $k_i = 1000$ . The closed loop pole ranges from 100 rad/s to  $2.75 \times 10^4$  rad/s, which is fast compared to the high frequency component 9.86 rad/s of the input signal.

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

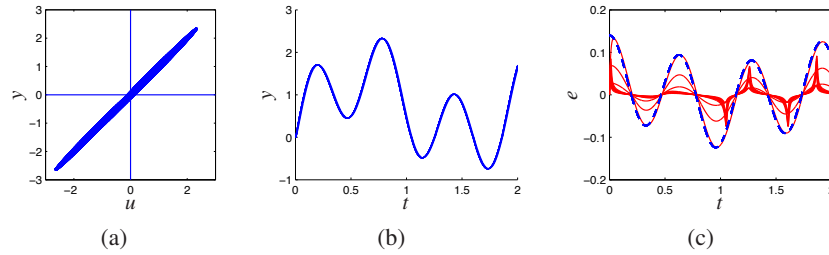
$$\frac{de}{dt} = -bk_i e + \frac{dr}{dt}, \quad \frac{dr}{dt} = \cos(t) + \pi \cos(\pi t) + \pi^2 \cos(\pi^2 t). \quad (2.33)$$

The fast frequency component of the input (2.30) has the frequency  $\pi^2 = 9.86$ ; it is slower than the process dynamics for all parameter variations. Neglecting the term  $de/dt$  in equation (2.33) gives

$$e \approx \frac{1}{bk_i} \frac{dr}{dt}. \quad (2.34)$$

The largest error is obtained when  $b$  has its smallest value 0.1. The error is then approximately  $\pi^2/(bk_i) \cos(\pi^2 t) \approx 0.1 \cos(\pi^2 t)$ , which is shown as the dashed line in Figure 2.9c.

This analysis has shown that it is possible to design an integrating controller



**Figure 2.9:** Responses of the system with integral feedback ( $k_i = 1000$ ). 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 2.7. Notice the dramatic improvement compared to Figure 2.7b. The dashed line in (c) corresponds to the approximate error based on linearization of the nonlinearity.

for a system whose dynamics can be approximated by a static model. Design is essentially the choice of a single parameter: the integral gain  $k_i$  of the controller. The closed loop transfer function from reference to output is given by equation (2.32) 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.

This 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.

## 2.5 Positive Feedback

Most of this book is focused on negative feedback because of its amazingly good properties, which have been illustrated in the previous sections. In this section we will discuss positive feedback, which has complementary properties. In spite of this positive feedback has found good use in several contexts.

Systems with negative feedback can be well understood by linear analysis. To understand systems with positive feedback it is necessary to consider nonlinear effects, because without the nonlinearities the instability caused by positive feedback will grow without bounds. The nonlinear elements contribute to create interesting and useful effects by limiting the signals.

Positive feedback is common: encouraging a student or a coworker when they have performed well encourages them to do even better. In biology, it is standard to distinguish inhibitory connections (negative feedback) from excitatory feedback (positive feedback) as illustrated in Figure 2.10. Neurons use a combination of positive and negative feedback to generate spikes.

Positive feedback may cause instabilities. Exponential growth is a typical example of positive feedback, where the rate of change of a quantity  $x$  is proportional

**Figure 2.10:** Schematic diagram of xxx which has both positive and negative feedback. An exhibitory connection (positive feedback) is denoted with a line ending with an arrow, an inhibitory interaction (negative feedback) is denoted with an arrow ending with a circle.

to the  $x$ . Hence

$$\frac{dx}{dt} = \alpha x,$$

which implies that  $x(t) = e^{\alpha t}$  grows exponentially. In nature, exponential growth of a species is limited by the finite amount of food. Another common example is when a microphone is placed close to a speaker in public address systems, resulting in a howling noise. Positive feedback can create stampedes in cattle heard, runs on banks and boom-bust behavior. In all these cases there is exponential growth that is finally limited by finite resources.

The notions of positive and negative feedback are highly intuitive but they are also too simplistic. Their interpretation is clear when we are not considering dynamics. But if dynamics is considered the sign of the feedback will depend on the frequency, and feedback can change from positive to negative when the frequency changes. A good understanding of the dynamic case can be obtained by tracing sinusoidal signals around the feedback loop as will be done in Section 10.2. We then obtain a more sophisticated notion that a block with feedback is positive if the input and output signals are in phase, and feedback is negative if signals are out of phase. With the simple interpretation it appears that a system with negative feedback will always be stable and that a system with positive feedback will always be unstable. None of these statements are true when dynamics are considered.

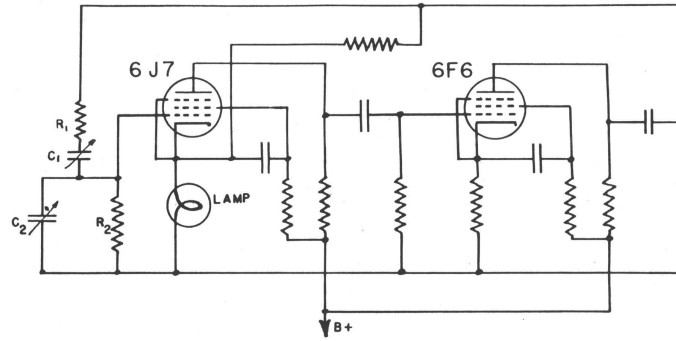
In spite of the drawbacks mentioned above, positive feedback has several useful applications. A typical example is the generation of periodic signals as illustrated by the following example.

### Hewlett's Oscillator

Since positive feedback tends to generate instability it can be used to construct oscillators. To limit the exponential growth it is necessary to introduce some nonlinearity that limits the amplitude of the oscillation. An example is given in Figure 2.11 from William Hewlett's 1939 PhD thesis at Stanford University. Hewlett used two vacuum tubes with positive feedback and a nonlinear element in the form of a lamp to maintain constant amplitude of the oscillation. The positive feedback in the basic loop creates an oscillation. The resistance of the lamp decreases as the signal amplitude increases and the amplitude is limited. Hewlett's oscillator was the beginning of a very successful company that Hewlett founded with David Packard.

### The Superregenerative Receiver

In the previous sections we have shown that negative feedback has some very useful properties. The negative feedback amplifier was an enabler for long-distance



**Figure 2.11:** Circuit diagram of William Hewlett's oscillator that gives a stable oscillation by using positive feedback and a nonlinearity (lamp) that stabilized the amplitude of the oscillation, [1].

telephony. A key idea was to design an amplifier with a very large open loop gain and to reduce the gain by negative feedback. The result was an amplifier that is robust and linear. Positive feedback has complementary properties: it is possible to create high gain but the closed loop system is sensitive to parameter variations.

To understand that positive feedback can generate high gains we consider an amplifier with gain  $A_{ol}$ . Neglecting dynamics and closing a feedback loop around the amplifier with positive feedback  $k$  gives a closed loop system with the closed loop gain

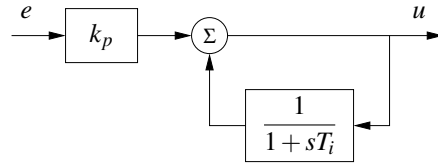
$$A_{cl} = \frac{A_{ol}}{1 - kA_{ol}}.$$

A very large closed loop gain  $A_{cl}$  can be obtained by selecting a feedback gain  $k$  that is just below the stability limit  $1/A_{ol}$ . Choosing the gain so that  $kA_{ol} = 0.999$  gives  $A_{cl} = 1000A_{ol}$ . Using this idea, Armstrong constructed a superregenerative radio receiver in 1914 when he was still undergraduate at Columbia University. He built a radio receiver with only one vacuum tube. The drawback by using positive feedback is that the system is highly sensitive and that the gain has to be adjusted carefully to avoid oscillations. It is still used in simple walkie-talkies, garage door openers and toys.

### Implementation of Integral Action

Positive feedback was used very in early controllers where integral action was provided by positive feedback around a system with first order dynamics, as shown in the block diagram of Figure 2.12. Intuitively the system can be explained as follows. Proportional feedback typically gives a steady state error. This can be overcome by adding a bias signal. In Figure 2.12 the bias is estimated by low pass filtering the control signal and adding it back into to the signal path. The circuit can be understood better by a little analysis. Using block diagram algebra we find





**Figure 2.12:** Implementation of integral action by *positive feedback* .

that the transfer function of the system is

$$G_{ue} = \frac{k_p}{1 - \frac{1}{1+sT}} = k_p + \frac{k_p}{sT},$$

which is a transfer function of a PI controller. This way of implementing integral action is still used in many systems. Notice that in this case the closed loop system is in fact unstable since it has an integrator. Since this is the desired behavior (integral action), it is not necessary to limit the signal.

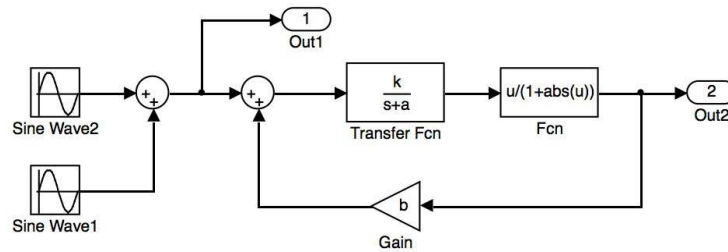
### Positive Feedback Combined with Saturation

Positive feedback is often combined with nonlinear elements as in Hewlett's oscillator. Figure 2.13, which shows the block diagram of a system with positive feedback and a nonlinearity in the form of a saturation is another example. The system has a forward path with a gain followed by a block with first order dynamics and a nonlinearity with saturation characteristics. We assume that the nonlinearity is given by

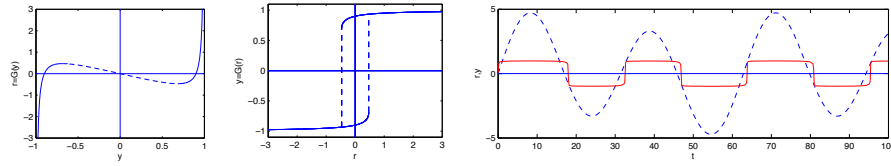
$$y = f(x) = \frac{x}{1 + |x|}, \quad x = f^{-1}(y) = \frac{y}{1 - |y|}. \quad (2.35)$$

The system is described by the equations

$$\frac{dx}{dt} = -ax + k(r + y) = k(r - G(y)), \quad G(y) = \frac{ay}{k(1 - |y|)} - y. \quad (2.36)$$



**Figure 2.13:** Block diagram of system with positive feedback and saturation .



**Figure 2.14:** System with positive feedback and saturation, (a) show the function  $G(y)$  used in the analysis, (b) shows the hysteretic input output map and (c) shows a simulation of the system. The input is the dashed curve and the output is the solid line.

The equilibria for a constant input  $r$  are given by

$$r = -y + \frac{a}{k} f^{-1}(y) = G(y) \quad \text{where} \quad G'(y) = -1 + \frac{a}{k(1-|y|)^2}. \quad (2.37)$$

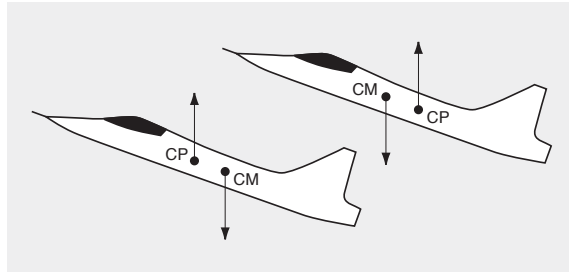
The graph of the function  $G$  is shown in Figure 2.13a for the parameters  $a = 1$ ,  $b = 1$  and  $k = 10$ . The function is monotone if  $k \leq a$  and there is a unique equilibrium. If  $k > a$  the function  $G$  has a maximum  $r_{\max} = 1 + a/k - 2\sqrt{a/k}$  at  $y = -1/\sqrt{1+a/k}$  and a minimum  $r_{\min} = -r_{\max}$  at  $y = 1/\sqrt{1+a/k}$ . For constant  $r$  the equation (2.37) has three solutions if  $r_{\min} < r < r_{\max}$ , and one equilibrium if  $r < r_{\min}$  or if  $r > r_{\max}$ . It is easy to understand the behavior of the solutions to differential equation (2.37), since it is of first order. Since  $x$  is a monotone function of  $y$  it follows from equation (2.36) that  $dx/dt$  is positive to the left of an equilibrium where  $G'(y)$  is positive, and positive to the left of the equilibrium. The stable equilibria thus corresponds the values of  $y$  where the slope of  $G(y)$  is positive, they are marked with full lines in Figure 2.14a. The differential equation thus has two stable equilibria when  $r_{\min} < r < r_{\max}$  and one stable equilibrium when  $|r| > r_{\max}$ . The input-output relation of the system will then have the hysteretic behavior shown in Figure 2.14b where the output switches between the values close to  $\pm 1$ . Results of a simulation of the system is shown in Figure 2.14c. The shape of the output signal depends on the parameters, the values  $a = 5$ ,  $b = 1$  and  $k = 50$  were used in the figure. The parameter  $a$  influences the dynamics of the switching, the parameter  $b$  is half the hysteresis width and the parameter  $k$  influences the sharpness of the switches..

The circuit shown in the Figure 2.13 is commonly used as a trigger to detect changes in a signal (a Schmitt trigger). It is also used as a memory element in solid state memories.

## 2.6 Using Feedback to Shape Behavior

The regulation and servo problems discussed in Sections 2.2 and 2.3 are classical applications of feedback. In Section 2.4 it was shown that feedback can be used to obtain essentially 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



**Figure 2.15:** Schematic diagram of two aircraft. 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.

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 high 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 being made unstable. Schematic pictures of two airplanes are shown in Figure 2.15. 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 an airplane will have superior performance, but it is unstable at subsonic speeds, typically at takeoff and landing. The control

system that stabilizes the aircraft in these operating conditions is mission-critical, with strong requirements on robustness and reliability.

The evolutionary biologist John Maynard Smith [Smi52] has claimed that while early flying animals were inherently stable they later developed unstable configurations when their sensory and nervous systems became more sophisticated and able to stabilize. The unstable configuration had significant advantages in manoeuvrability both for predator and prey.

Keeping an inverted pendulum in the upright position is a prototype example of stabilization. Consider the cart–pendulum system discussed in Examples 3.1 and 3.2. Neglecting damping, assuming that the cart is much heavier than the pendulum and assuming that the tilt angle  $\theta$  is small, equation (3.10) can be approximated by the differential equation

$$J_t \ddot{\theta} - mgl\theta = u. \quad (2.38)$$

The transfer function of the open loop system is

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

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

$$C(s) = k_d s + k_p. \quad (2.39)$$

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  $\omega_c$  and damping ratio  $\zeta_c$ , hence

$$k_d = 2\zeta_c \omega_c J_t, \quad k_p = J_t \omega_c^2 + mgl.$$

Choosing  $\omega_c = \omega_0$  moves the poles from  $\pm \omega_0$  in open loop to  $-\zeta_c \omega_0 \pm i\sqrt{1 - \zeta_c^2} \omega_0$  in closed loop. The controller gains are then  $k_p = 2mgl$  and  $k_d = 2\zeta_c \sqrt{mgl J_t}$ . The control law (2.39) stabilizes the pendulum but it 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 3.1 is essentially a pendulum on a cart and can be modeled by equation (3.9) with an added torque  $\tau$  on the pendulum that is exerted by the person leaning on the platform. Neglecting the damping terms  $c\dot{p}$ ,  $\gamma\dot{\theta}$  and linearizing equation (3.9) gives

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

where  $u$  is the horizontal force generated by the motor and  $\tau$  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 (2.39) can be used to stabilize the system. The closed loop system is then described by

$$M_t \ddot{p} - ml \ddot{\theta} = -k_d \dot{\theta} - k_p \theta, \quad -ml \ddot{p} + J_t \ddot{\theta} - mgl \theta = -\tau. \quad (2.40)$$

Multiplying the first equation by  $ml$ , the second by  $M_t$  and adding the equations, we obtain

$$(M_t J_t - m^2 l^2) \ddot{\theta} + ml k_d \dot{\theta} + ml(k_p - M_t g) \theta = -M_t \tau.$$

Since  $M_t J_t - m^2 l^2 = M_t J + n M l^2 > 0$ , this differential equation is stable if  $k_p > M_t g$ . Solving for  $\ddot{p}$  in equation (2.40) gives the following transfer functions from torque  $\tau$  to tilt angle  $\theta$  and horizontal acceleration  $\ddot{p}$ :

$$G_{\ddot{p}\tau}(s) = \frac{m^2 l^2 s^2 - ml k_d s - ml k_p}{(M_t J_t - m^2 l^2) s^2 + ml k_d s + ml(k_p - M_t g)},$$

$$G_{\theta\tau}(s) = \frac{M_t}{(M_t J_t - m^2 l^2) s^2 + ml k_d s + ml(k_p - M_t g)}.$$

The feedback (2.39), which stabilizes the Segway, thus creates a behavior where the horizontal acceleration  $\ddot{p}$  of the Segway can be controlled by the torque  $\tau$ , which can be generated by leaning on the Segway.

### 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 2.16 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



**Figure 2.16:** Haptic devices, the left figure shows the PHANTOM™ and the right a system is developed by Quanser.

operator be  $T_a$  and  $T$ , respectively. The equation of motion is

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

By measuring the angle  $\theta$  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  $J + k_a$ , virtual damping  $k_d$  and virtual spring action  $k_p$ . If no torque  $T$  is applied, the joystick will assume the orientation given by the reference signal  $\theta_r$ . If a the user applies a torque, the joystick will behave like a damped spring-mass system.

## 2.7 Feedback and Feedforward

Feedback and feedforward have complementary properties as was discussed in Section 1.4. Feedback only acts when there are deviations between the actual and the desired behavior, feedforward acts on planned behavior. Feedback and feedforward can be combined to improve response to command signals and to reduce the effect of disturbances that can be measured. In this section we will discuss design of feedforward control.

### Feedforward and System Inversion

To explore feedforward control we will first investigate command signal following. Consider the system modeled by the differential equation (2.1):

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

Assume that we want to find a control signal  $u$  that gives the response  $y_r$ . It follows from equation (2.1) that the desired control signal satisfies

$$b_1 \frac{d^{n-1}u}{dt^{n-1}} + \dots + b_n u = \frac{d^n y_r}{dt^n} + a_1 \frac{d^{n-1}y_r}{dt^{n-1}} + \dots + a_n y_r, \quad (2.41)$$

This equation is called the *inverse* of equation (2.1) 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  $P^{-1}(s)$ .

There are problems with system inversion since the inverse may require differentiations and it may be unstable. For example, if  $b_1 \neq 0$  we have  $P^{-1}(s) \approx s/b_1$  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

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

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

$$F_f(s) = P^{-1}(s)F_m(s) = \frac{\omega_c^2(s+1)^2}{s^2 + 2\zeta_c \omega_c s + \omega_c^2}. \quad (2.42)$$

Notice that  $F_f(0) = 1$  and  $F_f(\infty) = \omega_c^2$ . The initial value of the control signal for a unit step command is thus  $\omega_c^2$  and the final value is 1. Figure 2.17 shows the outputs  $y$  and the feedforward signals  $u_{ff}$  for a unit step reference signal  $r$  and different values of  $\omega_c$ . The parameter  $\omega_c$  determines the response speed and since  $F_f(\infty) = \omega_c^2$ , very large control signals are required fast responses, Achievable performance is thus limited by the size of admissible control signals.

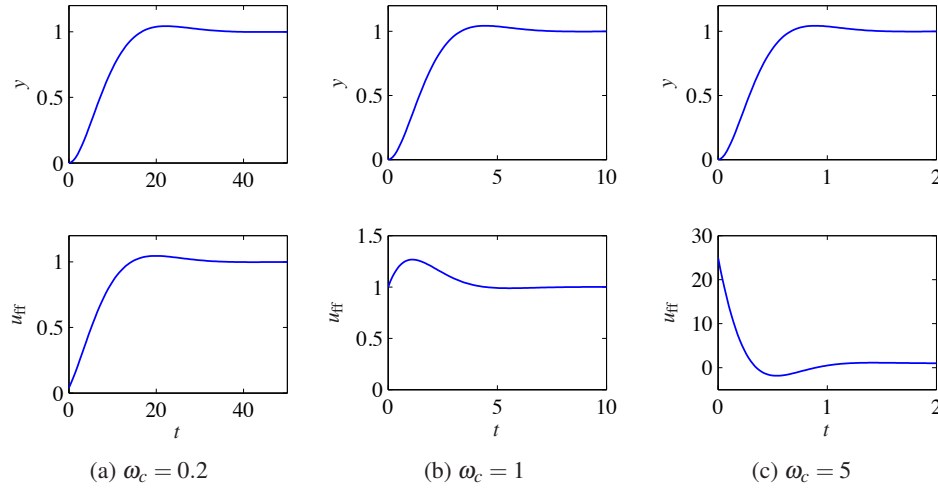
Let the process and the desired response be characterized by the transfer functions

$$P(s) = \frac{1-s}{(s+1)^2}, \quad F_m(s) = \frac{\omega_c^2(1-s)}{s^2 + 2\zeta_c \omega_c s + \omega_c^2}.$$

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

$$F_f(s) = \frac{\omega_c^2(s+1)^2}{s^2 + 2\zeta_c \omega_c s + \omega_c^2}, \quad (2.43)$$

which is the same as equation (2.42). Figure 2.18 shows the outputs  $y$  and the feedforward signals  $u_{ff}$  for different values of  $\omega_c$ . The response to the command signal 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



**Figure 2.17:** Outputs  $y$  (top plots) and feedforward signals  $u_{ff}$  (lower plots) for a unit step command signal. The values of the design parameter is  $\omega_c = 0.2$  (left) 1 (center) and 5 (right). The outputs are identical apart from the time scale, but the control signals required to generate the output differs significantly. The largest value of the feedforward signal  $u_{ff}$  is  $\omega_c^2$ , and it increases significantly with increasing  $\omega_c$ .

is slow ( $\omega_c = 1$ ) but increases with increasing response speed. For  $\omega_c = 5$  the undershoot is more than 200%. The right half plane zero thus severely limits the response time.

The behavior of the control signal changes qualitatively with  $\omega_c$ . To understand what happens we note that the zero frequency gain of the feedforward transfer function (2.43) is  $F_f(0) = 1$  and that its high frequency gain is  $F_f(\infty) = \omega_c^2$ . For a unit step reference signal  $r = 1$ , the initial value of the control signal is  $u_{ff}(0) = \omega_c^2$  and the final value is  $u_{ff}(\infty) = 1$ . 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 1. For  $\omega_c = 5$  the control signals starts at 25 and decays towards the final value 1 with an undershoot.

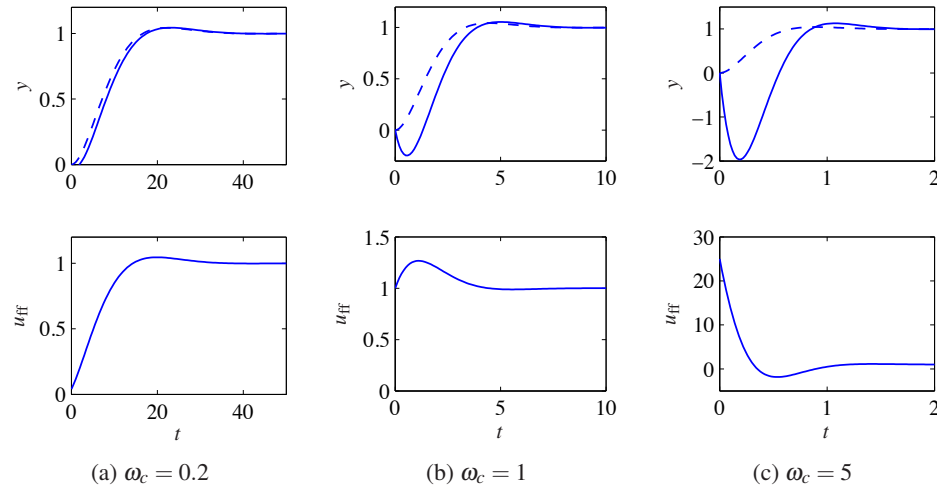
## Sensitivity to Process Variations

### 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 2.19. The controller has three blocks representing the feedback transfer function  $C(s)$  and the feedforward transfer functions  $F_m$  and  $F_f$ .

The controller architecture in Figure 2.19 admits a decoupling of the response to command signals and the response to disturbances. The feedback controller  $C(s)$  is designed to give robustness to process variations and attenuation of load



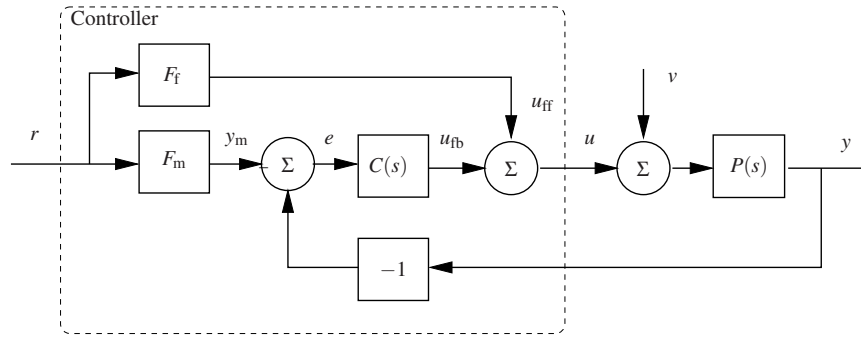


**Figure 2.18:** Outputs  $y$  (top plots) and feedforward signals  $u_{ff}$  (lower plots) for a unit step command signal. The design parameter has the values  $\omega_c = 0.2$  (left) 1 (center) and 0.5 (right) for a unit step command in the reference signal. The dashed curve shows the response that could be achieved if the process did not have the right half plane zero.

disturbances. The desired response to command signals is obtained by design of the feedforward transfer functions  $F_f$  and  $F_m$ .

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  $P_2 = P$ , and therefore  $P_2 = 1$ , the feedforward compensator is simply a proportional controller.

Noise cancellation is a common example of the 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 2.20. 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 2.20b. 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 5.16. Noise cancellation has many applications, in headphones, to create noise-free spaces by active noise control, or to measure electrocardiograms and the heartbeats of mother and

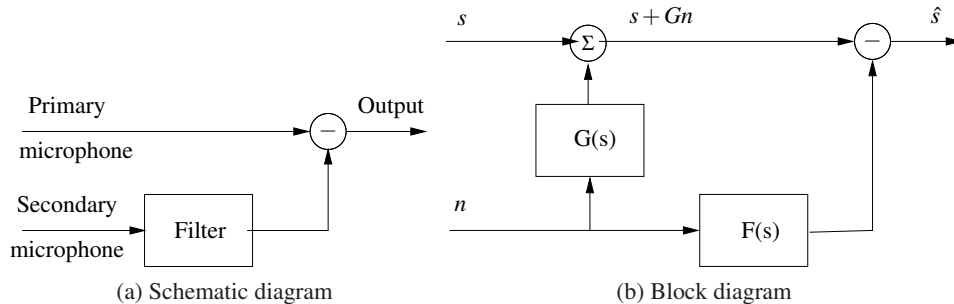


**Figure 2.19:** Block diagram of a closed loop system where the controller has an architecture with two degrees of freedom. The signals  $y_m$  and  $u_{ff}$  are generated by feedforward from the reference  $r$ . The feedback controller  $C(s)$  acts on the control error  $e = y_m - y$  and generates the feedback control signal  $u_{fb}$ .

fetus.

## 2.8 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 [CM51], [JNP47] 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 6 and 9, methods to investigate stability in Chapter 10. The simple method to find parameters of controllers based on matching of coefficients of the closed loop characteristic polynomial is developed further in Chapters 7, 8 and 13. Feedforward control is discussed in Section 8.5 and Section 12.2.



**Figure 2.20:** Schematic and block diagrams for noise cancellation.

### Exercises

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

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

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.

**2.2** 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  $G(s)$  and show that

$$y(t) = y_0(t) + T \frac{dy_0}{dt}, \quad (2.44)$$

Next consider 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 (2.44) to explore the effect of the zero  $s = -1/T$  on the step response of a system

**2.3** Consider a closed loop system with process dynamics and a PI controller modeled by

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

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.

**2.4** 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 controls the pupil opening is 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.

**2.5** 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  $b = 1 \text{ m/s}^2$  and  $a = 0.025 \text{ rad/s}$  (see Section ??). The control signal is normalized to the range  $0 \leq u \leq 1$ . 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  $\zeta$  and  $\omega_c$ ? Use your judgment to find suitable values. Hint: Investigate maximum acceleration and maximum velocity for step changes in the velocity reference.

**2.6** Consider the feedback system in Figure 2.2. 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%.

**2.7** The calculations in Section 2.2 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

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

Show that the controller parameters

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

give a closed loop system with the characteristic polynomial

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

**2.8** 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  $0 \leq u \leq 1$  and  $k = 5, 10$  and  $100$ .

**2.9** Consider the system in Section 2.2 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

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

The response of the closed loop system to step command signals has a settling time (time required to stay within 2% of the final value, see Figure 6.9) 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 that responds five times faster to command signals without overshoot. Determine the transfer functions of a controller with the architecture in Figure 2.19 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.

**2.10** 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 system changes significantly with the queuing length; see Example 3.12. Investigate the situation when a PI controller is used for admission control. The arrival intensity  $\lambda$  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  $r = 5 + 4 \sin(0.1t)$ .