5 PRINCIPLES OF CHEMICAL PROCESS CONTROL

by

H. HAJDÚ and P. MIZSEY

5.1 Introduction

The task of process control

Modern, technically advanced chemical industry cannot work without process control.

Process control is applied in order to

(i) maintain the value of process variables eliminating external disturbances, and thereby ensuring stable operation;

(ii) vary the value of process variables as a function of a measurement or time, or according to an optimizing function;

(iii) satisfy various constraints, e.g. prevent flooding in an extraction column.

The so-called control system, which can achieve these tasks, consists of well-designed equipment (the control hardware: controllers, control valves, valve motors, measuring devices, computers, etc.), as well as trained designers and operators.

The methods of control

The outlet temperature of the steam-heated heat exchanger in *Fig. 5.1* is to be held constant in spite of variations in feed rate and feed temperature. This can be accomplished in two ways:

(*a*) measuring the outlet temperature and manipulating steam rate *(feedback)*;
(*b*) measuring feed temperature and manipulating steam rate *(feedforward)*.

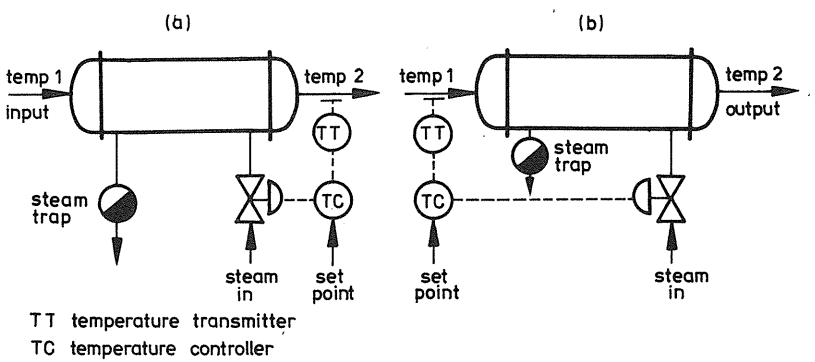

Fig. 5.1. Control of a heat exchanger: (*a*) feedback, (*b*) feedforward

In case (*a*) the measured value depends on the manipulation, in case (*b*) it does not. This is the basic difference between feedback and feedforward control.

Feedback control is able to eliminate the effects of changes in both feed temperature and feed rate, or any other disturbance (e.g. variation in steam pressure).

Feedforward control can eliminate only the effect of the disturbance on the measurement to which it applies; in some instances it does not even do that. In our heat exchanger example of *Fig. 5.1*, if the inlet temperature rises to a value for which the steam valve has to be closed down, the return of the outlet temperature to specification is not guaranteed. This is the main reason why feedforward is always used in combination with feedback.

Feedback has the disadvantage of requiring an offset in the control variable before control action. This implies the other main disadvantage of feedback: in case of erroneous application, feedback control may cause instability of the process controlled.

Stability

The definition of stability is known from mechanics. A process is *stable* if after a pulse-like disturbance it returns within some time to its original state. A step-like disturbance may force the stable system into a new steady state (or back to the original steady state, see *Fig. 5.2*), but the output of an unstable process will

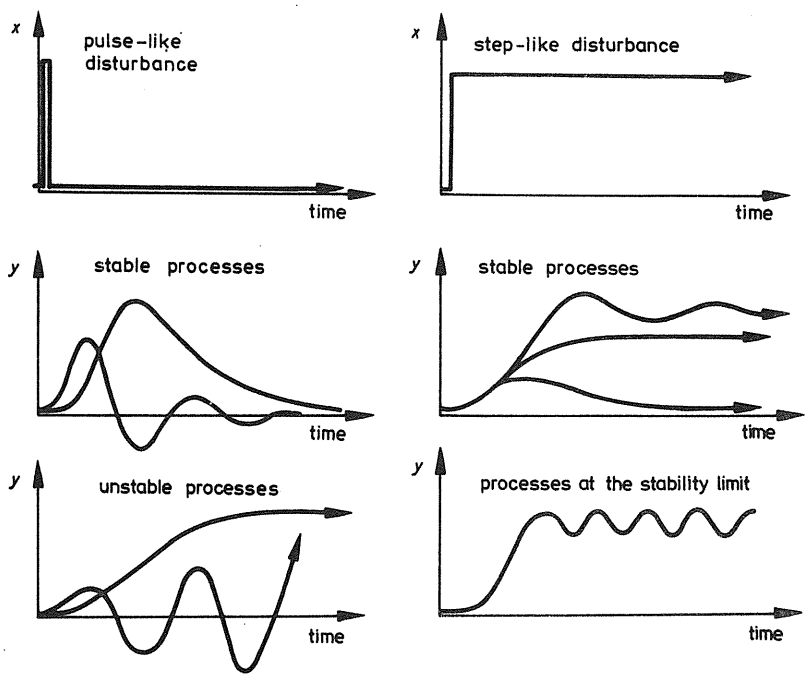

Fig. 5.2. Stable and unstable process responses

increase (or decrease) without limit (or until it reaches a natural limit, e.g. liquid level in a vessel will rise until it overflows). Some unstable processes show periodic instability: after a disturbance their output oscillates with increasing amplitude. At the stability limit the amplitude of the oscillations is constant.

Chemical processes are usually stable, but some unstable ones do exist. For example, irreversible exothermic reactions can be unstable and a reactor in which such a reaction takes place can only be operated in the steady state by the use of control.

The controlled process and its variables

A process has *input and output variables*: the *outputs are dependent* on the inputs, and on the characteristics of the process.

The general structure of a process is shown in *Fig. 5.3*.

The output variable to be controlled is the controlled variable. In general, it ought to be measured in order to control it; consider the output temperature of the heat exchanger in *Fig. 5.1*, for instance. However, it may occur that direct measurement of a controlled variable is impossible, e.g. the internal vapour velocity in a distillation column. In such cases, the variable is computed from inferential measurements.

The control variable is maintained at its desired value by changing a manipulated variable, thus eliminating the effect of uncontrolled variations in external load variables or disturbances. Both load and manipulated variables are inputs to the process.

Consider the heat exchanger in *Fig. 5.1* In this example,

(*i*) the *controlled variable* is the output temperature,
(*ii*) the *manipulated variable* is the flow rate of the heating steam, and
(*iii*) the *load variables* are the feed flow rate,
the feed temperature, etc.

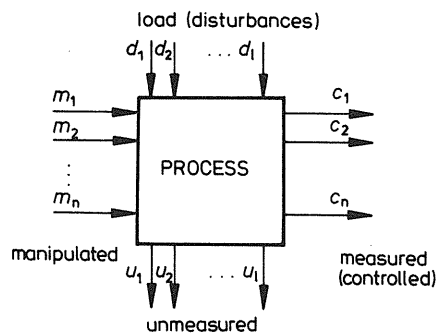

Fig. 5.3. Process and input/output variables

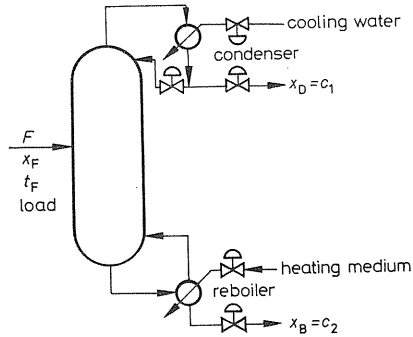

Fig. 5.4. Distillation column with possible manipulated variables

A system with one controlled output and one manipulated input is a Single Input, Single Output (SISO) system.

Many chemical processes have several controlled outputs and several possible manipulated inputs. A typical example is a distillation column. These Multiple Input, Multiple Output (MIMO) systems are a challenge to chemical engineers. They must consider the problems of: degrees of freedom (generally one controlled variable needs one manipulated variable); selection of the appropriate controlled variables, measurements and manipulations; proper pairing of controlled and manipulated variables.

Consider the distillation column in *Fig. 5.4*. The control objective is to obtain specified head and bottom products despite variations in the feed rate, composition and enthalpy. These latter are the load variables. The controlled variables are the distillate and the bottom compositions. To achieve their control, there is a choice of the following possible manipulated variables: distillate product rate, reflux rate, reboiler heat input, bottom product rate, etc.

The design of a control structure, i.e., the selection and pairing of controlled and manipulated variables requires a mathematical model of the process. For control system design purposes, the most practical model is the input-output model of the form:

$$c = f(m, d)$$

Control system and hardware

A measuring device is clearly needed in a control system in order to measure the controlled variable in a feedback control system (or, in a feedforward system, to measure the main load variable, or for inferential measurements in inferential control systems).

A transducer is used to convert measured physical quantities, such as temperatures, flow rates, levels or compositions, into signals which can easily be transmitted from the process to the controller, which may be installed some

hundreds of meters away. Easily transmitted signals may be electric current or voltage signals, or pneumatic signals (i.e. the pressure of compressed air or liquid). Measuring devices built together with transducers are called transmitters.

Transmission lines carry electric or pneumatic analogue signals from the process to the controller (and back). Sometimes amplifiers have to be installed into transmission lines. Sometimes it is more advantageous to transform analogue signals into digital ones and transmit several signals in series through a single transmission conduit. Pneumatic transmission lines, if long, may cause a considerable time delay.

The controller is the heart of the control process. It receives the control signal from the measuring device, compares it with its derived value, the set point, and according to the value and velocity of the error, gives a command signal to the process, telling it what manipulation is to be undertaken. The controller can be regarded as an analogue electric or pneumatic computer, on which simple control

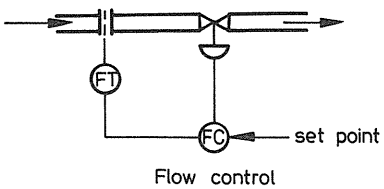

Fig. 5.5. Flow and level control loops and block diagram

laws can be implemented. Digital computers may also be used as controllers; they have the advantage that they can be programmed for more sophisticated control laws, and are capable of controlling more processes simultaneously. Digital controllers need an input and an output interface between the transmission line and the analogue instrumentation of the process.

The final control element is the device accomplishing the command given by the controller. In the chemical industry, the manipulated variable is most frequently a flow rate, which is manipulated by a control valve. The actuator of the control valve receives the command from the controller and moves the valve into the required position. Most valve actuators are pneumatic, since they are cheap and have a simple construction. Variable-speed pumps and compressors are also used as final control elements, as well as relay switches for on-off control.

A flow control and a level control system and the block diagram of a feedback control system are shown in *Fig. 5.5.*

5.2 Control dynamics

Feedback control is the basic regulating mechanism in control technology. However, negative feedback loops are capable of continuous oscillation and even instability — on the other hand, they may be sluggish and fail to control the controlled variable. It is thus of primary importance to understand the mechanism of feedback, and to examine its dynamics.

5.2.1 Basic concepts of feedback control

Negative feedback

Feedback may be negative or positive.

Positive feedback means that the controller works in the same direction as the disturbance. For example, in a heat exchanger where the output temperature is controlled *(Fig. 5.6)*, if the temperature is changing due to some disturbance, the

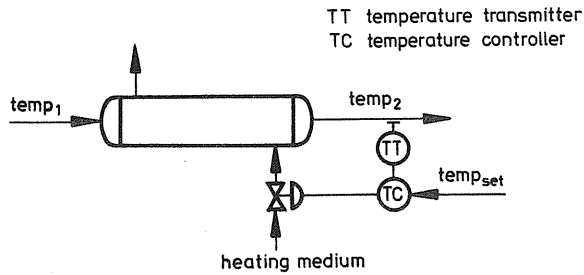

Fig. 5.6. Control of a heat exchanger

positive feedback controller will manipulate the flow of the heating medium so as to increase the temperature when it is above the set point and decrease when it is below. This feature is not conducive to regulation and causes instability.

Negative feedback works in the opposite sense. When the temperature, the controlled variable, is above the set point, i.e. it is higher than the desired value, the controller will reduce the heat input by manipulating some variable (called the manipulated variable). Of course if the temperature is below the set point it will increase the heat input. In short, negative feedback works always in the sense opposite to the measured deviation.

The block diagram

Figure 5.7 shows the block diagram and the simplified block diagram of a negative feedback loop, which may, for instance, be the heat exchanger control in *Fig. 5.6*. In the simplified block diagram the actuator, the control valve, the controlled process and the transmitter are all lumped together into the block labelled 'process'. Any classical simple control loop can be divided in this manner into two parts, the process and the controller, on the basis of the fact that the parameters of the controller are adjustable, but those of the process are not.

The mathematical treatment of networks similar to that represented in *Fig. 5.7*, and even of more complicated ones, is convenient if the signal transfer of all elements in the network can be represented by a simple linear input-output relationship of the form

$$dy = \underline{G}(\omega)dx \tag{5.1}$$

where y is the system output signal,
x is the system input signal,
$\underline{G}(\omega)$ is a function characterizing the signal transfer of the system. This is generally a function of the frequency of the input signal.

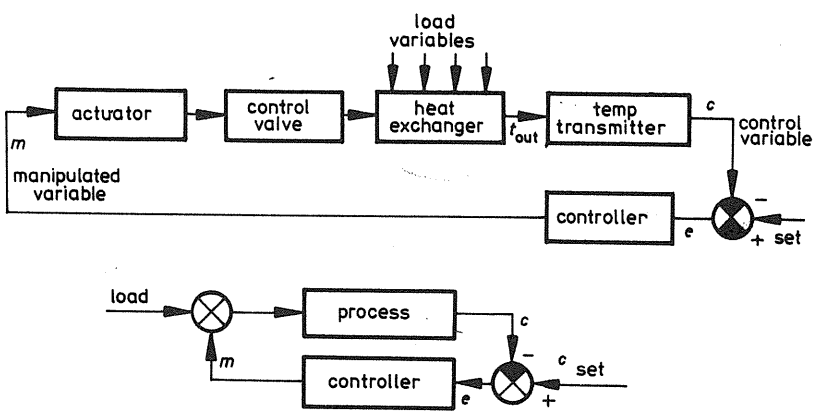

Fig. 5.7. Simplifying the block diagram

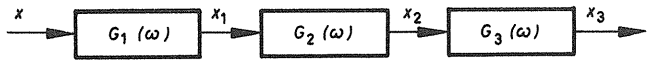

Fig. 5.8. Elements in series

In a linear network all linear transformations are valid. Thus the resulting frequency function of elements in series *(Fig. 5.8)* is the following:

$$dx_3 = G_3(\omega)dx_2 = G_3(\omega)G_2(\omega)dx_1 = G_3(\omega)G_2(\omega)G_1(\omega)dx$$

$$G(\omega) = \prod G_i(\omega) \tag{5.2}$$

The resulting frequency function of elements in parallel *(Fig. 5.9)* is:

$$dx_2 = G_1(\omega)dx_1 + G_2(\omega)dx_1 = [G_1(\omega) + G_2(\omega)]dx_1$$

$$G(\omega) = \sum G_i(\omega) \tag{5.3}$$

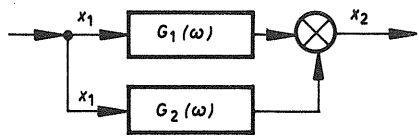

Fig. 5.9. Elements in parallel

The frequency function of a simple negative feedback control loop is derived as follows *(Fig. 5.10)*:

$$dx_S = G_L(\omega)dx_L + G_S(\omega)dx_m =$$

$$= G_L(\omega)dx_L + G_S(\omega)G_C(\omega)dx_{S,set} - G_S(\omega)G_C(\omega)dx_S$$

whence

$$dx_S = \frac{G_L(\omega)}{1 + G_S(\omega)G_C(\omega)}dx_L + \frac{G_S(\omega)G_C(\omega)}{1 + G_S(\omega)G_C(\omega)}dx_{S,set}. \tag{5.4}$$

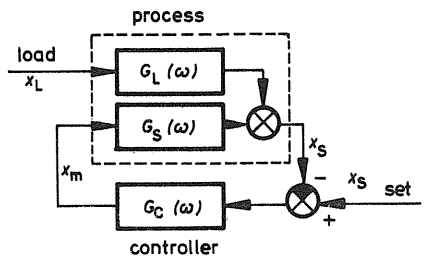

Fig. 5.10. Negative feedback control loop

Since the simple linear relationship in Eq.(5.1) greatly facilitates the treatment of control networks, one attempts to use linear relationships even for nonlinear systems, whenever possible. The technique of linearizing nonlinear input-output models will be presented in Section 5.3.3. It should be mentioned that most chemical engineering systems are nonlinear; some of them are suitable for linearization. Exceptions are, for instance, chemical reactors, where linearization may produce erroneous results.

The frequency function

The frequency response of a system is obtained experimentally by forcing the system, using a sinusoidal input (see *Fig. 5.11*):

$$x = a \sin \omega t \tag{5.5}$$

where a is the amplitude of the sine wave (–), and
ω is the frequency of the sine wave (radians/s).

After a certain time, depending on the characteristics of the forced system, the transients die out and the remaining steady-state output is a sine wave having the same frequency ω as the input:

$$y = b \sin (\omega t + \varphi) \tag{5.6}$$

but with a different amplitude b, and shifted by phase angle φ with respect to the input wave. The amplitude ratio and the phase angle are the components of $G(\omega)$, the frequency function vector:

the magnitude: $$|G(\omega)| = \frac{b}{a}(\omega) \tag{5.7}$$

the phase angle: $$\measuredangle G(\omega) = \varphi(\omega) \tag{5.8}$$

both depending on the frequency of the forcing function *(Fig. 5.11)*.

The frequency function vector can be represented graphically on a polar plot. $G(\omega)$ is conveniently expressed mathematically as a complex number, represented

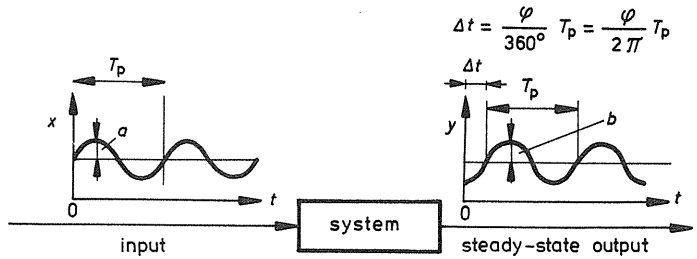

Fig. 5.11. Determination of the frequency response function

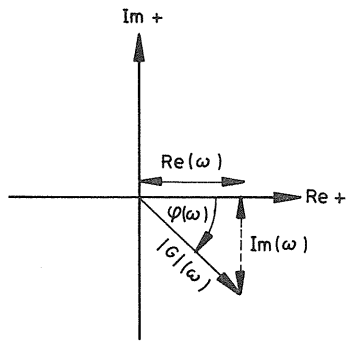

Fig. 5.12. Representation of the frequency function vector in the complex plane

in the complex plane *(Fig. 5.12)*:

$$\begin{aligned} G(i\omega) &= |G(i\omega)|\, e^{i\varphi(i\omega)} = \\ &= |G(i\omega)|\,[\cos\varphi(i\omega) + i\sin\varphi(i\omega)] = \\ &= \mathrm{Re}_{G(i\omega)} + i\,\mathrm{Im}_{G(i\omega)} \end{aligned} \tag{5.9}$$

This equation clearly shows that when frequency functions are in series (as in *Fig. 5.8*), then amplitude values multiply and phase shifts add. For parallel frequency functions *(Fig. 5.9)*, addition of the frequency functions is accomplished algebraically using the canonical form of the complex numbers (and graphically on a polar plot).

The most frequently used graphical representations of frequency functions over a broad frequency range are the Nyquist plot and the Bode plot. The Nyquist plot is a polar plot in which the end points of the individual vectors $G(\omega)$ are connected with a continuous line beginning at $\omega = 0$ and ending at $\omega = \infty$ *(Fig. 5.13)*. It has the drawback that frequencies ω are not indicated. The Bode plot consists of two

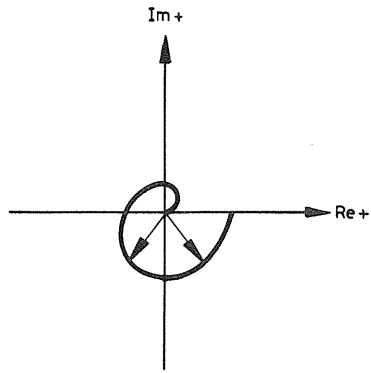

Fig. 5.13. Nyquist diagram

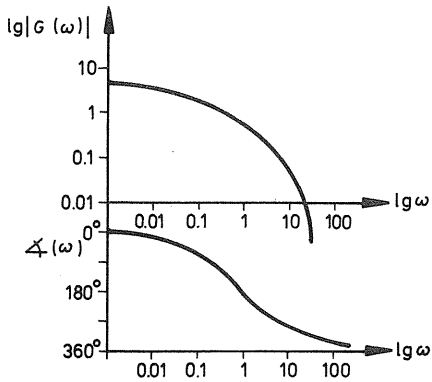

Fig. 5.14. Bode diagram

diagrams, one of which shows the logarithm of the amplitude ratio, and the other the phase lag (linear) both as a function of the logarithm of the frequency of the driving sine wave *(Fig. 5.14)*. The logarithmic scale of the amplitude diagram and the linear scale of the phase diagram are chosen to allow the multiplication of frequency functions in series by simple addition of distances in the diagram.

Oscillation in a feedback loop

Oscillation and resonance are characteristic of negative feedback loops. Resonance may be detrimental. A common example is the breakdown of a bridge caused by the rhythmical marching of a troop of soldiers.

The mechanism of oscillation is similar to a ball's bouncing. If a ball is being repeatedly struck (*i*) with the same force, (*ii*) at its highest position, it will continue to bounce to the same height (amplitude) at constant period. Should the force diminish, the amplitude will diminish, too. Should the ball be struck before its highest position, the period will be shortened.

In a feedback control loop a signal entering the loop circulates, as shown in *Fig. 5.15*. Oscillation is maintained if a signal which entered the loop returns with the same amplitude, shifted by a 360° phase angle. (The difference from ball bouncing is that the signal is generally continuous, instead of a series of pulses.)

There is an inherent $-180°$ phase shift in every negative feedback loop. Since the controller must always act in the opposite sense to the measured deviation, there must be a sign reversal somewhere in the loop. In *Fig. 5.15* the sign reversal is represented by the comparator. The sign reversal is equivalent to a $-180°$ phase shift. Thus, if the phase shifts exhibited by the process and the controller sum up to $-180°$, there is a phase shift of $-360°$ in the loop-condition (*ii*) for uniform oscillations. The frequency of the sinusoidal wave, for which

$$\measuredangle G_S(i\omega) + \measuredangle G_C(i\omega) = -180° \tag{5.10}$$

is the *critical frequency* of the loop.

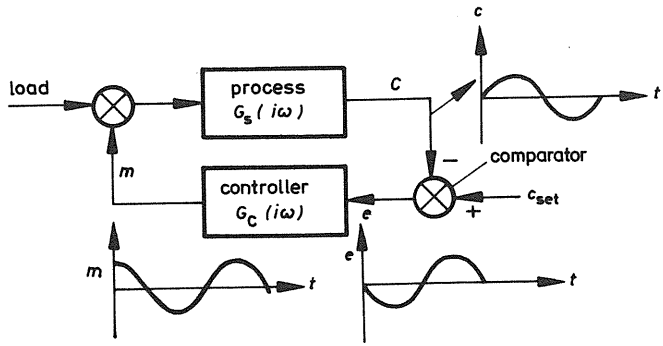

Fig. 5.15. Sign reversal in the negative feedback loop

Uniform oscillations are maintained in the control loop if the force exerted by the controller is the same after every cycle, that is, if the product of the amplitude ratios at the critical frequency is unity:

$$|G_C(i\omega_{cr})|\,|G_S(i\omega_{cr})| = 1 \tag{5.11}$$

Equations (5.10) and (5.11) can be combined into one equation, giving the representation of the vector in the complex plane:

$$G_C(i\omega_{cr})\,G_S(i\omega_{cr}) = -1 \tag{5.12}$$

This equation is a simple formulation of the Nyquist stability criterion. More exactly: for a stable loop, following the Nyquist curve of the open loop frequency function $(G_C(i\omega)\,G_S(i\omega))$ from $\omega=0$ in the direction $\omega\to\infty$, point (-1) must lie at the left. If point (-1) is at the right, the loop is unstable. If the Nyquist curve traverses point (-1), the loop is at the stability limit, exhibiting uniform oscillations at the critical frequency (see *Fig. 5.16*).

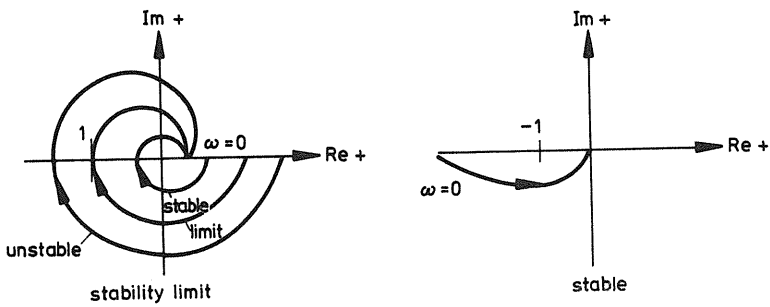

Fig. 5.16. The Nyquist stability criterion

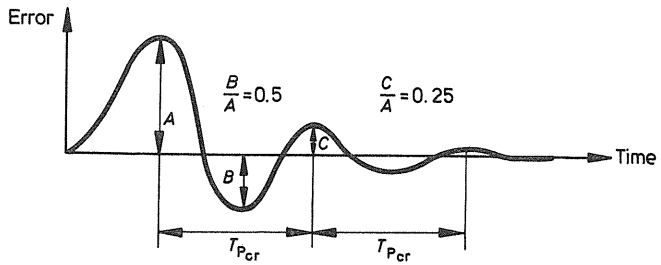

Fig. 5.17. Quarter-amplitude tuning

It is important to note that the Nyquist stability criterion uses the *open loop* frequency function to predict *closed loop stability!* Of course, it applies only to negative feedback loops.

How does a sine wave with the critical frequency enter the closed loop? Any pulse, impulse, or step signal may be decomposed into a sum of sine waves with different amplitudes, phases and frequencies. The sine components which have a frequency different from the critical frequency of the loop die out (because for these frequencies the loop gain is less than unity, or because for these frequencies the loop phase shift is different from $-360°$), and the loop will cycle with the critical frequency after a transient period. The forcing signal may enter the closed loop at any place, e.g. at the set point.

Damping

Damping of the oscillation (which has the critical frequency) is obtained by proper tuning of the controller. The controller is adjustable, whereas the parameters of the process are fixed.

Experience has shown that a gain product equal to 0.5 or less is generally advantageous. Then the amplitude of oscillation is attenuated by one-half each half cycle, and by one-quarter each full cycle. In *Fig. 5.17* A/B is 0.5, and A/C, the so-called decay ratio, is 0.25. This is the so-called 'quarter-amplitude damping', often used in industry as a rule of thumb for controller tuning.

With quarter-amplitude damping, any disturbance is virtually eliminated within three full cycles.

5.2.2 Control of a dead time process

Pure dead time processes are those that are most difficult to control. The reason for beginning with this process is to provide the reader with a grasp of the control problem.

Definition and characteristics of dead time

Definition

A dead time process has the special property that its response to any kind of disturbance is delayed by an interval called the dead time *(Fig. 5.18)*. During this interval no information about the occurrence of a disturbance is available. This is the cause of the difficulties encountered in the control of processes dominated by dead time. The output of a dead time process may be described by the following equation:

$$y[t]=A_S x[t-T_D] \tag{5.13}$$

that is, the output, y at time instant t is proportional to the input at time instant $[t-T_D]$. Proportionality factor A_S is the so-called steady-state gain of the process, its dimension is defined by those of y and x and by Eq.(5.13).

Dead time in the chemical industries is generally linked with transportation. As an example, consider a liquid product which is to be diluted with water to obtain

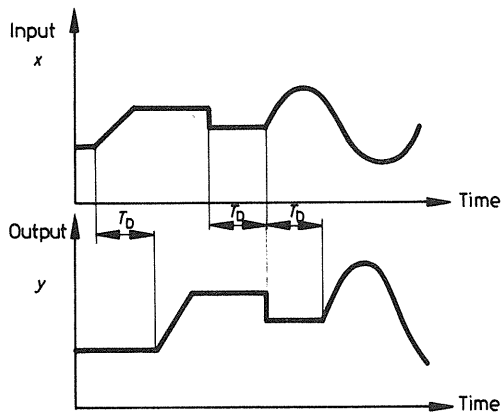

Fig. 5.18. Response of a dead time process

a 50% aqueous solution. Dilution water is fed directly into the product pipeline; concentration is measured by a density transmitter built downstream in the pipeline at a place where product and water are already thoroughly mixed *(Fig. 5.19)*. If backmixing in the pipeline is negligible, this is a pure dead time process, the dead time being equal to the transportation lag

$$T_D=\frac{L}{v} \tag{5.14}$$

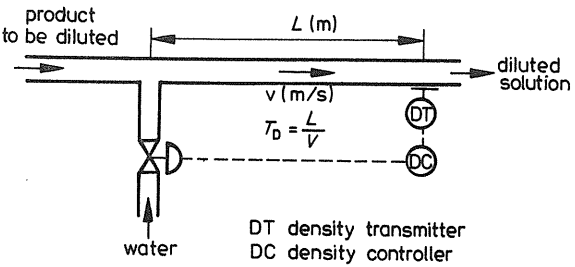

Fig. 5.19. Mixing in a pipeline is a dead-time process

The same phenomenon arises in tubular reactors, blending processes, and also in heat transfer processes; in fact, everywhere where molecules have to move through a distance in order to exert an effect. A typical dead time process is a conveyor belt transporting granulated solids. Sample pipelines to process analysers are also sources of dead time.

Frequency function: gain and phase shift

The frequency response of the dead time process is easily computed from its equation, as follows (refer to *Fig. 5.11*). The input to the process is a sine wave:

$$x = a \sin \omega t \tag{5.15}$$

where a is amplitude, ω is frequency, t is time. The steady-state output is also a sine wave. Its equation is obtained by substituting Eq.(5.15) into Eq.(5.13):

$$y = aA_S \sin [\omega(t - T_D)] = aA_S \sin (\omega t - \omega T_D) . \tag{5.16}$$

The output amplitude is $b = aA_S$.

The output lags behind the input by the phase angle:

$$\varphi = \measuredangle G(i\omega) = -\omega T_D . \tag{5.17}$$

The gain or magnitude of vector $G(i\omega)$ is the ratio of the amplitudes of the output and input sine waves:

$$|G(i\omega)| = \frac{b}{a} = \frac{aA_S}{a} = A_S \tag{5.18}$$

Thus, for a dead time process, $\measuredangle G(i\omega)$ depends on the frequency of the input, while $|G(i\omega)|$ does not.

We define the gain of the process as a product of the steady state and the dynamic gains:

$$|G(i\omega)| = A_S |g(i\omega)| \tag{5.19}$$

This is the definition of the dynamic gain, which will be found useful later. For the dead time process,

$$|g(i\omega)| = \frac{|G(i\omega)|}{A_S} = \frac{A_S}{A_S} = 1$$

We define the normalized frequency function vector of a process as

$$\underline{g}(i\omega) = \frac{\underline{G}(i\omega)}{A} \tag{5.20}$$

Of course, $\measuredangle g(i\omega) = \measuredangle G(i\omega)$ for all processes. A_S is a scalar quantity.

The normalized frequency function vector of a dead time process is shown in *Fig. 5.20* in the form of Nyquist and Bode plots, using coordinates that are independent of the parameters of the individual system.

Proportional control of dead time

A proportional controller (or *P*-controller) is a simple device with an output that is proportional to the error at every time instant

$$m = A_C e \tag{5.21}$$

where e is the error, m is the manipulation and A_C is the proportional gain of the controller.

The *P*-controller has no phase lag and its gain is constant for all input frequencies. Thus, for the proportional controller

$$|G_C(i\omega)| = A_C$$

$$|g_C(i\omega)| = 1$$

$$\measuredangle g_C(i\omega) = 0° .$$

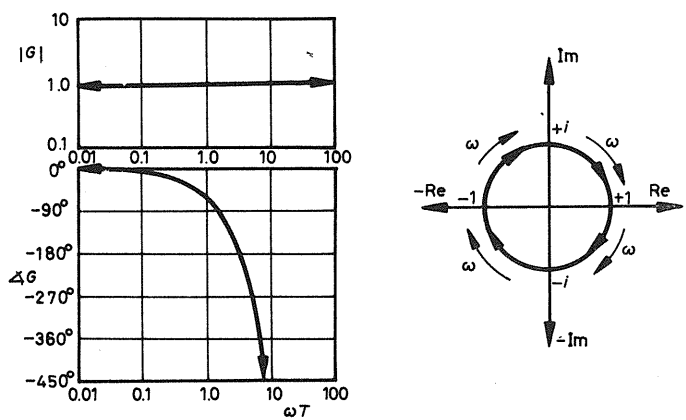

Fig. 5.20. Nyquist and Bode plots of dead time processes

Figure 5.21 illustrates a simplified block diagram of a dead time process controlled by a *P*-controller. The phase lags are as follows:

process $$\varphi_S = -\omega T_D$$

$$\left(\text{radians if } \omega \frac{\text{rad}}{\text{s}}, \text{ degrees if } \omega \frac{\text{deg}}{\text{s}}\right)$$

controller $$\varphi_C = 0$$

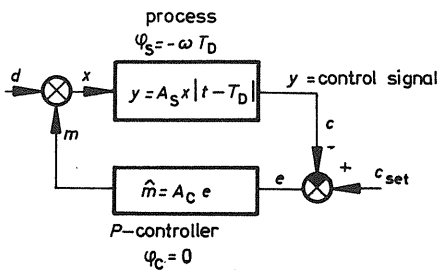

Fig. 5.21. Proportional control of dead time

Critical frequency

For a sine wave having the critical frequency, the phase lags of the process and the controller sum to $-180°$, which causes the sine wave to return to its starting point with $-360°$, i.e. zero phase shift:

$$\varphi_S + \varphi_C = -\omega_{cr} T_D + 0 = -180° \tag{5.22}$$

$$-\omega_{cr} T_D = -\frac{360°}{T_{P,cr}} T_D = -180° \tag{5.23}$$

From Eq.(5.23) a simple and exact expression for the critical period is obtained. Nothing has been neglected in its derivation.

$$T_{P,cr} = \frac{360°}{180°} T_D = 2T_D \tag{5.24}$$

Damping

To determine the conditions under which continuous cycling occurs (i.e. the control loop is at its stability limit), the Nyquist criterion is used, cf. Eqs (5.10) —(5.12). Thus, the stability limit is attained if the gain of the proportional control-

ler is set so that it satisfies the following equation:

$$|G_C(i\omega_{cr})|\,|G_S(i\omega_{cr})| = A_C A_S = 1 \tag{5.25}$$

This means that, if a signal composed of sinusoidal waves of a broad frequency spectrum (e.g. a pulse signal) enters the control loop in *Fig. 5.21* at any point, i.e. either as a disturbance, or as a set point change, then after a time necessary for the sine waves with other frequencies to die out, all signals in the loop, i.e. c, e and m will cycle with constant amplitude and a period equal to $2T_D$.

If the gain of the proportional controller is set so that

$$A_C A_S > 1$$

then the amplitudes of the successive cycles (with period $2T_D$) will be expanding, and the system is unstable.

As a rule of thumb, controllers are tuned for quarter-amplitude damping in order to obtain stable control which restores the steady-state within three full cycles *(Fig. 5.17)*.

For this damping a *P*-controller controlling a dead time process is tuned to produce a gain product equal to 0.5:

$$A_C A_S = 0.5\,.$$

With this tuning, the period of the signals in the control loop is essentially the critical one, $T_{P,cr} = 2T_D$ (it is, in fact, slightly longer), and the steady state is attained after a time of $3T_{P,cr} = 6T_D$.

Open-loop frequency function

To check the stability of the closed control loop by the Nyquist criterion, it is advisable to draw the Nyquist plot of the open-loop frequency function. In many cases it is sufficient to use the Bode plot, which is easier to construct.

For a dead-time process controlled by a *P*-controller: *a)* The Nyquist plot of the dead-time process is a circle with its centre at 0 and a radius of A_S. *b)* The frequency function of the *P*-controller is a scalar quantity:

$$G_C(i\omega) = A_C$$

Thus, the Nyquist plot of the open loop frequency function

$G_C(i\omega)G_S(i\omega)$ is also a circle,

with its centre at 0 and a radius of $A_S A_C$.

The Nyquist and the Bode plots for an open loop composed of dead time and a *P*-controller are shown in *Fig. 5.22*. For $A_C A_S < 1$, following the Nyquist plot in the direction of increasing frequencies, the point -1 is seen at the left, indicating that the loop is stable.

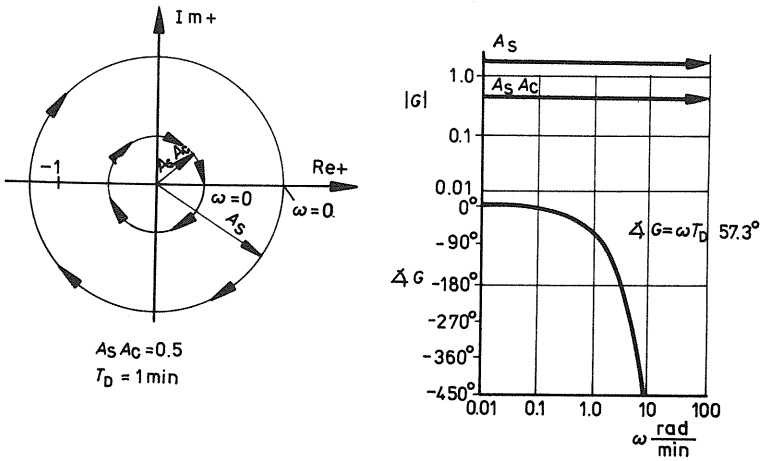

Fig. 5.22. Open loop frequency function for *P*-control of dead time process

Steady-state offset

A control system is expected to keep the controlled variable at its set point, i.e. the error signal is kept at a value of zero (cf. *Fig. 5.21*). A load disturbance or a set point change will cause a transient error, but the controller is expected to eliminate it by appropriate manipulation: in the case of a load disturbance, it must return the control signal to its previous steady-state value; in the case of a set point change it must transduce the control signal to its new command value.

The *P*-controller has the deficiency of being unable to accomplish the above task: it generally works with a steady-state offset. This follows from its control law represented by Eq.(5.21):

$$\hat{m} = A_C e \tag{5.26}$$

where $\hat{m}$ means the deviation of the manipulated variable from a pre-set value m_0 for zero error and normal operative conditions: $c_{set,0}$ and d_0. In order to balance a deviation from $c_{set,0}$ or d_0 a value m different from m_0 is needed, but this is produced only if $e \neq 0$ in the steady state.

The steady-state offset can be calculated from the steady-state gains in the loop. This calculation is valid only for stable systems. (Refer to the notation of *Fig. 5.20.*) The steady-state output of the process is

$$y = \hat{c} = A_S x = A_S \hat{m} + A_S \hat{d} \tag{5.27}$$

where $\hat{c} = c - c_0 = c - c_{set,0}$
$\hat{m} = m - m_0$
$\hat{d} = d - d_0$

Combine Eq.(5.26) and $e=\hat{c}_{set}-\hat{c}$ with Eq.(5.27):

$$\hat{c}=A_S A_C(\hat{c}_{set}-\hat{c})+A_S\hat{d} \tag{5.28}$$

Rearrange to give

$$\hat{c}=\frac{A_S A_C}{1+A_S A_C}\hat{c}_{set}+\frac{A_S}{1+A_S A_C}\hat{d} \tag{5.29}$$

Example 5.2.1

A *P*-controller controlling a dead time process with $A_S=0.3$ is set for quarter amplitude damping, i.e. $A_S A_C=0.5$. What is the steady-state change of the controlled variable after a unity change

(*a*) in the set point, $\hat{c}_{set}=1$

(*b*) in the load variable, $\hat{d}=1$?

(*a*) For quarter amplitude damping

$$\frac{A_S A_C}{1+A_S A_C}=\frac{0.5}{1+0.5} \quad \text{thus} \quad \hat{c}=\frac{0.5}{1.5}1=0.333;$$

(*b*)

$$\hat{c}=\frac{A_S\hat{d}}{1+0.5}=\frac{A_S\hat{d}}{1.5}=\frac{0.3}{1.5}=0.2$$

Without control, a unit change in d would cause a change in c, the controlled variable of

$$A_S\hat{d}=0.3\cdot 1=0.3$$

Thus, proportional control leaves $\frac{0.2}{0.3}=\frac{1}{1+A_S A_C}=0.66$, i.e., 66% of the offset without control!

To eliminate steady-state error, $A_S A_C=\infty$ would be needed, but this is prohibited by the stability requirement.

Integral control of dead time

Integral controller, definition

The *P*-controller is unable to eliminate offset, so another control mode is needed. The (integral) *I*-controller is a device producing a steady state only in the case of zero offset:

$$\frac{dm}{dt}=\frac{1}{I}e \tag{5.30}$$

$\left(\text{thus, } \frac{dm}{dt} = 0 \text{ if } e = 0\right)$, or, integrating Eq.(5.30):

$$\hat{m} = \frac{1}{I} \int_0^t e \, dt \tag{5.31}$$

Here I is the tuning parameter of the I-controller. The output of the I-controller is proportional to the error integral. The response to a step input is shown in *Fig. 5.23*. It is shown that I is the time necessary for the output to repeat the input: it is named *integral* or *reset* or *repeat* time.

Thus, the integral control mode eliminates offset in the steady state. There is, however, certainly a penalty for that.

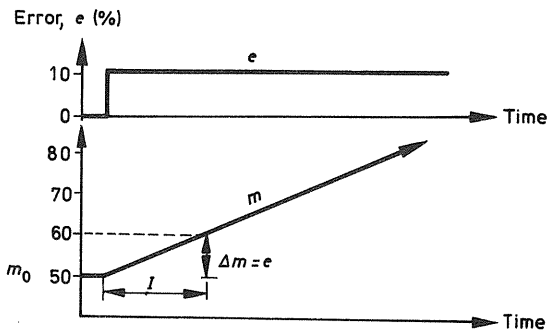

Fig. 5.23. Integral time of the I-controller

I-controller, frequency function

The gain and phase angle of the I-controller will be calculated before examining its effect on control loop behaviour. The frequency function is easily calculated using Eq.(5.31) (refer to *Fig. 5.11*). The input to the controller is the sine wave:

$$e = a \sin \omega t \tag{5.32}$$

Substitute this into Eq.(5.31)

$$\hat{m} = \frac{1}{I} \int a \sin \omega t \, dt = \frac{a}{I} (-\cos \omega t) \tag{5.33}$$

Since

$$-\cos \varphi = -\sin (90° - \varphi) = \sin (\varphi - 90°) \tag{5.34}$$

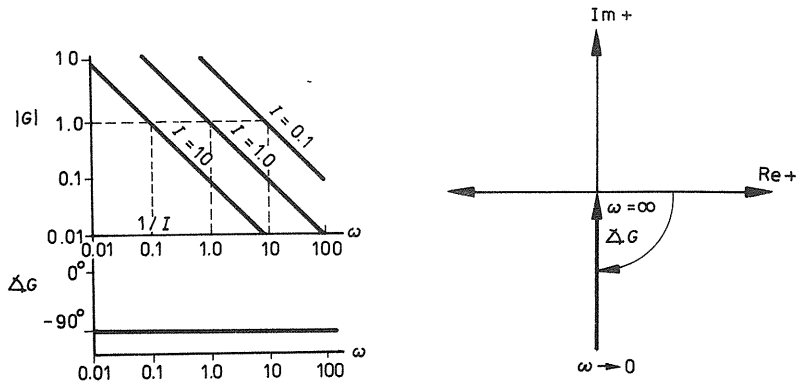

Fig. 5.24. Nyquist and Bode plot of the *I*-element

we obtain

$$\hat{m} = \frac{a}{I\omega} \sin(\omega t - 90°) \tag{5.35}$$

whence the magnitude of the frequency function vector, or gain, is

$$|G(i\omega)| = |g(i\omega)| = \left(\frac{a}{I\omega}\right) \Big/ (a) = \frac{1}{I\omega} \tag{5.36}$$

It is inversely proportional to the frequency of the input: when $\omega \to 0$, $|G| \to \infty$; when $\omega \to \infty$, $|G| \to 0$.

The phase angle of the frequency function vector is clearly

$$\measuredangle G(i\omega) = \measuredangle g(i\omega) = -90° \tag{5.37}$$

independent of the frequency of the input. Thus, all the frequency function vectors of the *I*-controller lie along the negative imaginary axis of the complex plane. The Nyquist and the Bode plots of the frequency function are shown in *Fig. 5.24.*

Recall that for the dead time process it is the magnitude of the frequency function which is independent; the phase lag is proportional to frequency.

The control loop with dead time process and I-controller

The simplified block diagram of the control loop is shown in *Fig. 5.25.*

In order to determine the *critical frequency* of the loop, the phase shifts are needed. Process [Eq.(5.17) with ω in degree/s]:

$$\measuredangle G_S(i\omega) = -\omega T_D$$

I-controller [Eq.(5.37)]:

$$\measuredangle G_C(i\omega) = -90°$$

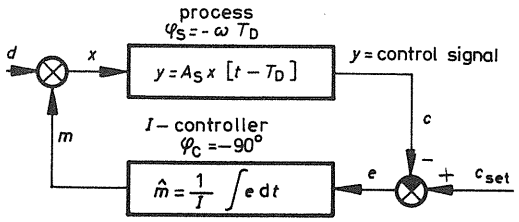

Fig. 5.25. *I*-control of a dead time process

Apply the Nyquist criterion (Eq.5.10)

$$-180° = \measuredangle G_S(i\omega) + \measuredangle G_C(i\omega) = -\omega_{cr} T_D - 90° \tag{5.38}$$

Whence:

$$\omega_{cr} = \frac{90°}{T_D}\,\frac{\text{degree}}{\text{s}} = \frac{\pi}{2T_D}\,\frac{\text{radian}}{\text{s}} \tag{5.39}$$

The critical period:

$$T_{P,cr} = \frac{360°}{\omega_{cr}} = 4T_D \tag{5.40}$$

The penalty for eliminating offset is found in Eq.(5.40): the critical period of the control loop with an *I*-controller is double that found with a *P*-controller, and so is the time of a transient. This result is quite general for any process and control loop: an *I*-controller slows control down.

Damping in the control loop with an *I*-controller is obtained by the proper setting of the sole tuning parameter *I*.

At the stability limit [cf. Eqs (5.11), (5.18) and (5.36)]

$$1 = |G_S(i\omega_{cr})|\,|G_C(i\omega_{cr})| = A_S \frac{1}{I\omega_{cr}} = \frac{2T_D A_S}{\pi I} \tag{5.41}$$

whence

$$I_{cr} = \frac{2}{\pi} T_D A_S \tag{5.42}$$

If quarter-amplitude damping is desired, then

$$0.5 = |G_S(i\omega_{cr})|\,|G_C(i\omega_{cr})|$$

and

$$I_{(0.5)} = \frac{4}{\pi} T_D A_S \tag{5.43}$$

In order to damp oscillations, the reset time has to be augmented. The minimum value of *I* is the critical one given by Eq.(5.42).

The choice of the reset time does not influence the critical frequency, which depends only on the dead time of the process.

The open loop frequency function for the integral control of dead time is plotted in the Bode diagram *(Fig. 5.26)* using the following equations:

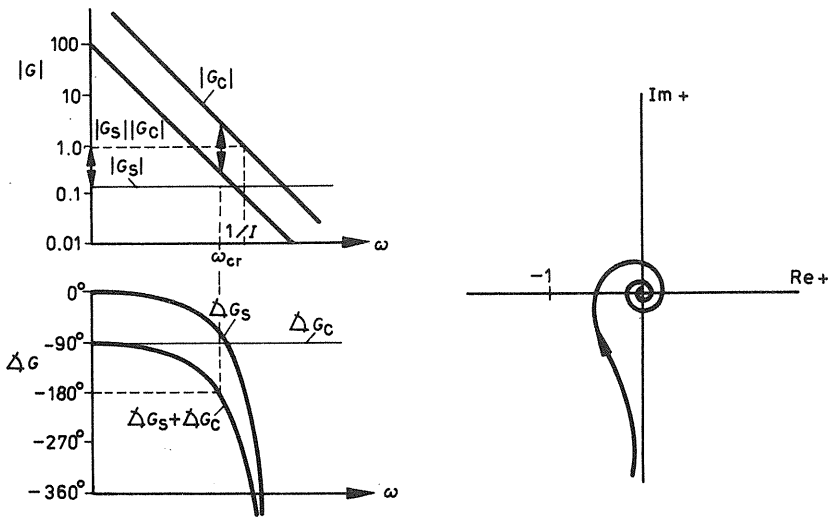

Fig. 5.26. Open loop Nyquist and Bode plots for *I*-control of a dead time process

gain: $$|G_S|\,|G_C| = A_S \frac{1}{I\omega}$$

phase shift: $$\measuredangle G_S + \measuredangle G_C = -T_D\omega \times 57.3° - 90°$$

ω in radians/s

Both gain and phase shift are monotonic functions of frequency, so stability can also be checked from the Bode plot. The Nyquist plot is constructed using the gain and phase values from the Bode plot. The point -1 is seen at the left, and thus the control loop is stable.

5.2.3 Control of capacity processes

By contrast with dead-time processes, capacity processes are easy to control. The input-output mathematical model of a pure capacity process is similar to that of the *I*-controller:

$$y = A_I \int x\,dt \tag{5.44}$$

where x is input, y is output, both being deviations from the steady-state values,

$$A_I \text{ is gain, its dimension being } \frac{[y]}{[x] \times \text{time}}$$

Pure capacity is thus an integrating element, as is the *I*-controller, for which $A_I = \frac{1}{I}$ [cf. Eq.(5.31)].

The frequency function of the pure capacity is similar to that of the *I*-controller, which has been derived in the previous section. With reference to Eq.(5.36), the magnitude of the frequency function vector is

$$|G_I(i\omega)| = \frac{A_I}{\omega} \tag{5.45}$$

The phase shift is independent of the frequency of the input sine wave [cf. Eq.(5.37)]:

$$\measuredangle G_I(i\omega) = -90° \tag{5.46}$$

An example of a pure capacity process is a liquid tank *(Fig. 5.27)*. The output of the process is essentially the quantity (mass or volume) of the liquid stored in the tank; it is measured by the liquid level. A metering pump delivers outflow (and thus outflow is independent of the liquid level). To control the level, the only possible manipulative variable is the inflow.

The mathematical model of the tank is obtained by writing down its total mass balance:

$$A\frac{dH}{dt} = W_{in} - W_{out} \tag{5.47}$$

whence the change in level is:

$$\hat{H} = \frac{1}{A}\int_0^t (W_{in} - W_{out})\,dt \tag{5.48}$$

Equation (5.48) is identical to Eq.(5.44), where $y = \hat{H}$, $A_I = \frac{1}{A}$, $x = (W_{in} - W_{out})$. Thus this tank is an integrating element, the frequency function of which is shown in *Fig. 5.24*, with a step response similar to that shown in *Fig. 5.23*.

The step response of the integrating process indicates that any small change in the flows, which implies a non-steady-state:

$$W_{in} - W_{out} \neq 0 \tag{5.49}$$

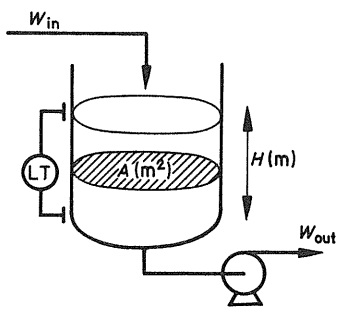

Fig. 5.27. Pure capacity element: liquid tank with fixed outflow

will cause the tank to flood or empty. This attribute is known as non-self-regulation and that is why an integrating process cannot be left for a long time without control. NOTE: Liquid tanks for which the outflow depends on liquid level, or gas tanks for which both the inflow and the outflow may depend on the pressure difference across a restriction (pipeline or valve) are capacities, too, but they are self-regulating capacities. Such processes will be treated in Section 5.3.

Control of a pure capacity by P-controller

The scheme of the control loop is shown in *Fig. 5.28.*

When applying the Nyquist stability criterion to this control loop, it can be seen that

$$\varphi_S + \varphi_C = -90°$$

for all frequencies. The loop phase lag never reaches 180°, thus there will be no oscillation. Therefore, the proportional band can be set without any theoretical limit.

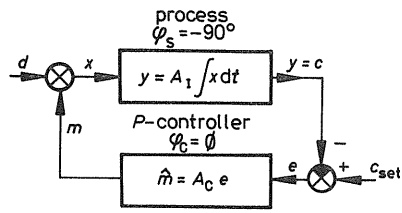

Fig. 5.28. *I*-element controlled by *P*-controller

In consequence, pure capacity processes can be controlled by a simple *P*-controller set to a high A_C proportional gain (narrow proportional band), with no oscillation or steady-state offset.

An *I*-controller cannot be used to control an *I*-process because both the *I*-controller and the pure capacity process exhibit a constant phase shift of −90° for all frequencies *(Fig. 5.29)*:

$$\varphi_S = \varphi_C = -90°$$

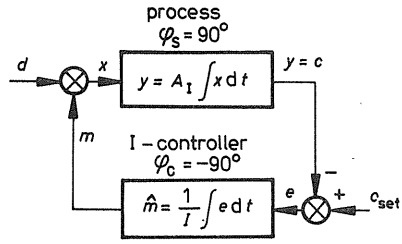

Fig. 5.29. *I*-element controlled by *I*-controller

thus the loop phase shift sums to $-180°$

$$\varphi_S + \varphi_C = -180°$$

independently of the frequency and of controller tuning.

5.2.4 The *PI*-controller

Due to the fact that the *I*-controller may produce unstable control if applied to a process with a free *I*-element, it is almost always combined with a proportional controller.

The ideal *PI*-controller consists of a *P*- and an *I*-controller in parallel *(Fig. 5.30)*. This combines the best features of the *P*- and *I*-controllers: the proportional offset is eliminated with little loss of response speed (compared to an *I*-controller) and stable operation can be achieved with almost any process. This controller combination is the one that is most frequently used. It is described by the following equation:

$$y = A_P \left(x + \frac{1}{I} \int_0^t x \, dt \right) \tag{5.50}$$

or

$$y = A_P x + A_I \int_0^t x \, dt \tag{5.51}$$

where y is output (for a controller, the manipulation),
x is input (for a controller, the error),
A_P is the static gain of the *P*-controller,
I is the integral time of *PI*-controller,
$A_I = \frac{A_P}{I}$ is the static gain of the *I*-controller, and t is time.

The step response of the *PI*-controller is obtained from Eq. (5.50) if the input is constant:

$$x = 0 \quad \text{for} \quad t \leqq 0$$

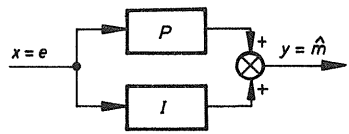

Fig. 5.30. Scheme of *PI*-controller

$$x=a \quad \text{for} \quad t>0$$
$$y=A_P\left(a+\frac{at}{I}\right). \tag{5.52}$$

For $t=I$, $y=A_P(2a)$ (see *Fig. 5.31*).

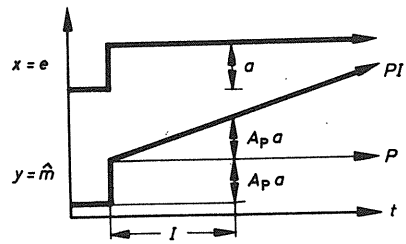

Fig. 5.31. Step response of a *PI*-controller

The integral, or reset, or repeat time I_I of a *PI*-controller is the time necessary for the *I* part to repeat the response of the *P*-part. (The repeat time of an *I*-controller is the time needed to repeat the input; cf. *Fig. 5.23.*)

The frequency function of the *PI*-controller can be obtained as the vectorial sum of those of the *P*- and the *I*-controllers [Eq. (5.3)]. This is best represented on a polar diagram *(Fig. 5.32.)*.

From *Fig. 5.32* it is evident that the magnitude of the vector $G_{PI}(i\omega)$ is:

$$\begin{aligned}|G_{PI}(i\omega)| &= \sqrt{|G_P(i\omega)|^2+|G_I(i\omega)|^2} \\ &= \sqrt{A_P^2+\left(\frac{A_P}{I\omega}\right)^2} \\ &= A_P\sqrt{1+\left(\frac{1}{I}\right)^2}\end{aligned} \tag{5.53}$$

The phase shift

$$\begin{aligned}\sphericalangle G_{PI}(i\omega) &= -\varphi_{PI}(i\omega)=\arctan\left(\frac{-G_I(i\omega)}{T_P(i\omega)}\right) \\ &= \arctan\left(\frac{-1}{I\omega}\right)\end{aligned} \tag{5.54}$$

Both the magnitude and the phase shift of a *PI*-controller depend on frequency.

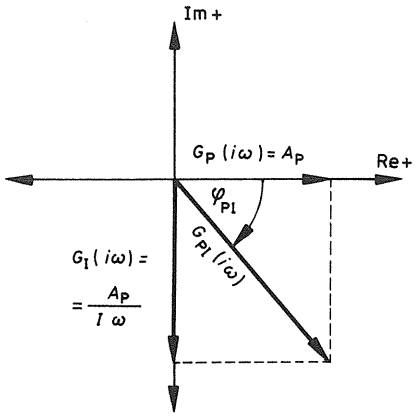

Fig. 5.32. Frequency function of a *PI*-controller

For

$$I\omega \to 0 \quad |G_{PI}| = \frac{A_P}{I} \to \infty \quad -\varphi_{PI} \to -90°$$

$$I\omega \to \infty \quad |G_{PI}| \to A_P \qquad -\varphi_{PI} \to 0°$$

Thus for low frequencies (near the steady state) the controller behaves as an *I*-controller, eliminating steady-state offset; and for high frequencies it behaves as a *P*-controller with very small phase lag — thus if properly set it will only slightly increase the phase lag of the control loop, the critical period and the settling time.

The phase lag of the *PI*-controller depends only on the parameter *I*. *I* should be set so that the phase lag of the controller is smaller than 10° at the critical frequency of the process.

Example 5.2.2

Consider a dead-time process with the parameters $A_S = 2$, $T_D = 2$ min. The task is to control it (*a*) with a *P*-controller; (*b*) with an *I*-controller; (*c*) with a *PI*-controller. The controller is to be tuned to quarter-amplitude damping. Offset and settling time are to be estimated.

Solution

(*a*) Tune the *P*-controller to quarter-amplitude damping:

$$A_C A_S = A_C \cdot 2 = 0.5$$

$$A_C = 0.25$$

Offset is calculated using Eq. (5.29):

$$\frac{\hat{c}}{\hat{c}_{set}} = \frac{A_S A_C}{1 + A_S A_C} = \frac{0.5}{1.5} = 0.333$$

$$\frac{\hat{c}}{d} = \frac{A_S}{1 + A_S A_C} = \frac{2}{1.5} = 1.333$$

The settling time is $3T_{P,\,cr}$. Use Eq. (5.24):

$$T_{set} = 3T_{P,\,cr} = 3 \cdot 2 \cdot T_D = 3 \cdot 2 \cdot 2 \text{ min} = 12 \text{ min}$$

(*b*) With an *I*-controller there is no offset. To tune the controller for quarter-amplitude damping, use Eq. (5.43):

$$I_{(0.5)} = \frac{4}{\pi} T_D A_S = \frac{4}{\pi} \cdot 2 \cdot 2 \text{ min} = 5.1 \text{ min}$$

The value of I (for an *I*-controller) does not affect the phase shift (which is always $-90°$ for an *I*-controller). The critical period of the control loop is obtained from Eq. (5.40):

$$T_{P,\,cr} = 4T_D = 8 \text{ min},$$

$$T_{set} = 3T_{P,\,cr} = 24 \text{ min}$$

(*c*) With a *PI*-controller, there is no offset. The critical frequency, where the dead-time process alone exhibits $-180°$ phase shift

$$\omega_{cr} T_D = \pi \rightarrow \omega_{cr} = \frac{\pi}{T_D} = \frac{\text{rad}}{\text{min}}$$

This is the frequency at which the *PI*-controller should produce a $-10°$ phase shift. Use Eq. (5.54):

$$-\varphi_{PI} = \arctan\left(\frac{-1}{I\omega}\right) = -10°,$$

$$I = 3.61 \text{ min}$$

The critical frequency of the control loop is where

$$\varphi_S + \varphi_C = -T_D\omega + \arctan\left(\frac{-1}{I\omega}\right) = -\pi$$

By trial and error, this is

$$\omega_{cr} = 1.40 \frac{\text{rad}}{\text{min}},$$

whence

$$T_{P,\,cr} = \frac{2\pi}{\omega_{cr}} = 4{,}49 \text{ min},$$

and the settling time is $3T_{P,\,cr} = 13.5$ min, which is slightly more than the 12-min value found for *P*-control. For quarter-amplitude damping A_P remains to be calculated.

$$|G_S|\,|G_C| = 0.5\,,$$

where [Eq. (5.53)]

$$|G_C| = A_P \sqrt{1+\left(\frac{1}{I\omega}\right)^2}$$

$$= A_P \sqrt{1+\left(\frac{1}{3.61 \cdot 1.40}\right)^2} = A_P \cdot 1.019$$

whence

$$A_P = \frac{0.5}{2 \cdot 1.019} = 0.245,$$

i. e. slightly smaller than 0.25, the value found for the *P*-controller.

5.2.5 *PD*- and *PID*-controllers

Dead-time processes are well controlled with *PI*-controllers as shown in the previous sections. In the case of multicapacity processes the critical period may be very slow and it would be advantageous to speed up control. In such cases control involves derivative action.

The input-output equation of the ideal derivative element is:

$$y = D\frac{dx}{dt} \tag{5.55}$$

where y is output, x is input, D is derivative time constant, and t is time.

Of course, Eq. (5.55) cannot be realized because the derivative of the input at time instant t cannot be known; it can be estimated only from previous measurements. So a real derivative element always contains a small time delay.

Moreover, for an ideal step input which would occur in zero time and for which $\frac{dx}{dt} = \infty$, the ideal derivative element should produce infinite output in zero time delay, which is also impossible.

The frequency function of the *D*-element is obtained by solving Eq. (5.55) for $x = a \sin \omega t$:

$$y = D\omega \cos\omega t = aD\omega \sin(\omega t + 90°)$$

whence $|G_D(i\omega)| = D\omega$ dependent on ω

$\measuredangle G_D(i\omega) = +90°$ independent from ω

The frequency function of the *D*-element in the complex plane in a series of vectors lying along the positive imaginary axis; its Nyquist and Bode plots are shown in *Fig. 5.33*.

The ideal *D*-element would thus reduce the loop lag by 90°; a real *D*-element can produce a phase shift of +60°.

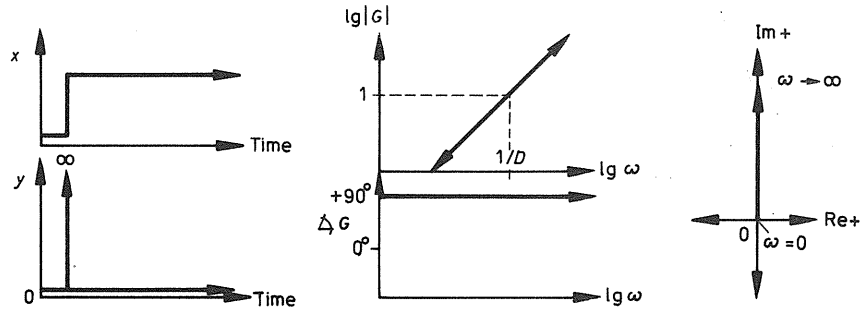

Fig. 5.33. Plots of a *D*-element

The derivative element cannot be applied in a controller as the only control mode, since it responds only to a change in its input, and a constant input has no effect at all on its output.

The *D*-element is therefore used only in combination with a *P*-controller (*PD*-controller) or with a *PI*-controller (*PID*-controller). Since the *PD*-controller does not contain the *I* operating mode, it cannot eliminate steady-state offset completely and is used only for special applications.

The *PID*-controller is the most general controller. The ideal *PID*-controller consists of a *P*, an *I* and a *D* element in parallel *(Fig. 5.34)*, its input-output equation is:

$$y = A_P x + A_I \int_0^t x\,dt + A_D \frac{dx}{dt} = A_P \left(x + \frac{1}{I} \int_0^t x\,dt + D \frac{dx}{dt} \right) \tag{5.56}$$

where $y = m - m_0$ is the output or change in the manipulated variable,

m_0 is the output of the controller in the steady state, when the error, its integral and its derivative are zero,

$x = e$ is the input of the controller: the error signal, A_P, I, D are tuning parameters of the controller, and

t is time.

The Nyquist plot of the *PID*-controller is obtained by adding the Nyquist plots

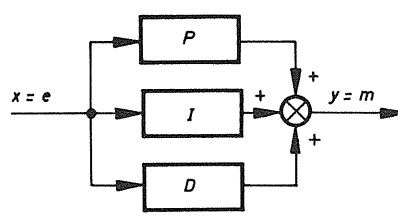

Fig. 5.34. Scheme of a *PID*-controller

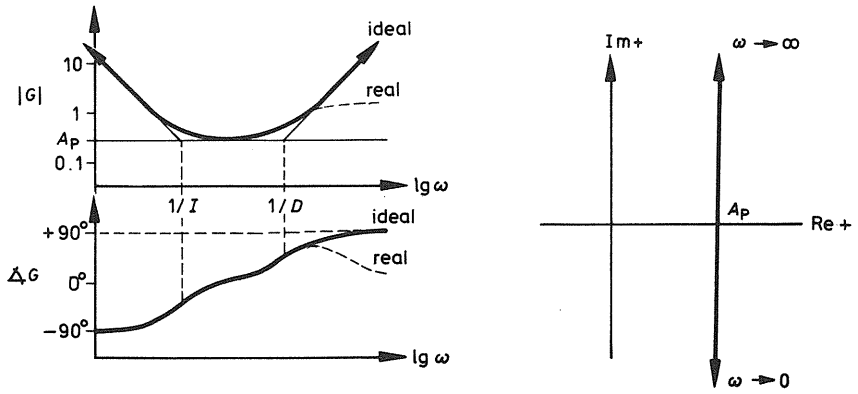

Fig. 5.35. Bode and Nyquist plots of a *PID*-controller

of the *P*, *I* and *D* elements (polar plot). Both the Nyquist and the Bode diagrams are shown in *Fig. 5.35*. It can be seen that

(*i*) for low frequencies, $\omega < \frac{1}{I}$, the *PID*-controller behaves like an *I*-controller, and may cause instability in a control loop by increasing both phase lag and dynamic gain;

(*ii*) for high frequencies $\omega > \frac{1}{D}$ the *PID*-controller behaves like a *D*-element and may cause instability in a control loop by increasing its dynamic gain (although it decreases phase lag);

(*iii*) thus parameters *I* and *D* should be tuned so that $\frac{1}{I} < \omega_{cr} < \frac{1}{D}$, where ω_{cr} is the critical frequency of the process along (or with a *P*-controller).

Conclusion. Negative feedback control loops are stable if their frequency function satisfies the Nyquist criterion:

$$\sphericalangle G_S(i\omega_{cr}) + \sphericalangle G_C(i\omega_{cr}) = -180°$$

$$|G_S(i\omega_{cr})|\,|G_C(i\omega_{cr})| < 1$$

Since the parameters are given and, with them, the frequency function of the process, $G_S(i\omega)$ is known, it is the controller that must be tuned in consideration of the characteristics of the process.

The proper control of a dead time process requires a *PI*-controller; the minimum settling time is $6T_D$.

A single-capacity process can be controlled well with a simple *P*-controller set to a narrow proportional band. A pure *I*-controller would cause instability and thus must not be applied.

5.2.6 Tuning the controller. Cycling method

The critical period of an existing control loop and the critical gain of the *P*-controller can be determined experimentally — this is the method used in practice to tune control loops. Once the critical period $T_{P,cr}$, and the critical gain $A_{C,cr}$ of the *P*-controller are known (both determined with the *P*-control mode only, any *I* and *D* control modes being eliminated), optimum controller settings for any control mode (i. e. for *P* only, for *PI*, for *PD*, or for *PID*) can be selected from tables.

The experiment is performed as follows:

1. Compile and close the control loop. Set the setpoint (command signal) of the controller.
2. Eliminate *I* and *D* control modes, possibly by setting *I* to its maximum and *D* to its minimum possible value.
3. Set A_C to its minimum value. (Most analogue controllers are scaled as $P=\frac{1}{A_C}$.)
4. Apply a small driving signal to one of the inputs of the control loop (this may be one of the load inputs as well; but it is most conveniently applied to the set point). Be sure to apply a driving signal containing a broad frequency spectrum (it must also contain the unknown critical frequency, e. g. a pulse or a stop).
5. Observe the output of the control loop (in most cases the controlled variable is the output, but the manipulated variable may also be used). If no oscillation occurs, increase the controller gain A_C and repeat step 4 until uniform oscillation results. If the amplitude of the oscillation increases, the controller gain A_C must be decreased.
6. The controller gain A_C causing constant amplitude oscillation of the control loop output signal is the critical one: $A_{C,cr}$, and the period of one oscillation is the critical period $I_{P,cr}$.
7. Use *Table 5.1* to calculate the optimum controller settings. Here K_{cr} is the critical and K the recommended loop gain. By definition

$$K=A_C A_S$$

and, since the process gain A_S is constant,

$$\frac{K}{K_{cr}}=\frac{A_C}{A_{C,cr}}$$

Table 5.2 gives the recommended controller settings given by Ziegler and Nichols, which are applicable to processes consisting of dead time and single capacity. It gives stable operation and quarter amplitude damping for most control loops. Controller tuning will be treated more thoroughly in Section 5.3.8.

The method just described can also be used for simulated control loops which may be tested using the *TACS* program.

Table 5.1. Optimum controller settings

Process parameters by cycling method: $A_{C,cr}$, $T_{P,cr}$
Zeigler—Nichols settings

Controller	A_C	I	D
P	$0.5\ A_{C,cr}$		
PI	$0.45\ A_{C,cr}$	$T_{P,cr}/1.2$	
PID	$0.6\ A_{C,cr}$	$T_{P,cr}/2$	$T_{P,cr}/8$

Table 5.2. Process parameters from process reaction curve: A_S, T_D, T

Controller	Ziegler—Nichols	Cohen—Coon
P	$A_S A_C = K = \frac{T}{T_D}$	$K = \frac{T}{T_D} + \frac{1}{3}$
PI	$k = 0.9\frac{T}{T_D}$ $I = \frac{T_D}{0.3}$	$K = 0.9\left(\frac{T}{T_D} + \frac{1}{11}\right)$ $\frac{I}{T} = \frac{3.33(T_D/T)\,(1+11(T_D/T))}{1+2.2(T_D/T)}$
PID	$K = 1.2(T/T_D)$ $I = 2T_D$ $D = 0.5T_D$	$K = 1.35(T/T_D)+0.27$ $\frac{I}{T} = \frac{2.5(T_D/T)\,(1+5(T_D/T))}{1+0.6(T_D/T)}$ $\frac{D}{T} = \frac{0.37(T_D/T)}{1+0.2(T_D/T)}$

5.3 Linear control theory

Before tackling more complex control problems it is necessary to survey the mathematical methods of linear control theory. The fundamentals of classical linear control theory are very important in practice, notwithstanding the fact that most chemical engineering systems are nonlinear. The mathematical models of most nonlinear systems may, however, be linearized around their operating points, in order to use the powerful linear techniques.

5.3.1 Classification of processes

There are several ways of classifying processes and their dynamics.

1. Mathematical model
   (a) Linear: if all equations and all function in the equations are linear.
   (b) Nonlinear: if the statement above is not true. In this case the principle of superposition does not apply. That is, the response of the system to several independent inputs cannot be determined independently and then added.
2. Independent variables
   (a) Lumped-parameter system: described by an ordinary differential equation; the dependent variable is a function of time only; for example, a perfectly mixed tank reactor.
   (b) Distributed-parameter system: described by a partial differential equation; the dependent variable is a function of time and of spatial variables; an example is the tubular reactor.
3. Order:

An nth order linear, lumped-parameter, single input single output process can be described by an nth order linear, ordinary differential equation of the form:

$$a_n \frac{d^n y}{dt^n} + a_{n-1} \frac{d^{n-1} y}{dt^{n-1}} + \ldots + a_1 \frac{dy}{dt} + a_0 a =$$

$$= b_m \frac{d^m x}{dt^m} + b_{m-1} \frac{d^{m-1} x}{dt^{m-1}} + \ldots + b_1 \frac{dx}{dt} + b_0 x \qquad (5.57)$$

where x is input,
y is output,
t is time, and
a, b are constants.

In real systems n is greater than m. Inequality $m > n$ would mean that the response for a step input must be infinite in zero time. This cannot be realized in a finite system (cf. the D-element, Section 5.2.5).

Multiple input—multiple output systems are characterized by a system of differential equations.

In control technology it is convenient to write Eq.(5.57) in the following form:

$$T_n^n \frac{d^n y}{dt^n} + T_{n-1}^{n-1} \frac{d^{n-1} y}{dt^{n-1}} + \ldots + T_1 \frac{dy}{dt} + y =$$

$$= A_S \left[\tau_m^m \frac{d^m x}{dt^m} + \tau_{m-1}^{m-1} \frac{d^{m-1} x}{dt^{m-1}} + \ldots + \tau_1 \frac{dx}{dt} + x \right] \qquad (5.58)$$

where $T_i = \sqrt[i]{\frac{a_i}{a_0}}$; $\tau_j = \sqrt[j]{\frac{b_j}{b_0}}$ are time constants, and

$$A_S = \frac{b_0}{a_0} = \frac{y(t=\infty)}{x(t=\infty)} \text{ is steady-state gain.}$$

Very important special cases are:

the first-order element:

$$T\frac{dy}{dt} + y = A_S x \tag{5.59}$$

the second-order element:

$$T_2^2 \frac{d^2y}{dt^2} + T_1 \frac{dy}{dt} + y = A_S x \tag{5.60}$$

A dead-time element (defined in Section 5.2.2) can be approximated only in this representation by nth order differential equation where $n \to \infty$.

5.3.2 Typical input signals

Linear processes can be characterized by their responses to typical input signals (besides being characterized by their differential equations).

Typical input signals are also applied for the experimental determination of the dynamic behaviour of an unknown process (identification).

The typical input signals are the following:

(*i*) step: a change in the input signal value during zero time;

(*ii*) pulse: a step upwards and, after time Δt, the same step downwards;

(*iii*) impulse: a pulse in time $\Delta t \to 0$;

(*iv*) ramp: a change in the input signal with constant speed; and

(*v*) a sinusoidal signal.

These deterministic signals are shown and defined in *Fig. 5.36*.

The sinusoidal input signal is used to obtain the frequency response that has already been treated in Section 5.2.1 (see *Fig. 5.11*). The sinusoidal input is to be repeated several times with the same frequency and amplitude, until a steady-state response is obtained; the steady state is indicated by at least three constant amplitude and phase sinusoidal outputs while step, impulse and ramp inputs may be applied only once or twice. Moreover, the experiment has to be carried out with different frequencies over a broad frequency range to obtain a meaningful frequency function, thus it is lengthy and expensive. However, in the presence of noise, it is advantageous that a single frequency output is to be measured.

Identification of an unknown process can also be performed using stochastic signals, such as white noise or plant noise.

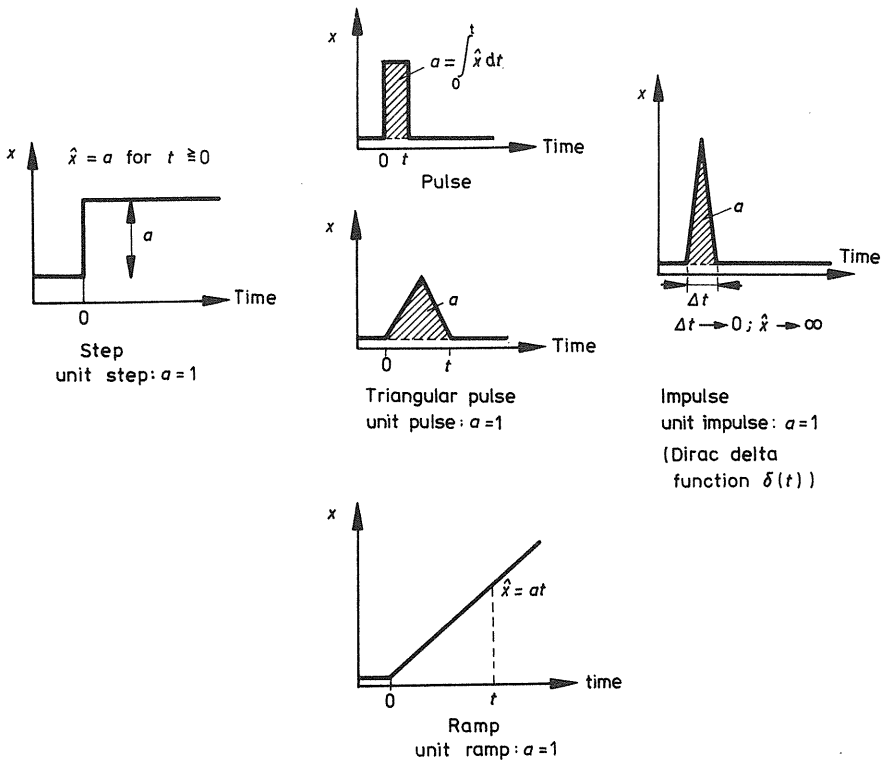

Fig. 5.36. Deterministic input signals

5.3.3 Linearization

Mathematical models of nonlinear processes are often linearized in the vicinity of an operating point, so that they may be tackled by linear mathematical techniques, e.g. the frequency function method.

Linearization of a nonlinear function is accomplished by expanding the nonlinear function in Taylor series and then truncating the series, leaving only the first derivative(s). The nonlinear function has to be differentiable at the operating point. Thus the nonlinear function is simply substituted by its tangent.

Consider a nonlinear function, $f(x, z)$, of two variables. $\bar{x}$ and $\bar{z}$ are the values of the variables at the operating point, at the steady state. The Taylor series of this function about point $(\bar{x}, \bar{z})$ is:

$$f(x,z)=f(\bar{x},\bar{z})+\left(\frac{\partial f}{\partial x}\right)_{\bar{x},\bar{z}}(x-\bar{x})+\left(\frac{\partial f}{\partial z}\right)_{\bar{x},\bar{z}}(z-\bar{z})+$$

$$+\frac{1}{2!}\left[\left(\frac{\partial^2 f}{\partial x^2}\right)_{\bar{x},\bar{z}}(x-\bar{x})^2+\left(\frac{\partial^2 f}{\partial x\partial z}\right)_{\bar{x},\bar{z}}(x-\bar{x})(z-\bar{z})+\left(\frac{\partial^2 f}{\partial z^2}\right)_{\bar{x},\bar{z}}(z-\bar{z})^2\right]+\ldots+$$

$$+\ldots+\frac{1}{n!}\left(\frac{\partial}{\partial x}(x-\bar{x})+\frac{\partial}{\partial z}(z-\bar{z})\right)^n f(\bar{x},\bar{z})+\ldots+R_n \qquad (5.61)$$

whence the linearized $f(x, z)$ function is:

$$f(x,z)\approx f(\bar{x},\bar{z})+\left(\frac{\partial f}{\partial x}\right)_{\bar{x},\bar{z}}\hat{x}+\left(\frac{\partial f}{\partial z}\right)_{\bar{x},\bar{z}}\hat{z} \qquad (5.62)$$

where $\hat{x}=x-\bar{x}$ and $\hat{z}=z-\bar{z}$

Example 5.3.1
Consider a second-order reaction:

$$2A\rightarrow P$$

for which the net reaction rate is:

$$r=\frac{\mathrm{d}c_\mathrm{P}}{\mathrm{d}t}=kc_\mathrm{A}^2$$

The function $r=r(c_\mathrm{A})$ is desired in linearized form. Apply Eq.(5.62) to a monovariate function:

$$r\approx\bar{r}+\left(\frac{\partial r}{\partial c_\mathrm{A}}\right)_{\bar{c}_\mathrm{A}}\hat{c}_\mathrm{A}=k\bar{c}_\mathrm{A}^2+2k\bar{c}_\mathrm{A}\hat{c}_\mathrm{A}$$

Example 5.3.2
The input of component A to a continuous reactor is $A=Wc_\mathrm{A}$, where both the volume rate W and its concentration c_A are variables. Obtain the function $A=A(W, c_\mathrm{A})$ in linearized form. Apply Eq.(5.62):

$$A\approx\bar{A}+\left(\frac{\partial A}{\partial W}\right)_{\bar{W},\bar{c}_\mathrm{A}}\hat{W}+\left(\frac{\partial A}{\partial c_\mathrm{A}}\right)_{\bar{W},\bar{c}_\mathrm{A}}\hat{c}_\mathrm{A}=\bar{W}\bar{c}_\mathrm{A}+\bar{c}_\mathrm{A}\hat{W}+\bar{W}\hat{c}_\mathrm{A}$$

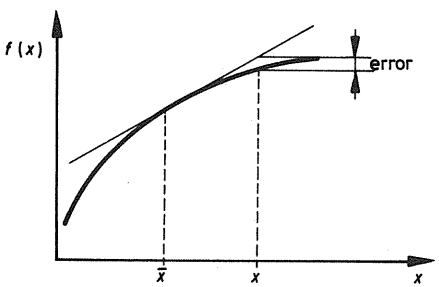

Fig. 5.37. Error of linearization

Linearization according to Eq. (5.62) gives rise to an error, which depends on the magnitude of the neglected elements of Eq. (5.61) which, in turn, depends on the magnitude of $\hat{x}=x-\bar{x}$ (and $\hat{z}$, resp.). Thus the error is small and the linearization by Eq. (5.62) is satisfactory if x is very close to $\bar{x}$ (and z to $\bar{z}$, resp.). The error introduced by linearization is shown in *Fig. 5.37* for a monovariate function.

It is evident that the linearized form of a nonlinear function varies with the operating point.

5.3.4 The transfer function method

The transfer function method is very convenient for obtaining the output of a process in response to any input, and for handling complicated networks. Transfer functions are closely related to the frequency functions introduced in Section 5.2.1, and are applied for the same purpose. The transfer function of a process is defined by the following expression:

$$G(s)=\frac{Y(s)}{X(s)} \tag{5.63}$$

where $G(s)$ is the transfer function,
$Y(s)$ is the Laplace transform of the output signal,
$X(s)$ is the Laplace transform of the input signal, and
s is a complex variable, the so-called Laplace transform variable.

Thus to use the transfer function method, Laplace transformation has to be performed.

Laplace transformation

Transform operations are used to transform difficult problems into more tractable forms. Laplace transformation is a method of transforming linear differential equations into algebraic equations.

Example 5.3.3

Transform the differential equation of the *PID*-controller:

$$m(t)=m(0)+A_P e(t)+A_I \int_0^t e(t)\,\mathrm{d}t+A_D \frac{\mathrm{d}e(t)}{\mathrm{d}t}$$

The Laplace transform is:

$$M(s)=A_P E(s)+A_I \frac{E(s)}{s}+A_D s E(s)$$

if

$$\int_{t=0} e\,dt = 0 \qquad \text{and} \qquad \left(\frac{de}{dt}\right)_0 = 0$$

The Laplace transform of a time domain function $f(t)$ is denoted by the symbol $F(s)$ and is defined as follows:

$$F(s) = \mathscr{L}[f(t)] = \int_0^\infty f(t)\, e^{-st} dt \tag{5.64}$$

where $\mathscr{L}[f(t)]$ is the symbol that indicates Laplace transformation of the function $f(t)$ in brackets,
$s = a + i\omega$ is a complex variable, the Laplace variable.

Laplace transformation is applicable to linear functions for which the integral in Eq. (5.64) is convergent. This is not a serious restriction.

Laplace transformation is a linear operation, as it can be seen from Example 5.3.3 which has been solved using the basic theorems listed in *Table 5.3*. These theorems are derived from the definition in Eq. (5.64), as well as the Laplace transforms of some functions most frequently used in control technique, listed in *Table 5.4*.

To perform Laplace transformation of a time function or the inverse transformation of a Laplace transform into a time function it is not necessary to resort to the definition given in Eq. (5.64); it is generally sufficient simply to look up the function in *Table 5.3* and/or *5.4*. The technique is illustrated by some examples.

Example 5.3.4. *First-order lag. Transfer function*

A first-order element is characterized by the first-order differential equation, Eq. (5.59):

$$T\frac{dy(t)}{dt} + y(t) = A_S x(t) \tag{5.65}$$

where $y(0) = 0$, and $\left(\frac{dy}{dt}\right)_0 = 0$ (steady state) and $x(0) = 0$.

Having defined the functions and initial conditions, we can proceed to the transformation. Applying the linearity theorem, each term in Eq. (5.65) can be transformed individually:

$$\mathscr{L}\left[T\frac{dy(t)}{dt}\right] + \mathscr{L}[y(t)] = \mathscr{L}[A_S x(t)] \tag{5.66}$$

Table 5.3. Laplace transform theorems

1. Linearity theorem

$$\mathscr{L}[Kf(t)]=K\,\mathscr{L}[f(t)]=KF(s) \qquad (K=\text{ constant})$$

$$\mathscr{L}[f_1(t)\pm f_2(t)]=F_1(s)\pm F_2(s)$$

2. Real differentiation theorem
First derivative

$$\mathscr{L}\left[\frac{\mathrm{d}}{\mathrm{d}t}f(t)\right]=sF(s)-f(0)$$

General nth derivative

$$\mathscr{L}\left[\frac{\mathrm{d}^n}{\mathrm{d}t^n}f(t)\right]=$$

$$=s^nF(s)-s^{n-1}f(0)-s^{n-2}\frac{\mathrm{d}}{\mathrm{d}t}f(0)-\ldots-s^2\frac{\mathrm{d}^{n-2}}{\mathrm{d}t^{n-2}}f(0)-s\frac{\mathrm{d}^{n-1}}{\mathrm{d}t^{n-1}}f(0)$$

3. Real integration theorem

$$\mathscr{L}\left[\int f(t)\,\mathrm{d}t\right]=\frac{F(s)}{s}$$

$$\mathscr{L}\left[\int^1\int^2\ldots\int^n f(t)\,\mathrm{d}t^n\right]=\frac{1}{s^n}F(s)$$

4. Initial value theorem

$$f(0)=\lim_{s\to\infty}sF(s)$$

5. Final value theorem

$$f(\infty)=\lim_{s\to 0}sF(s)$$

6. Time-shift theorem

$$\mathscr{L}[f(t-a)]=e^{-at}\,\mathscr{L}[f(t)]=e^{-at}\cdot F(s)$$

The transform of the first term is, using Theorems 1 and 2:

$$\mathscr{L}\left[T\frac{\mathrm{d}y(t)}{\mathrm{d}t}\right]=T\mathscr{L}\left[\frac{\mathrm{d}y(t)}{\mathrm{d}t}\right]=T\big(sY(s)-y(0)\big)$$

$$=TsY(s)$$

The transform of the second term is by definition

$$\mathscr{L}[y(t)]=Y(s)$$

The transform of the third term is by definition

$$\mathscr{L}[x(t)]=X(s), \qquad \text{whence} \qquad \mathscr{L}[A_S x(t)]=A_S X(s)$$

The entire Laplace domain equation is

$$TsY(s)+Y(s)=A_S X(s) \tag{5.67}$$

which can be rearranged to obtain the transfer function defined by Eq. (5.63):

$$\frac{Y(s)}{X(s)}=G(s)=\frac{A_S}{Ts+1} \tag{5.68}$$

The transfer function of a process can be used to obtain its response to any input:

$$Y(s)=G(s)\,X(s) \tag{5.69}$$

Table 5.4. Laplace transform pairs

$F(s)$	$f(t)$
1	$\delta(t)$ (impulse)
$\frac{1}{s}$	$u(t)$ (step)
$\frac{1}{s^n}$ $(n=1, 2, \ldots)$	$\frac{t^{n-1}}{(n-1)!}$
$\frac{1}{s\pm a}$	$e^{\mp at}$
$\frac{1}{s(s\pm a)}$	$\frac{1}{\pm a}(1-e^{\mp at})$
$\frac{s}{s^2+a^2}$	$\cos at$
$\frac{a}{s^2+a^2}$	$\sin at$
$\frac{s}{s^2-a^2}$	$\mathrm{ch}\, at$
$\frac{a}{s^2-a^2}$	$\mathrm{sh}\, at$
$\frac{1}{(s+a)^2+b^2}$	$\frac{1}{b}e^{-at}\sin bt$
$\frac{s+a}{(s+a)^2+b^2}$	$e^{-at}\cos bt$
$\frac{1}{(s+a)^n}$	$\frac{1}{(n-1)!}t^{n-1}e^{-at}$
$\frac{ab}{(s+a)(s+b)}$	$\frac{1}{b-a}(e^{-at}-e^{-bt})$

To obtain the impulse response, the Laplace transform of the unit impulse is to be substituted for $X(s)$ in Eq. (5.69). From *Table 5.4* it is found that

$$\text{for} \quad f(t)=1\cdot\delta(t) \qquad F(s)=1$$

Thus the Laplace transform of the impulse response of any process is

$$Y(s)=G(s) \tag{5.70}$$

in other words, the transfer function of a process is obtained by the Laplace transformation of its impulse response.

Example 5.3.5 *First-oder lag. Impulse response*

The Laplace transform of the impulse response of the first-oder lag is obtained by substituting Eq. (5.68) into Eq. (5.69):

$$Y(s)=\frac{A_S}{Ts+1}=\frac{\frac{A_S}{T}}{s+\frac{1}{T}}=\frac{A_S}{T}\left(\frac{1}{s+\frac{1}{T}}\right) \tag{5.71}$$

The time domain function of Eq.(5.68) is found in *Table 5.4* as

$$y(t)=\frac{A_S}{T}e^{-\frac{t}{T}} \tag{5.72}$$

The impulse response of a first-oder lag is shown in *Fig. 5.38*.

From Eq.(5.72):

$$y(0)=\frac{A_S}{T}\left(\frac{dy}{dt}\right)_0=-\frac{A_S}{T^2}$$

$$y(T)=\frac{A_S}{T}\cdot 0.368$$

$$y(\infty)\to 0$$

Example 5.3.6. *First-order lag. Step response*

The Laplace transform of the step response of the first-order element can be found in two ways:

1. by multiplying its transfer function by the Laplace transform of the unit step input;
2. by substituting $x(t)=1\cdot u(t)$, the expression for unit step input, into its differential equation given in Eq. (5.65).

The first method has been illustrated: to obtain the impulse response. Now use the second way.

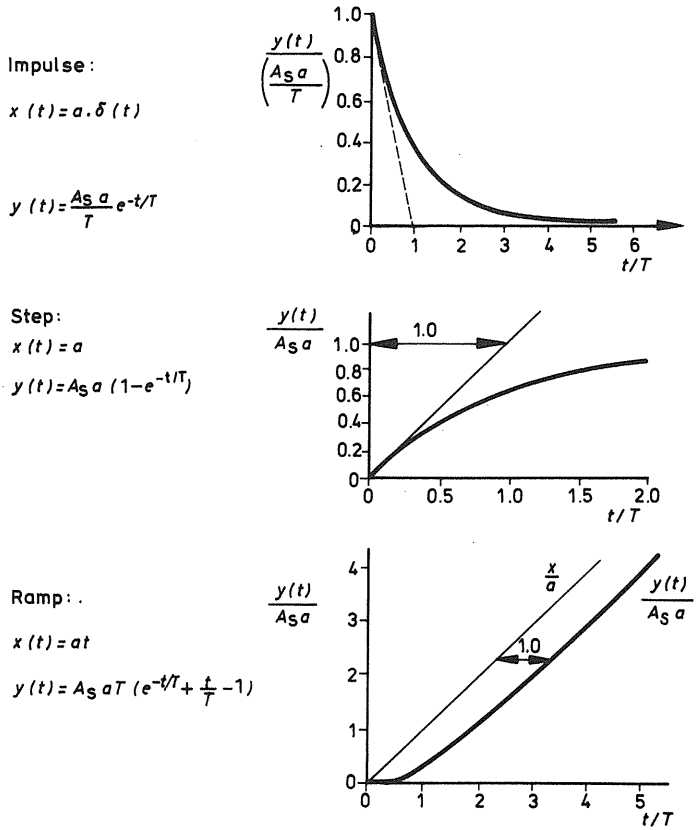

Fig. 5.38. Response functions of a first-order lag

Equation (5.65) is to be transformed, with $x(t)=1\cdot u(t)$, unit step input occurring at $t=0$.

The Laplace transforms of terms 1 and 2 were found in Example 5.3.4. Using *Table 5.4*:

$$\mathcal{L}[A_S u(t)] = A_S \frac{1}{s}.$$

The Laplace transform of Eq. (5.65) for $x(t)=u(t)$ is obtained as:

$$TsY(s) + Y(s) = A_S \frac{1}{s} \tag{5.73}$$

from which $Y(s)$ is expressed as:

$$Y(s) = \frac{A_S}{T} \frac{1}{s\left(s+\frac{1}{T}\right)} \tag{5.74}$$

This can be found in *Table 5.4*:

$$y(t)=A_S\left(1-e^{-\frac{t}{T}}\right) \tag{5.75}$$

The values of the step response function are:

$$y(0)=A_S(1-1)=0 \qquad \left(\frac{dy}{dt}\right)_0=\frac{A_S}{T}=\frac{y(\infty)}{T}$$

$$y(T)=A_S(1-e^{-1})=A_S\cdot 0.632$$

$$y(\infty)\to A_S$$

The function described by Eq. (5.75) is shown in *Fig. 5.38*.

Inversion of Laplace transforms by expansion in partial fractions

The inverse transforms of functions not listed in *Table 5.4* can be obtained by several methods.

Inversion of a Laplace transform is denoted by

$$\mathcal{L}^{-1}[F(s)]=f(t) \tag{5.76}$$

The inverse transform can be obtained by contour integration in the complex plane:

$$f(t)=\frac{1}{2\pi i}\int_{a-i\omega}^{a+i\omega} e^{st}F(s)\,ds \tag{5.77}$$

Other methods for inversion are the convolution method and expansion in partial fractions. The partial fraction method is presented here. It consists in converting the function to be inverted into the form of a sum of simpler functions, the inverse transforms of which can be found in the transform tables:

$$F(s)=F_1(s)+F_2(s)+\ldots+F_n(s) \tag{5.78}$$

The Laplace transform, for which we require the time inverse generally has the form

$$F(s)=\frac{P(s)}{Q(s)} \tag{5.79}$$

which is obtained by the Laplace transformation of the general ordinary differential equation, Eq.(5.57). Here $P(s)$ and $Q(s)$ are polynomials in s such that the order of the numerator is less than that of the denominator (if not, then it has to be converted into such a form by division).

Next the roots of polynomial $Q(s)$ are to be obtained. $Q(s)$ is factored into its roots:

$$Q(s)=(s-s_1)(s-s_2)\ldots(s-s_n)=0 \tag{5.80}$$

On the basis of Eq.(5.80) the function $F(s)$ can be expanded in terms of the roots of $Q(s)$:

$$F(s)=\frac{P(s)}{Q(s)}=\frac{C_1}{s-s_1}+\frac{C_2}{s-s_2}+\ldots+\frac{C_n}{s-s_n} \tag{5.81}$$

The inverse transform of Eq.(5.81) is simply (cf. *Table 5.4*)

$$f(t)=C_1e^{s_1t}+C_2e^{s_2t}+\ldots+C_ne^{s_nt} \tag{5.82}$$

The constants C_1, C_2 ... C_n can be determined by making use of the fact that Eq.(5.81) is an identity that is valid for every value of s or by applying the following relation for distinct roots:

$$C_i=\lim_{s\to s_i}[(s-s_i)F(s)], \qquad i=1\ldots n \tag{5.83}$$

Equation (5.83) can be checked by applying it to the right-hand-side of the identity [Eq.(5.81)]. Approaching the limit $s\to s_i$, all terms containing C_j (where $j\neq i$) become zero, except the term containing C_i, which is $C_i\cdot 1$.

If some roots of $Q(s)$ are repeated, the procedure is slightly different. Consider, for instance, that one of the roots of $Q(s)$ is repeated twice, thus $Q(s)$ is factored into the following equation [$Q(s)$ is of order $(n+1)$, it has n roots]:

$$Q(s)=(s-s_1)^2(s-s_2)\ldots(s-s_n) \tag{5.84}$$

In this case $F(s)$ is to be expanded as follows:

$$F(s)=\frac{P(s)}{Q(s)}=\frac{C_1}{(s-s_1)^2}+\frac{C_2}{s-s_1}+\frac{C_3}{s-s_2}+\ldots+\frac{C_{n+1}}{s-s_n} \tag{5.85}$$

Constants C_1 and C_2 are determined as:

$$C_1=\lim_{s\to s_1}[(s-s_1)^2F(s)] \tag{5.86}$$

$$C_2=\lim_{s\to s_1}\left[\frac{d}{ds}(s-s_1)^2F(s)\right] \tag{5.87}$$

Example 5.3.7. *First-order lag. Ramp response*

The unit ramp function is $x(t)=1\cdot t$ *(Fig. 5.36)*. Its Laplace transform is $X(s)=\frac{1}{s^2}$.

The Laplace transform of the ramp response of a first-order lag is obtained from its transfer function

$$Y(s)=G(s)X(s)=\left(\frac{A_S}{Ts+1}\right)\left(\frac{1}{s^2}\right) \tag{5.88}$$

Equation (5.88) is to be rearranged so that the coefficient of the highest power of s in the denominator is 1:

$$Y(s)=\frac{A_S}{T}\left(\frac{1}{s+\frac{1}{T}}\right)\left(\frac{1}{s^2}\right) \tag{5.89}$$

It is not necessary to write $Q(s)$ in the polynomial form, since it is already in the required factored form. It is evident that $Q(s)$ has three roots:

$$s_1=-\frac{1}{T}, \qquad s_2=0, \qquad s_3=0$$

Applying Eq.(5.85), $Y(s)$ is expanded in partial fractions:

$$Y(s)=\frac{A_S}{T}\left(\frac{C_1}{s+\frac{1}{T}}+\frac{C_2}{s^2}+\frac{C_3}{s}\right) \tag{5.90}$$

The transform of Eq.(5.90) is obtained by using *Table 5.3*:

$$y(t)=\frac{A_S}{T}\left(C_1 e^{-\frac{t}{T}}+C_2 t+C_3\right) \tag{5.91}$$

C_1 is determined by applying Eq.(5.89):

$$C_1=\lim_{s\to-\frac{1}{T}}\frac{s+\frac{1}{T}}{\left(s+\frac{1}{T}\right)s^2}=T^2 \tag{5.92}$$

C_2 is determined by applying Eq.(5.86) to Eq.(5.89):

$$C_2=\lim_{s\to 0}\frac{s^2}{\left(s+\frac{1}{T}\right)s^2}=T \tag{5.93}$$

C_3 is determined by applying Eq.(5.87):

$$C_3=\lim_{s\to 0}\left[\frac{\mathrm{d}}{\mathrm{d}s}\frac{s^2}{\left(s+\frac{1}{T}\right)s^2}\right]=\lim_{s\to 0}\left[\frac{\mathrm{d}}{\mathrm{d}s}\frac{1}{\left(s+\frac{1}{T}\right)}\right]=$$

$$=\lim_{s\to 0}-\frac{1}{\left(s+\frac{1}{T}\right)^2}=-T^2 \tag{5.94}$$

C_1, C_2 and C_3 are substituted into Eq.(5.91) to obtain the required time re-

sponse:

$$y(t)=\frac{A_S}{T}\left(T^2e^{-\frac{t}{T}}+T\cdot t-T^2\right)=A_S\left[Te^{-\frac{t}{T}}+(t-T)\right] \tag{5.96}$$

The response of a P-element with a gain A_S to a unit ramp input $1\cdot t$ is $y(t)=$ $=A_S t$. This can be compared to the response of the first order element, which is

$$y(t)\cong A_S(t-T) \qquad \text{for} \quad t>3T \tag{5.97}$$

since $e^{-3}=0.05$, $e^{-4}=0.018$, the first term in Eq.(5.96) dies out with time.

The ramp response is shown in *Fig. 5.38*. It can be seen from both the figure and from Eq.(5.97) that the ramp response of a first-order process lags by T, its time constant, behind that of a proportional element.

Block diagrams

The transfer function representation has the convenience that the output of a process is obtained by simply multiplying its transfer function by the input: $Y(s)=G(s)X(s)$, cf. Eq.(5.63). Owing to this feature it is very suitable for the treatment of networks. Moreover, since the Laplace transformation is linear, addition and subtraction of time domain signals also holds for their Laplace transforms.

A similar quality of the frequency functions has already been utilized in Section 5.2.1.

In the frequency function representation the elements were represented by their frequency functions. In the transfer function representation the elements are represented by their transfer functions, which are the blocks, and the signals are represented by their Laplace transforms, which are the arrows.

The overall frequency functions were derived for elements in series (Eq.5.2), in parallel (Eq.5.3) and for a simple negative control loop (Eq.5.4). Analogous equations are obtained for the Laplace transforms of the signals and the transfer functions of the elements.

Elements in series:

$$G(s)=G_1(s)\cdot G_2(s)\cdot G_3(s)\cdot \ \ldots \ \cdot G_n(s) \tag{5.98}$$

in parallel:

$$G(s)=G_1(s)+G_2(s)+\ \ldots \ +G_n(s) \tag{5.99}$$

The transfer functions for a simple negative feedback control loop, with a block diagram shown in *Fig. 5.39* can be derived as

$$W_D(s)=\frac{G(s)}{D(s)}=\frac{G_D(s)}{1+G_C(s)G_s(s)} \tag{5.100}$$

$$W(s)=\frac{G(s)}{R(s)}=\frac{G_C(s)G_s(s)}{1+G_C(s)G_s(s)} \tag{5.101}$$

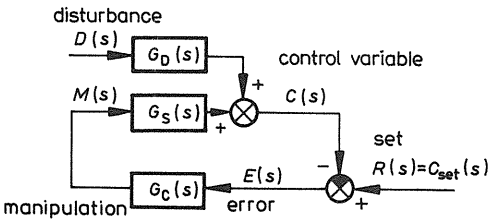

Fig. 5.39. Negative feedback loop. To the derivation of its transfer function

The block diagram of a positive feedback loop is shown in *Fig. 5.40*. The transfer function for this loop is derived as:

$$W(s)=\frac{Y(s)}{X(s)}=\frac{G(s)}{1-G(s)H(s)} \tag{5.102}$$

To facilitate calculations with the block diagram linear transformations are possible. Some of these are shown in *Fig. 5.41*.

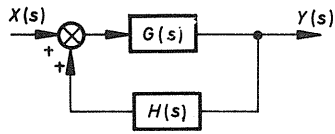

Fig. 5.40. Positive feedback loop

Stability analysis

Recall Eq. (5.82) for the time inverse of a Laplace transform:

$$f(t)=C_1e^{s_1t}+C_2e^{s_2t}+\ldots+C_ne^{s_nt} \tag{5.103}$$

Here, s_1, s_2,... s_n are complex numbers obtained as the roots of $Q(s)$, the characteristic polynomial [cf. Eq. (5.80)], and the roots are the characteristics of the time function $f(t)$, which may be unbounded, bounded or oscillating.

Case 1. All characteristic roots are real negative numbers,
e. g. $s_1=-k_1$, $s_2=-k_2$, etc.
Then $\lim_{t\to\infty} f(t)=0$, $f(t)$ is bounded, because $\lim_{t\to\infty} e^{-kt}=0$

Case 2. One of the characteristic roots is a positive real number, $s_i=+k$.
Then $\lim_{t\to\infty} f(t)=\infty$, unbounded, because $\lim_{t\to\infty} e^{+kt}=\infty$

Case 3. There is a complex pair among the characteristic roots; they always appear as conjugate pairs.

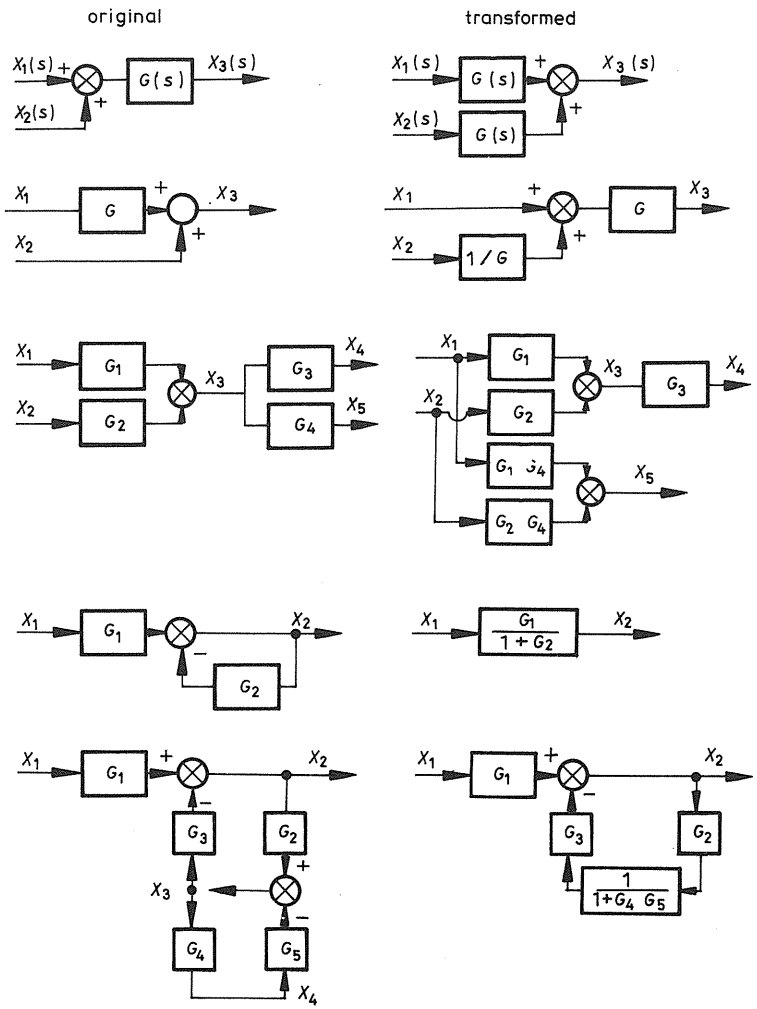

Fig. 5.41. Linear transformations of the block diagram

Case 3a. $s_k = -a+ib$, $s_m = -a-ib$, i. e. the real parts of the conjugate complex roots are negative.

Then $\lim_{t\to\infty} f(t)=0$, bounded, with damped oscillation, because

$\lim_{t\to\infty} (C_k e^{-a+ib} + C_m e^{-a-ib}) = 0$

and the factors e^{ib} and e^{-ib} indicate oscillations with frequency b.

Case 3b. $s_k = +a+ib$, $s_m = +a-ib$, i. e. the real parts of the conjugate complex roots are positive.

Then $\lim_{t\to\infty} f(t)=\infty$, unbounded (the oscillations are expanding), because $\lim_{t\to\infty}(C_k e^{+a+ib}+C_m e^{+a-ib})=\lim_{t\to\infty}(C_k e^{+a}e^{-ib}+C_m e^{+a}C_m e^{-ib})=\infty$, and the factors e^{ib} and e^{-ib} indicate oscillations with frequency b.

Note: For a pair of conjugate complex roots, $s_k=-a+ib$, $s_m=-a-ib$, the characteristic polynomial $Q(s)$ is conveniently factored as $(s-s_k)(s-s_m)=(s+a-ib)(s+a+ib)=s^2+2as+a^2+b^2=(s+a)^2+b^2$.

In the expansion in partial fractions this will give the two terms:

$$\frac{C_k b+C_m(s+a)}{(b+a)^2+b^2}$$

for which *Table 5.4* gives

$$\mathcal{L}^{-1}\left[\frac{C_k b}{(s+a)^2+b^2}\right]+\mathcal{L}^{-1}\left[\frac{C_m(s+a)}{(s+a)^2+b^2}\right]=$$

$$=C_k e^{-at}\sin bt+C_m e^{-at}\cos bt$$

Case 4. There is an imaginary pair among the characteristic roots: $s_k=ib$, $s_m=-ib$.
Then $f(t)$ is oscillating, undamped with a constant amplitude and frequency b.

Stability has been defined in Section 5.1. A process is stable when, after a pulse-like disturbance, it returns to its original state within some time. This is equivalent to the statement that the limit of the impulse response of a stable process is zero, for $t\to\infty$.

For a unit impulse, $f(t)=1\cdot\delta(t)$, $F(s)=1$.

The Laplace transform of the impulse response is:

$$Y(s)=G(s)\,X(s)=G(s)$$

Thus for a stable process:

$$\lim_{t\to\infty}(\mathcal{L}^{-1}[G(s)])=0 \tag{5.104}$$

The stability definition expressed by Eq. (5.104) suggests the stability criterion that, for a stable linear process, the real parts of all characteristic roots of the transfer function must be smaller than 0. If the characteristic roots of the transfer function are plotted in the complex plane, all roots must lie in the left half-plane *(Fig. 5.42)*.

Unfortunately, the stability criterion based on the examination of the characteristic roots is of only limited use.

The roots of the denominator of the transfer function must be known, which may cause problems for a higher-order process. For a control loop, even if the

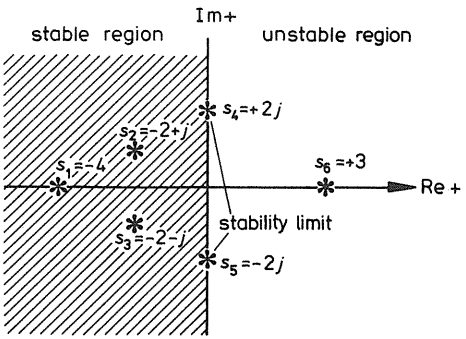

Fig. 5.42. Stable region of characteristic roots

roots of the open loop transfer function $G_L(s) = G_C(s)\, G_S(s)$ are known from the model of the process [refer to Eqs (5.100) and (5.101)], and even if the open loop is stable, the characteristic roots of the closed loop transfer function, i.e. the roots of the characteristic equation

$$1 + G_C(s)\, G_S(s) = 0 \tag{5.105}$$

are unknown, and the closed loop may be unstable. Equation (5.105) cannot be solved to find its roots if dead time is involved, because a dead time process can be approximated by a differential equation of infinite order.

Routh–Hurwitz stability criterion

The Routh–Hurwitz stability criterion is used to check the stability of a system for which the characteristic equation can be written and is of finite order, without solving the equation to find the roots. The procedure can also be applied to control loops, where the characteristic equation is:

$$Q(s) = 1 + G_C(s)\, G_S(s) = 0 \tag{5.106}$$

Expand the characteristic equation into the following form:

$$Q(s) = a_n s^n + a_{n-1} s^{n-1} + \ldots + a_1 s + a_0 = 0 \tag{5.107}$$

Let a_n be positive. If it is negative, both sides of Eq. (5.107) should be multiplied by -1.

First step: If any of the coefficients a_n, $a_{n-1} \ldots a_0$ is negative, there is at least one characteristic root which has a positive real part, and the system is unstable. No further analysis is needed.

Second step: If every coefficient is positive, then the following array (Routh array) should be constructed:

$$\begin{array}{lllll} 1 & a_n & a_{n-2} & a_{n-4} & \cdots \\ 2 & a_{n-1} & a_{n-3} & a_{n-5} & \cdots \\ 3 & b_1 & b_2 & b_3 & \cdots \\ 4 & c_1 & c_2 & c_3 & \cdots \\ 5 & d_1 & d_2 & d_3 & \cdots \\ \vdots & & & & \\ n-1 & w_1 & w_2 & w_3 & \end{array} \tag{5.108}$$

where the coefficients c, d, . . ., etc. are calculated as

$$b_1 = \frac{a_{n-1}a_{n-2} - a_n a_{n-3}}{a_{n-1}}$$

$$b_2 = \frac{a_{n-1}a_{n-4} - a_n a_{n-5}}{a_{n-1}}$$

$$b_3 = \frac{a_{n-1}a_{n-6} - a_n a_{n-7}}{a_{n-1}}$$

$$\vdots$$

$$c_1 = \frac{b_1 a_{n-3} - a_{n-1}b_2}{b_1}$$

$$c_2 = \frac{b_1 a_{n-5} - a_{n-1}b_3}{b_1}$$

$$\vdots$$

$$d_1 = \frac{c_1 b_2 - b_1 c_2}{c_1}$$

$$d_2 = \frac{c_1 b_3 - b_1 c_3}{c_1}$$

$$\vdots$$

etc. (5.109)

Examine the elements of the first column of the array in (Eq. 5.108):

$$a_n,\ n_{n-1},\ b_1,\ c_1,\ d_1,\ \ldots,\ w_1$$

A system is stable if all elements in the first column of the Routh array are positive.

Example 5.3.8

Check the stability of the process having the following transfer function:

$$G(s) = \frac{A}{1 - Ts} = -\frac{\frac{A}{T}}{s - \frac{1}{T}}$$

The characteristic equation of this process is:

$$Q(s)=s-\frac{1}{T}=0$$

The characteristic root is evident: $s_1=+1/T$. It is positive, thus the process is unstable.

Check the stability of this process by the Routh–Hurwitz criterion. $a_n=1$, i. e. it is positive, $a_{n-1}=-1/T$, i. e. negative. Thus, the process is found to be unstable by this method, too. The impulse response of this process is obtained as:

$$y(t)=\mathcal{L}^{-1}[G(s)\,X(s)]=\mathcal{L}^{-1}\left[\frac{\frac{A}{T}}{s-\frac{1}{T}}\cdot 1\right]$$

$$y(t)=-\frac{A}{T}e^{+\frac{t}{T}}$$

$y\to\infty$ when $t\to\infty$. Thus, the process is unstable.

Stabilize this unstable process using proportional control. Denote the proportional gain of the P-controller by A_C. The block diagram of the control loop is shown in *Fig. 5.43*. Eqs (5.100) and (5.101):

$$Q(s)=1+G_C(s)\,G_S(s)=1+\left(\frac{A}{1-Ts}\right)A_C=0$$

which can be arranged to

$$Q(s)=1-Ts+AA_C=-Ts+(1+AA_C)=0$$

The characteristic root is

$$s_1=+\frac{1+AA_C}{T}$$

s_1 is negative, and the control loop is stable if

$$A_C<0 \qquad \text{and} \qquad |A_C|>\frac{1}{A}$$

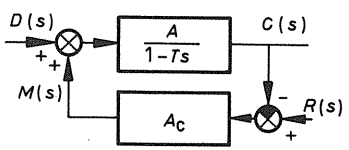

Fig. 5.43. To Example 5.3.8

5.3.5 The frequency function method

The analysis of frequency response is the third alternative method (besides the differential equation method and the transfer function method) for describing and examining the dynamics of linear systems. All these methods contain essentially the same information.

The frequency function method has already been mentioned and used in Section 5.2. It has the advantage of being directly applicable to dead time processes and to distributed parameter processes, and of being capable of utilizing direct experimental results. Frequency functions can be depicted as Nyquist and Bode plots; the resulting Bode plot of elements in series can easily be constructed by the simple addition of distances in the diagram. The stability of a negative feedback loop can easily be checked by the use of the Nyquist stability criterion (Section 5.2.1).

Determination of the frequency function

The frequency function of a process can be determined in three ways.

1. Experimentally, as already described in Section 5.2.1 and shown in *Fig. 5.11.* The process is driven by a continous sinusoidal signal until, after some time, a sine wave of the same frequency can be detected at the output. The result of the experiment is
   (*a*) the amplitude ratio of the output to the input sine wave, this is the magnitude of the frequency function vector $|G(i\omega)|$;
   (*b*) the phase shift of the output with respect to the input, which is the phase angle argument of the frequency function, $\measuredangle G(i\omega)$.

   The aim of the investigation is to obtain the frequency function over a broad frequency range, so it is a lengthy and expensive process (in the case of chemical processes, which are slow), but in the case of noisy signals it is advantageous that only one single frequency signal is to be detected.

   The frequency function can also be computed from responses to input signals containing a broad spectrum of frequencies, e. g. from impulse or step responses, or responses to white noise with the aid of Fourier analysis.
2. From the differential equation of the process, which can be solved substituting $x(t)=a \sin \omega t$, in simple cases, like in Section 5.2 for dead time, interval and differential elements.
3. From the transfer function, making the formal substitution $s=i\omega$.

Frequency function from transfer function

The procedure is illustrated by the example of the first-order process.

Example 5.3.9. *First-order lag. Frequency function*

The differential equation of a first-order element is given by Eq.(5.59):

$$T\frac{dy}{dt}+y=A_S x$$

By Laplace transformation:

$$TsY(s)+Y(s)=A_S X(s)$$

This can be rearranged to obtain the transfer function:

$$\frac{Y(s)}{X(s)} \equiv G(s)=\frac{A_S}{Ts+1}$$

Substitute $s=i\omega$

$$G(i\omega)=A_S\frac{1}{Ti\omega+1}\frac{(-iT\omega+1)}{(-iT\omega+1)}=A_S\left(\frac{1}{1+T^2\omega^2}-i\frac{T}{1+T^2\omega^2}\right) \quad (5.110)$$

This has the form $G(i\omega)=\mathrm{Te}_G+i\,\mathrm{Im}_G$ [cf. Eq.(5.9), *Fig. 5.12*]. The magnitude and the phase are obtained as:

$$\begin{aligned}|G(i\omega)| &= \sqrt{\mathrm{Re}_G^2+\mathrm{Im}_G^2}= \\ &= A_S\sqrt{\left(\frac{1}{1+T^2\omega^2}\right)^2+\left(\frac{T\omega}{1+T^2\omega^2}\right)^2}= \\ &= A_S\frac{1}{\sqrt{1+T^2\omega^2}}\end{aligned} \quad (5.111)$$

$$\sphericalangle G(i\omega)=\arctan\frac{\mathrm{Im}_G}{\mathrm{Re}_G}=\arctan(-\omega T). \quad (5.112)$$

The frequency function can be tabulated for different frequencies, e.g.

ω	ωT	$\lvert G(i\omega)\rvert$	$\sphericalangle G(i\omega)$
0	0	A_S	0°
$\frac{1}{T}$	1	$A_S\frac{1}{\sqrt{2}}$	−45°
$\to\infty$	$\to\infty$	$A_S\frac{1}{T\omega}$	$\to -90°$

For small frequencies ($\omega<1/T$) (e. g. close to the steady state) the first-order element resembles a *P* element; for high frequencies ($\omega>1/T$) (e. g. at the start of

a step or a pulse driving signal) it resembles to an I element. The Bode and the Nyquist diagrams of the first-order element are shown in *Fig. 5.44*. The amplitude characteristic in the Bode diagram is 'broken' at frequency $1/T$, called also the 'corner frequency' or 'breakpoint frequency'.

The general method for obtaining the frequency function from a transfer function can now be summarized:

1. $s=i\omega$ is substituted into the transfer function.
2. The expression obtained is arranged in the form:

$$G(i\omega)=\mathrm{Re}_G(i\omega)+i\,\mathrm{Im}_G(i\omega) \tag{5.113}$$

3. The magnitude of the frequency function is:

$$|G(i\omega)|=\sqrt{[\mathrm{Re}_G(i\omega)]^2+[\mathrm{Im}_G(i\omega)]^2} \tag{5.114}$$

4. The phase angle of the frequency function is:

$$\measuredangle G(i\omega)=\arctan\frac{\mathrm{Im}_G(i\omega)}{\mathrm{Re}_G(i\omega)} \tag{5.115}$$

5. The frequency function is evaluated for different frequencies.

It still remains to prove that a simple substitution of $s=i\omega$ into the transfer function of an arbitrary nth order system yields the same frequency function as that obtained experimentally by driving the system with a continuous sine wave.

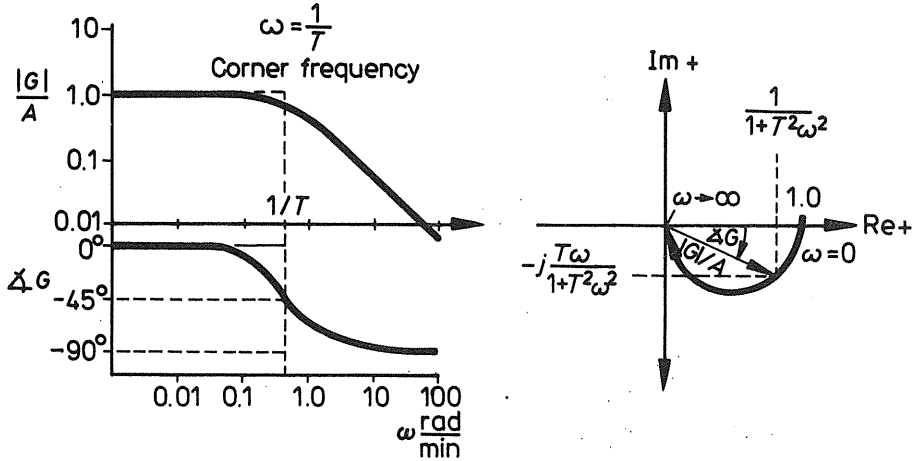

Fig. 5.44. Bode and Nyquist plots of first-order lag

The transfer function is:

$$G(s)=\frac{Y(s)}{X(s)} \tag{5.116}$$

where both $Y(s)$ and $X(s)$ are reciprocals of polynomials in s; $Y(s)$ are of nth and $X(s)$ are of mth order where $m<n$, and both polynomials can be factored into their roots:

$$G(s)=\frac{Y(s)}{X(s)}=\frac{(s-p_1)(s-p_2)\ldots(s-p_m)}{(s-s_1)(s-s_2)\ldots(s-s_n)} \tag{5.117}$$

The input is a sine wave the Laplace transform of which is, from *Table 5.4*,

$$X(a)=a\left(\frac{\omega}{s^2+\omega^2}\right) \tag{5.118}$$

$$Y(s)=X(s)G(s)=\left[\frac{a\omega}{s^2+\omega^2}\right]\left[\frac{(s-p_1)(s-p_2)\ldots(s-p_m)}{(s-s_1)(s-s_2)\ldots(s-s_n)}\right] \tag{5.119}$$

Expansion in partial fractions gives

$$Y(s)=\frac{aG(s)}{s^2+\omega^2}=\frac{B}{s+i\omega}+\frac{C}{s-i\omega}+\frac{D}{s-s_1}+\ldots+\frac{Q}{s-s_n} \tag{5.120}$$

Determine $B, C, D, \ldots Q$, using Eq.(5.83) in Section 5.3.4, and making use of the fact that Eqs (5.119) and (5.120) are identical equations for all values of s:

$$B=\lim_{s\to -i\omega}\left[(s+i\omega)\,Y(s)\right]=\lim_{s\to i\omega}\left[\frac{\omega a G(s)}{s-i\omega}\right]=-\frac{a}{2i}G(-i\omega)$$

$$C=\lim_{s\to i\omega}\left[(s-i\omega)\,Y(s)\right]=\lim_{s\to i\omega}\left[\frac{\omega a G(s)}{s+i\omega}\right]=\frac{a}{2i}G(i\omega)$$

$$D=\lim_{s\to s_1}\left[(s-s_1)\,Y(s)\right]=\lim_{s\to s_1}\left[(s-s_1)\left(\frac{a}{s^2+\omega^2}\right)G(s)\right],$$

etc.

Substituting $B, C, D\ldots$ into Eq.(5.120) and taking the inverse of the time domain:

$$y(t)=-\frac{a}{2i}G(-i\omega)\,e^{-i\omega t}+\frac{a}{2j}G(i\omega)\,e^{i\omega t}+\sum_{j=1}^{n} f_j e^{s_j t} \tag{5.121}$$

For a stable system all the characteristic roots $s_1, s_2 \ldots s_j \ldots s_n$ have negative real parts. Thus, with time, all the exponential terms in the sum in Eq.(5.121) decay to zero. So, as time passes, Eq.(5.121) becomes

$$y(t)=\frac{a}{2i}\left[G(i\omega)e^{i\omega t}-G(-i\omega)e^{-i\omega t}\right] \tag{5.122}$$

Note: $G(i\omega)$ and $G(-i\omega)$ are conjugate complex functions, which are characterized by equal real parts and imaginary parts that are also equal but opposite in sense. Represented in polar coordinates, the magnitudes are equal, and the arguments are equal too, but of opposite signs.

Using the Euler equations:

$$\begin{aligned} e^{ix} &= \cos x + i \sin x \\ e^{-ix} &= \cos x + i \sin x \end{aligned} \tag{5.123}$$

we obtain

$$\begin{aligned} \sin x &= \frac{1}{2i}(e^{ix} - e^{-ix}) \\ \cos x &= \frac{1}{2i}(e^{ix} + e^{-ix}) \end{aligned} \tag{5.124}$$

Write $G(i\omega)$ and $G(-i\omega)$ in the polar form of complex numbers:

$$G(i\omega) = |G(i\omega)|\, e^{i * G(i\omega)}$$

$$G(-i\omega) = |G(-i\omega)|\, e^{i * G(-i\omega)} = |G(i\omega)|\, e^{-i * G(i\omega)}$$

Using these forms:

$$y(t) = a|G(i\omega)| \frac{e^{i[\omega t + * G(i\omega)]} - e^{-i[\omega t + * G(i\omega)]}}{2i} \tag{5.125}$$

we obtain the result:

$$y(t) = a[|G(i\omega)| \sin(\omega t) + \text{arc } G(i\omega)] \tag{5.126}$$

Thus it has been proved that

(*i*) the amplitude ratio is the absolute value of $G(s)$ with "s" set equal to "$i\omega$";
(*ii*) the phase angle is the argument of $G(s)$ with "s" set equal to "$i\omega$".

Example 5.3.10. *Response of a first-order lag to a sine input*

The response of a first-order element with

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

to a sine input $x(t) = a \sin \omega t$; $X(s) = \dfrac{a\omega}{s^2+\omega^2}$ is obtained in the time domain by taking the time inverse

$$y(t) = \mathscr{L}^{-1}[X(s)G(s)] = \mathscr{L}^{-1}[Y(s)] \tag{5.127}$$

where

$$Y(s) = \left(\frac{a\omega}{s^2+\omega^2}\right)\left(\frac{A}{Ts+1}\right) = aA\frac{\omega}{T}\left[\frac{1}{(s^2+\omega^2)\left(s+\frac{1}{T}\right)}\right] \tag{5.128}$$

The roots of the characteristic equation

$$Q(s)=(s^2+\omega^2)\left(s+\frac{1}{T}\right)=0$$

are evidently $+i\omega$, $-i\omega$ and $-1/T$. However, to reduce complex algebraic work, it is more practical to factor $Y(s)$ as

$$Y(s)=aA\frac{\omega}{T}\left(\frac{B\omega+Cs}{s^2+\omega^2}+\frac{D}{s+\frac{1}{T}}\right) \tag{5.129}$$

which, transformed into the time domain, gives

$$y(t)=aA\frac{\omega}{T}\left(B\sin\omega t+C\cos\omega t+De^{-\frac{t}{T}}\right)$$

$$=aA\frac{\omega}{T}\left(\sqrt{B^2+C^2}\sin\left(\omega t+\arctan\frac{C}{B}\right)+De^{-\frac{t}{T}}\right) \tag{5.130}$$

The coefficients B and C are evaluated from

$$B\omega+Ci\omega=\lim_{s\to i\omega}\frac{(s^2+\omega^2)}{(s^2+\omega^2)\left(s+\frac{1}{T}\right)}=\lim_{s\to i\omega}\frac{1}{\left(s+\frac{1}{T}\right)}=$$

$$=\left(\frac{1}{i\omega+\frac{1}{T}}\right)\left(\frac{\frac{1}{T}-i\omega}{\frac{1}{T}-i\omega}\right)=\frac{T-i\omega T^2}{1+\omega^2T^2} \tag{5.131}$$

whence

$$B=\frac{\frac{\omega}{T}}{1+\omega^2T^2};\qquad C=-\frac{T^2}{1+\omega^2T^2} \tag{5.132}$$

$$\arctan\frac{C}{B}=\arctan(-\omega T)=\varphi \tag{5.133}$$

and

$$aA\frac{\omega}{T}\sqrt{B^2+C^2}=aA\sqrt{\frac{1+\omega^2T^2}{(1+\omega^2T^2)^2}}=aA\frac{1}{\sqrt{1+\omega^2T^2}} \tag{5.134}$$

Coefficient D is evaluated using Eq.(5.83):

$$D=\lim_{s\to -1/T}\frac{s+\frac{1}{T}}{(s^2+\omega^2)\left(s+\frac{1}{T}\right)}=\frac{1}{\frac{1}{T^2}+\omega^2}=\frac{T^2}{1+\omega^2T^2} \tag{5.135}$$

Substitute Eqs (5.133), (5.134) and (5.135) into Eq.(5.130):

$$y(t)=aA\left[\left(\frac{1}{\sqrt{1+\omega^2T^2}}\right)\sin\left(\omega t-\arctan\left(-\omega T\right)\right)+\left(\frac{\omega T}{1+\omega^2T^2}\right)e^{-t/T}\right] \quad (5.136)$$

Compare Eq.(5.136) with the result of Example 5.3.9 given in Eqs (5.111) and (5.112). After a time longer than $3T$ the transient decays to zero and the steady-state part of the response remains: this is a sine wave with the same frequency ω as the input. Its amplitude is $\frac{A}{\sqrt{1+\omega^2T^2}}$ times that of the input. Its phase lag with respect to the input is

$$\varphi=\arctan\left(-\omega T\right)$$

In the case of higher-order elements the transient part is a sum of functions similar to the transient term in Eq.(5.136). The time necessary for the transients to decay is longer than $3T_{max}$ where T_{max} is the greatest time constant. Since chemical engineering processes may have time constants of the order of several hours, the experimental determination of the frequency function by sinusoidal driving is rather expensive and is rarely applied.

5.3.6 Simple linear systems

Table 5.5 contains the differential equations and the transfer functions of the most important systems. Some of them, such as *P*, *I*, *D*, *PI* and *PID* elements, have been treated already in Section 5.2 and are not repeated here.

Figure 5.45 represents the impulse, step and ramp responses of these systems.

Figure 5.46 shows the frequency functions represented as Nyquist and Bode plots.

First-order element

A first-order element is described by a first-order ordinary differential equation.

$$T\frac{dy}{dt}+y=Ax \quad (5.137)$$

Its transfer function was derived in Example 5.3.4 as Eq.(5.68)

$$G(s)=\frac{A_S}{Ts+1} \quad (5.138)$$

The unit impulse response [Example 5.3.5; Eq.(5.72)]:

$$y(t)=\frac{A_S}{T}e^{-\frac{t}{T}}$$

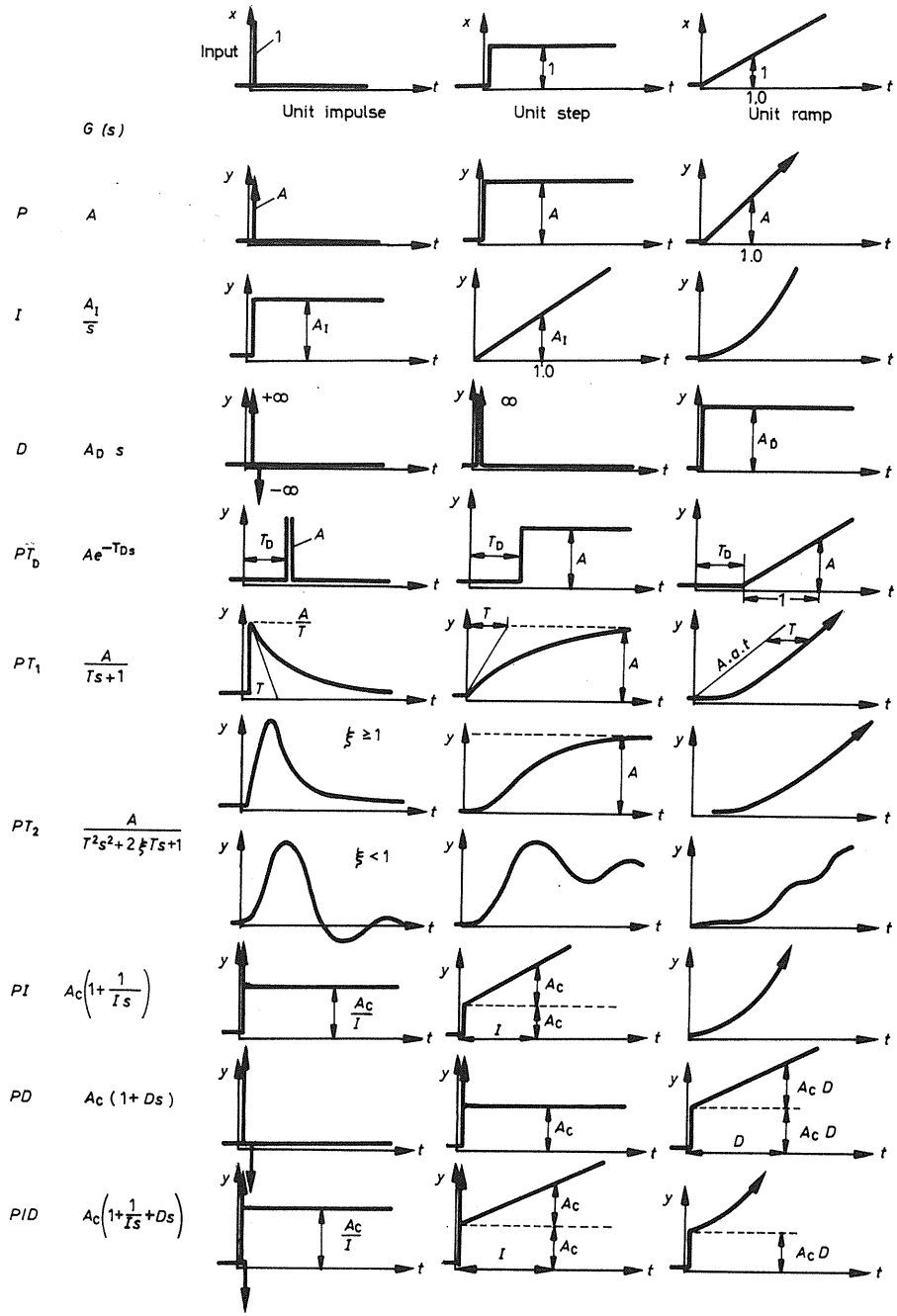

Fig. 5.45. Response functions of linear systems

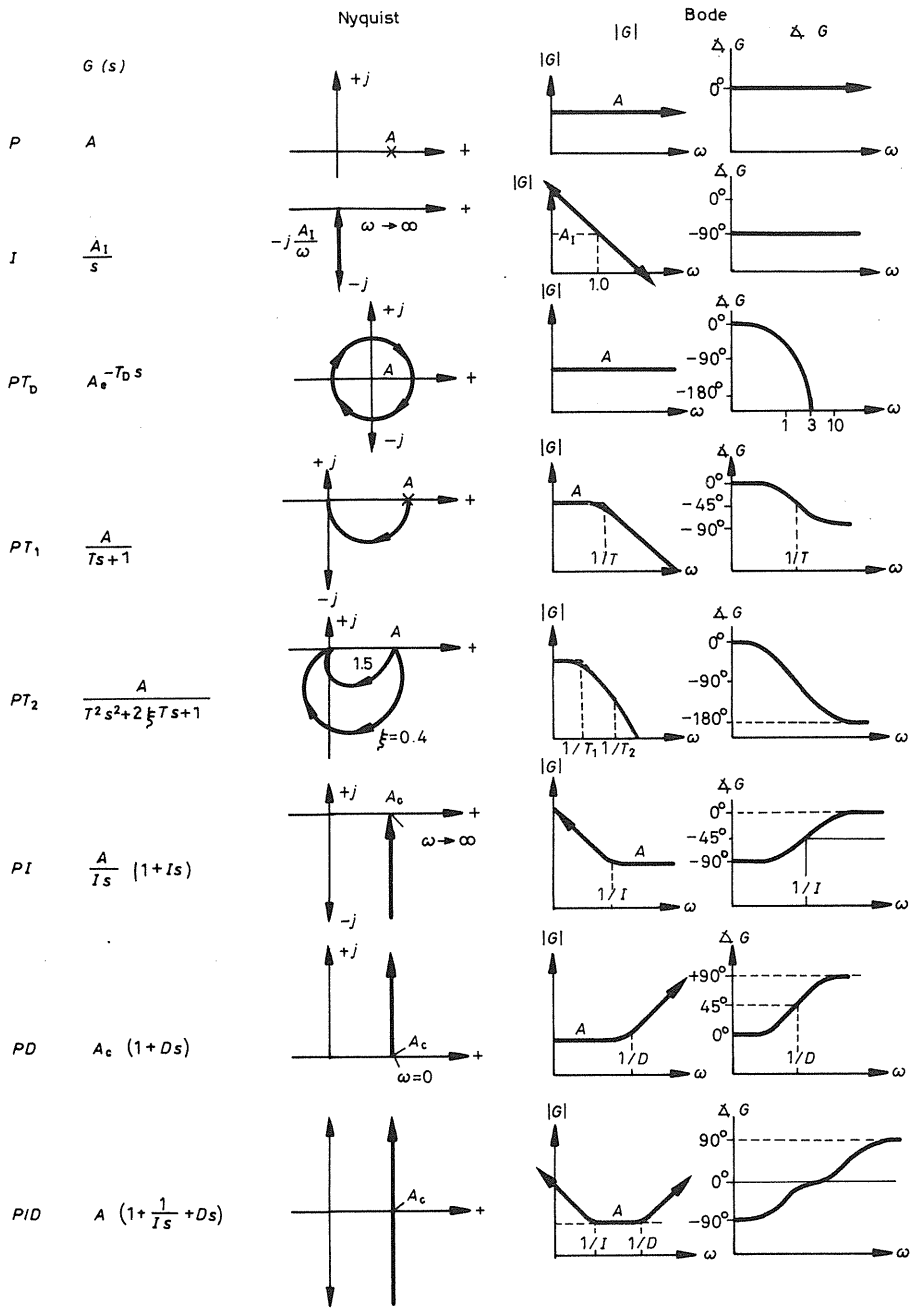

Fig. 5.46. Frequency functions of linear systems

Table 5.5. Equations of simple linear systems

	Differential equation	Transfer function $G(s)$	Frequency function	
			$\lvert g(i\omega)\rvert$	$\measuredangle g(i\omega)$
Proportional	$y=Ax$	A	1	0°
Integrating	$v=A_I\int x\,dt$	$\frac{A_I}{s}$	$\frac{A_I}{\omega}$	−90°
Dead time	$y(t)=Ax(t-T_D)$	$Ae^{-T_D s}$	1	$-T_D\omega$(rad)
First-order lag	$T\frac{dy}{dt}+y=Ax$	$\frac{A}{Ts+1}$	$\frac{1}{\sqrt{1+T^2\omega^2}}$	$\arctan(-T\omega)$
Second-order lag	$T^2\frac{d^2y}{dt^2}+2\xi T\frac{dy}{dt}+y=Ax$	$\frac{A}{T^2s^2+2\xi Ts+1}$	$\frac{1}{\sqrt{(1-T^2\omega^2)^2+(2\xi T\omega)^2}}$	$\arctan\left(-\frac{2\xi T\omega}{1-T^2\omega^2}\right)$
PI-controller	$y=A_C\left(x+\frac{1}{I}\int x\,dt\right)$	$\frac{A_C}{Is}(1+Is)$	$1+\left(\frac{1}{I\omega}\right)^2$	$\arctan\left(-\frac{1}{I\omega}\right)$
PD-controller (ideal)	$y=A_C\left(x+D\frac{dx}{dt}\right)$	$A_C(1+Ds)$	$\sqrt{1+(D\omega)^2}$	$\arctan(D\omega)$
PID-controller (ideal)	$y=A_C\left(x+\frac{1}{I}\int x\,dt+D\frac{dx}{dt}\right)$	$\frac{A_C}{Is}(1+Is+DI_S)$	$A_C\sqrt{1+\left(\frac{-1}{I\omega}+D\omega\right)^2}$	$\arctan\left(-\frac{1}{I\omega}+D\omega\right)$

The unit step response [Example 5.3.6; Eq.(5.75)]:

$$y(t)=A_S\left(1-e^{-\frac{t}{T}}\right) \tag{5.139}$$

The unit ramp response [Example 5.3.7; Eq.(5.96)]:

$$y(t)=A_S\left[Te^{-\frac{t}{T}}+(t-T)\right] \tag{5.140}$$

The response functions to these specific inputs are represented in *Fig. 5.38*. The two parameters of the system, A_S and T, can both be determined from any of these response functions, as shown in *Fig. 5.38*. It is important to note that a first-order element can be distinguished from higher-order and from dead time systems by its impulse response, which attains its maximum value at the instant of the disturbance, and by the slope of its step response, which is a maximum at the instant of the disturbance.

Another method of checking whether the system is truly first-order is using the response to a step input $x(t)=a$. Equation (5.139) is transformed into:

$$y(t)=A_S\left(1-e^{-\frac{t}{T}}\right)=y(\infty)\left(1-e^{-\frac{t}{T}}\right) \tag{5.141}$$

which can be arranged to give:

$$\ln\frac{y(\infty)-y(t)}{y(\infty)}=-\frac{1}{T}t \tag{5.142}$$

In the step response is plotted in a semi-logarithmic plot of $\ln\frac{y(\infty)-y(t)}{y(\infty)}$ against time, all points must lie on a straight line of slope $-\frac{1}{T}$.

The first-order element can also be identified from its frequency function (see *Fig. 5.44*). It is important to note that the minimum phase shift exhibited by a first-order element is $-90°$. If $-100°$ or $-120°$ phase shifts are measured, the system is not first-order. In order to determine the time constant, input sinusoids with frequencies around the breakpoint frequency, i.e. $\omega=\frac{1}{T}$, must be used.

Examples of first-order systems

There are many first-order systems in the chemical industry. Single-capacity systems which are self-regulating, i.e. which are not integrating elements, are first-order systems. By analogy with electrical practice, a first-order element is a capacitor charged through a resistor, and its time constant is interpreted as capacitance × resistance, $T=CR$.

Continuous stirred tank reactor (CSTR)

Consider the tank reactor shown in *Fig. 5.47*. It is perfectly mixed, thus the concentration in it is uniform, independent of the position in the tank. As a result, the composition of the output stream is equal to that in the tank: $c_0 = c$. First order, irreversible, isothermal reaction takes place in the tank with reaction rate k:

$$A \rightarrow P$$

The volume holdup in the tank, V, is constant. There is no density change, thus the mass balance may be written in terms of volume rates:

$$\text{input} = \text{output} + \text{accumulation}$$

$$W_i \, dt = W_0 \, dt + dV \tag{5.143}$$

However, $V = \text{const}$, thus

$$W_i = W_0 = W \tag{5.144}$$

Suppose that W is constant in time. The input variable is c_i, the output is c. The component balance is:

$$\text{input} = \text{output} + \text{consumed} + \text{accumulated}$$

$$W c_i \, dt = W c \, dt + V k c \, dt + V \, dc \tag{5.145}$$

The component balance can be arranged to give:

$$\left(\frac{W}{W + Vk}\right) c_i = c + \left(\frac{V}{W + Vk}\right) \frac{dc}{dt} \tag{5.146}$$

This is the differential equation of a first-order element, with

$$A_S = \frac{W}{W + Vk}; \qquad T = \frac{V}{W + Vk} \tag{5.147}$$

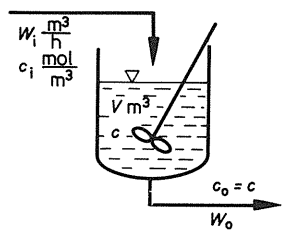

Fig. 5.47. Continuous stirred tank reactor

where V is capacitance and $1/(W+Vk)$ is resistance. In the steady-state $\frac{dc}{dt}=0$, and from Eq.(5.146):

$$c_A = \frac{W}{W+Vk} c_{A_i} \tag{5.148}$$

If the reaction in the *CSTR* is not first-order, but the reaction rate is a function of concentration, $r=f(c_A)$, then linearization and perturbation variables are to be used (cf. Section 5.3.3):

$$\hat{r} = \hat{c}\left(\frac{\partial r}{\partial c}\right)_{\bar{c}} \tag{5.149}$$

where $\bar{c}$ is the concentration at the steady state. The steady-state component balance is:

$$W\bar{c}_i = W\bar{c} + V\bar{r}, \tag{5.150}$$

where $\bar{c}_i$, $\bar{c}$ and $\bar{r}$ are the steady-state values. The component balance in the non-steady-state is:

$$\text{input} = \text{output} + \text{consumed} + \text{accumulated} \tag{5.151}$$

$$Wc_i\,dt = Wc\,dt + Vr\,dt + V\,dc$$

Rearrange Eq.(5.151):

$$Wc_i = Wc + Vr + V\frac{dc}{dt} \tag{5.152}$$

Subtracting Eq.(5.150) from Eq.(5.152) we obtain the differential equation in perturbation variables ($\hat{c}_i = c_i - \bar{c}_i$, $\hat{r} = r - \bar{r}$):

$$W\hat{c}_i = W\hat{c} + V\hat{r} + V\frac{d\hat{c}}{dt} \tag{5.153}$$

Substitute $\hat{r}$ from Eq.(5.149) and arrange into the form of the differential equation of a first-order element:

$$\left(\frac{W}{W+V\left(\frac{\partial r}{\partial c}\right)_{\bar{c}}}\right)\hat{c}_i = \hat{c} + \left(\frac{V}{W+V\left(\frac{\partial r}{\partial c}\right)_{\bar{c}}}\right)\frac{d\hat{c}}{dt} \tag{5.154}$$

whence

$$A_S = \frac{W}{W+V\left(\frac{\partial r}{\partial c}\right)_{\bar{c}}}; \qquad T = \frac{V}{W+V\left(\frac{\partial r}{\partial c}\right)_{\bar{c}}}$$

Parameters A_S and T both depend on the operation point.

Perfectly mixed tanks are generally first-order elements. It is conceivable that a perfectly mixed blending tank, where no reaction occurs, where the output is the

concentration of the blend, and where the input variables are the feed rate and the feed concentration, is a first-order lag. A perfectly mixed continuous heater (or cooler) with continuous inflow and outflow is a first-order lag, too: the output is the temperature of the outflow, the input variables are the temperature and the flow rate of the inflow, as well as the heat input. Other examples of first-order elements are:

— thermometer bulb,
— liquid tank discharging through a valve,
— gas tank.

Here the differential equation of the liquid tank will be derived. The derivation for the case of a gas tank is quite similar.

Self-regulating liquid tank

The liquid level system presented in Section 5.2.3 and *Fig. 5.27* is an integrating process, because its outflow is fixed by a metering pump: the outflow is independent of the liquid level in the tank.

The liquid tank in *Fig. 5.48* discharges through a valve, which represents a resistance. The flow through the valve depends on the valve opening, which is included in the valve coefficient k_v (k_v changes with the valve opening), and on the pressure drop $(p_1 - p_2)$ across the valve:

$$W_c = k_v \sqrt{\frac{p_1 - p_2}{\varrho}} \tag{5.155}$$

The dependence of the outflow rate on the liquid level is the strongest when the whole pressure drop $(p_1 - p_2)$ is covered by the liquid head in the tank:

$$p_1 - p_2 = \varrho g H . \tag{5.156}$$

Using Eq.(5.156), we obtain the outflow as a (nonlinear) function of the level:

$$W_0 = k_v \sqrt{\frac{\varrho g H}{\varrho}} = k \sqrt{H} \tag{5.157}$$

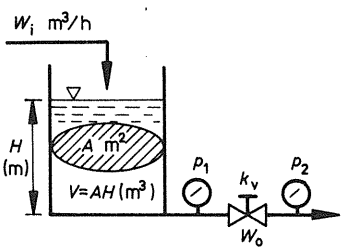

Fig. 5.48. Self-regulating liquid tank

Linearization around the steady state ($\bar{W}_0$, $\bar{H}$) is needed:

$$W_0 = \bar{W}_0 + \hat{H}\left(\frac{\partial W_0}{\partial H}\right)_{\bar{H}}. \tag{5.158}$$

The steady-state mass balance of the tank is:

$$\bar{W}_i = \bar{W}_0 \qquad \text{and} \qquad \bar{H} = \left(\frac{\bar{W}_0}{k}\right)^2 \tag{5.159}$$

thus, in the steady state, the level is determined by the inflow. The mass balance in the non-steady-state is:

$$\text{input} = \text{output} + \text{accumulation}$$

$$W_i\,\mathrm{d}t = W_0\,\mathrm{d}t \; + A\,\mathrm{d}H. \tag{5.160}$$

Substituting Eqs (5.159) and (5.158) into Eq.(5.160) and rearranging, the differential equation of a first-order lag is obtained with perturbation variables:

$$\frac{1}{\left(\frac{\partial W_0}{\partial H}\right)_{\bar{H}}}\hat{W}_i = \hat{H} + \frac{1}{\left(\frac{\partial W_0}{\partial H}\right)_{\bar{H}}}\frac{\mathrm{d}\hat{H}}{\mathrm{d}t} \tag{5.161}$$

The time constant is

$$T = \frac{A}{\left(\frac{\partial W_0}{\partial H}\right)_{\bar{H}}}$$

where A is capacitance, and $\left(\frac{\partial H}{\partial W_0}\right)_{\bar{H}}$ is resistance.

Note 1: In the case of pressurized tanks, when the pressure loss through the restriction is covered partly by the liquid head in the tank and partly by the pressure above the liquid (or by a centrifugal pump), the total effective pressure drop can be expressed in terms of the liquid head, i.e. $p_1 - p_2 = B\varrho gH$ is substituted into Eq.(5.156), and the final result is the same.

Note 2: Observe that in this nonlinear process both the steady-state gain and the time constant depend on the operating point ($\bar{W}$, $\bar{H}$). Compare this with the result obtained for the nonlinear *CSTR*, Eq.(5.154). However, the dynamic gains for high frequencies are independent of the operating point [cf. Eq.(5.111) and subsequent equations]:

$$|G(i\omega)| = \frac{A_S}{\sqrt{1+T^2\omega^2}}$$

for $\omega \to \infty$ $|G(i\omega)| \dfrac{A_S}{T\omega}$

For the *CSTR* $|G| \rightarrow \frac{V}{W}\frac{1}{\omega}$

For the liquid level tank $|G| \rightarrow \frac{1}{A}\frac{1}{\omega}$

This phenomenon is frequently encountered in chemical engineering processes.

Second-order systems

A second-order system is modelled by a second-order differential equation:

$$T^2\frac{d^2y}{dt^2}+2\xi T\frac{dy}{dt}+y=Ax \tag{5.162}$$

where y is output, x is input, T is time constant, A is process gain, and ξ is damping factor.

Second-order systems can be classified into three categories according to the value of the damping factor.

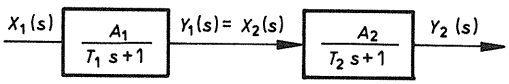

Fig. 5.49. Two first-order lags in series

1. Overdamped system

In this case the damping factor is greater than unity. This happens whenever two first-order systems are in series and their time constants are not equal. For example: two *CSTR*s in series, a thermometer in a heat exchanger, buffer tank and *CSTR*, and so on. *Figure 5.49* represents such a case.

The transfer function is:

$$\frac{Y_2(s)}{X_1(s)}=\left(\frac{A_1}{T_1s+1}\right)\left(\frac{A_2}{T_2s+1}\right)=\frac{A_1A_2}{T_1T_2s^2+(T_1+T_2)s+1}=\frac{A}{T^2s^2+2\xi Ts+1} \tag{5.163}$$

Comparing Eq.(5.163) with Eq.(5.162), the parameters are:

$$A=A_1A_2$$

$$T=\sqrt{T_1T_2} \tag{5.164}$$

$$\xi=\frac{1}{\sqrt{T_1T_2}}\left(\frac{T_1+T_2}{2}\right) \tag{5.165}$$

If $T_1 \neq T_2$ then $\xi<1$.

The responses to typical unit inputs are *(Fig. 5.45)*:

unit impulse $$y(t)=\frac{A}{T_1-T_2}\left(e^{-\frac{t}{T_1}}-e^{-\frac{t}{T_2}}\right) \tag{5.166}$$

The slope of the impulse response is a maximum at the instant of the disturbance $A/(T_1T_2)$. The impulse response has a maximum at time $t_{max}=\sqrt{T_1T_2}$.

unit step $$y(t)=A\left[1-\frac{1}{T_1-T_2}\left(T_1e^{-\frac{t}{T_1}}-T_2e^{-\frac{t}{T_2}}\right)\right] \tag{5.167}$$

$$\left(\frac{dy}{dt}\right)_{t=0}=0$$

unit ramp $$y(t)=A\left[\frac{1}{T_1-T_2}\left(T_1^2e^{-\frac{t}{T_1}}-T_2^2e^{-\frac{t}{T_2}}\right)+\left(t-(T_1+T_2)\right)\right]$$

$$\text{for}\quad t>3T_1\quad\text{and}\quad t>3T_2,\quad y(t)=A\left(t-(T_1+T_2)\right) \tag{5.168}$$

2. Critically damped system

This is the same as Case 1, presented in *Fig. 5.49*, with the transfer function in Eq.(5.163), but $T_1=T_2$. Thus the model parameters are

$$T=\sqrt{T_1T_2}=T_1=T_2 \tag{5.169}$$

$$\xi=\frac{1}{\sqrt{T_1T_2}}\left(\frac{T_1+T_2}{2}\right)=1 \tag{5.170}$$

The system is called critically damped, because it does not exhibit oscillations in response to the typical inputs but attains the steady state faster than over-damped systems with the same T parameter. The responses to typical inputs are *(Fig. 5.45)*:

unit impulse $$y(t)=\frac{A}{T^2}te^{-\frac{t}{T}} \tag{5.171}$$

$$\text{with}\quad\left(\frac{dy}{dt}\right)_0=\frac{A}{T^2}\quad\text{and}\quad t_{max}=t$$

unit step

$$y(t)=A\left[1-e^{-\frac{t}{T}}\left(1+\frac{t}{T}\right)\right] \tag{5.172}$$

$$\text{with} \quad \left(\frac{dy}{dt}\right)_0=0$$

unit ramp

$$y(t)=A\left[e^{-\frac{t}{T}}(t+2T)+(t-2T)\right] \tag{5.173}$$

$$\text{for} \quad t>3T \qquad y(t)=A(t-2T)$$

3. Underdamped system

The damping factor is less than unity, but always greater than zero. Second-order underdamped systems cannot be constructed from two first-order elements in series; they are second-order elements in nature, which is due to inertial effects and internal interactions. Examples of second-order underdamped systems are: a *U*-tube manometer; a mass suspended from a spring, etc.

The underdamped character manifests itself in the impulse or step response by overshooting the final value and reaching it after damped oscillations. The frequency of the damped oscillations is

$$\omega=\frac{1}{T}\sqrt{1-\xi^2} \tag{5.174}$$

whence the meaning of parameter T: $1/T$ would be the oscillation frequency if the system were undamped, i.e. for $\xi=0$.

The responses to typical inputs are presented in *Fig. 5.45.*

All second-order systems (independent of the value of the damping factor) can be distinguished from first-order systems by examining the initial slopes of the impulse and step responses. The initial slope of the impulse response of a first-order system is infinite, while that of a second-order system is a maximum (for higher-order systems it is zero). The initial slope of the step response of a first-order system is a maximum, that of second- (or higher-) order systems is zero.

The transfer function of second-order systems is obtained by the Laplace transformation of the differential equation, Eq.(5.162):

$$G(s)=\frac{Y(s)}{X(s)}=\frac{A}{T^2s^2\xi Ts+1}=\frac{A/T^2}{s^2\dfrac{2\xi}{T}s+\dfrac{1}{T^2}} \tag{5.175}$$

The magnitude and phase angle of the frequency function are obtained by substituting $s=i\omega$ into Eq.(5.175) and following the procedure described in Section 5.3.5:

$$|G(i\omega)|=\frac{A}{\sqrt{(1-\omega^2T^2)^2+(2\xi\omega T)^2}} \tag{5.176}$$

$$\sphericalangle G(i\omega) = \arctan\left(-\frac{2\xi\omega T}{1-\omega^2 T^2}\right) \tag{5.177}$$

Of course the response to a sine input with frequency ω is, in the case of underdamped systems, another sine wave with frequency ω, but the frequency function shown in *Fig. 5.46* reveals its underdamped character. There exists a resonance frequency $\omega = \frac{1}{T}$, where the dynamic gain of underdamped systems is several times the static gain. Important values of the frequency function are:

$$\omega \to 0 \; : \; |G| \to A \qquad \sphericalangle G \to 0° \tag{5.178}$$

$$T = 1 \; : \; |G| = \frac{A}{2\xi} \qquad \sphericalangle G = -90° \tag{5.179}$$

$$\omega \to \infty : \; |G| = \frac{A}{\omega^2 T^2} \qquad \sphericalangle G \to -180° \tag{5.180}$$

$$\omega \to \infty : \; \frac{\mathrm{d}\log|G|}{\mathrm{d}\log|\omega|} \to -2\,, \tag{5.181}$$

which represent the slope of the Bode plot.

Thus second-order systems may readily be identified from their frequency functions, which should be determined around the frequency $\omega = 1/T$.

It is important to note that the minimum phase of second-order elements is $2 \cdot (-90°) = -180°$.

Higher-order systems

Elements of order n are described by nth order ordinary differential equations, such as Eq.(5.57); their transfer functions are the reciprocals of nth order polynomials in s.

They usually consist of first- (and second-) order elements in series, in other words, they are multicapacity systems. The capacities of a multicapacity system can be interacting or noninteracting (isolated). In noninteracting multicapacity systems the signals are transferred in one direction only; the conditions in the mth capacity do not depend on the conditions in the $(m+1)$th capacity. A cascade of *CSTR*s is an example of a noninteracting multicapacity system. In such systems the time constants of the individual capacities can be calculated independently from their mathematical models (or $T_i = C_i R_i$).

In interacting multicapacity systems the signals are transferred in both directions, as in a series of interconnected gas tanks, where a pressure change at both ends of the system would cause the pressure to change in all tanks *(Fig. 5.50)*. Other examples of interacting multicapacity systems are: multitray distillation

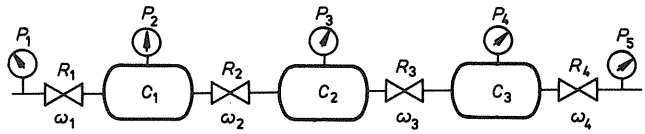

Fig. 5.50. Interacting gas tanks

columns, countercurrent extractors and many other chemical engineering processes. Consider the series of three interacting gas tanks presented in *Fig. 5.50.*

The input variables are p_1 and p_5. In the steady state

$$\bar{w}_1 = \bar{w}_2 = \bar{w}_3 = \bar{w}_4 = \frac{\bar{p}_1 - \bar{p}_2}{R_1} = \frac{\bar{p}_2 - \bar{p}_3}{R_2} = \frac{\bar{p}_3 - \bar{p}_4}{R_3} = \frac{\bar{p}_4 - \bar{p}_5}{R_4}$$

In the non-steady-state

$$\text{input} = \text{output} + \text{accumulation}$$

$$\frac{\hat{p}_1 - \hat{p}_2}{R_1} = \frac{\hat{p}_2 - \hat{p}_3}{R_2} + C_1 \frac{\mathrm{d}\hat{p}_2}{\mathrm{d}t}$$

$$\frac{\hat{p}_2 - \hat{p}_3}{R_2} = \frac{\hat{p}_3 - \hat{p}_4}{R_3} + C_2 \frac{\mathrm{d}\hat{p}_3}{\mathrm{d}t}$$

$$\frac{\hat{p}_3 - \hat{p}_4}{R_3} = \frac{\hat{p}_4 - \hat{p}_5}{R_4} + C_3 \frac{\mathrm{d}\hat{p}_4}{\mathrm{d}t}$$

This third-order system is described by a mathematical model consisting of three first-order differential equations which, however, are not independent of each other. If interaction is ignored, the equation for the first tank can be arranged as:

$$\left(\frac{1}{\frac{1}{R_1} + \frac{1}{R_2}}\right)\hat{p}_1 + \left(\frac{1}{\frac{1}{R_2} + \frac{1}{R_3}}\right)\hat{p}_3 = \hat{p}_2 + \left(\frac{1}{\frac{1}{R_1} + \frac{1}{R_2}}\right) \cdot \frac{\mathrm{d}\hat{p}_2}{\mathrm{d}t}$$

whence the time constant for the first tank can be expressed as:

$$T_1 = \frac{C_1}{\frac{1}{R_1} + \frac{1}{R_2}}; \qquad \text{and} \qquad T_i = \frac{C_i}{\frac{1}{R_i} + \frac{1}{R_{i+1}}}$$

If interaction is not ignored, the mathematical model can be solved to obtain three time constants $T_{i,\,\mathrm{eff}} \neq T_i$. The effect of interaction is to increase the greatest individual time constant and at the same time to reduce the smaller ones. Interaction will make the process more sluggish (it increases the sum of the effective time constants) but easier to control.

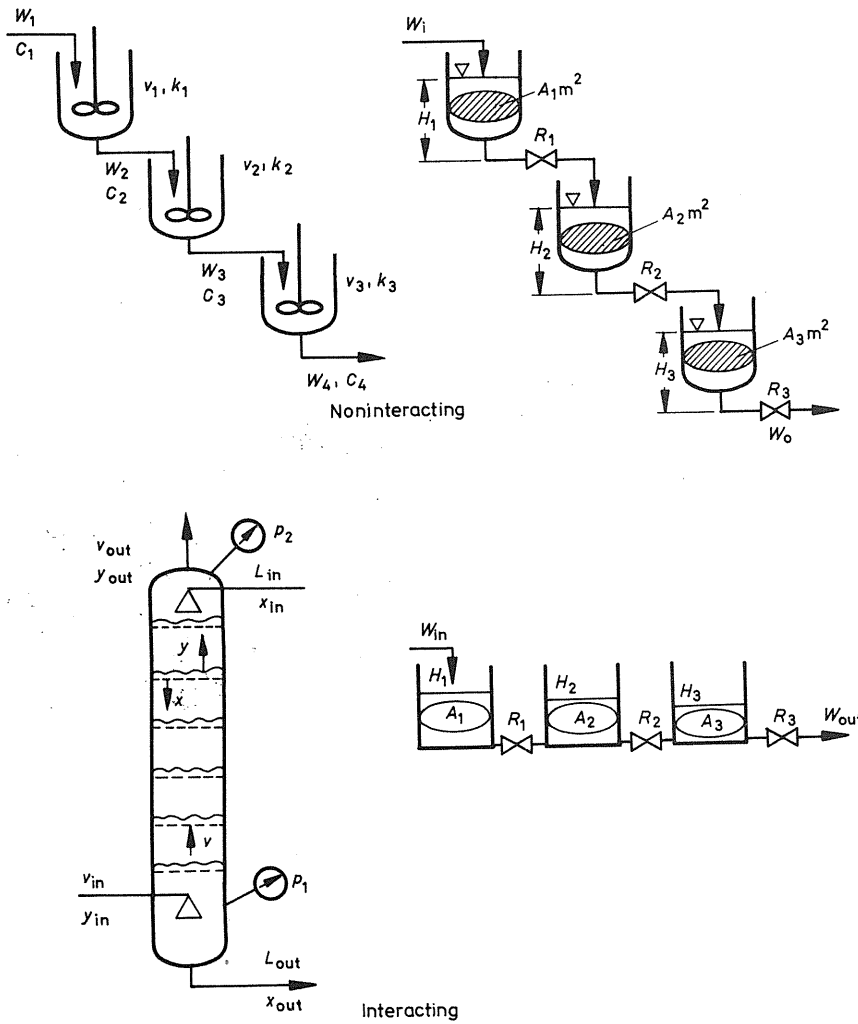

Fig. 5.51. Examples of interacting and noninteracting multicapacity systems

Some examples of interacting and noninteracting systems are shown in *Fig. 5.51*. The step responses are compared in *Fig. 5.52*.

Interacting or not, *n*-capacity systems are modelled by *n*th order differential equations and *n*th order transfer functions.

The time responses of higher-order systems may be calculated by inverse transformation of the Laplace transforms, but this needs considerable algebraic work. Typical responses in analytical form up to fourth-order may be found in handbooks. For higher-order systems numerical integration is easier, using analogue or

digital computers. The program *TACS*, the user manual of which is reproduced in the Appendix, can be applied to that end.

The responses to the typical input signals given in *Fig. 5.45* show the common characteristic that, at the start of the response, there is no change in the output signal, as if the system contained a dead time. This apparent dead time grows with the order of the system; it is first observed in the step response of the third-order system.

The order of the system cannot be determined from impulse, step or ramp responses by inspection, as was the case with first- and second-order systems. For this purpose frequency functions are needed.

The frequency function of an nth order system is characterized by its minimum phase and by the slope of its Bode plot:

for $\omega \to \infty$

$$\sphericalangle G(i\omega) \to -n \cdot 90° \tag{5.182}$$

$$\frac{\mathrm{d} \log |G(i\omega)|}{\mathrm{d} \log \omega} \to -n \tag{5.183}$$

The Nyquist curve of an nth order system passes through n quadrants.

The time constants of the system can be determined from the amplitude-frequency diagram of the Bode plot, by drawing tangents with slopes 0, -1, -2, -3, etc. The method, which is shown in *Fig. 5.53*, is easy to understand if one bears in mind how the Bode plot of a higher-order system is constructed from its 1st and

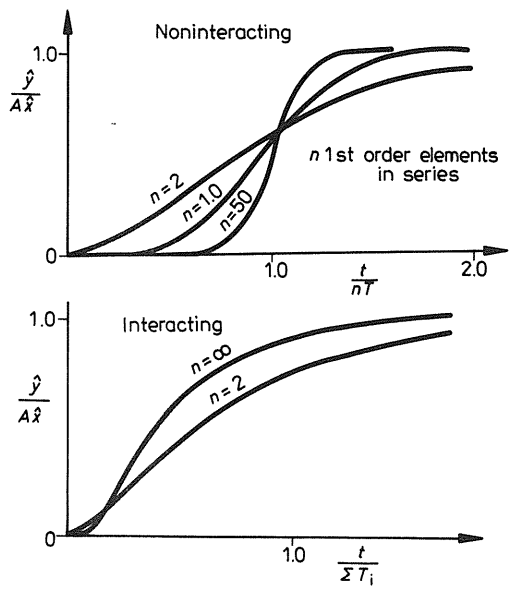

Fig. 5.52. Step responses of interacting and of noninteracting systems

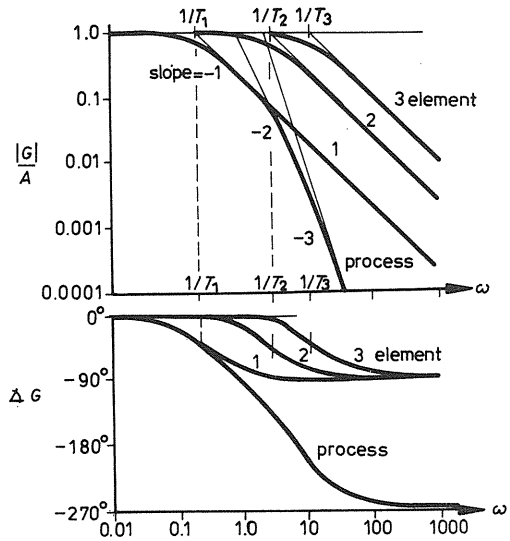

Fig. 5.53. Identification of time constants of a multicapacity system

2nd order elements. Recall that

$$|G(i\omega)| = \prod_i |G(i\omega)|\,, \qquad \text{thus}$$

$$\log|G| = \sum_i \log|G_i|\,,$$

and

$$\measuredangle G = \sum \measuredangle G_i\,.$$

Dead time

Dead time (or true time delay) systems have already been treated in detail in Section 5.2.2.

The differential equation method cannot be applied to dead time systems because they correspond to infinite-order systems.

The transfer function and the frequency function methods are applicable, however. The transfer function is obtained directly from the equation

$$y[t] = A_S x[t - T_D] \tag{5.184}$$

where $y[t]$ is output at moment t,
A_S is steady-state gain of the process, and
$x[t-T_D]$ is input at moment $(t-T_D)$.
The Laplace transform is

$$Y(s) = A_S X \mathcal{L}[x(t-T_D)] = A_S X(s) e^{-sT_D} \tag{5.185}$$

using the time shift theorem from *Table 5.3*. Thus the transfer function is:

$$G(s)=\frac{Y(s)}{X(s)}=A_S e^{-sT_D} \tag{5.186}$$

and the frequency function by substituting $s=i\omega$:

$$G(i\omega)=A_S e^{-i\omega T_D} \tag{5.187}$$

Recall the exponential form of the complex number [cf. Eq.(5.9)] thus

$$|G(i\omega)|=A_S \quad i\measuredangle G(i\omega)=-\omega T_D$$

as was found in Section 5.2.2.

When $\omega\to\infty, \qquad \measuredangle G(i\omega)\to-\infty$

This means that the dead time element is a non-minimum-phase element.

The effect of non-minimum-phase behaviour on the quality of control has been shown in Section 5.2.2. Thus it is important to identify dead time in processes.

Pure dead time is easily identified from typical responses *(Fig. 5.45)*. If dead time is in series with capacities (single or multicapacity systems), pure dead time and apparent dead time in a step response are hard to distinguish from each other.

For purposes of controller setting, capacity systems with or without pure dead time are generally approximated by a dead time–single capacity or a dead time–two capacity model. The determination of the parameters of the dead time–single capacity model from the step response is shown in *Fig. 5.54*. The approximate transfer function is

$$G(s)=\frac{A_S e^{-T_D s}}{T_1 s+1} \tag{5.188}$$

where T_D is the effective dead time,
T_1 is the effective time constant, and
A_S is the steady-state process gain.

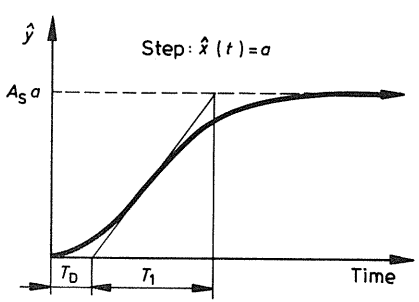

Fig. 5.54. Approximation of multicapacity step response

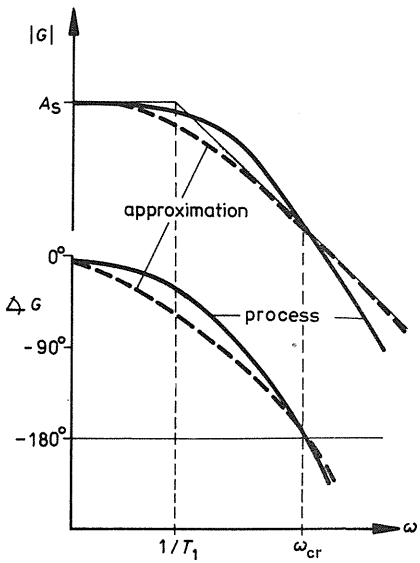

Fig. 5.55. Approximation of multicapacity process by first-order lag with dead time, Bode plot

The frequency function of the approximate model should fit that of the original process in the vicinity of the critical frequency, since it is used to determine controller settings *(Fig. 5.55)*.

Distributed parameter systems

So far we have discussed lumped parameter systems consisting of completely mixed units where the process variable is a function of time only. These processes are described by ordinary differential equations.

In distributed parameter systems the process variable is a function both of time and of position. Such systems are described by partial differential equations. The mathematical treatment of such systems is a little different. Applying Laplace transformation, ordinary differential equations are obtained, the solution of which gives the transfer functions directly.

The procedure is shown using a tubular reactor *(Fig. 5.56)* as an example. This is a unidimensional problem with axial symmetry. We suppose the following conditions: plug flow, no density change, no diffusion effects, no thermal effects (isothermal reaction).

The reaction rate $r = -\frac{1}{V}\frac{\mathrm{d}c}{\mathrm{d}t}$ is dependent on concentration, but an average value of $\left(\frac{\partial r}{\partial c}\right)_{t=0}$ can be used over the length of the reactor.

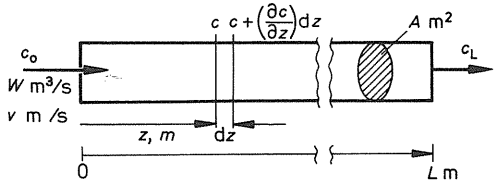

Fig. 5.56. Tubular reactor

The component balance is written for the volume element $dV = A\,dz$ of the reactor of lenght L.

In the steady state

$$\text{input} = \text{output} + \text{consumed}$$

$$W\bar{c} = W\left(\bar{c} + \left(\frac{\partial c}{\partial z}\right)dz\right) + \bar{r}\,dV \tag{5.189}$$

In the non-steady state

$$\text{input} = \text{output} + \text{consumed} + \text{accumulated}$$

$$Wc = W\left(c + \left(\frac{\partial c}{\partial z}\right)dz\right) + r\,dV + V\frac{\partial c}{\partial t} \tag{5.190}$$

The perturbation variables are defined as:

$$\hat{c}(z, t) = c(z, t) - \bar{c}(z, 0) \tag{5.191}$$

$$\hat{r}(z, t) = r(z, t) - \bar{r}(z, 0) \tag{5.192}$$

$\hat{r}(z, t)$ is function of $\hat{c}$

$$\hat{r}(z, t) = \left(\frac{\partial r}{\partial c}\right)_0 \hat{c}(z, t)\,. \tag{5.193}$$

By subtracting Eq.(5.189) from Eq.(5.190) and substituting Eq.(5.193) and $dV = A\,dz$, the linearized partial differential equation in perturbation variables is obtained:

$$0 = W\frac{\partial c}{\partial z}dz + A\left(\frac{\partial r}{\partial c}\right)_0 \hat{c}\,dz + A\frac{\partial \hat{c}}{\partial t}dz \tag{5.194}$$

giving:

$$-v\frac{\partial \hat{c}}{\partial z} = \left(\frac{\partial r}{\partial c}\right)_0 \hat{c} + \frac{\partial \hat{c}}{\partial t} \tag{5.195}$$

Apply Laplace transformation:

$$-v\frac{dC(s)}{dz} = \left(\frac{\partial r}{\partial c}\right)_0 C(s) + sC(s) \tag{5.196}$$

Separate the variables and integrate between the limits $z=0$ and $z=L$:

$$\int_{C_0(s)}^{C_L(s)} \frac{dC(s)}{\left[\left(\frac{\partial r}{\partial c}\right)_0 + s\right] C(s)} = -\int_0^L \frac{dz}{v} \tag{5.197}$$

$$\frac{1}{\left(\frac{\partial r}{\partial c}\right)_0 + s} \ln \frac{C_L(s)}{C_0(s)} = \frac{L}{v} = -T_D \tag{5.198}$$

then the transfer function from output c_L to input c_0 is

$$\frac{C_L(s)}{C_0(s)} = e^{-T_D\left(\frac{\partial r}{\partial c}\right)_0} \cdot e^{-T_D s} = A_S e^{-T_D s} \tag{5.199}$$

Thus a plug flow tubular reactor with no diffusion effects corresponds to a pure time delay, i.e. $n \to \infty$ perfectly-mixed volume elements. If the velocity is not uniform in the cross-section (e.g. the flow is laminar or turbulent, diffusion and/or back-mixing exist), the tubular reactor corresponds to a multicapacity system, the order of which depends on the mixing conditions.

The transfer and frequency functions of distributed parameter systems may also be approximated by first-order dead time models [Eq.(5.188)], or second-order dead time models.

5.3.7 Control loop behaviour

Open loop transfer and frequency function

The open loop frequency function is important for checking the stability of the closed control loop. The method has been introduced in Section 5.2.1 and was applied in Section 5.2.2 for dead time process and in Section 5.2.3 for pure capacity or integrating process. It is based on the simplified Nyquist stability criterion [see Eq.(5.12)]:

$$L(i\omega) = G_C(i\omega) G_S(i\omega) = -1 \tag{5.200}$$

or

$$\measuredangle L(i\omega) = \measuredangle G_C(i\omega) + \measuredangle G_S(i\omega) = -180° \tag{5.201}$$

and

$$\begin{aligned} |L(i\omega)| = 1 = |G_C(i\omega)|\,|G_S(i\omega)| = \\ = A_C |g_C(i\omega)|\, A_S |g_S(i\omega)| = \\ = K |g_C(i\omega)|\,|g_S(i\omega)| \end{aligned} \tag{5.202}$$

where $L(i\omega)$ is the open loop frequency function,

$\omega = \omega_{cr}$ is the critical frequency, if Eq.(5.201) is satisfied,

$K = A_C A_S$ is the loop gain,

$K = K_{max}$ and $A_C = A_{C,max}$ is the critical gain, if Eq.(5.202) is satisfied.

Open loop frequency function with P-controller

The *P*-controller (i.e. the ideal *P*-controller) has no phase lag at any frequency, thus it does not modify the phase-frequency function of the controlled process. This characteristic offers the possibility of determining the critical frequency of the process alone (called natural or ultimate frequency), i.e. the frequency at which

$$\measuredangle G_S(i\omega_0) = -180° \tag{5.203}$$

where ω_0 is the natural frequency.

For processes with known parameters, the frequency function of the process is calculated, plotted on a Bode diagram, and the natural frequency is simply read off.

Continuous cycling of the signals in a control loop occurs if Eq.(5.200) is satisfied:

$$A_C A_S |g(i\omega_0)| = 1 \tag{5.204}$$

i.e.

$$A_{C,\max} = \frac{1}{A_S |g_S(i\omega_0)|} \quad \text{or}$$
$$K_{\max} = \frac{1}{|g_S i\omega_0|} \tag{5.205}$$

$A_{C,\max}$ is read simply from the Bode plot of the process.

For industrial processes with unknown parameters $A_{C,\max}$ and ω_0 are determined experimentally by trial. The loop is closed through a *P*-controller and the gain of the controller is changed until continuous cycling of the process output (or the manipulated variable) occurs in response to a step input. The frequency of the output signal is ω_0, the natural frequency of the process, the gain set at the *P*-controller is $A_{C,\max}$, the ultimate gain. This method has been described in detail in Section 5.2.6.

The method fails for some processes: even with the controller set to its highest proportional gain, continuous cycling cannot be achieved. The reason for this is explained below.

The pure capacity, or integrating process with the transfer function

$$G_S(s) = \frac{A_S}{s} \tag{5.206}$$

has a phase lag of $\measuredangle G_S = -90°$, independent of the frequency of the input signal. As has already been discussed in Section 5.2.3, this process can be controlled safely by a *P*-controller set to the largest possible gain, since the phase shift cannot cross $-180°$. The first-order lag process with the transfer function

$$G_S(s) = \frac{A_S}{Ts+1} \tag{5.207}$$

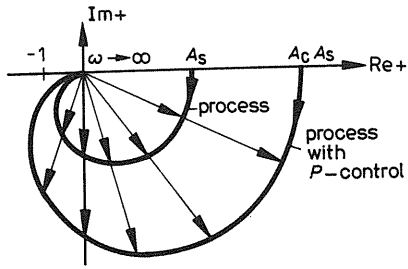

Fig. 5.57. Second-order process with P-controller is always stable

exhibits a phase lag (cf. *Table 5.3*)

$$\measuredangle G_S(i\omega) = \arctan(-\omega T)$$

For $\omega \to \infty$, the minimum phase $\measuredangle G_S \to -90°$, which is the same situation as was encountered with the pure capacity process.

A second-order lag process with the transfer function

$$G_S(s) = \frac{A_S}{T^2 s^2 + 2\xi T s + 1} \tag{5.208}$$

exhibits a phase lag (see Section 5.3.6)

$$\measuredangle G_S(i\omega) = \arctan\left(-\frac{2\xi T}{1-\omega^2 T^2}\right)$$

which has a minimum for $\omega \to \infty$: $\measuredangle G_S \to -2 \cdot 90°$. At the same time the dynamic gain of the process approaches zero:

$$\text{for} \quad \omega \to \infty \qquad |G_S| \to \frac{A_S}{\omega^2 T^2}$$

The situation is best visualized in the Nyquist diagram, *Fig. 5.57*. It is evident that a second-order process may be safely controlled by a P-controller set to the highest possible gain; oscillation will always be damped.

Processes with dead time or processes higher than second-order can be brought to continuous cycling with a P-controller.

Open loop frequency function with PID-controller

Any combination of controllers other than proportional ones will substantially modify the open loop frequency function as compared with that of the process.

In *Fig. 5.58* the Bode plot of a process is shown, together with that of a *PID*-controller. It is evident that the critical frequency of the open loop will be closest to the natural frequency of the process if the I and D parameters of the

controller are set so that

$$\frac{1}{I} < \omega_0 < \frac{1}{D} \tag{5.209}$$

At the same time this choice of parameters will also have less effect on the dynamic gain of the loop, in comparison with that obtained with a simple *P*-controller. If the Bode plot of the *PID*-controller were to be moved towards higher frequencies, i.e. if $1/I$ were increased, this would add to the phase lag of the open loop, decreasing the critical frequency. This implies a slower control loop. Simultaneously, the frequency-dependent gain of the *I*-controller part would add to the dynamic gain of the process, eventually causing instability with the proportional gain set to a level that may safely be used with *P*-control only.

When the Bode plot of the *PID*-controller is moved towards smaller frequencies, i.e. when $1/D$ is diminished, the *D*-part of the controller is effective in diminishing the phase lag of the open loop around ω_0, and thus it speeds up control. At the same time frequency dependent gain of the *D*-controller part adds to the process gain, and this can eventually cancel the effect of the phase lead. This depends, of course, on the slope of the amplitude-frequency function of the process.

The above considerations are also readily applicable to *PI*- and *PD*-controller combinations.

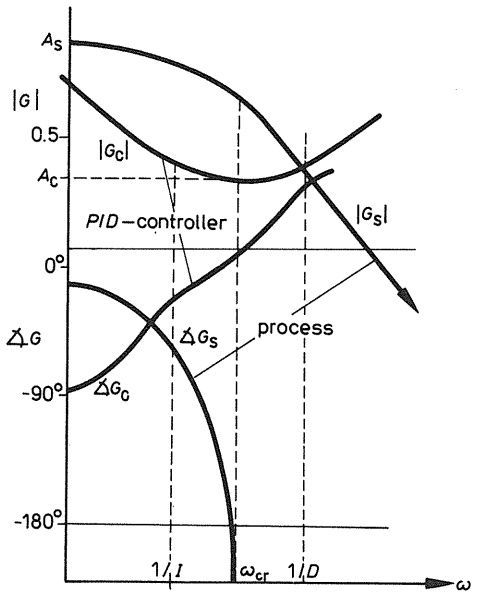

Fig. 5.58. Bode plot of process with *PID*-controller

Gain and phase margins

Controllers are, of course, never set to the stability limit defined by Eq.(5.200). In Section 5.2.1 we have seen that if quarter-amplitude damping is aimed for, the controller must be set to obtain a loop gain of 0.5:

$$|L(i\omega_{cr})| = A_S A_C |g_S(i\omega_{cr})|\,|g_C(i\omega_{cr})| = 0.5$$

This means a gain margin of $1/0.5 = 2.0$.

The definition of the gain margin is

$$\text{gain margin} = \frac{1}{|L(i\omega_{cr})|} \tag{5.210}$$

and it expresses the distance from the stability limit.

A similar value is needed for the phase lag of the open loop. This is found in the phase margin. Denote by $\varphi(L=1)$ the open loop phase lag at the frequency where $|L(i\omega)| = 1$, that is, at the gain crossover frequency. Then

$$\text{phase margin} = 180° - \varphi(L=1) \tag{5.211}$$

The phase margin is the phase shift which can be added to the phase shift of the open control loop to arrive at the stability limit of the closed control loop (at the gain crossover frequency).

The gain and the phase margins are easily read from the Bode and Nyquist plots of the open loop *(Fig. 5.59)*.

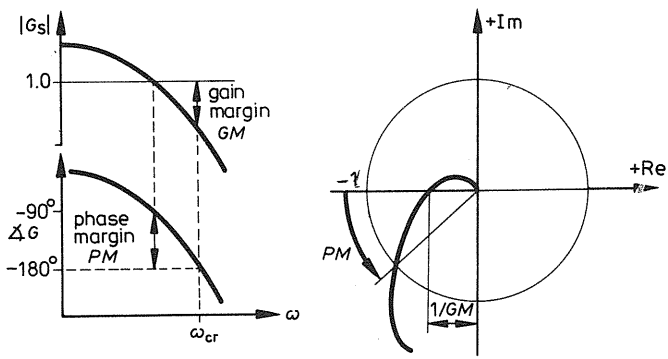

Fig. 5.59. Gain and phase margins

Both the gain and the phase margins account not only for the damping, but also for uncertainties in the knowledge of the process parameters. It is important to draw attention to the fact that most chemical engineering processes are nonlinear: their gain and also their time constants (including dead time) might be dependent

on the operating point. Gain variations in the range of 1:10 are also possible. The gain and phase margins (i.e. the controller parameters) must be set in consideration of these facts.

The Nyquist stability criterion

Hitherto the simplified Nyquist stability criterion formulated by Eq.(5.200) was used, which is readily applicable to monotonic open loop frequency functions. The rigorous Nyquist criterion is used to solve more complex problems.

The simplest example for which the rigorous Nyquist criterion has to be applied is the pure capacity process controlled by a *PI*-controller. The Bode and Nyquist plots of the open loop are shown in *Fig. 5.60.* For $\omega \to 0$, the loop phase lag is $-180°$ and the loop gain is infinite: there are two integrators in the loop. With increasing frequency the phase lag decreases towards $-90°$ and the loop gain decreases towards zero. The stability of this control loop can be checked by drawing the Nyquist plot. Following the plot from $\omega = 0$ towards $\omega = \infty$, the point $(-1, 0_{-j})$ is seen to be at the left, and thus the control loop is stable.

Conditional stability may occur in the case of chemical reactors in which exothermic reaction occurs. Some of these reactors are unstable without control. Continuous stirred tank reactors are first-order lag processes; at their unstable operating point both the steady-state gain and the time constant are negative. The transfer function is:

$$G(s) = \frac{-A_S}{-Ts+1} = \frac{A_S}{Ts-1}$$

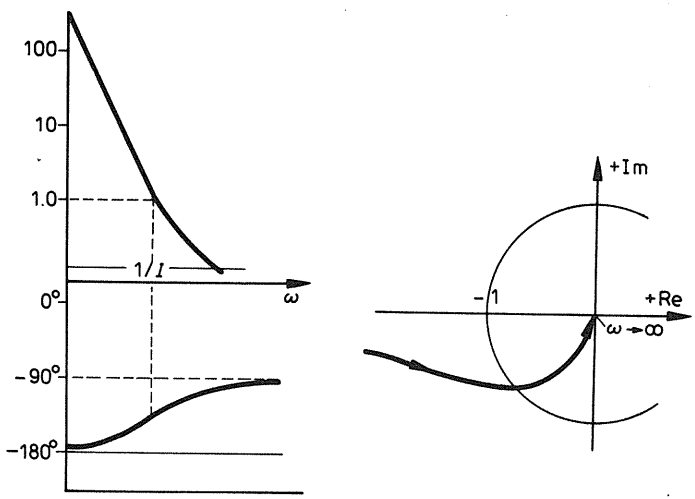

Fig. 5.60. *I* process controlled by *PI*-controller

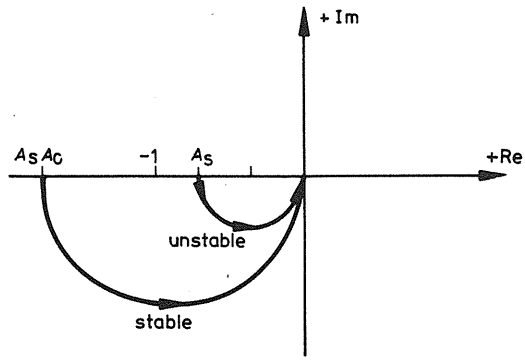

Fig. 5.61. Stabilizing an unstable reactor

This process was dealt with in Section 5.3.4, Example 5.3.8. The Nyquist plot of such a frequency function is shown in *Fig. 5.61*. It begins at point $(-A_S, 0)$ and ends at (0, 0). For a stable control loop $|A_S A_C| > 1$ is to be set, then the point $(-1, 0)$ is seen to be at the left. It is unusual that the controller gain must be increased to ensure stable operation.

When there are other dynamic elements in the control loop besides the reactor, the controller gain may have a high and a low limit for stable operation. Suppose that there are two first-order elements (e.g. the temperature sensor and the actuator of the control valve) both having smaller time constants than that of the reactor (of course, in absolute value). The transfer function of the process is:

$$G_S(s) = \left(\frac{-A_R}{-T_R s + 1}\right)\left(\frac{A_T}{T_T s + 1}\right)\left(\frac{A_V}{T_V s + 1}\right)$$

The subscripts refer to the reactor, the temperature transmitter and the valve. The Nyquist plot of the open loop frequency function of this process with *P*-control is shown in *Fig. 5.62*. The two additional first-order elements shift the end of the plot by $2 \cdot (-90°)$ to $-270°$. The control loop is unstable both with too high or too low controller gains.

Similar problems may occur with the control of multivariable systems.

The closed loop transfer and frequency function

The open loop transfer and frequency functions are obtained by multiplying the respective functions of all dynamic elements in the control loop. The open loop functions are not to be confused with those of the closed loop.

The frequency function of a negative feedback control loop is given in Eq.(5.4). Since every control loop has at least two inputs the set point (or command variable) and one load variable, there are at least two closed loop frequency functions:

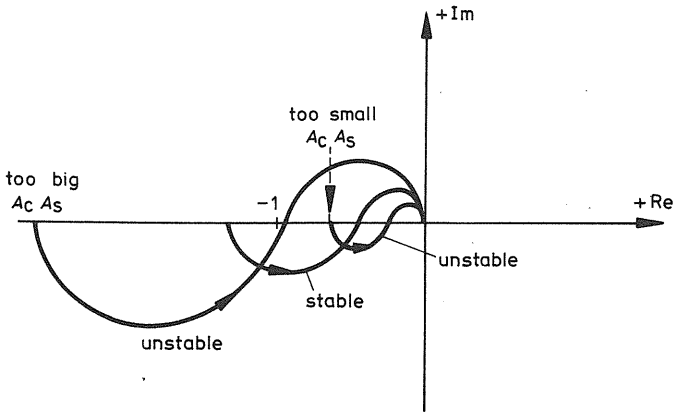

Fig. 5.62. Stability of control loop with unstable reactor and two small lags

for the set point:
$$W(i\omega)=\frac{G_S(i\omega)G_C(i\omega)}{1+G_S(i\omega)G_C(i\omega)} \tag{5.212}$$

for the load:
$$W_d(i\omega)=\frac{G_d(i\omega)}{1+G_S(i\omega)G_C(i\omega)} \tag{5.213}$$

The closed loop frequency functions are denoted by W to distinguish them from the corresponding open loop functions. All closed loop frequency functions of the same control loop have the same denominator: $1+L(i\omega)$, where the frequency dependent vector $L(i\omega)$ is to be added to the unit vector, which has zero phase angle. The closed loop frequency functions can easily be calculated using complex algebra, if the frequency functions of the process and the controller are known.

The closed loop frequency functions can be determined experimentally by driving the closed loop by a sine signal (varying the frequency) at the set point or at the load input, respectively, and recording the control signal, but this method is rather tedious and is rarely applied to chemical engineering processes.

The closed loop frequency functions, of course, are characteristic of the behaviour of the control loop. Since we require that the control loop causes the control signal to follow the set point signal exactly, with no phase shift at all frequencies, the required frequency function is

$$|W(i\omega)|=1, \qquad \measuredangle W(i\omega)=0^\circ \qquad \text{for all} \quad \omega.$$

In reality this is possible only for frequencies that are smaller than the open loop critical frequency. There is a maximum gain at the closed loop resonant frequency, which is a little lower than the critical frequency of the open loop. This situation is shown in *Fig. 5.63*.

At the stability limit, the maximum closed loop gain is infinite and the resonance frequency is equal to the open loop critical frequency.

The closed loop frequency functions may be used for the determination of controller settings. The load frequency function, Eq.(5.213) is less often used for this purpose, since it requires an additional knowledge of the frequency function $G_d(i\omega)$, but it does not furnish substantially more information.

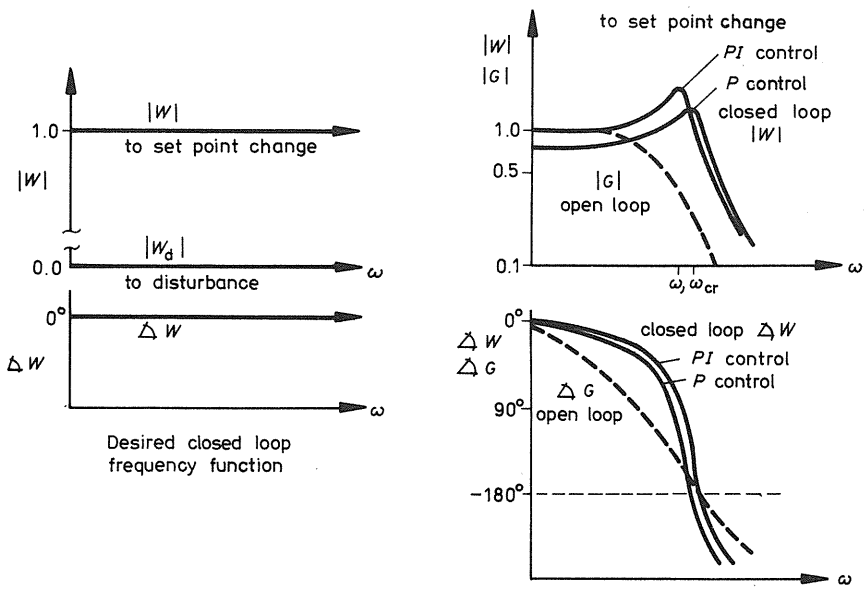

Fig. 5.63. Bode plots of closed control loop frequency functions

Steady-state offset. Final value theorem

We have seen in Section 5.2.2 that, if a dead time process is controlled by a *P*-controller, the control loop is unable to restore the control signal to its original value after a step-like disturbance of the load variable, and it is also unable to make the control signal exactly follow a change in the set point. In other words, the error signal is different from zero, not only during a transient, but also in the steady state. This steady-state offset forced us to incorporate integral action in the controller. However, we did not examine the problem of steady-state error for the general case.

The final steady-state value of any linear time function can be determined using the final value theorem, Item 5 in *Table 5.3*, if the Laplace transform of the function is known:

$$f(\infty)=\lim_{s\to 0} sF(s) \tag{5.214}$$

It is important to check the stability of the function, before applying the final value theorem, because it gives no information about stability.

The Laplace transform of c, the control variable for the closed negative feedback control loop, is calculated using Eqs (5.100) and (5.101) in Section 5.3.4, respectively:

for a set point change

$$C(s)=\left[\frac{G_C(s)G_S(s)}{1+G_C(s)G_S(s)}\right]R(s) \tag{5.215}$$

for a disturbance

$$C(s)=\left[\frac{G_D(s)}{1+G_C(s)G_S(s)}\right]D(s) \tag{5.216}$$

Tracking the set point

If the control signal follows the command, the error is zero (cf. *Fig. 5.39*):

$$e=c_{set}-c=r-c \tag{5.217}$$

Combining Eq.(5.215) with the Laplace transform of Eq.(5.217), we obtain the Laplace transform of the error function:

$$E(s)=R(s)-\left(\frac{G_C(s)G_S(s)}{1+G_C(s)G_S(s)}\right)R(s)$$

$$=\left(\frac{1}{1+G_C(s)G_S(s)}\right)R(s) \tag{5.218}$$

Suppose a unit step in the set point: $R(s)=\frac{1}{s}$, then the final value of the error is [Eq.(5.214)]

$$e(\infty)=\lim_{s\to 0} s\left(\frac{G_C(s)G_S(s)}{1+G_C(s)G_S(s)}\right)\frac{1}{s}=\lim_{s\to 0}\frac{G_C(s)G_S(s)}{1+G_C(s)G_S(s)} \tag{5.219}$$

To obtain $e(\infty)=0$, the product $G_C(s)G_S(s)$ must go to infinity if s goes to zero. This means that either the controller or the process must contain at least one integrator (the transfer function of which is A/s, cf. *Table 5.2*).

In the absence of an integrator, the controlled variable follows the set point with some error. The magnitude of the error is calculated from Eq.(5.219). To facilitate the calculation, consider that

$$\lim_{s\to 0} e^{-T_D s}=1$$

$$\lim_{s\to 0} \frac{1}{(T_1 s+1)^n (T_2 s+1)\,(T^2 s^2+2\xi Ts+1)}=1$$

$$\lim_{s\to 0} \frac{T_1 s+1}{T_2 s+1}=1$$

Thus, if the process is a higher-order lag, possibly also with dead time, but contains no integrating element, then

$$\lim_{s\to 0} A_S g_S(s)=A_S$$

If the controller is a *P*-controller

$$G_C=A_C$$

or a *PD*-controller:

$$G_C=A_C(1+Ds)$$

then

$$\lim_{s\to 0} A_C g_C(s)=A_C$$

Substitute into Eq.(5.219):

$$e(\infty)=\lim_{s\to 0} s\left(\frac{1}{1+A_C g_C(s) A_S g_S(s)}\right)\frac{1}{s}=\frac{1}{1+A_C A_S} \tag{5.220}$$

Equation (5.220) indicates that a *P*- or a *PD*-controller may be used to control such processes, for which a small steady-state error can be tolerated and the process stability allows the use of a high-proportional gain. Recall from Eq.(5.205) that

$$(A_C A_S)_{max}=\frac{1}{g_C(i\omega_{cr})g_S(i\omega_{cr})} \tag{5.221}$$

For the reader's orientation we may say that to a unit step a steady-state error $e(\infty)=0.05$, i.e. 5% of the step point change, is generally acceptable.

Disturbance rejection

The controller is expected to restore the controlled variable to the same steady-state value as it had before the disturbance: $\hat{c}(\infty)=0$.

Using Eq.(5.216), in response to a unit step disturbance: $D(s)=\frac{1}{s}$, we obtain

$$\hat{c}(\infty)=\lim_{s\to 0} s\left[\frac{G_d(s)}{1+G_C(s)G_S(s)}\right]\frac{1}{s}=\lim_{s\to 0}\frac{G_d(s)}{1+G_C(s)G_S(s)} \tag{5.222}$$

The following situations are possible:

1. $G_C(s)$ implies a *P*- or *PD*-controller. $G_S(s)$ contains a term $1/s$, but $G_d(s)$ does not. Then $\hat{c}(\infty)=0$.
2. $G_C(s)$ implies a *P*- or *PD*-controller. Both $G_D(s)$ and $G_S(s)$ are integrating processes containing a term $1/s$. Then $\hat{c}(\infty)=A_d/A_C A_S$.
3. $G_C(s)$ implies a *PI*- or *PID*-controller, containing a $1/s$ term. Then $\hat{c}(\infty)=0$, except in the case when $G_d(s)$ means an integrating process, but $G_S(s)$ not.

Thus the steady-state error for both command signal changes and disturbances is generally eliminated by using an integrating mode in the controller, and is diminished by the use of high loop gain, $K=A_C A_S$.

5.3.8 Controller tuning

Control performance criteria

In order to determine optimum controller settings we need some quantitative criteria to compare different control systems.

The performance criteria are not only used to determine the *P*, *I*, *D* parameters of the controller. Hitherto we have supposed that the parameters of the controlled process are given. In reality, even in the case of an existing industrial process, it is possible to choose different types of sensors and final control elements, to decide their placement, and thus, to influence the parameters of the process. As an example: the product composition of a distillation column may be controlled by (*a*) measuring the boiling temperature continuously at an intermediate tray; (*b*) measuring the composition intermittently by a process analyser from ($b\hat{a}$) a continuous vapour sample taken from the column head or ($b\hat{b}$) a continuous liquid sample taken from the reflux accumulator. In case (*b*) there is a dead time incorporated in the intermittent analysis, the effect of which can be evaluated by use of the performance criteria.

The desired performance of a closed-loop system can be specified in the time domain. The traditional test input signal is a step change in the set point. The

performance criteria based upon the response of the control signal are shown in *Fig. 5.64:*

(*i*) overshoot: A/B. This can be estimated from the phase margin of the open loop frequency function, or from the maximum, M, of the closed loop frequency function to the set point as follows:

Phase margin	M	Overshoot
$\leqq 30°$	$\geqq 1.5$	$M-0.1$
30°–40°	1.5–1.3	M
60°	1.0	M

(*ii*) decay ratio or damping: A/C. This can be estimated from the gain margin;

(*iii*) rise time: the time necessary to first reach the new steady-state value;

(*iv*) response time: the time it takes for the amplitude of the oscillations to decay to some fraction (generally $\pm 5\%$) of the final change of the control signal;

(*v*) period of oscillations, T_p. For a well-tuned control loop the response time is about 3 times the period. The period is estimated as the critical period of the open loop.

(*vi*) steady-state offset; estimated by the final value theorem.

The integral performance criteria are based upon the error $e = c_{set} - c$ during the response time *(Fig. 5.64)*. They may be calculated by computer for control loops with no offset or determined by experiment.

The error integral: (5.223)

$$EI = \int_0^\infty e\,dt$$

has the disadvantage of giving zero at the stability limit;

Integrated absolute error: (5.224)

$$IAE = \int_0^\infty |e|\,dt$$

given the total area at both sides of zero error;

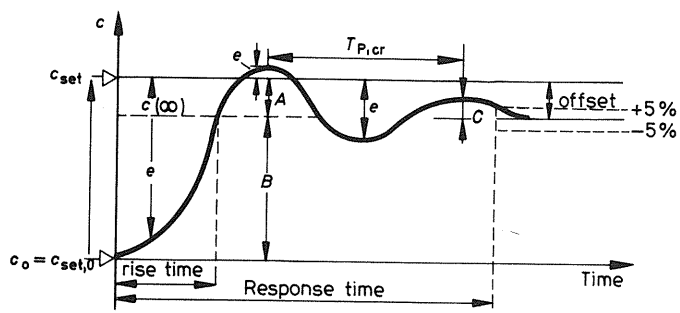

Fig. 5.64. Performance criteria based on closed loop step response

Integrated square error: (5.225)

$$ISE = \int_0^\infty e^2 \, dt$$ confers a penalty on overshoot;

Integrated time-weighted absolute error: (5.226)

$$ITAE = \int_0^\infty |e| t \, dt$$ implies a penalty at long response times.

The optimum controller settings determined by the various criteria are different. Which criterion is to be used depends on the process.

Quick estimation of the error integral

This method can be used for a quick choice between different control systems. The best control system is that for which:

$$K_{cr} \omega_{cr} = \max \tag{5.227}$$

This quick criterion is based upon the step response, *Fig. 5.64* and the error integral, Eq.(5.223). It is supposed that the decay ratios in the control systems to be compared are equal and thus the oscillations are similar, and only the error areas under the first peaks need to be considered.

The area of the first peak is proportional to the product of its height and its length. The length of a peak is half the oscillation period; thus it is inversely proportional to the oscillation frequency, which, in turn, can be roughly estimated as the critical frequency. As a consequence, the error integral is inversely proportional to the critical frequency. It is important to note that when comparing controller settings, the phase shift of the controller is to be included in the calculation of the critical frequency.

The height of the first peak depends only on $K = A_S A_C$, the open loop gain product, which has been demonstrated by experience. Similarly to the steady-state offset (with a *P*-controller only), which is proportional to $1/(1 + A_S A_C)$; cf. Eq.(5.220), the peak height is also proportional to $1/(1 + K)$, not only for *P*-, but also for *PI*-controllers; and this may be also used for *PID*-controllers for rough estimation. For high values of K we can neglect 1 as well as K, and since for a certain decay ratio K/K_{max} must be fixed, we may conclude that the height of the first peak is inversely proportional to K_{max}.

Thus we find that

$$EI \sim \frac{1}{K_{cr} \omega_{cr}}$$

and for the optimum control system

$$EI = \min, \qquad K_{cr} \omega_{cr} = \max$$

Controller selection

To make a decision upon the control mode to be selected for a given process, let us summarize the features of the various control modes:

Proportional mode:

(*i*) produces no phase lag, thus it does not slow down the response of the process;
(*ii*) is unable to eliminate offset, except in the case of an integrating process.

Integral mode:

(*i*) eliminates offset,
(*ii*) exhibits phase lag (max. $-90°$) as well as frequency dependent gain, thus it has a retarding and a destabilizing effect.

Derivative mode:

(*i*) unable to eliminate error as unique control mode;
(*ii*) produces phase lead (max. $+60°$ for real controllers), thus speeds up control and has a stabilizing effect.

A three-mode *PID*-controller seems to be the best since it confers all kinds of benefits. But, at the same time, it presents a more complex tuning problem, which is hard to overcome, particularly with the nonlinear processes frequently encountered in chemical engineering. Recall that the *PID* parameters ought to be tuned to have $\frac{1}{I} < \omega_0 < \frac{1}{D}$, which is hard to obtain if ω_0 changes with the operating point.

P-controller

This is the cheapest device and is the simplest to tune. It is used for control problems where offset is allowed or even desirable. An example of the latter is the level control in a buffer tank used to average out inflow fluctuations before a process which requires constant feed.

P-control may also be used for processes where offset is undesirable, but whose characteristics allow the use of a very high loop gain, K. Such processes are: single-capacity processes (integrating or first-order lag), which have a $-90°$ minimum phase; and second-order lag processes, with $-180°$ minimum phase.

It ought to be noted here, that control loops containing only one or two capacities rarely exist, since the sensing instrument, the signal transformer and the final control element also represent time lags. These types of device may be considered as first-order elements, their time constants ranging between 0.2 to 30 s. But, if the control loop comprises one capacity with a dominant time constant at least one order of magnitude greater than the second greatest, then the process may

be treated as a single capacity process and it may be controlled by a *P*-controller set to high loop gain, *K*. Such processes are: liquid tanks, stirred tank reactors, gas pressure controllers.

Example 5.3.11. *Control of a process with dominant time constant.*

Consider a liquid vessel having a time constant of 30 min. The time constant of the level transmitter, the pneumatic signal transmission line and the pneumatic control valve actuator sum together to 1 min. Although the transmitter and the valve actuator are first-order elements, the sum of the smaller time constants may be treated as pure dead time. This is an approximation, but it is on the safe side, since the dead time is less advantageous to control than a multicapacity system.

The natural period of the process is calculated as that of a single capacity, dead time process. Suppose the dominant time constant exhibits a $-90°$ phase lag at the natural period, then (cf. *Table 5.3*)

$$-\omega_0 T_D - 90° = -180°, \qquad \omega_0 = \frac{90°}{T_D}$$

$$T_{P,0} = \frac{2\pi}{\omega_0} = 4T_D = 4 \text{ min}$$

Verify:

$$\varphi(T = 30 \text{ min}) = \arctan(-\omega_0 T) =$$

$$= \arctan\left(-\frac{2\pi}{4} \cdot 30\right) = -98.6°$$

At $T_{P,0} = 4$ min period, the dynamic gain is:

for the dead time: 1

for the first-order lag:

$$|g(i\omega)| = \frac{1}{\sqrt{1 + T^2\omega_0^2}} = \frac{1}{\sqrt{1 + \left(30\frac{2\pi}{4}\right)^2}} = 0.021$$

The maximum loop gain is

$$K_{max} = \frac{1}{|g_S(i\omega_0)|} = \frac{1}{0.021} = 47.6$$

To obtain quarter amplitude damping, the loop gain is set to

$$K = 23.8$$

The error to a unit step change in the set point is

$$e(\infty) = \frac{1}{1 + K} = 0.04, \text{ acceptable}$$

PI-controller

This is used when offset is not tolerable and the process has no dominating time constant. A *PI*-controller may also be used for processes dominated by dead time. It has the disadvantage of rendering the control more sluggish due to the time lag of the *I*-controller.

PID-controller

The *D* action produces a phase lag of about 60° (in a real controller). This almost cancels the phase lag of one capacity (the minimum phase is $-90°$). Thus it may decrease loop phase lag and increase the critical frequency in the case of processes consisting of three or four capacities, and speed up control substantially. *D* action is of no help in the case of processes dominated by dead time, because a dead time element corresponds to an infinite capacity system, of which only one is cancelled by *D* action.

D action cannot be applied to processes with substantial measurement noise, e.g. flow control.

Controller tuning

Several tuning tables giving controller settings are offered in handbooks. These are based upon some performance criteria applied to an appropriate process model. The simplest process model is a first-order lag with dead time, with the transfer function

$$G(s) = \frac{A_S e^{-T_D s}}{Ts+1} \tag{5.228}$$

The Ziegler—Nichols and the Cohen—Coon tuning tables are based upon this model and give controller settings for quarter-amplitude damping. The Cohen—Coon settings were determined using the *ISE* criterion.

The parameters of the process model are determined from the process reaction curve, as shown in *Fig. 5.65*. The process reaction curve is the open loop response

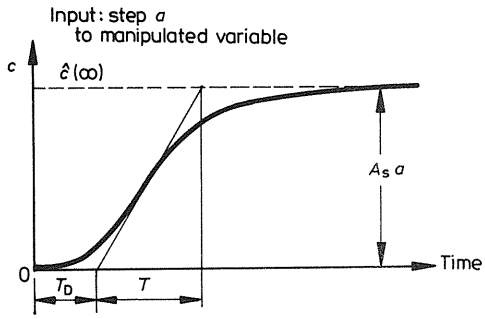

Fig. 5.65. Process reaction curve

of the controlled variable to a unit step input in the manipulated variable. All elements of the control loop must be included in the response, except the controller.

The determination of the parameters can be seen in *Fig. 5.65*. A tangent is drawn to the inflection point of the process reaction curve. The model dead time is determined as the time at which this tangent intersects the base line. The time at which the tangent intersects the 100% response line is the sum of the model time constants, $T+T_D$. The model gain factor, A_S, is, of course, the 100% response to unit input.

There are some difficulties in taking the process reaction curve. It may take a long time to obtain 100% response. It is difficult to draw the tangent line accurately. Thus the controller settings so obtained will be inaccurate, too.

The model parameters, critical period $T_{P,cr}$ and critical gain $A_{C,cr}$, may be determined more exactly. The experimental determination of these parameters involves the cycling method described in Section 5.2.6. They may also be determined from the open loop frequency function of the process, cf. Section 5.3.7. The controller settings based on these parameters are also given in *Table 5.1*.

Adaptive and optimal control

Most chemical engineering systems are nonlinear and if their operating point is changed, their gain factors and time constants change too. This may cause instability in a well-tuned control loop, and make it necessary to re-tune control loops frequently.

This difficulty may be overcome by adaptive control. Adaptive control means automatic adaption of the controller settings to the varying process parameters. Its simplest form is gain adaptive control, when only the controller gain is varied. Since process parameter changes are mostly connected with changes in the flow rate through the process, gain adaptive control may be implemented by measuring the position of the control valve which manipulates the flow rate, and feeding the correlation between valve position and controller gain into a microprocessor controller or into a computer, which adjusts the controller settings.

Self-tuning controllers are able to set all their tuning parameters by the use of some criterion, e.g. quarter-amplitude damping or the integral performance criterion. Adaptive control is best implemented in the case of direct digital control, i.e. when the controller functions are performed by a digital computer.

Optimal control may also be performed by computer control, but in this case by using supervisory control. In this case the computer furnishes the set point (command signal) to the controller, having determined the optimum operating point of the controlled process. The controller accepting the set point and controlling the process may be a digital or a conventional analogue controller.

5.4 The state space method

5.4.1 The state of a system

The state of any system means the past, present, and the future of the system. The state space method gives all this information. It is not limited to linear systems: it is capable of describing nonlinear and variable parameter systems, too. Using the state space method it is possible to investigate the system in the time domain, since it depends on the past history of the system. The classical linear methods are capable of describing linear systems especially in the frequency domain, and they neglect initial conditions.

The state space method uses the so-called state variables and state equations to determine the state and the dynamic behaviour of the system. The state variables are not always the output variables of the system and they cannot always be measured or observed. The set of the state variables $x_1(t)$, $x_2(t)$, ..., $x_n(t)$ must satisfy the following two conditions:

1. At any entirely randomly selected instant of time ($t=t_0$) the state variables $[x_1(t_0), x_2(t_0), \ldots, x_n(t_0)]$ describe the initial condition of the system $[x_1(t_0), x_2(t_0), \ldots, x_n(t_0)]$.
2. The inputs of the system have to be specified for all $t \geqq t_0$, then the state variables completely determine the behaviour, the state of the system at any time t.

Example 5.4.1

Determine the state variable of a perfectly mixed tank reactor using the first-order reaction equation and its state equation *(Fig. 5.66)*.

The well-known differential equation of such a system is:

$$T\frac{dc_1}{dt} + c_1 = A_S c_0 \tag{5.229}$$

The concentration in the reactor (c_1) completely specifies the state of the system, so it is the state variable (x).

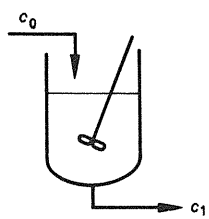

Fig. 5.66. State variables of a multivariable system

Laplace transform and solve the equation (c_i cannot be neglected at $t=0$, $c_i(0)$ is a constant):

$$A_S c_0(s) = c_i(s) + T[c_i(s)s - c_i(0)]. \tag{5.230}$$

Rearranging:

$$c_i(s) = \frac{A_S c_0(s)}{1+Ts} + \frac{Tc_i(0)}{1+Ts} \tag{5.231}$$

Inverse transform Eq.(5.231) in the case of step disturbance of c_0 at $t \geq 0$:

$$c_i(t) = A_S c_0\left(1 - e^{-\frac{t}{T}}\right) + c_i(0)e^{-\frac{t}{T}} \tag{5.232}$$

It is clear that if $c_i(0)$ and the input at $t \geq 0$ are known, the state of the system is determined. Thus c_i satisfies the basic requirements for a state variable. Rearranging Eq.(5.229) one obtains:

$$\frac{dc_i}{dt} = -\frac{1}{T}c_i + \frac{A_S}{T}c_0, \tag{5.233}$$

which is the state equation of a first-order element. It is usual to denote the state variable by x, the input variable by u, and the output variable by y. Generally, the state equation is:

$$\frac{dx(t)}{dt} = ax(t) + bu(t) \tag{5.234}$$

and the so-called output equation is:

$$y(t) = cx(t) + du(t) \tag{5.235}$$

In Example 5.4.1 the output equation is simply:

$$y(t) = x(t) \tag{5.236}$$

5.4.2 State equations of multivariable systems

Consider a linear multivariable system with k inputs and m outputs (see *Fig. 5.67*).

The state variables are $x_1, x_2 \ldots x_n$. The system dynamics are described by n first-order differential equations, the state equations being similar to Eq.(5.234)

$$\frac{dx_i(t)}{dt} = f[x_1(t), x_2(t) \ldots x_n(t), u_1, u_2 \ldots u_k] \tag{5.237}$$

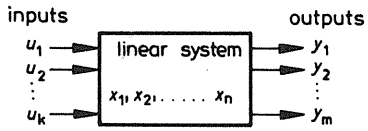

Fig. 5.67. To Example 5.4.1

or

$$\frac{dx_i(t)}{dt} = \sum_{j=1}^{n} a_{i,j}x_j(t) + \sum_{v=1}^{k} b_{i,j}u_v(t) \tag{5.238}$$

The output equations are:

$$y_p(t) = \sum_{j=1}^{n} c_{p,j}x_j(t) + \sum_{v=1}^{m} d_{p,v}u_v(t) \quad \text{where} \quad \begin{matrix} p=1,2\ldots m \\ v=1,2\ldots k \end{matrix} \tag{5.239}$$

It is much more convenient to use the matrix form of representation:

$$\frac{dx(t)}{dt} = \mathbf{A}\underline{x}(t) + \mathbf{B}\underline{u}(t) = f[x(t), u(t)] \tag{5.240}$$

or

$$\dot{x} = \mathbf{A}\underline{x}(t) + \mathbf{B}\underline{u}(t) \tag{5.241}$$

The output equation is

$$\underline{y}(t) = \mathbf{C}\underline{x}(t) + \mathbf{D}\underline{u}(t) \tag{5.242}$$

where

$$\underline{\dot{x}} = \begin{bmatrix} \frac{dx_1(t)}{dt} \\ \frac{dx_2(t)}{dt} \\ \vdots \\ \frac{dx_n(t)}{dt} \end{bmatrix} \qquad \underline{x} = \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{bmatrix}$$

is the column matrix of state variables

$$\underline{u} = \begin{bmatrix} u_1(t) \\ u_2(t) \\ \vdots \\ u_k(t) \end{bmatrix}$$

is the column matrix of input variables

$$\underline{y} = \begin{bmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_m(t) \end{bmatrix}$$

is the column matrix of output variables

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & & & \\ \vdots & & & \\ a_{n1} & \ldots\ldots & & a_{nn} \end{bmatrix}$$

is an $n \times n$ square matrix, the state matrix

$$\mathbf{B} = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1k} \\ b_{21} & & & \\ \vdots & & & \\ b_{n1} & \dots\dots & & b_{nk} \end{bmatrix}$$

is an $n \times k$ matrix, the input matrix

$$\mathbf{C} = \begin{bmatrix} c_{11} & c_{12} & \dots & c_{1n} \\ c_{21} & & & \\ \vdots & & & \\ c_{m1} & \dots\dots & & c_{mn} \end{bmatrix}$$

is an $m \times n$ matrix, the output matrix

$$\mathbf{D} = \begin{bmatrix} d_{11} & d_{12} & \dots & d_{1k} \\ d_{21} & & & \\ \vdots & & & \\ d_{m1} & \dots\dots & & d_{mk} \end{bmatrix}$$

is an $m \times k$ matrix, the transmission

5.4.3 State equations of higher-order elements

Consider an nth order linear system with single input and single output. Its general differential equation is:

$$\frac{d^n y(t)}{dt^n} + a_1 \frac{d^{n-1} y(t)}{dt^{n-1}} + a_2 \frac{d^{n-2} y(t)}{dt^{n-2}} + \dots + a_{n-1} \frac{dy(t)}{dt} + a_n y(t) = u(t) \tag{5.243}$$

The state variables are:

$$\begin{aligned} x_1(t) &= y(t) \\ x_2(t) &= \frac{dy(t)}{dt} \\ &\vdots \\ x_n(t) &= \frac{d^{n-1} y(t)}{dt} \end{aligned} \tag{5.244}$$

and the state equations:

$$\begin{aligned} \frac{dx_1(t)}{dt} &= x_2(t) \\ \frac{dx_2(t)}{dt} &= x_3(t) \\ &\vdots \\ \frac{dx_{n-1}(t)}{dt} &= x_n(t) \end{aligned} \tag{5.245}$$

$$\frac{dx_n(t)}{dt} = -a_n x_1(t) - a_{n-1} x_2(t) - \ldots - a_2 x_{n-1}(t) - \\ - a_1 x_n(t) + u(t)$$

The output equation is very simple:

$$y(t) = x_1(t) \tag{5.246}$$

The matrix representation of Eqs (5.244)—(5.245) is:

$$\begin{bmatrix} \frac{dx_1(t)}{dt} \\ \frac{dx_2(t)}{dt} \\ \vdots \\ \frac{dx_n(t)}{dt} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ -a_n & -a_{n-1} & \ldots\ldots & & -a_1 \end{bmatrix} \cdot \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \cdot u(t) \tag{5.247}$$

and the output matrices are:

$$\mathbf{C} = [1 \quad 0 \quad 0 \quad \ldots \quad 0]$$

and

$$\mathbf{D} = \mathbf{0} \text{ (null matrix)}$$

because the output equation is a simple scalar equation [Eq.(5.247)].

Real systems are often both higher-order and multivariable. The state space method offers the possibility to reduce the dynamic mathematical model of such real systems to a set of first-order differential equations. This is the method applied by *TACS* (Teaching Aid for Control Studies).

Example 5.4.2

Determine the state variables and state equations of a general second-order element.

The differential equation is:

$$T^2 \frac{d^2 y(t)}{dt^2} + 2\xi T \frac{dy(t)}{dt} + y(t) = u(t) \tag{5.248}$$

The state variables:

$$x_1(t) = y(t)$$
$$x_2(t) = \frac{dy(t)}{dt} \tag{5.249}$$

and the state equations:

$$\frac{dx_1(t)}{dt} = x_2(t)$$
$$\frac{dx_2(t)}{dt} = -\frac{1}{T^2}x_1(t) - \frac{2\xi}{T}x_2(t) + \frac{1}{T^2}u(t) \tag{5.250}$$

In matrix form

$$\begin{bmatrix} \frac{dx_1(t)}{dt} \\ \frac{dx_2(t)}{dt} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\frac{1}{T^2} & -\frac{2\xi}{T} \end{bmatrix} \cdot \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{1}{T^2} \end{bmatrix} \cdot u(t) \tag{5.251}$$

The output equation is:

$$y(t) = x_1(t) \tag{5.252}$$

Example 5.4.3

Consider a mixing unit *(Fig. 5.68)* and determine its state variables and state equations. c_1 and c_2 are constant. The other variables are time-dependent. Flows can change.

This is a multivariable (2) system. The total mass balance:

$$\frac{dV}{dt} = W_1 + W_2 - W_3 \tag{5.253}$$

The component balance:

$$\frac{d(Vc_3)}{dt} = W_1c_1 + W_2c_2 - W_3c_3 \tag{5.254}$$

$$V\frac{dc_3}{dt} + c_3(t)\frac{dV}{dt} = W_1c_1 + W_2c_2 - W_3c_3 \tag{5.255}$$

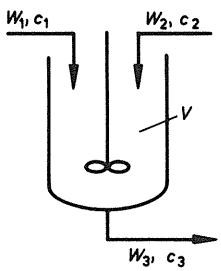

Fig. 5.68. To Example 5.4.3

Substituting the total mass balance into the component balance we obtain:

$$V\frac{dc_3}{dt} = W_1c_1 + W_2c_2 - W_3c_3 - [c_3(W_1 + W_2 - W_3)] \tag{5.256}$$

$$V\frac{dc_3}{dt} = W_1(c_1 - c_3) + W_2(c_2 - c_3) \tag{5.257}$$

The state variables are: V and c_3

The state equations are:

$$\frac{dV}{dt} = W_1 + W_2 - W_3 \tag{5.258}$$

$$\frac{dc_3(t)}{dt} = -\left(\frac{W_1(t)}{V} + \frac{W_2(t)}{V}\right)c_3(t) + \frac{c_1}{V}W_1(t) + \frac{c_2}{V}W_2(t) \tag{5.259}$$

This is a variable parameter equation.

The state equations in matrix form are:

$$\begin{bmatrix} \frac{dc_3(t)}{dt} \\ \frac{dV(t)}{dt} \end{bmatrix} = \begin{bmatrix} -\left(\frac{W_1}{V} + \frac{W_2}{V}\right) & 0 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} c_3(t) \\ V(t) \end{bmatrix} + \begin{bmatrix} \frac{c_1}{V} & \frac{c_2}{V} & 0 \\ 1 & 1 & -1 \end{bmatrix} \cdot \begin{bmatrix} W_1(t) \\ W_2(t) \\ W_3(t) \end{bmatrix} \tag{5.260}$$

The state and input matrices are also functions. They are not constant.

The output equation is simply

$$y(t) = c_3(t) \tag{5.261}$$

This example can be solved in a different way, so that the state and input matrices should be constant. In this case linearization and perturbation variables have to be used. The linearization, of course, brings its usual error with it, but it is much easier to solve.

The linearized state equations are:

$$\begin{bmatrix} \frac{d\hat{c}_3}{dt} \\ \frac{d\hat{V}}{dt} \end{bmatrix} = \begin{bmatrix} \frac{-\bar{W}_3}{\bar{V}} & 0 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} \hat{c}_3 \\ \hat{V} \end{bmatrix} + \begin{bmatrix} \frac{c_1 - \bar{c}_2}{\bar{V}} & \frac{c_2 - \bar{c}_3}{\bar{V}} & 0 \\ 1 & 1 & -1 \end{bmatrix} \cdot \begin{bmatrix} \hat{W}_1 \\ \hat{W}_2 \\ \hat{W}_3 \end{bmatrix} \tag{5.262}$$

5.4.4 The solution of the state equations

The solution of the state equations

$$\dot{\underline{x}} = \mathbf{A}\underline{x} + \mathbf{B}\underline{u}$$

can be performed by the Laplace transformation method or by a classical method, or some numerical method.

The most convenient way of solving the state equations is to use a computer. There are many suitable numerical methods, such as the trapezoid, Euler, or Runge–Kutta, etc. methods. The solution of the state equation is the dynamic behaviour of the modelled system in the time domain.

5.5 Notation

$a, b, c, \ldots$ $A, B, C, \ldots$	constants
A	gain factor
A	area
c	controlled variable
$C(s)$	its Laplace transform
c	concentration
d	disturbance, load variable
$D(s)$	its Laplace transform
D	derivative time constant
e	error
$E(s)$	its Laplace transform
$g(s)=\frac{G(s)}{A}$; $g(i\omega)$	normalized transfer or frequency function
$G(s)$, $G(I\omega)$	transfer or frequency function
H	height
$i=\sqrt{-1}$	
I	integral time constant
K	loop gain
$L(s)$	open loop transfer function
m	manipulated variable
p	pressure
$\mathscr{L}$	Laplace operator
t	time
T	time constant
T_D	dead time
T_P	period of oscillation

V	volume
W	flow rate
$W(s)$, $W(i\omega)$	closed loop transfer or frequency function
x	input
$X(s)$	its Laplace transform
y	output
$Y(s)$	its Laplace transform
φ	phase angle
ω	frequency
ω_0	natural frequency

Subscripts:

C	refers to the controller
cr	critical
S	refers to the process

Superscripts:

$\bar{}$	steady state
$\hat{}$	deviation

5.6 Literature

[1] Coughanowr, D. R. and Koppell, L. B. (1965): *Process Systems Analysis and Control.* McGraw-Hill, New York.
[2] Luyben, W. L. (1973): *Process Modeling, Simulation, and Control for Chemical Engineers.* McGraw-Hill, New York.
[3] Stephanopoulos, G. (1984): *Chemical Process Control.* Prentice-Hall, New York.
[4] Shinskey, F. G. (1979): *Process Control Systems.* (2nd ed.) McGraw-Hill, New York.
[5] Smith, C. L. (1972): *Digital Computer Process Control.* Intact Educational Publishers, Pa.
[6] Lipták, G. B. and Venczel, K. (eds) (1982): *Instrument Engineers Handbook.* Chilton, Pa.