Chapter 2
Plant Modelling

2.1 Introduction

2.1.1 Overview

A plant model is a mathematical model consisting of one or more equations whose solutions replicate the behaviour of the physical plant. Its purpose is to provide a means by which the behaviour of the plant may be studied to enable the determination of a suitable control technique and the subsequent design of a controller. The plant model also forms the basis of computer simulations for control system development.

This chapter develops the background theory and gives the knowledge needed to generate plant models. After an introduction to the basic character of plants and their components, a subsection on physical modelling is presented. This is based on the underlying science of the various applications. A comprehensive coverage of this, however, would occupy several volumes and therefore, within the space limitations, the main emphasis is on mechanical systems and electric motors as actuators to cater for a large proportion of control systems. Some introductory material on thermal and fluid systems is also given.

A case study of plant modelling in industry is presented in Appendix A2.

2.1.2 Dynamical and Non-Dynamical Systems

First a system can broadly be defined as a collection of interconnected objects that fulfil a specified function. In this context, a plant or its model may be described as a system. A part of a plant or its model will be termed a subsystem. A dynamical system is one whose outputs depend upon the past history of the inputs. Plants to which feedback control is applied are dynamical systems and most are continuous,

S.J. Dodds, *Feedback Control*, Advanced Textbooks in Control
and Signal Processing, DOI 10.1007/978-1-4471-6675-7_2

meaning that their variables are continuous functions of time. These are systems that obey differential equations in which time is the independent variable. The differential equation is the basic form of mathematical model of a plant from which other forms of model may be derived. The physical plant is usually separable into a number of constituent parts, or subsystems, each of which can be modelled separately. This is the physical modelling approach of Sect. 2.2.

The term 'dynamics' has two meanings in the field of control engineering. The first is the way in which the output of a dynamical system responds to its inputs. A system with fast (or high) dynamics is one whose outputs respond quickly to changes in its inputs. The second meaning is related to the modelling of mechanical systems and is defined in Sect. 2.2.2.

A non-dynamical system is one in which the present output depends only on the present input. These are subsystems such as a measurement device. A single input, single output (SISO) non-dynamical system can be represented by an equation defining the relationship between the input, $x(t)$ and the output, $y(t)$, in the form

$$y(t) = f\left[x(t)\right], \tag{2.1}$$

where $f(\bullet)$ is a single-valued and continuous function of its argument. Similarly, a multiple input, multiple output (MIMO or multivariable) non-dynamical subsystem can be represented by the set of equations

$$y_i(t) = f_i\left[x_1(t), x_2(t), \ldots, x_r(t)\right], \quad i = 1, 2, \ldots, m. \tag{2.2}$$

where x_j, $j = 1,\ 2, \ldots, r$, are the inputs and y_i, $i = 1,\ 2, \ldots,\ m$, are the outputs.

A dynamical system is one in which the present outputs depend on past values of the inputs and usually past values of the outputs. It is important to note that the present output of a controlled plant in the real world cannot respond instantaneously to a step change in the control input. The mathematical model therefore cannot contain algebraic dependence of the output on the input. To take a practical example, a step change in the torque demand of an electric drive on a train locomotive cannot cause a step change in the speed of the train, which is the controlled output.

Initially, consider a first-order SISO plant with input, $u(t)$, and output, $y(t)$. This obeys a first-order differential equation of the generic form

$$\dot{y}(t) = f\left[y(t), u(t)\right], \tag{2.3}$$

where $f(\bullet)$ is a single-valued and continuous function of its arguments. If (2.3) is written as an integral equation,

$$y(t) = \int_0^t f\left[y\left(\tau\right), u\left(\tau\right)\right] d\tau, \tag{2.4}$$

then it may be readily seen that the present output, $y(t)$, depends upon the continuum of past values of the input $u\left(\tau\right)$ and the output, $y\left(\tau\right)$, for $\tau \in [0,\ t]$.

An SISO plant of nth order can be modelled by an ordinary differential equation of the generic form,

$$y^{(n)} = f\left(y^{(n-1)}, \ldots, y^{(1)}, y, u^{(m)}, u^{(m-1)}, \ldots, u^{(1)}, u\right), \tag{2.5}$$

where terms such as $x^{(q)}$ mean $\mathrm{d}x^q/\mathrm{d}t^q$. A fundamental restriction on this model is $0 \leq m < n$. If $m = n$ then there would be a direct dependence of $y^{(n)}$ on $u^{(n)}$, implying a direct dependence of y on u. This, in turn, would imply that a step change in $u(t)$ would cause a step change in $y(t)$. Such behaviour is not found in physical plants. On the other hand, a step change in $u(t)$ will cause a step change in an output derivative of a certain order. To return to the train traction example, a step change in the drive torque demand will produce a step change in the acceleration of the train, which is the first derivative of the controlled speed. Considering the generic model (2.5), a step change in $u(t)$ will cause a step change in $y^{(r)}$, where $r = n - m$, but cannot cause a step change in any of the lower derivatives of y. As can be seen, $y^{(n)}$ depends algebraically on $u^{(m)}$ implying that a step change in $u^{(m)}$ will cause a step change in $y^{(n)}$, in turn implying that a step change in $u^{(0)}$, i.e. $u(t)$, will cause a step change in $y^{(n-m)}$. An important parameter in control system design that is related to this plant property is the *relative degree*, defined as

$$r = n - m. \tag{2.6}$$

This will be met in Chaps. 8 and 10. The term originates from its application to the transfer function model of a linear time-invariant plant, in which it is defined as the difference in degree between the denominator polynomial and numerator polynomial.

Since

$$n > m, \tag{2.7}$$

then, in view of (2.6),

$$r > 0. \tag{2.8}$$

The most general model of a multivariable plant of nth order is a set of interconnected ordinary differential equations of the generic form,

$$\begin{aligned} y_i^{(n_i)} = f_i\Big(&y_j^{(n_j-1)}, \ldots, y_j{}^{(1)}, y_j, u_k^{(m_{k,i})}, u_k^{(m_{k,i}-1)}, \ldots, u_k^{(1)}, u_k\Big), \\ &i = 1, 2, \ldots, m,\ j = 1, 2, \ldots, m,\ k = 1, 2, \ldots, p \end{aligned} \tag{2.9}$$

where

$$\sum_{l=1}^{m} n_i = n \tag{2.10}$$

and

$$0 \leq m_{k,i} < n_i \,. \tag{2.11}$$

Equation (2.10) merely states that the total order of the system is equal to the sums of the orders of the subsystems modelled by the individual ordinary differential equations of (2.9). Inequality (2.11) is analogous to (2.7) for SISO plants and represents similar practical limitations of real plants.

The relative degree of the plant with respect to the *i*th output is defined as

$$r_i = \min_k \left(n_i - m_{k,i}\right) . \tag{2.12}$$

It is the minimum order of the derivative of the output, y_i, that depends algebraically on *any* control input. This parameter is important in control system design, particularly when applying the techniques of sliding mode control, feedback linearisation or forced dynamic control.

2.1.3 Linearity and Nonlinearity

All systems are classified as linear or nonlinear. A linear system may be readily recognised through every mathematical expression being of the general form,

$$\sum_i C_i v_i(t) + B_i \,. \tag{2.13}$$

The scalar coefficients, C_i, are usually constant, in which case the term linear time-invariant (LTI) system applies. Some of the variables, $v_i(t)$, are derivatives in the case of dynamical systems. The constant bias, B_i, is included for generality but is not present in many cases. Occasionally, one or more of the coefficients are time varying, in which case the term linear time-varying (LTV) system applies. This provides a straightforward means of recognising a linear model. A nonlinear system is readily recognised as it contains at least one expression not of the general form (2.13). A simple example of a linear non-dynamic LTI subsystem is the model of a liquid level transducer that gives a voltage, y, proportional to the height, h, of the liquid in a cylindrical vessel, given by the linear equation,

$$y = K_h h \tag{2.14}$$

where K_h is the height measurement constant. This subsystem is an SISO one with input, h, and output, y. An example of a nonlinear non-dynamic subsystem is the model of the process in an electromagnet that produces the force, f, given the current, i, which may be written

$$f = K_f \frac{i^2}{(x - x_0)^2}, \tag{2.15}$$

where K_f is the electromagnetic force constant and x is the length of the air gap and x_0 is a positive constant. In this case the subsystem is a multiple input, single output (MISO) one with i and x as inputs and output f.

An LTI system exhibits the *scaling property* and the *superposition property*. It is important to note that if bias terms such as in (2.13) are present, then the equations must be reformulated in terms of *changes* in the variables to test for these properties. The scaling property is as follows. If the inputs of a system are multiplied by a constant, then the outputs will be multiplied by the same constant. The superposition property is as follows. Let a sequence of inputs be applied to a linear system one after the other and the corresponding outputs recorded. Then if a single input is applied that is the sum of the inputs previously applied, the output will be the sum of the previously recorded outputs. A nonlinear plant model will exhibit neither of these properties.

If the non-dynamical SISO subsystem modelled by (2.1) is linear, then it has the scaling property,

$$f\left[\lambda x(t)\right] = \lambda f\left[x(t)\right], \tag{2.16}$$

where λ is a scalar. It will also have the superposition property as follows. Let $x_k(t)$, $k = 1, 2, \ldots, p$, be a set of inputs applied separately and let $y_k(t)$ be the corresponding outputs. After this, let a single input, $x(t) = \sum_{k=1}^{p} x_k(t)$, be applied. Then the resulting output is $y(t) = \sum_{k=1}^{p} y_k(t)$. If the system is linear, then (2.1) is of the form,

$$y(t) = f\left[x(t)\right] = Cx(t), \tag{2.17}$$

where C is a coefficient that is usually constant. It is straightforward to confirm that the non-dynamical system (2.17) has the scaling and superposition properties. If, as is occasionally the case, C is time varying, the scaling property still holds.

Consider now the non-dynamical multivariable subsystem modelled by (2.2) . If the sets of inputs and outputs are expressed as input and output vectors, defined as $\mathbf{x}(t) \triangleq [x_1(t)\ x_2(t) \ldots x_r(t)]^{\mathrm{T}}$ and $\mathbf{y}(t) \triangleq [x_1(t)\ x_2(t) \ldots x_m(t)]^{\mathrm{T}}$, then (2.2) becomes

$$\mathbf{y}(t) = \mathbf{f}\left[\mathbf{x}(t)\right]. \tag{2.18}$$

The expression of the scaling property is then

$$\mathbf{f}\left[\lambda \mathbf{x}(t)\right] = \lambda \mathbf{f}\left[\mathbf{x}(t)\right]. \tag{2.19}$$

If the same subsystem obeys the superposition property, then the following is true.

Let $\mathbf{x}_k(t)$, $k = 1, 2, \ldots, q$, be a set of input vectors applied separately and let $\mathbf{y}_k(t)$ be the corresponding output vectors. After this, let a single input vector, $\mathbf{x}(t) = \sum_{k=1}^{q} \mathbf{x}_k(t)$, be applied. Then if the system exhibits the superposition property, the resulting output vector is $\mathbf{y}(t) = \sum_{k=1}^{q} \mathbf{y}_k(t)$. If the system is linear, then (2.18) has the form,

$$\mathbf{y}(t) = \mathbf{f}[\mathbf{x}(t)] = \mathbf{C}\mathbf{x}(t), \tag{2.20}$$

where $\mathbf{C}$ is a matrix of coefficients. In view of (2.20), the LHS of (2.19) is

$$\mathbf{f}[\lambda \mathbf{x}(t)] = \mathbf{C}\lambda \mathbf{x}(t) \tag{2.21}$$

and the RHS of (2.19) is

$$\lambda \mathbf{f}[\mathbf{x}(t)] = \lambda \mathbf{C}\mathbf{x}(t). \tag{2.22}$$

Since the RHS of (2.21) and (2.22) are equal, the system has the scaling property. If the input vectors, $\mathbf{x}_k(t)$, $k = 1, 2, \ldots, q$, are applied, one at a time, the output vectors will be $\mathbf{y}_k(t) = \mathbf{C}\mathbf{x}_k(t)$. Then let the single input vector, $\mathbf{x}(t) = \sum_{k=1}^{q} \mathbf{x}_k(t)$, be applied. This yields the output vector $\mathrm{y}(t) = \mathbf{C}\sum_{k=1}^{q} \mathbf{x}_k(t)$. This may be written as $\sum_{k=1}^{q} \mathbf{C}\mathbf{x}_k(t) = \sum_{k=1}^{q} \mathbf{y}_k(t)$. The system therefore has the superposition property.

Considering now the SISO dynamical system modelled by (2.5), if it is an LTI system, it will be of the form

$$\begin{aligned} y^{(n)} = {} & a_{n-1}y^{(n-1)} + \cdots + a_1 y^{(1)} + a_0 y + b_m u^{(m)} \\ & + b_{m-1}u^{(m-1)} + \cdots + b_1 u^{(1)} + b_0 u^{(0)}. \end{aligned} \tag{2.23}$$

where a_i, $i = 0, 1, \ldots, n-1$, and b_j, $j = 0, 1, \ldots, m$, are constant coefficients. Its scaling property may be stated as follows. If $y_1(t)$ is the output for given initial conditions, $y_1^{(i)}(0)$, $i = 0, 1, \ldots, n-1$ and a given input, $u_1(t)$, with finite derivatives, $u_1^{(j)}$, $j = 1, 2, \ldots, m$, then if the scaled input, $\lambda u_1(t)$, whose derivatives are $\lambda u_1^{(j)}$, $j = 1, 2, \ldots, m$, and the initial conditions are similarly scaled, i.e. $\lambda y_1^{(i)}(0)$, $i = 0, 1, \ldots, n-1$, then the output will be scaled by the same factor, i.e. it will be $\lambda y_1(t)$. This is tantamount to stating that scaling the input and its derivatives by λ scales the output and its derivatives by λ. That this is the case for system (2.23) is immediately apparent when both sides are multiplied by λ.

It will now be shown that system (2.23) has the superposition property. Let $u_k(t)$, $k = 1, 2, \ldots, q$, be a set of inputs applied separately and let $y_k(t)$ be the corresponding outputs with initial conditions, $y_k^{(i)}(0)$, $i = 0, 1, \ldots, n-1$. After this, let a single input,

$$u(t) = \sum_{k=1}^{q} u_k(t), \tag{2.24}$$

be applied with initial conditions, $\sum_{k=1}^{q} y_k^{(i)}(0)$, $i = 0,\ 1, \ldots,\ n-1$. Then the output is

$$y(t) = \sum_{k=1}^{n} y_k(t). \tag{2.25}$$

To show that system (2.23) has this property, first it is observed that each of the input-output pairs, $u_k(t)$ and $y_k(t)$, $k = 1,\ 2, \ldots, q$, satisfies (2.23), noting that $y_k(t) = y_k^{(0)}(t)$. If all q equations are added, while grouping the corresponding terms, then the result is as follows.

$$\begin{aligned}\sum_{k=1}^{q} y_k^{(n)}(t) &= a_{n-1}\sum_{k=1}^{q} y_k^{(n-1)}(t) + \cdots + a_1\sum_{k=1}^{q} y_k^{(1)}(t) + a_0\sum_{k=1}^{q} y_k^{(0)}(t)\\ &\quad + b_m\sum_{k=1}^{q} u_k^{(m)}(t) + b_{m-1}\sum_{k=1}^{q} u_k^{(m-1)}(t) + \cdots + b_1\sum_{k=1}^{q} u_k^{(1)}(t) + b_0\sum_{k=1}^{q} u_k^{(0)}(t).\end{aligned} \tag{2.26}$$

The summation terms in (2.23) may then be expressed in terms of $u(t)$ using (2.24) and $y(t)$ using (2.25). Thus

$$\begin{aligned}y^{(n)}(t) &= a_{n-1}y^{(n-1)}(t) + \cdots + a_1 y^{(1)}(t) + a_0 y^{(0)}(t)\\ &\quad + b_m u^{(m)}(t) + b_{m-1}u^{(m-1)}(t) + \cdots + b_1 u^{(1)}(t) + b_0 u^{(0)}(t).\end{aligned} \tag{2.27}$$

Since this is the differential equation of system (2.23), then the system has the superposition property.

It can be similarly shown that the multivariable linear system (2.9) has the scaling and superposition properties if it is in the LTI form,

$$y_i^{(n_i)} = \sum_{j=1}^{m}\sum_{k=0}^{n_j-1} a_{ijk}\, y_j^{(k)} + \sum_{j=1}^{p}\sum_{k=0}^{m_{j,i}} b_{ijk}\, u_j^{(k)}, \quad i = 1, 2, \ldots, m. \tag{2.28}$$

where a_{ijk} and b_{ijk} are constant coefficients.

The scaling and superposition properties of LTI systems enable several different forms of plant model to be derived that are useful in control system design. These will be met in Chap. 3. For a nonlinear plant, the only alternative to (2.5) or (2.9) is the state-space model of Chap. 3. The same is true for LTV dynamical systems.

The coil of an electromagnet with resistance, R, and inductance, L, is a linear dynamical subsystem in which the current, i, is related to the applied voltage, v, by

$$Li^{(1)} + Ri = v, \tag{2.29}$$

where $i^{(1)} = \mathrm{d}i/\mathrm{d}t$. This is an SISO subsystem with input, v, and output, i.

An example of a nonlinear dynamical subsystem is the dynamics model of a rigid body in free fall subject to externally applied torque components. This could be part of a spacecraft model. In this case, the subsystem is a multivariable one with three inputs, the torque components, γ_x, γ_y and γ_z, along the three mutually perpendicular principal axes of inertia, x, y and z, and three outputs, the body angular velocity components, ω_x, ω_y and ω_z. Thus

$$\begin{aligned} \dot{\omega}_x &= k_x\ \omega_y\omega_z + b_x\gamma_x \\ \dot{\omega}_y &= k_y\ \omega_z\omega_x + b_y\ \gamma_y \\ \dot{\omega}_z &= k_z\ \omega_x\omega_y + b_z\ \gamma_z \end{aligned} \tag{2.30}$$

where $k_x = \left(J_{yy} - J_{zz}\right)/J_{xx}$, $k_y = \left(J_{zz} - J_{xx}\right)/J_{yy}$, $k_z = \left(J_{yy} - J_{xx}\right)/J_{zz}$, $b_x = 1/J_{xx}$, $b_y = 1/J_{yy}$ and $b_z = 1/J_{zz}$ where J_{xx}, J_{yy} and J_{zz} are the principal axis moments of inertia.

2.1.4 Modelling Categories and Basic Forms of Model

2.1.4.1 White-Box Modelling and Differential Equations

White-box modelling, occasionally referred to as glass-box modelling, is a general term used to describe the process of studying each internal component of a system, forming a mathematical model of each and then assembling the whole using shared inputs and outputs to form an overall mathematical model. In control engineering, white-box modelling applies to plants for which control systems are to be created. This is the physical modelling of Sect. 2.2. All plants are continuous-time dynamical systems and therefore the fundamental form of model emerging from the physical modelling is the differential equation, many examples of which will be found throughout this chapter. All other forms of plant model stem from the differential equation.

2.1.4.2 Black Box Modelling and Transfer Functions

Black box modelling is a general term used to describe the process of creating a mathematical model of a system by collecting information from observations of the responses of its outputs to given inputs, without the study of its internal components. This approach is often taken in industry as it is less time consuming and therefore more cost effective than white-box modelling but is restricted to linear models. In control engineering, various control inputs are applied to the plant and its measurement variables observed. There are three basic approaches to the processing of these variables, covered in Sect. 2.2. One is the step response method of Sect. 2.3.2, applicable to first and second-order plants and leads to Laplace

transfer function models. The second is the frequency domain method of Sect. 2.3.3, which gives Laplace transfer function models and the third is the time domain method of Sect. 2.3.4, which gives z-transfer function models. The second and third methods are applicable to plants of arbitrary order.

2.2 Physical Modelling

2.2.1 Introduction

In physical plant modelling, the underlying science of each hardware component is applied to derive a mathematical component model in the form of algebraic and/or differential equations. The plant model is then the set of component models interconnected by means of common input and output variables. In the following subsections, some basic components are modelled that may be used to construct various plant models, starting with the most elementary. The equations of these basic elements may be used to build models of specific plants.

2.2.2 Mechanical Modeling Principles

2.2.2.1 Dynamic and Kinematic Subsystems

The *dynamic* subsystem is defined as the part of a mechanical system that relates translational and/or rotational velocity components to applied forces and/or torques. The *kinematic* subsystem is defined as the part of a mechanical system that relates the translational and/or rotational displacements to the translational and/or rotational velocities. The dynamic subsystem involves the inertial parameters of mass and/or moment of inertia, while the kinematic subsystem does not.

2.2.2.2 Degrees of Freedom of Motion

The number of degrees of freedom (d.o.f.) of movement of a mechanical system is the number of coordinates needed to define its position, as illustrated in Fig. 2.1. Many mechanisms such as the linear and rotational actuators of Fig. 2.1a, b, only have a single degree of freedom. The x-y positioning drive of Fig. 2.1c has many applications in industry. The two directions of movement are referred to as *axes* and for this reason the generic name for such mechanisms is the *multi-axis mechanism.*

Many multi-axis mechanisms have combinations of rotational and translational degrees of freedom. The axis for a rotational degree of freedom is that about which rotation of one component takes place relative to another component. The number of axes is equal to the number of degrees of freedom of movement.

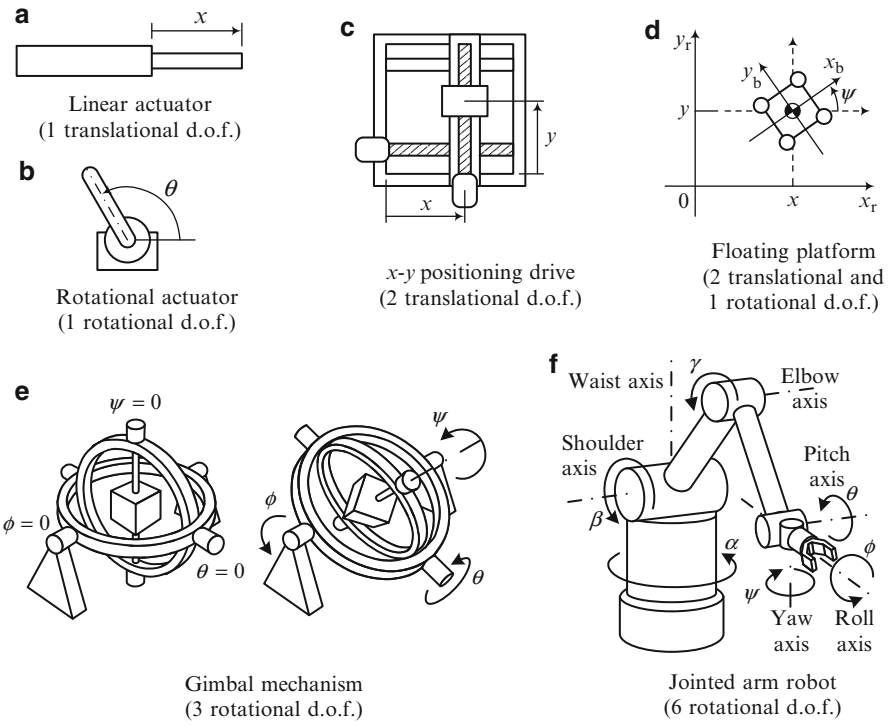


Fig. 2.1 Illustrations of translational and rotational degrees of freedom

Also, the number of degrees of freedom is equal to the number of control actuators, two worm drives being employed in the example of Fig. 2.1c.

Applications at sea such as oil rigs and wind turbines have to be positioned on platforms that are floated out to locations with translational coordinates, x and y, in a frame of reference, (x_r, y_r), and orientated about the yaw axis by an angle, ψ, as shown in Fig. 2.1d. prior to anchoring. In this case, there is no mechanism connecting the platform to the frame of reference so that movement between moving parts can be measured. Instead the two translational coordinates are measured using the global positioning system (GPS) and the rotational one by a compass.

The gimbal mechanism of Fig. 2.1e consists of frames mounted one within the other. The relative rotation axes and associated motors and angle sensors are arranged so that the object in the centre (shown as a cube) can be brought to any orientation by means of three control loops.

The universal multi-axis machine is the jointed-arm robot in which a workpiece held by the gripper is positioned in the finite three-dimensional work space with any orientation. Such robots can have various configurations. The one of Fig. 2.1f has six rotational degrees of freedom controlled using motors and joint angle sensors. Importantly, through the arm configuration, the purely rotational degrees of freedom of the joints are used to control the three *translational* and three rotational degrees of

freedom of the workpiece referred to a fixed frame of reference. Many mechanisms exist in which a given degree of freedom of motion entails both rotation and translation, an example being the crank mechanism.

Three of the rotational movements in Fig. 2.1f are named *roll*, *pitch* and *yaw*. This terminology originated in nautical applications, where the rotation of a ship about its longitudinal (x) axis is called *roll* motion, rotation about the vertical (z) axis perpendicular to the roll axis is called *yaw* motion and rotation about the horizontal (y) axis, perpendicular to the x and z axes, is called *pitch* motion. The same terminology is used for aircraft, underwater vehicles, spacecraft and automobiles.

In applications such as spacecraft and underwater vehicles it is necessary to position and orientate the vehicle body (considered to be rigid here) in the absence of a mechanism connecting the body to the frame of reference. Then it is possible to either select a *fictitious* mechanism connecting the body with the frame of reference, such as the gimbal mechanism of Fig. 2.1e, whose kinematic differential equations (KDEs) can be used as a basis for generating position and attitude coordinates, or select another set of KDEs not associated with any mechanism (Sect. 2.2.4.3).

For illustrative purposes, the gimbal mechanism-based roll, pitch and yaw attitude angles for a vehicle (represented by a cube) are shown in Fig. 2.2.

Starting with the body-fixed frame (x_b, y_b, z_b) aligned with the reference frame (x_r, y_r, z_r) as shown in Fig. 2.2a, a body rotation is made through an angle, ϕ, about the roll axis, bringing it to the orientation shown in Fig. 2.2b. Then a further body rotation is made through an angle, θ, about the newly orientated pitch axis, y_b, bringing it to the orientation shown in Fig. 2.2c. Then a final rotation of the body is made through an angle, ψ, about the newly orientated yaw axis, z_b, bringing it to the orientation shown in Fig. 2.2d. There are, in fact, twelve such attitude representations corresponding to the twelve permutations of the axis rotation orders, the one of Fig. 2.2 being r-p-y (roll–pitch–yaw), the remaining ones being r-y-p, p-r-y, p-y-r, y-r-p, y-p-r, r-p-r, r-y-r, p-r-p, p-y-p, y-r-y and y-p-y. The first six permutations have equivalent gimbal mechanisms. The last six permutations do not as they entail the first and last rotations about the same body-fixed axis. They do produce valid attitude representations, however, because this body-fixed axis has different orientations with respect to the reference frame for the first and third rotations.

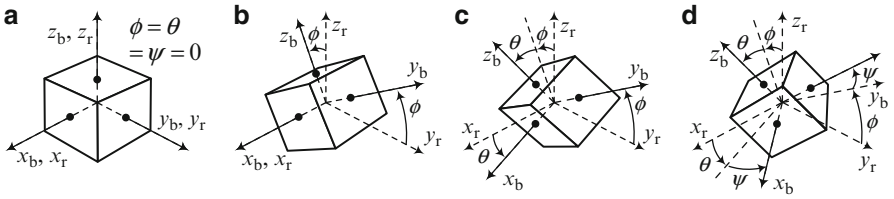


Fig. 2.2 A gimbal mechanism-based roll–pitch–yaw attitude representation for a vehicle. (**a**) Home attitude. (**b**) Roll rotation. (**c**) Pitch rotation. (**d**) Yaw rotation

The rotational dynamics and kinematics for three (x_i, y_i, z_i) degrees of freedom, which are much more complex than those for a single degree of freedom, are important as they are relevant to many applications in vehicle control and robotics.

2.2.2.3 Rigid Body with One Degree of Freedom

Figure 2.3a represents a rigid body of mass, M, constrained to move in a straight line to which is applied a control force, f_c, acting through the centre of mass (indicated by the standard symbol, ◓) opposed by a force, f_o, due to friction, drag (i.e. air or fluid resistance) or retention spring or a combination thereof, and subject to an external disturbance force, f_d, giving rise to a velocity, v, and displacement, x, relative to an inertial frame of reference.

An inertial frame of reference refers to a set of axes that are not undergoing either translational or rotational acceleration relative to inertial space. Hence Newton's laws of motion would hold in a laboratory fixed with respect to an inertial frame. The frame of reference is shown as three dimensional, which is usual, but in this case, only the x_i axis is relevant. Figure 2.3b similarly represents a rigid body with moment of inertia, J, about an axis parallel to the y_i axis, constrained to rotate about this axis subject to an applied torque, γ_a, opposed by a torque, γ_o, giving rise to an angular velocity, ω, and angular displacement, θ. The forces, torques, linear velocity and angular velocity are vector quantities as indicated by the arrows in Fig. 2.3 but in these simple cases are co-linear.

The dynamics equations are obtained for Fig. 2.3a by equating the net force to the first derivative of the linear momentum and for Fig. 2.3b by equating the net torque to the first derivative of the angular momentum. The KDEs for simple single-degree-of-freedom mechanical components are simply statements that the velocity is the first derivative of the displacement. Assuming that the mass is constant, the model for Fig. 2.3a is as follows.

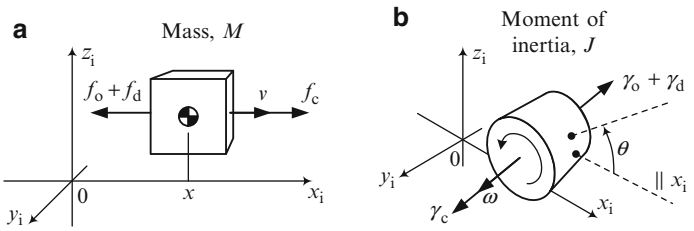


Fig. 2.3 Rigid body constrained to move with one degree of freedom. (**a**) Translational. (**b**) Rotational

$$\text{Dynamic subsystem}: \frac{\mathrm{d}}{\mathrm{d}t}(Mv) = f_\mathrm{a} - f_\mathrm{o} \Rightarrow \dot{v} = \frac{1}{M}(f_\mathrm{a} - f_\mathrm{o}). \tag{2.31}$$

$$\text{Kinematic subsystem}: \dot{x} = v. \tag{2.32}$$

The corresponding model for Fig. 2.3b is as follows.

$$\text{Dynamic subsystem}: \frac{\mathrm{d}}{\mathrm{d}t}(J\omega) = \gamma_\mathrm{a} - \gamma_\mathrm{o} \Rightarrow \dot{\omega} = \frac{1}{J}(\gamma_\mathrm{a} - \gamma_\mathrm{o}). \tag{2.33}$$

$$\text{Kinematic subsystem}: \dot{\theta} = \omega. \tag{2.34}$$

The opposing force of the translational model may be written as

$$f_\mathrm{o} = f_\mathrm{g} + f_\mathrm{f} + f_\mathrm{s} + f_\mathrm{I}, \tag{2.35}$$

where f_g is the drag force, f_f is the friction force, f_s is the spring force and f_I is the inertial force of any other connected masses. For specific applications, some or all of these may be zero. For the rotational model, the equivalent breakdown of the opposing torque is

$$\gamma_\mathrm{o} = \gamma_\mathrm{g} + \gamma_\mathrm{f} + \gamma_\mathrm{s} + \gamma_\mathrm{I}. \tag{2.36}$$

The use of (2.31) or (2.33) in conjunction with the material of the following three subsections is referred to, respectively, as the *force or torque balance methods*.

2.2.2.4 Drag Forces and Torques

This topic is relevant to applications such as aircraft, surface ships and underwater vehicles. With reference to Fig. 2.3a, Raleigh's equation for fluid drag [1] is

$$f_\mathrm{g} = \frac{1}{2}\rho v^2 C_\mathrm{d} A, \tag{2.37}$$

where ρ is the density of the fluid in which the body is immersed, A is the area of the orthographic projection onto a plane perpendicular to the direction of motion and C_d is the dimensionless drag coefficient of the fluid. To simplify any plant model of which this is part and at the same time to ensure that the drag force opposes the motion, (2.37) may be replaced by

$$f_\mathrm{g} = K_\mathrm{d} v^2 \mathrm{sgn}(v) = K_\mathrm{d} |v| v, \tag{2.38}$$

where $\text{sgn}(v) \triangleq \begin{cases} +1, v > 0 \\ 0, v = 0 \\ -1, v < 0 \end{cases}$ and $K_\text{d} = \frac{1}{2}\rho C_\text{d} A$.

For the rotational case of Fig. 2.3, the equivalent relationship,

$$\gamma_\text{g} = K_\text{g}\omega^2 \text{sgn}(v) = K_\text{d}\,|\omega|\,\omega, \tag{2.39}$$

may be used but in this case, the expression of K_g in terms of the parameters of Rayleigh's equation (2.37) is not straightforward as the relative velocity between a rotating body of arbitrary shape and the fluid in which it is immersed is a function of the position on its surface. In practice, this problem would be circumvented by determination of K_g experimentally.

2.2.2.5 Friction Forces and Torques

Many controlled plants embody mechanisms with relatively moving surfaces. A friction force is broadly divided into three components. Thus

$$f_\text{f} = f_\text{fs} + f_\text{fc} + f_\text{fv}, \tag{2.40}$$

where f_fs is the static friction component, f_fc is the Coulomb friction component and f_fv is the viscous friction component.

Detailed information on the underlying physical processes of friction, including explanations at the molecular level, may be found in works on mechanical engineering [2] but the models given here should be understandable from common experience with the behaviour of relatively moving objects in contact.

Suppose that the object of Fig. 2.3a has a flat base and is resting on a flat surface. Then starting from rest, if the applied force, f_a, is gradually increased from zero, at first there will be no relative movement, but above a certain level, $f_\text{a} = f_\text{fs}$, the object will suddenly move. This is the static friction force given by

$$f_\text{fs} = \mu f_\text{n} \text{sgn}\,(f_\text{a}) = F_\text{fs} \text{sgn}\,(f_\text{a})\,, \tag{2.41}$$

where f_n is the normal force keeping the surfaces in contact and μ is the coefficient of static friction, determined experimentally. Sometimes, static friction is referred to as *stiction* or *stick–slip friction* due to the sticking effect for $|f_\text{a}| < F_\text{fs}$. For rotating objects in contact such as bearings, the identification of the normal force is less obvious but a similar phenomenon exists in which the static friction torque is

$$\gamma_\text{fs} = \Gamma_\text{fs} \text{sgn}\,(\gamma_\text{a})\,, \tag{2.42}$$

where Γ_fs is the constant static friction torque magnitude.

In the following, v is the relative velocity between moving parts rather than the velocity of one part with respect to an inertial frame of reference.

For the translational case, once f_{fs} of (2.41) is exceeded, the physics of the friction changes as the relative movement begins and a model of the opposing force is

$$f_{\mathrm{f}} = f_{\mathrm{fc}} + f_{\mathrm{fv}} \tag{2.43}$$

where

$$f_{\mathrm{fc}} = F_{\mathrm{fc}}\mathrm{sgn}(v), \tag{2.44}$$

F_{fc} being constant and also $F_{\mathrm{fc}} < F_{\mathrm{fs}}$. This is the Coulomb friction force. Also

$$f_{\mathrm{fv}} = K_{\mathrm{v}} v. \tag{2.45}$$

This is the viscous friction force, with coefficient, K_{v}. Viscous friction is sometimes referred to as kinetic friction or dynamic friction, as it is a continuous function of v, in contrast to the other components whose magnitudes are constant.

Figure 2.4a shows an example of the transfer characteristic, i.e. the graph of f_{f} against v that results when the effects of static, coulomb and viscous friction described above are combined.

When experiments are conducted to measure the friction transfer characteristic, however, a similar result is obtained but with a continuous transition between the static friction and the combined Coulomb and viscous friction, as shown in Fig. 2.4b. The dot–dashed straight line segments are those of Fig. 2.4a for comparison. This form of transfer characteristic is preferred in view of its being more realistic and also better behaved regarding the accuracy of the numerical integration in simulations. The author has devised the following convenient function for this.

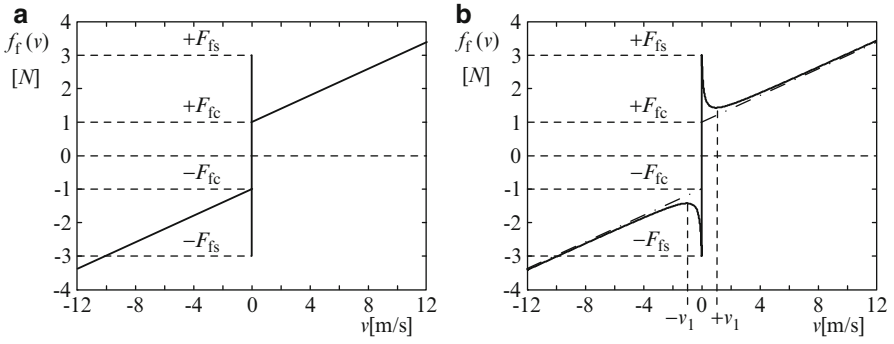


Fig. 2.4 Friction transfer characteristics. (**a**) Theoretical form. (**b**) Realistic form

$$f_{\mathrm{f}}(v) = K_{\mathrm{v}} v + \left(\frac{F_{\mathrm{fs}} + F_{\mathrm{fc}} K |v|}{1 + K |v|} \right) \mathrm{sgn}(v). \tag{2.46}$$

Here, K is a parameter that may be adjusted to obtain the required sharpness of transition, this being increased as K is increased. If v_1 is specified, then since the transfer characteristic is reflected in the origin, the required value of K is obtained by setting $\frac{\mathrm{d}}{\mathrm{d}v} f_{\mathrm{f}}(v) = 0$ for $v = v_1 > 0$ for which (2.46) becomes

$$f_{\mathrm{f}}(v) = K_{\mathrm{v}} v + \frac{F_{\mathrm{fs}} + F_{\mathrm{fc}} K v}{1 + K v} \tag{2.47}$$

Then

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}v} \left[K_{\mathrm{v}} v + \frac{F_{\mathrm{fs}} + F_{\mathrm{fc}} K v}{1 + K v} \right] &= K_{\mathrm{v}} + \frac{(1 + K v) F_{\mathrm{fc}} K - (F_{\mathrm{fs}} + F_{\mathrm{fc}} K v) K}{(1 + K v)^2} \\ &= \frac{\left(1 + 2K v + K^2 v^2\right) K_{\mathrm{v}} - (F_{\mathrm{fs}} - F_{\mathrm{fc}}) K}{(1 + K v)^2} \Rightarrow \\ K_{\mathrm{v}} v_1^2 K^2 + (2K_{\mathrm{v}} v_1 - F_{\mathrm{fs}} + F_{\mathrm{fc}}) K + K_{\mathrm{v}} &= 0 \Rightarrow \\ K = \frac{F_{\mathrm{fs}} - F_{\mathrm{fc}} - 2K_{\mathrm{v}} v_1 \pm \sqrt{(2K_{\mathrm{v}} v_1 - F_{\mathrm{fs}} + F_{\mathrm{fc}})^2 - 4K_{\mathrm{v}}^2 v_1^2}}{2K_{\mathrm{v}} v_1^2}. \end{aligned} \tag{2.48}$$

Since, $\sqrt{(2K_{\mathrm{v}} v_1 - F_{\mathrm{fs}} + F_{\mathrm{fc}})^2 - 4K_{\mathrm{v}}^2 v_1^2} < |F_{\mathrm{fs}} - F_{\mathrm{fc}} - 2K_{\mathrm{v}} v_1|$, a necessary condition for $K > 0$ is

$$F_{\mathrm{fs}} - F_{\mathrm{fc}} - 2K_{\mathrm{v}} v_1 > 0. \tag{2.49}$$

Then both roots are positive and therefore valid but the largest one will be chosen as this yields the sharpest transition between the static friction and the combined Coulomb and viscous friction. Hence

$$K = \frac{F_{\mathrm{fs}} - F_{\mathrm{fc}} - 2K_{\mathrm{v}} v_1 + \sqrt{(2K_{\mathrm{v}} v_1 - F_{\mathrm{fs}} + F_{\mathrm{fc}})^2 - 4K_{\mathrm{v}}^2 v_1^2}}{2K_{\mathrm{v}} v_1^2}. \tag{2.50}$$

In fact, the transfer characteristic of Fig. 2.1b was produced using (2.46) and (2.50) with $F_{\mathrm{fs}} = 6$ [N], $F_{\mathrm{fc}} = 1$ [N], $K_{\mathrm{v}} = 0.4$ [N/ (m/s)] and $v_1 = 2$ [m/s]. For the rotational case, F_{fs}, F_{fc}, K_{v} and v_1 are replaced, respectively, by Γ_{fs}, Γ_{fc}, K_{ω} and ω_1.

Static friction can cause steady-state errors in traditional control loops without an integral term and limit cycling with integral terms. The reader should be aware, however, that in most real applications the parameters of the friction model are highly dependent upon environmental conditions, particularly temperature, and therefore a model-based controller with inbuilt friction compensation would be

difficult to implement. The recommended approach is to use the model presented in this section with typical parameters for the application in hand in a simulation to predict its effect on a traditional control loop and change over to a robust control technique (Chaps. 9 and 10) if unsatisfactory performance is predicted.

2.2.2.6 Spring Force and Torque

In some mechanical systems, a rigid body such as illustrated in Fig. 2.3 is either physically connected to one or more components via springs or is part of a so-called lumped parameter model of a flexible structure consisting of rigid bodies interconnected by linear springs or torsion springs. In either case, with reference to (2.35) and (2.36),

$$f_{\mathrm{s}} = \sum_{i=1}^{m} f_{si} \quad \text{or} \quad \gamma_{\mathrm{s}} = \sum_{i=1}^{m} \gamma_{si} \tag{2.51}$$

where f_{si} or γ_{si}, $i = 1, 2, \ldots, m$, is the individual spring forces or torques. Figure 2.5 shows the translational and rotational single-degree-of-freedom cases in which all the displacements and velocities shown are with respect to the inertial frame of reference. The object could be another movable rigid body or a fixed point with respect to the inertial frame of reference.

In Fig. 2.5a, x_i is the displacement of the point on the ith object to which one end of the spring is attached and x_{0i} is the displacement of the same end of the spring if detached from the object and at rest with zero spring force. In these models, the springs are regarded as having zero mass. Hence, $x(t) - x_{0i}(t) = \text{const}$. Since, however, the physical structure does not have to be preserved in a mathematical model used only for control system design, the model is simplified by setting $x_{0i} = x$. Similar statements may be made for Fig. 2.5b, so $\theta_{0i} = \theta$. Assuming Hooke's law holds, the spring forces and torques are respectively,

$$f_{\mathrm{s}i} = K_{si}\,(x_i - x) \quad \text{and} \quad \gamma_{\mathrm{s}i} = K'_{si}\,(\theta_i - \theta)\,, \quad i = 1, 2, \ldots, n. \tag{2.52}$$

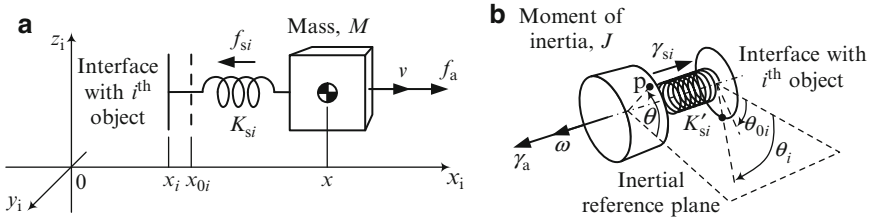


Fig. 2.5 Interaction of a mass with other objects via springs. (**a**) Translational. (**b**) Rotational

2.2.2.7 Inertial Force and Torque

The component of the opposing force or torque due to an *additional* rigid body fixed to either rigid body of Fig. 2.3 is given by

$$f_{\mathrm{I}} = M_{\mathrm{I}}\dot{v} \quad \text{or} \quad \gamma_{\mathrm{I}} = J_{\mathrm{I}}\dot{\omega} \tag{2.53}$$

where M_{I} and J_{I} are, respectively, the mass and moment of inertia of the additional rigid body. An example is an inertial mechanical load bolted to an electric motor.

2.2.2.8 Lagrangian Mechanics

Controlled mechanisms sometimes have several degrees of freedom of motion and, for modelling purposes, can be considered as a set of interconnected rigid bodies, possibly including springs, to which are applied actuator forces and torques. An example is the jointed-arm robotic manipulator introduced in Chap. 12. Derivation of the differential equations of motion of such systems using the force and torque balance methods introduced in Sect. 2.2.2.3 can be very laborious and time consuming, particularly if the number of chosen translational and rotational position coordinates exceeds the number of degrees of freedom, in which case the mechanical constraints have to be carefully incorporated. This task may be greatly eased, however, by applying Lagrangian mechanics [3]. First, the Lagrangian is defined as

$$L = T - V \tag{2.54}$$

where T is the total kinetic energy and V is the total potential energy. Then, if the ith mechanical displacement (either translational or rotational) corresponding to the ith degree of freedom is denoted by q_i, the equations of motion are given by:

$$\frac{d}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{q}_i}\right) - \frac{\partial L}{\partial q_i} + f\left(\dot{q}_i\right) = \tau_i, \quad i = 1, 2, \ldots, d, \tag{2.55}$$

where τ_i is the external force or torque and $f\left(\dot{q}_i\right)$ is the friction force or torque associated with the ith degree of freedom.

As an example, the equations of motion will be derived for the single-degree-of-freedom translational and rotational examples of Fig. 2.5. First, consider Fig. 2.5a. Let $x = q_1$. Then $\dot{q}_1 = v$. Also $\tau_1 = f_{\mathrm{a}}$. The kinetic energy is then

$$T = \frac{1}{2}M v^2 = \frac{1}{2}M \dot{q}_1^2 \tag{2.56}$$

and the potential energy is

$$V = \frac{1}{2}K_s[x - (x_i - x_{0i})]^2 = \frac{1}{2}K_s[q_1 - (x_i - x_{0i})]^2. \tag{2.57}$$

The Lagrangian is therefore

$$L = T - V = \frac{1}{2}\left\{M\dot{q}_1^2 - K_s[q_1 - (x_i - x_{0i})]^2\right\}. \tag{2.58}$$

To derive the equation of motion from (2.55), in this example, there is no friction, so

$$\frac{\partial L}{\partial \dot{q}_1} = M\dot{q}_1 \Rightarrow \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{q}_1}\right) = M\ddot{q}_1 \quad \text{and} \quad \frac{\partial L}{\partial q_1} = -K_s\left[q_1 - (x_i - x_{0i})\right]. \tag{2.59}$$

Hence (2.55) yields

$$M\ddot{q}_1 + K_s\left[q_1 - (x_i - x_{0i})\right] = \tau_1 \tag{2.60}$$

Noting that $\dot{x} = v$, (2.60) may be written

$$M\ddot{x} + K_s\left[x - (x_i - x_{0i})\right] = f_a. \tag{2.61}$$

The rotational example of Fig. 2.5b is similar. Letting $\theta = q_1$, $\tau_1 = \gamma_a$ and noting $\dot{\theta} = \omega$, the kinetic and potential energies are $T = \frac{1}{2}J\omega^2 = \frac{1}{2}J\dot{q}_1^2$ and $V = \frac{1}{2}K_s[\theta - (\theta_i - \theta_{0i})]^2 = \frac{1}{2}K_s[q_1 - (\theta_i - \theta_{0i})]^2$. Applying the Lagrangian method then yields

$$J\ddot{\theta} + K_s\left[\theta - (\theta_i - \theta_{0i})\right] = \gamma_a. \tag{2.62}$$

Another appropriate example is the cart and inverted pendulum mechanism sometimes used to demonstrate control techniques, shown in Fig. 2.6.

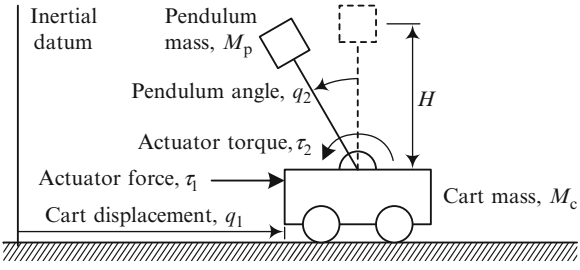


Fig. 2.6 Cart and pole mechanism

In this case,

$$L = T - V = \frac{1}{2}\left\{M_c\dot{q}_1^2 + M_p\left[(\dot{q}_1 - \dot{q}_2 H\cos(q_2))^2 + (\dot{q}_2 H\sin(q_2))^2\right]\right\} - M_p g H\cos(q_2)\,. \tag{2.63}$$

Again, the friction will be assumed negligible. Then applying (2.55) to obtain the equations of motion yields the following.

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{q}_1}\right) - \frac{\partial L}{\partial q_1} = \tau_1 \Rightarrow \frac{\mathrm{d}}{\mathrm{d}t}\left[M_c\dot{q}_1 + M_p(\dot{q}_1 - \dot{q}_2 H\cos(q_2))\right] - 0 = \tau_1 \Rightarrow\\ &\left(M_c + M_p\right)\ddot{q}_1 - M_p H\ddot{q}_2\cos(q_2) + M_p H\dot{q}_2^2\sin(q_2) = \tau_1\end{aligned} \tag{2.64}$$

and

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{q}_2}\right) - \frac{\partial L}{\partial q_2} = \tau_2 \Rightarrow\\ &\frac{\mathrm{d}}{\mathrm{d}t}\left\{-M_p\left[(\dot{q}_1 - \dot{q}_2 H\cos(q_2))\,H\cos(q_2) - \dot{q}_2 H^2\sin^2(q_2)\right]\right\}\\ &\quad + M_p g H\sin(q_2) = \tau_2 \Rightarrow\\ &\frac{\mathrm{d}}{\mathrm{d}t}\left\{M_p\left[\dot{q}_2 H^2 - \dot{q}_1 H\cos(q_2)\right]\right\} + M_p g H\sin(q_2) = \tau_2 \Rightarrow\\ &\quad M_p\left[H^2\ddot{q}_2 + H\dot{q}_1\dot{q}_2\sin(q_2) - H\ddot{q}_1\cos(q_2) + gH\sin(q_2)\right] = \tau_2\end{aligned} \tag{2.65}$$

Many of the terms in these equations would be difficult to deduce by inspection of Fig. 2.6 and application of the force and torque balance method.

2.2.3 Two Basic Mechanical Components

2.2.3.1 Gear Trains and Referred Mechanical Parameters

Many controlled mechanisms such as robot joint actuators employ gear trains to match the motor output to the mechanical load regarding the torque and speed requirements. A gear train is a set of several toothed wheels, i.e. gear wheels, that mesh with one another. For incorporation in a plant model for control system design, it is sufficient to represent a gear train comprising two or more gear wheels using an equivalent train of just two wheels as shown in Fig. 2.7.

Here, R_1 and R_2 are, respectively, the radii of the input and output wheels, γ_1 and γ_2 are the input and output torques, ω_1 and ω_2 are the angular velocities of the input and output shafts, while θ_1 and θ_2 are the corresponding angles of rotation. Starting with points, p_1 and p_2, on the wheel peripheries that are coincident with the point of

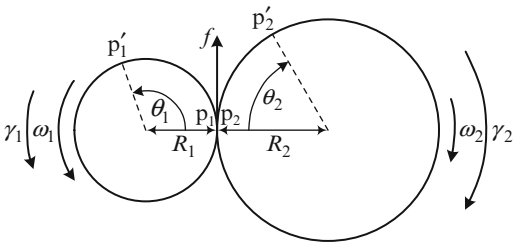


Fig. 2.7 Basic two-wheel representation of a gear system

contact, as the wheels rotate, these points move to new positions, $\mathrm{p}_1^{'}$ and $\mathrm{p}_2^{'}$, on arcs through the same distance, d, requiring

$$d = R_1\theta_1 = R_2\theta_2. \tag{2.66}$$

Differentiating (2.66) and letting $\dot{\theta}_1 = \omega_1$ and $\dot{\theta}_2 = \omega_2$ then yields

$$\omega_1 R_1 = \omega_2 R_2 \tag{2.67}$$

The *gear ratio* is defined as

$$G = \frac{\omega_1}{\omega_2} = \frac{R_2}{R_1}. \tag{2.68}$$

This is a basic parameter in modelling a mechanical system containing a gear train.

The tangential force, f, at the wheel interface produced by the applied torque, γ_1, satisfies $\gamma_1 = fR_1$. This also gives rise to an output torque of $\gamma_2 = fR_2$. Hence

$$f = \frac{\gamma_1}{R_1} = \frac{\gamma_2}{R_2} \Rightarrow \gamma_2 = \frac{R_2}{R_1}\gamma_1.$$

i.e.

$$\gamma_2 = G\gamma_1. \tag{2.69}$$

So if $R > 1$, the gear train achieves torque amplification. The mechanical input power, $\gamma_1\omega_1$, may be expressed in terms of γ_2 and ω_2 using (2.69) and (2.68) as follows.

$$\gamma_1\omega_1 = \frac{1}{G}\gamma_2 G\omega_2 = \gamma_2\omega_2. \tag{2.70}$$

This means that the gear system transmits mechanical power with zero loss, indicating that the model is of an ideal gear system. Power losses in a real system, however, may be modelled using viscous friction parameters, as shown below.

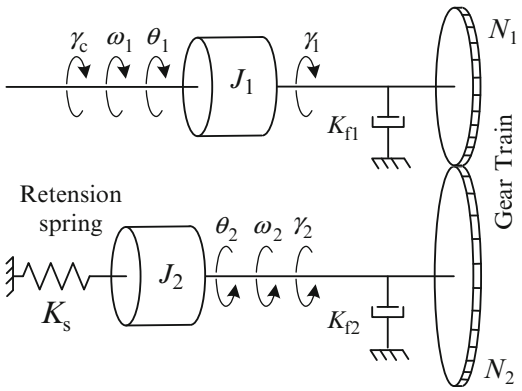


Fig. 2.8 Mechanical system with viscous friction and inertia containing a gear system

It may be observed that the gear system model described above is similar to that of an ideal electrical transformer, in that ω_1 and ω_2 are analogous to the primary and secondary currents, γ_1 and γ_2 are analogous to the primary and secondary voltages, and R_1 and R_2 are analogous to the numbers of primary and secondary turns. Hence, a gear system may be regarded as a mechanical transformer. Indeed, the viscous friction and inertial components of a mechanism of which the gear system is a part may be referred to either the input shaft side or the output shaft side to simplify the model, in a similar way to referring the inductive and resistive components of an electrical circuit to either the primary side or the secondary side of a transformer. Consider the mechanical system shown in Fig. 2.8.

This is a single-degree-of-freedom mechanism consisting of two balanced masses with moments of inertia, J_1 and J_2, connected by a gear train. A control torque, γ_1, is applied to mass 1 (the torque actuator not being included in this example) and the movement of the system is restrained by a torsion spring attached to mass 2. Also N_1 and N_2 are the numbers of gear teeth. Then the gear ratio is

$$G = N_2/N_1. \tag{2.71}$$

The torque balance equations for sides 1 and 2 of the system are

$$J_1\ddot{\theta}_1 + K_{f1}\dot{\theta}_1 = \gamma_c - \gamma_1 \tag{2.72}$$

$$J_2\ddot{\theta}_2 + K_{f2}\dot{\theta}_2 + K_s\theta_2 = \gamma_2 \tag{2.73}$$

It follows from (2.66) and (2.68) that

$$\theta_1 = G\theta_2 \tag{2.74}$$

and this, together with (2.69). completes the model by connecting (2.72) and (2.73). It also enables the model to be simplified by referring all the quantities to side 1 or side 2 of the gear train.

For side 1, it is necessary to substitute for θ_2 and γ_2 in (2.73) using, respectively, (2.74) and (2.69). Thus

$$J_2 \frac{1}{R} \ddot{\theta}_1 + K_{f2} \frac{1}{R} \dot{\theta}_1 + K_s \frac{1}{R} \theta_1 = G \gamma_1. \tag{2.75}$$

Then substituting for γ_1 in (2.72) using (2.75) yields

$$J_1 \ddot{\theta}_1 + K_{f1} \dot{\theta}_1 = \gamma_c - \left(J_2 \frac{1}{G^2} \ddot{\theta}_1 + K_{f2} \frac{1}{G^2} \dot{\theta}_1 + K_s \frac{1}{G^2} \theta_1 \right) \tag{2.76}$$

which may be written

$$J_{r1} \ddot{\theta}_1 + K_{fr1} \dot{\theta}_1 + K_{sr1} \theta_1 = \gamma_c, \tag{2.77}$$

where

$$J_{r1} = J_1 + \frac{1}{G^2} J_2, \; K_{fr1} = K_{f1} + \frac{1}{G^2} K_{f2} \text{ and } K_{sr1} = \frac{1}{G^2} K_s. \tag{2.78}$$

Then (2.77) is the simplified model, which is equivalent to a single mass moving against viscous friction and a torsion spring without a gear train, whose parameters, J_{11}, K_{f11} and K_{s1}, are, respectively, the moment of inertia, the viscous friction coefficient and the spring constant referred to side 1 of the gear train.

For the alternative of referring all the parameters to side 2 of the gear train, it is possible to start with (2.76) and substitute for θ_1 using (2.74), which gives

$$J_1 G \ddot{\theta}_2 + K_{f1} G \dot{\theta}_2 = \gamma_c - \left(J_2 \frac{1}{G} \ddot{\theta}_2 + K_{f2} \frac{1}{G} \dot{\theta}_2 + K_s \frac{1}{G} \theta_2 \right) \tag{2.79}$$

which may be written

$$J_{r2} \ddot{\theta}_2 + K_{fr2} \dot{\theta}_2 + K_{sr2} \theta_2 = G \gamma_c, \tag{2.80}$$

where

$$J_{r2} = J_2 + G^2 J_1, \; K_{fr2} = K_{f2} + G^2 K_{f1} \text{ and } K_{sr2} = K_s. \tag{2.81}$$

Equations (2.78) and (2.81) are similar to those referring reactive and resistive components to the primary or secondary circuits of an electrical transformer.

2.2.3.2 Hard Stops

Controlled mechanisms often have mechanical hard stops, limiting the range of movement to lie between maximum and minimum values. It is necessary to model these hard stops for simulation purposes if there is any likelihood of them being reached in the application under study. It will be assumed that the part of the mechanism constrained by the hard stops may be modelled as a rigid body when not in contact with these stops and has one degree of freedom of movement, which may be rotational or translational. Let the differential equation of motion of this body be

$$\ddot{x} = b\,(u_{\mathrm{b}} + u_{\mathrm{h}}) \tag{2.82}$$

where x is either the translational or rotational displacement, $b = 1/M$ or $b = 1/J$, where M is the body mass, J is the body moment of inertia about the axis of rotation, u_{b} is the total torque or force applied to the body without contacting either hard stop and u_{h} is the additional force or torque acting on the body as a result of contact with either of the two hard stops. Upon contacting a hard stop, small elastic deformations will occur in the body and the hard stop material. This will be modelled as a very stiff spring. Also a small proportion of the kinetic energy of the body will be lost as heat dissipated in the body due to the nature of its material or energy transfer to the structure upon which the hard stops are mounted. This energy dissipation, if considered significant, will be modelled as viscous damping. Thus

$$e_{\mathrm{h}} = \begin{cases} x_{\max} - x, x > x_{\max} \\ 0, x_{\min} \le x \le x_{\max} \\ x_{\min} - x, x < x_{\min} \end{cases}, u_{\mathrm{h}} = K_{\mathrm{sh}} e_{\mathrm{h}} + K_{\mathrm{vh}} \dot{e}_{\mathrm{h}}, \tag{2.83}$$

where K_{sh} and K_{vh} are, respectively, the spring constant and viscous damping coefficient representing the elastic deformation and the energy loss during the stop contact. Figure 2.9 shows a block diagram of the model based on (2.82) and (2.83).

The 'max' and 'min' functions shown are as in SIMULINK® and are defined as

$$\max(p,\ q) \triangleq \begin{cases} p,\ p \ge q \\ q,\ p < q \end{cases} \quad \text{and} \quad \min(p, q) \triangleq \begin{cases} p, p \le q \\ q, p > q \end{cases}. \tag{2.84}$$

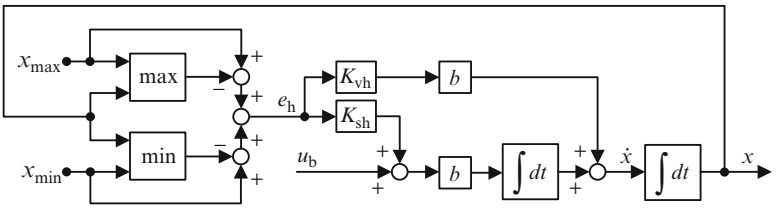


Fig. 2.9 Hard stop model for simulation

Also, the injection of the signal, $bK_{\mathrm{vh}}e_{\mathrm{h}}$, between the two integrators realises the term, $K_{\mathrm{vh}}\dot{e}_{\mathrm{h}}$, in (2.83), thereby avoiding the implementation of the differentiation.

2.2.4 Modelling for Vehicle Attitude and Position Control

2.2.4.1 Three-Axis Rotational Dynamics

An unconstrained rigid body, usually representing a vehicle such as a spacecraft or underwater vehicle has three rotational degrees of freedom as discussed in Sect. 2.2.2.2. It will be assumed that attitude control actuators are mounted on the body that produce torque components, $\gamma_{\mathrm{b}x}$, $\gamma_{\mathrm{b}y}$ and $\gamma_{\mathrm{b}z}$, about the body-fixed axes, x_{b}, y_{b} and z_{b}. The dynamic subsystem is obtained by equating the net applied torque to the first derivative of the body angular momentum vector. This is far more involved than the single-degree-of-freedom case due to the derivative of a vector having two parts, the rate of change of magnitude and the rate of change of direction. For the single-degree-of-freedom case, only the rate of change of magnitude of the angular momentum vector is needed, since it does not change direction, yielding (2.33). For an unconstrained rigid body, however, the motion in all three rotational degrees of freedom occurs and the theory of dynamics is needed [4]. Let the instantaneous angular velocity vector be

$$\boldsymbol{\omega} = \omega_x \widehat{i} + \omega_y \widehat{j} + \omega_z \widehat{k}, \tag{2.85}$$

where $\hat{\imath}$, $\hat{\jmath}$ and $\widehat{k}$ are the unit vectors directed along the mutually orthogonal body-fixed axes, x_{b}, y_{b} and z_{b} and ω_x, ω_y and ω_z are the angular velocity components along these axes. Let the body angular momentum be similarly represented as

$$\mathbf{L}_{\mathrm{b}} = l_{\mathrm{b}x} \widehat{i} + l_{\mathrm{b}y} \widehat{j} + l_{\mathrm{b}z} \widehat{k}, \tag{2.86}$$

Using the same notation, the net torque vector acting on the body is

$$\boldsymbol{\gamma} = \boldsymbol{\gamma}_{\mathrm{a}} - \boldsymbol{\gamma}_{\mathrm{o}}, \tag{2.87}$$

where

$$\boldsymbol{\gamma}_{\mathrm{a}} = \gamma_{\mathrm{a}x} \widehat{i} + \gamma_{\mathrm{a}y} \widehat{j} + \gamma_{\mathrm{a}z} \widehat{k} \tag{2.88}$$

is the actuator torque vector and

$$\boldsymbol{\gamma}_{\mathrm{o}} = \gamma_{\mathrm{o}x} \widehat{i} + \gamma_{\mathrm{o}y} \widehat{j} + \gamma_{\mathrm{o}z} \widehat{k} \tag{2.89}$$

is the opposing torque vector that could be due, for example, to hydrodynamic drag in an underwater vehicle or solar radiation pressure in the case of a space satellite. The dynamics equation can then be written as

$$\dot{\mathbf{L}}_{\mathrm{b}} + \boldsymbol{\omega} \wedge \mathbf{L}_{\mathrm{b}} = \boldsymbol{\gamma}_{\mathrm{a}} - \boldsymbol{\gamma}_0 \tag{2.90}$$

where

$$\dot{\mathbf{L}}_{\mathrm{b}} \triangleq \dot{l}_{\mathrm{b}x}\widehat{i} + \dot{l}_{\mathrm{b}y}\widehat{j} + \dot{l}_{\mathrm{b}z}\widehat{k} \tag{2.91}$$

is the rate of change of the magnitude of the angular momentum vector and the second term on the LHS of (2.90) is the rate of change of direction of the angular momentum vector. This is called the gyroscopic torque component since it is responsible for the behaviour of a gyroscope. A familiar example of gyroscopic torques at work is the prevention of a bicycle in motion falling over due to the wheel angular momentum vectors.

It is important to note that (2.90) is valid only for actuators such as thrusters that do not have angular momentums affecting the motion of the body. In spacecraft, however, reaction wheels or control moment gyros are commonly employed. These actuators accrue their own angular momentums through their principle of operation.

A reaction wheel consists of an electric motor with the stator bolted to the spacecraft body, directly driving a balanced flywheel. When the motor develops torque, the wheel undergoes an angular acceleration and its angular momentum magnitude therefore changes. The equal and opposite reaction torque acts on the spacecraft body via the stator, controlling the attitude as required. This also changes the angular momentum of the spacecraft body by an equal and opposite amount to the change in the wheel angular momentum. This is due to the principle of conservation of angular momentum, which states that the total angular momentum of a mechanical system is constant if no external torque acts on it. This is also called a conservative system because the angular momentum is conserved. Since a set of reaction wheels on a spacecraft effects momentum exchange between the spacecraft body and the wheels, these actuators are called momentum exchange actuators. At least three reaction wheels are needed to control the three rotational degrees of freedom, usually a set of four with their spin axes arranged so that any combination of three wheels can be used to maintain the mission with one wheel failure.

The control moment gyro (CMG) consists of a flywheel running at constant speed mounted in a single or a two-axis gimbal system, each gimbal axis equipped with an electromagnetic transducer that provides an attitude control torque component. Various configurations of control moment gyros can be employed to achieve controllability of the three rotational degrees of freedom. In this case, the wheel angular momentum magnitudes are constant but attitude control torques from the electromagnetic transducers produce equal and opposite torques acting at right angles to the wheel spin axes which change the direction of the angular momentum

vector of each CMG. Again, during attitude control manoeuvres, angular momentum is exchanged between the spacecraft body and the set of CMGs, the total angular momentum remaining constant if no external torques are acting.

If external torques do act on a spacecraft equipped with momentum exchange actuators, then the fundamental equation of rotational dynamics states that this is equal to the rate of change of the total angular momentum. If the attitude control maintains the spacecraft body stationary with respect to inertial space, then the actuators absorb the angular momentum until they reach a saturation condition. Either one of the reaction wheels reaches a maximum speed limit or one of the CMG gimbals reaches an angular limit. Then the stored angular momentum has to be removed by transferral to molecules of exhaust emission of a set of thrusters.

A model for control system simulation and design can be formed that is independent of the type of momentum exchange actuator and the configuration of the actuator set. This is done by expressing the actuator angular momentum vector as

$$\mathbf{L}_{\mathrm{a}} = l_{\mathrm{a}x}\widehat{i} + l_{\mathrm{a}y}\widehat{j} + l_{\mathrm{a}z}\widehat{k} \tag{2.92}$$

and then replacing the dynamic subsystem (2.90) with the following.

$$\dot{\mathbf{L}}_{\mathrm{b}} + \boldsymbol{\omega} \wedge (\mathbf{L}_{\mathrm{b}} + \mathbf{L}_{\mathrm{a}}) = \boldsymbol{\gamma}_{\mathrm{a}} - \boldsymbol{\gamma}_0, \;\; \dot{\mathbf{L}}_{\mathrm{a}} = -\boldsymbol{\gamma}_{\mathrm{a}}, \tag{2.93}$$

Next, to render the dynamic subsystems (2.90) and (2.93) useful for attitude control system design, they should be reformulated in the matrix–vector form with the vectors represented as 3×1 column vectors. First, however, the cross products will be expanded, noting that

$$\tilde{i} \wedge \tilde{i} = \tilde{j} \wedge \tilde{j} = \tilde{k} \wedge \tilde{k} = 0;\; \tilde{i} \wedge \tilde{j} = \tilde{k};\; \tilde{j} \wedge \tilde{i} = -\tilde{k};$$
$$\tilde{j} \wedge \tilde{k} = \tilde{i};\; \tilde{k} \wedge \tilde{j} = -\tilde{i};\; \tilde{k} \wedge \tilde{i} = \tilde{j};\; \tilde{i} \wedge \tilde{k} = -\tilde{j}$$

Then in (2.90)

$$\begin{aligned} \boldsymbol{\omega} \wedge \mathbf{L}_{\mathrm{b}} &= \left(\omega_x\widehat{i} + \omega_y\widehat{j} + \omega_z\widehat{k}\right) \wedge \left(l_{\mathrm{b}x}\widehat{i} + l_{\mathrm{b}y}\widehat{j} + l_{\mathrm{b}z}\widehat{k}\right) \\ &= \left(\omega_y l_{\mathrm{b}z} - \omega_z l_{\mathrm{b}y}\right)\widehat{i} + \left(\omega_z l_{\mathrm{b}x} - \omega_x l_{\mathrm{b}z}\right)\widehat{j} + \left(\omega_x l_{\mathrm{b}y} - \omega_y l_{\mathrm{b}x}\right)\widehat{k} \end{aligned} \tag{2.94}$$

Then the matrix–vector form of (2.90) follows.

$$\begin{bmatrix} \dot{l}_{\mathrm{b}x} \\ \dot{l}_{\mathrm{b}y} \\ \dot{l}_{\mathrm{b}z} \end{bmatrix} + \begin{bmatrix} 0 & -\omega_z & \omega_y \\ \omega_z & 0 & -\omega_x \\ -\omega_y & \omega_x & 0 \end{bmatrix} \begin{bmatrix} l_{\mathrm{b}x} \\ l_{\mathrm{b}y} \\ l_{\mathrm{b}z} \end{bmatrix} = \begin{bmatrix} \gamma_{\mathrm{a}x} \\ \gamma_{\mathrm{a}y} \\ \gamma_{\mathrm{a}z} \end{bmatrix} - \begin{bmatrix} \gamma_{\mathrm{o}x} \\ \gamma_{\mathrm{o}y} \\ \gamma_{\mathrm{o}z} \end{bmatrix}, \tag{2.95}$$

which, in compact form, may be written

$$\dot{\mathbf{L}}_{\mathrm{b}} + \boldsymbol{\Omega}_3 \mathbf{L}_{\mathrm{b}} = \boldsymbol{\gamma}_{\mathrm{a}} - \boldsymbol{\gamma}_{\mathrm{o}} \tag{2.96}$$

Similarly, (2.93) becomes

$$\dot{\mathbf{L}}_\mathrm{b} + \boldsymbol{\Omega}_3 [\mathbf{L}_\mathrm{b} + \mathbf{L}_\mathrm{a}] = \boldsymbol{\gamma}_\mathrm{a} - \boldsymbol{\gamma}_\mathrm{o}, \ \dot{\mathbf{L}}_\mathrm{a} = -\boldsymbol{\gamma}_\mathrm{a} \tag{2.97}$$

To obtain the dynamic subsystem (2.97) in a form equivalent to (2.33), the moment of inertia matrix, often called the moment of inertia tensor [4], is needed. This is

$$\mathbf{J} = \begin{bmatrix} J_{xx} & J_{xy} & J_{xz} \\ J_{yx} & J_{yy} & J_{yz} \\ J_{zx} & J_{zy} & J_{zz} \end{bmatrix} = \lim_{\substack{n \to \infty \\ \delta m_i \to 0}} \sum_{i=1}^{n} \delta m_i \begin{bmatrix} y_i^2 + z_i^2 & x_i y_i & x_i z_i \\ y_i x_i & x_i^2 + z_i^2 & y_i z_i \\ z_i x_i & z_i y_i & x_i^2 + y_i^2 \end{bmatrix}. \tag{2.98}$$

The diagonal terms are the moments of inertia of the body about the axes, x_b, y_b and z_b, while the off-diagonal terms are the products of inertia. The rigid body is divided into a large number of elements of mass, m_i with coordinates, x_i, y_i and z_i. Then the number of elements is allowed to become infinitely large, each of infinitesimal mass. The angular momentum vector is the product of the moment of inertia matrix and the angular velocity vector. Thus

$$\mathbf{L}_\mathrm{b} = \mathbf{J}\boldsymbol{\omega}. \tag{2.99}$$

Then (2.95), applicable when using thruster-based actuators, becomes

$$\mathbf{J}\dot{\boldsymbol{\omega}} + \boldsymbol{\Omega}_3 \mathbf{J}\boldsymbol{\omega} = \boldsymbol{\gamma}_\mathrm{a} - \boldsymbol{\gamma}_\mathrm{o} \tag{2.100}$$

and (2.97), applicable when using momentum exchange actuators, becomes

$$\mathbf{J}\dot{\boldsymbol{\omega}} + \boldsymbol{\Omega}_3 [\mathbf{J}\boldsymbol{\omega} + \mathbf{L}_\mathrm{a}] = \boldsymbol{\gamma}_\mathrm{a} - \boldsymbol{\gamma}_\mathrm{o}, \ \dot{\mathbf{L}}_\mathrm{a} = -\boldsymbol{\gamma}_\mathrm{a}. \tag{2.101}$$

2.2.4.2 Basic Three-Axis Rotational Kinematics

As already pointed out in Sect. 2.2.2.2, there are several sets of attitude coordinates that can be chosen for a rigid body. These are attitude representations [5]. For each attitude representation, there is a set of kinematic differential equations [KDEs] relating the derivatives of the attitude coordinates to themselves and the body angular velocity components, ω_x, ω_y and ω_z, defined *i*th in subsection 2.2.4.1. One of the twelve different attitude representations introduced in Sect. 2.2.2.2 based on three successive rotations about the body-fixed axes could be chosen, each with a different set of KDEs. These, however, have a common drawback. Consider, for example, the gimbal mechanism of Fig. 2.1e for orientations of the central cube requiring the two gimbal frames to be coplanar. Then the mechanism loses one degree of freedom of motion. This condition is referred to as gimbal lock. As will be seen, this manifests as a singularity in the corresponding set of KDEs. Fortunately,

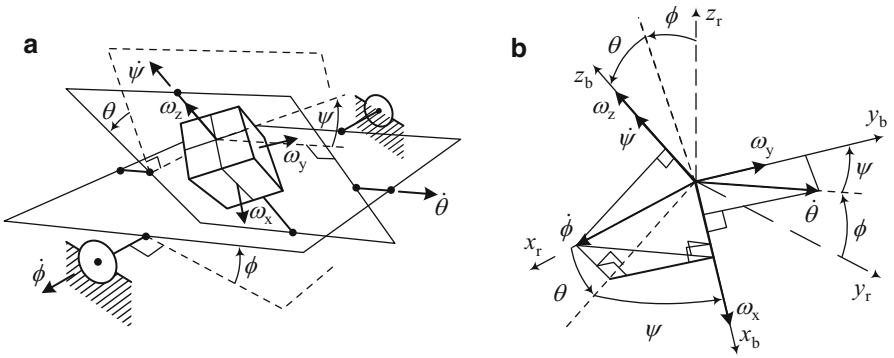


Fig. 2.10 Diagrams for derivation of kinematic differential equations for Fig. 2.2. (**a**) Equivalent gimbal system. (**b**) Angular velocity vector diagram

in addition to the 12 attitude representations already mentioned, there exist others whose sets of rotational KDEs do not exhibit singularities. One of these, based on the quaternion, is presented in the following subsection as it is suited to control system design.

Before introducing the quaternion-based attitude representation, the KDEs for the attitude representation of Fig. 2.2 will be derived to demonstrate the gimbal lock condition and the singularity. Figure 2.10a shows the fictitious gimbal system for this attitude representation that enables it to be visualised.

Figure 2.10b shows the geometry of the body angular velocity vectors and the gimbal joint angular velocity vectors. The gimbal joint angles are the attitude coordinates. The equations for ω_x, ω_y and ω_z, produced by given gimbal angular velocities, $\dot{\phi}$, $\dot{\theta}$ and $\dot{\psi}$, are obtained by studying Fig. 2.10b and are as follows.

$$\omega_x = \dot{\phi}\cos(\theta)\cos(\psi) + \dot{\theta}\sin(\psi) \tag{2.102}$$

$$\omega_y = \dot{\theta}\cos(\psi) - \dot{\phi}\cos(\theta)\sin(\psi) \tag{2.103}$$

$$\omega_z = \dot{\psi} + \dot{\phi}\sin(\theta) \tag{2.104}$$

The required KDEs are then obtained by solving (2.102), (2.103) and (2.104) for $\dot{\phi}$, $\dot{\theta}$ and $\dot{\psi}$. Hence (2.102) $\times \cos(\psi) -$ (2.103) $\times \sin(\psi)$ yields

$$\begin{gathered}\dot{\phi}\cos(\theta)\cos^2(\psi) + \dot{\theta}\sin(\psi)\cos(\psi) - \dot{\theta}\cos(\psi)\sin(\psi) + \dot{\phi}\cos(\theta)\sin^2(\psi) \\ = \omega_x\cos(\psi) - \omega_y\sin(\psi) \Rightarrow \dot{\phi}\cos(\theta) = \omega_x\cos(\psi) - \omega_y\sin(\psi) \Rightarrow \\ \dot{\phi} = \tfrac{1}{\cos(\theta)}\left[\omega_x\cos(\psi) - \omega_y\sin(\psi)\right]\end{gathered} \tag{2.105}$$

Similarly, (2.102) × sin(ψ) + (2.103) × cos(ψ) yields

$$\begin{aligned}\dot{\phi}\cos(\theta)\cos(\psi)\sin(\psi) + \dot{\theta}\sin^2(\psi) + \dot{\theta}\cos^2(\psi) - \dot{\phi}\cos(\theta)\sin(\psi)\cos(\psi) \\ = \omega_x \sin(\psi) + \omega_y \cos(\psi) \Rightarrow \\ \dot{\theta} = \omega_x \sin(\psi) + \omega_y \cos(\psi)\end{aligned} \tag{2.106}$$

Finally, substituting for in (2.104) using (2.105) and making $\dot{\psi}$ the subject of the resulting equation yields

$$\dot{\psi} = \omega_z - \left[\omega_x \cos(\psi) - \omega_y \sin(\psi)\right] \tan(\theta). \tag{2.107}$$

The set of three KDEs are (2.105), (2.106) and (2.107), and they constitute a *possible* kinematic subsystem for the rigid-body rotational model. The singularity is evident on the RHS of (2.105) and (2.107), as $\dot{\phi} \to \infty$ and $\dot{\psi} \to \infty$ as $\theta \to \pi/2$ provided $\omega_x \cos(\psi) - \omega_y \sin(\psi) \neq 0$, which will usually be true. Observing Fig. 2.1a, the inner gimbal is in the same plane as the outer gimbal for $\theta = \pi/2$ and the body cannot be rotated about an axis perpendicular to the gimbal plane for this condition. This is gimbal lock. For applications such as surface ships, civil airliners and mobile robots, however, this situation is tolerable as the attitude of the vehicle is limited. For applications, such as spacecraft in which the attitude is unlimited, singularity-free attitude representations exist, two of which are presented in the following subsection.

2.2.4.3 Singularity-Free Three-Axis Rotational Kinematics

Two sets of kinematic differential equations are derived from the first principles in Appendix A2 that, in contrast to the basic kinematic differential equations of Sect. 2.2.4.2, are free of singularities and trigonometric functions. The first is the set of direction cosine-based kinematic differential equations, as follows.

$$\begin{bmatrix} \dot{c}_{xx} & \dot{c}_{xy} & \dot{c}_{xz} \\ \dot{c}_{yx} & \dot{c}_{yy} & \dot{c}_{yz} \\ \dot{c}_{zx} & \dot{c}_{zy} & \dot{c}_{zz} \end{bmatrix} = \begin{bmatrix} 0 & \omega_z & -\omega_y \\ -\omega_z & 0 & \omega_x \\ \omega_y & -\omega_x & 0 \end{bmatrix} \begin{bmatrix} c_{xx} & c_{xy} & c_{xz} \\ c_{yx} & c_{yy} & c_{yz} \\ c_{zx} & c_{zy} & c_{zz} \end{bmatrix} \tag{2.108}$$

Here, c_{ij}, are the set of direction cosines of three mutually orthogonal unit vectors fixed in the vehicle body with respect to a set of three mutually orthogonal unit vectors fixed in the frame of reference. Although these constitute nine attitude coordinates, (2.108) obeys six constraint equations that reduce the total number of rotational degrees of freedom to three. In this case, (2.108) can be numerically integrated in a vehicle application and the three attitude coordinates taken as a suitable subset of three direction cosines for control purposes [5].

The second set of singularity-free kinematic differential equations is based on the quaternion. This interesting mathematical entity is introduced and fully discussed in Appendix A2, together with the derivation of the equations as follows.

$$\begin{bmatrix} \dot{q}_0 \\ \dot{q}_1 \\ \dot{q}_2 \\ \dot{q}_3 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 0 & -\omega_x & -\omega_y & -\omega_z \\ \omega_x & 0 & \omega_z & -\omega_y \\ \omega_y & -\omega_z & 0 & \omega_x \\ \omega_z & \omega_y & -\omega_x & 0 \end{bmatrix} \begin{bmatrix} q_0 \\ q_1 \\ q_2 \\ q_3 \end{bmatrix} \tag{2.109}$$

Here, the quaternion components, q_i, are the attitude coordinates which obey the single constraint equation,

$$q_0^2 + q_1^2 + q_2^2 + q_3^2 = 1, \tag{2.110}$$

thereby reducing the total number of rotational degrees of freedom to three, so a suitable subset of these coordinates, usually q_1, q_2 and q_3, can be used for control.

2.2.4.4 Translational Dynamics and Kinematics

In applications such as spacecraft and underwater vehicles, the translational dynamic subsystem is the differential equation relating the control force vector, $\mathbf{f}_\mathrm{c}$, the opposing force vector, $\mathbf{f}_\mathrm{o}$, and the external disturbance force vector, $\mathbf{f}_\mathrm{d}$, to the velocity vector, $\mathbf{v}_\mathrm{r}$, of the centre of mass. The translational kinematic subsystem is the differential equation relating the velocity and position vectors of the centre of mass. These are straightforward but the force vectors are usually formulated in the vehicle body-fixed frame, $(x_\mathrm{b}, y_\mathrm{b}, z_\mathrm{b})$, because the actuators are mounted on the body, while the position and velocity vectors are formulated in the reference frame $(x_\mathrm{r}, y_\mathrm{r}, z_\mathrm{r})$. In this case, the direction cosine matrix, $\mathbf{C}$, discussed in Appendix A2, is needed in the dynamic subsystem to convert the given force components along the body-fixed axes to components along the reference frame axes. Thus, if $\mathbf{f}_\mathrm{r}$ is the net force vector acting on the body centre of mass with components along the reference axes, then

$$\mathbf{f}_\mathrm{r} = \mathbf{C}^\mathrm{T} \left[\mathbf{f}_\mathrm{c} - \mathbf{f}_\mathrm{o} - \mathbf{f}_\mathrm{d} \right]. \tag{2.111}$$

The basic dynamics equation is then obtained by equating the rate of change of linear momentum to $\mathbf{f}_\mathrm{r}$. Thus,

$$\frac{\mathrm{d}}{\mathrm{d}t} \left(M \mathbf{v}_\mathrm{r} \right) = \mathbf{f}_\mathrm{r} \tag{2.112}$$

where M is the vehicle mass. If M is constant, then the dynamic subsystem equation is obtained by combining (2.111) and (2.112). In the component form, this is

$$M\begin{bmatrix}\dot{v}_{rx}\\ \dot{v}_{ry}\\ \dot{v}_{rz}\end{bmatrix}=\begin{bmatrix}c_{xx} & c_{yx} & c_{zx}\\ c_{xy} & c_{yy} & c_{zy}\\ c_{xz} & c_{yz} & c_{zz}\end{bmatrix}\left[\begin{bmatrix}f_{cx}\\ f_{cy}\\ f_{cz}\end{bmatrix}-\begin{bmatrix}f_{ox}\\ f_{oy}\\ f_{oz}\end{bmatrix}-\begin{bmatrix}f_{dx}\\ f_{dy}\\ f_{dz}\end{bmatrix}\right]. \tag{2.113}$$

The elements, c_{ij}, $i = x,\ y,\ z$, $j = x,\ y,\ z$, would have to be given in terms of the coordinates of the attitude representation for the rotational kinematic subsystem [5].

2.2.5 *Electric Motors*

2.2.5.1 Introduction

Many systems for controlling the position or velocity of a mechanical object employ electric motors as actuators. This trend is increasing. For example, electric drives, comprising motors, associated power electronics and digital processors, are replacing the internal combustion engine for vehicle propulsion and also replacing the hydraulic actuators on some aircraft control surfaces. Motors are therefore important electrical components to understand and model to be able to create plant models. It is important to note that these machines can operate as generators to return kinetic energy stored in controlled mechanisms to the power supply during deceleration, referred to as regenerative operation, but suitably designed power electronic circuits are needed for this purpose. This is an important practical feature of any system employing electric motors, such as an electric vehicle, designed to recycle energy that would otherwise be lost in the form of heat.

The following subsections present models of three electric motor types found in industry, i.e. the DC motor, and the two basic types of AC motor, i.e. the synchronous motor and the induction motor. The DC motor is relatively straightforward to model and incorporate in a plant model. On the other hand, AC motor models used by power systems engineers are usually equivalent circuits based on a sinusoidal power supply voltage at fixed amplitude and frequency, stemming from the era in which variable speed or position control was quite primitive and entailed varying the supply voltage amplitude only. For the effective use as actuators in feedback control systems, both the amplitude and frequency have to be variable. A plant model in which the motor supply voltage amplitude and frequency are input variables would, however, be nonlinear, also requiring the phasing to be varied for bidirectional control. These problems are circumvented in modern electric drives by the method of vector control [6] described in Sect. 2.2.6 that applies software-implemented transformations through which the AC motor appears similar to a DC motor. The induction and synchronous motor models presented in Sects. 2.2.6.4 and 2.2.6.6 have the DC motor-like input and output variables and therefore include the transformations.

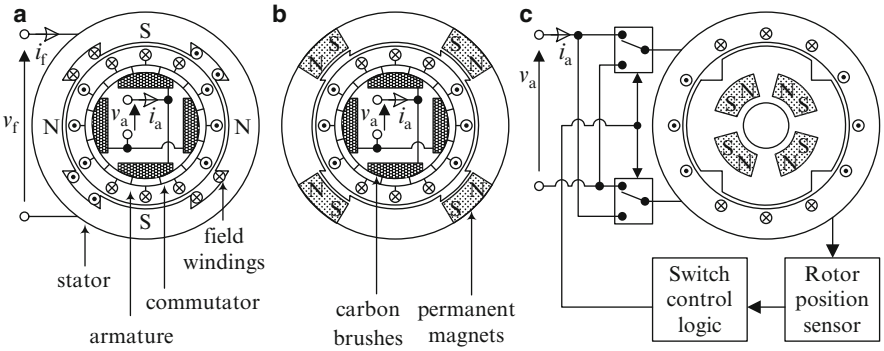


Fig. 2.11 DC motor types. (**a**) Separately excited. (**b**) Permanent magnet. (**c**) Brushless

Models of rotary motors are given, those for linear motors being similar. The models comprise a common mechanical subsystem given in Sect. 2.2.5.3 and specific electrical subsystems given in Sects. 2.2.5.4, 2.2.6.4 and 2.2.6.6.

2.2.5.2 The Three Basic Motor Types

Basic motor descriptions are given in this subsection in sufficient detail to enable the unfamiliar reader to understand the models presented subsequently. It should be noted that the detailed configurations, proportions and practical features are not given here but are available in specialist texts such as [8] and design aspects together with the underlying electromagnetic theory are covered by texts such as [9].

The basic forms of DC motor shown in Fig. 2.11a, b comprise a stator of magnetic material in which a cylindrical armature, also of magnetic material rotates, separated by a small air gap. The stators of large DC motors are configured to produce a magnetic field pattern with alternate North and South poles around the cylindrical air gap as illustrated in Fig. 2.11a, b. In relatively large motors rated in the Megawatt region, such as employed in steel rolling mills, the magnetic field is produced by applying a voltage, v_f, to drive a current, i_f, through field windings in the stator, as shown in Fig. 2.11a. In much smaller DC motors rated in the Kilowatt region and below, such as used in small positioning mechanisms, the magnetic field is produced by permanent magnets as shown in Fig. 2.11b. The crosses on the conductor sections indicate current direction away from the observer while the dots indicate current direction towards the observer.

The armature contains a set of conductors through which a controlled current, i_a, is passed to produce a tangential force, and hence torque, through interaction with the magnetic field. As the armature rotates, when its conductors move from South to North poles and vice versa, the current direction in those conductors is reversed by means of a commutator mounted on the armature shaft, to maintain the torque in the required direction despite the change of direction of the magnetic field as 'seen' by

the moving conductors. The commutator consists of a number of copper segments arranged on an insulated cylinder, on which carbon brushes delivering the armature current are kept in good contact by springs (not shown). The commutator is shown 'inside out' in Fig. 2.11a, b for clarity of illustration. In practice the brushes are placed on the outside of the cylindrical set of commutator segments. In brushless DC motors, which are rated in the Kilowatt region and below, the permanent magnets are mounted in the rotor instead of the stator and the torque producing conductors are mounted in the stator as shown in Fig. 2.11c. The commutation is electronic, being carried out with the aid of a rotor position sensor and switching logic driving power electronic switches. The control variable is the armature voltage, v_a. In the separately excited DC motor, v_f is normally kept constant so that the constant steady-state i_f maintains a constant magnetic flux, but this can be reduced as the speed increases to allow a higher maximum speed by reducing the armature back e.m.f. which, since it opposes the armature voltage, would cause loss of control of the armature current due to the voltage saturation limits.

Synchronous motors and induction motors, sometimes called asynchronous motors, have similar stators of magnetic material as illustrated in Fig. 2.12. As in the DC motor, the stator is configured to produce a field pattern with alternate North and South poles around the cylindrical air gap, but in contrast, this field pattern can also be made to rotate. This is achieved by means of a number of coils whose conductors are distributed in slots around the inside cylindrical surface of the stator. Each coil is referred to as a phase. There are at least two phases, usually three in AC motors used as control actuators.

The distribution of the conductors of each phase is such that if a constant direct current is passed through the coil, a magnetic field pattern results with an air gap flux density that has a nearly sinusoidal variation with angular position around the air gap. There are p cycles of flux density per 360° of angular position variation, the overall field pattern being equivalent to that produced by p bar magnets, each with a North pole and a South pole, as if they were buried in the stator in a symmetrical pattern instead of the coil. The integer, p, is therefore referred to as the number

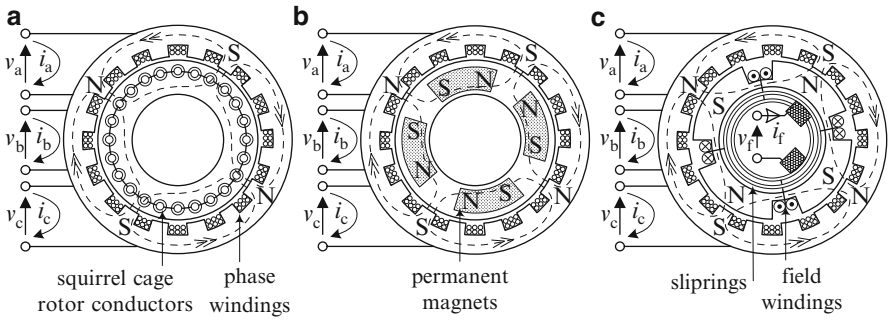


Fig. 2.12 Basic AC motor types. (**a**) Induction (asynchronous motor). (**b**) Permanent magnet synchronous motor (PMSM). (**c**) Separately excited synchronous motor

of pole pairs. The coil of each phase is similar but angularly separated from its neighbour. For a three-phase motor, the mechanical separation is $120°/p$. Then if sinusoidal currents at a frequency of ω_e [rad/s] are driven through the coils, the current in each phase being separated by 120° in electrical angle, then the magnetic field pattern will rotate about the rotor centre at an angular velocity of $\omega_s = \omega_e/p$ [rad/s], called the mechanical synchronous angular frequency, while maintaining its shape.

For all three AC motors, a rotor of magnetic material is placed inside the stator separated by a small air gap. The rotors of induction motors usually contain a set of conductors placed axially in a cylindrical configuration near the periphery. The ends of these conductors are electrically connected as shown. This type of rotor is called a squirrel-cage rotor. The induction motor is therefore similar to a transformer with a short-circuited secondary winding. Indeed, if the rotor is locked, the rotating field pattern generates e.m.f.s in the rotor conductors that give rise to relatively high circulating currents. With the rotor free to move, however, the torque generated by the interaction of the rotor currents with the rotating magnetic field causes the rotor to accelerate until it reaches a constant speed if the supply voltage amplitude and frequency are constant. If the mechanical load is purely inertial with zero friction (hypothetical in practice) then the rotor would reach ω_s [rad/s] for which there would be no relative movement between the magnetic field and the rotor conductors. In a real situation, however, a steady torque would be required to maintain a constant rotor speed, which would be ω_m, where $|\omega_m| < |\omega_s|$, as there has to be relative movement between the magnetic field and rotor conductors, called rotor slip, for there to be e.m.f.s driving currents through the rotor to produce the necessary torque.

In the permanent magnet synchronous motor (PMSM) illustrated in Fig. 2.12b, permanent magnets are buried in the rotor to produce a magnetic field with a set of poles equal in number to the set of poles of the rotating magnetic field. Then unlike poles of the rotating magnetic field attract like poles of the rotor and it is 'dragged' round at an angular velocity of ω_s, so there is no rotor slip. The magnetic field, however, is distorted under a mechanical load resulting in an angular displacement between the stator magnetic field and rotor called the load angle. The PMSMs are rated in the Kilowatt region or below. The separately excited synchronous motors are illustrated in Fig. 2.12c.

2.2.5.3 Mechanical Subsystem

The mechanical part is common to all motor types and effectively comprises the single-degree-of-freedom rotational dynamic and kinematic models of Sect. 2.2.2.3, which are as follows:

$$\begin{aligned} &\text{Dynamical subsystem} : \dot{\omega}_r = \tfrac{1}{J_r}(\gamma_e - \gamma_L) \\ &\text{Kinematic subsystem} : \dot{\theta}_r = \omega_r \end{aligned} \tag{2.114}$$

Here, J_r is the rotor moment of inertia, ω_r is the rotor angular velocity, θ_r is the rotor angle (relative to an inertial frame of reference), γ_e is the electromagnetic torque developed by the motor and γ_L is the load torque given by

$$\gamma_L = \gamma_d + \gamma_o, \tag{2.115}$$

where γ_d is the external disturbance torque and γ_o is the opposing torque defined in Sect. 2.2.2.3.

2.2.5.4 DC Motor Electrical Subsystem

The basic model of a DC motor may be developed by first considering a conductor of length of l metres, carrying a current of i Amperes at right angles to a uniform magnetic field of flux density B Tesla situated on a cylindrical armature at a mean radius of r metres. The torque developed is then

$$\gamma = rBli \quad [\text{Nm}]\,. \tag{2.116}$$

The back e.m.f., which enables the motor to operate as a generator when needed, is

$$e = Blr\,\omega_r [\text{V}] \tag{2.117}$$

It is evident from (2.116) and (2.117) that the torque and back e.m.f. constants are both equal. For a complete DC motor, similar equations hold that are written as

$$\gamma_e = C\Phi i_a \tag{2.118}$$

and

$$e_b = C\Phi\omega_r \tag{2.119}$$

where $K_m = C\Phi$, Φ is the magnetic flux interacting with the armature conductors and C is a constant determined by the configuration of the motor.

In addition, the inductive armature circuit has to be modelled. Also the magnetic field of relatively large DC motors, such as in steel rolling mills, is produced by another inductive circuit but by a permanent magnet in smaller DC motors such as those used in the reaction wheels of spacecraft. Both of these circuits are shown in Fig. 2.13.

Applying Kirchhoff's second law to the field and armature circuits then yields the following differential equations.

$$v_a = R_a i_a + L_a \frac{\mathrm{d}i_a}{\mathrm{d}t} + e_b \Rightarrow \frac{\mathrm{d}i_a}{\mathrm{d}t} = \frac{1}{L_a}\left(v_a - e_b - R_a i_a\right) \tag{2.120}$$

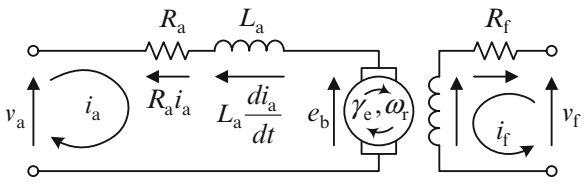


Fig. 2.13 Representation of DC motor including its equivalent circuit

and

$$v_f = R_f i_f + L_f \frac{di_f}{dt} \Rightarrow \frac{di_f}{dt} = \frac{1}{L_f}(v_f - R_f i_f)\,. \tag{2.121}$$

The DC motor model is then given by (2.114), (2.118), (2.119), (2.120) and (2.121).

2.2.6 Vector-Controlled AC Motors as Control Actuators

2.2.6.1 Concept of Vector Control

In a DC motor, the armature current, the magnetic flux linkage and the torque produced by the Lorenz force may be regarded as vectors, $\hat{\mathbf{i}}$, $\boldsymbol{\psi}$ and $\boldsymbol{\gamma}$. Then the torque equation (2.118) becomes

$$\boldsymbol{\gamma} = C'\boldsymbol{\psi} \wedge \mathbf{i} \tag{2.122}$$

noting that the constant, C' is not the same as C due to $\boldsymbol{\psi}$ being the flux linkage rather than the total flux. In literature on vector control, however, $\boldsymbol{\psi}$ is usually referred to simply as the magnetic flux. Since, in the vector cross product, $|\boldsymbol{\gamma}| = C'\,|\boldsymbol{\psi}|\,|\mathbf{i}|\sin(\alpha)$, where α is the angle between the vectors, $\boldsymbol{\psi}$ and $\mathbf{i}$, $|\boldsymbol{\gamma}|$ is maximised by maintaining $\alpha = \pi/2$, i.e. mutual orthogonality between these vectors. This is achieved in a DC motor through its physical design, but for AC motors it is achieved by vector control. For a synchronous motor, the rotor position is determined by measurement so that the orientation of $\boldsymbol{\psi}$ is known. Then the components of $\mathbf{i}$ are controlled to (a) keep $\mathbf{i}$ changing direction relative to the stator to follow rotor so that it is perpendicular to $\boldsymbol{\psi}$ and (b) its magnitude is set to produce the required torque. For an induction motor, the position of the rotor is measured (or estimated using a mathematical model of the motor in the so-called sensorless control). In this case, however, the magnetic flux vector, $\boldsymbol{\psi}$, results from the induced rotor currents and has to be estimated from a mathematical model of the motor. Then the components of $\mathbf{i}$ are determined, as for the synchronous motor, to maintain mutual orthogonality with $\boldsymbol{\psi}$ and produce the required torque.

2.2.6.2 The Transformations of Vector Control

The vectors representing the alternating voltages, currents and magnetic fluxes in an AC machine are expressed with respect to certain frames of reference, usually fixed to the stator or rotor of the motor. The stator currents and voltages are components of their vectors directed along stator-fixed axes and therefore alternate as the vectors rotate relative to the stator in a plane perpendicular to the rotation axis of the motor. The stator current vectors are usually controlled in the frame rotating with the rotor and their components along the axes of the rotating frame do not alternate in the same way. The same is true of the magnetic flux vector as this rotates with the rotor. The calculation of, for example, the stator current vector components in the stator-fixed frame is achieved by a rotational transformation similar to a two-dimensional version of the rotation matrix (direction cosine matrix) **C**, of Sect. 2.2.4.3. In the rotating frame, the components of the vectors appear as variable DC quantities and it is these that are controlled. The motor models needed to achieve this are in the form of differential equations and those available to the control system designer already incorporate the rotational transformations so that the input and output variables are the variable DC ones.

For a multiphase AC motor there exists an equivalent two-phase motor model and it is this that is used in vector control. Since most multiphase motors are three-phase motors, these are assumed in the following description. As shown in Fig. 2.12, the standard subscripts denoting the three phases are a, b and c. The corresponding phases of the two-phase equivalent motor model are denoted α and β. Let $\hat{\mathbf{x}}$ be a vector, corresponding to the applied stator voltage or the stator current, that is rotating at ω [rad/s] in a plane with components, $x_a(t)$, $x_b(t)$ and $x_c(t)$, along three axes, a, b and c equally separated in angle by $2\pi/3$ [rad], as shown in Fig. 2.14a. These are fixed with respect to the stator of the motor.

By convention, if the component of the vector along an axis is towards the arrow, then it is positive: otherwise, it is negative. If ω is constant, then $x_a(t)$, $x_b(t)$ and

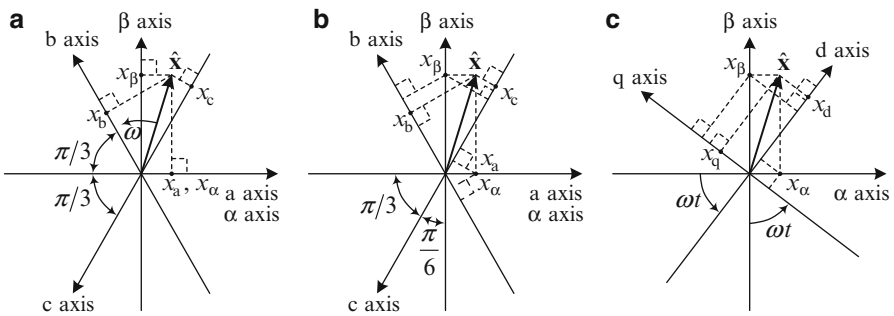


Fig. 2.14 Equivalent three and two-phase alternating variables generated by rotating vector. (**a**) Generation of 2 and 3 phase variables from the same vector. (**b**) Construction for derivation of Clarke transformation. (**c**) Construction for derivation of Park transformation

$x_c(t)$ are sinusoidal with angular frequency ω [rad/s] with amplitude $|\widehat{\mathbf{x}}|$ separated in phase by $2\pi/3$ [rad], i.e. balanced three-phase variables. Since the α and β axes are perpendicular, then the components, $x_\alpha(t)$ and $x_\beta(t)$ are sinusoidal variables, also at an angular frequency of ω [rad/s] with amplitudes equal to $|\widehat{\mathbf{x}}|$ separated in phase by $\pi/2$ [rad]. Figure 2.14b may then be used to write down equations for $x_a(t)$, $x_b(t)$ and $x_c(t)$ in terms of $x_\alpha(t)$ and $x_\beta(t)$, as follows:

$$\begin{aligned} x_a(t) &= x_\alpha(t) \\ x_b(t) &= -x_\alpha(t)\sin(\pi/6) + x_\beta(t)\cos(\pi/6) = -\frac{1}{2}x_\alpha(t) + \tfrac{\sqrt{3}}{2}x_\beta(t) \\ x_c(t) &= -x_\alpha(t)\sin(\pi/6) - x_\beta(t)\cos(\pi/6) = -\frac{1}{2}x_\alpha(t) - \tfrac{\sqrt{3}}{2}x_\beta(t) \end{aligned} \tag{2.123}$$

i.e.

$$\begin{bmatrix} x_a(t) \\ x_b(t) \\ x_c(t) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -1/2 & \sqrt{3}/2 \\ -1/2 & -\sqrt{3}/2 \end{bmatrix} \begin{bmatrix} x_\alpha(t) \\ x_\beta(t) \end{bmatrix}. \tag{2.124}$$

The left pseudo inverse of the matrix on the RHS may then be used to obtain the two-phase variables in terms of the three-phase variables. Thus

$$\begin{aligned} \begin{bmatrix} x_\alpha(t) \\ x_\beta(t) \end{bmatrix} &= \left[\begin{bmatrix} 1 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & \frac{\sqrt{3}}{2} & -\frac{\sqrt{3}}{2} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{1}{2} & -\frac{\sqrt{3}}{2} \end{bmatrix} \right]^{-1} \begin{bmatrix} 1 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & \frac{\sqrt{3}}{2} & -\frac{\sqrt{3}}{2} \end{bmatrix} \begin{bmatrix} x_a(t) \\ x_b(t) \\ x_c(t) \end{bmatrix} \\ &= \begin{bmatrix} \frac{3}{2} & 0 \\ 0 & \frac{3}{2} \end{bmatrix}^{-1} \begin{bmatrix} 1 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & \frac{\sqrt{3}}{2} & -\frac{\sqrt{3}}{2} \end{bmatrix} \begin{bmatrix} x_a(t) \\ x_b(t) \\ x_c(t) \end{bmatrix} \end{aligned}$$

i.e.

$$\begin{bmatrix} x_\alpha(t) \\ x_\beta(t) \end{bmatrix} = \begin{bmatrix} \frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} \\ 0 & \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} \end{bmatrix} \begin{bmatrix} x_a(t) \\ x_b(t) \\ x_c(t) \end{bmatrix}. \tag{2.125}$$

Transformation (2.125) is the Clarke transformation and (2.124) is called the inverse Clarke transformation.

Figure 2.14c represents a single-degree-of-freedom rotational transformation in which the components, $x_d(t)$ and $x_q(t)$, in a new frame of reference with axes, d and q, are expressed in terms of the components, $x_\alpha(t)$ and $x_\beta(t)$, in the frame of reference with axes, α and β, already introduced. If the d - q frame rotates with the vector, $\widehat{\mathbf{x}}$, and $|\widehat{\mathbf{x}}|$ is constant, then x_d and x_q are constant. The transformation equations follow from the figure and may be written as

$$\begin{bmatrix} x_{\mathrm{d}}(t) \\ x_{\mathrm{q}}(t) \end{bmatrix} = \begin{bmatrix} \cos(\omega t) & \sin(\omega t) \\ -\sin(\omega t) & \cos(\omega t) \end{bmatrix} \begin{bmatrix} x_{\alpha}(t) \\ x_{\beta}(t) \end{bmatrix}. \tag{2.126}$$

This is called the Park transformation. As this is a rotational transformation, it is orthogonal and therefore the matrix of the inverse Park transformation is the transpose of that of the RHS of (2.126). Thus

$$\begin{bmatrix} x_{\alpha}(t) \\ x_{\beta}(t) \end{bmatrix} = \begin{bmatrix} \cos(\omega t) & -\sin(\omega t) \\ \sin(\omega t) & \cos(\omega t) \end{bmatrix} \begin{bmatrix} x_{\mathrm{d}}(t) \\ x_{\mathrm{q}}(t) \end{bmatrix}. \tag{2.127}$$

2.2.6.3 Vector Control Implementation

In vector control, the transformations, (2.124), (2.125), (2.126) and (2.127) are implemented on a digital processor interfaced with the motor according to Fig. 2.15. The d-q frame is fixed in the rotor and 'd' denotes the direct axis along which the magnetic flux vector should be directed and 'q' denotes the quadrature axis along which the current component producing the torque is directed. The purpose of vector control is to keep the current and flux vectors mutually perpendicular, which produces the maximum torque for given vector magnitudes, as in a DC motor in which the armature current direction is perpendicular to the magnetic flux direction. It is important to realise that the current and flux vectors referred to in vector control are independent of the machine geometry and correspond to, rather than equal, the physical fluxes and currents [6].

A few practical features in Fig. 2.15 require explanation. First, nearly all electric drive applications employ *switched mode* power electronics to minimise the energy loss in the physical devices used to control the motor by regulating its electrical power input. The inverter is a set of six electronic switches that are controlled by a pulse modulator to apply physical stator voltages, $v_{\mathrm{as}}(t)$, $v_{\mathrm{bs}}(t)$ and $v_{\mathrm{cs}}(t)$, that rapidly switch between $\pm V_{\mathrm{DC}}$ with continuously varying mark space ratios such that the short-term mean values equal $v_{\mathrm{a}}(t)$, $v_{\mathrm{b}}(t)$ and $v_{\mathrm{c}}(t)$, the frequency being high enough for the system performance to be indistinguishable from a hypothetical one in which these continuously varying stator voltages were to be directly applied.

Pulse modulation is covered in some detail in Chap. 8. Second, it is usual for shaft encoders to be employed as speed and position sensors on the shafts of motors used in controlled electric drives. These provide digital outputs with pulse patterns enabling direction of motion to be detected. The frequency of the pulse trains can be determined by pulse timing and this yields an angular velocity measurement. The pulse count yields the angle of rotation. Software-implemented signal processing provides the position and velocity measurements. Third, to minimise instrumentation, it is usual to measure only two stator-phase currents, such as i_{b} and i_{c} and calculate the third using the well-known constraint equation of a balanced three-phase load, $i_{\mathrm{a}} + i_{\mathrm{b}} + i_{\mathrm{c}} = 0$, yielding $i_{\mathrm{a}} = -(i_{\mathrm{b}} + i_{\mathrm{c}})$. Specialist texts such as [7] may be read for more details.

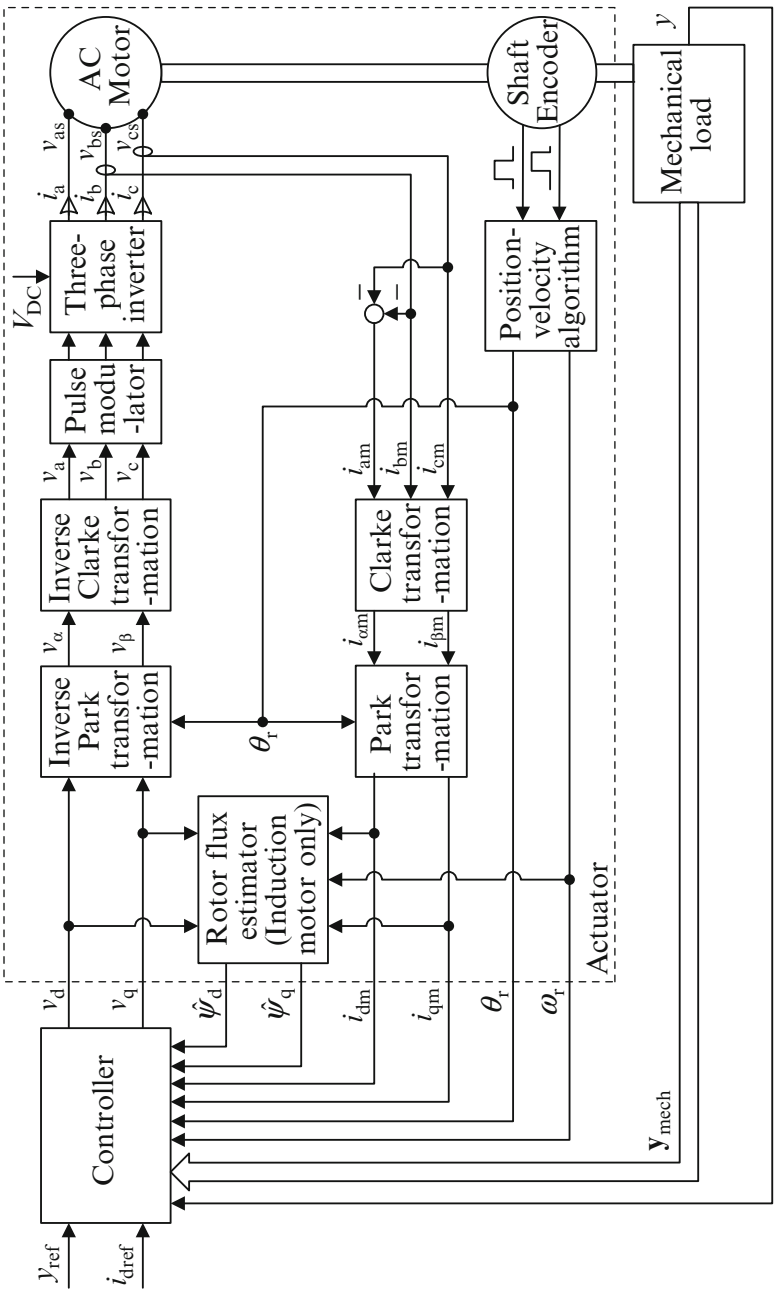


Fig. 2.15 AC motor with vector control transformations for use as a control actuator

Table 2.1 Variables used in vector control of AC motors and their models

Variables	Units	Description
v_a, v_b, v_c	[V]	Continuous three-phase stator voltage demands
V_{DC}	[V]	DC power supply voltage
v_{as}, v_{bs}, v_{cs}	[V]	Switched physical stator voltages with mean values, v_a, v_b, v_c
i_a, i_b, i_c	[A]	Physical stator-phase currents
i_{am}, i_{bm}, i_{cm}	[A]	Measured stator-phase currents
$i_{\alpha m}, i_{\beta m}$	[A]	Equivalent two-phase stator current measurements (Fig. 2.14b)
i_{dm}, i_{qm}	[A]	Measured stator current vector direct and quadrature axis components (Fig. 2.14c)
v_d, v_q	[V]	Continuous stator voltage vector direct and quadrature axis components to be applied
v_α, v_β	[V]	Continuous two-phase stator voltage demands
i_{dref}	[A]	Reference input value of direct-axis stator current vector component.
y, y_{ref}	[a]	Controlled plant output and corresponding reference input
$\mathbf{y}_{mech}$	[a]	Optional measurements, $y_i, i = 1, 2, \ldots$, from controlled mechanism

[a]Units dependent upon application

Many vector control schemes in industrial electric drives employ more than one of the traditional PI controllers. One controls the direct component, i_{dm}, of the transformed measured stator current in an induction motor to control the magnetic field, using the direct component, v_d, of the transformed stator voltage. In a PMSM, this PI controller is used to keep i_{dm} as close to zero as possible by setting the reference current to $i_{dref} = 0$. Exceptionally, i_{dref} is made a function of ω_r to reduce the magnetic field at high speeds to extend the speed range by reducing the stator back e.m.f. for a given speed to avoid stator voltage saturation due to the finite DC power supply voltage, V_{DC}, but this technique, known as flux weakening, is carried out with extreme care to avoid demagnetising the permanent magnets. Another PI controller is employed to control the rotor speed, ω_r, using the quadrature component, v_q, of the transformed stator voltage. Typically, if the rotor position is to be controlled, a third PI controller is added using the reference input of the speed control loop as its control variable.

The variables of Fig. 2.15 are described in Table 2.1.

This traditional arrangement of PI controllers, however, often requires much time-consuming tuning at commissioning time and retuning during the lifetime of an electric drive. Also, the traditional philosophy is to control the position or the speed of the motor with the additional torque due to the mechanical load regarded as external disturbance torque. For some applications, such as those entailing mechanical vibration modes, acceptable control is difficult to attain in this way, certainly not a specified dynamic response of the closed-loop system to the reference inputs. The more general control structure of Fig. 2.15 can overcome these problems with a suitable choice of the single controller shown, the freedom of choice being wide with modern digital implementation. In this spirit, the block arrow signal, $\mathbf{y}_{mech}$, represents the additional measurements, such as flexural deflections in

mechanical structures, that enable the best control to be attained within the hardware limitations. As far as this chapter is concerned, the commonly found PI-based vector-controlled electric drives are not included as complete actuators. Instead, the actuator is regarded as just the motor and the transformations, as shown in Fig. 2.15, which can be accurately modelled as part of the plant to be controlled. The controller choice is left open so the mechanical load, the motor and the transformations together constitute the plant. Control techniques other than the traditional ones are advantageous in electric drives [10].

When the motor is viewed through the input and output signals of the 'actuator' block in Fig. 2.15, its behaviour resembles that of the DC motor, at least in that the variables are not required to oscillate, and this demonstrates the great advantage of vector control in enabling AC motors to be used as actuators with relatively sophisticated controllers. The alternating voltages required by the motor are automatically produced by the inverse Park transformation, due to its time-varying elements, $\sin(\omega t)$ and $\cos(\omega t)$. As the motor accelerates and decelerates, however, the frequencies and amplitudes of its alternating voltages, currents and magnetic fluxes will change, in contrast with such motors used directly with AC power supplies. Also, the time-varying Park transformation removes the oscillations of the alternating variables of the motor from the measured current components, i_{dm} and i_{dm}. The oscillations of the AC variables in the motor are therefore 'invisible' to the control engineer in the d-q models in the following subsections, which are in the form of differential equations that may be used directly for control system design. Detailed derivations of these models may be found in specialist texts [11].

2.2.6.4 Induction Motor d-q Model

The complete d-q induction model comprises the following set of first-order differential equations.

$$\begin{aligned}
\tfrac{\mathrm{d}i_{\mathrm{d}}}{\mathrm{d}t} &= -Ai_{\mathrm{d}} + B\psi_{\mathrm{d}} + C\omega_{\mathrm{r}}\psi_{\mathrm{q}} + Dv_{\mathrm{d}} + p\omega_{\mathrm{r}}i_{\mathrm{q}} \\
\tfrac{\mathrm{d}i_{\mathrm{q}}}{\mathrm{d}t} &= -Ai_{\mathrm{q}} + B\psi_{\mathrm{q}} - C\omega_{\mathrm{r}}\psi_{\mathrm{d}} + Dv_{\mathrm{q}} - p\omega_{\mathrm{r}}i_{\mathrm{d}} \\
\tfrac{\mathrm{d}\psi_{\mathrm{d}}}{\mathrm{d}t} &= -E\psi_{\mathrm{q}} + Fi_{\mathrm{d}} \\
\tfrac{\mathrm{d}\psi_{\mathrm{q}}}{\mathrm{d}t} &= -E\psi_{\mathrm{q}} + Fi_{\mathrm{q}} \\
\tfrac{\mathrm{d}\omega_{\mathrm{r}}}{\mathrm{d}t} &= G\left(\Psi_{\mathrm{d}}i_{\mathrm{q}} - \Psi_{\mathrm{q}}i_{\mathrm{d}}\right) - H\gamma_L, \quad \tfrac{\mathrm{d}\theta_{\mathrm{r}}}{\mathrm{d}t} = \omega_{\mathrm{r}}
\end{aligned} \tag{2.128}$$

where

$$\begin{aligned}
&D = \tfrac{L_{\mathrm{r}}}{L_{\mathrm{s}}L_{\mathrm{r}} - L_{\mathrm{m}}^2};\ A = D.\left(R_{\mathrm{s}} + \tfrac{L_{\mathrm{m}}^2}{L_{\mathrm{r}}^2}R_{\mathrm{r}}\right)\ B = D.\tfrac{L_{\mathrm{m}}R_{\mathrm{r}}}{L_{\mathrm{r}}^2}\quad C = D.\tfrac{L_{\mathrm{m}}}{L_{\mathrm{r}}}.p \\
&E = \tfrac{R_{\mathrm{r}}}{L_{\mathrm{r}}} \qquad F = \tfrac{L_{\mathrm{m}}}{L_{\mathrm{r}}}R_{\mathrm{r}} \qquad G = \tfrac{1}{J_{\mathrm{r}}}.\tfrac{3p}{2}.\tfrac{L_{\mathrm{m}}}{L_{\mathrm{r}}}\ H = \tfrac{1}{J_{\mathrm{r}}}
\end{aligned} \tag{2.129}$$

Here, ψ_d and ψ_q are the rotor magnetic flux vector components, and γ_L is the load torque defined in Sects. 2.2.5.3 and 2.2.2.3. L_s, L_r and L_m are, respectively, the stator, rotor and mutual inductances. R_s and R_r are the stator and rotor resistances, p is the number of stator pole pairs and J_r is the rotor moment of inertia.

2.2.6.5 Induction Motor α-β Model

The induction motor model equivalent to that of Sect. 2.2.6.4 but formulated in terms of the vector components along the α and β axes fixed with respect to the stator are presented here as they could be useful in simulations to display the alternating variables of the motor. Also, it is possible to create a controller based directly on this model with an internal oscillatory mode that automatically creates the alternating variables of the machine without the aid of the time-varying Park and inverse Park transformations, only the Clarke and inverse Clarke transformations being necessary in Fig. 2.15 [10]. Thus,

$$\begin{aligned}
\frac{\mathrm{d}i_\alpha}{\mathrm{d}t} &= -Ai_\alpha + B\psi_\alpha + C\omega_r\psi_\beta + Dv_\alpha \\
\frac{\mathrm{d}i_\beta}{\mathrm{d}t} &= -Ai_\beta + B\psi_\beta - C\omega_r\psi_\alpha + Du_\beta \\
\frac{\mathrm{d}\psi_\alpha}{\mathrm{d}t} &= -E\psi_\alpha - p\omega_r\psi_\beta + Fi_\alpha \\
\frac{\mathrm{d}\psi_\beta}{\mathrm{d}t} &= -E\psi_\beta + p\omega_r\psi_\alpha + Fi_\beta \\
\frac{\mathrm{d}\omega_r}{\mathrm{d}t} &= G\left(\Psi_\alpha i_\beta - \Psi_\beta i_\alpha\right) - H\gamma_L, \quad \frac{\mathrm{d}\theta_r}{\mathrm{d}t} = \omega_r
\end{aligned} \tag{2.130}$$

The constants are as defined in (2.129).

2.2.6.6 Synchronous Motor d-q Model

The complete d-q permanent magnet synchronous motor model comprises the following set of first-order differential equations:

$$\begin{aligned}
\frac{\mathrm{d}i_d}{\mathrm{d}t} &= -\frac{R_s}{L_d}i_d + p\omega_r\frac{L_q}{L_d}i_q + \frac{1}{L_d}v_d \\
\frac{\mathrm{d}i_q}{\mathrm{d}t} &= -p\omega_r\frac{L_d}{L_q}i_d - \frac{R_s}{L_q}i_q - \frac{p\omega_r}{L_q}\Psi_{PM} + \frac{1}{L_q}v_q \\
\frac{\mathrm{d}\omega_r}{\mathrm{d}t} &= \frac{1}{J_r}\left(\gamma_e - \gamma_L\right) \\
\frac{\mathrm{d}\theta_r}{\mathrm{d}t} &= \omega_r
\end{aligned} \tag{2.131}$$

where

$$\gamma_e = \frac{3p}{2}\left[\Psi_{PM}i_q + \left(L_d - L_q\right)i_d i_q\right]. \tag{2.132}$$

Here, Ψ_{PM} is the permanent magnet flux, R_s is the stator resistance, L_d and L_q are the direct and quadrature axis inductances and p is the number of pole pairs.

2.2.7 Fluid and Thermal Subsystems

2.2.7.1 Introduction

Some plants involve heat flow and/or fluid flow and this subsection presents some relevant models. Specialist texts such as [1, 12] may be consulted for a comprehensive coverage.

2.2.7.2 Coupled-Tank Systems

Many industrial processes involve one or more interconnected tanks through which liquid is passed and it is necessary to control the liquid heights in the tanks and the flow rates. In such cases, the liquid may be regarded incompressible. The general coupled-tank system of Fig. 2.16 covers several specific examples.

Pumps, P_1 and P_2, supply the liquid at controlled volume flow rates of q_1 and q_2 via the control variables, u_1 and u_2. Assuming that these pumps and their electric drives are linear and the dynamical effects of these drives are negligible, then

$$q_i = b_i u_i, \quad i = 1, 2. \tag{2.133}$$

The fluid pressures at the bases of the tanks are

$$p_i = \rho g h_i, \quad i = 1, 2. \tag{2.134}$$

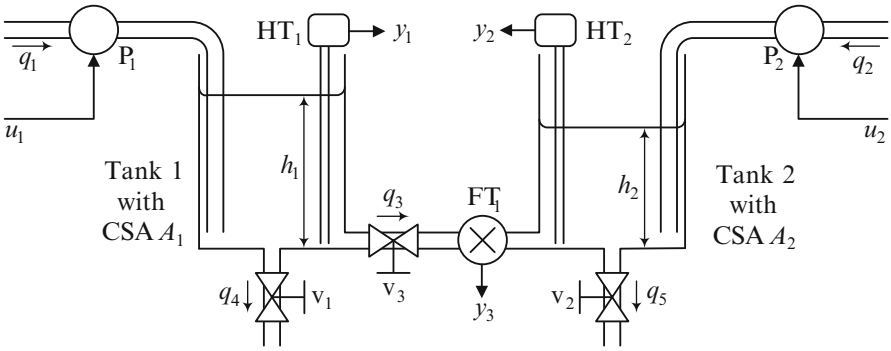


Fig. 2.16 A coupled-tank system

Here, ρ is the liquid density and $g = 9.81$ [m/s/s] is the acceleration due to gravity. The transducers, HT_1 and HT_2, measure these pressures but in view of (2.134) they are calibrated to measure h_1 and h_2, which are also referred to as the liquid heads. Then, assuming linearity of the transducers, the liquid height measurements are

$$y_i = K_h h_i, \quad i = 1, 2 \tag{2.135}$$

where K_h is the height measurement constant. The volume flow rate, q_3, between the tanks, which can be positive or negative, is measured by the transducer, FT_1, and assuming this is linear, the measurement is

$$y_3 = K_f q_3. \tag{2.136}$$

The valves, V_1, V_2 and V_3, can be preset to yield different flow rates for given values of h_1 and h_2. According to the theory of fluid dynamics [1], the Reynolds numbers of the valve orifices are dimensionless parameters given by

$$N_{Rei} = \frac{\rho v_i L_i}{\mu}, \quad i = 1, 2, 3, \tag{2.137}$$

where μ is the fluid dynamic viscosity, v_i is the fluid velocity and L_i is a characteristic linear dimension dependent on the valve setting. So the Reynolds numbers vary with the flow rates and the valve settings, but if they remain sufficiently small ($N_{Rei} < 2,000, \quad i = 1, 2, 3$), then the flow is laminar and the relationship between the pressure drop across each valve and the flow rate through it is linear, yielding

$$q_4 = h_1 / R_{f1}, \tag{2.138}$$

$$q_5 = h_2 / R_{f2} \tag{2.139}$$

and

$$q_3 = (h_1 - h_2) / R_{f3}. \tag{2.140}$$

where R_{fi}, $i = 1, 2, 3$, are defined as the fluid resistances of the orifices. An electrical analogy is immediately apparent in which volume flow rate is equivalent to electric current and the liquid heads are equivalent to voltages. If, on the other hand, the Reynolds numbers of the valve orifices are relatively large ($N_{Rei} > 4,000, \quad i = 1, 2, 3$), then the flow is turbulent and the relationship between the pressure drop across each valve and the flow rate through it becomes nonlinear. Thus

$$q_4 = K_{V1} \sqrt{h_1}, \tag{2.141}$$

$$q_5 = K_{V2}\sqrt{h_2} \tag{2.142}$$

and

$$q_3 = K_{V3}\sqrt{|h_1 - h_2|}\text{sgn}\,(h_1 - h_2)\,, \tag{2.143}$$

where $\text{sgn}(x) \triangleq \{+1, x > 0; 0, x = 0; = 1, x < 0\}$. In this case, the fluid resistances are defined as the changes in the liquid heads divided by the changes in the volume flow rates, i.e.

$$R_1 = 1/\frac{\mathrm{d}q_4}{\mathrm{d}h_1} = \frac{2}{K_{V1}}\sqrt{h_1}, \tag{2.144}$$

$$R_2 = 1/\frac{\mathrm{d}q_5}{\mathrm{d}h_2} = \frac{2}{K_{V2}}\sqrt{h_2} \tag{2.145}$$

and

$$R_3 = 1/\left.\frac{\mathrm{d}q_3}{\mathrm{d}h}\right|_{h=|h_1-h_2|} = \frac{2}{K_{V3}}\sqrt{|h_1 - h_2|}. \tag{2.146}$$

The model is completed by relating the rates of change of the liquid heights to the rates of change of liquid volume in the tanks, using the cross-sectional areas. Thus,

$$\dot{h}_1 = [q_1 - (q_3 + q_4)]\,/A_1 \tag{2.147}$$

and

$$\dot{h}_2 = (q_3 + q_2 - q_5)\,/A_2. \tag{2.148}$$

2.2.7.3 Thermal Systems

Plants involving heat flow have continuous spatial temperature distributions for which partial differential equations would be needed to form a precise mathematical model [12]. Such models are referred to as distributed parameter models. While they may be numerically integrated on a computer to predict the system behaviour, they are not convenient for control system design. For this purpose, it is usual to replace a partial differential equation with a finite set of ordinary differential equations whose solutions are accurate at a number of discrete points. Fortunately many thermal systems may be divided into subsystems in which the temperature is nearly uniform, substantial temperature gradients being restricted to the interfaces between the subsystems. Then the number of ordinary differential equations required can be quite small, just one for each subsystem. The complete model is then referred to as a lumped parameter model.

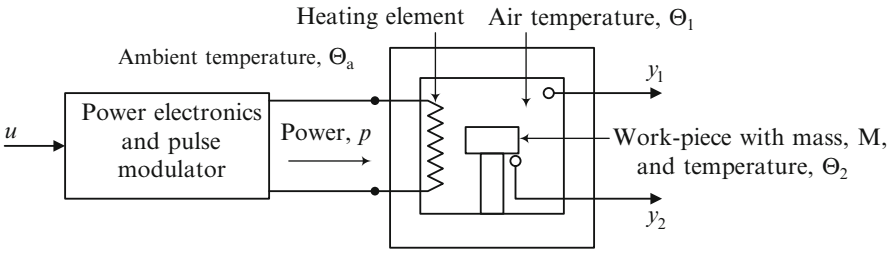


Fig. 2.17 Electric kiln

An example of a controllable heating system with two subsystems is an electric kiln such as illustrated in Fig. 2.17. It is assumed that the pulse modulator, present to operate the power electronics in a switched mode to minimise the energy loss, is designed so that the mean power dissipation in the heating element is directly proportional to the control variable, u. Thus

$$p = K_e u \tag{2.149}$$

where K_e is the heating element power constant. It will be assumed that the convection currents ensure a uniform air temperature within the kiln.

Let q_a be the total amount of heat contained in the air within the kiln, $\dot{q}_s$ be the rate of supply of heat from the heating element and $\dot{q}_w$ be the rate of flow of heat from the kiln wall, which will be negative since heat is actually being lost through the wall due to its imperfect insulation. Then

$$\dot{q}_s = p \tag{2.150}$$

and Fourier's law of heat conduction gives

$$\dot{q}_w = -k_w A_w \frac{d\Theta}{dx}, \tag{2.151}$$

where k_w is the wall conductivity, A_w is the inside area of the kiln wall and $\frac{d\Theta}{dx}$ is the temperature gradient in the wall. Assuming this is constant, then

$$\frac{d\Theta}{dx} = \frac{\Theta_1 - \Theta_a}{D}, \tag{2.152}$$

where D is the wall thickness. The differential equation governing the air temperature is then obtained as follows. First

$$\dot{q}_a = \dot{q}_s + \dot{q}_w = p - k_w A_w \frac{d\Theta}{dx} = K_e u - \frac{k_w A_w}{D} (\Theta_1 - \Theta_a) . \tag{2.153}$$

Then, if C_a is the specific heat capacity of air and M_a is the mass of air contained in the kiln,

$$q_a = M_a C_a \Theta_1 \Rightarrow \Theta_1 = \frac{1}{M_a C_a} q_a \Rightarrow \dot{\Theta}_1 = \frac{1}{M_a C_a} \dot{q}_a. \tag{2.154}$$

This enables (2.153) to be written as

$$\dot{\Theta}_1 = \frac{1}{T_1} \left[bu - \Theta_1 + \Theta_a \right], \tag{2.155}$$

where $T_1 = \frac{M_a C_a D}{k_w A_w}$ is the air heating time constant and $b = \frac{K_e D}{k_w A_w}$ is the aiming temperature constant. The first subsystem of the plant model is given by (2.155).

Let the total amount of heat in the workpiece be q_p, the heat transfer coefficient between the surrounding air and the work-piece be h_p and the surface area of the workpiece be A_p. Then Newton's Law of heating yields

$$\dot{q}_p = h_p A_p \left(\Theta_1 - \Theta_2 \right). \tag{2.156}$$

If C_p is the specific heat capacity of the workpiece and M_p is its mass, then

$$q_p = M_p C_p \Theta_2 \Rightarrow \Theta_2 = \frac{1}{M_p C_p} q_p \Rightarrow \dot{\Theta}_2 = \frac{1}{M_p C_p} \dot{q}_p, \tag{2.157}$$

enabling (2.156) to be written as

$$\dot{\Theta}_2 = \frac{1}{T_2} \left(\Theta_1 - \Theta_2 \right) \tag{2.158}$$

where $T_2 = \frac{M_p C_p}{h_p A_p}$ is the workpiece time constant. The second subsystem of the plant model is (2.158).

Finally, the temperature measurement transducers are usually linear so that $y_1 = K_T \Theta_1$ and $y_2 = K_T \Theta_2$, where K_T is the temperature measurement constant.

2.3 Identification of LTI Plants from Measurements

2.3.1 Overview

Plant (or system) identification is the determination of a mathematical model of the plant using measured data. In contrast to the physical modelling of Sect. 2.1.4, the plant is thought of as a 'black box' with input and output signals. The approach is then to determine a model that fits the observed output responses to given inputs.

Here, time domain and frequency domain methods are introduced for obtaining such models in the form of transfer functions.

For some identification methods, an input of known form is applied to the plant, which implies that the operation has to be on an open-loop basis. In these cases, the plant has to be known in advance to be stable. On the other hand, it is possible to identify an unstable plant if a feedback controller can be applied yielding closed-loop stability. Then given reference inputs are applied and the resulting control and measurement variables are observed to perform the identification.

The following subsections commence with elementary methods for the simplest SISO LTI plants, assuming open-loop stability, and progress to more sophisticated methods for the identification of general LTI plants.

2.3.2 Plant Model Determination from Step Response

2.3.2.1 First-Order Plant

Consider a first-order plant characterised by its DC gain, K_{dc}, and time constant, T. The Laplace transfer function model is then

$$\frac{Y(s)}{U(s)} = \frac{K_{\mathrm{dc}}}{1+sT}. \tag{2.159}$$

Let a step input, $u(t) = Ah(t)$, be applied, where A is a constant and $h(t)$ is the unit step function. Then $U(s) = A/s$ and

$$Y(s) = \frac{K_{\mathrm{dc}}}{1+sT}.\frac{A}{s}. \tag{2.160}$$

Then using the table of Laplace transforms and their inverses [Table 1],

$$y(t) = \mathcal{L}^{-1}\left\{\frac{K_{\mathrm{dc}}}{1+sT}.\frac{A}{s}\right\} = K_{\mathrm{dc}}A\left(1-e^{-t/T}\right). \tag{2.161}$$

The steady-state value of the response is then

$$y_{\mathrm{ss}} = \lim_{t\to\infty} y(t) = K_{\mathrm{dc}}A. \tag{2.162}$$

Then (2.161) may be written as

$$y(t) = y_{\mathrm{ss}}\left(1-e^{-t/T}\right). \tag{2.163}$$

Figure 2.18 illustrates an experimental step response.

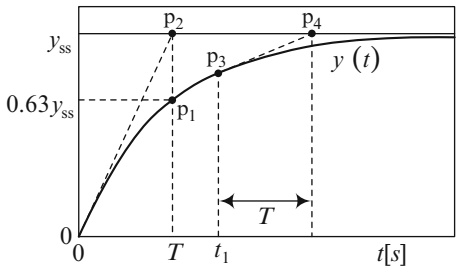


Fig. 2.18 Step response and parameters for estimation of first-order transfer function

The measured steady-state value of $y(t)$ may be used to obtain K_{dc} from (2.162) as

$$K_{dc} = \frac{y_{ss}}{A} \tag{2.164}$$

The time constant, T, may be determined by three methods for cross-checking:

Method 1 From (2.163),

$$y(T) = y_{ss}\left(1 - e^{-1}\right) = 0.6321 y_{ss} \cong 0.63 y_{ss}. \tag{2.165}$$

The time constant, T, may therefore be estimated as the time at which the graph of $y(t)$ intersects the line, $y = 0.63y_{ss}$, at the point, p_1, as shown in Fig. 2.18.

Method 2 With reference to Fig. 2.18, the time constant may be estimated as the time at which the tangent to the graph of $y(t)$ at $t = 0$ intersects the line, $y = y_{ss}$, at the point, p_2. This relationship may be proven as follows. Differentiating (2.163) yields

$$\dot{y}(t) = \frac{y_{ss}}{T} e^{-t/T}. \tag{2.166}$$

The slope of the tangent, $0\text{-}p_2$, is therefore $\dot{y}(0) = y_{ss}/T$ and its equation is

$$f_0(t) = \frac{y_{ss}}{T} t. \tag{2.167}$$

This straight line intersects the horizontal straight line, $y = y_{ss}$, when $f_0(t) = y_{ss}$ and by inspection of (2.167) this is when $t = T$.

Method 3 This method is a generalisation of Method 2 and, with reference to Fig. 2.18, is based on the fact that the tangent to the graph of $y(t)$ at an arbitrary point, p_3, and time, t_1, intersects the horizontal straight line, $y = y_{ss}$, at a point, p_4, where $t = t_1 + T$. This enables T to be estimated at any point on the graph of $y(t)$. The proof of this relationship is as follows. The equation of the tangent, $p_3 - p_4$, is

$$f_{t_1}(t - t_1) = \dot{y}(t_1)(t - t_1) + y(t_1). \tag{2.168}$$

From (2.163) and (2.166),

$$y(t_1) = y_{ss}\left(1 - e^{-t_1/T}\right) \quad \text{and} \quad \dot{y}(t_1) = \frac{y_{ss}}{T} e^{-t_1/T} \tag{2.169}$$

Substituting for $\dot{y}(t_1)$ and $y(t_1)$ in (2.168) using (2.169) then yields

$$f_{t_1}(t - t_1) = \frac{y_{ss}}{T} e^{-t_1/T}(t - t_1) + y_{ss}\left(1 - e^{-t_1/T}\right) \tag{2.170}$$

By inspection of (2.170), $f_{t_1}(t - t_1) = y_{ss}$ when $t - t_1 = T \Rightarrow t = t_1 + T$.

It is recommended that the three methods are applied, Method 3 for several points, and the set of time constant estimates averaged.

If the plant transfer function model is required in the standard form,

$$\frac{Y(s)}{U(s)} = \frac{b_0}{s + a_0} \tag{2.171}$$

and the model has been obtained in the form of (2.159), then (2.171) can be manipulated into the same form to obtain b_0 and a_0 in terms of K_{dc} and T, as follows.

$$\frac{Y(s)}{U(s)} = \frac{b_0/a_0}{s/a_0 + 1} = \frac{K_{dc}}{1 + sT} \Rightarrow a_0 = \frac{1}{T} \text{ and } \frac{b_0}{a_0} = K_{dc} \Rightarrow b_0 = a_0 K_{dc} = \frac{K_{dc}}{T}. \tag{2.172}$$

2.3.2.2 Underdamped Second-Order Plant

In this case, the plant is characterised by the undamped natural frequency, ω_n, the damping ratio, ζ, where $0 < \zeta < 1$ and the DC gain, K_{dc}, the transfer function being

$$\frac{Y(s)}{U(s)} = \frac{K_{dc}\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2}. \tag{2.173}$$

If the input is $u(t) = Ah(t)$, where A is a constant and $h(t)$ is the unit step function, then $U(s) = A/s$ and the Laplace transform of the output is

$$Y(s) = \frac{K_{dc}\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} \cdot \frac{A}{s} \tag{2.174}$$

Using the final value theorem, the steady-state output is

$$y_{ss} = \lim_{t \to \infty} y(t) = \lim_{s \to 0} sY(s) = K_{dc}A \tag{2.175}$$

Then using the table of Laplace transforms and their inverses (Table 1 in Tables),

$$y(t) = \mathcal{L}^{-1}\left\{\frac{K_{\mathrm{dc}}\omega_{\mathrm{n}}^2}{s^2 + 2\zeta\omega_{\mathrm{n}}s + \omega_{\mathrm{n}}^2}.\frac{A}{s}\right\} = y_{\mathrm{ss}}\left[1 - \frac{1}{\sqrt{1-\zeta^2}}e^{-\zeta\omega_{\mathrm{n}}t}\sin\left(\omega_{\mathrm{d}}t + \phi\right)\right] \tag{2.176}$$

where $\omega_{\mathrm{d}} = \omega_{\mathrm{n}}\sqrt{1-\zeta^2}$ is the damped natural frequency and $\phi = \cos^{-1}(\zeta)$.

The parameters, K_{dc}, ω_{n} and ζ, may be estimated from an experimentally obtained step response of the form shown in Fig. 2.19.

The DC gain can be estimated using y_{ss} obtained from Fig. 2.19 and the known step input level, A, using the following equation from (2.175).

$$K_{\mathrm{dc}} = \frac{y_{\mathrm{ss}}}{A}. \tag{2.177}$$

An expression for the peak output, y_{p}, will now be derived using the step response of (2.176). This will first be converted to a more convenient form as follows.

$$\begin{aligned}\sin\left(\omega_{\mathrm{d}}t + \phi\right) &= \sin\left(\omega_{\mathrm{d}}t\right)\cos\left(\phi\right) + \cos\left(\omega_{\mathrm{d}}t\right)\sin\left(\phi\right)\\ &= \zeta\sin\left(\omega_{\mathrm{d}}t\right) + \sqrt{1-\zeta^2}\cos\left(\omega_{\mathrm{d}}t\right),\end{aligned} \tag{2.178}$$

recalling that $\cos(\phi) = \zeta$ and therefore $\sin(\phi) = \sqrt{1-\cos^2(\phi)} = \sqrt{1-\zeta^2}$. Then (2.176) may be rewritten as

$$y(t) = y_{\mathrm{ss}}\left[1 - e^{-\zeta\omega_{\mathrm{n}}t}\left(\frac{\zeta}{\sqrt{1-\zeta^2}}\sin\left(\omega_{\mathrm{d}}t\right) + \cos\left(\omega_{\mathrm{d}}t\right)\right)\right] \tag{2.179}$$

The first peak occurs at $t = T_{\mathrm{p}}$, which is the smallest value of $t > 0$ for which $\dot{y}(t) = 0$. Again with the aid of (2.174) and the Laplace transform tables (Table 1 in Tables),

$$\dot{y}(t) = \mathcal{L}^{-1}\left\{sY(s)\right\} = \mathcal{L}^{-1}\left\{\frac{y_{\mathrm{ss}}\omega_{\mathrm{n}}^2}{s^2 + 2\zeta\omega_{\mathrm{n}}s + \omega_{\mathrm{n}}^2}\right\} = \frac{y_{\mathrm{ss}}\omega_{\mathrm{n}}}{\sqrt{1-\zeta^2}}\sin\left(\omega_{\mathrm{d}}t\right). \tag{2.180}$$

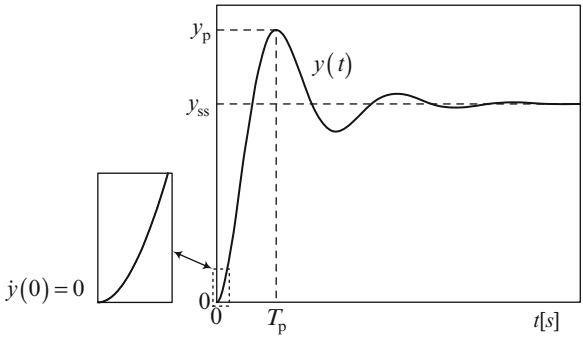


Fig. 2.19 Step response and parameters for estimation of second-order transfer function

It is clear from (2.180) that the peak time is half the oscillation period. Thus

$$T_{\mathrm{p}} = \frac{1}{2}.\frac{2\pi}{\omega_{\mathrm{d}}} = \frac{\pi}{\omega_{\mathrm{d}}} = \frac{\pi}{\omega_{\mathrm{n}}\sqrt{1-\zeta^2}}. \tag{2.181}$$

Then from (2.179),

$$y\left(T_{\mathrm{p}}\right) = y_{\mathrm{p}} = y_{\mathrm{ss}}\left[1 + e^{-\frac{\zeta\pi}{\sqrt{1-\zeta^2}}}\right] \Rightarrow -\frac{\zeta\pi}{\sqrt{1-\zeta^2}} = \ln\left(\frac{y_{\mathrm{p}}}{y_{\mathrm{ss}}} - 1\right) \Rightarrow$$

$$\zeta^2\pi^2 = \left(1-\zeta^2\right)\ln^2\left(\frac{y_{\mathrm{p}}}{y_{\mathrm{ss}}} - 1\right) \Rightarrow \zeta^2\left[\pi^2 + \ln^2\left(\frac{y_{\mathrm{p}}}{y_{\mathrm{ss}}} - 1\right)\right] = \ln^2\left(\frac{y_{\mathrm{p}}}{y_{\mathrm{ss}}} - 1\right) \Rightarrow$$

$$\zeta = -\ln\left(\frac{y_{\mathrm{p}}}{y_{\mathrm{ss}}} - 1\right) \Big/ \sqrt{\pi^2 + \ln^2\left(\frac{y_{\mathrm{p}}}{y_{\mathrm{ss}}} - 1\right)}, \tag{2.182}$$

noting that $0 < y_{\mathrm{p}}/y_{\mathrm{ss}} - 1 < 1 \Rightarrow \ln\left(y_{\mathrm{p}}/y_{\mathrm{ss}} - 1\right) < 0$ and ζ must be positive. This enables the damping ratio, ζ, to be estimated from y_{p} and y_{ss} obtained from Fig. 2.19. Then the undamped natural frequency can be estimated using T_{p} from Fig. 2.19 and ζ from (2.182) using the following equation from (2.181).

$$\omega_{\mathrm{n}} = \frac{\pi}{T_{\mathrm{p}}\sqrt{1-\zeta^2}}. \tag{2.183}$$

In many cases, the form of the transfer function will have been established by physical modelling as covered in section 2.1.4. If only the experimental step response is available, then it is important to examine it to check that it is suitable for fitting the model of (2.173). Apart from the need for the step response to be oscillatory, as shown, it is wise to check that it commences with zero slope, i.e. $\dot{y}(0) = 0$, as shown in the insert of Fig. 2.19. This is an indication that the plant satisfies the requirement of having no finite zeros. That fact that the initial slope is non-zero if the plant has a finite zero is proven as follows. Let the plant transfer function have the following transfer function with a finite zero at $s = -1/T_{\mathrm{z}}$.

$$\frac{Y'(s)}{U(s)} = \frac{K_{\mathrm{dc}}\omega_{\mathrm{n}}^2\left(1 + sT_{\mathrm{z}}\right)}{s^2 + 2\zeta\omega_{\mathrm{n}}s + \omega_{\mathrm{n}}^2} \tag{2.184}$$

The Laplace transform of the step response is therefore

$$Y'(s) = \frac{K_{\mathrm{dc}}\omega_{\mathrm{n}}^2\left(1 + sT_{\mathrm{z}}\right)}{s^2 + 2\zeta\omega_{\mathrm{n}}s + \omega_{\mathrm{n}}^2}.\frac{A}{s} \tag{2.185}$$

and that of the first derivative is

$$\mathcal{L}\left\{\dot{y}(t)\right\} = sY(s) = \frac{AK_{\mathrm{dc}}\omega_{\mathrm{n}}^2\left(1 + sT_{\mathrm{z}}\right)}{s^2 + 2\zeta\omega_{\mathrm{n}}s + \omega_{\mathrm{n}}^2}. \tag{2.186}$$

Applying the initial value theorem then yields

$$\dot{y}(0) = \lim_{s\to\infty} s\mathcal{L}\{\dot{y}(t)\} = \lim_{s\to\infty} \frac{AK_{\mathrm{dc}}\omega_{\mathrm{n}}^2\left(s + s^2T_{\mathrm{z}}\right)}{s^2 + 2\zeta\omega_{\mathrm{n}}s + \omega_{\mathrm{n}}^2} = AK_{\mathrm{dc}}\omega_{\mathrm{n}}^2T_{\mathrm{z}}. \tag{2.187}$$

Hence $\dot{y}(0) \neq 0$ for $T_{\mathrm{z}} \neq 0$. If transfer function (2.182) does not have a finite zero, then $T_{\mathrm{z}} = 0$. In this case (2.187) yields $\dot{y}(0) = 0$, which is correct.

2.3.3 Plant Model Determination from Frequency Response

2.3.3.1 Introduction

Experimental plant data is often available in the frequency domain. The aim here is to use this data to estimate a linear SISO plant model in the form,

$$\frac{Y(s)}{U(s)} = G(s) = K_{\mathrm{dc}}\frac{\prod_i\left(1 + \frac{s}{\nu_i}\right)\prod_i\left(1 + \frac{2\eta_i}{\nu_{\mathrm{n}i}}s + \frac{1}{\nu_{\mathrm{n}i}^2}s^2\right)}{s^q\prod_i\left(1 + \frac{s}{\omega_i}\right)\prod_i\left(1 + \frac{2\zeta_i}{\omega_{\mathrm{n}i}}s + \frac{1}{\omega_{\mathrm{n}i}^2}s^2\right)}, \tag{2.188}$$

where K_{dc} is the DC gain, ω_i, $\omega_{\mathrm{n}i}$, ν_i and $\nu_{\mathrm{n}i}$ are the corner frequencies, ζ_i are the damping ratios of the complex conjugate poles and η_i are equivalent parameters for the complex conjugate zeros. All these parameters, except K_{dc}, are always positive and therefore all the poles and zeros are assumed to lie in the left half of the s-plane. Plants with poles and/or zeros in the right half of the s-plane, however, are dealt with in Sect. 2.3.3.10.

The identification method used depends upon the relative positions of the poles and zeros. These, of course, are not known in advance, but an initial examination of the measured data reveals features that enable an appropriate method to be chosen.

Some background theory in the frequency domain will now be reviewed in preparation for developing the methods of transfer function model determination. First consider the frequency domain transfer function,

$$G(j\omega) = |G(j\omega)|\angle G(j\omega) = M(\omega)e^{\phi(\omega)}. \tag{2.189}$$

This can be displayed graphically in the form of the magnitude, $M(\omega)$, and the phase angle, $\phi(\omega)$. The simplest way to obtain this data is to carry out tests using sinusoidal plant excitation over the frequency range, $\omega \in (0,\ \omega_{\mathrm{b}})$, where ω_{b} is the specified bandwidth for the control system to be designed. Assuming plant linearity, once the initial transients have decayed to negligible proportions the output is a sinusoid, as shown in Fig. 2.20. This is the steady-state sinusoidal response.

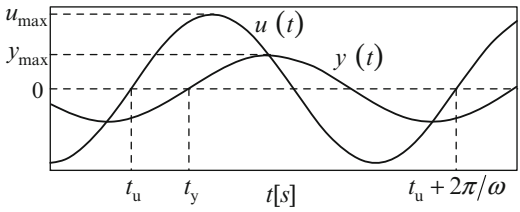


Fig. 2.20 Illustration of steady-state sinusoidal response of a linear SISO plant

Then

$$M\,(\omega) = \frac{y_{max}}{u_{max}} \quad \text{and} \quad \phi\,(\omega) = \omega\left(t_y - t_u\right) \tag{2.190}$$

Applying sinusoids at many different frequencies, however, could be laborious and time consuming. A faster computer aided alternative is therefore introduced below.

It will be recalled that, in theory, $G(s)$ is the Laplace transform of the unit impulse response of the plant. The Laplace transform becomes the Fourier transform if $s = j\omega$. Then $G\,(j\omega)$ is the Fourier transform of the unit impulse response of the plant. The Fourier transform of the unit impulse, $\delta(t)$, itself is

$$\int_0^\infty \delta(t) e^{-j\omega t} \mathrm{d}t = 1 \tag{2.191}$$

indicating that $\delta(t)$ has a flat Fourier spectrum, meaning that it is composed of an infinite continuum of sinusoidal components having the same magnitude over an infinite frequency range. The unit impulse function, however, cannot be applied to a plant in practice, because it is infinite in magnitude for an infinitesimal time, but an alternative method is possible using a realisable input. Let the Fourier transforms of $y(t)$ and $u(t)$ be, respectively, $Y\,(j\omega)$ and $U\,(j\omega)$. Then by analogy with the Laplace transfer function, the frequency domain transfer function is

$$G\,(j\omega) = \frac{Y\,(j\omega)}{U\,(j\omega)}. \tag{2.192}$$

In principle, any $u(t)$ could be used provided it has sufficiently rich frequency content over the frequency range, $\omega \in (0,\ \omega_b)$, to adequately excite the plant. Then $u(t)$ would be applied in real time, while both $y(t)$ and $u(t)$ are data logged. The Fourier transforms, $Y\,(j\omega)$ and $U\,(j\omega)$, would then be computed numerically to yield points on the corresponding magnitude and phase functions.

$$M_u\,(\omega) = |U\,(j\omega)|\,, \phi_u\,(\omega) = \angle M_u\,(\omega) \tag{2.193}$$

and

$$M_y\,(\omega) = |Y\,(j\omega)|\,, \phi_y\,(\omega) = \angle Y_u\,(\omega) \tag{2.194}$$

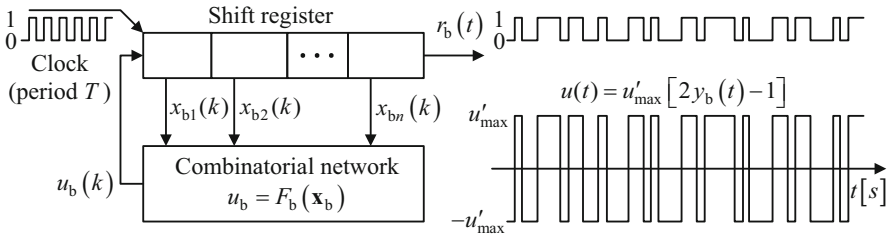


Fig. 2.21 Pseudo-random binary sequence (PRBS) generator

Then the required magnitude and phase of $G(j\omega)$ would be obtained as

$$M(\omega) = \frac{M_y(\omega)}{M_u(\omega)} \quad \text{and} \quad \phi(\omega) = \phi_y(\omega) - \phi_u(\omega). \tag{2.195}$$

It remains to find a realisable form of $u(t)$ for which $|U(j\omega)|$ is nearly constant for $\omega \in (0, \omega_b)$. White noise has a perfectly flat Fourier spectrum for $\omega \in (0, \infty)$ and therefore a realisable signal approximating this would be suitable. Such a signal is the pseudo-random binary sequence (PRBS). This is a discrete binary signal, $r_b(t)$, produced by an algorithm emulating a shift register with a combinatorial network in the feedback loop [13]. The input of this network is the binary state, $\mathbf{x}_b = [x_{b1}\ x_{b2} \ldots x_{bn}]^T$, of the register and its binary output, u_b, is the register input, as shown in Fig. 2.21. The integer, k, increases by 1 every clock pulse.

The term, pseudo-random, applies because the sequence of register states repeats every n_c clock pulses and is strictly deterministic. The Boolean function, $F_b(\mathbf{x}_b)$, is chosen so that $n_c = 2^n - 1$, which is the maximal length sequence to achieve the best approximation to randomness for a given shift register length. Importantly, for a good approximation to the continuous $G(j\omega)$ to be obtained for the highest angular frequencies approaching, ω_b, the clock frequency, $1/T$, must be chosen several times greater than $f_b = \omega_b/(2\pi)$. The recommendation is $1/T > 5\omega_b/(2\pi) \Rightarrow$

$$T < \frac{2\pi}{5\omega_b}. \tag{2.196}$$

The amplitude, u'_{max}, of the excitation signal, $u(t)$, sent out to the plant should be less than the physical control saturation limit, u_{max}, to ensure nominally linear operation. A safe value would be

$$u'_{max} = 0.5u_{max}. \tag{2.197}$$

The computationally efficient discrete fast Fourier transform [14] is employed for the numerical computation of $M(\omega)$ and $\phi(\omega)$.

The frequency domain transfer function, $G(j\omega)$, will be expressed in the form of Bode plots. These can be produced by real time hardware 'in the loop' systems

whose software implements the PRBS and Fourier transform based method such as MATLAB® used with dSPACE® for on-line identification. This consists of two plots, one displaying $M(\omega)$ and the other $\phi(\omega)$. In both plots, the abscissa is ω plotted on a logarithmic scale. $M(\omega)$ is plotted in decibels, meaning

$$M_{\mathrm{dB}}(\omega) \triangleq 20\log_{10}\left[M(\omega)\right]. \tag{2.198}$$

This is called the Bode magnitude plot and this alone enables the parameters of transfer function (2.188) to be estimated using the methods presented in the following subsection. As will be seen in Sect. 2.3.3.10, all these methods are useful for modeling plants with poles and/or zeros in the right half of the s-plane but only give the magnitudes of the real parts of the poles and zeros. To determine which poles and/or zeros lie in the right half of the s-plane, more information is required and this is obtained from the graph of $\phi(\omega)$.
The frequency domain transfer function corresponding to (2.188) is

$$G(j\omega) = K_{\mathrm{dc}} \frac{\prod_i \left(1 + j\frac{\omega}{\nu_i}\right) \prod_i \left(1 - \frac{\omega^2}{\nu_{\mathrm{n}i}^2} + j\frac{2\eta_i\omega}{\nu_{\mathrm{n}i}}\right)}{(j\omega)^q \prod_i \left(1 + j\frac{\omega}{\omega_i}\right) \prod_i \left(1 - \frac{\omega^2}{\omega_{\mathrm{n}i}^2} + j\frac{2\zeta_i\omega}{\omega_{\mathrm{n}i}}\right)}. \tag{2.199}$$

The corresponding magnitude function is therefore

$$M(\omega) = K_{\mathrm{dc}} \frac{\prod_i \sqrt{1 + \frac{\omega^2}{\nu_i^2}} \prod_i \sqrt{\left(1 - \frac{\omega^2}{\nu_{\mathrm{n}i}^2}\right)^2 + 4\frac{\eta_i^2\omega^2}{\nu_{\mathrm{n}i}^2}}}{\omega^q \prod_i \sqrt{1 + \frac{\omega^2}{\omega_i^2}} \prod_i \sqrt{\left(1 - \frac{\omega^2}{\omega_{\mathrm{n}i}^2}\right)^2 + 4\frac{\eta_i^2\omega^2}{\omega_{\mathrm{n}i}^2}}}. \tag{2.200}$$

The expression for the Bode magnitude plot of this general transfer function, using definition (2.198), is as follows.

$$\begin{aligned} M_{\mathrm{dB}}(\omega) =& 20\log_{10}(K_{\mathrm{dc}}) - 20q\log_{10}(\omega) + \sum_i 10\log_{10}\left(1 + \frac{\omega^2}{\nu_i^2}\right) \\ & - \sum_i 10\log_{10}\left(1 + \frac{\omega^2}{\omega_i^2}\right) + \sum_i 10\log_{10}\left[\left(1 - \frac{\omega^2}{\nu_{\mathrm{n}i}^2}\right)^2 + 4\frac{\eta_i^2\omega^2}{\nu_{\mathrm{n}i}^2}\right] \\ & - \sum_i 10\log_{10}\left[\left(1 - \frac{\omega^2}{\omega_{\mathrm{n}i}^2}\right)^2 + 4\frac{\eta_i^2\omega^2}{\omega_{\mathrm{n}i}^2}\right]. \end{aligned} \tag{2.201}$$

The following subsections present methods for estimating the plant parameters from experimentally obtained Bode magnitude plots.

2.3.3.2 DC Gain of a Plant with No Pure Integrators

In this case, $q = 0$ in (2.201), which then satisfies

$$M_{\mathrm{dB}}(0) = 20\log_{10}(K_{\mathrm{dc}}). \tag{2.202}$$

If $M_{\mathrm{dB}}(0)$ could be measured, K_{dc} could be estimated but the logarithmic frequency scale of the Bode plot does not permit this to be shown. If, however, the minimum frequency, $\omega_{\min}$, of the Bode magnitude plot satisfies $\omega_{\min} < \omega_{\mathrm{a}}$ where

$$\omega_{\mathrm{a}} \ll \nu_i, \omega_{\mathrm{a}} \ll \nu_{\mathrm{n}i}, \omega_{\mathrm{a}} \ll \omega_i \ \text{ and } \ \omega_{\mathrm{a}} \ll \omega_{\mathrm{n}i}, \forall i, \tag{2.203}$$

then for $\omega \in (\omega_{\min},\ \omega_{\mathrm{a}})$, (2.201) is approximated by

$$M_{\mathrm{dB}}(\omega) = 20\log_{10}(K_{\mathrm{dc}}) \tag{2.204}$$

which is the equation of the low-frequency asymptote of the Bode magnitude plot. This enables K_{dc} to be approximated by reading $M_{\mathrm{dB}}(\omega_{\min})$ and then calculating

$$K_{\mathrm{dc}} = 10^{M_{\mathrm{db}}(\omega_{\min})/20}. \tag{2.205}$$

If necessary, $\omega_{\min}$, should be reduced to ensure that (2.205) is accurate.

The method is illustrated in Fig. 2.22 for the simple plant model,

$$\frac{Y(s)}{U(s)} = \frac{K_{\mathrm{dc}}}{1 + s/\omega_1} = \frac{10}{1 + s/2}. \tag{2.206}$$

Model (2.206) is initially unknown and its DC gain has to be determined from the Bode magnitude plot. Both asymptotes are shown, one of which is the straight line with a slope of -6 [dB/octave] (or -20 [dB/decade]) passing through the point,

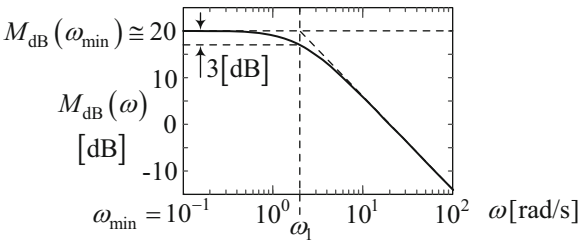


Fig. 2.22 Bode magnitude plot of first-order plant

$(\omega_1,\ M_{\mathrm{dB}}(0))$, which the plot approaches as $\omega \to \infty$. The other asymptote is the horizontal straight line given by (2.204). Provided the lowest frequency, $\omega_{\min}$, on the horizontal axis (0.1 [rad/s] in Fig. 2.22) is an order of magnitude less than the corner frequency, ω_1, or lower, the plot is so close to the asymptote as $\omega \to \omega_{\min}$ that reading $M_{\mathrm{dB}}(\omega_{\min})$ from the plot yields a close approximation to the required $M_{\mathrm{dB}}(0)$. In Fig. 2.22, $M_{\mathrm{dB}}(\omega_{\min}) \cong 20$ [dB]. Assuming $M_{\mathrm{dB}}(0) = M_{\mathrm{dB}}(\omega_{\min})$, (2.205) gives the correct value of

$$K_{\mathrm{dc}} = 10^{20/20} = 10. \tag{2.207}$$

2.3.3.3 The Number of Pure Integrators

Consider now a plant containing q pure integrators, where $q \geq 1$. As ω is reduced to values below ω_{a} of (2.203), then (2.201) is approximated by

$$M_{\mathrm{dB}}(\omega) = 20\log_{10}(K_{\mathrm{dc}}) - 20q\log_{10}(\omega). \tag{2.208}$$

If necessary, $\omega_{\min}$ should be reduced until a nearly straight line segment of the Bode magnitude plot is visible at the low-frequency end. The corresponding asymptote is the straight line of (2.208) with a slope of $-20q$ [dB/decade] (or $-6q$ [dB/octave]). Then if the slope of this line is estimated as S_{d} [dB/decade] or S_{o} [dB/octave], the number of integrators is the nearest integer to $S_{\mathrm{d}}/20$ or $S_{\mathrm{o}}/6$.

2.3.3.4 DC Gain of a Plant with Pure Integrators

Once the Bode magnitude plot of Sect. 2.3.3.3 has been obtained then (2.208) applies. With $\omega = \omega_{\min}$ this yields

$$K_{\mathrm{dc}} = 10^{\left[M_{\mathrm{dB}}(\omega_{\min}) + 20q\log_{10}(\omega_{\min})\right]/20}. \tag{2.209}$$

The method will be illustrated for the plant model

$$\frac{Y(s)}{U(s)} = \frac{K_{\mathrm{dc}}}{s^2(1 + s/\omega_1)} = \frac{10}{s^2(1 + s/2)}. \tag{2.210}$$

Model (2.210) is initially unknown. The Bode magnitude plot from which the DC gain is to be found is shown in Fig. 2.23. The slope, S_{d}, of the plot measured between $\omega = 10^{-2}$ [rad/s] and $\omega = 10^{-1}$ [rad/s] is -40 dB/decade. The number of pure integrators, q, is then the nearest integer to $S_{\mathrm{d}}/20 = 40/20 = 2$.

Thus $q = 2$, agreeing with (2.210). Then $M_{\mathrm{dB}}(\omega_{\min}) = 100$ dB and $\omega_{\min} = 10^{-2}$ [rad/s]. Equation (2.209) then yields $K_{\mathrm{dc}} = 10^{[100+40\times(-2)]/20} = 10$, also agreeing with (2.210).

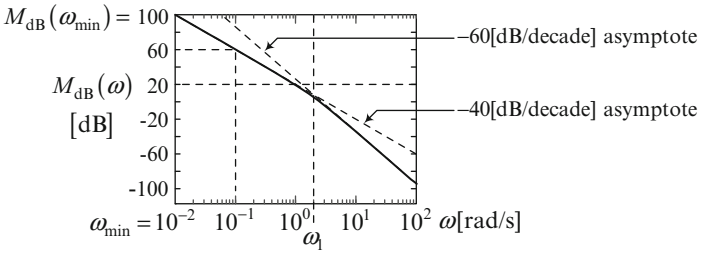


Fig. 2.23 Bode magnitude plot of third order plant containing two pure integrators

2.3.3.5 Asymptotes and the Asymptotic Approximation

Each term under the summation signs in (2.201) is a function having two asymptotes, one for $\omega \to 0$ and the other for $\omega \to \infty$. First consider the terms,

$$M_{\mathrm{dB}i}(\omega) = 10Q\log_{10}\left(1 + \frac{\omega^2}{q_i^2}\right), \tag{2.211}$$

where $Q = 1$ and $q = \nu$ for real negative zeros while $Q = -1$ and $q = \omega$ for real negative poles. As $\omega \to 0$, (2.211) approaches the asymptote,

$$M^0_{\mathrm{dB}i}(\omega) = 10Q\log_{10}(1) = 0, \tag{2.212}$$

which is a straight line with zero slope. As $\omega \to \infty$, $\omega \gg q_i$ and therefore (2.211)approaches the asymptote,

$$M^\infty_{\mathrm{dB}i}(\omega) = 10Q\log_{10}\left(\frac{\omega^2}{q_i^2}\right) = 20Q\left[\log_{10}(\omega) - \log_{10}(q_i)\right], \tag{2.213}$$

which is a straight line satisfying $M^\infty_{\mathrm{dB}i}(q_i) = 0$ with slope $20Q$ [dB/decade], i.e. $6Q$ [dB/octave], noting that the abscissa is $\log_{10}(\omega)$ on the Bode magnitude plot. The Bode magnitude plot and asymptotes for these terms are shown in Fig. 2.24. Importantly, the asymptotes intersect at the corner frequencies, at $\omega = \omega_i$ for the pole and at $\omega = \nu_i$ for the zero. At these frequencies, $M_{\mathrm{dB}i}(\omega)$ has fallen by 3 [dB] for the pole and increased by 3 [dB] for the zero. These features enable the pole magnitudes, which are the corner frequencies, to be estimated from the Bode plot.

An important feature is that $M_{\mathrm{dB}i}(\omega)$ nearly coincides with the asymptote, $M^0_{\mathrm{dB}i}(\omega)$, for $\omega < \omega_{\mathrm{n}i}/10$ and $\omega < \nu_{\mathrm{n}i}/10$. Similarly $M_{\mathrm{dB}i}(\omega)$ nearly coincides with the asymptote, $M^\infty_{\mathrm{dB}i}(\omega)$, for $\omega > 10\omega_{\mathrm{n}i}$ and $\omega > 10\nu_{\mathrm{n}i}$. As will be seen in the following subsections, the parameter estimation process is simplified if the corner frequencies, ω_i, ν_i, $\omega_{\mathrm{n}i}$ and $\nu_{\mathrm{n}i}$, are separated by at least two orders of magnitude.

The piecewise linear function, $A_i(\omega)$, is the concatenation of two segments of the asymptote functions, $M^0_{\mathrm{dB}i}(\omega)$ and $M^\infty_{\mathrm{dB}i}(\omega)$, and is defined as follows.

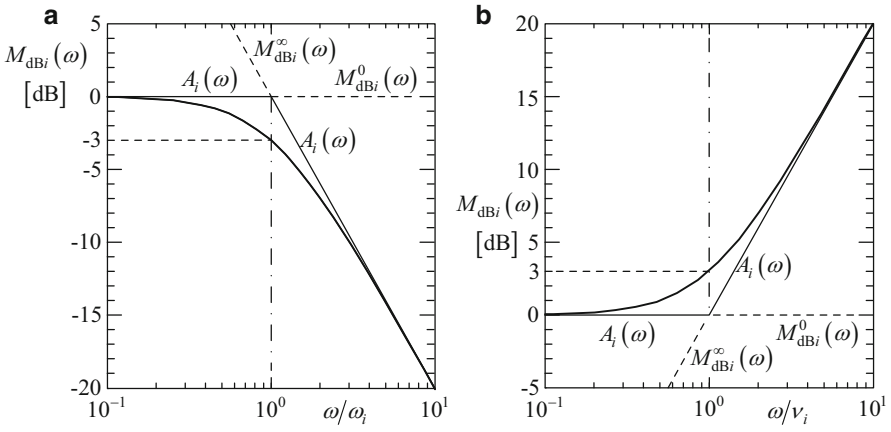


Fig. 2.24 Bode magnitude contributions of pole and zero and their asymptotes. (**a**) Real negative pole. (**b**) Real negative zero

$$A_i(\omega) = \begin{cases} M^0_{\mathrm{dB}i}(\omega), 0 < \omega < q_i \\ M^{\infty}_{\mathrm{dB}i}(\omega), \omega \geq q_i \end{cases}. \tag{2.214}$$

It is evident that this is an approximation to the Bode magnitude plot, $M_{\mathrm{dB}i}(\omega)$. It will be called a *concatenated asymptote function*. The approximation appears better when viewed on larger frequency and amplitude scales due to the closer approach of $M_{\mathrm{dB}i}(\omega)$ to the asymptotes at frequencies far removed from the corner frequency.

Next, consider the terms,

$$M_{\mathrm{dB}i}(\omega) = 10Q\log_{10}\left[\left(1-\frac{\omega^2}{q_{\mathrm{n}i}^2}\right)^2 + 4\frac{d_i^2\omega^2}{q_{\mathrm{n}i}^2}\right], \tag{2.215}$$

of (2.201) where $Q = 1$, $q = \nu$ and $d = \eta$ for the complex conjugate zeros and $Q = -1$, $q = \omega$ and $d = \zeta$ for the complex conjugate poles. As $\omega \to 0$, (2.215) approaches the asymptote,

$$M^0_{\mathrm{dB}i}(\omega) = 10Q\log_{10}(1) = 0, \tag{2.216}$$

which is a straight line with zero slope. As $\omega \to \infty$, $\omega \gg q_{\mathrm{n}i}$ and therefore (2.215), which can be expanded as

$$M_{\mathrm{dB}i}(\omega) = 10Q\log_{10}\left(1 - 2\frac{\omega^2}{q_{\mathrm{n}i}^2} + \frac{\omega^4}{q_{\mathrm{n}i}^4} + 4\frac{d_i^2\omega^2}{q_{\mathrm{n}i}^2}\right),$$

approaches the asymptote,

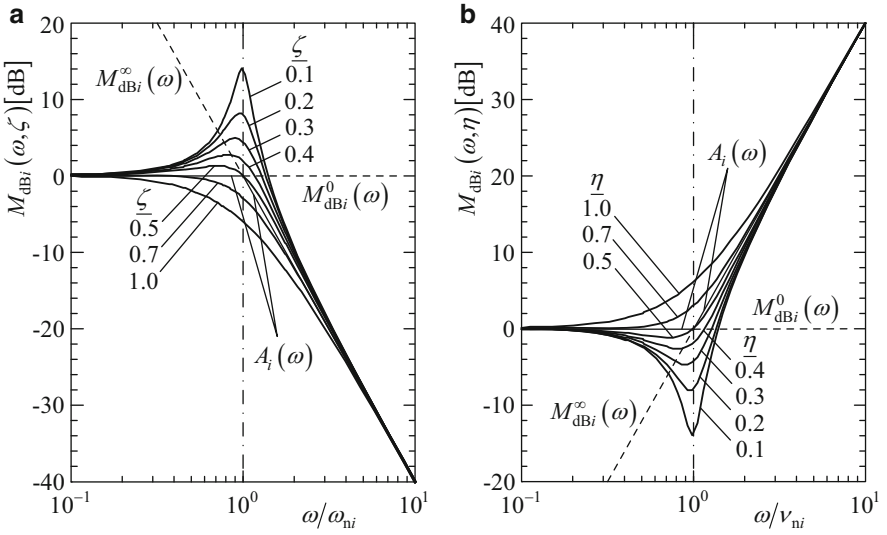


Fig. 2.25 Bode magnitude contributions of complex conjugate pole and zero pairs in the left half of the s-plane together with the asymptotes. (**a**) Complex conjugate pole pair. (**b**) Complex conjugate zero pair

$$M^{\infty}_{\mathrm{dB}i}(\omega) = 10Q\log_{10}\left(\omega^4/q^4_{\mathrm{n}i}\right) = 40Q\left[\log_{10}(\omega) - \log_{10}(q_{\mathrm{n}i})\right], \qquad (2.217)$$

which is a straight line satisfying $M^{\infty}_{\mathrm{dB}i}(q_{\mathrm{n}i}) = 0$ with slope $40Q$ [dB/decade].

The Bode magnitude plot for a range of damping ratios is shown in Fig. 2.25 together with the two asymptotes and the concatenated asymptote function, $A_i(\omega)$.

In this case, the asymptotes intersect at the undamped natural frequency, $\omega_{\mathrm{n}i}$, for the poles and at the frequency, $\nu_{\mathrm{n}i}$, for the zeros.

As in Fig. 2.24, the Bode magnitude plots of Fig. 2.25 closely approach $A_i(\omega)$ at frequencies more than an order of magnitude different from the corner frequency.

The peaks visible in Fig. 2.25a are referred to as resonance peaks, while the dips in Fig. 2.25b are referred to as the anti-resonance dips. The parameters, ζ_i and η_i, can be estimated by determining the amplitudes of the resonance peaks and the anti-resonance dips from the graph of $M_{\mathrm{dB}i}(\omega)$.

Another feature that should be observed is that the frequencies at which the maxima of the peaks in Fig. 2.25a occur, which are the resonance frequencies, $\omega_{\mathrm{r}i}$, are lower than $\omega_{\mathrm{n}i}$. Similarly, the frequencies at which the minima of the dips in Fig. 2.25b occur, which are the anti-resonance frequencies, $\nu_{\mathrm{r}i}$, are lower than $\nu_{\mathrm{n}i}$. This effect is more pronounced as ζ_i and η_i increase and must be taken into account when determining the asymptote intersections by examination of the Bode magnitude plot for estimation of $\omega_{\mathrm{n}i}$ and $\nu_{\mathrm{n}i}$. This is addressed in Sect. 2.3.3.8. For lightly damped cases, however, where $0 < \zeta_i \leq 0.1$ and $0 < \eta_i \leq 0.1$,

$\omega_{ri} \cong \omega_{ni}$ and $\nu_{ri} \cong \nu_{ni}$. Then the asymptote intersections occur approximately at the frequencies of the maxima and minima in Fig. 2.25, rendering the parameter estimation from the Bode magnitude plot more straightforward.

The concatenated asymptote function, $A_i(\omega)$, may be considered to be a good approximation to $M_{\mathrm{dB}i}$ for ζ_i and η_i between 0.5 and 0.7. For lower damping ratios the approximation is not so good but (Appendix A2) the error, $P_i(\omega) = M_{\mathrm{dB}i}(\omega) - A_i(\omega)$, referred to as the resonance peak function, is useful in the parameter estimation process when the complex conjugate pole or zero pairs are close enough for the resonance peak functions to overlap.

In view of the close approach of $M_{\mathrm{dB}i}(\omega)$ to $A_i(\omega)$ at frequencies removed from the associated corner frequency by more than an order of magnitude, if the corner frequencies, ω_i, ν_i, ω_{ni} and ν_{ni}, of a complete plant model are separated by at least two orders of magnitude, a piecewise linear asymptotic approximation, $L_{\mathrm{dB}}(\omega)$ to $M_{\mathrm{dB}}(\omega)$ can be formed by summing the two-segment concatenated asymptote functions, $A_i(\omega)$, from all of the terms in (2.201) together with the DC gain term and the linear term contributed by any pure integrators. Thus,

$$L_{\mathrm{dB}}(\omega) = 20\log_{10}(K_{\mathrm{dc}}) - 20q\log_{10}(\omega) + \sum_i A_i(\omega). \tag{2.218}$$

The vertices of $L_{\mathrm{dB}}(\omega)$ occur at the corner frequencies, ω_i, ν_i, ω_{ni} and ν_{ni}. Since these are parameters of the required transfer function, finding an estimate, $\widehat{L}_{\mathrm{dB}}(\omega)$, of $L_{\mathrm{dB}}(\omega)$, using the graph of $M_{\mathrm{dB}}(\omega)$ is the first step of the parameter estimation.

The closeness of approach of $\widehat{L}_{\mathrm{dB}}(\omega)$ to $L_{\mathrm{dB}}(\omega)$, is of paramount importance to ensure an accurate transfer function model. The task is aided by the knowledge that each of the straight line segments of $L_{\mathrm{dB}}(\omega)$ has a slope that is an integral multiple of -20 [dB/decade], i.e. -6 [dB/octave]. It is usual for all the segment slopes to be zero or negative due to the domination of the poles of the transfer function. Bode plots may be found, however, containing segments of $L_{\mathrm{dB}}(\omega)$ with positive slopes but these are usually those of the open loop transfer function, $G(s)G_{\mathrm{p}}(s)$, of a control system designed with the aid of classical methods including a compensator with transfer function, $G_{\mathrm{p}}(s)$, having the corner frequencies of its zeros lower than those of its poles.

If the corner frequencies are separated by at least two orders of magnitude, it is evident from the foregoing that $M_{\mathrm{dB}}(\omega)$ will have almost straight portions with slopes of nearly $-20n$ [dB/decade], where n is an integer, enabling $\widehat{L}_{\mathrm{dB}}(\omega)$ to be found by simply fitting tangents to $M_{\mathrm{dB}}(\omega)$. If the corner frequencies are closer, however, the nearly straight portions of $M_{\mathrm{dB}}(\omega)$ are less well defined. Although $\widehat{L}_{\mathrm{dB}}(\omega)$ cannot be a good approximation to $M_{\mathrm{dB}}(\omega)$ in these cases, $L_{\mathrm{dB}}(\omega)$ still exists whose corner frequencies are the required plant parameters. It is therefore still important to find $\widehat{L}_{\mathrm{dB}}(\omega)$. Means of calculating the required corner frequencies, using the graph of $M_{\mathrm{dB}}(\omega)$, are developed in Appendix A2.

As all parameter estimation methods are subject to errors, it is recommended, in any case, to generate a Bode magnitude plot, $\widehat{M}_{\mathrm{dB}}(\omega)$, from the estimated transfer

function and compare this with $M_{\mathrm{dB}}(\omega)$. If necessary, adjustments may be made to the transfer function parameters to bring $\widehat{M}_{\mathrm{dB}}(\omega)$ closer to $M_{\mathrm{dB}}(\omega)$.

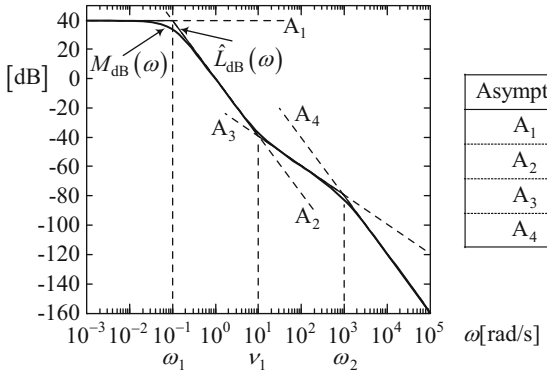
2.3.3.6 Well-Separated Real Poles and Zeros

As shown in Sect. 2.3.3.5, the magnitudes of the real poles and zeros are the corner frequencies, ω_i and ν_i, of the Bode magnitude plot. If these are separated by at least two orders of magnitude, then the Bode magnitude plot contains segments that are nearly linear with relatively sharp changes of slope between them, which enable the piecewise linear approximation, $\widehat{L}_{\mathrm{dB}}(\omega)$, to be fitted easily and accurately to the graph of $M_{\mathrm{dB}}(\omega)$. Also, the vertices of $\widehat{L}_{\mathrm{dB}}(\omega)$ may be assumed to be displaced vertically from $M_{\mathrm{dB}}(\omega)$ by ± 3 [dB]. Moving from left to right along the Bode magnitude plot, a positive change of slope of $\widehat{L}_{\mathrm{dB}}(\omega)$ at a vertex indicates a pole while a negative change of slope indicates a zero. Single poles or zeros cause, respectively, a decrease or increase in slope by 20 [dB/decade] (or 6 [dB/octave]). Occasionally a plant has a real pole of multiplicity, m, which can be detected by a decrease in slope of 20 m [dB/decade] (or 6 m [dB/octave]).

To demonstrate the method, suppose that the Bode magnitude plot, $M_{\mathrm{dB}}(\omega)$, of the plant with transfer function,

$$\frac{Y(s)}{U(s)} = \frac{100\,(1+s/10)}{(1+s/0.1)^2\,(1+s/1,000)}, \tag{2.219}$$

is given and it is required to find the corner frequencies by estimating $L_{\mathrm{dB}}(\omega)$ using the tangent fitting. The Bode magnitude plot is shown in Fig. 2.26. The asymptotes required to form $\widehat{L}_{\mathrm{dB}}(\omega)$ are denoted A_1, A_2, A_3 and A_4. These are drawn tangential to the nearly linear portions of $M_{\mathrm{dB}}(\omega)$. The change of slope from asymptote, A_1, to asymptote, A_2, is $S_{\mathrm{d2}} - S_{\mathrm{d1}} = -40 - 0 = -40$ [dB/decade]. This therefore indicates a double pole. The corresponding corner frequency is $\omega_1 = 0.1$ [rad/s].



Asymptote	Slope [dB/decade]
A_1	$S_{\mathrm{d1}} = 0$
A_2	$S_{\mathrm{d2}} = -40$
A_3	$S_{\mathrm{d3}} = -20$
A_4	$S_{\mathrm{d4}} = -40$

Fig. 2.26 Bode magnitude plot of third-order plant with a double pole, single pole and a zero

The change of slope from asymptote, A_2, to asymptote, A_3, is $S_{d3} - S_{d2} = -20 - (-40) = +20$ [dB/decade].

This therefore indicates a distinct zero. The corresponding corner frequency is $\nu_1 = 10$ [rad/s]. The change of slope from asymptote, A_3, to asymptote, A_4, is $S_{d4} - S_{d3} = -40 - (-20) = -20$, indicating a distinct pole. The corresponding corner frequency is $\omega_1 = 1,000$ [rad/s]. To complete the determination of the transfer function, the DC gain may be calculated using (2.205). Thus $K_{dc} = 10^{M_{db}(10^{-3})/20} = 10^{40/20} = 100$. The transfer function is therefore of the form,

$$\frac{Y(s)}{U(s)} = \frac{K_{dc}\left(1 + s/\nu_1\right)}{\left(1 + s/\omega_1\right)^2 \left(1 + s/\omega_2\right)}. \tag{2.220}$$

Inserting the parameter values calculated above yields transfer function (2.219).

2.3.3.7 Plants with Complex Conjugate Poles and Zeros

The following subsections develop methods for estimation of the parameters of transfer function factors corresponding to complex conjugate poles and zeros, by extracting information from the Bode magnitude plot. The next subsection focuses on second-order plants with one complex conjugate pair of poles and no finite zeros. The procedure developed for this model is applicable, without modification, to higher-order plants provided all the corner frequencies are separated by at least an order of magnitude, as shown in Sect. 2.3.3.9. In Appendix A2, this procedure is extended to cater for plants in which the corner frequencies corresponding to the complex conjugate poles and zeros may be made arbitrarily close.

Regarding notation, with reference to (2.199) the general term representing ω_{ni} or ν_{ni} is υ_{ni} and that representing ζ_i or η_i is d_i.

2.3.3.8 Second Order Underdamped Model

Plants that can be modelled by the second-order underdamped plant model,

$$\frac{Y(s)}{U(s)} = \frac{K_{dc}}{1 + \frac{2\zeta}{\omega_{ni}} s + \frac{1}{\omega_{ni}^2} s^2}, \tag{2.221}$$

are considered here. The corresponding frequency domain transfer function is

$$\frac{Y\left(j\omega\right)}{U\left(j\omega\right)} = \frac{K_{dc}}{1 - \frac{\omega^2}{\omega_n^2} + \frac{2\zeta}{\omega_n} j\omega} \tag{2.222}$$

The magnitude function is therefore

$$M(\omega) = \frac{K_{\mathrm{dc}}}{\sqrt{\left(1-\omega^2/\omega_{\mathrm{n}}^2\right)^2 + 4\zeta^2\omega^2/\omega_{\mathrm{n}}^2}}. \tag{2.223}$$

It will now be shown that a resonance peak occurs for $0 < \zeta < 1/\sqrt{2}$, an expression for which is derived below, enabling ζ to be determined from the Bode magnitude plot. For simplification, if any maximum of (2.223) exists then a minimum of

$$x(\omega) = \left(1-\omega^2/\omega_{\mathrm{n}}^2\right)^2 + 4\zeta^2\omega^2/\omega_{\mathrm{n}}^2 \tag{2.224}$$

also exists for the same value of ω. This satisfies

$$\begin{aligned} \frac{\mathrm{d}x(\omega)}{\mathrm{d}\omega} &= \frac{8\zeta^2\omega}{\omega_{\mathrm{n}}^2} - 4\left(1-\frac{\omega^2}{\omega_{\mathrm{n}}^2}\right)\frac{\omega}{\omega_{\mathrm{n}}^2} = 0 \Rightarrow 2\zeta^2 - 1 + \frac{\omega^2}{\omega_{\mathrm{n}}^2} = 0 \\ &\Rightarrow \omega = \pm\omega_{\mathrm{n}}\sqrt{1-2\zeta^2}. \end{aligned} \tag{2.225}$$

The positive root is the required value of ω, which will be referred to as the *resonance frequency*, ω_{r}. This has to be real and non-zero for the resonance peak to exist, thereby restricting the damping ratio to the range,

$$0 < \zeta < 1/\sqrt{2}. \tag{2.226}$$

Then

$$\omega_{\mathrm{r}} = \omega_{\mathrm{n}}\sqrt{1-2\zeta^2}. \tag{2.227}$$

The resonance peak magnitude is given by (2.223) with $\omega = \omega_{\mathrm{r}}$ of (2.227). Thus

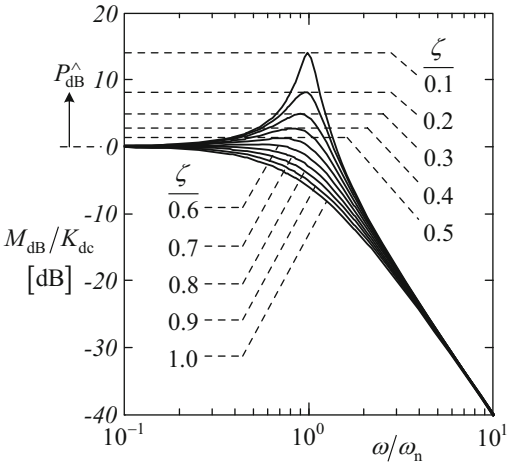
$$M_{\mathrm{p}} = \frac{K_{\mathrm{dc}}}{\sqrt{\left(1-\left(1-2\zeta^2\right)\right)^2 + 4\zeta^2\left(1-2\zeta^2\right)}} = \frac{K_{\mathrm{dc}}}{2\zeta\sqrt{1-\zeta^2}}. \tag{2.228}$$

Note that for $\zeta = 1/\sqrt{2}$, (2.228) yields $M_{\mathrm{p}} = K_{\mathrm{dc}}$ and according to (2.227) this occurs at $\omega_{\mathrm{r}} = 0$ so that $M(\omega)$ monotonically decreases with ω for $\omega > 0$ and therefore, in this case, there is no resonance peak.

On the Bode magnitude plot, the peak of (2.228) for $0 < \zeta < 1/\sqrt{2}$ is

$$M_{\mathrm{pdB}} = 20\log_{10}\left(M_{\mathrm{p}}\right) = 20\log_{10}\left(K_{\mathrm{dc}}\right) + 20\log_{10}\left(\frac{1}{2\zeta\sqrt{1-\zeta^2}}\right). \tag{2.229}$$

The resonance peak w.r.t. the DC level (for $\omega \to 0$) is therefore as follows.



ζ	$P_{\text{dB}}^{\wedge}$
0.1	14.0230
0.2	8.1361
0.3	4.8466
0.4	2.6954
0.5	1.2494
0.6	0.3546

ζ	$P_{\text{dB}}^{\wedge}$
0.05	20.01
0.01	33.97
0.002	47.95
0.0004	61.93
0.00008	75.92
0.000016	89.92

Fig. 2.27 Normalised Bode magnitude plots for underdamped second-order plant models

$$P_{\text{dB}}^{\wedge} = 20\log_{10}\left(\frac{1}{2\zeta\sqrt{1-\zeta^2}}\right). \tag{2.230}$$

Figure 2.27 shows a family of normalised Bode magnitude plots and the resonance peak values for different damping ratios.

Plots for $\zeta > 1/\sqrt{2} \cong 0.7071$ are included to demonstrate the lack of the resonance peak for these cases. Furthermore it is evident that the method is only practicable for $\zeta \leq 0.5$, since P_{dB} could not be accurately read for higher damping ratios. The first inset table of Fig. 2.27 indicates the values of P_{dB} calculated using (2.230) for the Bode magnitude plots shown. The second inset table indicates the values of P_{dB} for much lower damping ratios, which could be relevant to mechanical structures requiring vibration control, space satellites with flexible appendages being a case in point.

It remains to derive an equation for ζ using (2.230) to enable its estimation, given a reading from the Bode magnitude plot. First let

$$P = 10^{P_{\text{dB}}^{\wedge}/20}. \tag{2.231}$$

Then (2.230) becomes

$$P = \frac{1}{2\zeta\sqrt{1-\zeta^2}} \Rightarrow 4P^2\zeta^2\left(1-\zeta^2\right) = 1 \Rightarrow \zeta^4 - \zeta^2 + \frac{1}{4P^2} = 0. \tag{2.232}$$

Solving (2.232) for ζ^2 then yields

$$\zeta^2 = \frac{1 \pm \sqrt{1 - 1/P^2}}{2} \tag{2.233}$$

Since $P_{\text{dB}} \geq 0$, then according to (2.231), $P \geq 1$. If $P = 1$ then (2.233) yields $\zeta^2 = 1/2$. As P increases from 1, then ζ^2 must *reduce* from 1/2.

Hence the negative root is taken in (2.233). Finally the positive root is taken for ζ, resulting in

$$\zeta = \sqrt{\frac{1 - \sqrt{1 - 1/P^2}}{2}}. \tag{2.234}$$

Next, the resonance frequency, ω_{r}, is read from the Bode magnitude plot and ω_{n} is calculated using (2.227) with the estimate of ζ from (2.234). Thus

$$\omega_{\text{n}} = \frac{\omega_{\text{r}}}{\sqrt{1 - 2\zeta^2}}. \tag{2.235}$$

To complete the estimation of the transfer function, K_{dc} is determined using the method of Sect. 2.3.3.2.

2.3.3.9 Well-Separated Complex Conjugate Poles and Zeros

The procedure of Sect. 2.3.3.8 can be applied to estimate the parameters of complex conjugate pole factors, $1/\left(1 + \frac{2\zeta_i}{\omega_{\text{n}i}}s + \frac{1}{\omega_{\text{n}i}^2}s^2\right)$, or complex conjugate zero factors, $1 + \frac{2\eta_i}{\nu_{\text{n}i}}s + \frac{1}{\nu_{\text{n}i}^2}s^2$, of the transfer function if the corner frequencies are separated by at least an order of magnitude. First the amplitude, $P_{\text{dB}i}^{\wedge}$, of the measured resonance peak or anti-resonance dip is determined from the Bode magnitude plot, $M_{\text{dB}}(\omega)$. From (2.231) and (2.234),

$$d_i = \sqrt{\frac{1 - \sqrt{1 - \frac{1}{10^{P_{\text{dB}i}^{\wedge}/10}}}}{2}}. \tag{2.236}$$

where $d_i = \zeta_i$ for the pole factors and $d_i = \eta_i$ for the zero factors. Figure 2.28 shows graphs of this function that can be used to determine d_i directly from $P_{\text{dB}i}^{\wedge}$.

Figure 2.29 shows a Bode magnitude plot for the transfer function model,

$$\frac{Y(s)}{U(s)} = K_{\text{dc}}\left(\frac{1}{1 + \frac{2\zeta_1}{\omega_{\text{n}1}}s + \frac{1}{\omega_{\text{n}1}^2}s^2}\right)\left(\frac{1}{1 + \frac{2\zeta_2}{\omega_{\text{n}2}}s + \frac{1}{\omega_{\text{n}2}^2}s^2}\right). \tag{2.237}$$

The parameters are $\zeta_1 = 0.2$, $\zeta_2 = 0.1$, $\omega_{\text{n}1} = 1$ [rad/s], $\omega_{\text{n}2} = 100$ [rad/s] and $K_{\text{dc}} = 1$. The peak (or dip) magnitude may be determined by measuring the vertical

distance of the tangent from the peak (or the dip) on $M_{\mathrm{dB}}(\omega)$ parallel to one of the two asymptotes intersecting at the vertex of that peak (or dip). There are two alternatives for measuring each peak or dip magnitude. For each resonance peak (or anti-resonance dip) the tangent parallel to the left-hand asymptote meets $M_{\mathrm{dB}}(\omega)$ at the resonance frequency while the tangent parallel to the right-hand asymptote meets $M_{\mathrm{dB}}(\omega)$ at the mirrored resonance frequency. Using the left-hand tangents, on the vertical scale of this figure, the two resonance peak magnitudes are seen to be $P^{\wedge}_{\mathrm{dB1}} \cong 8$ [dB] and $P^{\wedge}_{\mathrm{dB2}} \cong 14$ [dB]. The left-hand graph of Fig. 2.28 then gives the correct values of $\zeta_1 = 0.2$ and $\zeta_2 = 0.1$. Reading the two resonance frequencies from Fig. 2.29 yields $\omega_{\mathrm{r1}} \cong 0.96$ [rad/s] and $\omega_{\mathrm{r2}} \cong 98$ [rad/s].

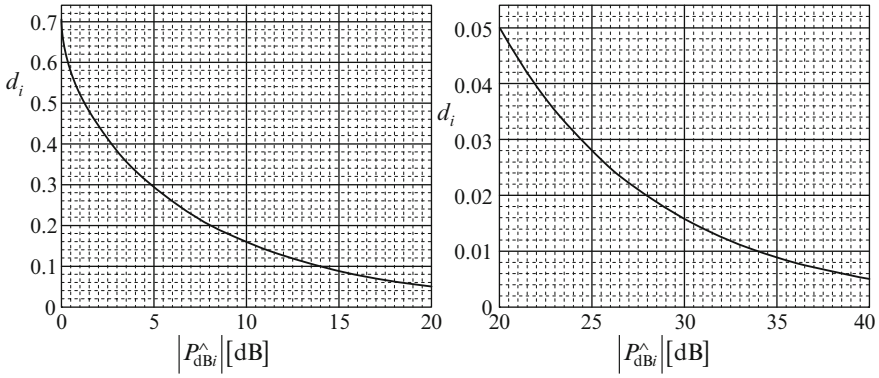


Fig. 2.28 Peak to damping ratio graphs

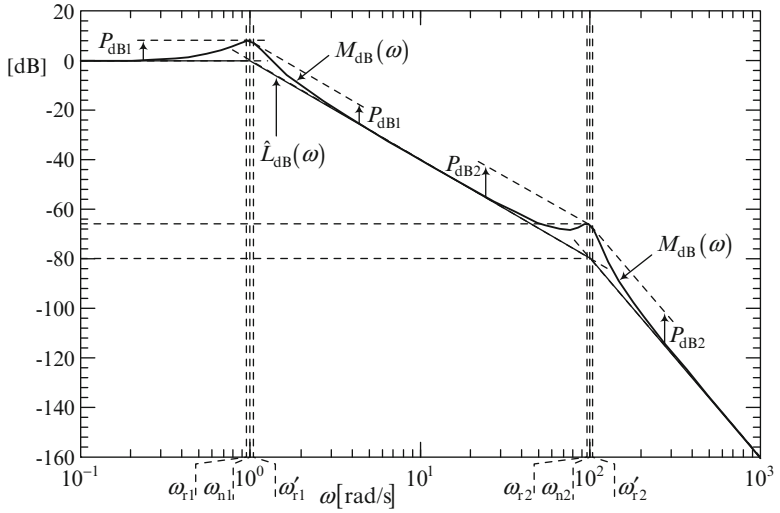


Fig. 2.29 Bode magnitude plot of plant with two underdamped well-separated modes

The undamped natural frequencies are then given by (2.235) as $\omega_{n1} = 1.001$ [rad/s] and $\omega_{n2} = 98.99$ [rad/s]. This precision is acceptable for this graphical method.

With reference to (2.201), the contributions of any complex conjugate zeros to the Bode magnitude plot are of the same form as the contributions from the complex conjugate poles but are opposite in sign. Complex conjugate zeros may therefore be recognised from dips in the plot, their contributions being reflections of those shown in Fig. 2.27 about the horizontal line, $P_{dB} = 0$. The frequencies at which the dips occur will be called *anti-resonance frequencies,* ν_{ri}. It follows that the parameters, η_i and ν_{ni}, may be estimated by a similar method to that above for ζ and ω_n.

Plant modelling with relatively close real poles and zeros or complex conjugate pole or pole–zero pairs is addressed in Appendix A2.

2.3.3.10 Zeros in the Right Half of the *s*-Plane

The transfer function model (2.188) may be modified with a few additions to cater for poles or zeros in the right or left halves of the *s*-plane, as follows.

$$\frac{Y(s)}{U(s)} = G(s) = K_{dc}\frac{\prod_i \left(1 + B_i \frac{s}{\nu_i}\right)\prod_i \left(1 + D_i \frac{2\eta_i}{\nu_{ni}}s + \frac{1}{\nu_{ni}^2}s^2\right)}{s^q \prod_i \left(1 + A_i \frac{s}{\omega_i}\right)\prod_i \left(1 + C_i \frac{2\zeta_i}{\omega_{ni}}s + \frac{1}{\omega_{ni}^2}s^2\right)} \tag{2.238}$$

Setting either A_i or B_i, to -1 then indicates that the ith real pole or zero is a right half plane [RHP] pole at $s_i = \omega_i$ or an RHP zero at $s_i = \nu_i$. Similarly setting C_i or D_i to -1 yields complex conjugate RHP pole pairs at $s_{i,i+1} = \omega_{ni}\left(\zeta_i \pm j\sqrt{1-\zeta_i^2}\right)$ or RHP zero pairs at $s_{i,i+1} = \nu_{ni}\left(\eta_i \pm j\sqrt{1-\eta_i^2}\right)$. Of course, if any coefficient is set to $+1$ the associated poles or zeros are left half plane (LHP) poles or zeros.

In the frequency domain, the magnitude of (2.238) is given by (2.200) since $|A_i| = |B_i| = |C_i| = |D_i| = 1$. Since $|G(j\omega)|$ is independent of A_i, B_i, C_i and D_i, the Bode magnitude plot is insufficient alone to determine the transfer function, unless the poles and zeros are all known to lie in the left half plane of the *s*-plane, which is often the case. Otherwise, the phase information may be used to determine in which half of the *s*-plane every pole and zero lies. Assuming $K_{dc} > 0$, the phase angle of transfer function (2.238) is given by

$$\phi(\omega) = \sum_i \phi_{rzi}(\omega) + \sum_i \phi_{czi}(\omega) + \sum_i \phi_{rpi}(\omega) + \sum_i \phi_{cpi}(\omega) \tag{2.239}$$

where

$$\left.\begin{aligned} \phi_{\mathrm{rz}i}(\omega) &= \tan^{-1}\left(B_i\frac{\omega}{\nu_i}\right), \quad \phi_{\mathrm{cz}i}(\omega) = \tan^{-1}\left(\frac{2\eta_i D_i\frac{\omega}{\nu_{\mathrm{n}i}}}{1-\frac{\omega^2}{\nu_{\mathrm{n}i}^2}}\right) \\ \phi_{\mathrm{rp}i}(\omega) &= -\tan^{-1}\left(A_i\frac{\omega}{\omega_i}\right), \quad \phi_{\mathrm{cp}i}(\omega) = -\tan^{-1}\left(\frac{2\zeta_i C_i\frac{\omega}{\omega_{\mathrm{n}i}}}{1-\frac{\omega^2}{\omega_{\mathrm{n}i}^2}}\right) \end{aligned}\right\}. \tag{2.240}$$

It is evident that as ω goes from 0 to ∞, the changes in the phase angle contributed by the various poles and zeros are as indicated in Table 2.2. Thus, if any real pole or zero, or any complex conjugate pole or zero pair is transferred to its mirror image location, reflected in the $j\omega$ axis, then the Bode magnitude plot will not change but the phase angle contribution of the transferred poles or zeros will change sign.

Figure 2.30 shows the graphs of the individual phase angle contributions of real poles and zeros and complex conjugate poles and zeros.

Four features may be observed in these phase functions as follows.

1. Each of the phase functions of (2.240) reaches half the maximum contribution indicated in Table 2.2 at the corner frequency. This is confirmed by substitution. Thus

$$\phi_{\mathrm{rp}i}(\omega_i) = -A_i\frac{\pi}{4}, \phi_{\mathrm{rz}i}(\nu_i) = B_i\frac{\pi}{4}, \phi_{\mathrm{cp}i}(\omega_{\mathrm{n}i}) = -C_i\frac{\pi}{2} \text{ and } \phi_{\mathrm{cz}i}(\nu_{\mathrm{n}i}) = D_i\frac{\pi}{2}.$$

2. It appears that the slopes of the phase angle contributions on the RHS of (2.240) have maximum magnitudes at the corner frequencies, ω_i, ν_i, $\omega_{\mathrm{n}i}$ and $\nu_{\mathrm{n}i}$, due to ω being on a logarithmic scale. This will now be proven. Let $x = \log(\omega)$. Then

$$\frac{\mathrm{d}x}{\mathrm{d}\omega} = \frac{1}{\omega} \Rightarrow \frac{\mathrm{d}\omega}{\mathrm{d}x} = \omega. \tag{2.241}$$

First consider the real zeros. Let the ith contribution be denoted

$$\phi_{\mathrm{rz}i}(\omega) = \tan^{-1}(B_i\omega/\nu_i). \tag{2.242}$$

Then, noting that $B_i^2 = 1$,

$$\frac{\mathrm{d}\phi_{\mathrm{rz}i}(x)}{\mathrm{d}x} = \frac{\mathrm{d}\phi_{\mathrm{rz}i}(\omega)}{\mathrm{d}\omega}.\frac{\mathrm{d}\omega}{\mathrm{d}x} = \frac{\omega}{1+\omega^2/\nu_i^2}.\frac{B_i}{\nu_i} \tag{2.243}$$

Using (2.241), a maximum or minimum value of $\mathrm{d}\phi_{\mathrm{rz}i}(x)/\mathrm{d}x$ occurs if

$$\begin{aligned} &\frac{\mathrm{d}^2\phi_{\mathrm{rz}i}(x)}{\mathrm{d}x^2} = 0 \Rightarrow \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\omega}{1+\frac{\omega^2}{\nu_i^2}}\right) = 0 \Rightarrow \left[\frac{\mathrm{d}}{\mathrm{d}\omega}\left(\frac{\omega}{1+\frac{\omega^2}{\nu_i^2}}\right)\right]\frac{\mathrm{d}\omega}{\mathrm{d}x} = 0 \Rightarrow \\ &\omega\frac{\mathrm{d}}{\mathrm{d}\omega}\left(\frac{\omega}{1+\frac{\omega^2}{\nu_i^2}}\right) = 0 \Rightarrow \frac{\mathrm{d}}{\mathrm{d}\omega}\left(\frac{\omega}{1+\frac{\omega^2}{\nu_i^2}}\right) = 0 \Rightarrow 1+\frac{\omega^2}{\nu_i^2} - \omega.\frac{2\omega}{\nu_i^2} = 0 \Rightarrow \omega = \nu_i. \end{aligned} \tag{2.244}$$

Table 2.2 Phase angle contributions of poles and zeros as ω goes from 0 to ∞

Key :	Real poles		Real zeros		Complex conjugate poles		Complex conjugate zeros	
LHP $\equiv$ Left Half Plane RHP $\equiv$ Right Half Plane	LHP: $A_i = +1$	RHP: $A_i = -1$	LHP: $B_i = +1$	RHP: $B_i = -1$	LHP: $C_i = +1$	RHP: $C_i = -1$	LHP: $D_i = +1$	RHP: $D_i = -1$
Contribution [deg]	0 to $-90°$	0 to $+90°$	0 to $+90°$	0 to $-90°$	0 to $-180°$	0 to $+180°$	0 to $+180°$	0 to $-180°$

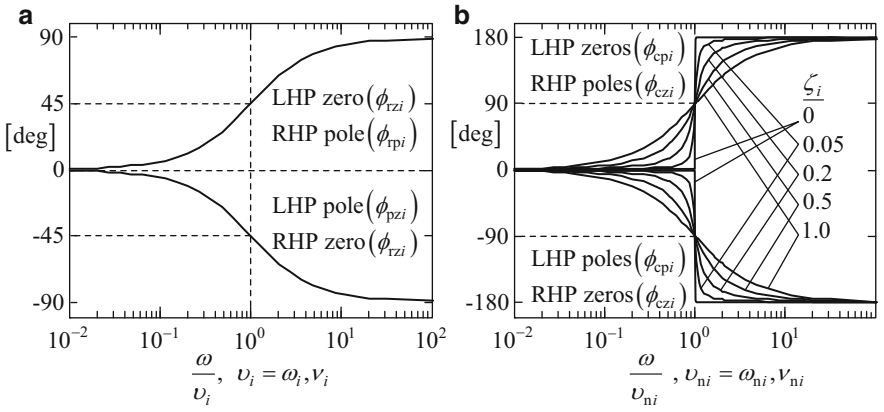


Fig. 2.30 Phase angle contributions of poles and zeros. (**a**) Real pole/zero. (**b**) Complex conjugate pole/zero pair

Since each contribution from the real poles is identical in form to (2.242), it follows that its slope has a maximum magnitude at $\omega = \omega_i$.

Now consider the complex conjugate zeros. Let each phase angle contribution from (2.239) be denoted

$$\phi_{czi}(\omega) = \tan^{-1}\left(\frac{2\eta_i D_i \omega/\nu_{ni}}{1-\omega^2/\nu_{ni}^2}\right) \tag{2.245}$$

Then using (2.241) and noting that $D_i^2 = 1$,

$$\begin{aligned}\frac{\mathrm{d}\phi_{czi}(\omega)}{\mathrm{d}x} &= \frac{\mathrm{d}\phi_{czi}(\omega)}{\mathrm{d}\omega}.\frac{\mathrm{d}\omega}{\mathrm{d}x}\\ &= \omega.\left[\frac{1}{1+\left(\frac{2\eta_i D_i \omega/\nu_{ni}}{1-\omega^2/\nu_{ni}^2}\right)^2}\right].\frac{\left[\left(1-\omega^2/\nu_{ni}^2\right)-\omega.\left(-2\omega/\nu_{ni}^2\right)\right]}{\left(1-\omega^2/\nu_{ni}^2\right)^2}.\frac{2\eta_i}{\nu_{ni}}D_i.\end{aligned}$$

Hence

$$\frac{\mathrm{d}\phi_{czi}(\omega)}{\mathrm{d}x} = \omega.\frac{1+\omega^2/\nu_{ni}^2}{\left(1-\omega^2/\nu_{ni}^2\right)^2+\left(2\eta_i\omega/\nu_{ni}\right)^2}.\frac{2\eta_i}{\nu_{ni}}D_i. \tag{2.246}$$

A maximum or minimum of $\frac{\mathrm{d}\phi_{czi}(\omega)}{\mathrm{d}x}$ occurs if $\frac{\mathrm{d}^2\phi_{czi}(x)}{\mathrm{d}x^2} = 0 \Rightarrow$

$$\left[1-2\left(1-2\eta_i^2\right)\frac{\omega^2}{\nu_{\mathrm{n}i}^2}+\frac{\omega^4}{\nu_{\mathrm{n}i}^4}\right]\left(1+\frac{3\omega^2}{\nu_{\mathrm{n}i}^2}\right) - \left(\omega+\frac{\omega^3}{\nu_{\mathrm{n}i}^2}\right).4\left(\frac{\omega^3}{\nu_{\mathrm{n}i}^4}-\left(1-2\eta_i^2\right)\frac{\omega}{\nu_{\mathrm{n}i}^2}\right)=0. \tag{2.247}$$

Inserting $\omega = \nu_{\mathrm{n}i}$ in the LHS of (2.247) yields $16\eta_i^2 - 16\eta_i^2 = 0$. Hence (2.247) is satisfied. The maximum slope magnitude of this phase contribution therefore occurs at $\omega = \nu_{\mathrm{n}i}$. Since the phase contribution of each complex conjugate pole pair is of the same form as (2.245), then by inspection of (2.239), this contribution has the maximum slope magnitude at $\omega = \omega_{\mathrm{n}i}$. The sign of the derivative in any case is equal to the sign of the corresponding contribution given in Table 2.2.

Setting $\omega = \nu_{\mathrm{n}i}$ in (2.246) yields the derivative of maximum magnitude as

$$\left.\frac{\mathrm{d}\phi_{\mathrm{cz}i}\left(\nu_{\mathrm{n}i}\right)}{\mathrm{d}x}\right|_{\max} = \frac{1}{\eta_i}D_i. \tag{2.248}$$

This explains the increase in the sharpness of the transition between 0 and the extreme value, $D_i\pi/2$, as the damping ratio, η_i, is reduced, which is visible in Fig. 2.30b. In the limit, as $\eta_i \to 0$, the maximum derivative of (2.248) becomes infinite and the phase function becomes a step function of ω, switching between 0 and $D_i\pi/2$ at $\omega = \nu_{\mathrm{n}i}$. Similar relationships hold for $\phi_{\mathrm{cp}i}(\omega)$.

3. With reference to Fig. 2.30, each phase function appears to be an odd function, meaning $\phi(-q) = -\phi(q)$, for all real q, if the origin were to be moved to the point where half the extreme value is reached at the corner frequency. This will now be proven. Consider first the real-zero phase function of (2.240). Thus

$$\phi_{\mathrm{rz}i}(\omega) = \tan^{-1}\left(B_i\frac{\omega}{\nu_i}\right) \tag{2.249}$$

The origin has to be shifted to the point, $[\omega, \phi_{\mathrm{rz}i}(\omega)] = [\nu_i, (\pi/4)\,B_i]$, with ω on a logarithmic scale. To achieve this, the variables are changed to

$$q = \log(\omega) - \log(\nu_i) = \log(\omega/\nu_i) \tag{2.250}$$

and

$$\theta_{\mathrm{rz}i}(\omega) = \phi_{\mathrm{rz}i}(\omega) - (\pi/4)\,B_i \tag{2.251}$$

From (2.250),

$$\omega/\nu_i = e^q. \tag{2.252}$$

Substituting for ω/ν_i in (2.249) using (2.252) and then inserting the resulting expression for $\phi_{\mathrm{rz}i}(\omega)$ in (2.251) then yields the following.

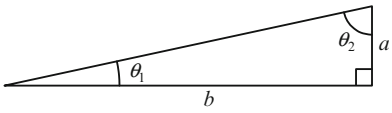


Fig. 2.31 A simple geometric aid

$$\theta_{\mathrm{rz}i}(q) = \tan^{-1}(B_i e^q) - (\pi/4)\, B_i = B_i \left[\tan^{-1}(e^q) - \pi/4\right]. \tag{2.253}$$

If $\theta_{\mathrm{rz}i}(q)$ is an odd function, then

$$\theta_{\mathrm{rz}i}(q) + \theta_{\mathrm{rz}i}(-q) = 0 \Rightarrow \tan^{-1}(e^q) + \tan^{-1}(e^{-q}) - \pi/2 = 0. \tag{2.254}$$

This can be proven geometrically using the right-angled triangle of Fig. 2.31.

Let the side lengths, a and b, be chosen such that

$$\frac{a}{b} = e^q \Rightarrow \frac{b}{a} = e^{-q} \tag{2.255}$$

Then using (2.255) and the geometry of Fig. 2.31,

$$\theta_1 + \theta_2 = \frac{\pi}{2} \Rightarrow \tan^{-1}\left(\frac{a}{b}\right) + \tan^{-1}\left(\frac{b}{a}\right) = \frac{\pi}{2} \Rightarrow \tan^{-1}(e^q) + \tan^{-1}(e^{-q}) = \frac{\pi}{2} \tag{2.256}$$

which satisfies (2.254).

Q.E.D.

This also applies to the real-pole function of (2.240) since it is of the same form as (2.249).

Similar analysis of the complex conjugate zero function of (2.240) will now be carried out. Thus

$$\phi_{\mathrm{cz}i}(\omega) = \tan^{-1}\left(2\eta_i D_i \frac{\omega}{\nu_{\mathrm{n}i}} \Big/ \left(1 - \frac{\omega^2}{\nu_{\mathrm{n}i}^2}\right)\right). \tag{2.257}$$

In this case, the origin has to be shifted to the point, $[\omega, \phi_{\mathrm{rz}i}(\omega)] = [\nu_{\mathrm{n}i}, (\pi/2)\, D_i]$.

The variables are therefore changed to

$$q = \log(\omega) - \log(\nu_{\mathrm{n}i}) = \log(\omega/\nu_{\mathrm{n}i}) \tag{2.258}$$

and

$$\theta_{\mathrm{cz}i}(\omega) = \phi_{\mathrm{cz}i}(\omega) - (\pi/2)\, D_i \tag{2.259}$$

From (2.258),

$$\omega/\nu_{\mathrm{n}i} = e^q \tag{2.260}$$

Substituting for ω/ν_{ni} in (2.257) using (2.260) and then inserting the resulting expression for $\phi_{czi}(\omega)$ in (2.259) then yield

$$\theta_{czi}(q) = D_i\left[\tan^{-1}\left(\frac{2\eta_i e^q}{1-e^{2q}}\right) - \frac{\pi}{2}\right] = -D_i \cot^{-1}\left(\frac{2\eta_i e^q}{1-e^{2q}}\right) \tag{2.261}$$

If $\theta_{czi}(q)$ is an odd function, then so would be

$$f(q) = -D_i \cot\left[\theta_{czi}(q)\right] = \frac{2\eta_i e^q}{1-e^{2q}} \tag{2.262}$$

Then

$$f(q) + f(-q) = 0 \Rightarrow \frac{e^q}{1-e^{2q}} + \frac{e^{-q}}{1-e^{-2q}} = 0. \tag{2.263}$$

The LHS of (2.263) is $\dfrac{\left(1-e^{-2q}\right)e^q + \left(1-e^{2q}\right)e^{-q}}{\left(1-e^{2q}\right)\left(1-e^{-2q}\right)} = \dfrac{e^q - e^{-q} + e^{-q} - e^q}{2-e^{2q}-e^{-2q}}$

$= 0.$

Q.E.D.

4. The observation of Fig. 2.30 indicates that if ω is lower than the corner frequency by at least an order of magnitude, the phase angle contribution is negligible but if ω is greater than the corner frequency by at least an order of magnitude, the phase angle contribution is near its maximum value. Consequently, if the corner frequencies of a transfer function are separated by an order of magnitude or more, the changes in the phase angle due to the individual real poles and zeros and the individual complex conjugate pole and zero pairs will be clearly visible, thereby identifying which half of the s-plane the poles and zeros lie.

The following demonstration stresses that the phase information is essential in frequency domain-based plant identification if it is not known in advance which half of the s-plane the poles and/or zeros lie. It also shows how the LHP and RHP allocations are made. The following three different plant transfer functions are taken since they have the same Bode magnitude functions.

$$\left.\begin{aligned} G_1(s) &= \frac{1}{s\left(1+\frac{s}{\omega_1}\right)}\cdot\frac{1+\frac{2\eta}{\nu_n}s+\frac{1}{\nu_n^2}s^2}{1+\frac{2\zeta}{\omega_n}s+!\frac{1}{\omega_n^2}s^2}, \quad G_2(s) = \frac{1}{s\left(1+\frac{s}{\omega_1}\right)}\cdot\frac{1+\frac{2\eta}{\nu_n}s+\frac{1}{\nu_n^2}s^2}{1-\frac{2\zeta}{\omega_n}s+\frac{1}{\omega_n^2}s^2} \\ G_3(s) &= \frac{1}{s\left(1+\frac{s}{\omega_1}\right)}\cdot\frac{1-\frac{2\eta}{\nu_n}s+\frac{1}{\nu_n^2}s^2}{1+\frac{2\zeta}{\omega_n}s+\frac{1}{\omega_n^2}s^2} \end{aligned}\right\}. \tag{2.264}$$

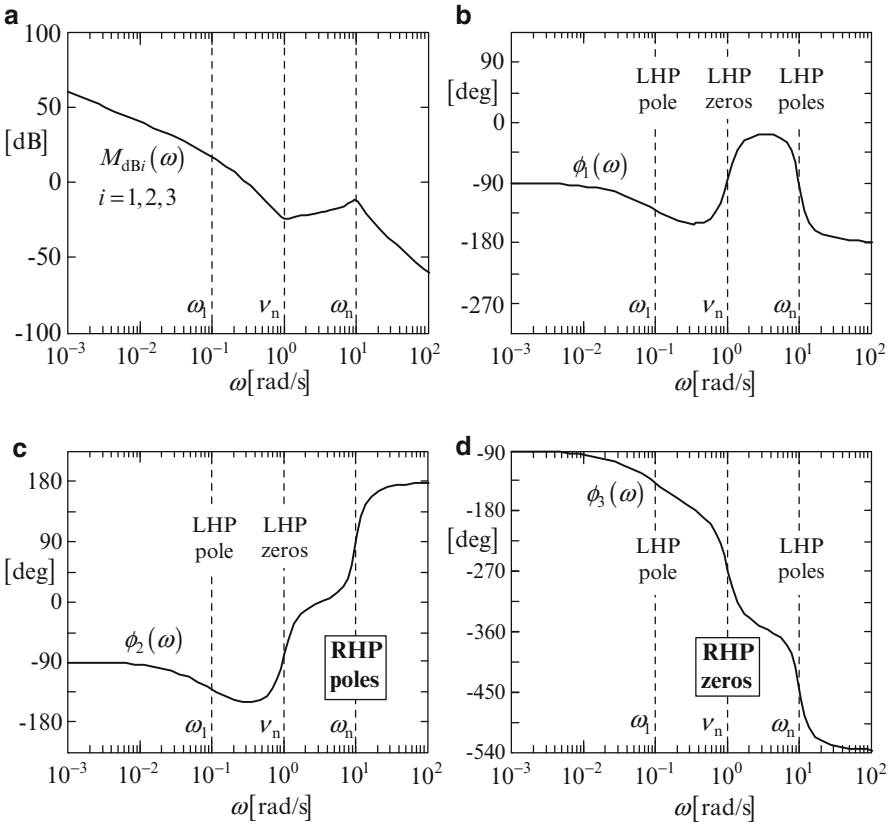


Fig. 2.32 Bode plots of plants with poles or zeros in the RHP. (**a**) Common Bode magnitude plot. (**b**) Phase angle: all LHP poles and zero. (**c**) Phase angle: RHP complex conj. poles. (**d**) Phase angle: RHP comples conj. zero

Here, $\omega_1 = 0.1$ [rad/s], $\omega_n = 10$ [rad/s], $\zeta = 0.2$, $\nu_n = 1$ [rad/s] and $\eta = 0.3$. Figure 2.32 shows the Bode plots.

First, the break frequencies are determined from the Bode magnitude plot of Fig. 2.32a using the methodology of the previous subsections. This enables the identification of a real pole with a corner frequency of ω_1, a complex conjugate zero pair with a corner frequency of ν_n and a complex conjugate pole pair with a corner frequency of ω_n. The damping ratios, ζ and η, can also be estimated using the methods of the previous sections.

The differences between the phase plots are very pronounced due to the separation of the corner frequencies by an order of magnitude. Essentially, the slope of the phase plot for each pole or zero at its corner frequency indicates the half of the s-plane in which the poles or zeros lie. LHP poles have phase plots with negative slopes and LHP zeros have phase plots with positive slopes at the corner frequency.

RHP poles have phase plots with positive slopes and LHP zeros have phase plots with negative slopes at the corner frequency.

2.3.4 Recursive Parameter Estimation: An Introduction

2.3.4.1 The z-Transfer Function Model

A digital controller provides a piecewise constant control input, usually with a constant update period of h seconds. Then a sample/hold unit is placed at the output of an LTI plant with the samples taken at the same instants as the control updates. It is then found that the plant, when observed through the piecewise constant input and sample/hold output obeys a linear difference equation. If the corresponding Laplace transform relationship between these signals is formed, then it is found that the complex variable, s, appears in the form, e^{sh}, throughout the transfer function. This is then simplified by defining a new complex transform variable,

$$z = e^{sh}. \tag{2.265}$$

This gives the z-transfer function of the plant as the ratio of two polynomials in z, the general form of which is given at the beginning of the following subsection. This basic knowledge regarding the z-transfer function is all that is needed here but the underlying theory is fully developed in Chap. 3, Sect. 3.4.3.

2.3.4.2 Statement of the Problem and Approach

In this subsection, the problem of determining the z-transfer function model of a nominally linear time invariant plant by correlating sampled inputs and outputs is addressed. The z-transfer function model is in the general form,

$$\mathbf{Y}(z) = \frac{\mathbf{B}_1 z^{-1} + \mathbf{B}_2 z^{-2} + \cdots + \mathbf{B}_n z^{-n}}{1 + a_1 z^{-1} + a_2 z^{-2} + \cdots + a_n z^{-n}} \mathbf{U}(z), \tag{2.266}$$

where a_i, $i = 1, 2, \ldots, n$, are constant characteristic polynomial coefficients, $\mathbf{B}_i \in \Re^{m \times r}$, $i = 1, 2, \ldots, n$, are constant matrices and n is the plant order. This models SISO plants by setting $m = r = 1$.

Recursive parameter estimation schemes exist that instead estimate the parameters of a discrete state space model once the state representation has been chosen. This, however, requires the theory covered in Chaps. 3 and 5, and is less straightforward than working with (2.266). In any case, a state space model (Chap. 3) can be formed from (2.266).

It is not possible to offer a universal solution to the recursive parameter estimation problem, which is rendered particularly difficult in practice by (a) the presence of

plant and measurement noise, the latter being particularly troublesome, and (b) the uncertainty of the model order in some cases. Many different algorithms may be devised, all of which work 'perfectly' in an ideal noise-free environment and if the order of the chosen model is correct. In practice, the best approach depends upon the particular application. An exhaustive treatment of the subject warrants a dedicated work, such as that by Ljung [16]. The purpose of this subsection is to introduce the subject by developing a particular approach and presenting simulation-based demonstrations of applications. There is actually much room for invention here and the reader is encouraged to devise schemes and try them out.

The approach taken here is to form a difference equation corresponding to (2.266), which consists of one or more linear equations relating the regularly sampled input and output measurements to the transfer function coefficients. These equations are repeated for past input and output measurements to form a completely determined set that may be expressed in the matrix form,

$$\mathbf{M}(k)\mathbf{p} = \mathbf{v}(k), \tag{2.267}$$

where $\mathbf{M}(k)$ is a matrix, which will be referred to as the *solution matrix*, and $\mathbf{v}(k)$ is a column vector, whose elements are the input and output measurements, k is the sample number and $\mathbf{p}$ is a column vector consisting of the plant parameters to be estimated. The estimate of $\mathbf{p}$ will be denoted by $\widehat{\mathbf{p}}$.

2.3.4.3 The Condition Number

Let the solution matrix, $\mathbf{M}(k)$ be square and non-singular. Consider the solution,

$$\widehat{\mathbf{p}}(k) = \mathbf{M}^{-1}(k)\mathbf{v}(k). \tag{2.268}$$

Although, in theory, $\widehat{\mathbf{p}}$ is constant, in practice it will vary due to measurement noise (already introduced in Chap. 1) and plant noise (originating in the actuators and additive to the control inputs). Hence it is shown as a function of k. Under certain circumstances even in the hypothetical case of zero measurement and plant noise sources, $\widehat{\mathbf{p}}$ will vary significantly due to imperfect calculations imposed by the finite wordlength of any digital processor. This effect is marked if $\mathbf{M}(k)$ is *ill conditioned*. Conditioning is a term in numerical analysis pertaining to the accuracy of calculations. In general it is the proportional change in the result of a computation due to an erroneous proportional change of an input parameter due, for example, to rounding errors. Thus, (2.268) may or *may not* produce an accurate result and the conditioning is quantified by the *condition number* of the matrix, $\mathbf{M}(k)$, given by

$$\text{cond}\,(\mathbf{M}(k)) = \frac{|\lambda_{\max}(k)|}{|\lambda_{\min}(k)|}, \tag{2.269}$$

where $\lambda_{\max}(k)$ and $\lambda_{\min}(k)$ are, respectively, the eigenvalues of $\mathbf{M}(k)$ with the maximum and minimum magnitude [15]. The condition number therefore varies between 1 and ∞, the smaller being the better. For this reason, constant or slowly varying inputs are unsuitable. If the control input was constant and the plant stable, then all the variables would settle to constant values, resulting in singularity of $\mathbf{M}(k)$ and an infinite condition number due to at least one of the eigenvalues being zero. It is therefore necessary to excite the plant, preferably by random signals. These must be within the control saturation limits to ensure linear plant operation.

2.3.4.4 The Simultaneous Equations

The z-transfer function model of (2.266) is taken rather than the version in terms of positive powers of z, in order that the corresponding difference equation is in terms of accessible present and past sampled values of $\mathbf{y}(t)$ and $\mathbf{u}(t)$. Thus

$$\left(1 + a_1 z^{-1} + a_2 z^{-2} + \cdots + a_n z^{-n}\right) \mathbf{Y}(z) = \left(\mathbf{B}_1 z^{-1} + \mathbf{B}_2 z^{-2} + \cdots + \mathbf{B}_n z^{-n}\right) \mathbf{U}(z). \tag{2.270}$$

In the discrete domain, this becomes

$$\begin{aligned} &\mathbf{y}(k) + a_1 \mathbf{y}(k-1) + a_2 \mathbf{y}(k-2) + \cdots + a_n \mathbf{y}(k-n) \\ &\quad = \mathbf{B}_1 \mathbf{u}(k-1) + \mathbf{B}_2 \mathbf{u}(k-2) + \cdots + \mathbf{B}_n \mathbf{u}(k-n) \Rightarrow \\ &\mathbf{y}(k) = -a_1 \mathbf{y}(k-1) - a_2 \mathbf{y}(k-2) - \cdots - a_n \mathbf{y}(k-n) \\ &\quad + \mathbf{B}_1 \mathbf{u}(k-1) + \mathbf{B}_2 \mathbf{u}(k-2) + \cdots + \mathbf{B}_n \mathbf{u}(k-n). \end{aligned} \tag{2.271}$$

The basic approach will be to form sets of linear simultaneous equations in the plant parameters that are not underdetermined by taking sufficient input–output samples and solving them for the plant parameters. They will be expressed in the matrix–vector form and the first step towards this is to write the component equations of (2.271) as

$$y_i(k) = \left[a_1 \ \mathbf{b}_{i\,1}^{\mathrm{T}} \ a_2 \ \mathbf{b}_{i\,2}^{\mathrm{T}} \ \ldots \ a_n \ \mathbf{b}_{i\,n}^{\mathrm{T}}\right] \begin{bmatrix} -y_i(k-1) \\ \mathbf{u}(k-1) \\ -y_i(k-2) \\ \mathbf{u}(k-2) \\ \vdots \\ -y_i(k-n) \\ \mathbf{u}(k-n) \end{bmatrix}, \quad i = 1, 2, \ldots, m, \tag{2.272}$$

where $\mathbf{b}_{ij}^{\mathrm{T}}$ is the ith row of $\mathbf{B}_j$. Alternatively, (2.272) will be written as

$$y_i(k) = \underbrace{\left[-y_i(k-1)\ \mathbf{u}^{\mathrm{T}}(k-1)\ -y_i(k-2)\ \mathbf{u}^{\mathrm{T}}(k-2)\ \cdots\ -y_i(k-n)\ \mathbf{u}^{\mathrm{T}}(k-n)\right]}_{\mathbf{w}_i^{\mathrm{T}}} \times \underbrace{\begin{bmatrix} a_1 \\ \mathbf{b}_{i1} \\ a_2 \\ \mathbf{b}_{i2} \\ \vdots \\ a_n \\ \mathbf{b}_{in} \end{bmatrix}}_{\mathbf{p}_i}. \tag{2.273}$$

Then this may be written compactly as

$$y_i(k) = \mathbf{w}_i^{\mathrm{T}}(k)\mathbf{p}_i, \quad i = 1, 2, \ldots, m. \tag{2.274}$$

Since there are $N = (r+1)n$ elements in the plant parameter vector, $\mathbf{p}_i$, N linearly independent equations are needed to calculate $\mathbf{p}_i$. The first equation is (2.274) and a further $N-1$ equations are obtained by repeatedly forming (2.274) for past values of y_i, as follows.

$$\underbrace{\begin{bmatrix} y_i(k) \\ y_i(k-1) \\ \vdots \\ y_i(k-N+1) \end{bmatrix}}_{\mathbf{y}_i(k)} = \underbrace{\begin{bmatrix} \mathbf{w}_i^{\mathrm{T}}(k) \\ \mathbf{w}_i^{\mathrm{T}}(k-1) \\ \vdots \\ \mathbf{w}_i^{\mathrm{T}}(k-N+1) \end{bmatrix}}_{\mathbf{W}_i(k)} \mathbf{p}_i, \quad i = 1, 2, \ldots, m \tag{2.275}$$

The complete set of equations in the plant parameters is therefore

$$\mathbf{W}_i(k)\mathbf{p}_i = \mathbf{y}_i(k), \quad i = 1, 2, \ldots, m. \tag{2.276}$$

For multivariable plants, the coefficients, a_i, $i = 1, 2, \ldots, n$, are estimated m times but advantage can be taken of this by averaging to obtain more accurate results.

2.3.4.5 Algorithm Avoiding Matrix Inversion

An apparently elegant method of solution of (2.276) that avoids the inversion of the solution matrix, $\mathbf{W}_i(k)$, in the directly calculated estimate,

$$\widehat{\mathbf{p}}_i = \mathbf{W}_i^{-1}(k)\mathbf{y}_i(k), \quad i = 1, 2, \ldots, m, \tag{2.277}$$

is to premultiply both sides of (2.276) by $\mathbf{W}_i^{\mathrm{T}}(k)$ and form a matrix-vector differential equation, as follows:

$$\dot{\widehat{\mathbf{p}}}_i + \mathbf{W}_i^{\mathrm{T}}(k)\mathbf{W}_i(k)\widehat{\mathbf{p}}_i = \mathbf{W}_i^{\mathrm{T}}(k)\mathbf{y}_i(k). \tag{2.278}$$

Since the eigenvalues of $\mathbf{W}_i^{\mathrm{T}}(k)\mathbf{W}_i(k)$ are positive and real, then if they are non-zero, $\dot{\widehat{\mathbf{p}}}(t) \to \mathbf{0}$ as $t \to \infty$ and (2.278) approaches $\mathbf{W}_i(k)\widehat{\mathbf{p}}_i = \mathbf{y}_i(k)$. It then follows by comparison with (2.276) that $\widehat{\mathbf{p}}_i(t) \to \mathbf{p}_i$ as required. If the plant is well excited then, in general, the smallest eigenvalues of $\mathbf{M}_i(k) = \mathbf{W}_i^{\mathrm{T}}(k)\mathbf{W}_i(k)$ increase in value and the convergence rate increases. Conversely, if the plant approaches a steady-state condition, then the matrix, $\mathbf{M}_i(k)$, approaches singularity and, as already mentioned, at least one of its eigenvalues approaches zero. In this case the system 'gracefully fails' through its convergence rate reducing to zero instead of a numerical overflow upon attempting to compute (2.277).

The author has found this algorithm works for every plant he has tried but only in simulations in which noise contamination of the measured signals is absent.

2.3.4.6 Noise Contamination

The approach of Sect. 2.3.4.5 could only be successful in practice in cases where the matrix, $\mathbf{W}_i(k)$, is sufficiently well conditioned for plant noise, and particularly measurement noise, to have little effect. Suppose the plant is approaching a settled condition. Then in the presence of noise contamination, the minimum eigenvalue of the matrix, $\mathbf{M}_i(k)$, will tend to increase. In the extreme, if the plant states were constant, $\mathbf{M}_i(k)$ would, in theory, be singular but not so in practice, causing the plant parameter estimates to 'wander' towards incorrect values when using algorithm (2.278). Such random errors in the parameter estimates could be reduced, however, by taking more than the minimum of N component equations in (2.275). This would give $\mathbf{W}_i(k)$ more rows than columns, but $\mathbf{M}_i(k) = \mathbf{W}_i^{\mathrm{T}}(k)\mathbf{W}_i(k)$ would still be square and of dimension, $N \times N$, rendering (2.278) workable. The redundant measurement samples would effectively be filtered in the process of forming $\widehat{\mathbf{p}}_i$, the algorithm performing a type of moving window averaging. In addition, passing $\widehat{\mathbf{p}}_i(k)$ through a low-pass filter with output, $\widehat{\mathbf{p}}_{\mathrm{f}i}(k)$, will result in significant attenuation of the fluctuations in $\widehat{\mathbf{p}}_{\mathrm{f}i}(k)$ compared with those of $\widehat{\mathbf{p}}_i(k)$. A suitable filter would be the IIR (infinite impulse response) filter with unity DC gain defined by

$$\widehat{\mathbf{p}}_{\mathrm{f}i}(k+1) = \alpha\widehat{\mathbf{p}}_{\mathrm{f}i}(k) + (1-\alpha)\widehat{\mathbf{p}}_i(k), \tag{2.279}$$

with $\alpha = e^{-h/T_{\mathrm{f}}}$, where T_{f} is the time constant of the equivalent continuous low-pass filter.

There is, however, another problem that the filtering cannot remove alone. Errors in $\widehat{\mathbf{p}}(k)$ with *non-zero mean values* can occur due to propagation of the noise contamination through squared elements of $\mathbf{W}_i(k)$ within $\mathbf{M}_i(k) = \mathbf{W}_i^{\mathrm{T}}(k)\mathbf{W}_i(k)$. This phenomenon is referred to as *biasing*.

As a simple illustration, consider a single element, $w = \overline{w} + n$, where $\overline{w}$ is the signal that would have occurred without the noise and $_n$ is a noise signal with zero mean value. Incidentally, if the noise contamination is biased, then so will be the parameter estimation. Hence zero mean value of all noise signals is a mandatory condition. The squared element then gives

$$(\overline{w} + n)^2 = \overline{w}^2 + 2\overline{w}n + n^2, \tag{2.280}$$

in which the term, n^2, is responsible for the biasing.

Biasing is not a real problem in some applications but is very significant in others. Hence the following subsection is included.

2.3.4.7 Biasing Minimisation

The diagonal elements of $\mathbf{W}_i^{\mathrm{T}}(k)\mathbf{W}_i(k)$ are w_{jji}^2, $j = 1, 2, \ldots, n$. Biasing can therefore never be eliminated in algorithm (2.278). Hence reverting to (2.277) is considered but with the protection of monitoring the condition number using (2.269), for which practicable real-time algorithms are available. This protection would be given by an algorithm of the following basic form.

$$\left\{\begin{array}{l}\text{If cond}\,(\mathbf{W}_i(k)) < C_{\max},\ \text{then compute}\ \mathbf{p}_i = \mathbf{W}_i^{-1}(k)\mathbf{y}_i(k)\\ \text{else set}\ \mathbf{p}_i\ \text{to the last computed value and skip the inversion.}\end{array}\right\}, \quad i = 1, 2, \ldots, m, \tag{2.281}$$

where $C_{\max} > 1$ is a selected threshold that can be very large, even of the order of 10^8 with floating processors.

Next, the question of biasing in (2.281) must be addressed. For this purpose, (2.277) can be written as

$$\mathbf{p}_i = \frac{\text{adj}\ [\mathbf{W}_i(k)]}{\det[\mathbf{W}_i(k)]}\mathbf{y}_i(k). \tag{2.282}$$

As will be recalled from (2.273) and (2.275), each row of $\mathbf{W}_i(k)$ is formed by shifting the row below two elements to the right, losing the last two elements and reforming the first two elements with new data samples. There are therefore $N - 2$ common elements between every adjacent row. As an example, consider an SISO third-order plant, for (2.273) becomes (2.283) below. The common elements between the rows of $\mathbf{W}_1(k)$ and between $\mathbf{y}_1$ and $\mathbf{W}_1(k)$ can clearly be seen. This, as it stands, would give rise to biasing in (2.282), partially due to the product, adj $[\mathbf{W}_1(k)]\mathbf{y}_1(k)$.

$$\underbrace{\begin{bmatrix} y(k) \\ y\,(k-1) \\ y\,(k-2) \\ y\,(k-3) \\ y\,(k-4) \\ y\,(k-5) \end{bmatrix}}_{\mathbf{y}_1} = \underbrace{\begin{bmatrix} -y\,(k-1) & u\,(k-1)-y\,(k-2) & u\,(k-2)-y\,(k-3) & u\,(k-3) \\ -y\,(k-2) & u\,(k-2)-y\,(k-3) & u\,(k-3)-y\,(k-4) & u\,(k-4) \\ -y\,(k-3) & u\,(k-3)-y\,(k-4) & u\,(k-4)-y\,(k-5) & u\,(k-5) \\ -y\,(k-4) & u\,(k-4)-y\,(k-5) & u\,(k-5)-y\,(k-6) & u\,(k-6) \\ -y\,(k-5) & u\,(k-5)-y\,(k-6) & u\,(k-6)-y\,(k-7) & u\,(k-7) \\ -y\,(k-6) & u\,(k-6)-y\,(k-7) & u\,(k-7)-y\,(k-8) & u\,(k-8) \end{bmatrix}}_{\mathbf{W}_1(k)} \underbrace{\begin{bmatrix} a_1 \\ b_1 \\ a_2 \\ b_2 \\ a_3 \\ b_3 \end{bmatrix}}_{\mathbf{p}_1}. \tag{2.283}$$

Biasing due to this product, however, could be eliminated by forming the successive rows from the component equations (2.273) of (2.275) separated by a delay of $n+1$ sampling periods instead of just one sampling period. Then $\mathbf{y}_i(k)$ and $\mathbf{W}_i(k)$ would be redefined and in the $n = 3$ example, (2.283) would be replaced by

$$\underbrace{\begin{bmatrix} y(k) \\ y\,(k-4) \\ y\,(k-8) \\ y\,(k-12) \\ y\,(k-16) \\ y\,(k-20) \end{bmatrix}}_{\mathbf{y}_1} = \underbrace{\begin{bmatrix} -y\,(k-1) & u\,(k-1) & -y\,(k-2) & u\,(k-2) & -y\,(k-3) & u\,(k-3) \\ -y\,(k-5) & u\,(k-5) & -y\,(k-6) & u\,(k-6) & -y\,(k-7) & u\,(k-7) \\ -y\,(k-9) & u\,(k-9) & -y\,(k-10) & u\,(k-10) & -y\,(k-11) & u\,(k-11) \\ -y\,(k-13) & u\,(k-13) & -y\,(k-14) & u\,(k-14) & -y\,(k-15) & u\,(k-15) \\ -y\,(k-17) & u\,(k-17) & -y\,(k-18) & u\,(k-18) & -y\,(k-19) & u\,(k-19) \\ -y\,(k-21) & u\,(k-21) & -y\,(k-22) & u\,(k-22) & -y\,(k-23) & u\,(k-23) \end{bmatrix}}_{\mathbf{W}_1(k)} \underbrace{\begin{bmatrix} a_1 \\ b_1 \\ a_2 \\ b_2 \\ a_3 \\ b_3 \end{bmatrix}}_{\mathbf{p}_1}. \tag{2.284}$$

Now no common elements reside in the rows of $\mathbf{W}_1(k)$ or the vector $\mathbf{y}_1$. By analogy with (2.284) and with reference to (2.275) and (2.273), the general set of equations is as follows.

$$\underbrace{\begin{bmatrix} y_i(k) \\ y_i\,(k-(n+1)) \\ y_i\,(k-2\,(n+1)) \\ \vdots \\ y_i\,(k-(N-1)\;\;(n+1)) \end{bmatrix}}_{\mathbf{y}_i(k)} = \underbrace{\begin{bmatrix} -y_i\,(k-1) & \cdots & \mathbf{u}^{\mathrm{T}}(k-n) \\ -\,y_i\,(n-(n+1)-1) & \cdots & \mathbf{u}^{\mathrm{T}}(k-(n+1)-n) \\ y_i\,(k-2\,(n+1)-1) & \cdots & \mathbf{u}^{\mathrm{T}}(k-2\,(n+1)-n) \\ \vdots & \cdots & \vdots \\ y_i\,(k-(N-1)\;\;(n+1)-1) & \cdots & \mathbf{u}^{\mathrm{T}}(k-(N-1)\;\;(n+1)-n) \end{bmatrix}}_{\mathbf{W}_i(k)} \underbrace{\begin{bmatrix} a_1 \\ \mathbf{b}_1 \\ \vdots \\ a_n \\ \mathbf{b}_n \end{bmatrix}}_{\mathbf{p}_i} \tag{2.285}$$

To consider the impact of noise contamination, in the same vein as (2.280), $\det[\mathbf{W}_i(k)]$ will contain products of the form, $(\overline{w}_i + n_i)\;\;(\overline{w}_j + n_j)\ldots, i \neq j \neq \ldots$, where $n_i, n_j, \ldots$ are uncorrelated noise signals with zero mean values. Then it is reasonable to suppose that each of the resulting noise signal product combinations, $n_i, n_j, \ldots$, will have zero mean values. Hence $\det[\mathbf{W}_i(k)]$ alone will not contribute any biasing. This is also true of the product, $\mathrm{adj}\,[\mathbf{W}_1(k)]\mathbf{y}_1(k)$, since $\mathrm{adj}\,[\mathbf{W}_1(k)]$ is made up of determinants, each of which will not produce biasing and the post-multiplication by $\mathbf{y}_1(k)$ just generates more product terms of the form already discussed, each of which will have zero mean errors due to the noise signals. The errors in $\det[\mathbf{W}_i(k)]$ and $\mathrm{adj}\;[\mathbf{W}_1(k)]\mathbf{y}_1(k)$ are, however, correlated, since they are formed from the same noisy signal samples. Unfortunately, therefore, it cannot be concluded that algorithm (2.282) is entirely free of biasing, but the author has not yet encountered an example in which this has been troublesome.

2.3.4.8 Filtering

In view of the foregoing, a practicable algorithm could be based on (2.281) with first-order low-pass filtering operating on the parameter estimates defined by (2.279). The filter (2.279) will be transformed into the z-domain (Chap. 3) in preparation for the system block diagram to be presented. Thus

$$z\widehat{\mathbf{P}}_{\mathrm{f}i}(z) = \alpha\widehat{\mathbf{P}}_{\mathrm{f}i}(z) + (1-\alpha)\,\widehat{\mathbf{P}}_i(z) \Rightarrow \widehat{\mathbf{P}}_{\mathrm{f}i}(z) = \left(\frac{1-\alpha}{z-\alpha}\right)\mathbf{I}_N\widehat{\mathbf{P}}_i(z), i = 1, 2, \ldots, m, \tag{2.286}$$

where $\mathbf{I}_N$ is the unit matrix of dimension $N \times N$, which is introduced to remind the reader that in the physical implementation there are N identical first-order filters.

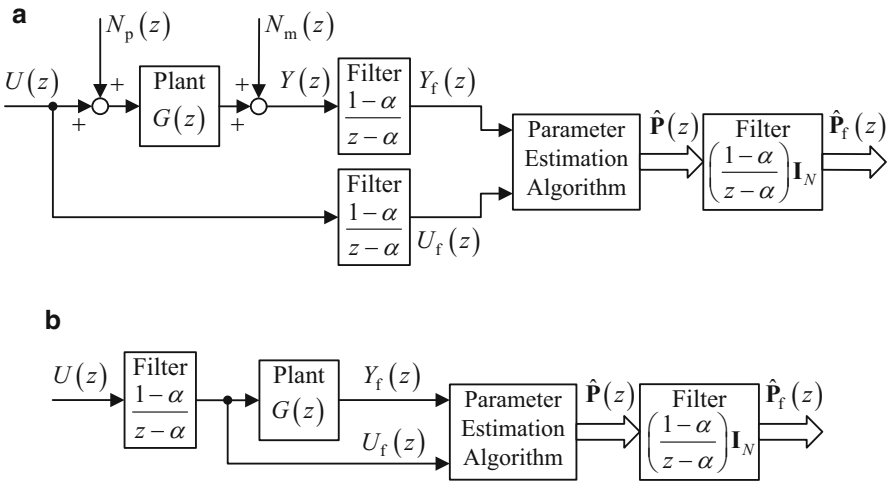


Fig. 2.33 Block diagram manipulation illustrating validity of input–output signal filtering. (**a**) Parameter estimation system with filtering. (**b**) Equivalent simplified deterministic system

In addition, it would be highly desirable to apply similar filtering to the noisy signals before they are processed by the parameter estimation algorithm (2.281), but not with an excessive time constant, T_f; otherwise, there will be insufficient short-term plant excitation to yield a well-conditioned data matrix, $\mathbf{W}_i(k)$. This form of filtering is valid only if identical filtering is applied to the control signals and the measurement signals, as illustrated in Fig. 2.33 for an SISO plant.

In Fig. 2.33a, $N_p(z)$ is the plant noise originating in the control actuators, referred to the plant input and $N_m(z)$ is the measurement noise originating in the measurement instrumentation and referred to the plant output. It is important to realise that since the parameter estimation algorithm is nonlinear, a z-transfer function relationship for it does not exist. The transformed variables are only valid mathematically for the transfer function relationships of the linear parts of the system, i.e. the plant and the filters. To show the filtering arrangement of Fig. 2.33a consider the hyperthetical system operation in the absence of the plant and measurement noise signals. It may then be seen that Fig. 2.33b is obtained from Fig. 2.33a by block diagram manipulation for linear systems. In Fig. 2.33b, the plant input, $U_f(z)$, and the plant output, $Y_f(z)$ are fed directly to the parameter estimation algorithm, which will yield the correct estimates. Since these are the same signals as would be applied to the parameter estimator in the physical system of Fig. 2.33a, the correct estimates will be obtained by identically filtering the plant input and output separately, as shown.

2.3.4.9 Sampling Interval and Modal Timescale Ratio

In order for the set of input and output samples to contain sufficient information about the plant for accurate estimation of the transfer function coefficients, not only must the signal levels be sufficiently; high to mask the effects of any noise contamination and the limited number representation in the digital processor; but they must also capture the changes in the plant variables that result from its dynamical behaviour. This behaviour is characterised by the plant modes. Oscillatory and exponential modes have already been introduced in Chap. 1 as describing the behaviour of feedback control systems and are treated more mathematically in Chap. 3. The term, however, also applies to other dynamical systems, including plants taken in isolation without feedback control, whose modelling is the subject of this chapter. Two examples are briefly discussed in the following paragraphs.

In civil engineering, vibration modes occur in building structures. These are particular types of motion due to the combined elasticity and mass of connected elements in the structure, occurring at specific frequencies, called the eigenfrequencies. These are oscillatory modes.

In thermal systems, if the electrical power supply to an electric kiln is switched on, the temperature of the heating element will rise exponentially with a relatively short time constant towards a steady-state value. On the overall timescale of the kiln operation, the heating element time constant is usually negligible compared with the heating time constant of the workpiece. Then the workpiece temperature will rise nearly exponentially. In any case, the workpiece temperature rise has two exponential components. These are exponential modes.

The estimation window period, over which the recursive parameter estimation algorithm acquires a complete set of input–output samples is, with reference to (2.285), given by

$$T_{\mathrm{w}} = [1 + (N-1) \quad (n+1) - n]\, h = (n+1)\, N h \tag{2.287}$$

where h is the sampling period. In order to collect sufficient information about the dynamic response of the plant to the changing input, T_{w}, should not be very much less than the longest period of the oscillatory modes or the longest time constant of the exponential modes. If T_{w} is too small, in absence of noise, there would be insufficient changes of the variables over the window duration for the solution matrix, $\mathbf{W}_i(k)$, to be sufficiently well conditioned and the noise sources would cause unacceptable errors in the parameter estimates. On the other hand, T_{w} must not be very much greater than the shortest period of the oscillatory modes or the shortest time constant of the exponential modes; otherwise, the input–output samples will be too infrequent to capture sufficient information about the plant behaviour, again yielding inaccurate parameter estimation.

The period of an oscillatory mode or the time constant of an exponential mode are of the same order of magnitude as the reciprocals of the associated poles. These will be defined as the modal timescales. It will be recalled from Chap. 1 that (a) an exponential mode is a first-order mode and therefore only a single pole is associated

with the mode and (b) an oscillatory mode is a second-order mode associated with a complex conjugate pole pair, both poles sharing the same magnitude, given by the undamped natural frequency. The modal timescale of an exponential mode is therefore equal to its time constant, while the modal timescale of an oscillatory mode is $T/(2\pi)$, where T is the period of the oscillation at the undamped natural frequency.

As a general guideline, the estimation window period should satisfy

$$1 < T_{\mathrm{w}}/T < 100 \tag{2.288}$$

for each exponential or oscillatory mode.

For a linear plant model, the modal timescale ratio may be defined as

$$R_{\mathrm{m}} \triangleq \frac{|T_{\max}|}{|T_{\min}|} = \frac{|s_{\max}|}{|s_{\min}|}, \tag{2.289}$$

where $T_{\min} = 1/s_{\max}$ and $T_{\max} = 1/s_{\min}$, $s_{\max}$ and $s_{\min}$ being the modal poles with the largest and smallest *non-zero* magnitudes. As a general guideline,

$$1 < R_{\mathrm{m}} < 100, \tag{2.290}$$

It must be born in mind, however, that the most appropriate limits of (2.288) and (2.290) could vary significantly from one plant to another.

Pure integrators, either distinct or multiple, are associated with particular forms of polynomial exponential modes (Chap. 1), which do not have modal impulse responses that decay on a finite timescale and are not found to pose problems. They are therefore not considered when determining the estimation window period, T_{w}.

If there is only one oscillatory mode or one exponential mode with a modal timescale of T, then only (2.288) has to be satisfied.

The modal timescale ratio is related to another similarly defined quantity called the stiffness ratio, which is the ratio between the largest and smallest real parts of the poles. Most readers will be familiar with the term stiffness in connection with elastic elements of mechanical systems and, by analogy, with control loops in which it is defined as $(de/du_{\mathrm{d}})_{\mathrm{ss}}$, where e is the error between the reference input and the controlled output, u_{d} is a constant external disturbance referred to the control input and the suffix, ss, refers to the steady-state condition. Such stiffness is generally brought about in control loops by means of high gain values. As will be seen in Chap. 10, this is often associated with a closed-loop pole with a large negative real value in the s-domain relative to a group of dominant poles. This gives a large stiffness ratio as defined above. This form of stiffness, not always associated with high control-loop gains, is a property of some dynamical systems that poses a challenge in obtaining an accurate numerical solution to the ordinary differential equations that model them [17]. The reader may have observed that some SIMULINK® simulations run slowly with systems containing fast and slow modes. In fact, various numerical integration algorithms are provided that are specifically

designed to simulate stiff systems. If changing to one of these substantially speeds up the simulation, this is an indication that the system under study may be stiff. In view of this, it is not surprising that stiff plants pose problems when attempting recursive parameter estimation.

Unfortunately, for many plants, $R_m \gg 100$, but in some cases, model order reduction may be possible since the modes associated with poles having relatively large negative real parts will contribute far less to the dynamic response of the plant than modes associated with poles with much smaller real parts, which are dominant poles (Chap. 1). In these cases, recursive parameter estimation is possible with a reduced-order model, but it must be realised that the accuracy of the derived model will deteriorate to some extent for values of R_m on the borderline between having to estimate with a full or reduced-order model, i.e. $R_m \cong 100$.

Of the following two examples, the first is a plant for which the order of the estimated z-transfer function can be the same as the known plant order. The second is an example requiring model order reduction.

Example 2.1 Recursive parameter estimation for attitude control of a flexible spacecraft

The Laplace transfer function between the reaction wheel drive input, $U(s)$, and the rate gyro output, $Y(s)$, for attitude control of a flexible spacecraft is

$$\frac{Y(s)}{U(s)} = K_{DC} \frac{1 + s^2/\nu^2}{s\,(1 + s^2/\omega^2)}, \tag{2.291}$$

where the free natural frequency is $\omega = 1$ [rad/s], the encastre natural frequency is $\nu = 0.8$ [rad/s] and the DC gain is $K_{DC} = 1$.

To determine a suitable estimation window duration, T_w, since there is just one oscillatory mode, the integrator not being considered, the modal time-scale ratio of (2.289) does not apply.

Since the plant order is $n = 3$, the number of plant parameters is $N = 2n = 6$. The algorithm is therefore given by (2.284). Evaluating (2.287) with $h = 0.5$ [s] then yields $T_w = (n + 1)\,Nh = 12$ [s]. The modal timescale of the oscillatory mode is $T = 1/\omega = 1$ [**s**], which is of the same order as T_w. So according to (2.288), the choice of h is suitable for the application. Figure 2.34 shows some simulation results.

The model of (2.291) is implemented in MATLAB®–SIMULINK® with zero-order sample and hold unit placed at the output with a sampling period of $h = 0.5$ [s] and a simulation run with the parameter estimation algorithm of Sect. 2.3.4.7, shown in Fig. 2.33a, with $C_{max} = 10^4$ and the filtering of Sect. 2.3.4.8 with $T_f = 40$ [s]. This value was arrived at by repeated estimation runs, increasing T_f in steps until the random variations of the filtered parameter estimates were reduced to acceptable proportions. A total estimation time of 400 [s] is taken to allow the filtered parameter estimates to reach steady-state (practically). The z-transfer function model obtained is then simulated alongside the sampled continuous model to check that the parameter estimation has been successful.

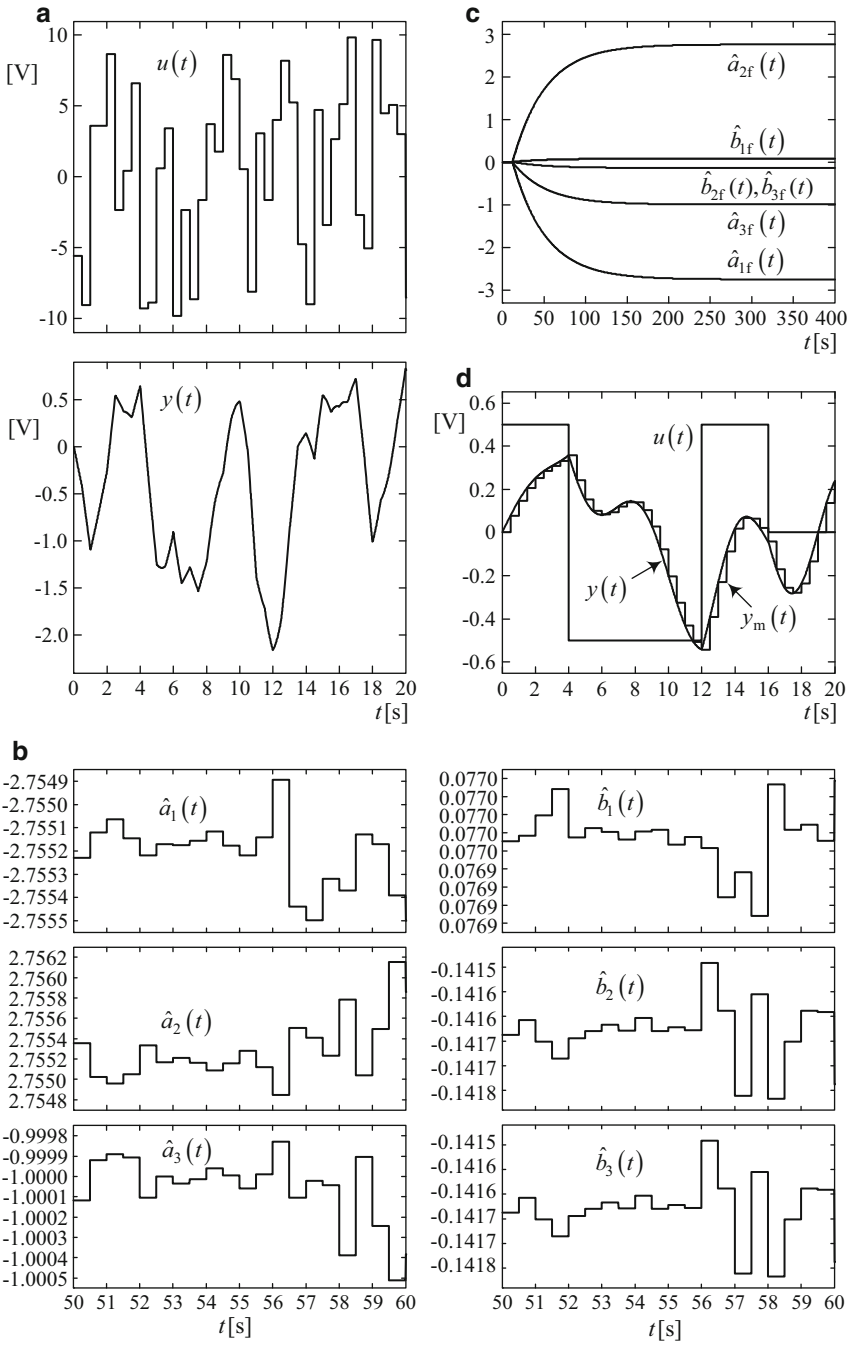


Fig. 2.34 Recursive parameter estimation of a flexible spacecraft. (**a**) Plant input and output variables. (**b**) Unfiltered parameter estimates. (**c**) Filtered parameter estimates. (**d**) Responses of plant and model

The stimulation signal is a piecewise constant random sequence input voltage with a uniform statistical distribution between the reaction wheel drive saturation limits of ± 10 [V], synchronised with the 0.5 [s] sampling. This is shown in Fig. 2.34a.

The reaction wheel noise added to this signal has a Gaussian distribution with a 3σ value of 1 [mV]. The measurement noise signal added to the sensor output is also Gaussian at 1.5×10^{-4} [V] $\equiv 0.54$ [arc sec/s], 3σ, referred to the attitude measurement, noting that 1 [arcsec] $= 1/3600$ [deg], which is realistic with a rate gyro of moderate accuracy. Figure 2.34b shows the unfiltered estimates of the z-transfer function coefficients, updated at $h = 0.5$ [s]. A ten second window is shown so that the piecewise constant estimates can be seen. The fluctuations are, of course, due to the measurement and plant noise inputs. The filter initial conditions are zero, resulting in the exponential convergence of the filtered parameter estimates towards the required constant values, as shown in Fig. 2.34c.

These values were inserted in a SIMULINK® z-transfer function and a piecewise constant input, $u(t)$, applied over a 20 [s] period, resulting in a discrete model output, $y_m(t)$. This is shown together with the output, $y(t)$, of the continuous-time plant simulation in Fig. 2.34d, indicating no visible errors on the scale of the graph.

Example 2.2 Recursive parameter estimation for plate angle control in a throttle valve

There are many examples of electromagnetic control actuators throughout industry that have very small time constants associated with the electrical part of the model and much longer time constants associated with the mechanical part of the model. In such cases model order reduction effectively ignores the electrical time constant. This example is a case in point.

The air-to-fuel ratio of an internal combustion engine is controlled by means of a throttle valve consisting of a pivoted plate mounted in the air intake tube, driven by a DC motor through a gear system. A pre-windup coil spring applies a residual torque ensuring the valve is open in case of an electrical failure. The plate position is measured by a position sensor attached to the plate. The linearised continuous-time model supplied courtesy of Delphi Diesel Systems Ltd has the transfer function

$$\frac{Y(s)}{U(s)} = \frac{K_{dc}}{(1 + sT_1)\ (1 + sT_2)\ (1 + sT_3)}, \tag{2.292}$$

where $K_{dc} = 1.8136$, $T_1 = 0.5464$ [s], $T_2 = 0.0266$ [s] and $T_3 = 3.03 \times 10^{-4}$ [s]. The time constant, T_3, is associated with the armature circuit of the DC motor, the time constant, T_2, is associated with the inertia and viscous friction of the moving plate and the time constant, T_3, results from the coil spring, whose spring constant is insufficient to give the system a complex conjugate pole pair. As it stands, this model has a modal time-scale ratio of $R_m = T_1/T_3 = 1803.3$, which by far violates (2.290). In fact, attempts at estimating a third-order model (not displayed) proved to be totally intolerant of the realistic noise levels. A reduced-order model, however,

would ignore T_3 and yield a new modal time-scale ratio of $R_m = T_1/T_3 = 1803.3$, which satisfies (2.290). The z-transfer function to be estimated is then

$$\frac{Y(z)}{U(z)} = \frac{b_1 z^{-1} + b_2 z^{-2}}{1 + a_1 z^{-1} + a_2 z^{-2}} \tag{2.293}$$

For $n = 2$, the number of plant parameters is $N = 2n = 4$. Then taking $h = 0.1$ [s], (2.287) gives the estimation window duration as $T_w = (n+1)Nh = 1.2$ [s]. Then $T_w/T_1 = 2.196$ and $T_w/T_2 = 45.11$, both of which satisfy (2.288).

The model of (2.292) will be implemented in MATLAB®–SIMULINK® with a zero-order sample and hold unit c sampling period of $h = 0.5$ [s]. A simulation is then run with the parameter estimation algorithm of Sect. 2.3.4.7, shown in Fig. 2.33a, with $C_{max} = 10^4$ and the filtering of Sect. 2.3.4.8 with $T_f = 50$ [s]. This value was arrived at by repeated estimation runs, increasing T_f in steps until the random variations of the filtered parameter estimates were reduced to acceptable proportions. A total estimation time of 300 [s] is taken to allow the filtered parameter estimates to reach a sufficiently close approximation to their steady-state values. The stimulation signal is a piecewise constant random input voltage sequence with levels having a uniform statistical distribution between limits of 0 [V] and 5 [V], to keep the plate between its end stops corresponding to an output voltage range of $0 \text{ [V]} \leq y \leq 10 \text{ [V]}$.

This is updated to synchronise with the 0.1 [s] sampling, as shown in Fig. 2.35a. The plant noise added to this signal has a Gaussian distribution with a 3σ value of 1 [mV]. Similarly, the measurement noise signal added to the plate angle sensor output is also Gaussian with a 3σ value of 1 [mV]. Figure 2.35b shows the unfiltered estimates of the z-transfer function coefficients, updated at $h = 0.1$ [s]. A five second window is shown so that the piecewise constant estimates can be seen. The filter initial conditions are zero, resulting in the exponential convergence of the filtered parameter estimates towards the required constant values, as shown in Fig. 2.35c. These values were inserted in a SIMULINK® z-transfer function with a ramped input, $u(t)$, which is often used in this application, over a 3 [s] period, resulting in a discrete model output, $y_m(t)$. This is shown together with the output, $y(t)$, of the continuous-time plant simulation in Fig. 2.35d. In this case, a small error between the original plant output, $y_m(t)$, and the estimated discrete model output, $y_m(t)$, is visible, which is attributed to the model order reduction rather than biasing. This has been confirmed by carrying out the whole of the above parameter estimation simulation again but with zero noise sources, no change of the error relative to that of Fig. 2.35d being visible.

In some applications, the sampling periods required for recursive parameter estimation may be too long for satisfactory discrete control to be attainable. To solve this problem, means of changing the coefficients of a z-transfer function plant model to cater for a reduced sampling period are given in Chap. 3, Sect. 3.4.4.

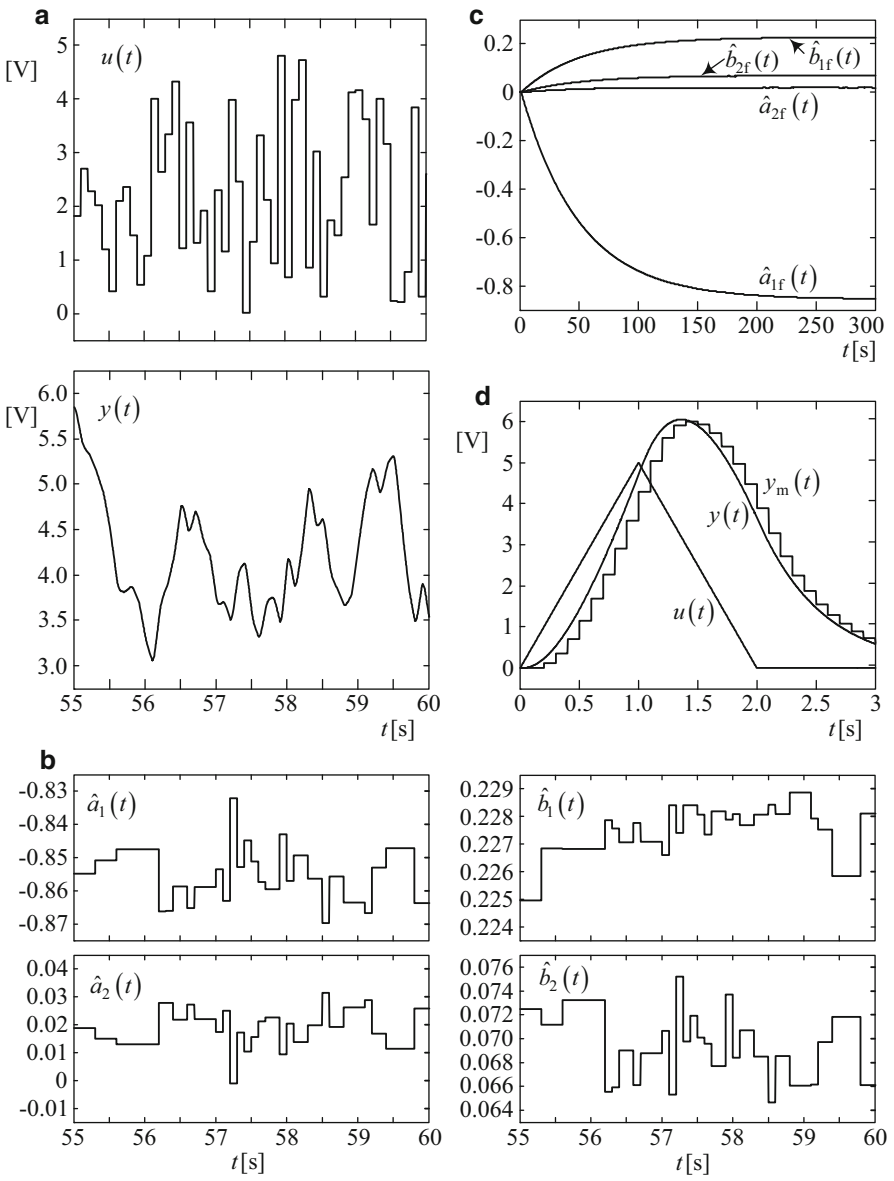


Fig. 2.35 Recursive parameter estimation of a throttle valve for Diesel engines. (**a**) Plant input and output variables. (**b**) Unfiltered parameter estimates. (**c**) Filtered parameter estimates. (**d**) Responses of plant and model

References

1. Batchelor GK (2000) An introduction to fluid dynamics, 2nd edn. Cambridge University Press, Cambridge
2. Popov VL (2010) Contact mechanics and friction. Springer-Verlag, Berlin Heidelberg
3. Taylor JR (2005) Classical mechanics. University Science Books, Herndon, USA. ISBN 1-891389-22
4. Kasdin NJ, Paley DA (2011) Engineering dynamics: a comprehensive introduction. Princeton University Press, Princeton, New Jersey
5. Junkins J, Turner JD (1986) Optimal spacecraft attitude maneuvers. Elsevier, Amsterdam, The Netherlands
6. Quang NP, Dittrich JA (2008) Vector control of three-phase AC machines. Springer, Berlin Heidelberg
7. Haitham AR et al (2012) High performance control of AC drives with Matlab/Simulink models. Wiley, Hoboken, New Jersey
8. Kothari DP, Nagrath IJ (2004) Electric machines. Tata McGraw Hill, New Delhi
9. Pyrhonen J et al (2013) Design of rotating electrical machines. Wiley, Chichester, England
10. Vittek J, Dodds SJ (2003) Forced dynamics control of electric drives. University of Zilina Press, Zilina, Slovakia. ISBN 80-8070-087-7
11. Chiasson JN (2005) Modelling and high performance control of electric machines. Wiley, Hoboken, New Jersey
12. Jaluria Y (2007) Design and optimization of thermal systems, 2nd edn. CRC Press/Taylor and Francis, Boca Raton, Florida
13. Landau ID, Zito G (2006) Digital control systems: design, identification and implementation. Springer-Verlag, London. ISBN 1846280567
14. James JF (2011) A students guide to fourier transforms with applications in physics and engineering. Cambridge University Press, New York. ISBN 0 521 80826/00428
15. Cheney W, Kincaid D (2008) Numerical mathematics and computing. Thomson Brooks/Cole, Monterey. ISBN 978-0-521-17683-5
16. Ljung L (1999) System identification: theory for the user. Prentice Hall, New Jersey
17. Iserles A (2009) A first course in the numerical analysis of differential equations. Cambridge University Press, New York

http://www.springer.com/978-1-4471-6674-0

Feedback Control
Linear, Nonlinear and Robust Techniques and Design with Industrial Applications
Dodds, S.J.
2015, XXV, 1012 p. 674 illus., 3 illus. in color. With online files/update., Softcover
ISBN: 978-1-4471-6674-0