CHAPTER 2

Data Gathering

2.1 INTRODUCTION

The process of performing experiments and recording system inputs and outputs constitutes data gathering. Besides the ability to efficiently extract unknown system parameters applying advanced estimation techniques, which we will cover in the following chapters, data gathering is the other crucial aspect of flight vehicle system identification because of this basic rule: *"If it is not in the data, it cannot be modeled."* This rule applies to all exercises that attempt parameter estimation from experimental data. This is true irrespective of the type of flight vehicle we might attempt to model. Gathered data basically limits, both in terms of scope and accuracy, the model development and parameter estimation. Although data-acquisition-related aspects like parameters (aircraft motion variables and control surface deflections) to be recorded, quality of sensors in terms of accuracy and noise, sampling rate, signal conditioners and data recording equipment play a role in the overall process of data analysis, the most important aspects of data gathering are to do the following:

1. Define the scope of the flight testing
2. Define a suitable sequence of flight maneuvers to be performed at each test point
3. Choose an adequate form of the inputs to excite the aircraft motion in some optimum sense

The last mentioned topic is commonly called in the literature as "optimal input design."

Accuracy and reliability of parameter estimates, obtained applying either the recent modern methods like maximum likelihood or the methods of the past elaborated in Secs. 1.7 and 1.6 respectively, depend heavily on the amount of information available in the vehicle response, which can be maximized through

suitable excitations. This fact was recognized early during the evolution stages as evident from Milliken's statement in 1951 [1]:

> It would appear that the optimum input in a given case is that which best excites the frequency range of interest, and hence the harmonic content of the input should be examined before the test to ensure that it is suitable.

Throughout the history of flight vehicle system identification, this has been, in essence, the guiding principle in designing a proper flight-test maneuver for extracting aerodynamic parameters [2]. The early development in the mid-1940s and early 1950s led to heuristic input signals such as continuous sinusoidal input or pulse input. These designs were mainly governed by the method of analysis applied to extract stability and control derivatives from flight data. In the general field of system identification, theoretical developments on input design started during the 1960s [3–5]. Around the same time, optimal design for flight vehicle system identification started with the work of Gerlach [6, 7]. During the early 1970s, the subject of optimal input designs for aircraft parameter estimation has been extensively researched by Mehra, Gupta, Mulder, Koehler, Marchand, and Plaetschke et al., both from theoretically motivated design as well as more practical oriented approaches [8–13]. They led to inputs based on some suitable combinations of harmonic signals or of multiple pulses. An overview on optimal input and maneuver design is provided by Mulder et al. [14].

Although a lot of effort has been and continues to be spent on optimal input design [15–17], the methodology is plagued by a fundamental problem, namely, the design is based on an a priori model of the flight vehicle. Better a priori models lead to better designs; in other words, for a good input design a good model must be known ahead of flight testing. However, in such a case the estimation from flight data would no longer be needed. There is no exact solution to this paradoxical situation; the only way out is to approach the problem pragmatically and attempt to design an input that is sufficiently broadband in frequency content to cover uncertainties in the a priori model and at the same time easy to adapt to changing flight conditions.

In this chapter we discuss the scope and types of flight testing, providing explanation of maneuvers in a nutshell. The rigorous approaches of optimal input design based on theoretical formulations are briefly summarized without going deeply into mathematical details. A detailed explanation is, however, provided of the basic ideas behind a simple, yet efficient, engineering approach to design a multistep input. Simple software programs are provided to trace the discussion on multiple step inputs. Based on the basic knowledge about the natural frequencies of various modes of motion, typical empirical rules to arrive at time steps for multistep inputs are elaborated. From our experience of modeling different types of flight vehicles, we aim at suggesting a general procedure for performing dynamic flight maneuvers for system identification purposes. Finally, we take a look at other related aspects of data acquisition like a set of measurements required for estimation of stability and control derivatives, sensors, sampling rate, and signal preprocessing.

2.2 FLIGHT TESTING AND MANEUVERS

Pertinent to our specific interest, flight testing can be broadly classified into two categories, namely, the following:

1. Flight testing for performance evaluation
2. Flight testing for system identification

The flight testing for aircraft certification usually falls in the first category whereas that for the aerodynamic database development is in the second group. Usually, a large number of specific maneuvers is required in each case. A part of these maneuvers is common to both whereas the rest is characteristically different. Flight maneuvers used in the proof-of-match to demonstrate the fidelity of aerodynamic database for flight simulators are defined in the Acceptance Test Guide (ATG). As defined by FAA (Federal Aviation Administration) and JAA (Joint Aviation Authorities), roughly 100 to 120 test cases, covering various configurations and maneuvers including special test cases for military aircraft, are usually needed for this purpose [18, 19]. Although special care is exercised in performing these maneuvers in terms of exactness and to minimize external disturbances, basically these maneuvers fall within the two classes already mentioned. Flight testing and its broad spectrum is a vast subject in itself; a detailed account is found, for example, in [20–22]. Here, it is not our aim to give a consolidated account of these two types of testing; rather, we will take a cursory look at them to understand the basic procedures and differences, providing a brief summary of common flight maneuvers.

2.2.1 FLIGHT TESTING FOR PERFORMANCE EVALUATION

Flight testing for performance evaluation is aimed at determining the ability to fulfill the mission in terms of range, fuel consumption, achievable maximum speed, rate of climb, altitude, etc. These tests are usually longer duration tests, requiring specific maneuvers, for example:

1. Acceleration and deceleration
2. Pushover–pull-up
3. Wind-up turn
4. Climb/sawtooth climb
5. Bank-to-bank roll
6. Steady sideslip
7. Landings and takeoffs

A level flight acceleration-deceleration maneuver is performed as follows: starting from a trimmed horizontal level flight, the aircraft is decelerated by

changing the power lever to idle position and maintaining altitude using the pitch control. Once the minimum speed (just above the stall speed) is reached, the power lever is moved from idle to full throttle position, accelerating the aircraft. Once the maximum speed is reached, the power setting is changed to idle, decelerating up to the initial trim speed, at which the test is terminated. In some cases the test is performed for smaller speed variations about the trim condition. The acceleration-deceleration maneuver covers the large range of speed and is usually performed at different altitudes. It allows determination of acceleration capability and longitudinal speed stability; hence, it is also called as longitudinal static stability or speed stability test. The speed variations from minimum to maximum results in a large variation in the angle of attack, and as such the acceleration-deceleration maneuver is used for system identification purposes as well. For such purposes, during the acceleration phase elevator inputs, usually doublet, which we will discuss in the next section, are applied at regular interval to increase the information content in the data.

Pushover–pull-ups, also called roller-coaster, are primarily performed to determine lift and drag characteristics, longitudinal stability, and the elevator trim requirements. The maneuver is initiated from a trimmed level flight by slowly pushing the stick, reducing the angle of attack continuously, which leads to a slow diving motion. Once the maximum speed is reached, the stick is pulled slowly to result in increasing angle of attack, which leads to a slow climbing motion. Once the minimum speed (just above the stall speed) is reached, the stick is once again pushed slowly to return to the starting trim condition. In contrast to the acceleration-deceleration maneuver, pushover–pull-up maneuvers are performed at constant thrust. This maneuver results in a significant variation in the vertical acceleration, typically covering the range of 0 to 2 g and as such is used mostly to determine the drag polars. The maneuver is performed in about 40 to 50 s.

Wind-up turns are used to determine the gradient of "stick force per g," which is a design and certification criterion. It is a measure of the stick force required to change the load factor; it is an indicator of the compromise between the two conflicting requirements of longitudinal stability and maneuverability. A higher slope of stick force vs load factor curve implies higher apparent stability and less maneuverability because a larger pilot force will be necessary to move away from the trim and vice versa. The maneuver starts by initiating a level turn and continuously increasing the bank angle. Increasing bank angle leads to a nose-down motion at increasing speed. Through pitch control the angle of attack is increased to maintain the speed constant. The throttle position is kept fixed. The process of simultaneously increasing the bank angle and angle of attack results in a descending spiraling and wind-up motion with decreasing turn radius, resulting in an increasing load factor. The maneuver is terminated when a g-limit is reached. This is one of the demanding maneuvers for the pilot, particularly for fighter aircraft with higher g-limits. The duration of a wind-up turn is roughly 60 to 80 s and repeated at different nominal speeds. Besides the primary purpose of finding the

gradient of "stick force per *g*," this maneuver also aids in determination of drag and lift coefficients at different angles of attack for a constant Mach number and also in the system identification.

The main aim of the climb/sawtooth climb is to determine the best climb rate at different altitudes to arrive at a best climb schedule to reach a desired altitude in a minimum amount of time. Starting from a level flight, repeated climbs and dives are performed to result in predefined variations in the altitude by applying maximum and idle power alternatively. The airspeed is maintained constant using pitch control. This maneuver is rarely used in system identification; however, it is a part of the proof-of-match process that is necessary to demonstrate the model fidelity.

Determination of the maximum roll capability is important for fighter aircraft, as it mainly governs the ability to turn quickly and change direction of flight. In the case of transport aircraft, adequate roll performance is necessary to quickly compensate for disturbances such as turbulence or asymmetric configurations. Determination of aerodynamic coupling between roll and yaw axes and ability to compensate adverse yaw following an aileron input is part of this test. This test is performed at several trim conditions. The bank-to-bank roll, also called rapid roll, tests are widely used for parameter estimation purposes.

The primary aim of steady sideslip maneuvers is to determine the gradient of rudder deflection and aileron deflection per degree of sideslip. These maneuvers are particularly of interest at low speeds for the landing configuration because larger side slipping might be necessary to compensate for strong crosswind during landing phases. Starting from a horizontal level flight, the maneuver is initiated by applying and holding a constant pedal force, leading to a gradual increase in the angle of sideslip. Constant heading flight is maintained in a banked position by applying ailerons. These maneuvers are widely used in system identification.

The landings and takeoffs are critical flight phases. A careful planning is necessary during these tests because the performance is greatly affected by wind gradients, particularly during landing phases. The ground effect also affects the aircraft performance at low altitudes. Besides the flybys at different low altitudes over the runway, landing maneuvers are also used in the estimation of ground effect. Landings under different external weather conditions and one-engine out on takeoff are an important part of the data gathering for simulator validation.

Besides the preceding elaborated performance evaluation specific tests, other stability-related maneuvers enabling frequency and damping determination are also performed as a part of overall flight-test program. They are primarily used for handling qualities evaluations and for estimation of aerodynamic derivatives. Such maneuvers and those commonly used for system identification purpose, for example, steady heading sideslips, are elaborated in the next section.

2.2.2 FLIGHT TESTING FOR SYSTEM IDENTIFICATION

Estimation of stability and control derivatives is carried out primarily from the dynamic response of an aircraft to specific control inputs. In general, dynamic motion is excited by applying control pulse, step, multistep, or harmonic inputs. A variety of maneuvers is usually necessary to excite dynamic motion about different axes using independent inputs on every control. Dynamic maneuvers allow extraction of a larger number of static as well as dynamic parameters from a single test. Combining several stationary maneuvers at different trim conditions principally allows determination of some of the aerodynamic derivatives, for example, slope of the lift coefficient or the longitudinal static stability parameter. Compared to dynamic maneuvers, this approach is less efficient because a larger number of test points will be necessary and the uncertainties in the estimates might be larger. Usually, the best choice is to combine the dynamic maneuvers and different stationary conditions to extract a single set of aerodynamic derivatives.

A set of common flight maneuvers for system identification consists of the following:

1. Short-period maneuver
2. Phugoid maneuver
3. Pushover–pull-up
4. Level turn
5. Thrust variation
6. Bank-to-bank roll
7. Dutch roll maneuver
8. Steady heading steady sideslip

Besides these maneuvers, as already mentioned in Sec. 2.2.1, acceleration-deceleration is also used for the purpose at hand. The main source for estimation of ground effect is the flybys at different altitudes over the runway and the landings. Generally, for control applications, for safety purposes, and for passenger comfort, excitation of the eigenmodes is suppressed. However, for parameter estimation purposes, it is essential that the dynamic response exhibits frequencies and damping of the oscillatory modes. In general, it is recommended to start each maneuver from a trimmed level flight and allow about 5 to 10 s of steady flight before applying specific control inputs and, depending upon the mode of motion, allow sufficient time after the input to allow the aircraft to oscillate.

The short-period motion, which is a fast responding longitudinal mode, provides the most information to enable estimation of derivatives pertaining to the vertical and pitching motion. The short-period maneuver is initiated from a horizontal level flight by applying an elevator or other longitudinal control input,

which is usually a doublet or multistep input. Doublet input is a two-sided pulse input whereas multistep input consists of a suitable combination of more than two pulses. The time step of such inputs is to be chosen such that the short-period mode is excited well. We will deal with the aspects of arriving at a suitable form of multistep input and choosing proper time steps for such input in the next section. Input amplitude should be such as to result in about $\pm$3- to 4-deg variation in angle of attack about the trimmed value, or in terms of the load factor the variation should be typically $\pm$0.4 to 0.5 *g*. A typical duration of such a maneuver is about 15–20 s. It is performed at different trim speeds. Larger amplitude maneuvers might be necessary to enable estimation of nonlinearities. While applying a multistep elevator input, pilots have to pay attention because in special cases they tend to maintain the angle of attack, particularly during approach; this practice is in contrast to the basic principle of exciting the longitudinal mode resulting in variation in the angle of attack.

The phugoid mode is a long-duration mode, which can be excited either through an elevator pulse or thrust variation. Typically, one full cycle of phugoid motion is required for parameter estimation purposes. For the validation of databases for simulators, the model accuracy has to be demonstrated for three full cycles.

As already mentioned in Sec. 2.2.1, the pushover–pull-ups result in variation in the vertical acceleration and are useful to determine lift and drag characteristics. A level-turn maneuver is initiated from a horizontal level flight by banking the aircraft smoothly into 30- (45- or 60-) deg bank at a rate of 10 deg/s, holding the bank attitude steady for several seconds, then rolling the aircraft to the opposite direction, again holding steady the bank angle for several seconds, and finally returning to the wings-level conditions. Altitude is maintained during the maneuvers through additional power. It results in an S-type flight path in the horizontal plane, the heading at the start and at the end of the maneuver roughly the same. As will be elaborated in Sec. 12.2, this maneuver combined with an elevator doublet input in the banked position is useful to separate out the pitch rate and rate of change of angle-of-attack components of longitudinal motion.

Dynamic thrust input maneuvers are performed applying a longer-duration doublet input to throttle. They are useful to determine the dynamic effects due to thrust variations on the longitudinal motion, for example, on the downwash effect on the horizontal tail. They also aid determination of the effect of the thrust coefficient on the various aerodynamic derivatives, particularly for propeller aircraft where such effects are dominant. The dependencies of the aerodynamic derivatives on the thrust coefficient can also be determined by performing other longitudinal and lateral-directional dynamic maneuvers starting from trims with different thrust levels.

To enable estimation of lateral-directional derivatives, bank-to-bank roll and Dutch roll maneuvers are necessary. A bank-to-bank maneuver is performed applying a series of aileron pulses. Starting from a level flight, sharp aileron input is applied, holding it for some time to roll the aircraft from wings level to

at least 30-deg bank on one side; this is followed by rapidly changing the input and going smoothly through wings level to the opposite bank angle, followed by returning to the wings-level conditions. The test is repeated for different levels of bank angles (30, 45, and 60 deg). The bank-to-bank maneuver is also sometimes called rapid-roll maneuver because the abrupt changes in aileron pulses result in rapid variation in roll rate and acceleration. A typical duration of such a maneuver is 30 to 40 s. The doublet or multistep aileron input with short time steps might not be quite adequate because sufficient time should be allowed for bank-angle buildup; rolling motion corresponds to a first-order system.

In many aircraft, ailerons are coupled to roll spoilers, which are additional flaps on the upper surface of the wings. They are drawn out either on the left or the right wing corresponding to the upper aileron deflection, that is, on the wing going down, and help to improve the overall roll efficiency without being too sensitive at small aileron deflections. For aileron deflections less than a threshold value, the spoilers are inactive; they are deflected proportional to aileron deflection larger than the threshold. For a given aircraft, the threshold aileron deflection for roll spoiler activation is a fixed value, which usually ranges from 3 to 5 deg. Because of this basic construction, ailerons and roll spoilers are deflected proportionately, except for small aileron deflections. In case we are interested in estimating from flight data both aileron and roll-spoiler effectiveness, then it is necessary to perform the bank-to-bank maneuver with small aileron deflection (smaller than the threshold values) during which roll spoilers are inactive, followed by larger aileron input in which case both control surfaces would contribute to the rolling motion. A similar situation arises for aircraft having aileron-to-rudder interconnect. In this case, automatic rudder deflection following an aileron input is detrimental to extraction of independent derivatives, although both surfaces move. Basically, it amounts to the fundamental aspect of generating independent inputs on every control; otherwise, it would not be possible to separate out the effects due to multiple surfaces moving proportionately.

The Dutch roll maneuver provides maximum information on the frequency and damping of this oscillatory mode. It is excited by applying rudder inputs. Usually, several cycles of oscillations are recorded; the duration of the total maneuver could be typically 30–35 s or even longer. The resulting maximum peak-to-peak variation in the angle of sideslip is typically of the order of ± 4 deg, or 0.1-g lateral acceleration. The Dutch roll mode of the basic aircraft is usually lightly damped and can be excited very well with a simple doublet input. Many aircraft are equipped with a yaw damper, which is turned off during maneuvers performed for system identification purposes, because we are mainly interested in the aerodynamic characteristics of the bare airframe. It is required to perform the Dutch roll and bank-to-bank maneuvers at different trim speeds because most of the lateral-directional derivatives depend on the angle of attack.

Both aileron and rudder input maneuvers are necessary not only to allow estimation of the respective control surface effectiveness, but also to improve

estimation of other derivatives due to roll and yaw rates. The aileron input maneuver yields significantly larger roll rates, which are characteristically different in shape than compared to those from the rudder input. In general, it is our experience that better estimation results are obtained by performing the aileron and rudder input maneuvers separately starting with a bank-to-bank maneuver followed by a Dutch roll. This sequence is easy for pilots to execute manually. In some cases we have also applied computerized aileron and rudder inputs simultaneously. In such cases, the starting points of the two inputs should be time skewed, that is, displaced in time. If the starting point is time synchronized, the duration of the two inputs should be different, or an independent maneuver with a single control surface input would be desirable to improve separation of the components due to these two control surfaces.

The steady heading steady sideslip (SHSS) maneuver provides additional information on the directional stability, lateral-directional control, and cross coupling effects. It is usually flown in both the positive and negative sideslipping conditions to allow detection of any asymmetric effects. Such tests are particularly important for landing configurations. SHSS are performed in different variations to cover a large range in angle of sideslips in both directions and at different angles of attack and thrust levels. For example, starting from a horizontal level flight, SHSS is performed by applying abrupt pedal step and maintaining the sideslip angle at some chosen level for roughly 10–15 s, followed by releasing the pedal to neutral position allowing few seconds for dynamic recovery response. This is followed by repeating the procedure in the other direction of side slipping. In each case, the track angle is maintained by adjusting the bank angle applying aileron input. The side slipping conditions should be achieved rapidly, ideally as shown schematically in Fig. 2.1a. However, in practice some overshoot is unavoidable; strong excitation of the Dutch roll is to be suppressed by regulating the pedal input. For larger sideslips, the procedure is similar. The alternative is to perform building up the larger sideslip angles in two steps before releasing the controls to neutral. Yet another variation is to change abruptly from one steady sideslip in one direction to the other direction, without a stabilizing phase for no side slipping conditions, as shown in the two plots on the right side of Fig. 2.1a. If SHSS are carried out for performance evaluations, then the overall procedure is similar, but for maintaining the steady side slipping conditions for a considerably longer time than indicated earlier.

Besides SHSS with abrupt changes in asymmetric trim conditions, a quasi-steady type of maneuver, called wings-level steady sideslip, with gradual variation in side slipping is also useful. Starting from horizontal level flight, rudder is applied gradually to increase angle of sideslip from zero to some chosen value (say 5 or 10 deg), then gradually releasing the pedal pressure to reduce the side slipping, crossing over through zero to angle of sideslip in the opposite direction, finally returning to neutral; see Fig. 2.1b. Throughout the maneuver, the wings-level conditions are maintained applying lateral-stick cross-control. In

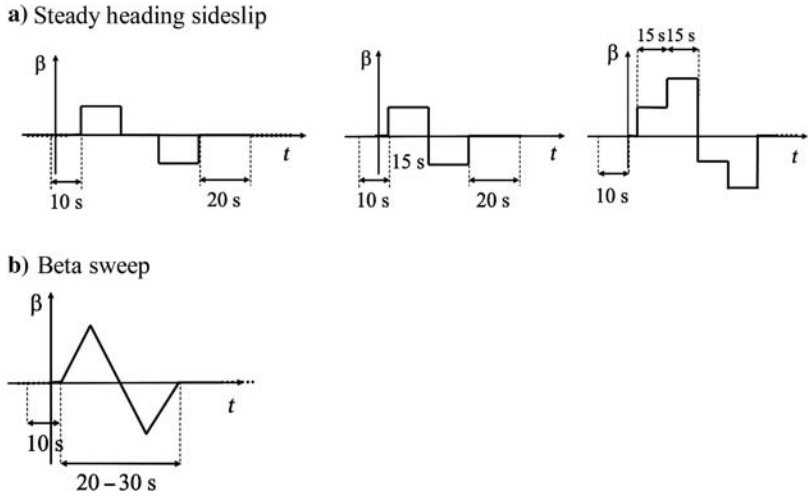

Fig. 2.1 Schematic of steady heading sideslip and sweep input maneuvers.

steady heading sideslip as well as wings-level steady sideslip maneuvers, the variations in speed are compensated through the longitudinal control.

Precaution and constant monitoring is necessary during large constant sideslipping conditions due to large aerodynamic loads acting on the vertical tail and to make sure that the structural limits are not exceeded. The piloting techniques to perform SHSS are demanding because considerable power is necessary for rudder deflections required to achieve asymmetric trim conditions. New safety recommendations have been defined to avoid structural damages resulting from large rudder deflections for transport-category airplanes [23].

Depending upon the aircraft, if there are additional controls available, for example direct-lift-control flaps or all-moving trimmable horizontal tail, then maneuvers with appropriate inputs (either multistep or longer duration pulses) will be necessary. In the case of unstable aircraft, the flight control laws prevent dynamic excitation of the aircraft motion and result in highly correlated control surface deflections and aircraft motion. Special techniques are necessary to generate flight data suitable for parameter estimation; we defer this specific case to Chapter 9, dealing exclusively with unstable aircraft identification.

The frequency sweep test techniques, although rarely used for fixed-wing aircraft, are used more routinely in the field of rotorcraft identification [24]. These techniques are useful and necessary not only for the next generation specification requirements [25] but also are an integral part of the interdisciplinary modeling aspects. Critical flight incidences have, however, occurred while sweep testing due to, for example, exceeding the aeroservoelastic range or the flight permissible maximum loads [26]. Proper coordination is, hence, necessary through careful

preparation, buildup, real-time monitoring, and analysis to prevent possible structural damage and to avoid any increase in the risk factor.

The central idea behind system identification maneuvers described here is to excite pertinent modes of aircraft motion independently and sufficiently. In general, while exciting a particular mode, excitation of the other modes is to be minimized. The amplitudes are usually based on heuristic considerations, which have been discussed. Thus, the task now boils down to finding suitable shapes and durations of inputs that determine the frequency content.

2.3 OPTIMAL INPUT DESIGN

There are two approaches that have been applied in the past to design optimal inputs. The first one, based on the estimation error criterion, is more rigorous, involved, and tends to be theoretical. The second one is an engineering approach, based on the spectral behavior of the model.

2.3.1 INPUT DESIGN BY ESTIMATION ERROR ANALYSIS

As will be described in Chapter 4, statistical properties of the estimates are most conveniently expressed in terms of the bias and covariances of the estimates. It will be shown in the same chapter that the maximum likelihood estimation is based on the maximization $p(z|\Theta)$, the conditional probability density function of measurements z for a given parameter vector Θ, and that the Fisher information matrix, which is an indicator of the information content in the data being analyzed, is given by

$$\mathcal{F}_{ij} = E\left\{-\frac{\partial^2 L(z|\Theta)}{\partial\Theta_i \partial\Theta_j}\right\} \tag{2.1}$$

where $L(z|\Theta) = \ell n(p(z|\Theta))$ and $E\{\}$ denotes the expected value. It will also be shown that the maximum likelihood estimation is bias free and efficient in a statistical sense. For further development of the input signals, we simply accept these properties at this point. Under such conditions, the inverse of the Fisher information matrix given by Eq. (2.1) provides a good approximation to P, the parameter error covariance matrix [27, 28]. Without going into the exact mathematical derivation which will follow in the subsequent chapters, we also draw on the expression for the information matrix $\mathcal{F}$ in terms of $\partial y/\partial\Theta$, the gradient of the system responses with respect to the parameters.

$$P \approx \mathcal{F}^{-1} \approx \left\{\sum_{k=1}^{N}\left[\frac{\partial y(t_k)}{\partial\Theta}\right]^T R^{-1}\left[\frac{\partial y(t_k)}{\partial\Theta}\right]\right\}^{-1} \tag{2.2}$$

Thus, the parameter error covariance matrix P depends upon the response sensitivities to the model parameters $\partial y/\partial\Theta$, the data points being analyzed N, and a

weighting matrix R, which depends on the measurement noise. The system responses y and its gradients $\partial y/\partial\Theta$ are obtained from the a priori model used for designing the inputs. At this stage the parameters Θ are known and kept fixed at their a priori values. Thus, the information content in $\mathcal{F}$ will be mainly determined by the response gradients that depend on the input excitation. Proper tuning of the input shapes will result in an optimal excitation of the modes of a system defined by the a priori model. In other words, it would lead to maximization of the information matrix. In turn, this is the same as minimizing the error covariance matrix P, which suggests that the parameters that will be subsequently estimated from such inputs will have lowest statistical errors and hence will be more accurate. This is the basic idea behind designing the optimal inputs based on the theoretical approach.

Different measures, all based on the Fisher information matrix, can be chosen as criteria to design optimal inputs [11, 28]. The first criterion is

$$\min_{u(t)}\{\det(P)\} \Longrightarrow u_{\text{opt}}(t) \tag{2.3}$$

which is the same as maximizing the determinant of the Fisher information matrix $\mathcal{F}$. The criterion of Eq. (2.3) is the overall measure in terms of the volume of the estimation error ellipsoid, which is proportional to the determinant of P. Input signals that maximize the determinant of $\mathcal{F}$ (or equivalently minimize the determinant of P) are commonly called in the literature as D-optimal, where "D" stands for "determinant." It can be shown that the optimum is invariant to scaling of the system states. Let us take a look at the physical significance of D-optimality. The columns of the information matrix $\mathcal{F}$ represent contributions of the individual parameters. If the columns are independent, the determinant of the matrix will have a maximum value subject to the given constraints in terms of the postulated model. On the other hand, if parameters are correlated, the determinant will be smaller; in the worst case, linear dependency of two parameters results in a zero determinant, indicating that the two parameters cannot be estimated independently. Thus, D-optimality minimizes the redundancy and leads to better identifiable parameters.

The other criteria are the sum or product of the diagonal elements of the matrix P:

$$\min_{u(t)}\{tr(P)\} \Longrightarrow u_{\text{opt}}(t) \tag{2.4}$$

and

$$\min_{u(t)}\left\{\prod_i P_{ii}\right\} \Longrightarrow u_{\text{opt}}(t) \tag{2.5}$$

The diagonal elements of P represent the variances of the estimates and are indicators of the accuracy of the estimates. As it will be discussed in Chapters 4 and 11,

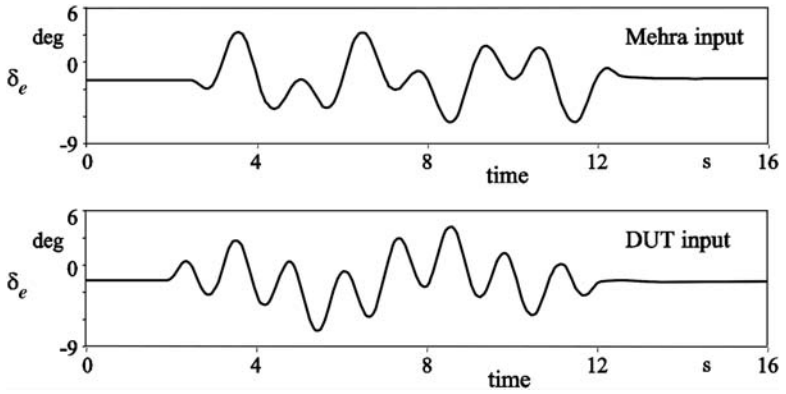

Fig. 2.2 Time histories of typical harmonic inputs.

the standard deviations are given by the square root of the variances. Thus, the last two criteria attempt to minimize the standard deviations or, in other words, increase the accuracy of the estimates. Input signals resulting from Eq. (2.4), that is, by minimizing the sum of the variances of the parameter estimation errors, are commonly called in the literature as A-optimal, where "A" stands for "average." In contrast to the D-optimal design, the A-optimality criterion of Eq. (2.4) or (2.5) is strongly dependent on the units and is not scale invariant. This aspect needs to be carefully considered while designing A-optimal inputs.

Both time-domain and frequency-domain methods have been applied to solve the preceding optimal design problem. To apply any of these criteria, it is necessary to define an input space (in terms of forms and amplitudes) to be searched for the optimal input. Different researchers have followed different approaches. The best known two designs shown in Fig. 2.2 are the following [13]: 1) Mehra input, based on optimization of harmonic signals selecting integer multiples of the basic frequency, where the basic frequency corresponds to the observation time and 2) Delft University of Technology (DUT) input, based on sum of sine functions. Both of these methods and other similar approaches lead to continuous and smoothly varying inputs. These signals are tuned to the a priori model used in their design, but they are not quite suitable to be flown manually. In all of these cases, an onboard computer implementation is necessary to generate such harmonic inputs.

2.3.2 DESIGN OF MULTISTEP INPUT SIGNALS

The design of multistep input signals is a two-step procedure: first, investigating the range of frequencies needed for accurate estimation of parameters, and second, designing suitable multiple step input to cover the desired frequency

range. As in the preceding case, the design procedure requires an a priori model of the aircraft.

Investigation of the range of frequencies for accurate estimation of parameters is based on synthesizing the contributions due to each of the parameters appearing in each of the force and moment equations. This is conveniently done using the Bode diagram, which helps to determine which of the frequencies must be included in the input signal to enable extraction of particular derivatives. The following procedure developed by Marchand [10, 29] is briefly illustrated here considering the following linearized model pertaining to the longitudinal motion:

$$\begin{bmatrix} \dot{u} \\ \dot{\alpha} \\ \dot{q} \\ \dot{\theta} \end{bmatrix} = \begin{bmatrix} X_u & X_\alpha & X_q & -g \\ Z_u/U_0 & Z_\alpha/U_0 & 1 & 0 \\ M_u & M_\alpha & M_q & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} u \\ \alpha \\ q \\ \theta \end{bmatrix} + \begin{bmatrix} X_{\delta e} \\ Z_{\delta e} \\ M_{\delta e} \\ 0 \end{bmatrix} \delta_e \tag{2.6}$$

where U_0 is nominal trim speed, u the component of true airspeed along the x axis, α the angle of attack, q the pitch rate, θ the pitch attitude, δ_e the elevator input, and ($X_{()}$, $Z_{()}$, $M_{()}$) the dimensional derivatives. For the case considered here for demonstration, the a priori values of the dimensional derivatives for a forward speed of 98 m/s lead to system state and control matrices in Eq. (2.6), which can be denoted as A and B, given by

$$A = \begin{bmatrix} -0.0091 & 9.43 & 0 & -9.80665 \\ -0.0022 & -0.867 & 1 & 0 \\ 0 & -3.49 & -2.04 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

$$B = [0 \quad -0.11 \quad -5.09 \quad 0]^T$$

Now, for each equation in Eq. (2.6), the frequency response magnitudes of the various terms in that equation are plotted as a function of the input signal frequency. To demonstrate the procedure, we consider here the pitching moment equation. We compute the frequency response magnitudes for each of the terms, namely, M_u, M_α, M_q, and $M_{\delta e}$ as well as for the total pitch acceleration $\dot{q}$ with respect to the elevator input δ_e. In other words, it amounts to computing the magnitudes:

$$\left|\frac{\tilde{\dot{q}}(\omega)}{\tilde{\delta}_e(\omega)}\right|, \quad \left|\frac{M_u\tilde{u}(\omega)}{\tilde{\delta}_e(\omega)}\right|, \quad \left|\frac{M_\alpha\tilde{\alpha}(\omega)}{\tilde{\delta}_e(\omega)}\right|, \quad \left|\frac{M_q\tilde{q}(\omega)}{\tilde{\delta}_e(\omega)}\right|, \quad \left|\frac{M_{\delta e}\tilde{\delta}_e(\omega)}{\tilde{\delta}_e(\omega)}\right| \tag{2.7}$$

where $\sim$ denotes the Fourier transform. The individual components for $\dot{q}$, α, q, and δ_e can be computed from the output equation $y = C[u \;\; \alpha \;\; q \;\; \theta]^T + D\delta_e$

by defining the observation matrices C and D appropriately as follows: 1) $C_{\dot{q}} = [0\ \ -3.49\ \ -2.04\ 0]$, $D_{\dot{q}} = [-5.09]$; 2) $C_{\alpha} = [0\ \ -3.49\ 0\ 0]$, $D_{\alpha} = [0]$; 3) $C_{q} = [0\ 0\ \ -2.04\ 0]$, $D_{q} = [0]$; and 4) $C_{\delta e} = [0\ 0\ 0\ 0]$, $D_{\delta e} = [-5.09]$, where the subscripts denote matrices corresponding to the respective variables. These components are computed using the program "/FVSysID2/chapter02/FreqCompInput.m" and plotted in Fig. 2.3. The contribution due to the term M_u is not shown because the a priori value of this derivative is zero. The magnitude of the control term corresponding to $M_{\delta e}$ with respect to δ_e is largest compared to the other terms in the higher frequency region. This suggests that the control effectiveness is best identifiable at such frequencies. Identifiability of other parameters depends on the frequency range as will be discussed next. The preceding elaborated procedure evaluates the contributions due to control surface deflections, flow angles, velocity components, and angular rates, but does not include the indirect influence of attitude angles, which are used in the output-error method that we will study in detail subsequently. In Sec. 2.5, it has been demonstrated that attitude angles are also useful in the estimation of aerodynamic models.

The procedure is repeated for the X and Z equations for the longitudinal and vertical motion. From such plots, identifiability of the derivatives can be ascertained. At any given frequency, a large magnitude of any particular term compared to the other contributions suggests a dominant influence of that

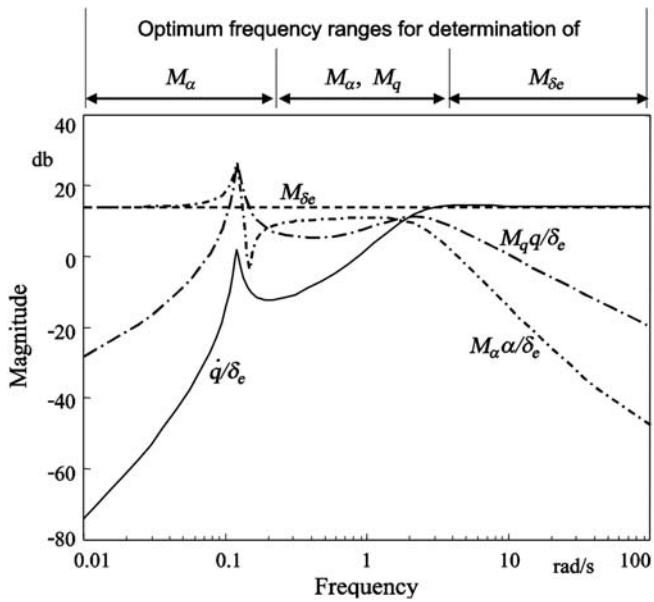

Fig. 2.3 Bode magnitude plot of the pitching moment equation terms.

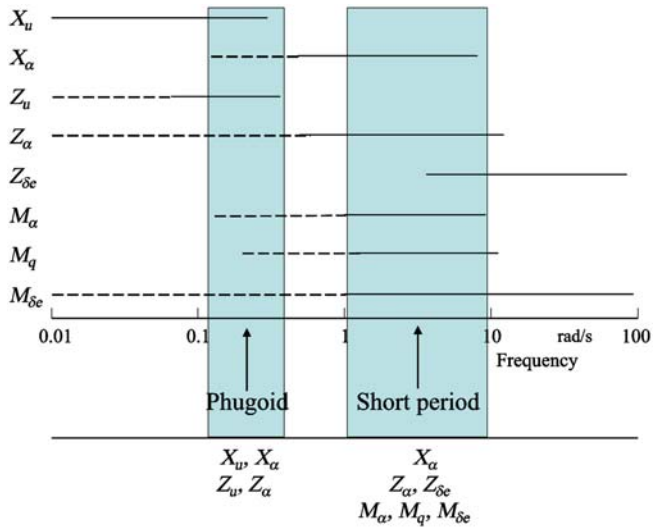

Fig. 2.4 Regions of identifiability.

derivative, which indicates good information content necessary for estimation of the parameter. Conversely, the derivative cannot be accurately estimated if the contribution is small. It is now necessary to specify a measure to decide upon relative contributions. As a rule of thumb, a derivative is considered identifiable when its term has a magnitude of at least 10% of the largest term's magnitude [13].

From frequency response magnitude plots for the X and Z forces and for the pitching moment, the ranges of frequencies are determined in which contributions due to each derivative are appreciable in the sense just described. These ranges are plotted by horizontal lines in Fig. 2.4, which shows that all of the derivatives can be identified from excitation in the two regions marked in gray. The solid lines indicate regions where the derivatives are directly identifiable and dashed line the regions in which they can be estimated only as the ratio of two parameters. The regions surround the natural frequencies of the phugoid and the short-period mode. This is not really surprising because from our basic intuition we know that the system excited at its natural mode will exhibit dominant dynamic motion. Excitation of the long-period phugoid mode is easily obtained through a longer duration pulse input, but the excitation of the fast responding short-period motion is more critical. The preceding analysis combined with additional issues provides a good basis for designing elevator input signals for longitudinal motion, namely: 1) the optimum range for frequencies covers a range below and above the natural frequency of the short-period mode, 2) the eigenfrequencies of the modes based on the a priori model are subjected to uncertainties, and 3) the eigenfrequency changes with the flight

condition. These considerations imply that besides the eigenfrequency it is also necessary to excite the frequencies around the eigenfrequency as well. Based on these considerations, a bandwidth of 1:10 is recommended for the design of a multistep input signal. Thus, the task now is to design an input signal covering this required frequency range.

A multistep input signal of arbitrary shape can be synthesized by a suitable sequential combination of pulse inputs. For an input signal consisting of an arbitrary number of equidistant time steps, in any combination of input levels, as derived in Appendix A, the power (or energy) spectrum is given by [30]

$$\mathrm{E}(\omega) = 2\Delta t^2 \, \frac{1-\cos\Omega}{\Omega^2} \left[\sum_{i=1}^{N} V_i^2 + 2 \sum_{j=1}^{N-1} \cos j\Omega \sum_{i=1}^{N-j} V_i \, V_{i+j} \right] \tag{2.8}$$

where $\Omega = \omega\Delta t$ is the normalized frequency, $T = N\Delta t$ the total duration of the input consisting of N impulses each of duration Δt, and V_i the amplitude for the current impulse ($t_{i-1} < t < t_i$). Equation (2.8) has been coded in the function "/FVSysID2/chapter02/EnergySpectrum.m," which is used in the following to compute the spectra of various input signals. It is also possible to compute the spectra by fast Fourier transform (FFT) using the function "InputSigPSD.m," which, however, requires the signal processing toolbox of Matlab®.

The simplest input to excite the oscillatory motion is a pulse input. After the input over a specified time, the control is released to allow the aircraft to oscillate freely without pilot inputs. Equation (2.8) leads to the energy spectra shown in Fig. 2.5 for such an input corresponding to different durations of Δt, but with the same input amplitude. The spectra in Fig. 2.5 are computed and plotted using the program "/FVSysID2/chapter02/EnergyImpulse.m." From this figure, we observe that a pulse with $\Delta t = 0.4$ s has energy spread over a larger range of frequencies, but the magnitude is very small, which is not enough to excite the system adequately. Increasing the duration leads to larger energy content in the signal, but at the same time decreasing the spread to lower frequencies. For $\Delta t = 1.2$ s, it has large energy, but unfortunately concentrated at low frequencies with rapid decay. The pulse input is asymmetric about the starting trim values, which leads to nonzero energy at zero frequency. The asymmetric excitation can lead to changes in the stationary conditions, in many cases deviating significantly compared to the starting trim conditions from which the maneuver

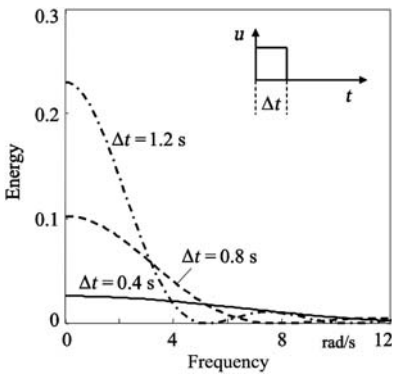

Fig. 2.5 Energy spectra of pulse inputs.

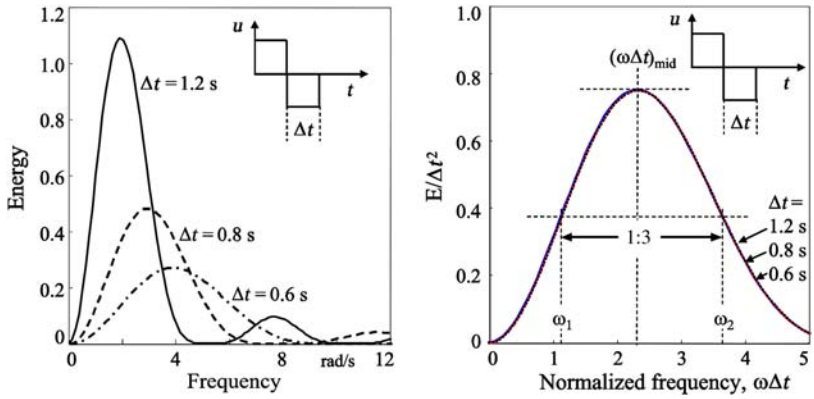

Fig. 2.6 Energy spectra of doublet inputs.

might have been initiated. This in itself is not a serious drawback, provided we can account for such changes in the identified model, which may turn out to be flight condition dependent or nonlinear. If linear models are to be identified, it may lead to violation of the range of validity. The major limitation of this input is that it is less suitable for exciting the rapid eigenmodes at higher frequencies, like short period or Dutch roll.

A more commonly used input to excite the dynamic motion is a doublet: control moved abruptly in one direction, held fixed for certain time Δt, then moved abruptly to other direction, held fixed for same specified time before releasing it to neutral. The doublet input, which is a two-sided pulse, results in a symmetrical signal, having energy concentrated at a frequency that varies with Δt. Figure 2.6 shows the energy spectra for three different time steps $\Delta t = 0.6$, 0.8, and 1.2 s, which are computed and plotted using the program "/FVSysID2/chapter02/EnergyDoublet.m." The drop in the energy content around the peak is relatively fast. As observed on the right side of Fig. 2.6, the normalized frequency $\omega\Delta t$ for the peak is roughly 2.3, that is,

$$\omega_n \Delta t_{\mathrm{DBLT}} \approx 2.3 \tag{2.9}$$

where ω_n is the short period frequency. Equation (2.9) can be expressed in terms of the period of oscillation, yielding an estimate of the time step for the doublet given by

$$\Delta t_{\mathrm{DBLT}} \approx \frac{2.3}{\omega_n} \approx \frac{2\pi}{2.7\ \omega_n} \approx \frac{1}{2.7} \cdot \text{ period of oscillation} \tag{2.10}$$

Although Eq. (2.10) shows a factor of 1/2.7, in practice, a simpler factor of 1/2 is often used, which corresponds to choosing the total length of the doublet ($=2\Delta t$) equal to the period of the eigenmotion. Through this very simple rule of

thumb, the corresponding mode based on the a priori model can be easily and adequately excited. It is observed from Fig. 2.6 that the doublet input has a bandwidth of roughly 1:3, which is the ratio of the normalized frequencies ω_2 ($\approx$3.63) to ω_1 ($\approx$1.1, cf. Fig. 2.6). In this range the corresponding spectrum is at least half of its maximum value. Because of its simplicity, the doublet input is widely used for estimation of stability and control derivatives. A combination of doublet and pulse inputs is also quite useful in some cases; a somewhat smaller step size than given by Eq. (2.10) leads to a slightly higher central frequency of the doublet, which is followed by a pulse to cover the lower frequency range.

Extending the logic that progressing from single-step input to two-step input (doublet) leads to a spread of the power spectrum, a much broader band signal can be achieved through a multistep input. Without going into exact details, a multistep input can be designed by minimizing the deviation of the computed spectrum for a design from the desired spectrum suitably defined from the results of frequency response magnitude plots (Fig. 2.3). In such cases it is desirable to minimize the variations in the energy contents in the range of frequencies covered, with drop on the two sides being smooth and rapid, and ideally zero or very small energy at zero frequency. The parameters to be adjusted during the minimization are the number of time steps, N in Eq. (2.8), and amplitudes at each step. Design of such a multistep input was performed by Koehler, resulting in the widely known 3-2-1-1 input [12]. The 3-2-1-1 input is $7\Delta t$ long and consists of alternating positive and negative, equal amplitude, steps of relative duration 3, 2, 1, and 1. The energy spectrum of such a 3-2-1-1 input, computed from Eq. (2.8) using the program "/FVSysID2/chapter02/EnergyImpDbtS3211M3211.m," is shown in Fig. 2.7.

The time step Δt for the 3-2-1-1 input is usually chosen such that the natural frequency of the mode being excited lies in the center or in the upper third of the

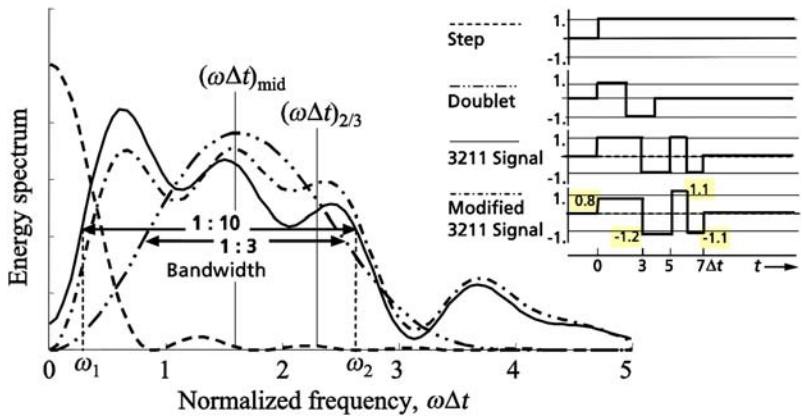

Fig. 2.7 Frequency-domain comparison of standard inputs.

input spectrum. As seen from Fig. 2.7, the normalized frequencies $\omega\Delta t$ for these two conditions are roughly $(\omega\Delta t)_{\text{mid}} = 1.6$ and $(\omega\Delta t)_{2/3} = 2.1$, respectively. This leads to an estimate for Δt_{3211} given by

$$\Delta t_{3211} \approx \frac{1.6}{\omega_n} \approx \frac{\pi}{2\,\omega_n} \approx \frac{1}{4} \cdot \text{ period of oscillation} \tag{2.11}$$

or

$$\Delta t_{3211} \approx \frac{2.1}{\omega_n} \approx \frac{1}{3} \cdot \text{ period of oscillation} \tag{2.12}$$

Thus, the time step Δt_{3211} is somewhat smaller than Δt_{DBLT}. As a rule of thumb, a time step Δt_{3211} given by $\Delta t \approx 0.3/f_C$ is found to be quite effective in several aircraft cases, where f_C in hertz is the frequency of the mode to be excited. This matches well with the range of 0.25 to 0.35 based on an empirical rule and numerical studies [31].

The spectra for various signals are computed with suitable time steps using the program "/FVSysID2/chapter02/EnergyImpDbtS3211M3211.m"; for comparison purposes, they are plotted in Fig. 2.7 as a function of normalized frequency of the 3-2-1-1 signal. Assuming a time step of $\Delta t_{3211} = 1$ s for the 3-2-1-1 signal, the equivalent time step for the doublet is computed from Eqs. (2.11) and (2.9), which turns out to be roughly 1.44 s ($\Delta t_{\text{DBLT}} \approx 2.3/\omega_n$; $\omega_n \approx 1.6/\Delta t_{3211}$). It is observed from Fig. 2.7 that the 3-2-1-1 input has a bandwidth of roughly 1:10, ($=\omega_2/\omega_1 \approx 2.7/0.3$) having at least half of its maximum energy. Thus, the 3-2-1-1 input has a much wider spectrum compared to that of the impulse or doublet inputs. For reference purposes, another equivalent program based on FFT, "PSDImpDbtS3211M3211.m," is provided to compute the various spectra.

Because the natural frequency of the mode is flight condition dependent, strictly speaking the optimal Δt varies from flight condition to flight condition. However, the time variations and step reversals in the 3-2-1-1 signal result in a broad frequency content, and in spite of some deviations from the optimal value of Δt, the input is adequate to excite a band around the natural frequency. On the other hand, tuning of Δt for a doublet input is a little more critical because it has a smaller spread of frequency content around the central frequency. It is primarily due to this reason that the 3-2-1-1 signal is superior to the doublet input.

The main advantage of the 3-2-1-1 input lies in its simplicity and ability to execute it manually. It can be fairly easily remembered without any external aid. Different procedures have been adopted in practice to apply the 3-2-1-1 input manually, matching the desired Δt or for repeatability. For example, just counting "twenty-one, twenty-two, twenty-three" to keep the stick pushed, then "twenty-one, twenty-two" to keep it pulled, then "twenty-one" to pushed, and finally once again "twenty-one" to pulled position, and each time changing the stick position abruptly has produced remarkably good quality 3-2-1-1 inputs.

Another procedure is to provide audio-cuing, consisting of series of beeps. Optical cuing has also been used in a few cases, by providing on the display the multistep input form which the pilot tries to track.

Two minor aspects of 3-2-1-1 inputs are as follows:

1. It is asymmetric about the trim deflection (with four time steps in one direction and three time steps in the other). As a consequence, it has nonzero energy at zero frequency.
2. With the first step being of larger duration, namely, three units of Δt, it may lead to motions far from the initial trim condition, before the following steps are applied.

These undesirable effects can be minimized by modifying the input amplitudes of the 3-2-1-1 input; such a modified 3-2-1-1 input is shown in Fig. 2.7. It has much reduced energy at zero frequency, and the variations in the spectrum are much less. Precise manual application of the modified 3-2-1-1 is, however, difficult and calls for onboard computer implementation. Another way is to apply 3-2-1-1 twice, and once starting by pulling up the aircraft and second time by pushing the stick.

Another variation of the 3-2-1-1 input is the 1-1-2-3 input obtained by time twisting the steps. The 1-1-2-3 input first excites the higher frequencies at the trimmed condition and prevents the vehicle from going far away from the trim condition before the larger duration time steps are applied. Like the 3-2-1-1, the 1-1-2-3 input can be easily applied manually with little practice. The 1-1-2-3 has the same energy spectrum as the 3-2-1-1 input. In some cases a 1-2-1 input, which is a combination of a normal doublet followed by an inverted doublet, is used. By using a time step for the 1-2-1 signal that is half of that of the doublet, it becomes possible to shift the spectrum to higher frequencies. This can be easily verified using the sample program "/FVSysID2/chapter02/EnergyDblt121.m." Much more complex multistep maneuvers have also been designed in some cases, which of course cannot be flown manually [15, 17]. In general, for more complex multistep inputs, the time step has to be chosen carefully; otherwise, there will be pronounced variations with valleys in the input spectrum.

Following a similar procedure based on analyzing the frequency response magnitude plots and approximations given in Eqs. (2.10) or (2.11), we can also design a 3-2-1-1 input for Dutch roll excitation. However, the Dutch roll is lightly damped compared to the short period, and hence a doublet input is found adequate to excite the Dutch roll adequately. Furthermore, although based on the a priori knowledge the time step of a doublet can be optimized, mostly the duration as well as the exact shape is of secondary importance. This is evident from the typical Dutch roll responses shown in Fig. 2.8 for clean configuration at four different speeds. The durations of the oscillations as measured from the recorded data and the actual time steps of the inputs are also provided in

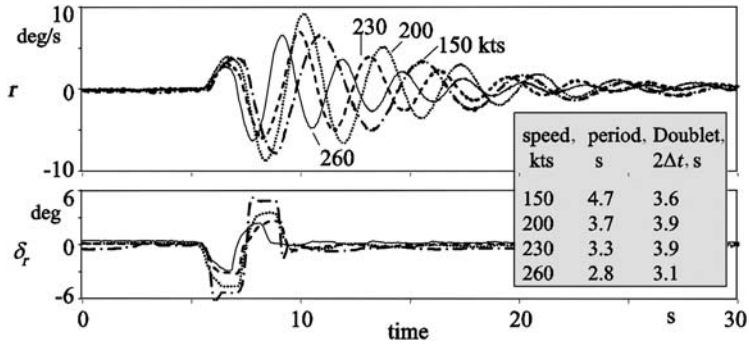

Fig. 2.8 Typical Dutch roll responses.

the same figure. Based on the very good results obtained for several aircraft cases, we recommend a doublet input on the rudder to excite the Dutch roll motion and 3-2-1-1 or 1-1-2-3 on the elevator to excite the short-period motion.

A detailed investigation based on flight data [32], comparing five different types of inputs, has shown that the simpler 3-2-1-1 or doublet inputs yield comparable and in many cases even better estimates compared to the optimal harmonic inputs such as Mehra or DUT elaborated in Sec. 2.3.1. This is mainly because the multistep inputs are richer in frequency contents than the harmonic inputs. For the multistep inputs, if applied automatically using an onboard computer, it becomes possible to change the input levels within one sampling time. Such abrupt changes can lead to rate saturation in the electrohydraulic actuators or can result in much higher load factors at off center-of-gravity locations, for example, in the cockpit, than at the center of gravity. They can also lead to excitation of much higher frequencies and of aeroelastic modes. The optimal harmonic inputs are limited in frequency range and tuned to the desired rigid-body frequencies. This is particularly of importance when the rigid-body and structural frequencies are close to each other, as tends to be the case for very large aircraft. To avoid such undesirable excitations from multistep inputs, the slopes of the input changes should be limited. On the other hand, if they are applied manually, the inputs do not have extreme sharp edges because the pilot acts as a filter. This does not affect the estimation of the stability and control derivatives because, as already mentioned, the exact shape and time step are not critical.

Thus, based on our discussion in Secs. 2.2.2 and 2.3.2, a variety of maneuvers is usually necessary to excite dynamic motions about different axes. We recommend independent control inputs applied one at a time and prefer those that can be flown manually by the pilot. The excitation level should be adjusted to the flight condition being tested; at lower dynamic pressures larger amplitudes are necessary, whereas at higher dynamic pressures lower amplitudes can be applied. We have performed a large number of flight tests, first with a small

amplitude input followed by a larger amplitude input, but still within the range of the linear model. The large amplitude maneuvers, in which nonlinear aerodynamic effects are encountered, are performed separately. Figure 2.9 summarizes schematically the most important system identification maneuvers.

The various multistep inputs just elaborated belong to the class of binary signals, which consist of a series of optimally spaced pulses, yielding desired frequency spectrum. It is also possible to realize an arbitrary power spectral density spread across a specified bandwidth through superposition of phased harmonic signals [33–36]. Such inputs belong to the class of multisine signals. One such input that is used in some applications is called the Schroeder-phased multisine signal. By introducing orthogonalization, simultaneous multiaxis optimal excitation signals can be generated. Such input sequences are preferred in modal estimation of structural systems and for estimation of aerodynamic derivatives of hypersonic flight vehicles because the total test time is usually very short and the flight condition (Mach number) changes rapidly. Furthermore, such input signals have a lot of energy, keeping the input amplitude levels low. In the case of helicopters too, they provide a broadband excitation with improved low-frequency contents [37]. As already mentioned in Sec. 2.3.1, an onboard computer implementation is necessary to apply such multisine inputs. Other signals in the class of frequency sweeps, such as chirp or Fresnel chirp input signal, are also possible.

Although the control inputs such as impulse, doublet, and 3211 are sometimes used for ASE applications and for flight flutter investigation, it is more common to adopt the frequency sweep testing in this case [38]. In general, the frequency sweeps might not be time optimal, but they have a broader bandwidth. The

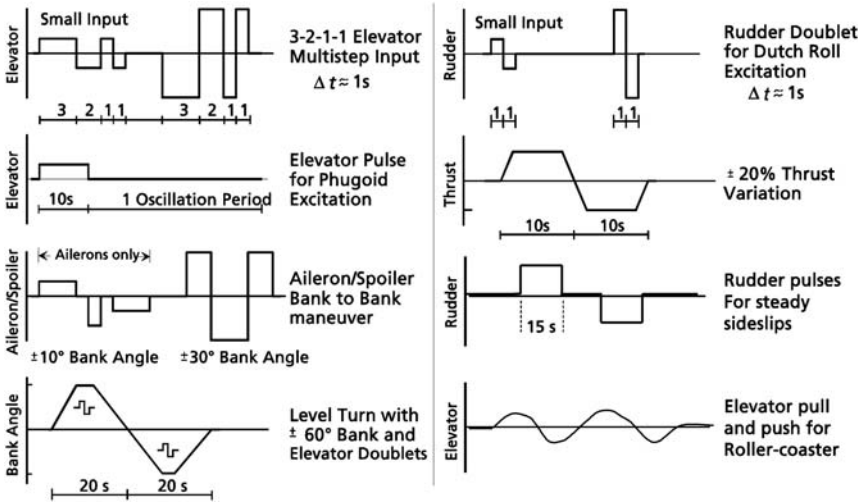

Fig. 2.9 Representative manual control inputs.

frequency sweep testing has been applied, for example, to X-29, EAP aircraft [39, 40] and to identify several structural modes of an XV-15 tilt-rotor aircraft [41].

2.4 SCOPE OF FLIGHT TESTING

Flight testing is costly; hence, it is necessary to limit the number of flight tests to a minimum and also to optimize the test procedure and maneuver sequence within each flight. As pointed out in Sec. 1.4, the goals of system identification could be different, each of which poses different requirements on the range of validity and fidelity of the model to be identified. The scope of the flight testing for system identification depends mainly on these requirements, that is, on the specific goal for which the model is intended. To some extent, it also depends upon the type of flight vehicle being investigated. The larger the number of configurations to be modeled, the larger is the number of tests to be performed. Thus, the following important aspects need to be considered while organizing a flight-test program:

1. Define the specific goal of flight testing.
2. Select the appropriate configurations to be flown.
3. Choose the trim conditions to be tested.
4. Define maneuvers to be performed and the input signals to be applied.

Extensive flight testing is required for generating aerodynamic databases for flight simulators meeting the FAA Level-D fidelity requirements. Estimation of a so-called global model valid over the complete operational envelope needs a comprehensive flight-test program covering all possible configurations. Because the flight vehicle response characteristics depend on flight conditions (angle of attack, sideslip angle, Mach number), dynamic maneuvers will have to be performed at various trim conditions. Furthermore, depending upon the aircraft, special configurations and effects will have to be investigated, which may call for additional tests. As already mentioned in Sec. 2.2, the data for the simulator certification might have to be gathered separately.

As summarized in Fig. 2.9, a variety of maneuvers is usually necessary to allow collection of data necessary to excite significantly all modes that we will be analyzing. The maneuvers are usually performed in blocks, repeated at different trim conditions defined by Mach number, angle of attack, altitude, configuration, which will cover flaps, thrust levels (particularly for propeller aircraft), landing gear, center of gravity, etc. The general procedure consists of the following:

1. Fly to the desired level and speed.
2. Select the configuration and trim the aircraft at a reference flight condition.
3. Apply the sequence of inputs; choose the amplitudes such that operational limits are not exceeded.
4. Repeat the sequence multiple times at the same trim condition.

Between maneuvers, the aircraft is trimmed to horizontal level flight. The same input maneuver is repeated a number of times to enable comparison of estimates, to minimize the influence of external disturbances, and to determine the amount of variation in the measurements. If possible, the same maneuver should be performed three or more times. If the maneuver is performed twice, we face the common dilemma of experimental investigations; namely, if the estimates from the two repeat runs deviate significantly, it becomes difficult to choose or discard one or the other. External disturbances are to be minimized, which means that maneuvers should be performed in calm air. Atmospheric disturbances due to thermal activities are less early in the morning. The reference point for the estimates obtained from the dynamic maneuvers is usually the starting trim condition. Another possibility is to treat reference point as the average of the initial trim and flight condition to which the flight vehicle returns after the dynamic maneuver; ideally, both should the same. In a few cases, average over the complete maneuver has also been considered.

The reference flight conditions are defined in terms of altitude and calibrated airspeed. Figure 2.10 shows the flight envelope for the Transall C-160 military transport aircraft. As an example, various test flight conditions are shown spread over four altitude levels and several speeds at each level. The equidistant stapling of trim points over altitude and speed, shown in the figure, might not be optimal and needs to be adapted to each case appropriately. For example, fewer altitude levels would suffice or more number of points at lower speeds might provide more data at operationally critical flight conditions. Likewise, trim points at several thrust levels should be part of the testing although not

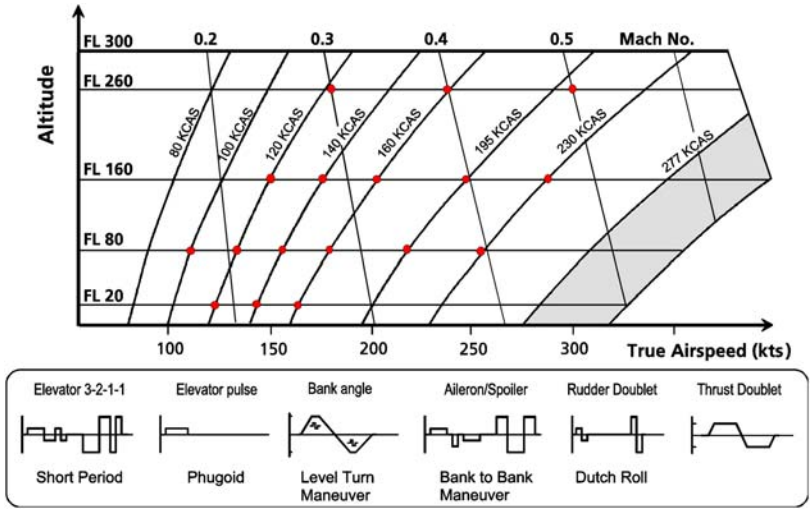

Fig. 2.10 Typical flight-test program for system identification.

explicitly shown in the figure. Different altitude levels allow separation of the Mach number and dynamic pressure effects in the model because the same trim speed at two different altitudes results in two different Mach numbers. At each point the sequence of maneuvers to be performed is shown at the bottom of the same figure. Usually, it is efficient to perform all of the maneuvers at different trim conditions at one flight level, including the repeat runs and then change the flight level to follow the same procedure. During the actual flight testing, the flight levels might have to be adapted to the prevailing weather conditions.

In some cases, at high Mach numbers it might not be possible to fly level; we may have to then perform the tests in descending flight. Special flight conditions and those pertaining to flight at low altitudes above ground, for example, flybys, takeoffs, and landings, are not marked in Fig. 2.10. Similarly, tests at all of the altitude levels and speeds will not be possible for each configuration. The test points will have to be chosen accordingly, for example, dynamic tests with landing gear extended will be possible only up to certain altitudes and speeds. It is also not meaningful to perform such tests with landing gear extended at high altitudes, even if possible, because such test conditions are not realistic.

Development of a global model valid over the complete operational envelope requires flight testing for different landing flaps. The common approach is to perform dynamic maneuvers at each landing flap and identify models for each configuration separately. Interpolation over the flaps is incorporated in the final database to cater to the flap change dynamics, which is one of the tests to be covered in validating the models for flight simulators. If the models identified for the individual flap settings are accurate enough, usually the preceding procedure of interpolation is adequate to meet the fidelity requirements. If not, then some model adjustments will be necessarily late in the database development, which is a drawback of this approach. On the other hand, the preceding approach yields more accurate models for each flap configuration, covering significantly larger flight durations than the transition from one flap setting to another. The other approach of identifying the global model from simultaneous evaluation of flight maneuvers at different flap settings is not considered efficient although it might theoretically be possible, given sufficient computing power and computer memory for data processing. Irrespective of the approach we adopt, it is important to note that the input sequence does not include specific dynamic inputs to the landing flaps because they are rate limited. Instead, various flaps are chosen as reference configuration, performing the dynamic tests at each of them.

In a few specific cases, we deviate from this basic philosophy of performing various flight maneuvers at constant flap setting as a chosen reference configuration. If the aircraft is equipped with fast responding direct-lift flaps, which are modified parts of the landing flaps, then we apply the dynamic inputs to the flaps as well. The direct lift flaps provide an independent vertical control necessary in the aircraft that serve as in-flight simulators. Similarly, in the case of glider the flaps are purely mechanically actuated, limited only by ergonomics; the flaps can be deflected dynamically with multistep inputs to provide a direct generation of

lift on the wing, which is expected to be more efficient to excite the wing symmetric bending modes than the indirect effect obtained through elevator deflection. We turn to these two cases in Chapter 12 to discuss in more details the relevance and advantages of dynamic flap inputs.

In the foregoing discussion we recommended exciting one mode at a time through appropriate input; this results from the practical aspects of performing flight maneuvers. In practice, it is difficult to excite optimally all of the modes simultaneously. Thus, flight testing will generate a large set of flight data with several dynamic maneuvers at a chosen trim point, which are repeated the same at different points over the flight envelope. However, estimation of the global model will require combining multiple maneuvers to estimate a set of aerodynamic parameters common to different modes and account for nonlinearities and dependencies on flight conditions. Even when we restrict ourselves to single trim point, it will be necessary to analyze data from several maneuvers. This requirement calls for special capabilities in processing flight data. We turn our attention to this aspect in Sec. 3.3.

2.5 FLIGHT-TEST INSTRUMENTATION AND MEASUREMENTS

Accuracy of parameter estimates is directly dependent on the quality of flight measured data, and hence high accuracy measurements of control inputs and of motion variables are a prerequisite for successful application of the modern methods of flight vehicle system identification. Classical information on flight-test instrumentation for aircraft parameter estimation is provided by Wolowicz [42] and also found in other references, for example, [43]. During the 1970s, the various aspects of flight-test instrumentation have been investigated in detail [44]. Dedicated flight-test instrumentation systems have been developed in several applications of flight vehicle system identification, mainly because 1) several signals from the basic aircraft systems can be tapped, that is, no additional sensors are necessary (for example, engine parameters, static pressure, rate gyros, etc.); and 2) control surfaces to be instrumented are flight vehicle dependent. Although more recently flight instrumentation systems based on commercially available sensors and standard signal processors have been developed to simplify this task, flight instrumentation is still a laborious and time-consuming job. Instead of providing a detailed description of a particular setup, we take a look at the basic requirements in terms of the necessary signals and their importance in the estimation of parameters.

A typical set of measurements required for aerodynamic model extraction from flight data consists of the following:

1. Control surface deflections
2. Linear accelerations
3. Angular rates

4. Attitude angles
5. Air data
6. Static pressure
7. Engine parameters
8. Pilot forces and inceptor positions

If the scope of modeling is extended to validation of CFD simulations, additional measurements of flowfield velocity, other surface variables such as pressures, and temperatures, and of loads will become necessary [45].

Control surface deflections are inputs to the aerodynamic model being identified; inaccuracies in their measurements directly affect estimation results. The deflections are usually measured using potentiometers or LVDT (linear variable differential transducer), mounted directly at the respective hinges. Besides the primary controls (elevator, ailerons, and rudder), other control surfaces like flaps, speed brakes, spoilers, etc. need to be measured.

Accelerations and angular rates along and about the three axes provide very good information about aerodynamic effects. The inertial navigation system (INS) or sometimes called IMU (inertial measuring unit) is an integral unit providing these measurements and those of attitude angles. Some INS incorporate low-pass filters, mainly because such systems are used for long-term navigation purposes whereas short-term dynamic response is not of particular importance. For aerodynamic modeling purposes, however, dynamic response is of primary importance; hence, it needs to be ascertained that raw data from the INS are available or at least the cutoff frequency of the filter included is high enough. A dedicated tri-axis accelerometer is usually recommended. Accelerations are also measured at several other locations such as cockpit or the wing tips, horizontal tail, rudder tip, and engines, which are useful for other investigations such as flutter. For rigid-body aerodynamic model identification, a single of set of acceleration is sufficient.

The measurements of attitude angles are mostly provided by INS, usually through internal integration of measured angular rates; in rare cases they are obtained using magnetometers. These measurements are primarily useful in the data compatibility check, which we will discuss in Chapter 10 to ascertain quality of the recorded data and to determine biases in the measurements of the angular rates. Usually for aerodynamic model identification, the attitude angles are considered to be unnecessary or of secondary importance because aerodynamic forces and moments do not depend on the aircraft attitude relative to Earth axes. They are not required for the least-squares method discussed in Chapter 6, which also includes preprocessing of data to compute aerodynamic coefficients from the measured translational accelerations and angular rates; see Sec. 6.9. On the other hand, the attitude angles are useful in the output-error method discussed in Chapter 4, as they directly aid the determination of aerodynamic zero terms and of the other derivatives indirectly because they contain

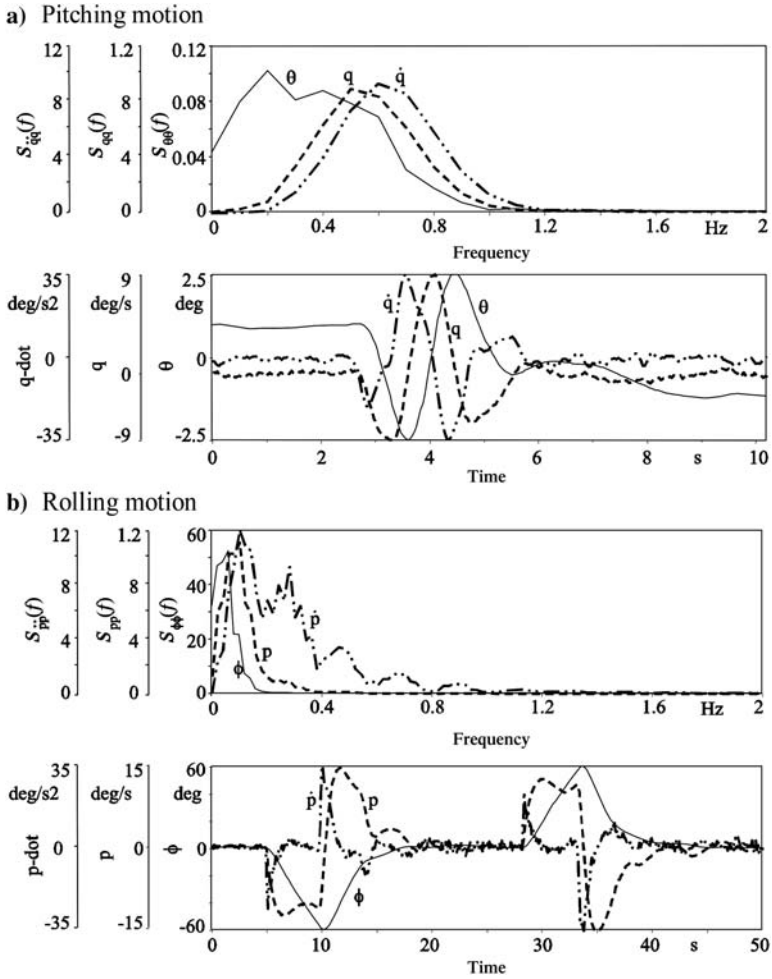

Fig. 2.11 Time histories and power spectra.

information at lower frequencies; see Fig. 2.11 showing typical plots of time histories and power spectra for attitude angle, angular rate, and angular acceleration related to pitching and rolling motion.

The angular accelerations are generally not measured directly. In many exercises related to system identification, they are derived by numerical differentiation of the measured angular rates. It is our experience that inclusion of angular accelerations, besides the angular rates and attitude angles, improves accuracy of parameter estimates and overall convergence of the optimization methods significantly. The contributions due to control surface deflections appear directly in the

accelerations, and hence they come out much better from the acceleration signals. As seen from Fig. 2.11, these signals contain frequencies higher than those from the corresponding attitude angles and angular rates. Their inclusion leads to a model based on a wider range of frequencies. Based on these considerations, although angular accelerations may not be measured using dedicated sensors and aerodynamic forces and moments do not depend on the attitude angles, all three variables (angular accelerations, angular rates, and attitude angles) are considered useful in estimating aerodynamic derivatives.

Air data measurements, consisting of angle of attack, angle of sideslip, and airspeed, are also very useful in the parameter estimation. The flow angles are commonly measured using mechanical vanes and the airspeed by stagnation of the flow using a pitot tube or by propeller. Swiveling probes minimize the alignment errors. These variables are also measured quite often using a five-hole probe, measuring the difference and total pressures. In specific cases, a more advanced flush air data sensing system avoiding an intrusive boom is necessary to measure these parameters. If possible, the flow angles should be measured at multiple locations, typically in front of the aircraft or near pilot station and wing tips. A static pressure measurement is necessary to derive the information on the dynamic pressure, which is the difference between the total and static pressure and also for thrust computations. The static pressure varies along the length of the aircraft, and these measurements usually contain position errors. Availability of additional measurements such as altitude above ground measured using a radio altimeter, the ground speed, and the geodetical positions might be necessary in specific cases.

The engine parameters are necessary to correctly account for the thrust, which is necessary to enable estimation of drag. During aerodynamic model identification, the engine-manufacturer supplied data are used to compute the thrust, which is then treated as known input. In other words, we do not estimate the engine-thrust model. The identified drag model will, of course, depend upon the thrust model. Validating and updating engine thrust (performance) model from flight tests is possible, but it is not the subject addressed here [46, 47]. The exact set of required engine parameters depends on the aircraft (jet or propeller engines). For example, measurements of power lever and condition lever positions, propeller rpm, low-/high-pressure rpm, torque, and turbine gas temperature are necessary for propeller engines whereas those of low-pressure engine rpm are necessary for jet engines. The engine-thrust model also needs other variables such as Mach number, airspeed, and static pressure or altitude.

The pilot inputs (stick, wheel, and pedal positions) and pilot forces are necessary only for the modeling of the flight control system, particularly in the case of reversible controls, driven through mechanical rods and linkages. For estimation of rigid-body aerodynamics, the measured control surface deflections are treated as inputs to the model.

Besides the foregoing aspect of measurements required for extracting aerodynamic derivatives, other aspects related to data recording also affect data

analysis. We simply mention them here without going into any details. They are as follows:

1. The sampling rate is the rate at which data are sampled and recorded. The minimum sampling rate is given by the Nyquist frequency, which should be twice the frequency of interest. In general, the data are sampled at a much higher frequency. A sampling frequency of 20–25 Hz is usually sufficient for rigid-body aerodynamic model estimation. For an extended model comprising rigid-body and structural modes, a higher sampling rate will be necessary.
2. The anti-aliasing filter for all measurements should have the same cutoff frequency, so that all signals would have the same time delays introduced by such filters.
3. Recording of raw data is usually preferable.
4. Ideally, all data channels should be recorded at the same sampling rate. Some of the data channels tapped from the basic aircraft systems may be available at lower sampling rates, for example, slowly varying parameters like altitude. But, they are less critical in the estimation of aerodynamic parameters. More critical measurements like translational accelerations, angular rates, and control surface deflections should be sampled at higher and uniform rates.
5. Ideally, all data channels should be time synchronized. This might not be possible due to serial data recording or the onboard system might introduce time delays. It is usually sufficient to know which data channel is recorded fastest; the others can be time synchronized through data compatibility checking.
6. The signal-to-noise ratio of 10:1 is desirable.
7. All sensors should be calibrated in laboratory and in situ as far as possible.
8. Data reduction should be avoided at the time of recording; if necessary, data reduction can be done while reading the data stored with higher sampling rate by skipping every point or the other during the off-line analysis. Such a procedure is a better one because once the information is lost due to data compression in the recording, it can never be recovered.

Finally, on-site preliminary verification of recorded data is recommended to ensure the adequacy of the data gathered for off-line analysis.

2.5.1 DIGITAL FILTER

Because recording of raw data is recommended, in a few cases it might be necessary to suppress high-frequency noise before using the signals. The high-frequency noise components are usually unrelated to the signal of interest. While generating the angular accelerations by numerical differentiation of measured angular rates,

it is preferable to filter the measured signals for this specific purpose only. Digital filters based on a linear combination of input values are widely used for this purpose. They have the advantage that due to symmetry, they do not introduce any lag in the filtered values. In our applications, we have used a low-pass digital filter, developed by Spencer, based on 15 points, the current value, and preceding as well as following seven data points [48–50]. It belongs to a class of weighted moving average filter, implying that the averaging window moves point by point through the time series, whereby the number of data points for averaging remains the same. Weighted averaging leads to a smoother signal. Such a 15-point Spencer filter is given by

$$y_n = \frac{1}{320}\big[-3\,u_{n-7} - 6\,u_{n-6} - 5\,u_{n-5} + 3\,u_{n-4} + 21\,u_{n-3} + 46\,u_{n-2} + 67\,u_{n-1} + 74\,u_n + 67\,u_{n+1} + 46\,u_{n+2} + 21\,u_{n+3} + 3\,u_{n+4} - 5\,u_{n+5} - 6\,u_{n+6} - 3\,u_{n+7}\big] \tag{2.13}$$

where u is the noisy time series, y the filtered signal, n the discrete time index, u_{n-i} the ith point in the past, and u_{n+i} the ith point in the future from the nth discrete point being processed.

The magnitude response of a discrete filter, such as the one given by Eq. (2.13), can be computed as

$$H(\omega) = C(0) + 2\sum_{j=1}^{m} C(j)\cos(j\,\omega) \tag{2.14}$$

where $m = (N-1)/2$, N the number of data points for weighted averaging, that is, the window size, and C the filter coefficients, also called weights. For the Spencer's 15-point moving average filter given by Eq. (2.13), N is 15, and the filter weights $C = [-3\ \ -6\ \ -5\ \ 3\ \ 21\ \ 46\ \ 67\ \ 74\ \ 67\ \ 46\ \ 21\ \ 3\ \ -5\ \ -6\ \ -3]$, with $C(0) = 74$ at the window midpoint. The magnitude response plot for the 15-point Spencer filter is shown in Fig. 2.12, which shows that it has negligible oscillatory behavior at high frequencies and that the separation between low and high frequencies is well defined and sharp. The magnitude responses are plotted using the function "/FVSysID2/chapter02/filter_Sp_Hen_FreqResp.m."

Another commonly found Spencer filter with 21 data points is given by

$$y_n = \frac{1}{350}\big[-u_{n-10} - 3\,u_{n-9} - 5\,u_{n-8} - 5\,u_{n-7} - 2\,u_{n-6} + 6\,u_{n-5} + 18\,u_{n-4} + 33\,u_{n-3} + 47\,u_{n-2} + 57\,u_{n-1} + 60\,u_n + 57\,u_{n+1} + 47\,u_{n+2} + 33\,u_{n+3} + 18\,u_{n+4} + 6\,u_{n+5} - 2\,u_{n+6} - 5\,u_{n+7} - 5\,u_{n+8} - 3\,u_{n+9} - u_{n+10}\big] \tag{2.15}$$

Figure 2.12 shows the magnitude response of this filter as well. It can be shown that the Spencer filter leaves the local quadratic polynomials untouched and,

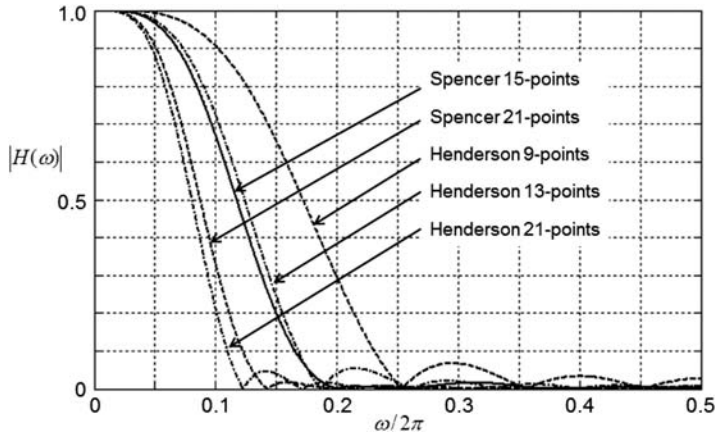

Fig. 2.12 Magnitude responses of Spencer and Henderson filters.

due to symmetry, the cubic polynomials as well. This implies that they can match peaks and troughs well. In all of our applications we have used the 15-point Spencer filter given by Eq. (2.13).

Yet another filter commonly applied in the trend estimation is due to Henderson, mostly based on 7, 9, 13, or 23 points [51–53]. The general formula to compute weights of Henderson filter is given by

$$C_j = \frac{315[(m+1)^2 - j^2][(m+2)^2 - j^2][(m+3)^2 - j^2][3(m+2)^2 - 11j^2 - 16]}{8(m+2)[(m+2)^2 - 1][4(m+2)^2 - 1][4(m+2)^2 - 9][4(m+2)^2 - 25]} \tag{2.16}$$

for $j = -m$ to m, where $m = (N-1)/2$, N being the window size (filter length), that is, number of data points. We consider here only symmetrical filter, that is, N must be an odd number. Henderson's filter is designed to follow local cubic polynomial and can fit different curvatures in the noisy time series, whereby smoothing is obtained through weighted moving averaging process [53]. The preceding formula yields weights for the 7-, 9-, and 13-point Henderson filters as (−0.0587, 0.0587, 0.2937, 0.4126, 0.2937, 0.0587, −0.0587), (−0.0407, −0.0099, 0.1185, 0.2666, 0.3311, 0.2666, 0.1185, −0.0099, −0.0407), and (−0.0193, −0.0279, 0, 0.0655, 0.1474, 0.2143, 0.2401, 0.2143, 0.1474, 0.0655, 0, −0.0279, −0.0193), respectively. They are computed using the function "/FVSysID2/chapter02/HendersonCoeff.m." The magnitude responses of these Henderson filters with different filter lengths are shown in Fig. 2.12. It is observed that 13-point Henderson filter has characteristics similar to 15-point Spencer filter, but with small oscillations around zero level at higher frequencies. The

same is true for the two filters with 21 points. The 9-point Henderson filter has a wider bandpass, but at the same time somewhat larger oscillations at higher frequencies than those with more data points. Thus, there is tradeoff between the required range of bandpass and the oscillations around zero at higher frequencies that can be tolerated, which might have to be investigated in the case of very noisy data.

The preceding equations can be correctly implemented for a truncated time series for all data points except for few data points, namely, $(N-1)/2$ points, from the two ends; for example, for the 15-point Spencer filter for the first and the last seven points from the two ends of the time series. To smooth the two ends of a time series, a filter based on fewer points is applied. Because of the lack of sufficient information, the first two and the last two data points cannot be filtered accurately. This needs to be kept in mind while using preceding procedure. In the case of 15-point Spencer filter given by Eq. (2.13), for the third to seventh data points from the two ends the following filter using two data points can be incorporated:

$$y_n = \frac{1}{96}[7\,u_{n-2} + 24\,u_{n-1} + 34\,u_n + 24\,u_{n+1} + 7\,u_{n+2}] \tag{2.17}$$

Equations (2.13) and (2.17) have been coded for a single time segment in the function "/FVSysID2/chapter02/smooth.m" and for multiple time segments in "smoothMulTS.m," which is used in the following chapters. Similarly the Henderson filter can be coded based on Eq. (2.16), if the need arises.

2.5.2 NUMERICAL DIFFERENTIATION

As already pointed out earlier in this section, angular accelerations are obtained by numerical differentiation of the measured angular rates. Time derivatives of other measurements may also be needed in specific cases. The simplest way to obtain such time derivatives is to use the one-sided forward or two-sided central difference formulas. We will deal in more detail with such formulas in Sec. 3.5.3 in connection with approximating the system functions. Such simple formulas are not quite suitable for numerical differentiation of measured time histories because they tend to amplify the noise. In other words, small errors in measured data can induce large deviations in the numerically approximated derivative. In spite of high quality sensors, the presence of measurement noise is unavoidable. In general, there are two ways to go about doing differentiation of noisy data: 1) filter the raw data first using suitable filter such as a 15-point Spencer or 9-point Henderson filter discussed in the foregoing section and then differentiate the smoothed data using one of the differentiator having near-ideal response characteristics and 2) choose numerical differentiator incorporating a suitable low-pass digital filter that suppresses the noisy high frequencies. We study both the approaches.

Without going into mathematical details, the frequency response of a differentiating smoother given by $i\Omega/\Delta t$, where $\Omega = \omega\Delta t$, having no real part, can be constructed by the following digital filter of arbitrary order given by [54]

$$y(k) = \frac{1}{\Delta t} \sum_{i=1}^{N} C_i[x(k+i) - x(k-i)] \tag{2.18}$$

where k is the discrete time point, x the measured signal, y the time derivative of the measured signal, Δt the sampling time, N the order of the filter incorporated within a differentiator, and C_i the corresponding coefficients. The coefficients C_i are obtained from the equation:

$$A\ C = b \tag{2.19}$$

where the vector b and the elements of n–by–n matrix A are given by

$$b^T = [1/2 \quad 0 \quad 0 \quad 0 \quad \cdots]; \quad a_{ij} = (-1)^{i+1}\ j^{2i-1} \tag{2.20}$$

Equations (2.19) and (2.20) yield the following coefficients for differentiator from first to fourth order, respectively: (0.5), (0.666666667, −0.08333333), (0.75, −0.15, 0.016666667), and (0.8, −0.2, 0.038095238, −0.003571428). These coefficients and those for the higher-order smoother differentiator can be generated using the program "/FVSysID2/chapter02/numDiffCoeff.m." Utility functions "ndiff_Filter02," "ndiff_Filter04," "ndiff_Filter08," and "ndiff_Filter12" for differentiation, catering to multiple time segments, have been developed based on second-, fourth-, eighth-, and twelfth-order formulas. They have the magnitude responses shown in Fig. 2.13, whereby the magnitude response of a differentiator is given by

$$H(\omega) = \frac{i}{\Delta t}\ 2\sum_{j=1}^{N} C(j)\ \sin(j\omega) \tag{2.21}$$

and computed using the function "/FVSysID2/chapter02/nDiff_FreqResp.m." The higher the order (that is, number of data points used to compute the derivative), then the more closer is the differentiator to the ideal response of $H(\omega) = i\omega$. These four differentiators and those discussed hereafter can be easily verified on a simple test case using "ndiffVerify.m." It is fairly obvious that for $N = 1$, Eq. (2.18) corresponds to the standard central difference formula. These formulas with more data points are, in general, termed as N-point central differences. Other ways to arrive at these centrally pivoted algorithms for smoothing and differentiation are possible [55]. We have used eighth- and twelfth-order differentiator formulas in several applications. We will be using the differentiator with eighth-order central differences ("ndiff_Filter08.m") preceded by "smoothMulTS.m" based on the 15-point Spencer filter in some of the test cases analyzed in this book. From Eq. (2.18), we note that these formulas are based on the use of past and future data about a point. The number of data points used to compute the

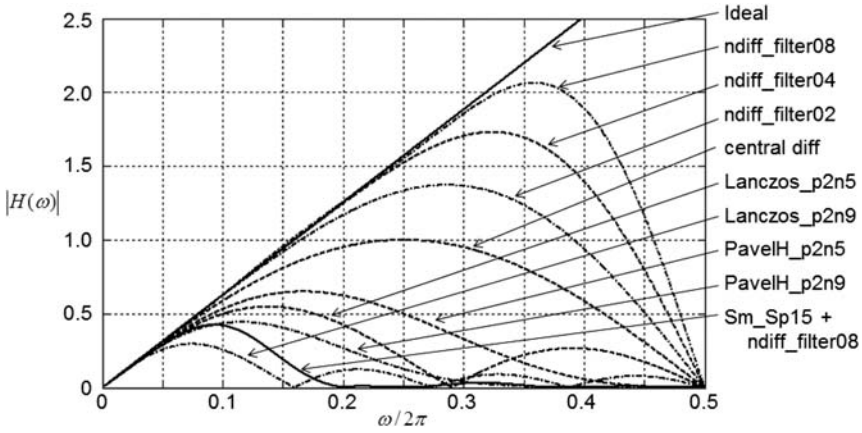

Fig. 2.13 Magnitude responses of various differentiators.

derivative at any point is twice the order plus one for the data point being evaluated. As these differentiators do not suppress sufficiently the high-frequency noise, as already pointed out, they are usually preceded by smoothing the raw data using a suitable filter.

Another approach based on first fitting a polynomial to a few data points by the least-squares method and then analytically differentiating the approximated polynomial is also frequently used to compute the first derivative. Fitting a polynomial automatically leads to smoothing out the noisy data. The method, called Lanczos differentiator, is once again centrally pivoted and needs an equal number of data points on either side of the data point at which the derivative is being computed. It can be shown that for five data points the formula for the first derivative at the midpoint, fitting a parabola by least-squares minimization, is given by [56, 57]

$$y(k) = \frac{2x(k+2) + x(k+1) - x(k-1) - 2x(k-2)}{10\,\Delta t} \tag{2.22}$$

where, as before in Eq. (2.18), k is the index of the data point, y the first derivative of the measured signal x, $x(k+1)$, $x(k+2)$ the two point in the future, and $x(k-1)$, $x(k-2)$ the two point in the past from the kth point.

It is also possible to use a greater number of data points in the Lanczos differentiator. The general formula for the Lanczos differentiator based on a second-order polynomial using N data points is given by [58]

$$y(k) = \frac{3}{\Delta t}\sum_{i=1}^{m} i\frac{x(k+i) - x(k-i)}{m(m+1)(2m+1)}, \quad m = \frac{N-1}{2} \tag{2.23}$$

where N must be an odd number. It can be easily verified that for $N = 5$, Eq. (2.23) leads to the formula in Eq. (2.22).

Note that the coefficients for this Lanczos differentiator with five data points are (0.1, 0.2), which are different from (0.666666667, −0.083333333) for the second-order differentiator with similar structure given by Eqs. (2.18–2.20). For the case with nine data points, the coefficients for the Lanczos differentiator turn out to be (0.01666666666667, 0.03333333333333, 0.050, 0.06666666666667), which are different from (0.8, −0.2, 0.038095238, −0.003571428) for the fourth-order formula given by Eqs. (2.18–2.20). As such, these differentiator filters show different magnitude response and noise suppression characteristics. The Lanczos differentiator based on a second-order polynomial using five data points given by Eq. (2.22) is coded in a function "ndiff_Lanczos_p2n5.m" and for nine data points in "ndiff_Lanczos_p2n9.m." These differentiator filters based on a second-order polynomial approximation are termed as *low-noise Lanczos differentiators* because they are less sensitive to noise due to the low gain at high frequencies. At low frequencies their characteristics approach the ideal response. In other words, they have a good suppression of high frequencies. The limitations are that the linear frequency range is somewhat limited and that they lead to wavy nonzero response at higher frequencies as seen in Fig. 2.13. The Lanczos formulas can also be derived for higher-order polynomials [58, 59]. They are usually termed as *super Lanczos low-noise differentiators*, which have a wider linear range, but the high-frequency suppression is poorer than that of the low-noise Lanczos differentiator just discussed. We do not pursue further the super Lanczos low-noise differentiators, but their formulas can be found in [58–60].

A more general polynomial fitting approach is commonly termed in the field of digital signal processing as Savitzky–Golay filter [61–64]. Such a general-purpose, more flexible version of Savitzky–Golay digital differentiators is available as a Matlab® function "sgolay" found in the Signal Processing Toolbox [65]. The same function can also be readily used to obtain higher-order derivatives or just for smoothing the noisy data, in which case it provides an alternative to Spencer digital filter given by Eq. (2.13). It can be verified that the coefficients given by Eqs. (2.22) for a second-order polynomial with five data points, and those by Eq. (2.23) for nine data points, are exactly the same as those provided by the function sgolay for the same order of the polynomial and the same number of data points; the second column of the differentiation filter in sgolay provides the coefficients for the first derivative. For our specific application involving data analysis with multiple time segments, we provide two functions, namely, "ndiff_SGolay_p2n5" and "ndiff_SGolay_p2n9" based on the function sgolay for the preceding choice of polynomial order and data points. They are numerically equivalent to "ndiff_Lanczos_p2n5.m" and "ndiff_Lanczos_p2n9.m," respectively.

For all of the differentiators, as in the case of digital filter for smoothing discussed in the preceding section, for a truncated time series an adequate procedure needs to be developed for a few data points at the two ends. This aspect has been

suitably taken care of in the various functions for numerical differentiation provided herewith. As the algorithms elaborated here are based on the use of past and future data, they are noncausal and suitable for off-line data analysis only. For real-time application, a different class of filters needs to be developed. We do not go into details of such algorithms. For a simple test of differentiating a data with sine wave without any noise in "ndiffVerify.m," all the differentiators yield, as expected, very similar results. With data gathered from different aircraft, which showed noise characteristics commonly found with typical sensors and data acquisition systems, different noise-suppressing differentiators discussed hitherto provided comparable and acceptable results. In general, smoothing the data first by Spencer filter "smoothMulTS.m" followed by "ndiff_Filter08.m" differentiator discussed in Eqs. (2.18–2.20) showed marginally better performance. This conclusion is also supported by the magnitude response of the combined 15-point Spencer filter followed by ndiff_Filter08 shown in Fig. 2.13 and enlarged in Fig. 2.14, which shows much better suppression of high frequencies and negligible oscillations around zero at higher frequencies. Figure 2.14 is generated using the function "/FVSysID2/chapter02/nDiff_FreqResp2.m." If larger linear range is desired, we will have to go for the nine-point Henderson filter to smooth out the noise first and then apply the ndiff_Filter08 to differentiate the filtered data.

In a few cases, particularly with unmanned aerial vehicle with a very high level of noise with wider frequency spread, the choice of differentiator was more involved. The differences in the performance in such a case are mostly observed at the peaks of the rapidly changing responses in terms of peak heights and width and around the zero level in the form of oscillations. For smoothing out the noise, longer window length is preferable; the longer the length, the better the

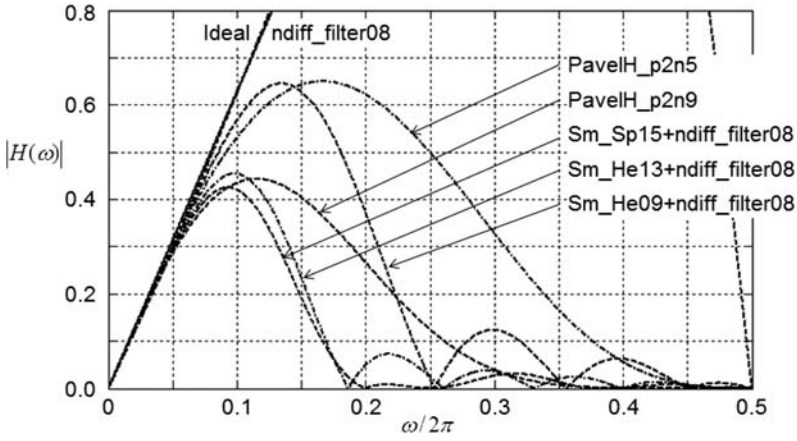

Fig. 2.14 Magnitude responses of smooth noise-robust differentiators.

smoothing. However, for a fast responding dynamic system, longer window length may distort the actual fast system response by smoothing out the desired signal itself. As smoothing and noise-suppressing differentiators may, in general, reduce the information content, it is a tradeoff between noise suppression, mostly determined by the window length, that is, the data points, and acceptable reduction in the information, which needs to be carefully considered. Optimal choice of differentiator depends on the noise contained in the data and may have to be investigated in each case.

The just-discussed differentiator filters might show poor performance in specific cases because the least-squares polynomial approximation does not guarantee complete noise suppression of high frequencies. A smooth noise-robust differentiator developed recently by Holoborodko guarantees suppression of high frequencies and yet is precise on low frequencies and retains exactness on the polynomial [58]. The general formula for such a smooth noise-robust differentiator is given by

$$y(k) \approx \frac{1}{\Delta t} \sum_{i=1}^{M} c_i (x(k+i) - x(k-i))$$
$$c_i = \frac{1}{2^{2m+1}} \left[\binom{2m}{m-k+1} - \binom{2m}{m-k-1} \right], \quad m = \frac{N-3}{2}; \quad M = \frac{N-1}{2} \tag{2.24}$$

where, as before, k is the index of the data point, y the first derivative of the measured signal x, and N the number of data points used in the filter differentiator; it must be an odd number. The terms inside the two brackets within the square bracket are the binomial coefficients.

We note that the structure of this smooth noise-robust differentiator is basically the same as that of the other differentiators, with coefficients as defined in Eq. (2.24). In the present case, they turn out to be (0.25, 0.125) for differentiator with five data points and (0.109375, 0.109375, 0.046875, 0.0078125) for differentiator with nine data points. Functions for differentiators based on the preceding formula using five and nine data are provided as utility functions "ndiff_PavelH_p2n5.m" and "ndiff_PavelH_p2n9.m," respectively, and can be used for multiple time segments analyzed simultaneously. The magnitude responses of these differentiators are shown in Figs. 2.13 and 2.14. They appear to have a wider linear range than the Lanczos differentiators and suppress the high frequencies better without showing wavy characteristics. Similar characteristics were observed for the combined 15-point Spencer filter followed differentiator using 9-point or 17-point central differences.

From Fig. 2.14, it is observed that the smooth noise-robust differentiator given by Eq. (2.24) has no nonzero wavy response at higher frequencies, but the separation between the low and high frequencies is gradual compared to the procedure recommended and followed in this book, namely, filter the data

first using 15-point Spencer filter followed by differentiation using 9-point or 17-point central differences. If a wider linear range is desired, then the same two-step procedure using the 9-point Henderson filter can be followed, but at the cost of a small nonzero wavy response at higher frequencies. The optimal choice in a particular case will depend upon the noise characteristics. As pointed out in [58], however, the smooth noise-robust differentiator given by Eq. (2.24) is suitable in special cases when a certain range of frequencies is corrupted with noise whereas Savitzky–Golay is optimal for Gaussian noise suppression. In connection with flight data analysis, we have not investigated the performance of this smooth noise-robust differentiator filter in detail and hence refrain from comparing its performance with other formulas, mainly because the majority of the test cases showed noise frequencies for which the two-step procedure of "smoothMulTS.m" followed by "ndiff_Filter08.m" was performed adequately.

In spite of great care in choosing sensors, their installation and calibration, and in recording the data, instrument errors such as bias and time delays are unavoidable. In some cases verification of scale factors is necessary, particularly for flow angles, because they can be configuration dependent and realistic conditions might not be duplicated in a laboratory. Such a step dealing with data checking, estimating instrument errors, and generating refined measurements is part of the overall data gathering process. We defer the discussion of this step, called data compatibility checking, to Chapter 10 because it requires some of the estimation algorithms that we will discuss in the following chapters.

2.6 CONCLUDING REMARKS

In this chapter we have discussed an important aspect of data gathering for system identification purposes. In terms of the Quad-M basics introduced in the preceding chapter, we covered in detail the aspects pertaining to the first M for maneuvers and a part of the second M for measurements. In essence, optimal input design boils down to two simple basic principles: 1) inputs must be such as to excite significantly a range of frequencies around the eigenmodes that are to be modeled and 2) independent input on every control is necessary. Two approaches to design optimal inputs have been discussed; the one based on the estimation error criterion is more rigorous, and the other is the more practical one based on a simple engineering process of synthesizing the contributions due to each of the parameters appearing in each of the force and moment equations. Because both designs are based on the a priori model involving parameter uncertainties, it is necessary to provide for a broader band input signal and one that can be easily adapted to changing flight conditions.

It is our experience that from the basic a priori knowledge about the modes of motion, which we mostly possess, adequate inputs can be designed for

conventional aircraft without detailed investigations by following a few simple guidelines:

1. The multistep 3-2-1-1 input is preferable for short-period mode whereas a simple doublet is quite sufficient for Dutch roll; rolling motion is adequately excited by a bank-to-bank maneuver.
2. The exact shape of the doublet or of the multistep inputs and the optimal time step for these inputs are not critical.
3. The duration of time steps is obtained from simple empirical rules suggested here.

To further improve upon the estimation results, it is preferable to apply one control input at a time during dynamic maneuvers and to minimize excitation of the unmodeled modes. We believe that the gains obtained through a more detailed design of input signals are marginal; furthermore, the improvements in the statistical accuracies as predicted by the error covariance matrix are associated with certain limitations with which we will deal in one of the following chapters. Thus, we have followed the simple philosophy in many applications, namely, stick to the basics just elaborated, and use a combination of different system identification maneuvers briefly discussed in this chapter.

In special cases such as hypersonic flight vehicles, however, more advanced experiment design will be required because the duration of the total test time is usually short and the flight condition (Mach number) changes rapidly. In such cases it is desirable to have input signals with a lot of energy, keeping the input amplitude levels low. Multi-axis orthogonal phase-optimized sweeps having a wide frequency band are more suitable in these applications. Similarly, inputs other than the multistep signals might be more suitable for aeroservoelastic applications.

The scope of the flight testing depends on the desired range of validity and fidelity of the model; hence, defining the goal of system identification prior to data gathering is necessary to optimize the scope of flight testing. The larger the number of configurations to be modeled and the range of the desired validity, the larger is the number of tests to be performed. A brief summary of various flight maneuvers has been provided to gain insight into experimental techniques. It is necessary to point out here that the primary goal of flight testing is to certify flight worthiness and safety; however, conforming to the subject of this book we have focused here only a very specific part, namely, that which allows more accurate validation and update of aerodynamic database.

A brief summary of typical measurements required for parameter estimation has also been provided. It has been argued that besides the control surface deflections, translational accelerations, angular rates, and flow angles that are necessary, the angular accelerations and attitudes angles are useful in extracting aerodynamic derivatives, yielding a model covering a broader frequency range. Practical aspects of data smoothing using digital filters and of differentiating noisy signals have been discussed in some depth.