OPTIMAL TEST INPUTS FOR HELICOPTER SYSTEM IDENTIFICATION

A Thesis submitted to the

Faculty of Engineering

of the University of Glasgow

for the degree of

Doctor of Philosophy

by

DOUGLAS JAMES LEITH

October 1989

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

ABSTRACT

The test input applied to a helicopter, or any other system, for the purpose of system identification can have a substantial effect on the parameter estimates obtained. It is therefore important that an appropriate input is chosen. Inputs must take account of the requirements, and restrictions, of the application. For example, in the rotorcraft case studied a linearised model is being identified, and it is therefore essential that the input produces a linear response.

A straightforward method has been developed for the design of multi-step inputs. This method is based in the frequency-domain, and involves tailoring the auto-spectra of the inputs to give long, linear test records, and parameter estimates with reasonably low variances. In flight trials using the Lynx helicopter at RAE (Bedford), the double-doublet input, designed with this method, has been found to be a significant improvement over more traditional inputs.

Using the data from the flight trials of the double-doublet, both equation-error and output-error identification have been carried out. Several discrepancies were found between the theoretical and identified models. More work is required to clarify this. Numerical difficulties were encountered during the output-error identification, and these were attributed to ill-conditioning resulting from the use of an unstable system.

The design of optimal inputs has also been investigated. In particular, constraints have been developed which are suitable for ensuring that the optimal inputs produce linear responses, and are robust. Conventional energy constraints were found to be of little use for these purposes. Algorithms have been developed for the design of optimal inputs with a variety of constraints, and simulation studies have been made to gain an understanding of the effect of these constraints on the form of the inputs.

With the constraints obtained from this work, an optimal input has been designed for use with the Lynx helicopter. This input is robust, and yet is predicted to give significantly improved parameter estimates. Unfortunately, at the time of writing, flight trials of this input could not be performed.

ACKNOWLEDGEMENTS

I sincerely thank my supervisors Professor D.J Murray– Smith (E. & E.E. Dept) and Dr G.D. Padfield (RAE) for their encouragement, help, and guidance during the past three years.

My thanks to Professor J. Lamb for allowing me to use the many facilities of the Electronics and Electrical Engineering Department.

I would like to thank my friend Dr S. Houston for his help during the flight trials, and for many useful discussions.

I would like to acknowledge :

The computing staff A. McKinnon, E. McArthur, J. Sutherland (E. & E.E. Dept) and L. McCormick (CAE Centre) for their assistance throughout this work.

My thanks to M. Manness, R.Henderson, Li Ping and M. Foad for their friendship.

Finally, I would like to acknowledge the support of the Science and Engineering Research Council and the Royal Aerospace Establishment (Bedford) through the provision of an SERC CASE studentship. The extensive help provided by the Flight Management Department at RAE (Bedford) is also gratefully acknowledged.

TABLE OF CONTENTS

CHAPTER ONE

INTRODUCTION

1.1 INTRODUCTION

One of the major shortcomings of modern helicopters is the high pilot workload involved in performing even simple flight manoeuvers. As a result, even although a helicopter may be capable of rapid manoeuvering, the pilot may be unable to use such agility to the full. In the nap–of–the–earth environment, for example, where a helicopter flies extremely low and fast, and has to avoid ground obstacles such as trees, this loss of manoeuverability can cause serious problems.

At present, production helicopters provide the pilot with direct mechanical controls, together with a simple, limited authority stability augmentation system. Research has therefore concentrated on designing improved control systems which will modify the dynamics of the helicopter in order to reduce the pilot workload.

In order to design high performance flight control systems which can meet current and future handling qualities requirements, it is essential to have available good theoretical flight mechanics models of the helicopter. Such models, which are in general non–linear and multi–variable, can then provide the linearised state–space descriptions needed for control system design. In non–linear form, the models can also be used, through simulation studies, to predict the effect of the proposed control systems on the helicopter's performance throughout the flight envelope.

Typically, helicopter models can be divided into three levels of complexity [1]. The simplest, or level one models, are generally suitable for the prediction of handling qualities and low bandwidth control. The most complex, or level three, models are used for detailed analysis of the rotor. Level two models lie between these extremes, and are simpler than the full level three models, yet are of sufficient detail that they can be used for the development and evaluation of high bandwidth controllers.

Currently the most widely used models are level one, with level two models still under research. Since level one models are not of sufficient accuracy for the design of high bandwidth controllers, it is important that suitable level two models are developed. As part of the efforts to obtain improved models, system identification techniques are used to obtain empirical models based on flight measurements. Theoretical models can be compared with such empirical ones, and any differences between the two can be used to gain insight into the shortcomings of the theoretical models. Moreover, the empirical models can also be used directly in control system design, and for handling qualities evaluation.

In the literature, much research has been focussed on developing system identification methods for both rotorcraft and a wide variety of other applications. Two steps are involved in system identification. Firstly, a known test input is applied to the system of interest, and the response is measured. These measured responses are then processed to obtain a suitable mathematical model for the system. In the helicopter case, the structure of this identified model is assumed to be the same as that of the theoretical model. The system identification problem then simplifies to finding the parameters of this model.

However, when attempting system identification of rotorcraft without stability augmentation, serious difficulties are encountered. These are largely due to three factors : the complexity of the system, inherent instabilities, and measurement problems (see section 1.4 below). As a result, rotorcraft identification has been of mixed success to date.

The aim of the current research has been to tackle these difficulties by designing suitable identification test inputs, and so lead to more successful identification. The test input applied to a helicopter for the purposes of system identification can have a substantial effect on the parameter estimates obtained. It is therefore important that an appropriate input is chosen. In the past, considerable work has been directed towards input design for fixed–wing aircraft, with good results. However, relatively little work has been carried out on the design of rotorcraft inputs.

1.2 THE HELICOPTER MODEL

Before the issues involved in system identification and input design can be discussed, it is necessary to consider the theoretical model that is being validated.

A helicopter can be divided into several major sub– systems, each of which must be considered when developing a mathematical model of the helicopter. These are :

1) Main– Rotor

2) Tail– Rotor or other anti– torque device

3) Fuselage

4) Power plant

While providing lift, the magnitude and direction of the force produced by the main– rotor can be modified by the pilot to permit manoeuvering. The thrust generated by the tail– rotor can also be altered to give yaw motion. Conventionally, four controls are available to the pilot for the handling of the main and tail rotors. These controls are as follows:

a) Main– Rotor Collective, η_{0e}
Controls the magnitude of the thrust produced by the main– rotor.

b) Longitudinal Cyclic, η_{1s}
Controls the longitudinal thrust produced by the main– rotor.

c) Lateral Cyclic, η_{1c}
Controls the lateral thrust produced by the main– rotor.

d) Tail– Rotor Collective, η_{0tre}
Controls the magnitude of the the tail– rotor thrust.

These controls are incorporated in the main and tail rotor sub– systems. The theoretical flight mechanics model, HELISTAB [2], developed at the Royal Aerospace Establishment (Bedford) was used. This is a level one model [1], and the various sub– systems are represented as follows :

1) **Main– Rotor**

The HELISTAB model incorporates several main– rotor descriptions. For the present work, only the very simplest of these was used. This consisted of a quasi– static representation, i.e. with the rotor dynamics neglected.

Such a simplified representation is justified on the basis that the rotor dynamics are typically significantly faster than those of the rigid– body fuselage. On the time scales of the rigid– body fuselage motions, the rotor therefore appears to act instantaneously. Several authors e.g. [11] have found, however, that the rotor dynamics cannot be neglected in this way. Nevertheless, the quasi– static representation is sufficiently accurate for many situations.

2) **Tail–Rotor**

A quasi– static model is also used for the tail– rotor. Main– rotor downwash effects are ignored.

3) **Fuselage**

The fuselage is modelled as a point mass with six degrees of freedom. The six degrees of freedom used are the longitudinal, lateral, and vertical translational velocities (u, v, w) and the roll, pitch, and yaw rotational rates (p, q, r). It is possible to obtain the Euler roll, pitch, and yaw angles (φ, θ, ψ) from the rotational rates.

4) **Power plant**

The power plant is considered to be ideal i.e. able to maintain a constant rotor rpm.

The resulting helicopter model is non–linear. There is therefore provision in the HELISTAB software package [3] for linearisation of this model about a given flight condition. The resulting linear model can be described in the following state–space form :

$$dx(t)/dt = A\,x(t) + B\,u(t) \qquad (1.1)$$

where,

$x = (u\ w\ q\ \theta\ v\ p\ \varphi\ r)^T$

$u = (\eta_{1s}\ \eta_{1c}\ \eta_{0e}\ \eta_{0tre})^T$

A,B are the system and control matrices, respectively.

In the present validation work, the matrices A and B are estimated from flight measurements of x(t) and u(t) using system identification techniques. These are then compared with the theoretical A and B matrices. Any significant differences between the theoretical and estimated matrices can then be used to gain insight into the shortcomings of the theoretical model.

However, it should be noted that this linear model is only valid for small perturbations about the flight condition used in the linearisation. It is important that this restriction be taken into consideration when use is made of the model.

1.3 SYSTEM IDENTIFICATION TECHNIQUES

1.3.1 Introduction

A wide variety of system identification techniques are available which enable a model to be estimated from measured data. Each has particular strengths and weaknesses, and each is often more suited to certain applications than to others. The particular identification techniques in most widespread use with rotorcraft can be separated into two distinct types: equation–error methods, and output–error methods (see, for example, [4,5,6]).

Under certain conditions, these methods act as maximum likelihood estimators. The conditions required are different for each method, and relate to the characteristics of the system being identified (see sections 1.3.3 and 1.3.4 below for details). However, the equation–error and output–error methods are also widely used in applications where these requirements are not met. In such cases, the behaviour of the methods will depend on the specific situation prevailing in that application, and will not be maximum likelihood.

1.3.2 Maximum Likelihood Estimators

Firstly, take the general maximum likelihood case [7]. An input vector, u taken from the set U of possible input vectors, is applied to the system concerned. The system response is a random variable :

$$z = z(\underline{\theta}, u, \eta) \qquad (1.2)$$

where,

$\underline{\theta} \in \Theta$ is the true value of the parameter vector,

η is the random component in the system.

An estimator is then any function of z and u with range in Θ. The value of the function is called the estimate, θ'. Thus,

$$\theta' = \theta'(z, u) \qquad (1.3)$$

The maximum likelihood estimate is defined as the value of θ which maximises the likelihood function, $p(z|\theta)$ i.e.

$$\theta' = \arg\max_{\theta} p(z|\theta) \qquad (1.4)$$

where,

$p(z|\theta)$ is the probability distribution function of the input, u producing the response, z when θ is the set of parameters used. Note that this function is also dependant on η, the random component in the system, although this is not explicity shown in the notation used.

This can be interpreted as choosing that value of θ which makes the observed measurements most plausible.

An alternative to maximising the likelihood function is to maximise the log– likelihood function, log $p(z|\theta)$. This produces the same estimates, since the log function is monotonic, but often has the advantage of leading to a simpler optimisation procedure.

Maximum likelihood estimates exhibit several important properties [8]. These include :

1) **Estimates are asymptotically unbiased**

The bias of an estimate is defined as,

$$b = E\{ \theta' | \underline{\theta} \} - \underline{\theta} \qquad (1.5)$$

i.e. bias measures the consistent error between the parameter estimates and the true parameter value.

For maximum likelihood estimates, the bias is always zero when an infinite number of measurements are made. However, for a finite number of measurements, the estimates may still be biased.

2) **Estimates are asymptotically efficient i.e. have minimum covariance**

The covariance of an unbiased estimator is defined as :

$$\text{cov}\ (\theta') \quad = \quad E\{\ (\theta' - \underline{\theta})\ (\theta' - \underline{\theta})^T\ \} \tag{1.6}$$

i.e. the covariance measures the spread of the parameter estimates about the true parameter value. This spread is fundamental, and is caused by η, the random component in the system.

An 'efficient' estimator gives estimates which have the minimum possible covariance i.e. which extract the maximum information from the measurements of z and u. The minimum covariance is given by the Cramer– Rao bound [9], which states that :

$$\text{cov}\ (\theta') \quad \geqslant \quad D \tag{1.7}$$

where,

$$D \quad = \quad M^{-1} \tag{1.8}$$

$$M \quad = \quad E\{\ \nabla_\theta\ (\log\ p(z \mid \theta))^T\ \nabla_\theta\ (\log\ p(z \mid \theta))\ \} \tag{1.9}$$

D is known as the dispersion matrix, and M as the information matrix.

3. **Estimates are invariant**

Invariance is the property that, given θ' is the maximum likelihood estimate of $\underline{\theta}$, then $f(\theta')$ is the maximum likelihood estimate of $f(\underline{\theta})$, when f is a linear function.

1.3.3 The Equation– Error Estimator

It is now possible to consider the particular maximum likelihood cases of equation– error and output– error.

In equation– error [7], the response function $z(\theta, u, \eta)$ is assumed to be linear in θ, and to have the following form,

$$z = f(u)\ \theta + \eta \tag{1.10}$$

where,

η is a zero mean Gaussian process with covariance, R.

f is some function of u, which may be non– linear.

For this case, it can be shown [7] that the log– likelihood function is given by,

$$\log p(z \mid \theta) = -\tfrac{1}{2} (z - f(u)\theta)^T R^{-1} (z - f(u)\theta) - \tfrac{1}{2} \log |2\pi R| \tag{1.11}$$

This is maximised by finding the minimum of the term,

$$(z - f(u)\theta)^T R^{-1} (z - f(u)\theta) \tag{1.12}$$

Notice that this term represents the square error between the measured response, z and the predicted response, $f(u)\theta$. Hence, the equation– error estimator is also widely known as a 'least squares' estimator. This estimator is often used with systems which do not meet the assumptions made above concerning the form of $z(\theta, u, \eta)$. However, in these cases it no longer behaves as a maximum likelihood estimator.

As described in section 1.2, the theoretical helicopter model used in this work is of the form :

$$dx(t)/dt = A\ x(t) + B\ u(t) \tag{1.13}$$

The matrices A and B contain the parameters to be identified. To use the equation–error method with this model, measurements of dx(t)/dt, x(t), and u(t) must be available. Typically, x(t), and u(t) are obtained by direct measurement, and dx(t)/dt either by measurement or by differentiation of x(t).

The theoretical model can be written as :

$$\frac{dx}{dt}(t) = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix} \begin{bmatrix} x(t) \\ u(t) \end{bmatrix} \qquad (1.14)$$

$$\text{i.e.} \qquad z^*(t) = F\, u^*(t) \qquad (1.15)$$

where,

$$z^*(t) = dx(t)/dt,$$

$$u^*(t) = \begin{bmatrix} x(t) \\ u(t) \end{bmatrix},$$

$$F = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}$$

The equation–error method can then be used to estimate the elements of F, i.e. of A and B. However, in helicopter applications the measurements of x(t) and dx(t)/dt contain significant noise components. Hence, both the input, u^* and the output, z^* contain noise. In the ideal equation–error model (equation (1.10)), it is assumed that the outputs contain noise, but that the input does not. When used in the helicopter case, the equation–error method is therefore not acting as a maximum likelihood estimator. In particular, it can be shown that the resulting parameter estimates will be biased [10].

Returning to equation (1.12), this can be minimised analytically [7], and the result may be expressed as follows :

$$\theta' = \left(f(u)^T R^{-1} f(u) \right)^{-1} f(u)\, R^{-1}\, z \qquad (1.16)$$

Given the measurements u, and z, and the values for R and f(u), this expression can be evaluated to obtain the parameter estimates. Hence, the equation–error case leads to relatively straightforward, fast estimation algorithms. Difficulties can occur, however, if there is ill–conditioning in the matrix [8],

$$f(u)^T R^{-1} f(u)$$

since this needs to be inverted to obtain θ'. The most common cause of such ill–conditioning is correlations between the elements of the input vector, u. Various techniques [8] can be used to reduce the effects of ill–conditioning, but in severe cases the identification may fail. The only true solution to this problem lies in designing suitable inputs, u which have low correlations.

1.3.4 The Output–Error Estimator

Now, in the output–error case, the response function, z is assumed to have the following form [7] :

$$\frac{dx}{dt}(t) = A\,x(t) + B\,u(t) \qquad (1.17)$$

$$z(\theta, u, \eta, t) = C\,x(t) + \eta(t)$$

The matrices A, B, and C are functions of θ, and $\eta(t)$ is a zero mean Gaussian process with covariance, R. Time, t is now included in the formulation. In the equation–error case this was not necessary. However, it is required for output–error, since the model now involves the derivative dx(t)/dt.

It can be shown [7] that the log–likelihood function for this case is as follows :

$$\log p(z|\theta) = \frac{1}{2}(z(t)-Cx(t))^T R^{-1}(z(t)-Cx(t)) - \frac{1}{2}\log|R| - \frac{1}{2}\log 2\pi \qquad (1.18)$$

Despite the apparent similarity between the log–likelihood functions of the equation–error and output–error cases, the output–error method is significantly more complicated. This is because the model response, Cx(t) is no longer linear in the parameters, θ. It is therefore necessary to use a numerical optimisation algorithm to maximise the output–error log–likelihood function [7].

The model assumed in the output–error method is a more accurate description of the conditions prevailing in helicopter identification than is the equation–error model. To use the output–error method, measurements must be made of the outputs, z, and the inputs, u. The method assumes, correctly, that there is noise in the measurement of z. It is also takes u to be noise free, which is untrue. However, the input measurements typically contain little noise, and so this assumption is not unreasonable.

For helicopter applications, the output–error method can therefore be expected to produce more accurate parameter estimates than the equation–error method. As a result, the equation–error method is often used to obtain initial parameter estimates. These are then used as the starting point for the more powerful output–error algorithm [11].

Finally, combining equations (1.8), (1.9) and (1.18) gives the following expression for the output–error dispersion matrix :

$$D = M^{-1}$$

$$M = \int_0^T (C\, dx(t)/d\theta)^T \quad R^{-1} \quad (C\, dx(t)/d\theta)\, dt \qquad (1.19)$$

where,

T is the length of the test record used.

1.4 DIFFICULTIES OF HELICOPTER SYSTEM IDENTIFICATION

While identification of fixed–wing aircraft has in the past been relatively successful [12], particular problems have been encountered in rotorcraft applications which have complicated the identification task. These problems are mainly due to three factors : the complexity of the system, inherent instabilities, and data measurement difficulties [5,11].

Rotorcraft are complex, highly–coupled systems. Due to the coupling, motion in one axis will excite motion on several other axes. This results in the system responses being highly correlated, which causes severe problems when identification is attempted. Moreover, high order models are needed to describe such responses. There are therefore a large number of parameters to be identified.

Secondly, as mentioned in section 1.2, the theoretical model of interest is obtained by linearising a more general non–linear model about an operating point. However, the system is usually poorly damped, and often exhibits instabilities. Hence, only very short flight test records are obtained before the system departs too far from the operating point and the linear model becomes invalid.

Finally, the signal–to–noise ratio of flight test data is poor due to the high vibration levels in rotorcraft. Other instrumentation problems can also occur e.g. the well known difficulties involved in accurately measuring airspeed at hover and in low speed flight [5].

As a result of these difficulties, system identification of rotorcraft has been of only mixed success to date.

1.5 REVIEW OF INPUT DESIGN APPROACHES

1.5.1 Non– Optimal Inputs

In the past, various semi– intuitive approaches have been proposed by a number of authors for the design of system identification test inputs.

Perhaps the most straight– forward of these uses the inputs applied to the system during normal operation as the identification test inputs. However, problems can arise if these inputs are not persistently exciting i.e. if they are zero (or constant) for lengthy periods of time, or if their frequency components do not adequately span the pass– band of the system [13]. Also, more subtle problems can occur if there are high correlations between the input and output of the system. These are often present when the input is being used as a control to maintain some specified output i.e. in a closed– loop manner [13].

For example, consider the simple closed– loop system given in figure 1.1. This closed– loop system responds to the command input, u(s) and the noise, n(s). In practice, the noise, n(s), typically results from external disturbances to the system, e.g. air turbulence acting on an aircraft. To identify the open– loop transfer function, H(s) the output, y(s) and the error signal, x(s) must be available. The transfer function can be expressed as y(s)/x(s).

Now,

$$y(s)/x(s) = y(s)/(u(s) - G(s)y(s)) \tag{1.20}$$

When $G(s)y(s) >> u(s)$,

$$y(s)/x(s) \rightarrow -1/G(s) \tag{1.21}$$

In this case, the inverse of the feedback transfer function will be obtained as the open– loop transfer function. This is clearly incorrect.

Figure 1.1

A simple closed–loop system with process noise

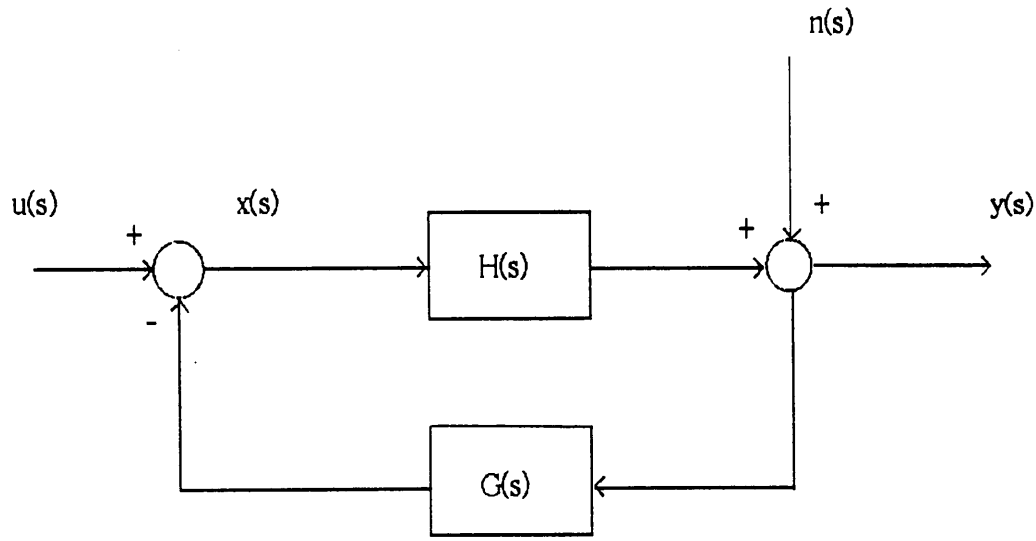
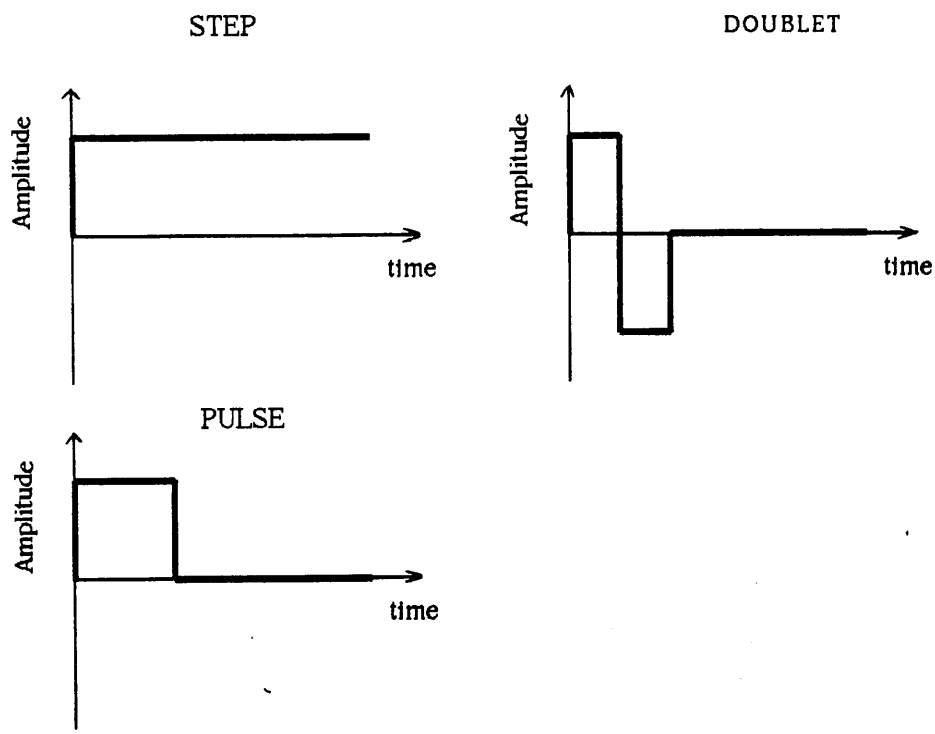


Figure 1.2

Time histories of the typical multi–step inputs conventionally used in system identification (step, pulse, and doublet)



This limiting condition is only reachable in practice when the command input, u(s) is held at zero, and the system is excited by the noise. However, biases will occur in the open–loop transfer function estimate even if u(s) is non–zero, as a result of the correlation between x(s) and the noise n(s). There are therefore serious difficulties involved in the identification of open–loop information from closed–loop data.

Sine Wave Inputs

A more successful alternative to using normal operating inputs is to use pure sine waves as inputs. By applying several different frequencies, and allowing the system to reach a steady state after each change of frequency, it is possible to obtain information about the gain and phase of the system transfer function at those frequencies. However, this can be a very time–consuming, and expensive procedure if many frequencies are required [14].

This problem may be overcome by using swept sine waves as inputs. In a swept sine wave, the frequency starts at the begining of the range that is of interest, and is then continuously increased until the desired final frequency has been reached. This type of input can therefore be used to excite a range of frequencies at once, greatly reducing the time required when compared with using individual pure sine waves. Such swept sine waves have in the past been found to be good general–purpose inputs for many applications [14].

In the aerospace field, pure sine wave inputs have been widely used in the past with fixed–wing aircraft e.g. [15]. More recently, swept sine waves have been successfully used of the identification of the XV–15 tilt–rotor aircraft [16], and the Aerospeciale Puma and Westlands Lynx helicopters [17,18].

However, these particular rotorcraft applications involved identifying the closed–loop transfer functions of the system i.e. with stability augmentation in use. When these rotorcraft are used without augmentation, in order to permit identification of their open–loop characteristics, they are only marginally stable at best. The swept sine wave inputs produce too large an excitation at the frequencies where the unstable modes are located, resulting in only short test records before the response becomes non–linear.

In principle, this can be overcome by modifying the swept sine waves so that they are selective and avoid exciting those frequencies corresponding to the unstable modes. However, to the author's knowledge this has not been attempted in practice in either the aerospace field or any other area of application, since the resulting input is very complex. This complexity means, in particular, that the input cannot be applied manually by the pilot. Moreover, as shown in Chapter 2 below, such selective excitation may be easily achieved with simple multi–step inputs, avoiding the need for the more complex swept–sine based approach.

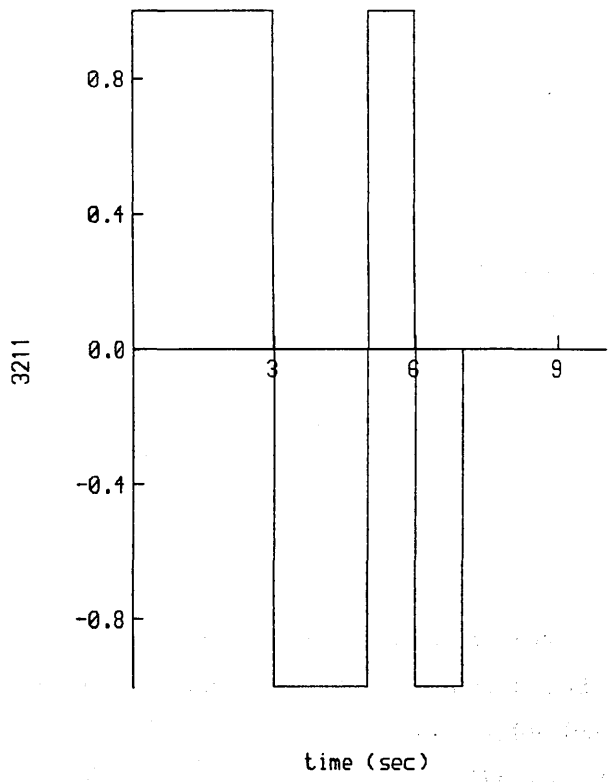
Swept–sine inputs may nevertheless still be useful for some open–loop rotorcraft applications. One case is where the frequencies of interest are not located near the unstable modes. Typically this is true for the rotor dynamics, since these are concentrated at higher frequencies while the unstable fuselage modes are at lower frequencies.

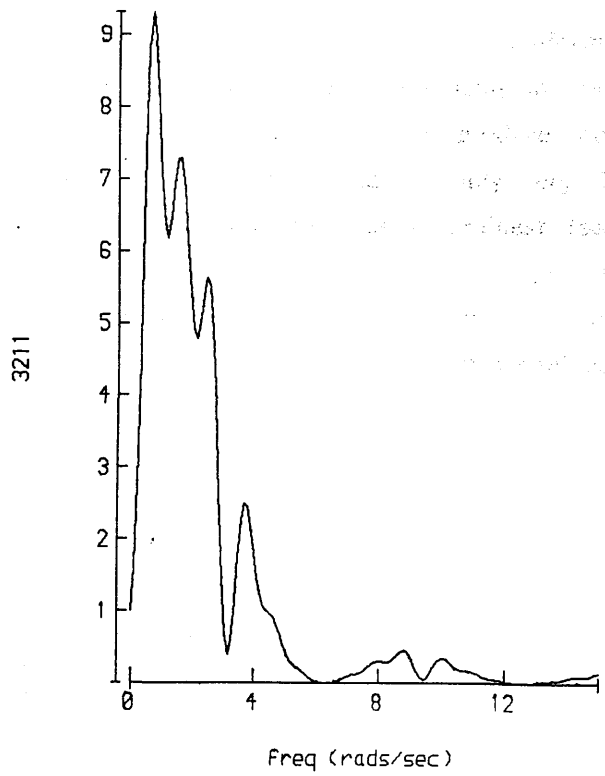
Multi–Step Inputs

While swept sine waves are often useful inputs, in some practical situations they can be difficult to use due to their complex shape. As a result, there has been interest in designing simpler forms of input. In particular, a large effort has been directed towards designing binary multi–step inputs, which have only two amplitude levels 'on' and 'off', and which consist of a sequence of step transitions between these levels.

Figure 1.3

Time history and auto–spectrum of the 3211 PRBS–based multi–step input





Wide use is made of the simplest multi–step inputs, i.e. steps, pulses, and doublets (see figure 1.2) [14]. Despite their simplicity, these inputs frequently give good results, especially where low order systems with relatively slow dynamics are used. For example, it is standard practice to use step inputs for the identification of chemical plants and other large, slow installations [19]. However, where more complex, higher bandwidth systems are involved, these inputs can produce much poorer results.

More sophisticated types of multi–step input can be used to greatly improve the success of identification for both simple and complex systems. In particular, many authors have concentrated on using pseudo–random binary sequences (PRBS) as inputs. These are periodic inputs which have an auto–spectrum which approximates that of band–limited white noise, and therefore can be used to excite a specified range of frequencies [20,21]. Good results have been obtained with PRBS inputs in many applications outside the aerospace field, e.g. [13,21,22,23].

In fixed–wing and rotorcraft applications, a wide variety of multi–step inputs have been used, including steps, pulses, doublets, and PRBS inputs. These appear to have been found to be relatively successful for fixed–wing aircraft [12]. For example, a PRBS–based input called the 3211 (see figure 1.3) has been reported to give particularly good results [24].

However, these types of multi–step input aim to uniformly excite a range of frequencies. Hence, they suffer from similar problems to swept sine wave inputs when they are used with rotorcraft, i.e. they can produce too large an excitation of the rotorcraft unstable modes. This leads to only very short test runs being obtained before the system response becomes non–linear (see section 1.4). In Chapter 2 below, a method is described for the design of multi–step inputs which avoid exciting the unstable modes and so lead to longer test runs. Previously, other authors have tried various different techniques in an attempt to overcome the problem of short test runs.

Firstly, several different test runs can be combined together to give greater information about the system than is contained in any single run. It is possible to simply concatenate these runs, and this approach is advocated by some authors [e.g. 5]. Unfortunately, discontinuities will be present if $x_i(T) \neq x_{i+1}(0)$, where $x_i(T)$ is the response at the end of time history i, and $x_{i+1}(0)$ is the response at the start of the subsequent time history, i+1. Such discontinuities can result in difficulties. In the equation–error approach, the identification problem is treated as a regression problem. Hence, time–domain equation–error identification is unaffected by any discontinuities, although it has been found that some errors can be introduced if frequency–domain equation–error identification is used [11]. However, with output–error methods, the dynamics of the system under investigation are explicity taken into account. As a result, discontinuities cause severe difficulties for output–error identification.

An alternative approach called the method of 'successive residuals' has also been suggested for combining the data from several runs [26]. In this approach, the model to be identified is partitioned into suitable sub–sets. The full model is then built up by identifying each of these subsets from appropriate test runs, and combining these in a mathematically consistent manner to obtain the full model. In this way, the identification problem is broken down into several smaller problems which can be tackled individually. Several authors have found that this method gives better results than simple concatenation of the test runs [6,26].

Finally, the duration of test runs can also be increased by using stability augmentation of the rotorcraft to improve the damping of the system. Unfortunately, as explained above, identifying open–loop information from closed–loop data is fraught with difficulties, and can lead to severely biased estimates. Nevertheless, some authors appear to have successfully used closed–loop testing [27], and this area requires further work.

1.5.2 Optimal Inputs

Introduction

In some applications, the inputs discussed in the previous section are not able to produce identification results of sufficiently high quality. In other cases, the collection of test data for identification purposes is expensive, e.g. with aircraft, and it is therefore desirable to use efficient inputs which reduce the amount of test data required. Much attention has therefore been directed at developing more rigorous methods for identification test input design.

In all of these more rigorous approaches, the task of input design is cast as an optimisation problem. Their aim is to design the best, or optimal, input for a given problem.

Input Design Criteria

The first step in designing an 'optimal' input is to obtain some quantitative cost function that provides a measure of the 'goodness' of any particular input. It is then possible to choose that input which maximises this 'goodness' function or, equivalently, minimises its inverse.

In order to decide on such a criterion for designing inputs, the uses for which the identification is intended must be considered. One of the principal aims of the present work on identification is to help in the validation of theoretical flight mechanics models. The identification must therefore produce accurate values for the model parameters.

When parameter accuracy is the criterion used, much of the work on input design reported in the literature has been concerned with designing inputs to minimise some function of the dispersion matrix. This is because for an efficient estimator [8], the Cramer–Rao bound [9] states that :

$$\text{cov}\ (\theta') = D \qquad (1.22)$$

where,

θ' are the parameter estimates obtained from the identification.

D is the dispersion matrix (see equation (1.8), (1.9)).

Now in many situations an efficient estimator does not exist, and the dispersion matrix then merely gives a lower bound on the covariance of the parameter estimates. In such cases, these input designs can no longer be relied upon.

However, inputs designed using the dispersion matrix are useful in situations where sufficiently long test records are available, since at least one class of estimators – maximum likelihood estimators – are asymtotically efficient [8].

Several special functions of the dispersion matrix have been used in the past as performance measures [28]. These include :

1) **A– Optimality** : minimise tr(D) i.e minimise the average covariance of the parameter estimates.

2) **E– Optimality** : minimise $\lambda_{max}(D)$, where λ_{max} is the maximum eigenvalue of D.

3) **D– Optimality** : minimise $|D|$

The choice of which of these criteria to use is, unfortunately, relatively arbitary. However, D– Optimality possesses several important advantages over A– and E– Optimality.

Firstly, D– Optimality is invariant under scale changes in the parameters, and linear transformations of the system being used [28]. This is not true for A– and E– Optimality.

Another advantage of D– Optimality is that it implies G– Optimality [28]. The A–, E–, and D– Optimality criteria are based on the accuracy of the parameter estimates. However, in other applications the emphasis may lie on obtaining a model which accurately predicts the response of the system to an input. G– Optimality caters for such situations.

G-Optimality : minimise $\max_{u} \text{cov}(z(t,\theta'))$

where,

z is the response of the model to an input u,

θ' are the parameter estimates.

Hence, D−Optimality produces both accurate parameter estimates and accurate responses.

Finally, in addition to the more widely used optimality criteria mentioned above, particular authors have in the past approached the input design problem using other criteria. For example, Ramachandran [29] attempted to minimise the parameter correlations i.e. minimise the off−diagonal elements of the dispersion matrix. Whereas Chen [30] worked on generating inputs which would give a specified dispersion matrix.

Noise Characteristics

It is widely assumed that the noise in the system under consideration is Gaussian. However, in practice this is rarely true. As a result, growing attention has been given to alternative types of noise statistic. For example, in the bounded noise approach [31], the only assumption is that the noise lies between known upper and lower bounds. No knowledge of the distribution between these bounds is required.

However, in aerospace applications, all input design and identification work to date has concentrated on Gaussian noise. For rotorcraft in particular, the noise distribution is extremely complex. Moreover, the distribution varies with the loads placed on the rotor, and so will alter during a manoeuver. Given this complex behaviour, any simple distribution is inadequate. However, the Gaussian form often leads to significant savings in computing time, and has been found to be a useful approximation until a more accurate distribution is developed.

Constraints

In practical systems, where the noise is finite, an input of infinitely large magnitude will produce a response with an infinite signal– to– noise ratio. The noise can then be neglected, and perfectly accurate parameter estimates will be obtained. It can therefore be seen that the input which optimally minimises some function of the dispersion matrix will be of infinite magnitude. This is clearly impractical.

It is therefore essential that constraints are included in such optimisation, in order to restrict attention to practical inputs. The simplest, and least restrictive, constraint is merely to insist that inputs have finite energy. This has been the most widely used constraint reported in the literature, and is usually expressed as follows [28] :

$$\int_0^T u^T(t)\, u(t)\, dt \quad = \quad 1 \qquad (1.23)$$

where,

u(t) is the input,

T is the duration of the input.

Equation (1.23) restricts the input to have an energy equal to unity. Now the dispersion matrix varies linearly with input energy. The input which is optimal at one energy can therefore simply be scaled to give the input which is optimal at a different energy. The particular energy used in (1.23) is therefore arbitary, and so is chosen for convenience to be unity [28].

A more restrictive constraint is to insist that the input amplitude lies below a certain level, that is :

$$|u(t)| \leqslant k, \quad k \text{ some constant} \qquad (1.24)$$

The resulting optimal input will be of a bang– bang nature i.e. a multi– step [28].

Finally, the following constraint has been proposed for situations where it is undesirable for the input to produce a large system response [28] :

$$\int_0^T (u^T(t)\ P\ u(t) + y^T(t)\ Q\ y(t)\ )dt \quad = \quad 1 \qquad (1.25)$$

where,

u(t) is the input,

y(t) is the corresponding noise– free system response.

P, Q are weighting matrices,

T is the duration of the input.

This constraint restricts the combined energy of the input and output. However, this constraint appears to have only been considered from a theoretical viewpoint, and no reports have been found of it being used in practice.

The question of constraints has been largely neglected by many authors in favour of other areas of optimal input design. However, the importance of such constraints in determining the characteristics of the optimal inputs, and the practicality and usefulness of these inputs must be stressed. Constraints will therefore be considered later in greater depth later.

Optimal Inputs for Aircraft and Rotorcraft Applications

Optimal identification test inputs have been studied for several fixed– wing applications. A major advantage of using optimal inputs in aircraft applications is that due to the greater efficiency of such inputs, a reduction is obtained in the amount of flight testing required to obtain sufficiently accurate parameter estimates. Since flight testing is extremely expensive, any reduction can produce significant cost savings.

A study by Gupta et al [32] investigated the use of A– and D– optimal inputs. The simple energy constraint given in (1.23) was used. It was concluded from simulation results that these inputs produced more accurate parameter estimates than traditional multi– steps. However, no flight trials were performed to verify these results.

Chen [30] examined the use of optimal inputs which gave the same accuracy of parameter estimates as multi–steps, but which required flight tests of shorter duration. Once again. a theoretical study showed that significant flight test reductions could be achieved.

However, less encouraging results were obtained by Plaetschke [24]. In this case, optimal inputs were generated and compared with the conventional 3211 PRBS input in flight trials using a De Havilland DHC–2 Beaver. From these trials, it appeared that the 3211 and the optimal inputs produced comparable results.

Work has also been carried out with rotorcraft on the design of optimal inputs. The main aim of this work has been to improve the success of rotorcraft identification and obtain more accurate parameter estimates.

In a study of optimal inputs for rotorcraft, Hall et al [25] encountered similar difficulties to those found with fixed–wing aircraft. A UH–1H helicopter was used for the flight trials, and A– and D– optimal inputs were compared against standard doublet inputs. The doublet inputs appeared to be more successful than the optimal inputs. However, these results were felt to be inconclusive by Hall et al.

From this work, it appears that optimal inputs have given excellent results in simulations, but have been less successful in flight trials. There therefore may be practical difficulties involved in the use of optimal inputs that need to be resolved before the full potential of these inputs can be realised. Several proposals are made on this subject later by the present author.

1.6 STATEMENT OF ORIGINALITY

The following results of the research work presented in this thesis are, as far as is known, original and, as noted below, some of these results have been published.

CHAPTER 2

Section 2.2 – The design approach for multi–step inputs.

Section 2.3 – The design and flight testing of multi–step inputs for the Lynx helicopter at RAE (Bedford).

Section 2.4 – The identification results for the double–doublet input. The ill–conditioning encountered during the output–error identification of an unstable system.

Leith, D.J.; Murray–smith, D.J. 'Experience with multi–step test inputs for helicopter parameter identification' Paper no. 68, presented at the 14th European Rotorcraft Forum, Milan, 1988

CHAPTER 3

Section 3.1 – The application of the following constraints to input design :

i) Linearity constraint obtained by repeatedly linearising a non–linear model

ii) Output amplitude constraint

iii) Response robustness constraint

iv) Parameter robustness constraint

Section 3.2 – The time–domain design algorithm, and the simulation results for energy constrained optimal inputs.

Section 3.3 – The time–domain design algorithm, and the simulation results for output and input energy constrained optimal inputs.

Section 3.4 – The time–domain design algorithm, and the simulation results for output amplitude constrained optimal inputs.

Section 3.5 – The time–domain design algorithm, and the simulation

results for response robust optimal inputs.

Section 3.6 – The simulation results for parameter robust optimal inputs.

Section 3.7 – Combined response robustness and parameter robustness.

CHAPTER 4

Section 4.2 – The optimal Lynx input.

APPENDIX B

Section B.2 – The recursive method for calculating h_k^i and g_k^i for Chebyshev polynomials.

Section B.3 – The stabilised basis functions for use with unstable systems.

Section B.4 – The implementation of amplitude constraints using a smooth, continuous scalar function.

REFERENCES FOR CHAPTER ONE

1. Bradley, R.; Padfield, G.D.; Murray–smith, D.J. 'Validation of helicopter mathematical models' presented at the 2nd one day symposium on model validation, Inst Meas and Control, London, May 1989

2. Padfield, G.D. 'A theoretical model of helicopter flight mechanics for application to piloted simulation' Royal Aerospace Establishment Tech Report 81048, 1981

3. Smith, J. 'An analysis of helicopter flight mechanics : part 1 – users guide to the software package HELISTAB' Royal Aerospace Establishment Tech Memo FS(B) 569, 1984

4. Black, C.G.; Murray–smith, D.J.; Padfield, G.D. 'Experience with frequency–domain methods in helicopter system identification' Paper no. 76, 12th European Rotorcraft Forum, 1986

5. Kaletka, J. 'Rotorcraft identification experience' Lecture no. 7, AGARD lecture series no. 104, 1979

6. Wang, J.C.; DeMiroz, M.; Talbot, P. 'Flight test planning and parameter extraction for rotorcraft system identification' 3rd Flight Testing Conference, Las Vegas, 1986

7. Maine, R.; Iliff, R. 'Identification of dynamic systems – theory and formulation' NASA RP–1138, 1985

8. Norton, J.P. 'An introduction to identification' Academic Press, London 1986

9. Rao, C.R. 'Information and the accuracy attainable in the estimation of statistical parameters' Bull Calc Math Soc, Vol 37, pp81–91, 1945

10. Klein, V. 'Identification evaltuation methods' Lecture no. 2, AGARD lecture series no. 104, 1979

11. Black, C.G. 'A methodology for the identification of helicopter mathematical models from flight data based on the frequency domain' PhD Thesis, University of Glasgow, 1988

12. Marchand, M. 'Determination of aircraft derivatives by automatic parameter adjustment and frequency response methods' Paper no. 17, AGARD Conference Proceedings No. 172, 1974

13. Godfrey, K.R. 'Practical problems in identification' Signal Processing for Control, pp358–386, Ed Godfrey, K.R.; Jones, P. Springer–Verlag, Berlin 1986

14. Godfrey, K.R.; Brown, R.F. 'Practical aspects of the identification of process dynamics' Trans Inst M C Vol 1, pp85–95, 1979

15. Donegan, J.J 'Determination of lateral stability derivatives and transfer function coefficients from frequency domain data for lateral measurements', NASA Report 1225, 1955

16. Tischler, M.B. 'Frequency–response identification of XV–15 tilt–rotor aircraft dynamics' NASA TM 89428, 1987

17. Houston, S.; Horton, R. 'The identification of reduced order models of helicopter behaviour for handling qualities studies' Royal Aerospace Establishment Tech Memo FS(B) 682, 1987

18. Houston, S.; Tarttelin, P. 'Theoretical and experimental correlation of helicopter aeromechanics in hover' Proc 45th AHS annual forum, 1989

19. Godfrey, K.R.; Shackcloth, B. 'Dynamic modelling of a steam reformer and the implementation of feedforward/feedback control' Meas and Control, Vol 3, pp65–72, 1970

20. Briggs, P.A.; Godfrey, K.R.; Hammond, P. 'Estimation of process dynamic characteristics by correlation methods using pseudo–random signals' Preprints IFAC Prague Symposium on System Identification, Paper 3.10, 1967

21. Godfrey, K.R. 'The theory of the cross−correlation method of dynamic analysis and its applications to industrial processes and nuclear power plant' Meas and Control, Vol 2, pp65−72, 1969

22. Cumming, I.G. 'On−line identification of a steel mill' Automatica Vol. 8, pp531−541, 1972

23. Billings, S.A. 'Modelling and identification of a 3−phase electric arc furnace' in 'Modelling of dynamical systems' Vol. 2, Ed. H. Nicholson, Peregrinus, London 1981

24. Plaetschke, E.; Schulz, G. 'Practical input design' Lecture no. 3, AGARD lecture series no. 104, 1979

25. Hall, W.E. 'Development of advanced techniques for rotorcraft state estimation and parameter identification' NASA CR 159297, 1980

26. DuVal, R.W.; Wang, J.; DeMiroz, M. 'A practical approach to rotorcraft systems identification' 39th AHS National Forum, St Louis, 1983

27. Kloster, M.; Kaletka, J.; Schaufele, H. 'Parameter identification of a hingeless rotor helicopter in flight conditions of increased instability' Vertica, Vol 5, pp261−278, 1981

28. Mehra, R.K. 'Optimal input signals for parameter estimation in dynamic systems − survey and new results' IEEE Trans Aut Control, Vol AC−19, pp753−768, 1974

29. Ramachandran, S. 'Optimal input design for parameter identification of dynamic systems' NASA STAR Issue 6, PhD Thesis

30. Chen, R.T. 'Input design for aircraft parameter identification' Paper no. 13, AGARD Conference Proc 172, 1974

31. Pronzato, L.; Walter, E. 'Experiment design in a bounded−error context : comparison with D−optimality' Automatica, Vol 25, pp383−391, 1989

32. Gupta, N.; Hall, W. 'Input design for the identification of aircraft stability and control derivatives' NASA CR 2493, 1975

CHAPTER TWO

A SIMPLE DESIGN APPROACH FOR MULTI–STEP INPUTS

2.1 INTRODUCTION

Much of the work on input design reported in the literature has been concerned with developing algorithms to minimise some function of the dispersion matrix. This is because, for an efficient estimator, the Cramer– Rao bound states that,

$$\text{cov}\ (\theta) = D \tag{2.1}$$

where, θ is a vector containing the parameter estimates,
and D is the dispersion matrix.

Now, in many situations an efficient estimator does not exist, and the dispersion matrix then merely gives a lower bound on the covariance of the parameter estimates. In such cases, these input design algorithms can no longer be relied upon.

However, inputs designed using the dispersion matrix are useful in situations where sufficiently long test records are available, since at least one class of estimators – maximum likelihood estimators – are asymptotically efficient.

Using conventional inputs, typically only short rotorcraft test records are obtained before the response departs too far from the operating point of the linearised model being used, and so becomes non– linear (e.g. see section 2.3 below).

Hence, as a first step, it was decided to concentrate on trying to design sub– optimal inputs which would give longer test records while at the same time giving a reasonably 'small' dispersion matrix. The next step would then be to design optimal inputs to strictly minimise some function of the dispersion matrix, while still giving sufficiently long test records. The results of the first step – sub– optimal inputs – are covered in this chapter.

Consideration of the robustness of the inputs was also included in the input designs. Robustness is an important, though often neglected, aspect of input design. Only an approximate model of the system involved is available beforehand, and so the inputs used must be insensitive to errors in the model. In addition, on the particular helicopter used in this work inputs are currently applied by the pilot via the normal controls, and so they must also be insensitive to being applied inaccurately. Since they are applied manually, they must also be kept relatively simple e.g. multi–steps. However, work is currently underway at RAE (Bedford) to develop an input device which will in the future allow inputs to be applied directly to the helicopter without pilot intervention [1].

2.2 INPUT DESIGN METHOD

2.2.1 Input Auto–Spectrum Design

The aim is to design an input which is robust, gives long linear test records, and which gives a reasonably 'small' dispersion matrix. Several general guidelines can be developed concerning the features that should be present in the auto–spectum of such an input.

Firstly, consider obtaining longer time histories. Typically, the transfer function between a given rotorcraft model state and a given control input contains large peaks. These peaks correspond to resonances in the system. If an input excites these resonances, then the response will be large, and will rapidly become non–linear, so leading to a short test record. Hence, by designing inputs which avoid exciting these resonances, longer test records can be obtained.

In addition, inputs should not contain a dc component. If a large dc component is present, then the aircraft response to the input will also contain a dc component. This is undesirable when using a model linearised about a particular operating point, since a dc component essentially means that the response is about a different operating point. If this operating point is significantly different from that used when linearising the model, then the model will be invalid.

Now, the next step is to consider how to obtain a reasonably 'small' dispersion matrix. Take the model,

$$\frac{dx(t)}{dt} = A\,x(t) + B\,u(t) \qquad z(t) = C\,x(t) + \eta(t)$$

where, $z(t)$ = measured responses
$\eta(t)$ = Gaussian process with zero mean and covariance, R
$u(t)$ = single control input
$x(t)$ = model states
A,B,C = model matrices

(2.2)

Let θ be a vector containing the model parameters that are to be identified, and let the true values of these parameters be given by $\underline{\theta}$. This system corresponds to the model used in output-error identification, and it was shown in section 1.3.4 that the dispersion matrix for this is given by,

$$D = M^{-1}$$

$$M = \int_0^T (dy(t)/d\theta)^T\, R^{-1}\, (dy(t)/d\theta)\, dt \qquad (2.4)$$

where,

$y(t) = C\,x(t)$

T is the length of the test record used.

This expression can be simplified by letting $T \to \infty$ [2]. Parseval's theorem then gives,

$$\int_0^{\infty} \frac{dy}{d\theta}^T(t)\, R^{-1}\, \frac{dy}{d\theta}(t)\, dt = \int_{-\infty}^{\infty} \frac{dY}{d\theta}^*(\omega)\, R^{-1}\, \frac{dY}{d\theta}(\omega)\, d\omega \qquad (2.5)$$

where,

$dY(\omega)/d\theta$ is the Fourier transform of $dy(t)/d\theta$,

ω is the complex frequency.

Now,

$$Y(\omega) = G(\omega)\, U(\omega) \tag{2.6}$$

$$\Rightarrow \quad dY(\omega)/d\theta = dG(\omega)/d\theta\, U(\omega) \tag{2.7}$$

$$= F(\omega)\, U(\omega) \tag{2.8}$$

where,

$Y(\omega)$, $U(\omega)$ are the Fourier transforms of $y(t)$, $u(t)$

$G(\omega) = C(j\omega - A)^{-1}B$, transfer function matrix

$F(\omega) = dG(\omega)/d\theta\,|_{\theta=\underline{\theta}}$

Hence,

$$M = \int_{-\infty}^{\infty} (F(\omega)U(\omega))^{*}\, R^{-1}\, (F(\omega)U(\omega))\, d\omega \tag{2.9}$$

$$= \int_{-\infty}^{\infty} F^{*}(\omega)\, R^{-1}\, F(\omega)\, S_{uu}(\omega)\, d\omega \tag{2.10}$$

where,

$S_{uu}(\omega)$ = auto-spectrum of input $u(t)$

$= E(U^{*}(\omega)\, U(\omega))$

$= U^{*}(\omega)\, U(\omega)$, since the input, u is completely deterministic in the current application.

In practical systems, $F(\omega)$ becomes negligible above some frequency, ω_c. Now consider the scalar case of a single-input single-output system with only one parameter to be identified. This gives,

$$M = R^{-1} \int_{-\omega_c}^{\omega_c} |F(\omega)|^2\, S_{uu}(\omega)\, d\omega \tag{2.11}$$

$$D = 1/M \tag{2.12}$$

Notice that the expression for M does not involve the phase of the input or the phase of $F(\omega)$. Hence, attention can be restricted to the magnitudes of the input and $F(\omega)$. It can then be seen that to obtain a small dispersion, D, a large M is needed i.e. the area under the curve $|F(\omega)|^2 S_{uu}(\omega)$ should be large.

Now, in order to obtain long test records, the input auto–spectrum, $S_{uu}(\omega)$ should avoid exciting the transfer function resonances. However, by exciting the frequencies away from the resonances, $|F(\omega)|^2 S_{uu}(\omega)$ can still be made to have a fairly large area.

Returning to the more general case of a vector model, exciting the frequencies away from the resonances will result in the elements of M being reasonably large, and intuitively, this should lead to a 'small' dispersion matrix, D. It should be noted that while it is well known that maximising M in this way can in fact produce a singular dispersion matrix [3], for the present application this can be neglected since the inputs used are significantly sub–optimal, and therefore do not maximise M sufficiently to cause difficulties.

The arguments concerning the dispersion matrix in particular, and those concerning the other guidelines in general, are very crude. However, the system transfer functions are typically only known approximately, and $F(\omega)$ is usually even less well known. Hence, when designing inputs only the general characteristics of $F(\omega)$ can be relied upon to any extent, and therefore sophisticated algorithms using detailed knowledge of $F(\omega)$ are redundant. Of course, if $F(\omega)$ is better known, then more sophisticated algorithms become worthwhile.

Finally, the robustness of the inputs has to be considered. Since the model is not known exactly, the frequencies of the resonances are not known exactly. To allow for errors in these, inputs should avoid exciting a range of frequencies around the predicted position of each resonance. This will also make the inputs less sensitive to errors introduced during the application of the inputs by the pilot, since errors in the input's auto–spectrum can then be tolerated to a greater extent.

So, by intuitive arguments, the features that should be present in the auto–spectrum of a reasonable input have been determined,

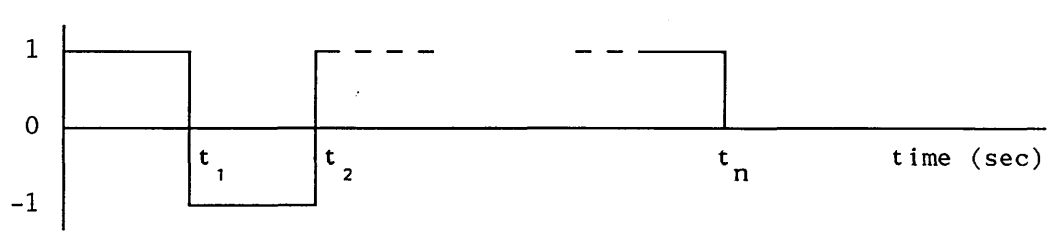
a) Auto–spectrum should avoid exciting transfer–function resonances, to give longer test records.
b) Auto–spectrum should also avoid exciting frequencies around the resonances, to give robustness.
c) Auto–spectrum should excite the remaining frequencies, however, to give a fairly 'small' dispersion matrix.

It is of interest to note that if there is no constraint on the magnitude of the system's response to an input, then features (a) and (b) can be discarded. The resulting input is then the standard wide–bandwidth type, such as an impulse or a pseudo–random binary input, which is in widespread use for identification.

2.2.2 Optimal Spectrum Program

Now that the criteria for designing the auto–spectrum of an reasonable input have been obtained, the next stage is to produce an input with the desired auto–spectrum characteristics. Since inputs must be applied by the pilot, they must be kept relatively simple e.g. a sequence of steps. A program has therefore been written that will generate the binary multi–step input whose auto–spectrum best meets a given specification.

Consider the general aperiodic binary multi–step input,



The Fourier Transform of this input is,

$$F(\omega) = \frac{1}{j\omega}\left[1 + 2\sum_{i=1}^{n-1}(-1)^i \exp(-j\omega t_i) + (-1)^n \exp(-j\omega t_n)\right] \tag{2.13}$$

where,

$F(\omega)$ is the Fourier Transform

ω is the frequency in radians/second.

$j = \sqrt{(-1)}$

$n+1$ is the number of steps in the input.

t_i is the time in seconds of the i^{th} step in the input; $t_0 = 0$ sec.

Now define the cost function,

$$I = \sum_{k=1}^{m} a_k \, |F(\omega_k)|^2 \tag{2.14}$$

where,

a_k are constants, k= 1, 2, ... m

ω_k are frequencies in rads/sec, k= 1, 2, ... m

The optimal spectrum program takes as input,

1) The number of steps, n, in the input.
2) The number, m, of constants in the cost function, I
3) The values of the weightings, a_k, and the frequencies, ω_k, in the cost function, I.

The program will then calculate the times, t_i, of the steps in the input that will result in the cost function, I being maximised. Specifying a large, positive a_k results in an input with a large auto–spectrum component at frequency ω_k. Conversely, specifying a large negative a_k results in an input with a small auto–spectrum component at the corresponding frequency, ω_k.

This permits the straightforward synthesis of inputs with auto–spectra meeting the guidelines developed in section 2.2.1. The optimal spectrum program, in conjunction with the auto–spectra guidelines, therefore forms a simple input design approach.

2.3 LYNX FLIGHT TRIALS

2.3.1 Design of Lynx Test Inputs

In order to assess the input design method presented above, inputs were designed for the Lynx helicopter at RAE (Bedford), and subsequently underwent flight trials.

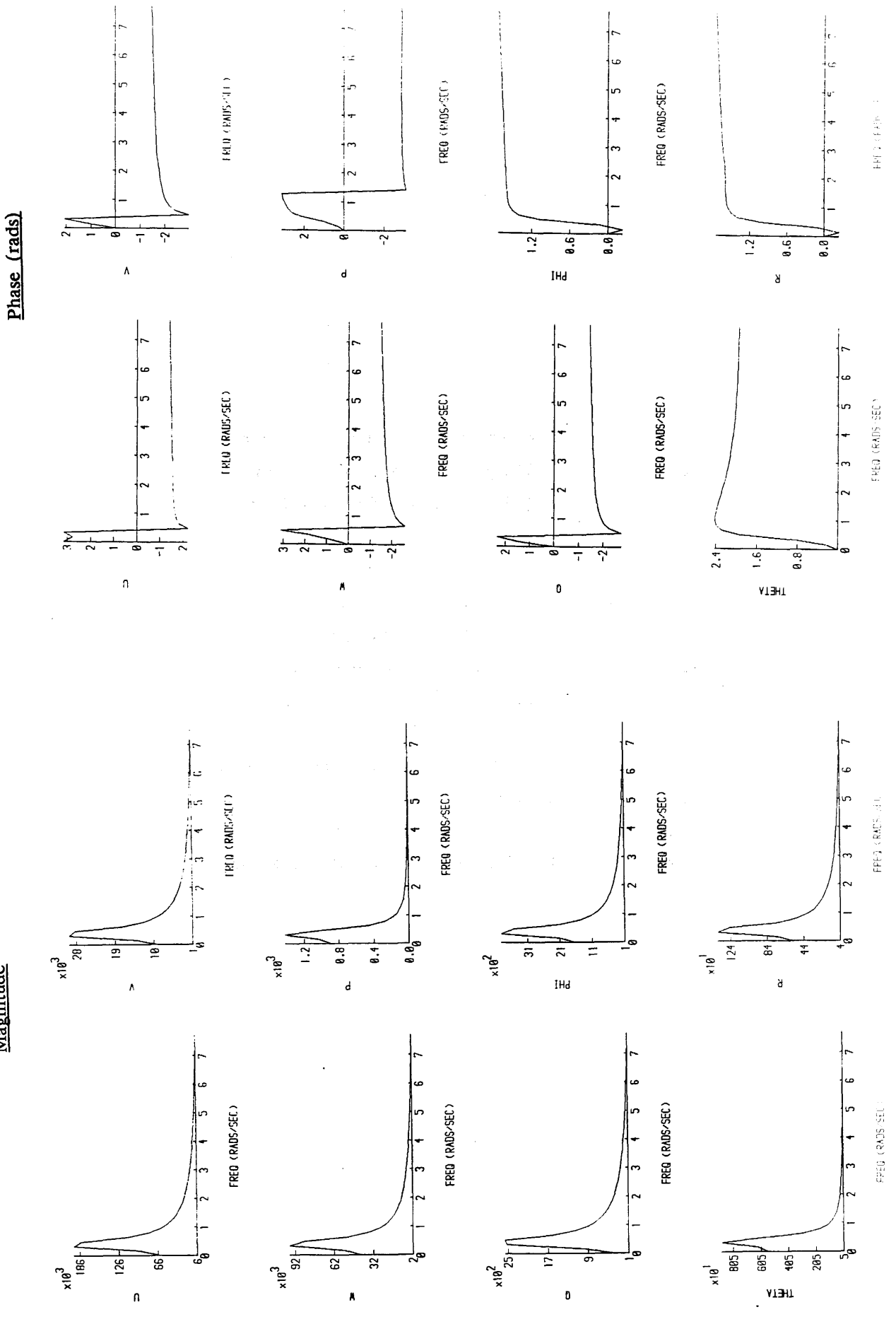
Figure 2.1 shows the magnitude and phase of the theoretical HELISTAB pseudo transfer functions associated with the Lynx longitudinal cyclic control for 80 knots level flight [3]. At this flight condition, the Lynx helicopter is unstable. Hence, the impulse response of the helicopter diverges to infinity as time goes to infinity. The transfer functions are therefore given by the Laplace transform of the impulse response rather than the Fourier transform, since the Fourier transform cannot be used with a divergent function. However, the results shown in figure 2.1 were obtained by taking the Fourier transform of the first 60 seconds of the theoretical impulse response of the Lynx. Hence, they are not the true magnitude and phase of the transfer functions. Nevertheless, they reflect the characteristics of the system, and are sufficient for the present purpose.

It can be seen that there is a large peak in the magnitude transfer functions around 0.3 rads/sec, and that there is also a rapid change in phase. This corresponds to the unstable phugoid–like mode of the Lynx, which has a theoretical natural frequency of 0.36 rads/sec.

The optimal spectrum program was therefore used with the weightings shown in table 2.1.

Figure 2.1

Magnitude and phase transfer functions between the longitudinal cyclic input and the outputs of the theoretical HELISTAB model. Linear 8th order model used, with Lynx at 80 knots level flight.



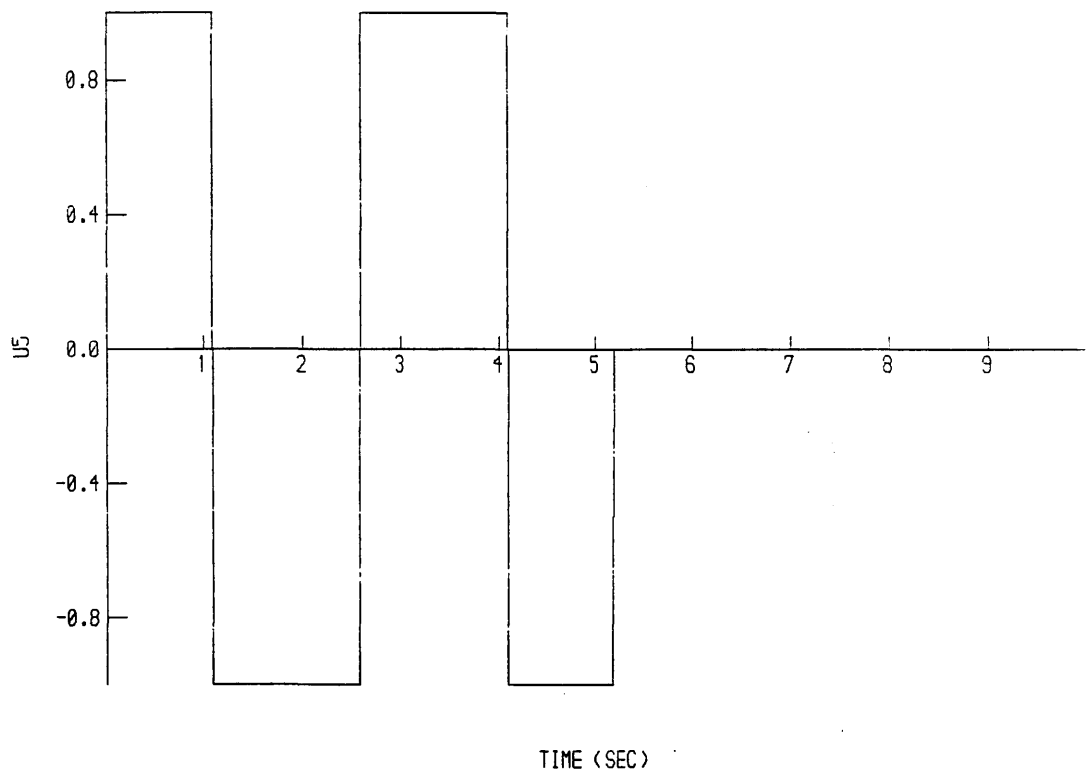
Frequency (rads/sec)	Weighting
0.00	-5.0
0.30	-5.0
0.36	-5.0
0.50	-5.0
0.60	-5.0
1.00	-2.0
2.00	5.0
2.50	5.0
3.00	0.5
3.50	-2.0
6.00	-2.0

Table 2.1 - *The weightings used with the optimal spectrum program to generate longitudinal cyclic inputs for a Lynx helicopter at 80 knots level flight.*

The first weighting was selected to ensure that the input did not contain a dc component. The following five weightings were chosen to ensure that the input avoided exciting the resonance at around 0.3 rads/sec. The next three weightings were then used to produce a rise in the input auto–spectrum between 2 and 3 rads/sec. Finally, the second last weighting was used to prevent this rise in the auto–spectrum from spreading out to higher frequencies, and the last weighting was similarly used to reduce the power at high frequencies. This was because previous experience with the Puma at RAE (Bedford) suggested that the theoretical model was only valid up to around 3 rads/sec. Above this frequency, rotor dynamics appear to dominate the response, and these are not included in the simplified model that is being used [4].

Figure 2.2

Time history and auto–spectrum of the input generated by the optimal spectrum program in accordance with the weightings given in Table 2.1



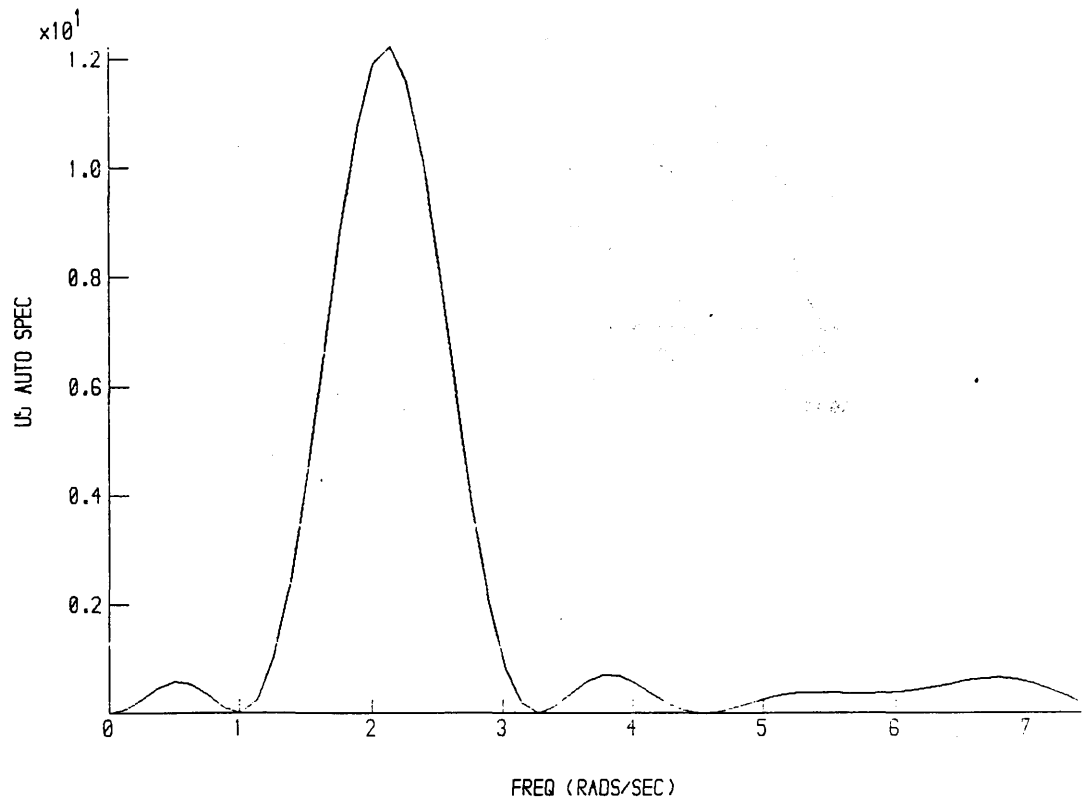


Figure 2.3

Time histories and auto–spectra of double–doublet and 1221 inputs.

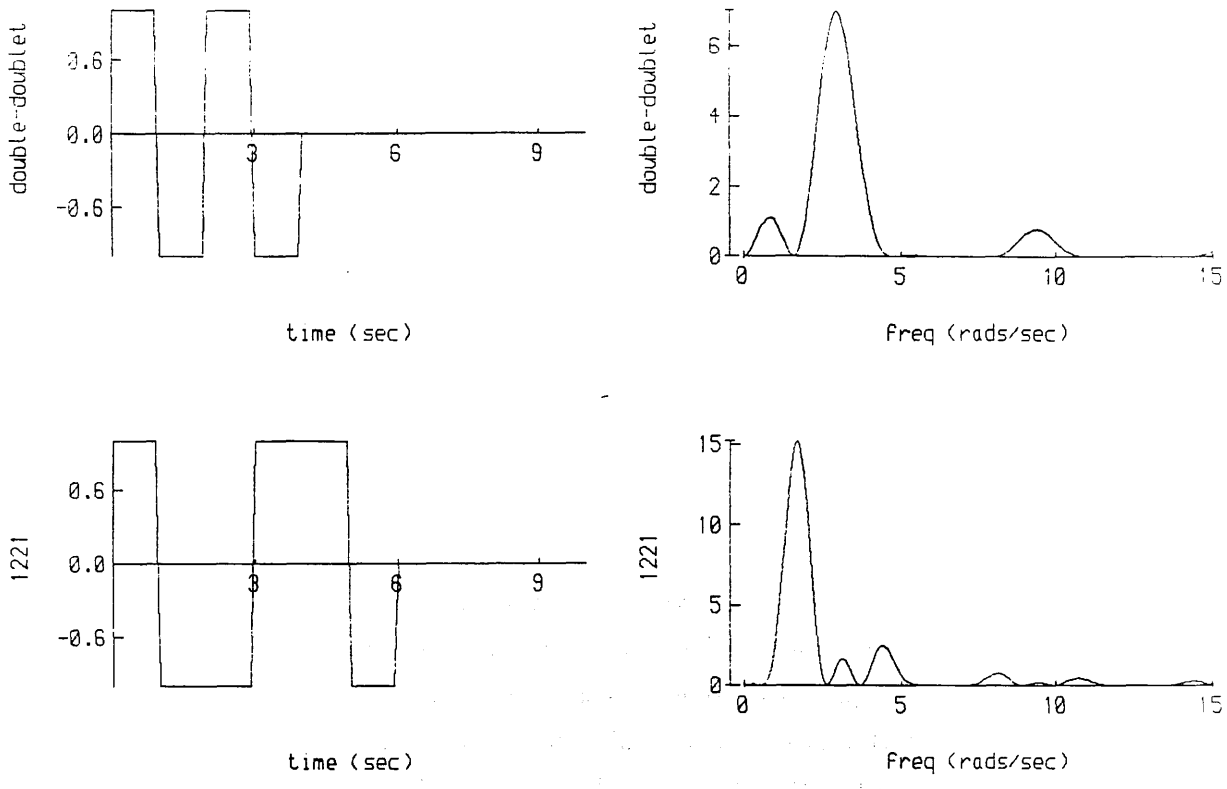
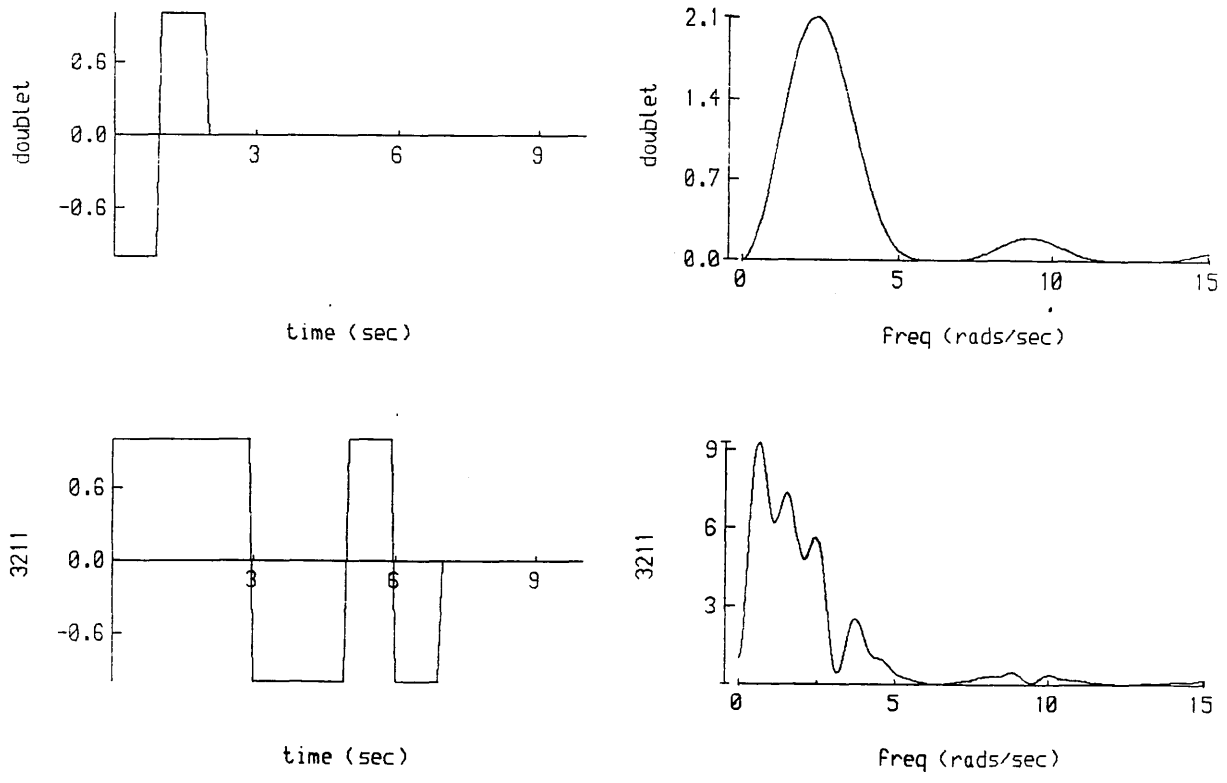


Figure 2.4

Time histories and auto–spectra of doublet and 3211 inputs.



Using these weightings, an input consisting of 5 steps was generated by the optimal spectrum program (see figure 2.2). It can be seen that the auto–spectrum of this input has little power below 1 rad/sec, and that most of the power is concentrated between 1 and 3 rads/sec, as desired.

The timings of the steps in this input are as follows,

t_1 = 0.00 sec
t_2 = 1.08 sec (t_2-t_1=1.08 sec)
t_3 = 2.59 sec (t_3-t_2=1.51 sec)
t_4 = 4.10 sec (t_4-t_3=1.51 sec)
t_5 = 5.18 sec (t_5-t_4=1.08 sec)

It is unrealistic to expect a pilot to apply an input with such awkward timings, so this input was modified to give two versions with much simpler timings. This should be straightforward, since the original input was designed to be robust, and in particular to withstand being applied inaccurately by the pilot. Hence, moderate changes in the timings of the input should have only a small effect.

The modified inputs were called the *double–doublet* and the *1221*, and have timings as follows,

1) Double-Doublet

t_1 = 0.0 sec
t_2 = 1.0 sec (t_2-t_1=1.0 sec)
t_3 = 2.0 sec (t_3-t_2=1.0 sec)
t_4 = 3.0 sec (t_4-t_3=1.0 sec)
t_5 = 4.0 sec (t_5-t_4=1.0 sec)

2) 1221

t1 = 0.0 sec
t2 = 1.0 sec (t_2-t_1=1.0 sec)
t3 = 3.0 sec (t_3-t_2=2.0 sec)
t4 = 5.0 sec (t_4-t_3=2.0 sec)
t5 = 6.0 sec (t_5-t_4=1.0 sec)

Note that the timings of these inputs have been arranged so that the inputs have no dc component.

It can be seen (figure 2.3) that the auto–spectra of these inputs are very similar to that of the original input. However, certain differences are apparent.

In the case of the double–doublet, it can be seen that the input contains little power below 2 rad/sec. Hence, it can accomodate a larger error in the predicted frequency of the unstable resonance of the Lynx than the original input. However, the double–doublet contains slightly more power below 1 rad/sec than the original input, and so may excite the unstable mode more and so give shorter test records.

In contrast, the 1221 has less power below 0.75 rad/sec than the original input, and so should give longer test records. However, since the region around the unstable resonance where the input power is low is smaller than in the original input, the 1221 is more sensitive to errors in the frequency of the resonance.

Hence, the double–doublet and the 1221 complement each other, and the original input is a compromise between the two. It was therefore decided to study both of these inputs in more detail.

2.3.2 Flight Trials of Lynx Test Inputs

The inputs developed were used in flight trials with the Lynx helicopter at RAE (Bedford).

In past rotorcraft identification work at the University of Glasgow (e.g. [4]), and in much of the identification work reported in the literature, the inputs used have consisted of doublets and 3211's (see figure 2.4). These inputs were therefore used as references against which to compare the double–doublet and 1221 inputs.

The HELISTAB 8^{th} order helicopter model used in this work has potentially 96 parameters to be identified. In order to make the problem more manageable, attention was initially restricted to the pitching moment equation of this model. This equation describes the behaviour of the pitch rate of the helicopter, and contains 7 parameters, as follows.

$$\frac{dq}{dt}(t) = M_u u(t) + M_w w(t) + M_q q(t) + M_v v(t) + M_p p(t) + M_{\eta_{1s}} \eta_{1s} + M_{\eta_{1c}} \eta_{1c}$$

where,

$u(t)$ = longitudinal velocity — $w(t)$ = vertical velocity

$q(t)$ = pitch rate — $v(t)$ = lateral velocity

$p(t)$ = roll rate

$\eta_{1s}(t)$ = longitudinal cyclic control input

$\eta_{1c}(t)$ = lateral cyclic control input

M_u, M_w, M_q, M_v, M_p, $M_{\eta_{1s}}$, $M_{\eta_{1c}}$ are the parameters to be estimated in the system identification.

A frequency–domain equation–error program developed by Black [5] was used to perform the initial identification.

Four inputs were used in the flight trials,

1. 3211
2. Double–Doublet
3. Doublet
4. 1221

An input amplitude of 0.02 radians was used in all cases. This corresponds to the smallest practical pilot input.

The results obtained for these inputs are now given.

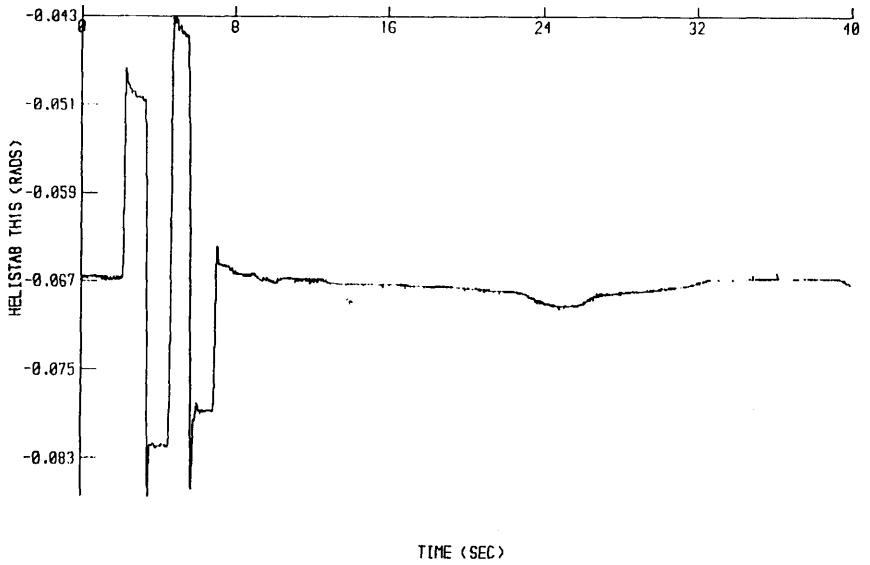
3211 Input

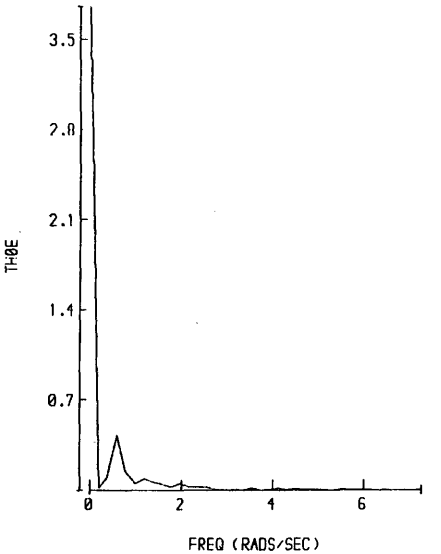
The 3211 input gave very short test records – typically only 3 seconds long – before the pilot was forced to recover control of the aircraft. Since the 3211 input itself is 7 seconds long, it was not possible to fully apply the input in the available time.

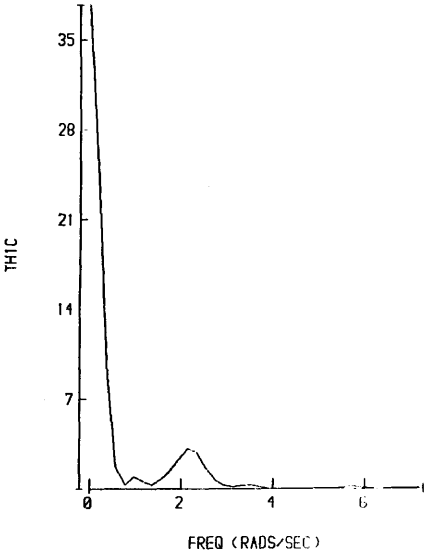
Figure 2.5

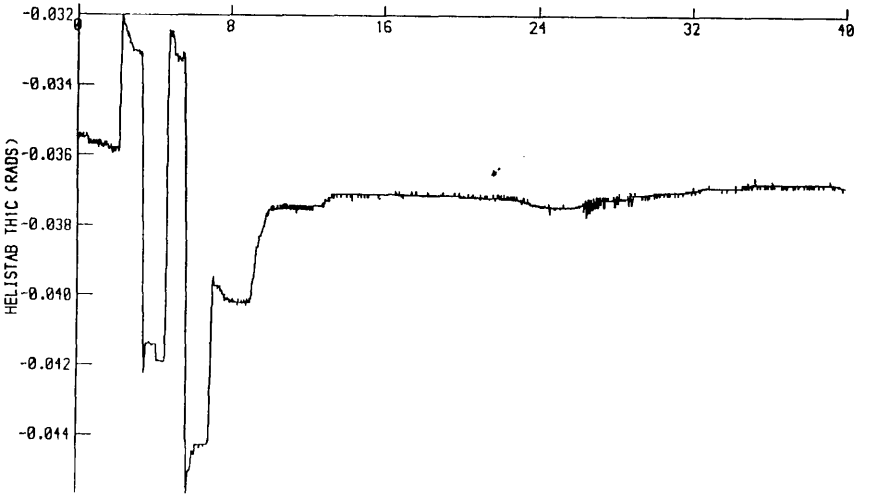
Time histories and auto–spectra of control inputs applied by pilot during flight 190/12.

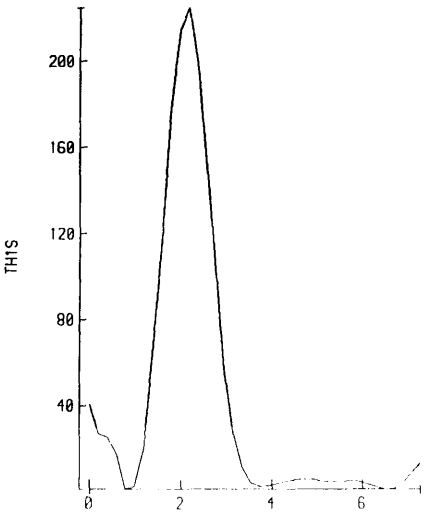
TH1S - longitudinal cyclic input
TH1C - lateral cyclic input
TH0E - collective input
TH0TRE - pedals











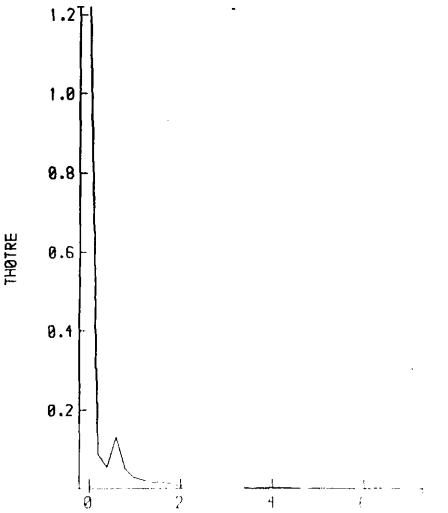
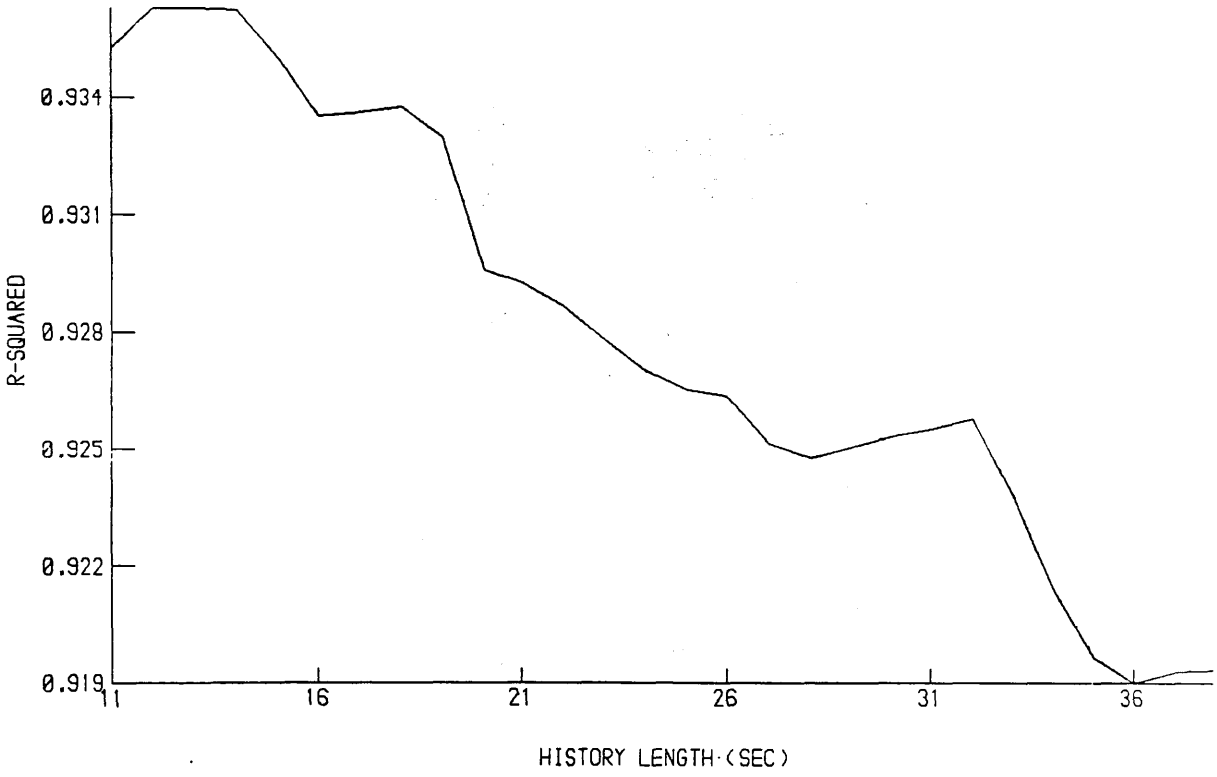


Figure 2.6

Variation of the Pitching Moment Equation squared–correlation coefficient, R^2 and F_{total} with the length of test record used for flight 190/12.



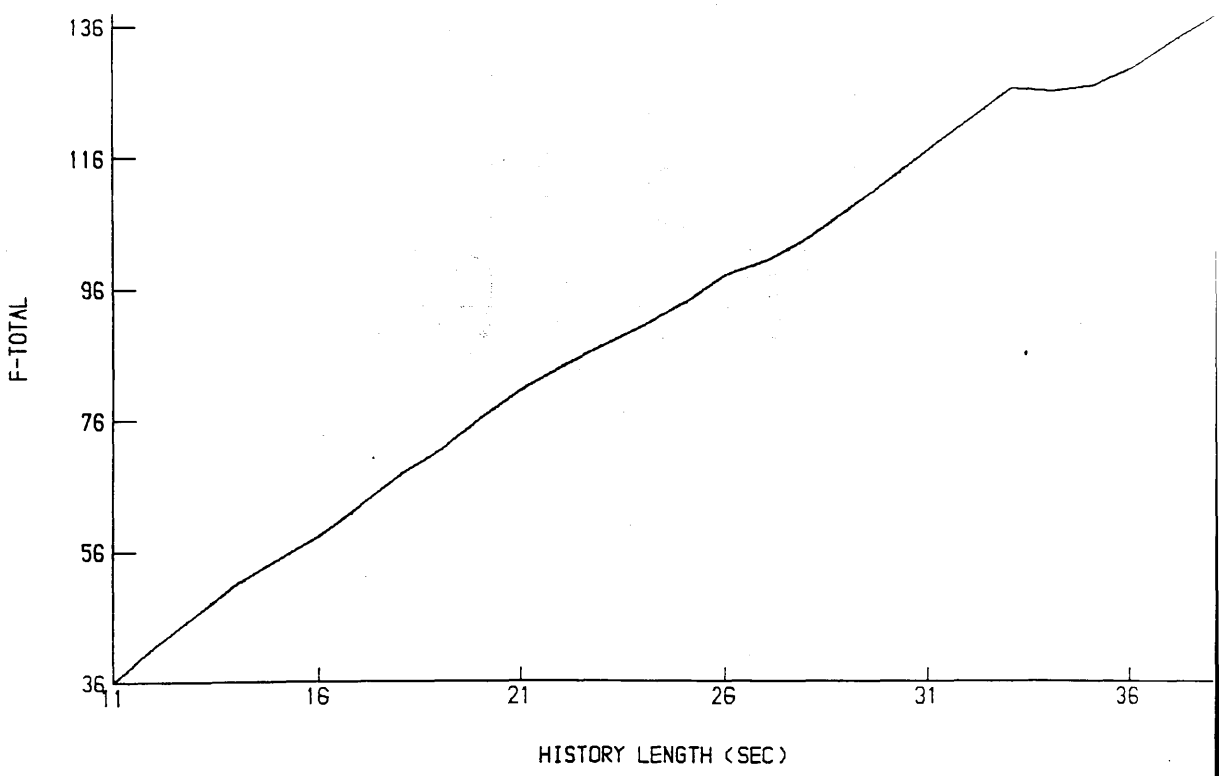


Figure 2.7

Coherency functions of the 8 helicopter states to the longitudinal cyclic input for the full 40 second test record and for the first 30 seconds of the record (flight 190/12).

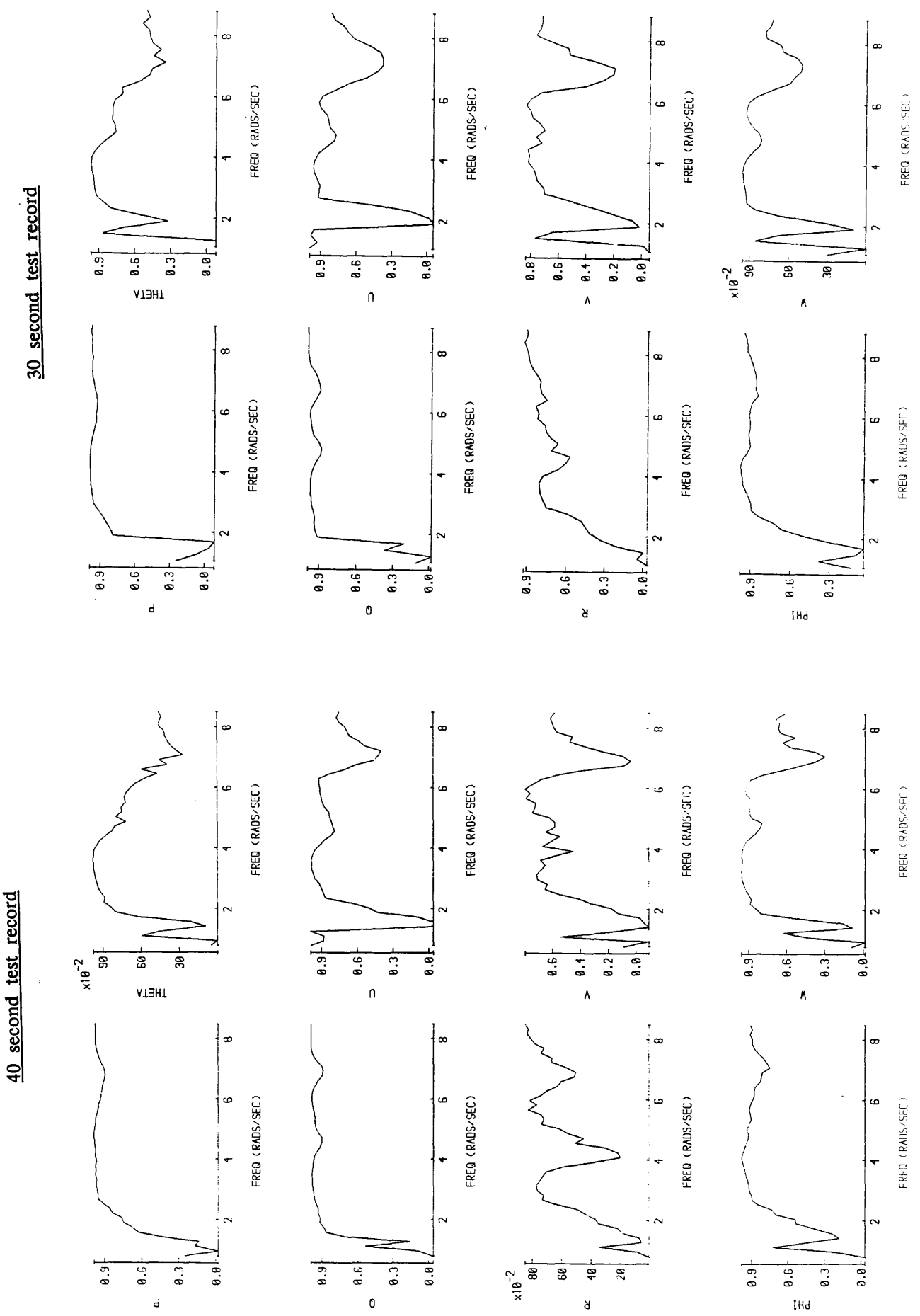


Figure 2.8

Variation of the estimates of the parameters of the Pitching Moment Equation with the length of test record used for flight 190/12.

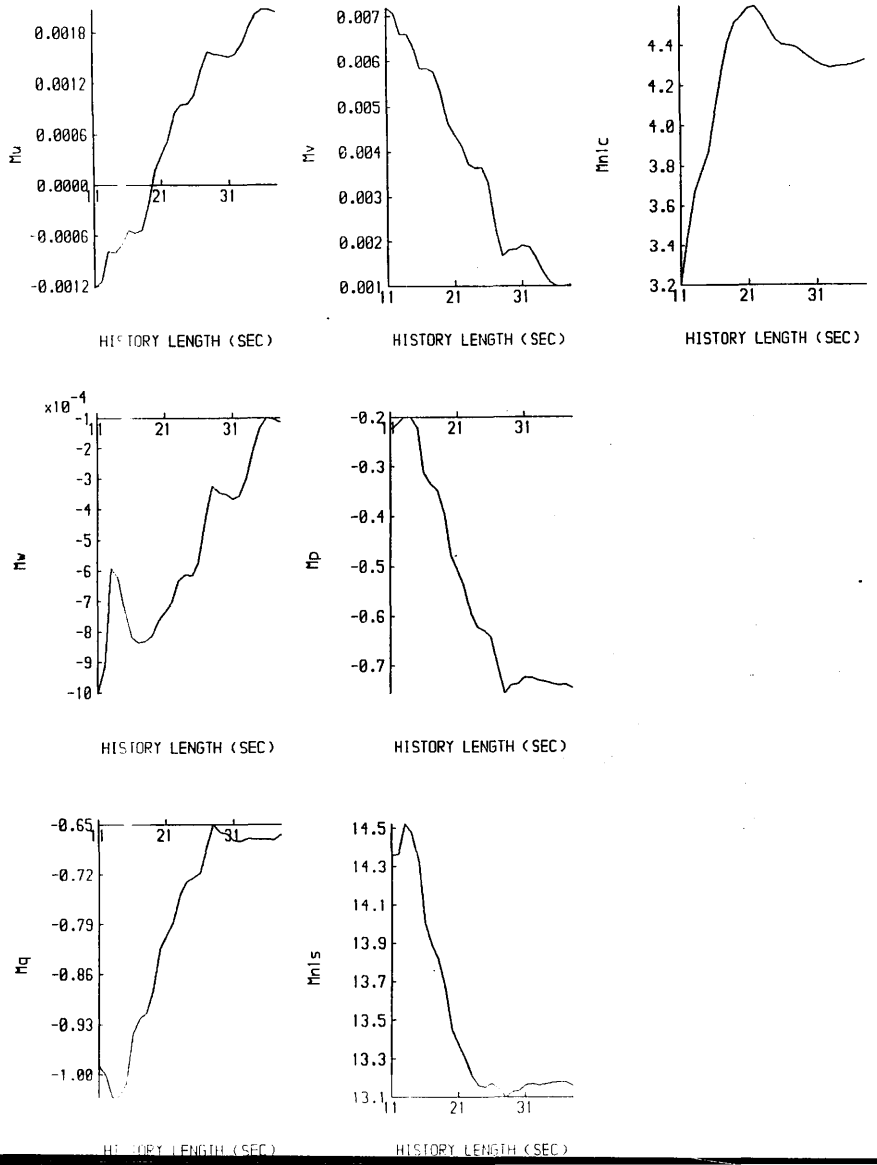


Figure 2.9

Variation of the Pitching Moment Equation parameter standard deviations with the length of test record used (flight 190/12).

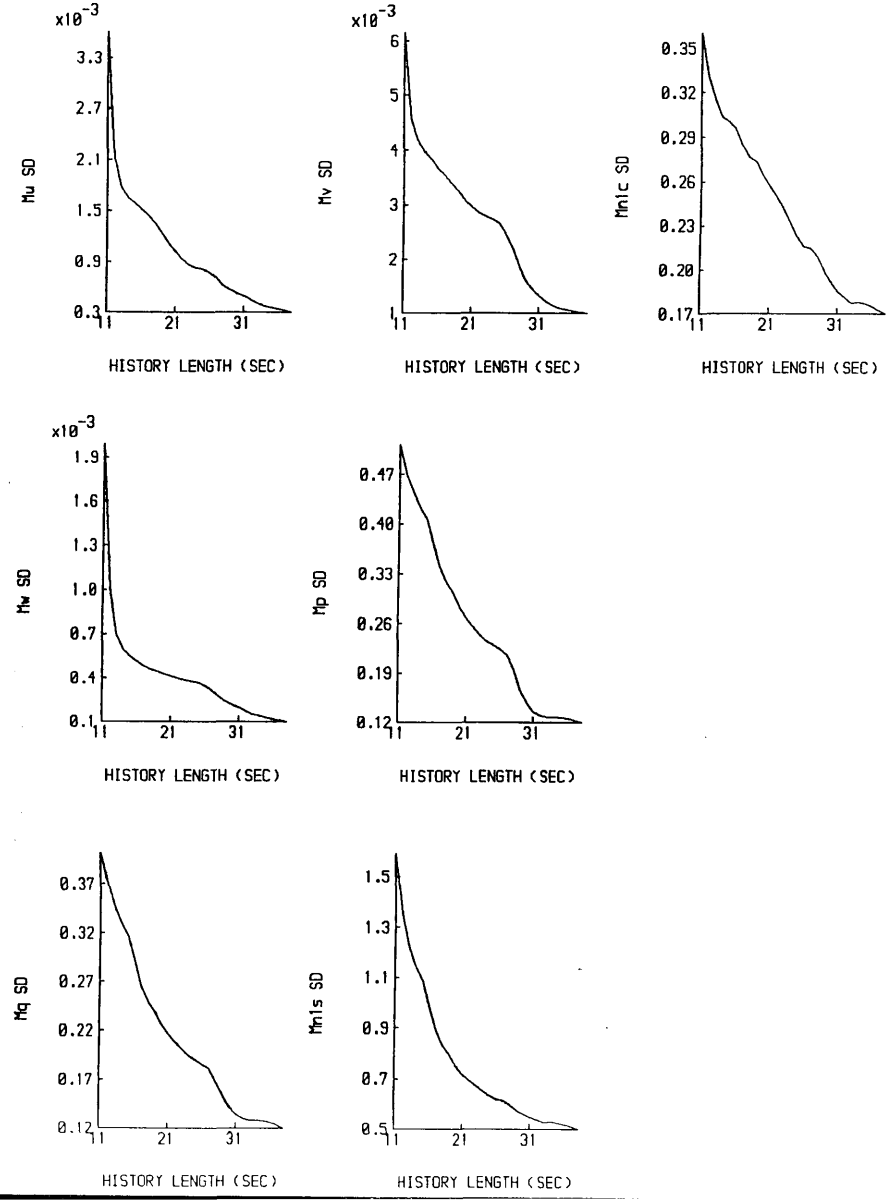


Figure 2.4 shows the auto–spectrum of an ideal 3211, and it can be seen that it has a large peak below 1 rad/sec. Hence, this input will greatly excite the unstable resonance of the Lynx, so leading to short test records.

In conclusion, the 3211 is an unsuitable input for use with the Lynx.

Double–Doublet Input

In contrast to the 3211, the double–doublet gave long test records – typically 40 seconds of data were obtained before the pilot was forced to recover control of the aircraft.

Three flight trials were obtained using double–doublets, and these gave results which were in extremely close agreement. Figure 2.5 shows the double–doublet input applied by the pilot in flight 190/12. It can be seen that the amplitude of the steps in the input vary, and that the timing is not precise. However, due to the robustness of this input, it can be seen that the auto–spectrum is still very similar to that of an ideal double–doublet.

Figure 2.6 shows the variation of the squared–correlation coefficient, R^2 and the F_{total} coefficient with the length of the test record used. The coherency functions [6] between the eight measured states and the longitudinal cyclic input are given in figure 2.7.

From these results it was decided that the first 32 seconds of the test records could be taken as linear. Various factors were taken into account.

Firstly, it can be seen that there is a large drop in R^2 after 32 seconds, and this corresponds to a noticeable peak in F_{total}. This could be due to non–linearities becoming significant for records longer than 32 seconds. The coherency function helps to confirm this possibility. For the full 40 seconds of the test record, the coherence is high (around 0.9) for all of the states except for yaw rate, r and lateral velocity, v. These two states have coherences of around 0.6, indicating significant non–linearities.

Moving to the coherence of the first 30 seconds of the record, it can be seen that the coherence of r and v rises to around 0.8, which can be taken as linear for most purposes. The coherences therefore indicate that significant non– linearities are excluded when going from 40 seconds to 30 seconds of the test record, in agreement with R^2 and F_{total}. Hence the decision was made to use the first 32 seconds of the test record.

It should be noted that several interesting numerical effects [7] can be seen in the coherences shown in figure 2.7. Each one of the coherence functions has troughs at around 0.3, 1, 4, and 7 rads/sec.

Firstly, recall that the unstable resonance of the Lynx occurs at around 0.3 rads/sec, and that there is a rapid phase change in the transfer– functions at this point. This rapid phase change introduces a large bias into the coherence at this frequency, and so produces the trough at around 0.3 rads/sec.

The remaining troughs in the coherence can be understood by considering the auto– spectrum of the double– doublet input used (see figure 2.5). It can be seen that the auto– spectrum is very low at around 1, 4, and 7 rads/sec. Hence, the response of the states will be low at these frequencies, and so the signal– to– noise ratio will also be low. The coherence will therefore be strongly biased at these frequencies. Notice that there is also a slight shift in the frequencies of these troughs for shorter test records, due to insufficient frequency resolution.

Lastly, notice that some of the coherency functions take values below zero at certain frequencies. This is purely a numerical artifact, and should be taken as a coherence of zero – the coherence function should lie strictly between 0 and 1.

It is therefore the author's opinion that due to these effects the coherency function should be used with some care, and coherence results interpreted with caution.

Finally, figure 2.8 shows the variation of the parameter estimates with the length of record used. It can be seen that the estimates of M_q, M_p, $M_{\eta 1s}$, and $M_{\eta 1c}$ appear to have converged after about 28 seconds. The estimates of M_u, M_v, and M_w also appear to reach a plateau after 28 seconds. However, they then start to vary rapidly again after about 32 seconds, corresponding to non-linearities becoming significant. In the theoretical model, 28 seconds is the length of approximately four cycles of the slowest mode of the Lynx. Intuitively, it is perhaps reasonable to expect that a few cycles of this mode will be needed before its parameters can be properly identified.

From these results, it can be seen that unless an input provides fairly long linear test records (greater than about 28 seconds) then the parameter estimates will not converge, and poor estimates will result. Long records are also necessary to give efficient estimation i.e. with the minimum parameter variance (given by the dispersion matrix). Figure 2.9 shows the variation of the standard-deviations of the parameter estimates with the length of test record. It can be seen that these appear to have converged for M_q, M_p, $M_{\eta 1s}$, and $M_{\eta 1c}$, but not for M_u, M_v, and M_w. Hence, still longer records would be desirable.

In conclusion, the double-doublet has given good identification results, and appears to be a useful identification input.

Doublet Input

Three flight trials were obtained using doublets. The longest doublet test run gave 38 seconds of data before the pilot recovered control of the aircraft. However, the other doublet test records were typically 10 to 20 seconds long.

It was decided to look at the long 38 second run first of all – flight 183/24. The input applied by the pilot in this run is shown in figure 2.10. It can be seen that the auto-spectrum of this input is similar to that of the ideal doublet input.

Figure 2.10

Time histories and auto-spectra of control inputs applied by pilot during flight 183/24.

TH1S - longitudinal cyclic input
TH1C - lateral cyclic input
TH0E - collective input
TH0TRE - pedals

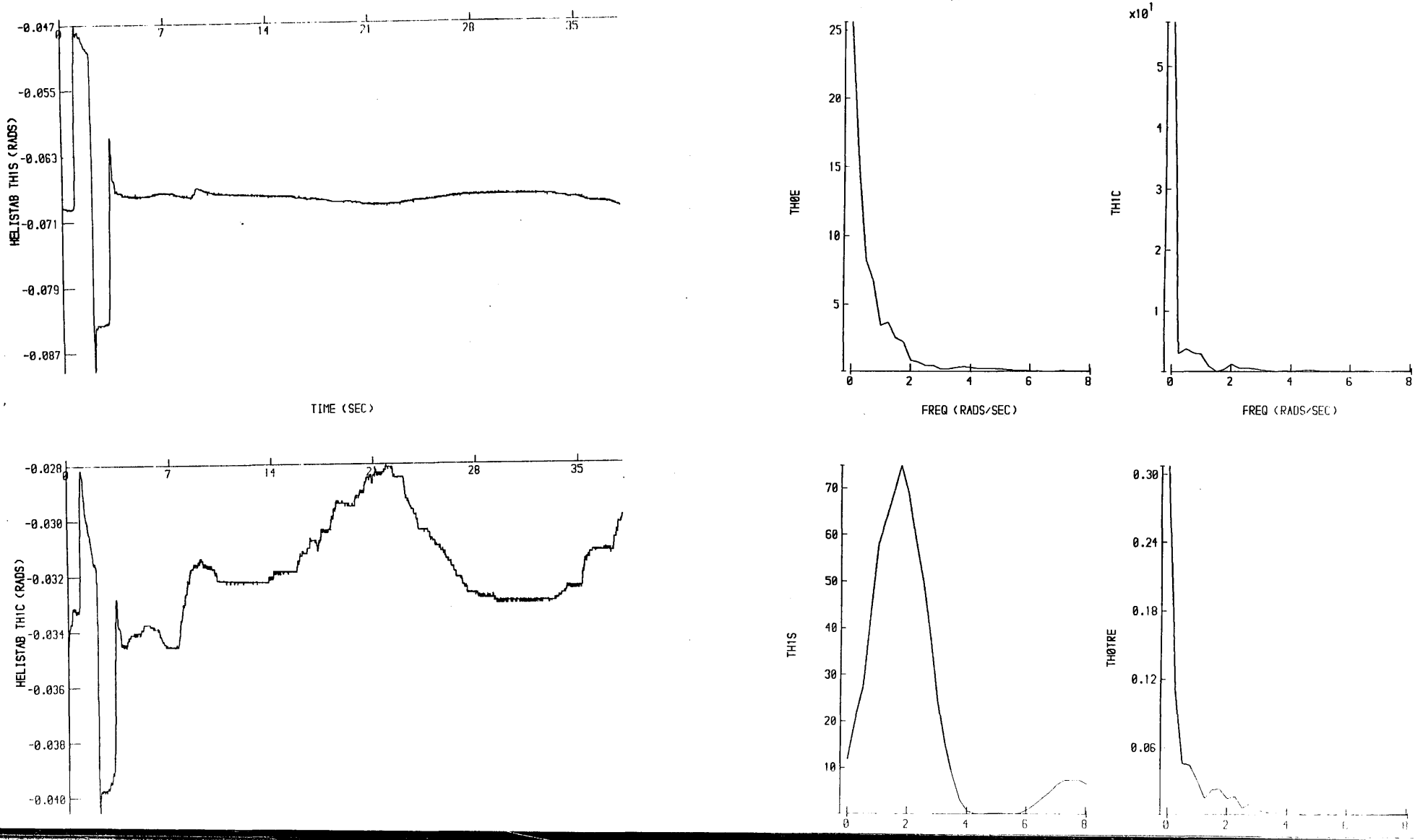


Figure 2.11

Variation of the Pitching Moment Equation squared–correlation coefficient, R^2 and F_{total} with the length of test record used for flight 183/24.

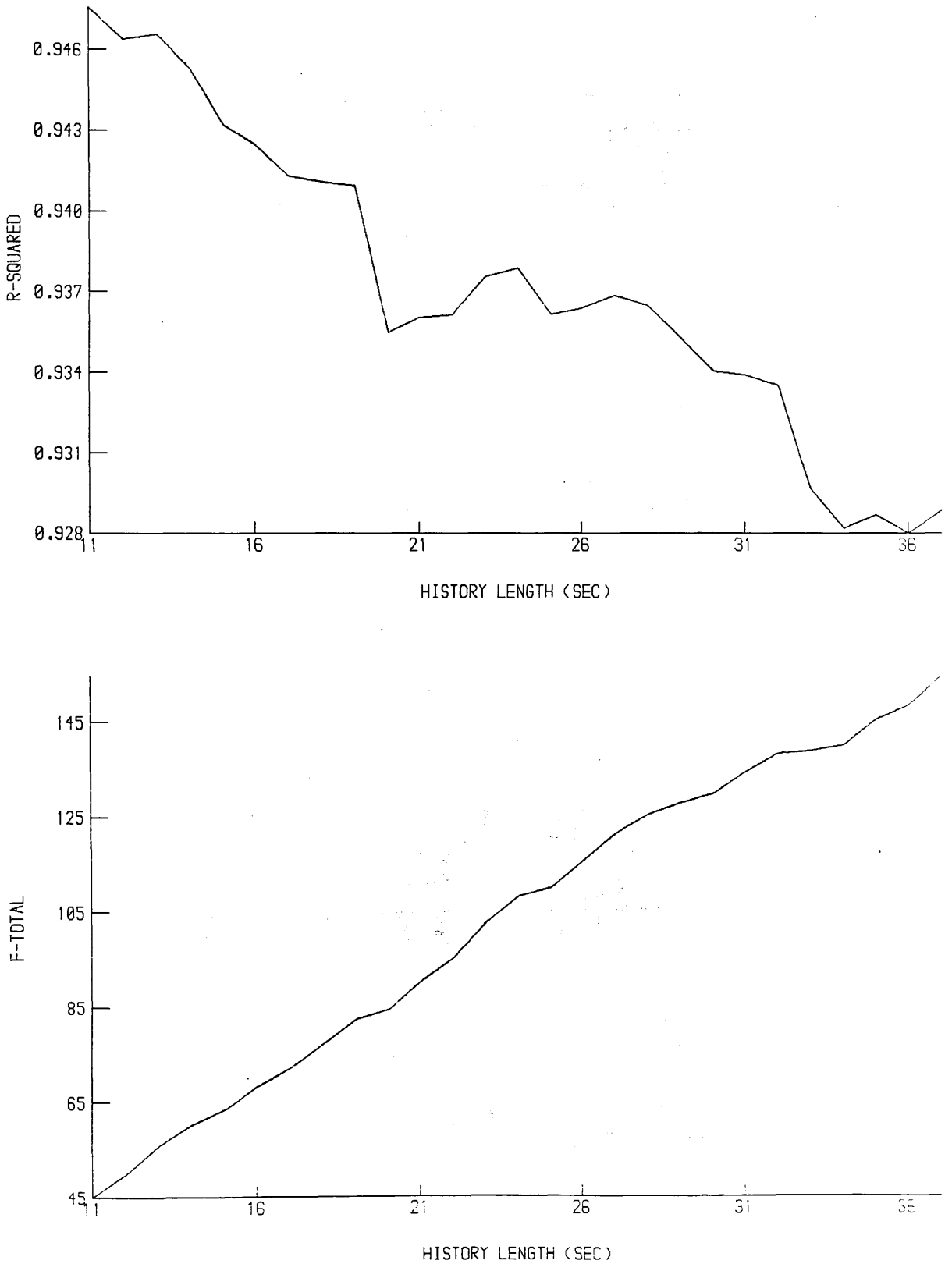


Figure 2.12

Coherency functions of the 8 helicopter states to the longitudinal cyclic input for the full 38 second test record and for the first 25 seconds of the record (flight 183/24).

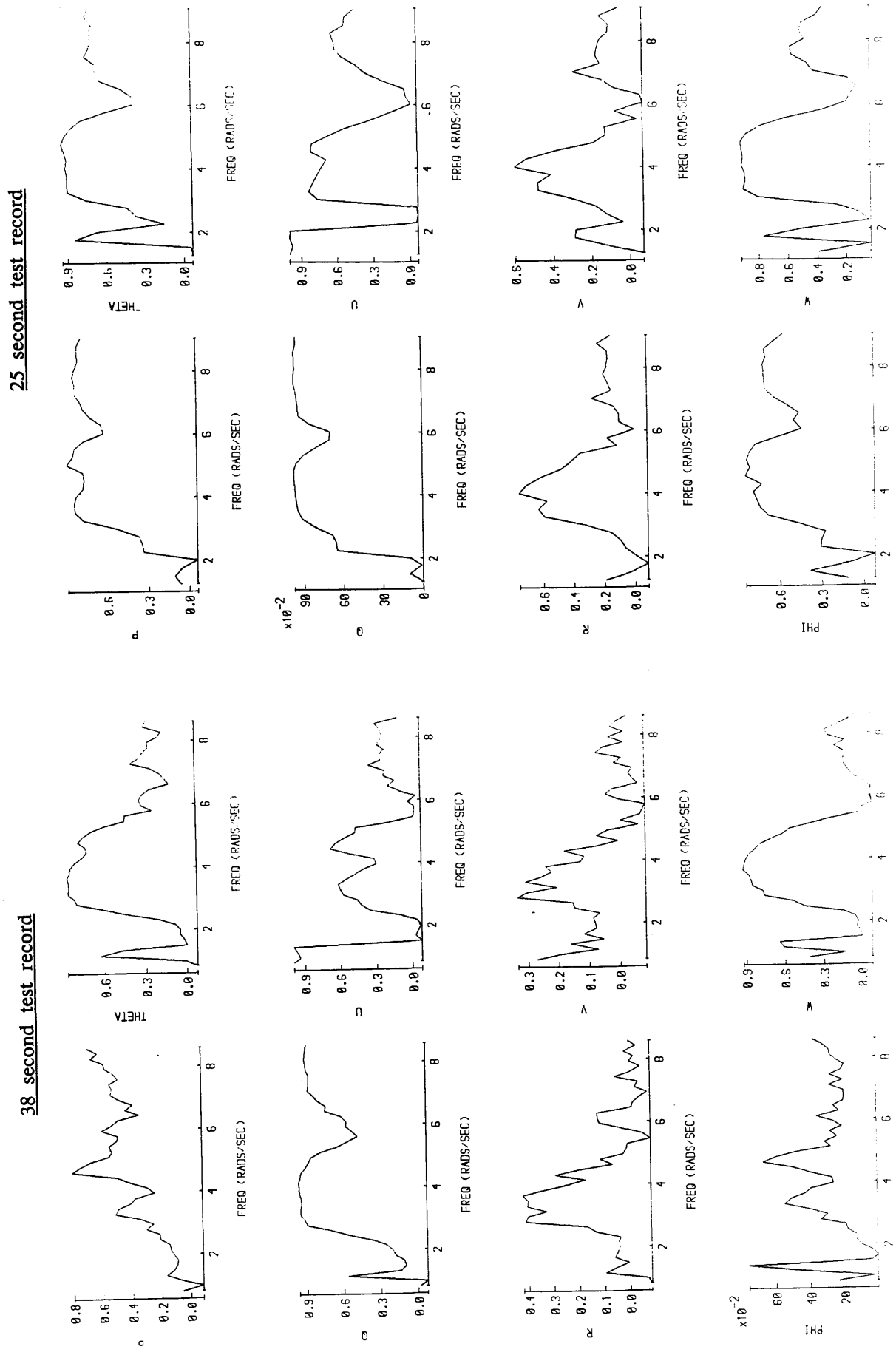


Figure 2.13

Variation of the estimates of the parameters of the Pitching Moment Equation with the length of test record used for flight 183/24.

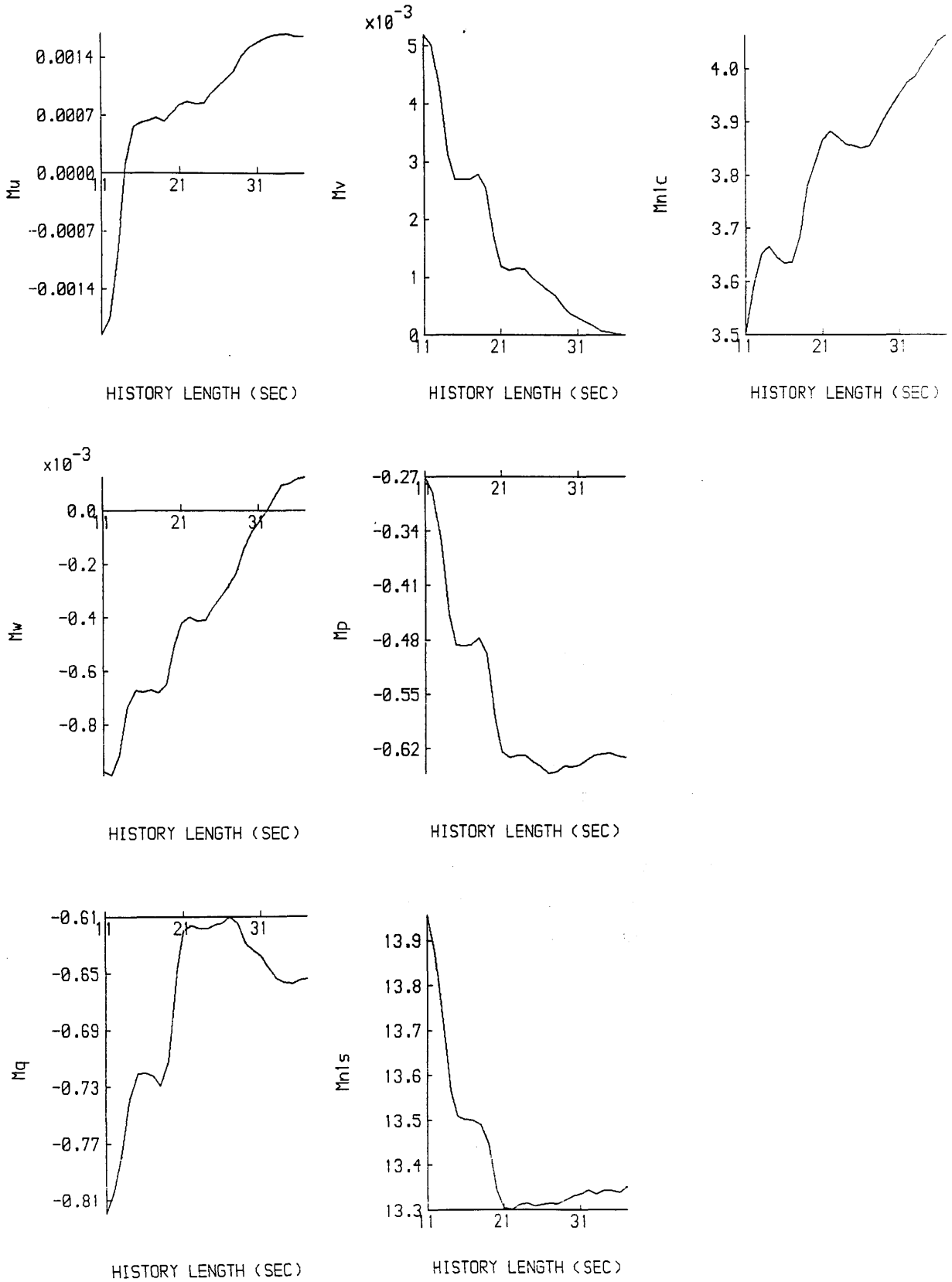


Figure 2.11 shows the variation of the squared–correlation coefficient, R^2 and the F_{total} coefficient with the length of test record used. The coherency functions are given in figure 2.12 for 38 seconds and 25 seconds of the test record. It can be seen that there are no significant peaks or troughs in R^2 or F_{total} which would indicate the onset of non–linearities. However, it can be seen that the coherence for the full 38 second test record is very low for most of the states, suggesting that serious non–linearities are present.

The coherence for the first 25 seconds of the record is, however, considerably better. The coherence is generally around 0.9 for all of the states except for yaw rate, r and lateral velocity, v, which have coherences of around 0.2. Hence, r and v are still very non–linear. Several large troughs are also present in all of the coherences, in particular at around 0.3, 1.25, and 5 rads/sec. These troughs are shifted to around 1, 2.25, 6 rads/sec for the 25 second records, due to insufficient frequency resolution with this shorter time history.

The trough at around 0.3 rads/sec is produced by the rapid phase change in the transfer functions at this frequency due to the unstable resonance of the Lynx. While the trough at 5 rads/sec results from poor excitation by the input at this frequency (see figure 2.10). These effects are similar to those observed for the double–doublet.

The remaining trough, at 1.25 rads/sec, corresponds to a peak in the doublet's auto–spectrum, and appears to represent at genuine non–linearity. The peak in the doublet's spectrum is lower than for the double–doublet, but is much wider and extends into the frequencies below 1 rads/sec. The doublet therefore excites the Lynx's unstable resonance more than the double–doublet.

Hence, for 25 seconds of the available test record the responses still contain significant non–linearities. It has been found that these non–linearities remain present even when only 15 seconds of the test record are used.

Figure 2.14

Time histories and auto–spectra of control inputs applied by pilot during flight 190/10.

TH1S - longitudinal cyclic input
TH1C - lateral cyclic input
TH0E - collective input
TH0TRE - pedals

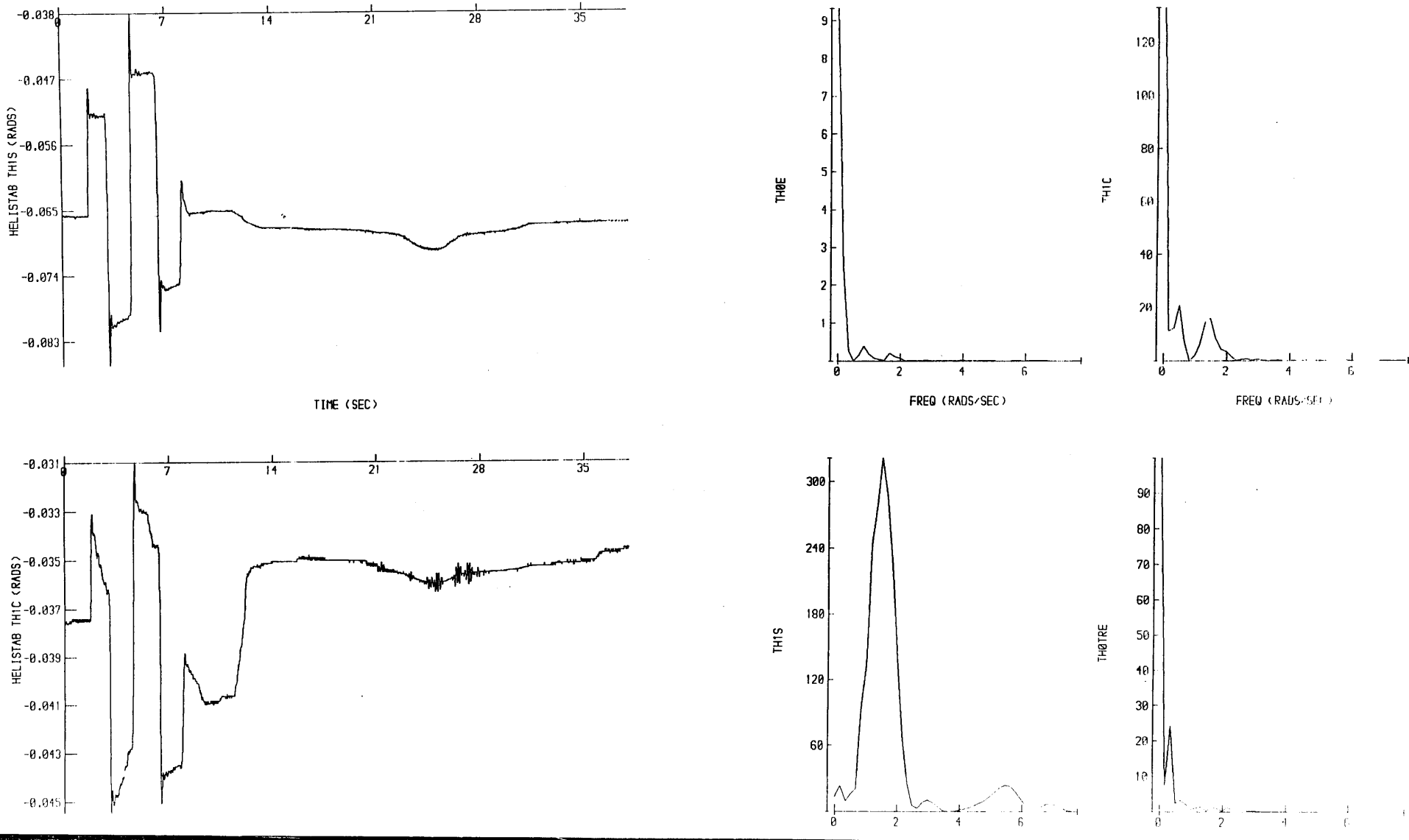
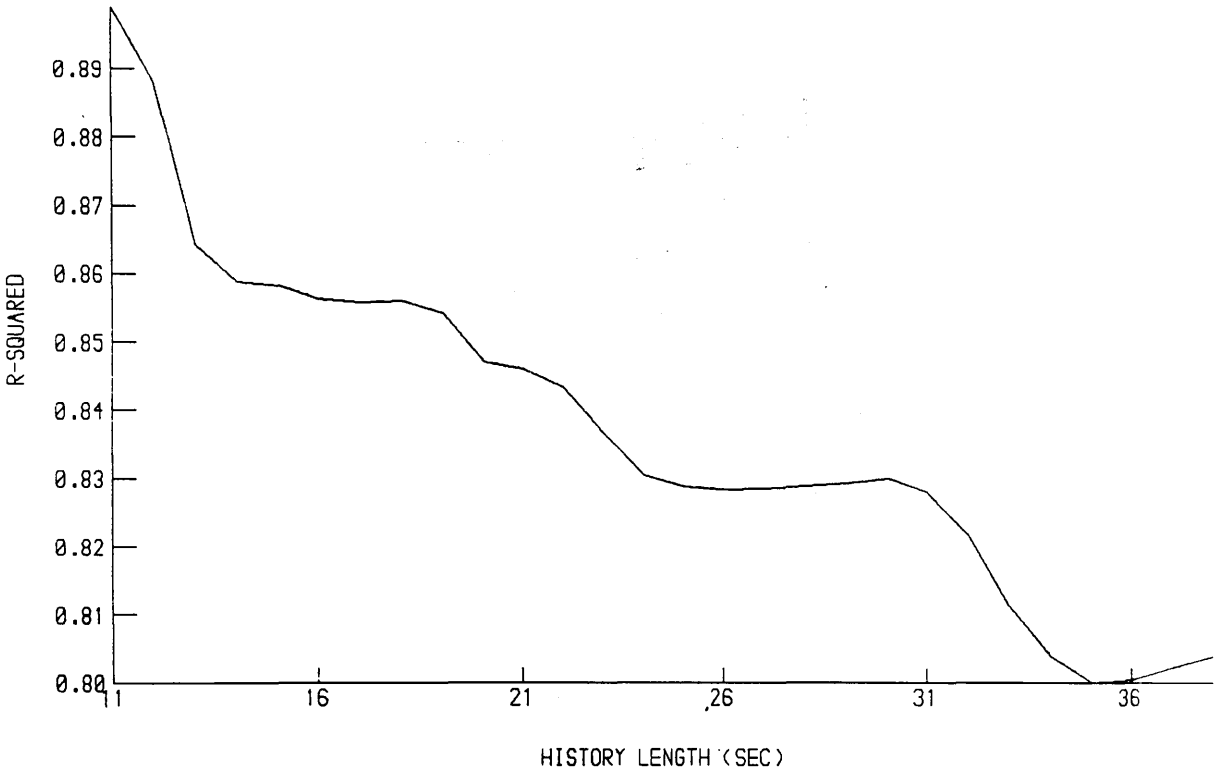


Figure 2.15

Variation of the Pitching Moment Equation squared–correlation coefficient, R^2 and F_{total} with the length of test record used for flight 190/10.



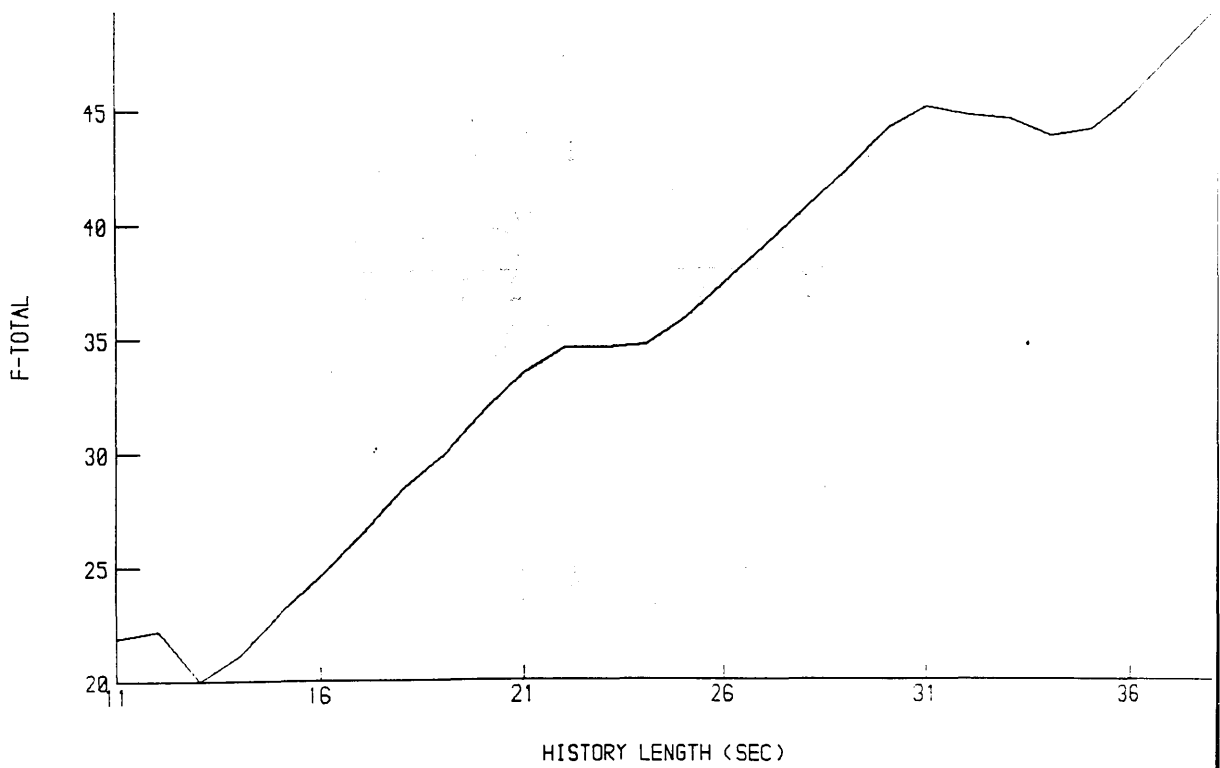


Figure 2.16

Coherency functions of the 8 helicopter states to the longitudinal cyclic input for the full 38 second test record and for the first 20 seconds of the record (flight 190/10).

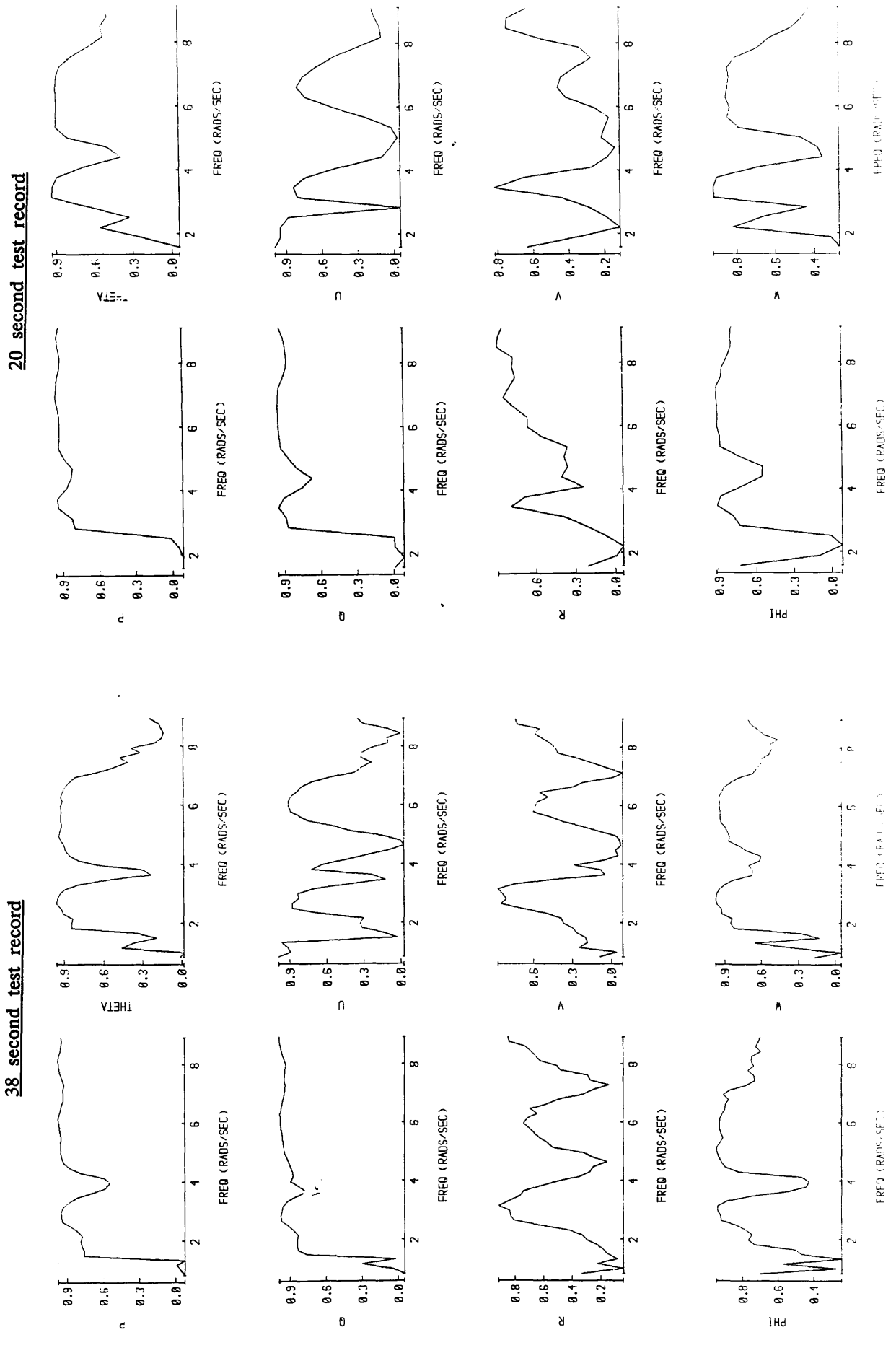


Figure 2.17

Variation of the estimates of the parameters of the Pitching Moment Equation with the length of test record used for flight 190/10.

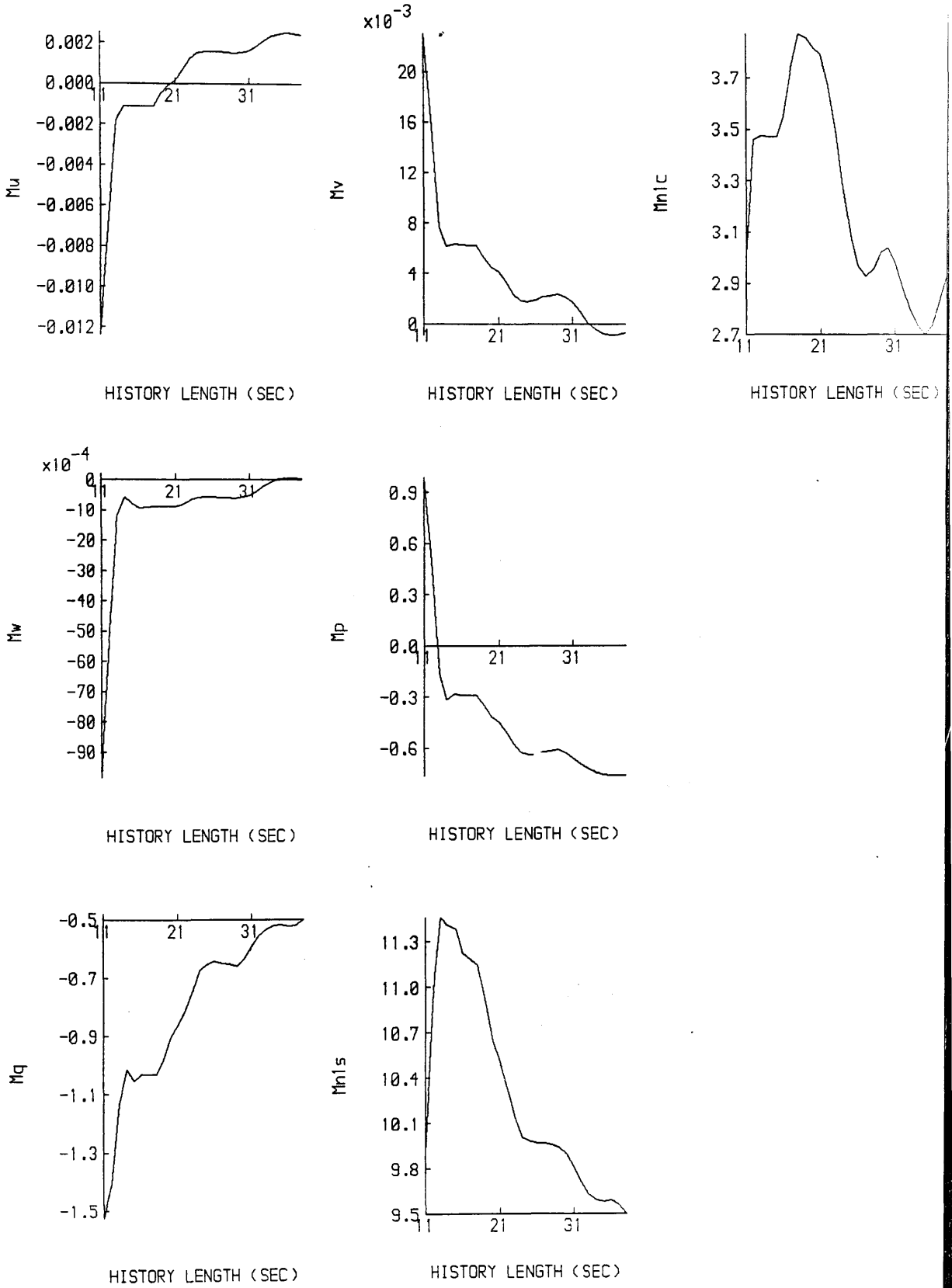


Figure 2.13 shows the variation of the parameter estimates with the length of test record used. It can be seen that there are several plateaus, but that the estimates have not converged satisfactorily. In particular, below 15 seconds the estimates are varying rapidly. This doublet test record therefore appears to be unsuitable for identification, and gives only poor estimates.

Other test runs using doublets also typically produced test records with only 10 to 20 seconds of data before the pilot recovered control of the aircraft.

In conclusion, doublets do not appear to give linear test records which are long enough to produce good parameter estimates. Moreover, the doublet flight trials resulted in test records varying in length from 38 seconds to around 10 seconds. It therefore appears to be difficult to obtain consistent results using doublet inputs, which suggests that this input is not sufficiently robust.

1221 Input

Finally, several test runs were performed using the 1221 input. These runs typically gave 32 seconds of data before the pilot recovered control of the aircraft, with the longest run being 38 seconds. Figure 2.14 shows the input applied by the pilot in this 38 second run (flight 190/10).

The variation of the squared correlation coefficient, R^2 and the F_{total} coefficient with the length of the test record used are shown in figure 2.15. The coherency functions are given in figure 2.16 for 38 seconds and 20 seconds of the test record.

It can be seen that there are several sharp increases in R^2 as the length of the test record used is reduced, with corresponding peaks in F_{total}. The results discussed above for other test inputs suggest that these changes in R^2 and F_{total} may be due to non-linearities being excluded from the identification as the record length falls.

For the full 38 second record, the coherency functions are generally high (around 0.9). However, several troughs can be observed.

The troughs at around 4 and 7 rads/sec can be attributed to the low power of the input at these frequencies producing a poor signal– to– noise ratio, and hence biased coherence.

The trough at approximately 0.75 rads/sec may be the result of either, or both, of two effects. The input auto– spectrum is low at this frequency, and so the trough may be caused by a poor signal– to– noise ratio as for the other troughs. However, recall that a rapid phase change occurs in the system transfer functions at around 0.3 rad/sec due to the unstable resonance of the Lynx. Hence, the coherence has a trough at 0.3 rads/sec, and this trough may be shifted to around 0.75 rads/sec on the coherence plots due to insufficient frequency resolution. In either case the effect is not due to non– linearities.

The trough at around 1.5 rads/sec, however, corresponds to a large peak in the input auto– spectrum, and appears to indicate a genuine non– linearity.

This peak in the auto– spectrum of the 1221 input is of similar magnitude to the peak in the double– doublet's auto– spectrum. However, the 1221 peak occurs at a significantly lower frequency, and hence will excite the system to a greater extent.

When only the first 20 seconds of the test record are used, it can be seen that the coherence is little different from that of the full 38 second record, except that the frequency resolution is somewhat poorer.

Hence, the coherence functions display no evidence of non– linearities being excluded at the record lengths corresponding to the sharp increases in R^2. Rather, the coherence suggests that non– linearities are still present in the shorter test records.

Perhaps the sharp increases in R^2 do correspond to a drop in the magnitude of the non– linearities. However, the non– linearities appear to remain significant, and so are still registered by the coherency functions. which may have insufficient precision to detect the reduction in the non– linearities.

Finally, figure 2.17 gives the variation of the parameter estimates with the length of the test record used. It can be seen that the estimates do not converge, and that their rate of change corresponds to the rate of change in R^2.

Similar identification results were obtained for the other 1221 test runs.

2.3.3 Conclusions from Flight Trials of Lynx Test Inputs

In conclusion, it appears that the 3211, the 1221 and the doublet suffer from a similar type of problem: the inputs' auto–spectra contain too much power at low frequencies in the vicinity of the Lynx's unstable resonance, resulting in non–linear test records. This is particulary pronounced in the case of the 3211 input.

The double–doublet therefore appears to be the best input, giving fairly long linear test records, and being reasonably robust.

Hence, these flight trials have demonstrated the effectiveness of the input design method presented, despite its simplicity. The importance of obtaining sufficiently long test records has been shown, both in terms of the convergence of the parameter estimates and in terms of convergence of the parameter variances to give efficient estimation.

It is interesting to take note of the results concerning the 1221 and double–doublet inputs. These inputs complement each other, as discussed in section 2.3.1 above – the 1221 theoretically produces less excitation of the unstable resonance, while the double–doublet is more robust to errors in the theoretical model, and to pilot errors in applying the input. Since the double–doublet has been found to give the best results, it appears that robustness has been the deciding factor. When designing inputs, the degree of robustness required should therefore be taken carefully into account.

2.4 Identification Results for Double– Doublet Input

2.4.1 Introduction

As discussed in section 2.3.2 above, the most satisfactory flight trials were obtained using double– doublet inputs. More detailed double– doublet identification results are therefore presented below.

The parameters of the pitching moment equation described in section 2.3.2 were estimated. Results from both equation– error and output– error identification are presented.

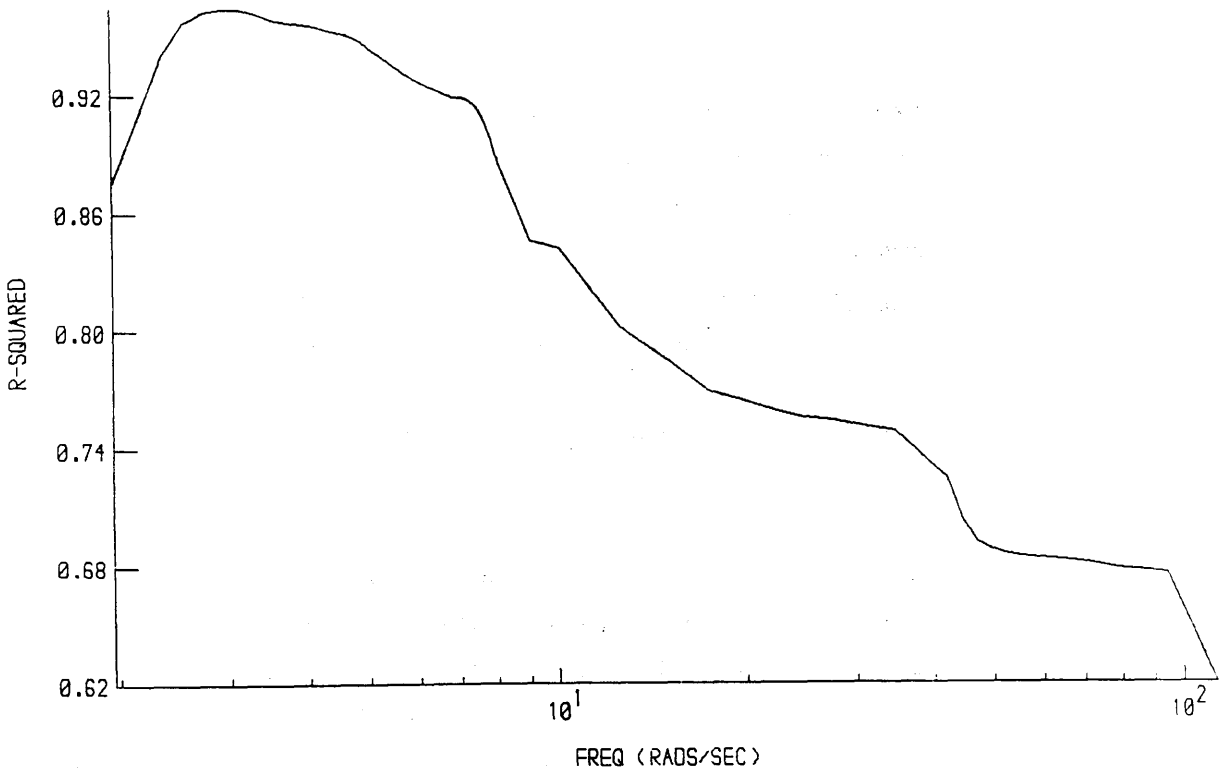
Before estimates of the parameters of the pitching moment equation could be obtained, however, for each test record it was necessary to resolve certain issues.

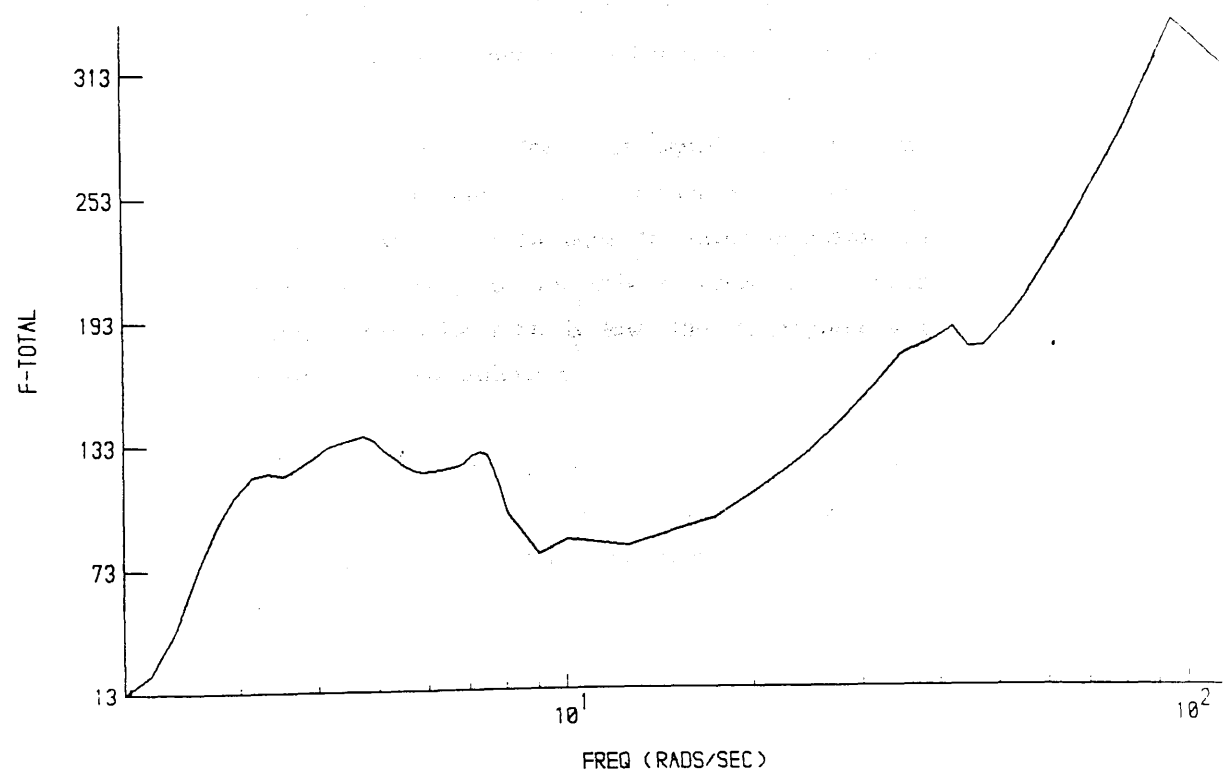
a) How much of the available test record can be used ?
The response will become significantly non-linear at some point, and this point needs to be determined.

b) What range of frequencies should be used in the identification ?
The model being identified ignores the dynamics of the helicopter's rotor. These dynamics tend to have short time constants compared with the dynamics of the fuselage. Hence, in the frequency-domain the rotor dynamics tend to dominate the response at higher frequencies, while the fuselage dynamics tend to dominate the low frequency response. The model used is therefore only valid at low frequencies. Hence, it is necessary to determine the range of valid frequencies that can be used.

Figure 2.18

Variation of the Pitching Moment Equation squared–correlation coefficient, R^2 and F_{total} with the upper frequency used in the identification. The lower frequency was chosen to exclude dc, and depends on the frequency resolution available i.e on the length of test record – a 32 second record was used, giving a lower frequency of 0.196 rads/sec. (flight 190/12).





2.4.2 Equation–Error Identification Results

For flight 190/12 it was decided to use the first 32 seconds of the test records, for the reasons explained in section 2.3.2.

The next step was therefore to determine the frequency range to be used. At lower frequencies, dc is excluded from the identification in order to deal with any biases in the measurements. It then remains to decide what range of higher frequencies should be used. Figure 2.18 shows the variation of the squared–correlation coefficient, R^2 and F_{total} with the frequency range.

Several interesting features can be seen. Firstly, as the frequency range is reduced there is a sharp increase in R^2, and a peak in F_{total} at about 40 rads/sec. This is probably due to a rotor mode being excluded from the identification, resulting in a better model fit. The theoretical model used, HELISTAB, is in good agreement with this, and predicts a rotor coning mode at 35 rads/sec.

As the frequency range is reduced further, there is a second sharp increase in R^2 and a peak in F_{total} at about 7 rads/sec. Again this is probably due to a rotor mode being excluded, and in particular HELISTAB predicts a longitudinal flapping mode at 10.41 rads/sec.

HELISTAB predicts no rotor modes below this frequency, and it can be seen that there are no sharp increases in R^2 at lower frequencies, and that F_{total} is fairly level, except for a fall in R^2 and F_{total} below about 2.5 rads/sec.

This fall can be attributed to the poor signal–to–noise ratio around 1 rad/sec. At this frequency the input auto–spectrum is very low, and hence the system's response is also low. Now an equation–error estimator was used, and this assumes that there is no noise on the driving terms of the Pitching Moment Equation. If the signal–to–noise ratio is low, this assumption will be invalid, and will result in poor parameter estimates.

Since the low signal–to–noise ratio is localised around 1 rad/sec, its effect will be small if a fairly large frequency range is used. However, when smaller frequency ranges are used its effect will become more significant, and so result in the fall observed in R^2 and F_{total}.

The input auto–spectrum is also low at other frequencies, in particular, at 4 and 7 rads/sec. As discussed in section 2.3.2, this has affected the coherence functions : the poor signal–to–noise ratios at 1, 4, and 7 rads/sec resulting in biased coherences at these frequencies. Now, the effect on identification at 1 rad/sec has been described above. At 4 and 7 rads/sec, however, R^2 and F_{total} do not appear to show any low auto–spectrum effects. It can also be seen that the bias in the coherence does not appear to be as large at 4 and 7 rads/sec as at 1 rad/sec. This suggests that the signal–to–noise ratio is not as poor at these frequencies, and hence the effect on identification will be less.

From these results, it was decided that a frequency range of 7 rads/sec should be used in the identification.

Notice that this is more than double the 3 rads/sec range of the Puma helicopter. This is reasonable, given the greater agility of the Lynx, and the differing roles that these helicopters are intended for. However, since no previous results were available for the Lynx, the 3 rads/sec range of the Puma was used when designing the Lynx inputs (see section 2.3.1). The inputs could therefore be improved upon in future flight trials by using the greater frequency range now known to be available for the Lynx.

The parameter estimates obtained are shown in table 2.2, together with the parameter values predicted by the theoretical model, HELISTAB.

Parameter	Estimate	HELISTAB
M_u	0.00155	0.00774
M_w	-0.00034	0.0098
M_q	-0.677	-2.382
M_v	0.0020	0.0027
M_p	-0.708	0.363
$M_{\eta 1s}$	13.13	28.15

Table 2.2 - *Equation-error estimates and HELISTAB values for the parameters of the pitching moment equation. (Lynx, 80 knots level flight)*

It can be seen that several of the parameter estimates differ significantly from the HELISTAB values. In particular, M_w and M_p are of different signs from HELISTAB, while M_q and $M_{\eta 1s}$ are underestimated. These results are discussed in more detail below.

2.4.3 Output–Error Identification Results

As in the equation–error case, the first 32 seconds of flight 190/12 were used for the output–error identification, with a frequency range of 7 rads/sec. A frequency–domain output–error algorithm developed by Black [4] was used, and the mathematical model given below in equation (2.15) formed the basis for the identification. This model is 5^{th} order, and contains the main parameters affecting the longitudinal dynamics. In particular, the pitching moment equation parameters are included in this model.

$$\begin{bmatrix} \dot{u}(t) \\ \dot{w}(t) \\ \dot{q}(t) \\ \dot{\theta}(t) \\ \dot{p}(t) \end{bmatrix} = \begin{bmatrix} X_u & X_w & 0 & X_\theta & 0 \\ Z_u & Z_w & Z_q & 0 & 0 \\ M_u & M_w & M_q & 0 & M_p \\ 0 & 0 & 1 & 0 & 0 \\ L_u & L_w & L_q & 0 & L_p \end{bmatrix} \begin{bmatrix} u(t) \\ w(t) \\ q(t) \\ \theta(t) \\ p(t) \end{bmatrix} + \begin{bmatrix} X_{\eta 1s} & 0 \\ Z_{\eta 1s} & 0 \\ M_{\eta 1s} & 0 \\ 0 & 0 \\ L_{\eta 1s} & L_{\eta 1c} \end{bmatrix} \begin{bmatrix} \eta_{1s}(t) \\ \eta_{1c}(t) \end{bmatrix} \qquad (2.15)$$

Table 2.3 shows the parameter estimates obtained, together with the theoretical values predicted by HELISTAB.

Parameter	Estimate	HELISTAB
X_u	0.0348	-0.033
X_w	0.110	0.043
X_θ	-33.53	-32.179
$X_{\eta 1s}$	-24.62	-24.43
Z_u	-0.0027	-0.0068
Z_w	-0.061	-0.838
Z_q	90.0	134.8
$Z_{\eta 1s}$	-114.8	-104.19
M_u	0.0031	0.00774
M_w	0.00065	0.0098
M_q	-1.076	-2.382
M_p	-0.85	0.363
$M_{\eta 1s}$	17.91	28.15
L_u	-0.0065	-0.12
L_w	0.00288	0.068
L_q	4.239	-1.731
L_p	-4.92	-10.9
$L_{\eta 1s}$	10.232	-26.54
$L_{\eta 1c}$	2.01	-150.51

Table 2.3- *Output-error estimates and HELISTAB values for the parameters of the 5^{th} order model. (Lynx, 80 knots level flight)*

Concentrating firstly on the pitching moment equation parameters, it can be seen that compared with the equation–error results, the output–error estimates are generally in closer agreement with HELISTAB. M_p is an exception to this, and is estimated with a different sign to the HELISTAB value. This parameter has an important effect on the degree of coupling between the pitch and roll rates, and so was investigated in more detail.

Figure 2.19

Comparison of response predicted using the 5th order identified model with measured response (flight 190/12, double–doublet input).

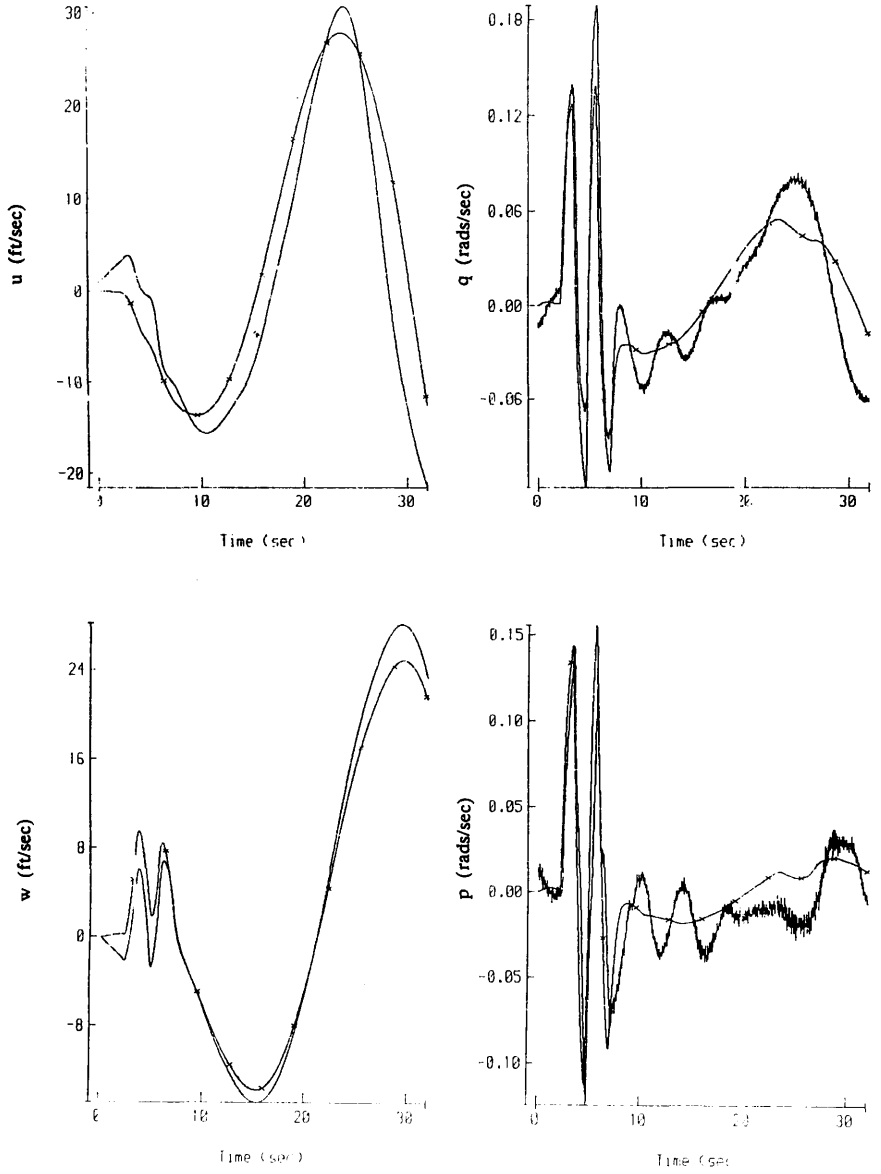


Figure 2.20

Comparison of response predicted using the 4th order identified model with measured response (flight 190/12, double–doublet input).

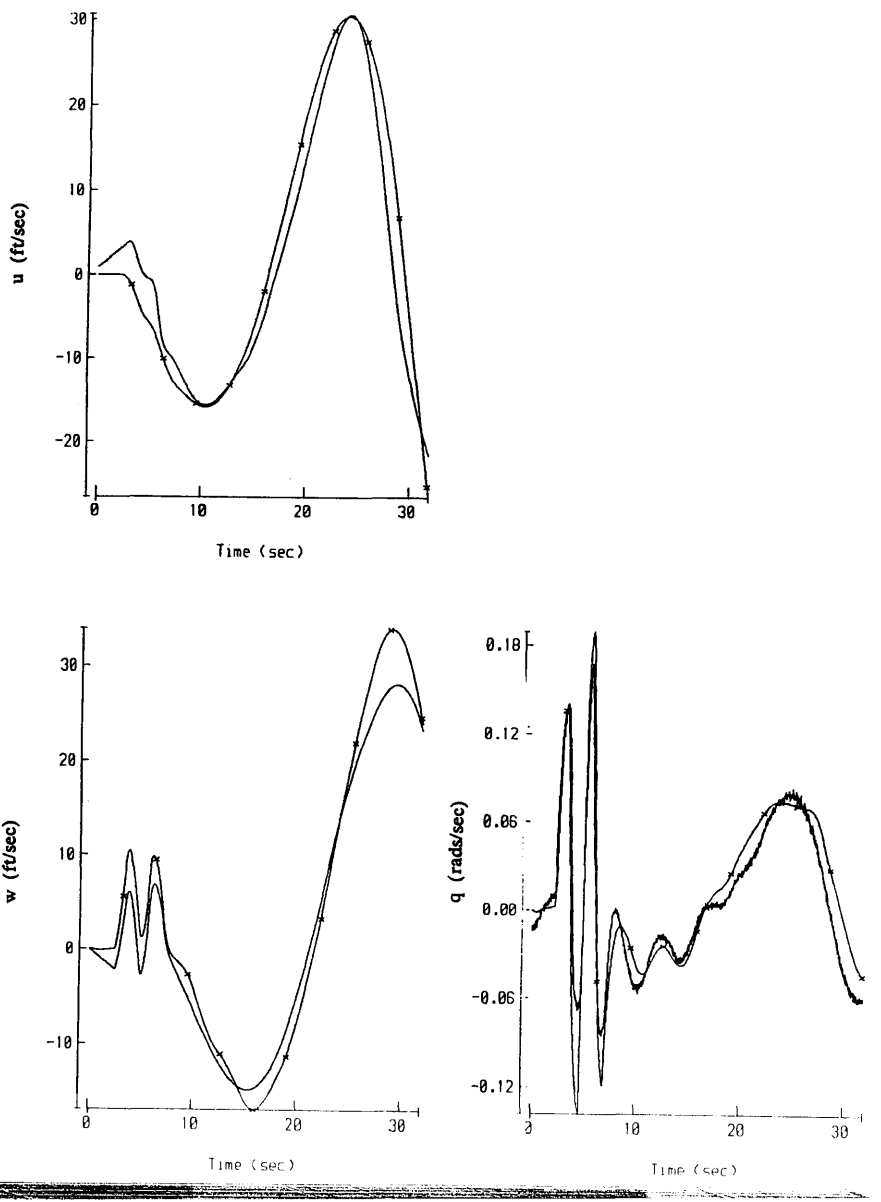


Figure 2.21

Comparison of response predicted using the 5th order identified model with measured response (flight 190/10, 1221 input).

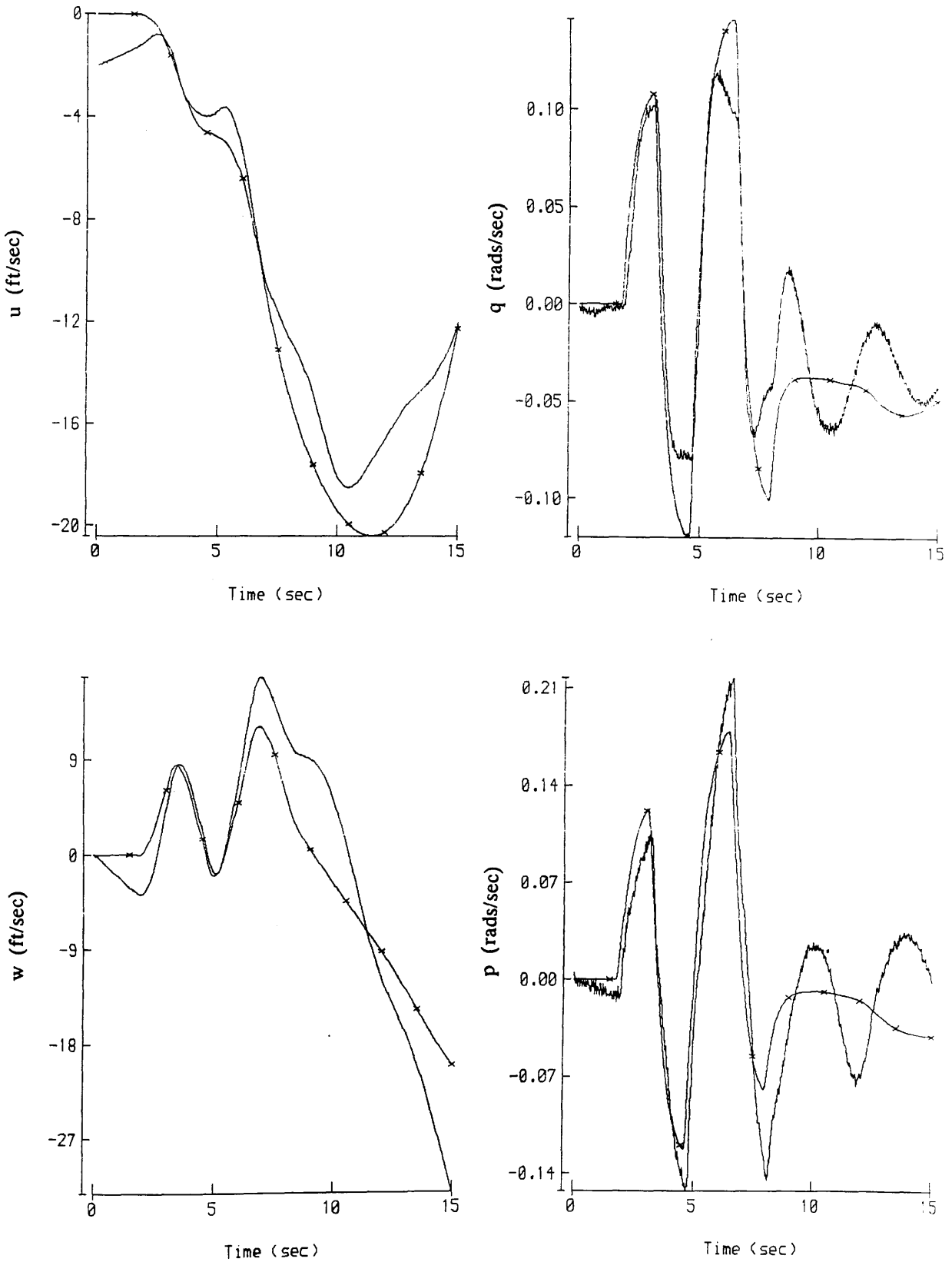


Figure 2.19 shows a time–domain comparison between the measured responses and those predicted using the identified 5^{th} order model. There is generally a reasonable agreement between these responses. However, this comparison highlights a significant oscillatory component in the measured responses for pitch rate, q and roll rate, p which is not present in the responses predicted by the identified model. To study this further, the 5^{th} order model was modified to use roll rate, p as a pseudo–input rather than using it as one of the model states. This gave the following 4^{th} order model.

$$\begin{bmatrix} \dot{u}(t) \\ \dot{w}(t) \\ \dot{q}(t) \\ \dot{\theta}(t) \end{bmatrix} = \begin{bmatrix} X_u & X_w & 0 & X_\theta \\ Z_u & Z_w & Z_q & 0 \\ M_u & M_w & M_q & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} u(t) \\ w(t) \\ q(t) \\ \theta(t) \end{bmatrix}$$

$$+ \begin{bmatrix} X_{\eta 1s} & 0 & 0 \\ Z_{\eta 1s} & 0 & 0 \\ M_{\eta 1s} & 0 & M_p \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \eta_{1s}(t) \\ \eta_{1c}(t) \\ p(t) \end{bmatrix} \qquad (2.16)$$

Measurements of η_{1s}, η_{1c}, and p were used to drive this model. Figure 2.20 shows a comparison between the measured responses and those predicted by this 4^{th} order model. It can be seen that the pitch rate, q predicted by the 4^{th} order model contains an oscillatory component similar to that present in the measured pitch rate. The oscillation therefore appears to originate in the roll–rate, p, and is present in the pitch rate due to the coupling between roll and pitch. The 5^{th} order model does not include this lateral oscillatory roll mode, since it concentrates on the longitudinal dynamics. Hence the poor estimate obtained for M_p.

Moving to those parameters not included in the pitching moment equation, several estimates agree closely with HELISTAB, while others are in much poorer agreement. In particular, the parameter estimates for the rolling–moment equation are significantly different from the HELISTAB values. This can be at least partially attributed to the difficulties described previously concerning unmodelled lateral dyamics.

Of the other parameters, the estimates of X_θ, $X_{\eta 1s}$, and $Z_{\eta 1s}$ are in good agreement with the theoretical values. While the estimates of X_u, X_w, and Z_w are a much poorer match. X_u, X_w, and Z_w are important parameters affecting the Lynx phugoid–like unstable mode. HELISTAB predicts that this mode has a frequency of 0.368 rads/sec, and time constant of 7.69 seconds, while in the identified model the frequency is 0.233 rads/sec and the time constant 22.37 seconds. HELISTAB therefore appears to underestimate the stability of this mode. However, further work is required to investigate this more fully.

Finally, in order to verify the identified model it was used to predict the responses to a pilot input other than a double–doublet. Figure 2.21 shows a comparison between the measured and predicted responses to a 1221 input. it can be seen that these are in fair agreement, and suggest that the identified model is a reasonable representation of the system.

2.4.4 Difficulties of Output–Error Identification of Unstable Systems

When performing the output–error identification for the Lynx, serious numerical difficulties were encountered as a result of the unstable mode present in the system. An insight into these difficulties can be gained by considering a simple first order output–error system, as follows :

$$\frac{dx(t)}{dt} = a\,x(t) + u(t)$$

$$z(t) = x(t) + \eta(t) \qquad (2.17)$$

where,

a is the parameter to be estimated

u(t) is the input,

x(t) is the model state,

z(t) is the output

$\eta(t)$ is a Gaussian noise process with zero mean and unity variance

If the input, u is a step, then the system response can be expressed in the following form :

$$z(t) = -1/a\ (1 - e^{at}) + \eta(t) \qquad (2.18)$$

With this system, in time–domain output–error the likelihood function is maximised by minimising the following cost function :

$$J(a) = \sum_{i=1}^{N} (z(t_i) - x(t_i))^2 \qquad (2.19)$$

$$= \sum_{i=1}^{N} (z(t_i) + \frac{1}{a} (1-e^{at_i}))^2 \qquad (2.20)$$

$$\Rightarrow \frac{dJ}{da}(a) = \sum_{i=1}^{N} 2\ \varepsilon(a,i)\ s(a,i) \qquad (2.21)$$

where,

$$\varepsilon(a,i) = z(t_i) - x(t_i)$$

$$s(a,i) = -\frac{1}{a^2} + e^{at_i}(\frac{1}{a^2} - \frac{t_i}{a})$$

For an unstable system, the parameter, a is positive, and the scaling factor s(a,i) is therefore divergent with time. This produces ill–conditioning in the derivative, dJ(a)/da, which in turn leads to difficulties when attempting to minimise the cost function, J(a) using numerical techniques. The use of output–error identification with unstable systems therefore involves a significantly more demanding numerical problem than is the case for stable systems.

For the present work, these difficulties were overcome by the simple expedient of reducing the size of the integration step used in the routines for calculating J(a) and dJ(a)/da in order to improve the accuracy. However, while this permitted parameter estimates to be obtained, it also greatly increased the computing time required, and was found to be less than satisfactory. There is therefore a need for more work on the problems involved in identifying unstable systems.

2.4.5 Dispersion Matrices of Multi– step Inputs

Using the 5^{th} order model obtained by output– error identification, |D| was calculated for an ideal doublet, 1221, and double– doublet. The results are shown in table 2.4.

Input	|D|
Doublet	136.2
1221	159.7
Double-Doublet	3.95

Table 2.4 – *|D| for various multi–step inputs, calculated using 5^{th} order output–error identified Lynx model*

It can be seen that the doublet and 1221 both give determinants around 150, while the double– doublet gives the significantly lower value of 3.95. These results appear to support those presented earlier in this chapter, suggesting that the double– doublet is superior to both the doublet and 1221 inputs.

REFERENCES FOR CHAPTER TWO

1. Padfield, G.D.; Thorne, R.; Murray–smith, D.J.; Black, C.G.; Caldwell, A. 'UK Research into system identification for helicopter flight mechanics' Paper no. 82, presented at the 11th European Rotorcraft Forum, 1985

2. Mehra, R.K. 'Frequency–domain synthesis of optimal inputs for linear system parameter estimation' Trans ASME (J. Dynamic Systems, Meas, and Control) Vol 98, 1976 pp130–138

3. Mehra, R.K. 'Optimal input signals for parameter estimation in dynamic systems – survey and new results' IEEE Trans Aut Control, Vol AC–19, 1974 pp753–768

4. Black, C.G.; Murray–smith, D.J.; Padfield, G.D. 'Experience with frequency–domain methods for helicopter system identification' Paper no. 76, presented at the 12th European Rotorcraft Forum, 1986

5. Black, C.G. 'A methodology for the identification of helicopter mathematical models from flight data based on the frequency domain' PhD Thesis, University of Glasgow, 1988

6. Bendat, J.; Piersol, A. 'Engineering applications of correlation and spectral analysis' Wiley, Toronto 1980

7. Leith, D.J. 'A set of FORTRAN routines to calculate the ordinary coherence function – user's guide and internal description' Dept of Electronics and Electrical Engineering Internal Report FS 87005, University of Glasgow, 1987

CHAPTER THREE

CONSTRAINTS FOR USE WITH ROTORCRAFT OPTIMAL INPUTS

3.1 INTRODUCTION

With the encouraging results obtained using sub– optimal inputs (Chapter 2), a closer study was consequently made of optimal inputs. The D– optimality criterion has been used, for the reasons outlined in Chapter 1, with an output– error model of the system. This model is as follows :

$$dx(t)/dt = A\,x(t) + B\,u(t) \qquad (3.1)$$

$$y(t) = C\,x(t)$$

$$z(t) = y(t) + \eta(t)$$

where,

A,B,C are suitable matrices containing the model parameters, θ

$\eta(t)$ is a zero mean Gaussian process with covariance, R

y(t) is noise– free system response, z(t) is noisy system response

For this model, the dispersion matrix is given by equation (1.19) as :

$$D = M^{-1} \qquad (3.2)$$

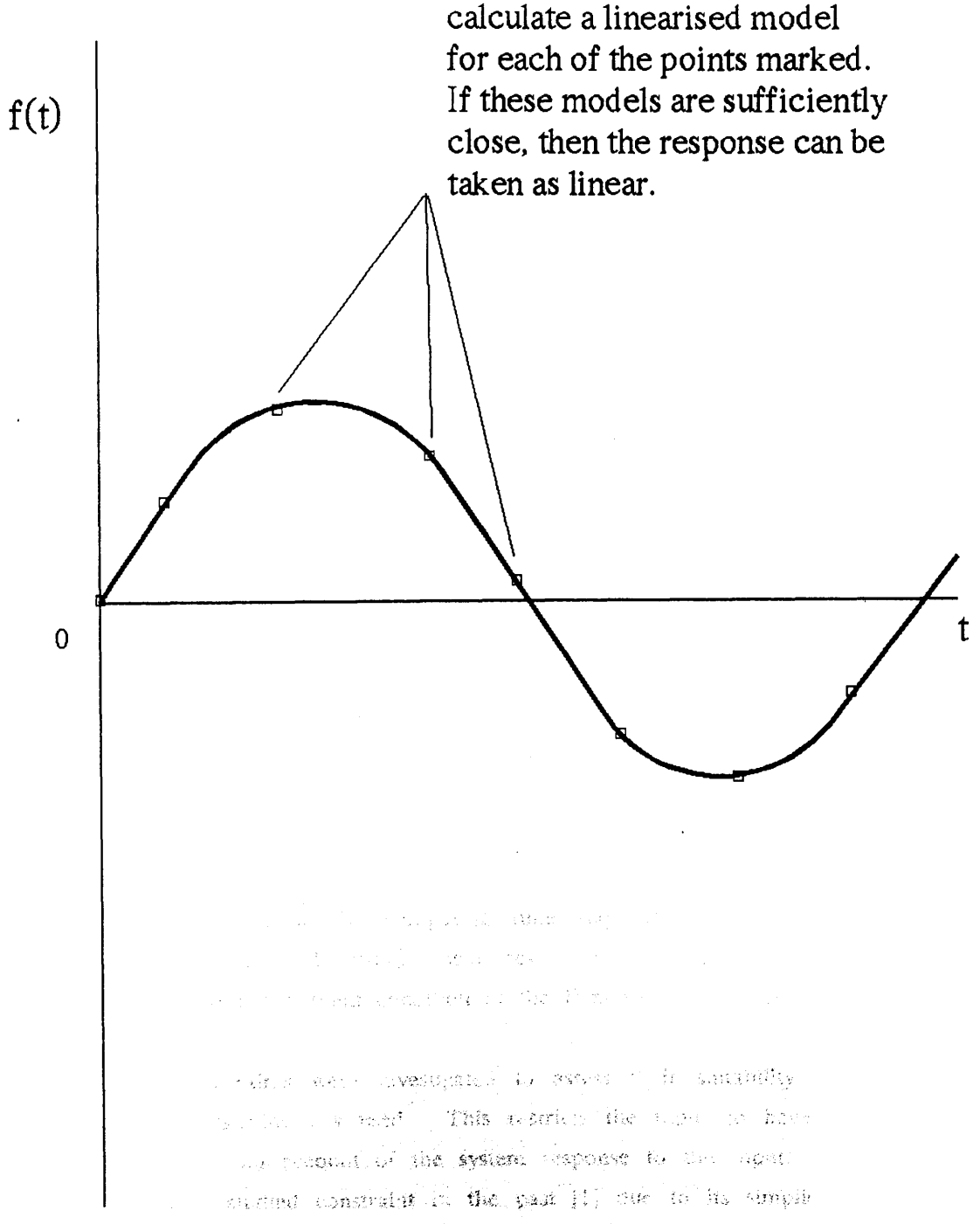
$$M = \int_0^T (dy(t)/d\theta)^T\, R^{-1}\, dy(t)/d\theta\; dt$$

where T is the length of the test record used.

In the rotorcraft case, the matrices A, B, and C are obtained by linearising the non– linear HELISTAB model about a particular flight condition. If the system response departs too far from this flight condition then it becomes non– linear, and the linearised model is invalid. It is therefore necessary to constrain the inputs to give linear test records, while minimising $|D|$.

Figure 3.1

An outline of the arrangement for a linearity constraint based on obtaining several linearised models along the trajectory of a response.



Unfortunately, determining the linearity of a response is not straightforward. As described in Chapter 2, it is possible to use the coherence function to assess linearity. However, this function is affected by a large number of factors, and can be subject to severe biases. These make the coherence function difficult to use quantitatively, and it was therefore thought to be unsuitable for use in a numerical optimisation scheme.

An alternative method of measuring linearity is to obtain the system response using the non–linear HELISTAB model. This non–linear model can then be linearised at flight conditions corresponding to several points in the calculated system response. If the parameters of these linearised models agree to within the expected variance of the estimates of these parameters obtained by identification, then the response may be taken a linear for practical purposes. This arrangement is outlined in figure 3.1.

However, this is a complex constraint, and requires a large amount of computing time. For example, obtaining a linearised model typically requires 2 seconds of computing time with HELISTAB (on a DEC VAX 11/750). Typically a test record contains approximately 1000 sample points, and a linearised model needs to be generated at every tenth point. Hence, around 100 such models are required to check the linearity of a response. It has been the author's experience that the constraint is evaluated around 1000 times in an input design run. This gives 200,000 seconds, or 55 hours of computing time, which is clearly unrealistic.

In order to reduce the computing time required, a simpler constraint is required, which will still ensure linear responses i.e. responses which do not depart too far from the flight condition of the linearised model used.

Various constraints were investigated to assess their suitability. Firstly, an input energy constraint was used. This restricts the input to have a particular energy, but takes no account of the system response to the input. It has been the most widely studied constraint in the past [1] due to its simplicity, and can be expressed as follows (see equation (1.23)) :

$$\int_0^T u^T(t)\, u(t)\, dt = 1 \qquad (3.3)$$

An extension of this constraint takes account of the energy of the system response, to give a combined input and output energy constraint [1], as follows (see equation (1.25)) :

$$\int_0^T (u^T(t)\, u(t) + y^T(t)\, Q\, y(t))\, dt = 1 \qquad (3.4)$$

where Q is a weighting matrix.

Unfortunately, despite their relative simplicity, neither of the energy constraints described directly address the problem of ensuring low amplitude responses. It is proposed that this can be achieved by imposing an explicit amplitude constraint on the system response, as follows :

$$|y(t)| \leqslant L_y, \qquad 0 \leqslant t \leqslant T \qquad (3.5)$$

where L_y is the amplitude limit.

This constraint guarantees that the reponses are of the required maximum amplitude. However, the constraint is essentially discontinuous in nature, and so leads to a more difficult optimisation problem than the energy constraints considered previously.

Once a constraint has been chosen, the number of inputs that it is necessary to design before a satisfactory result is obtained can be greatly reduced by making use of the property that D− optimality is invariant under scaling of the system input. It is then possible to simply scale an input to give a desired energy or response amplitude, avoiding the need to design a completely new input each time.

Finally, the question of robustness must be considered. From the results obtained using sub− optimal inputs in Chapter 2, it was found that robustness was an important factor in the success of an input, in addition to the linearity constraints discussed so far. Extra constraints may therefore be needed to ensure robustness.

Several types of robustness are required. Firstly, when inputs are applied manually, they must be insensitive to errors in amplitude and timing introduced at this stage. For the present work, it is intended that the control input device developed at RAE (Bedford) will be used to apply the optimal inputs automatically, without pilot intervention. Robustness to errors in applying the inputs is therefore less important.

Secondly, the model used to design the inputs will be inaccurate, otherwise the system identification would be unneccessary. Inputs must therefore be able to tolerate such errors, and still give linear responses and high quality parameter estimates.

Considering this second type of robustness, in order to ensure that the responses are robust and remain linear under changes in the model parameters, it is proposed that the following constraint be included in the input design process :

$$\left|\frac{dy}{d\theta_i}(t)\right| \leqslant R_{y_i} \qquad i = 1, 2, \ldots q \qquad (3.6)$$

where,

q is the number of parameters to be identified

θ_i is the i^{th} parameter

R_{yi} is the robustness limit required.

This limits the sensitivity of the responses to changes in the model parameters. However, care must be taken when using this constraint, since in some situations $dy(t)/d\theta$ can be coupled to y(t), which in turn couples this robustness constraint with the output energy and output amplitude 'linearity' constraints described above. This can be shown by the following simple example.

Consider the scalar system,

$$\dot{x}(t) = a\,x(t) + b\,u(t) \quad \text{(a,b are the system parameters)}$$

$$y(t) = x(t)$$

$$\Rightarrow \quad \frac{d\dot{x}}{db}(t) = a\frac{dx}{db}(t) + u(t), \qquad \frac{dy}{db}(t) = \frac{dx}{db}(t)$$

If b = 1, then dy(t)/db = y(t)

In addition to ensuring that the response is robust and remains linear, it is also important that high quality parameter estimates are obtained. A second robustness constraint may therefore be needed. The following constraint is suggested as being suitable for this purpose :

$$\frac{\theta_i}{|D|}\ \frac{d|D|}{d\theta_i} \leqslant R_{D_i} \qquad i = 1, 2, \ldots q \qquad (3.7)$$

where, R_{Di} is the robustness limit required.

This constraint limits the sensitivity of $|D|$ to changes in the parameters i.e. limits the change in the variances of the parameter estimates.

In the past, other methods have been proposed for ensuring that $|D|$ is robust. These have usually involved modifying the D– optimality criterion itself to include an element of robustness. For example, instead of minimising $|D|$, the average value of $|D|$ over all parameter values can be minimised [1]. Alternatively, the maximum value of $|D|$ over all the parameter values can be minimised [2]. However, except for very simple systems, these criteria are too complex to be implemented in practice with the level of computing power currently available.

3.2 INPUT ENERGY CONSTRAINED OPTIMAL INPUTS

3.2.1 Frequency– Domain Designs

The design of input energy constrained inputs can be greatly simplified by assuming that sufficiently long test records are available [3,4,5]. This leads to the following frequency– domain problem :

$$\underset{u \epsilon U}{\text{minimise}}\ |D|$$

where,

$$D = M^{-1}$$

$$M = \int_0^\infty (dy(t)/d\theta)^T \, R^{-1} \, dy(t)/d\theta \, dt \quad \text{(from equation (2.4))}$$

$$= \int_{-\infty}^{\infty} F^*(\omega) \, R^{-1} \, F(\omega) \, S_{uu}(\omega) \, d\omega \quad \text{(from equation (2.10))}$$

$$U = \left\{ u : \int_0^\infty u^T(t) \, u(t) \, dt = 1 \right\} \tag{3.8}$$

$$= \left\{ u : \int_{-\infty}^{\infty} S_{uu}(\omega) \, d\omega = 1 \right\} \tag{3.9}$$

Now consider an input, u_{n+1} formed from the combination of inputs $u_0 \epsilon U$ and $u_n \epsilon U$ according to the following expression,

$$S_{u_{n+1}u_{n+1}}(\omega) = \alpha \, S_{u_0u_0}(\omega) + (1-\alpha) \, S_{u_nu_n}(\omega) \tag{3.10}$$

The information matrix of the input, u_{n+1} is given by,

$$M_{n+1} = \alpha \int_{-\infty}^{\infty} F^*(\omega) \, R^{-1} \, F(\omega) \, S_{u_0u_0}(\omega) \, d\omega$$

$$+ (1-\alpha) \int_{-\infty}^{\infty} F^*(\omega) \, R^{-1} \, F(\omega) \, S_{u_nu_n}(\omega) \, d\omega$$

$$= \alpha \, M_0 + (1-\alpha) \, M_n \tag{3.11}$$

and the dispersion matrix is given by,

$$D_{n+1} = M_{n+1}^{-1}$$

Now it is well known [5] that for any square matrix, A the following relation is true,

$$\frac{d \log |A|}{dx} = Tr \left(A^{-1} \frac{dA}{dx} \right) \tag{3.12}$$

Hence,

$$\frac{d\log |D_{n+1}|}{d\alpha} = -\frac{d\log|M_{n+1}|}{d\alpha} = -Tr\left(M_{n+1}^{-1} \frac{dM_{n+1}}{d\alpha} \right)$$
$$= -Tr\ (M_{n+1}^{-1}\ (M_0 - M_n)) \qquad (3.13)$$

Finally, consider the limit as $\alpha \to 0$. This gives,

$$\left.\frac{d\log |D_{n+1}|}{d\alpha}\right|_{\alpha=0} = -Tr\ (\ M_n^{-1}M_0 - I\ )$$
$$= -Tr\ (\ M_n^{-1}M_0) - q \qquad (3.14)$$

where,

q = dimension of information matrices
= number of parameters in system.

Or, for sufficiently small α,

$$\log |D_{n+1}| \simeq \log |D_n| - \alpha\ (\ Tr\ (M_n^{-1}M_0) - q\ ) \qquad (3.15)$$

If $\alpha\ (Tr\ (M_n^{-1}M_0) - q\ ) > 0$, then $|D_{n+1}| < |D_n|$, and u_{n+1} is a better input than u_n. This can be used as the basis for an optimisation algorithm that successively improves upon an input until the optimum is reached.

To simplify the optimisation algorithm, consider input u_0 to be a pure sine wave of frequency ω_0. Then,

$$S_{u_0u_0}(\omega) = \begin{cases} 0, & \omega \neq \omega_0 \\ 1, & \omega = \omega_0 \end{cases} \qquad (3.16)$$

and,

$$M_0 = F^*(\omega_0)\ R^{-1}\ F(\omega_0) \qquad (3.17)$$

Now,

$$\int_{-\infty}^{\infty} \mathrm{Tr}\ M_n^{-1}\ F^*(\omega)\ R^{-1}\ F(\omega)\ S_{u_n u_n}(\omega)\ d\omega\ = \mathrm{Tr}\ M_n^{-1} M_n\ = q \tag{3.18}$$

Hence, since $u_n \epsilon U$,

$$\max_{\omega} \mathrm{Tr}\ M_n^{-1}\ F^*(\omega)\ R^{-1}\ F(\omega)\ \geq\ q \tag{3.19}$$

i.e. we can choose ω_0 such that $\mathrm{Tr}(M_n^{-1} M_0) - q \geq 0$, and equality holds if and only if u_n is optimal.

Moreover, since ω_0 is a scalar quantity, a simple line search can be used to maximise $\mathrm{Tr}(M_n^{-1} M_0)$. A line search can also then be used to find α to minimise $|D_{n+1}|$. Finally, practical systems have finite bandwidth. Hence, the infinite limits of integration used in the expressions above can be replaced by a suitable finite value.

This leads to the following simple, efficient algorithm which will converge to the global D-optimum input.

1. Start with any input, $u_1 \epsilon U$ which has a non-singular information matrix, M_1. Let $n=1$.

2. Find the input $u_0 \epsilon U$ of frequency ω_0 which maximises $\varphi(u_0)$, where $\varphi(u_0) = \mathrm{Tr}\ (M_n^{-1} M_0)$

3. Update u_{n+1} to,

$$S_{u_{n+1} u_{n+1}}(\omega) = \alpha\ S_{u_0 u_0}(\omega) + (1-\alpha)\ S_{u_n u_n}(\omega)$$

where, α is chosen to minimise $|D_{n+1}|$ such that $u_{n+1} \epsilon U$

4. If $|D_n| - |D_{n+1}| < \varepsilon\ |D_n|$ for some specified ε, then stop, else $n = n+1$; goto step 2.

The convergence of this algorithm can be further improved by minor changes to steps 2 and 3 [6,7].

However, as was mentioned previously, in order to use this frequency domain approach long time histories must be available. Moreover, the system concerned must be stable. This latter condition results from the use of $F(\omega)$, the Fourier transform of f(t). For stable systems, f(t) is convergent, and the Fourier transform exists. However, f(t) is divergent for unstable systems, and in this case has no Fourier transform. Hence, the frequency–domain algorithms cannot be used with rotorcraft, which are both unstable and give short test records.

Nevertheless, the frequency–domain algorithms form the basis of the more powerful time–domain algorithms to be discussed later.

3.2.2 Time–Domain Designs

The time–domain algorithms for designing energy constrained inputs act as the complement of those based on the frequency–domain. Whereas the frequency–domain algorithms are simple and efficient, those in the time–domain are significantly more complex, and generally require more computing time. However, the time–domain methods do not suffer from the restrictions of those in the frequency–domain. The time–domain algorithms can be used with both stable and unstable systems, and with any length of test records.

In the time–domain, the energy constrained input design problem is cast as follows [8]:

$$\underset{u \epsilon U}{\text{minimise}} \; |D|$$

where,

$$D = M^{-1}$$

$$M = \int_0^T (dy(t)/d\theta)^T \, R^{-1} \, dy(t)/d\theta \; dt \quad \text{(from equation (2.4))}$$

$$U = \left\{ u : \int_0^T u^T(t) \, u(t) \; dt = 1 \right\} \tag{3.20}$$

T = length of test record

As in the frequency–domain, consider an input u_{n+1} formed from the combination of inputs $u_0 \epsilon U$ and $u_n \epsilon U$, now according to the expression,

$$u_{n+1}(t) = \alpha\, u_0(t) + \beta\, u_n(t) \tag{3.21}$$

For $u_{n+1} \epsilon U$, require,

$$\alpha^2 + \beta^2 + 2\alpha\beta\gamma = 1, \quad \text{where } \gamma = \int_0^T u_0^T(t)\, u_n(t)\, dt \tag{3.22}$$

The information matrix, M_{n+1} of the input u_{n+1} is given by,

$$M_{n+1} = \alpha^2 M_0 + \beta^2 M_n + 2\alpha\beta M_{0n} \tag{3.23}$$

where,

$$M_{0n} = \int_0^T (dy_0(t)/d\theta)^T R^{-1} dy_n(t)/d\theta\, dt \tag{3.24}$$

and,

$$\frac{d\log |D_{n+1}|}{d\alpha} = -Tr\, (M_{n+1}^{-1} \left[2\alpha M_0 + 2\beta\frac{d\beta}{d\alpha} M_n + 2(\beta+\alpha\frac{d\beta}{d\alpha})M_{0n}\right] \tag{3.25}$$

Now,

$$d\beta/d\alpha = -(\alpha+\beta\gamma)/(\beta+\alpha\gamma) \tag{3.26}$$

Hence, letting $\alpha \to 0$,

$$\left.\frac{d\log |D_{n+1}|}{d\alpha}\right|_{\alpha=0} = -2 \left[Tr(M_n^{-1}M_{0n}) - \gamma q \right] \tag{3.27}$$

Hence, if $Tr\ (M_n^{-1}M_{0n}) > \gamma q$, then, for sufficiently small α, can obtain $|D_{n+1}| < |D_n|$.

It has been shown by [8] that Tr $(M_n^{-1} M_{0n}) \geq \gamma q$ if u_0 is chosen to maximise Tr $(M_n^{-1} M_0)$ such that $\gamma \geq 0$. However, the proof of this is not straightforward. It is the opinion of the present author that a simpler and more elegant approach is to choose u_0 to directly maximise Tr $(M_n^{-1} M_{0n}) - \gamma q$. This leads to the following input design algorithm which will converge to the global optimum.

1. Start with any input, $u_1 \epsilon U$ which has a non-singular information matrix, M_1. Let $n=1$.

2. Find the input $u_0 \epsilon U$ which maximises $\varphi(u_0)$, where
$\varphi(u_0) = \text{Tr } (M_n^{-1} M_{0n}) - \gamma q$

3. Update u_{n+1} to,

$$u_{n+1}(t) = \alpha\, u_0(t) + \beta\, u_n(t)$$

where,

$\alpha^2 + \beta^2 + 2\alpha\beta\gamma = 1$

α, β chosen to minimise $|D_{n+1}|$

4. If $|D_n| - |D_{n+1}| < \epsilon\, |D_n|$ for some specified ϵ, then stop, else $n = n+1$; goto step 2.

Most of the computing time required by this algorithm is absorbed in step 2. Further details concerning the implementation of this step can be found in Appendix B.

Figure 3.2

|D| versus the length of test record used for input energy constrained optimal inputs (simple first order system, a=−1, b=1).

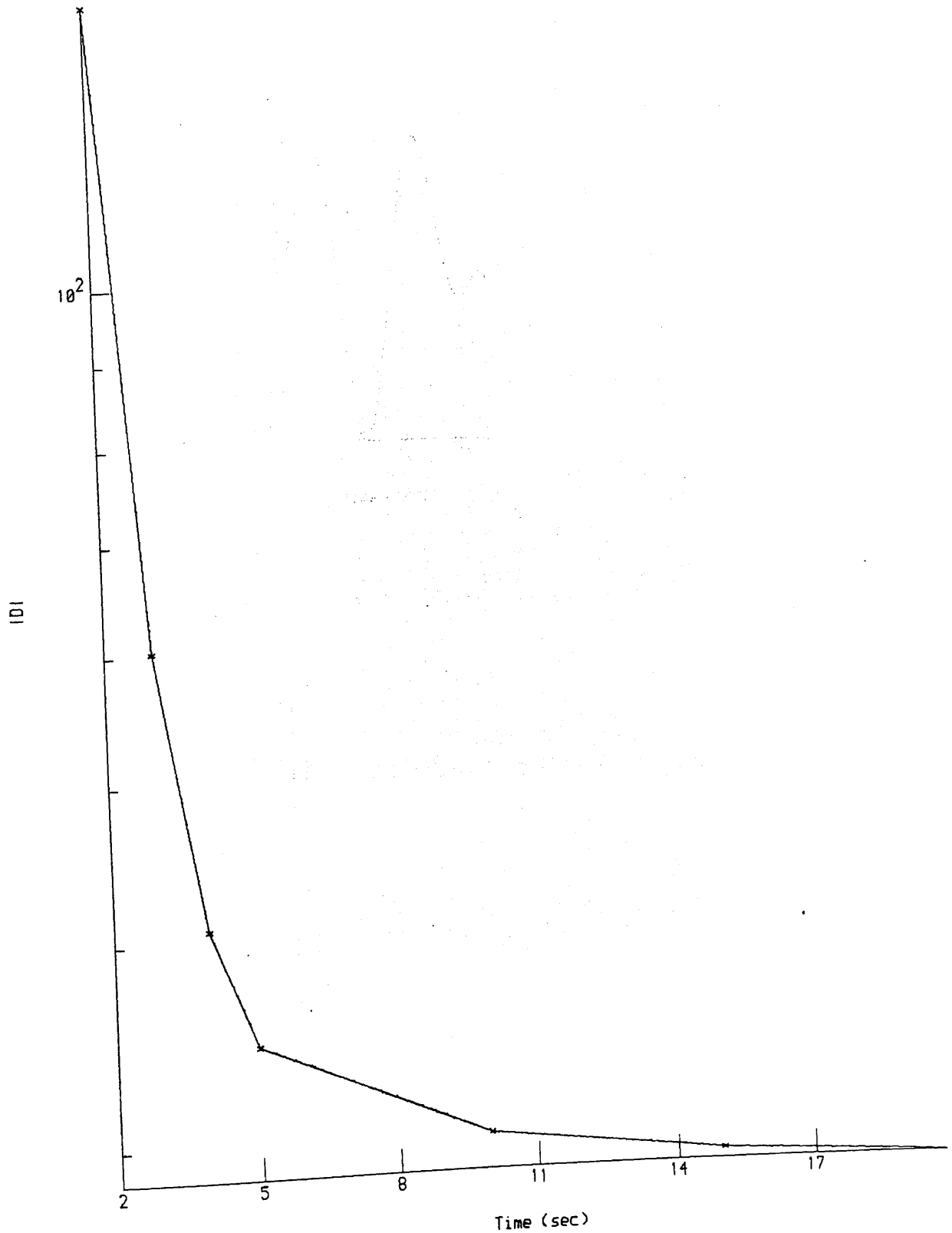
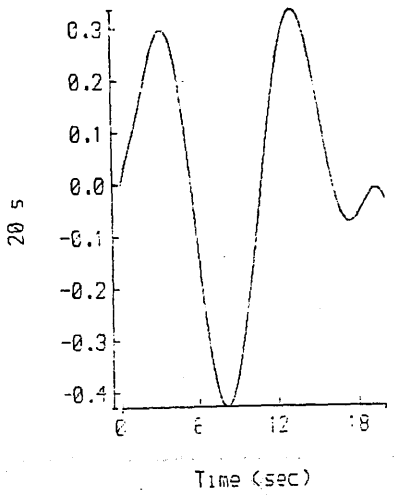


Figure 3.3

The time history and auto-spectrum of the input energy constrained optimal input for a 20 second test record (simple first order system, a=−1, b=1).



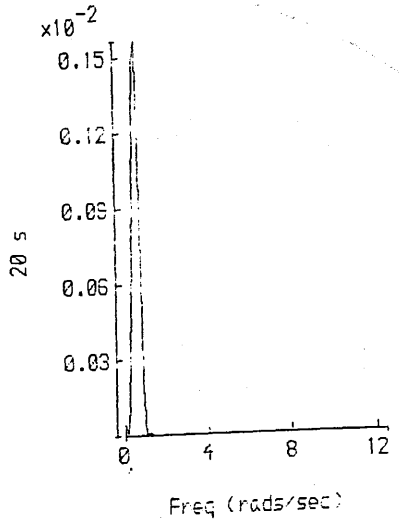
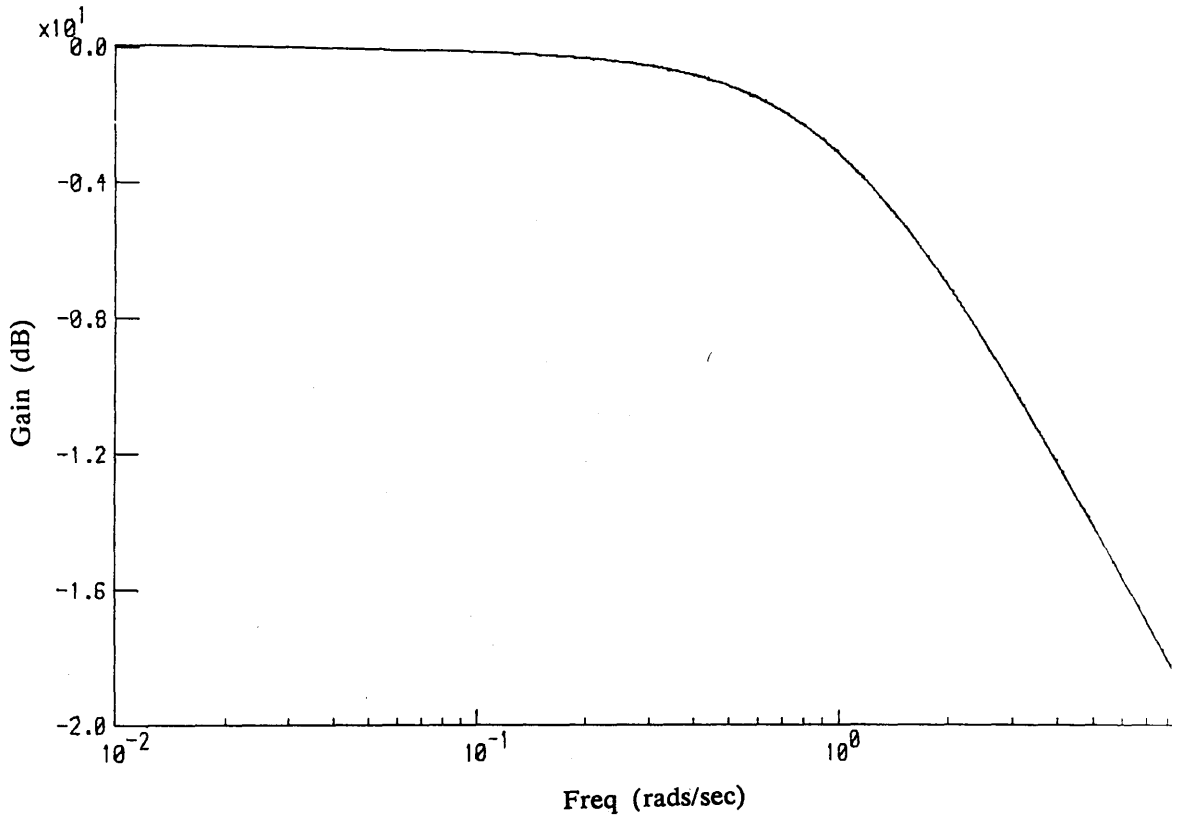


Figure 3.4

Bode plot of the gain and phase of the simple first order system studied (a=−1, b=1)



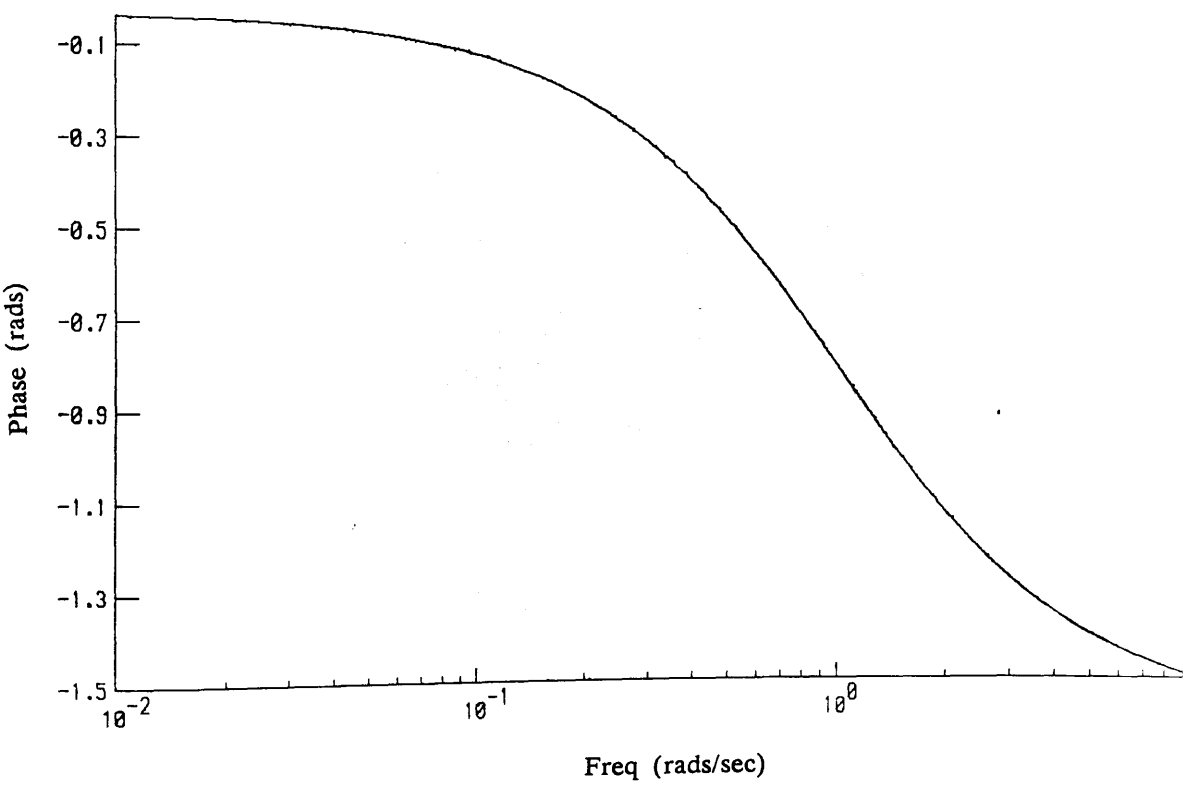


Figure 3.5

|D| versus the length of test record used for input energy constrained optimal inputs (simple first order system, a=+1, b=1)

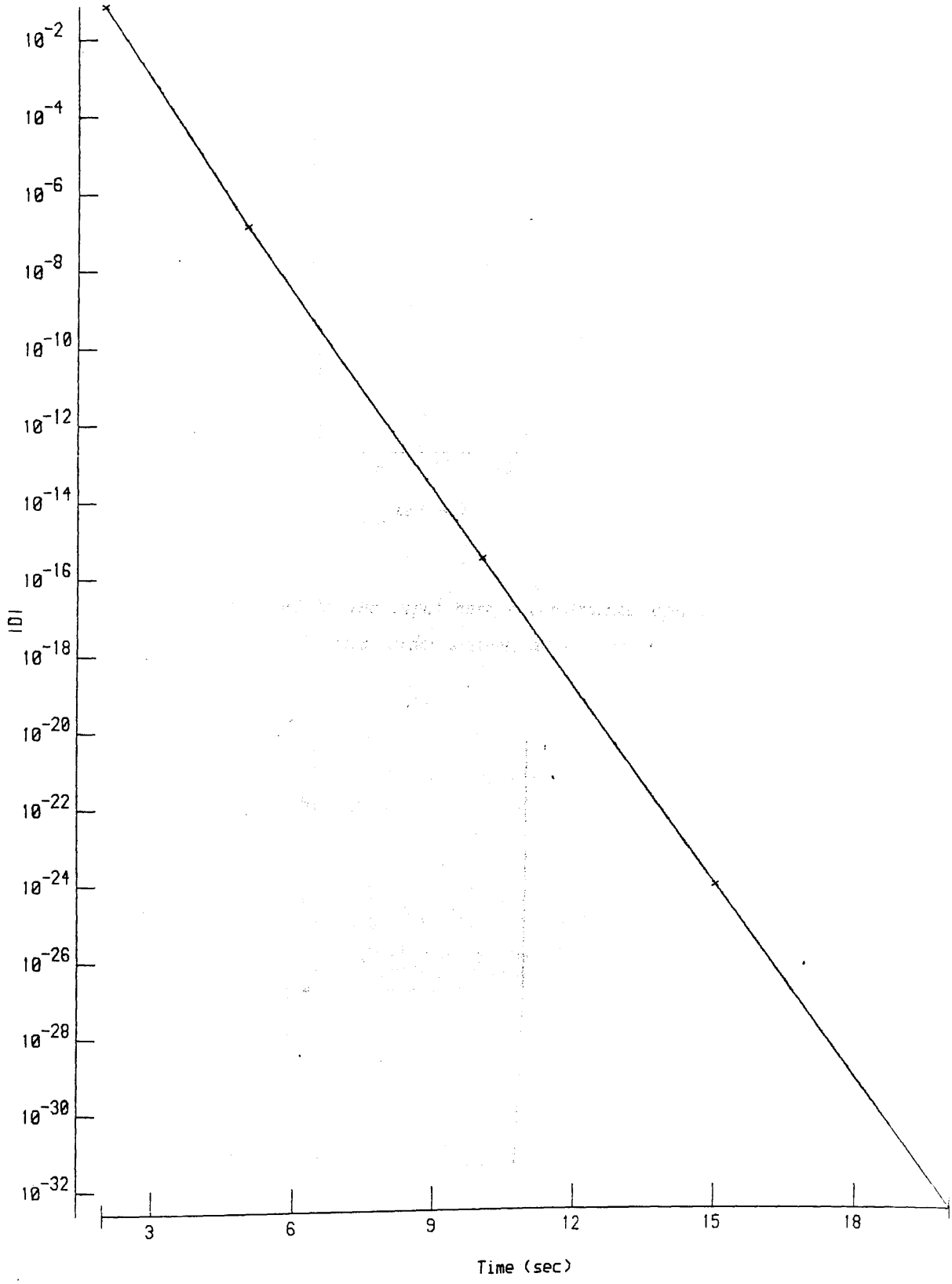


Figure 3.6

The input energy constrained optimal input for a 20 second test record (simple first order system, a=+1, b=1)

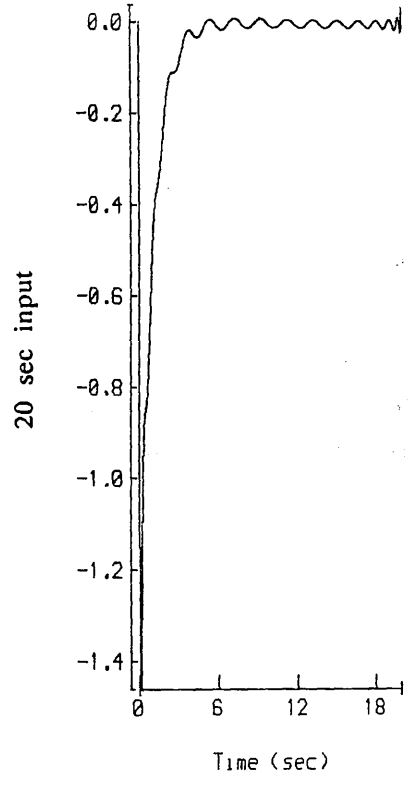
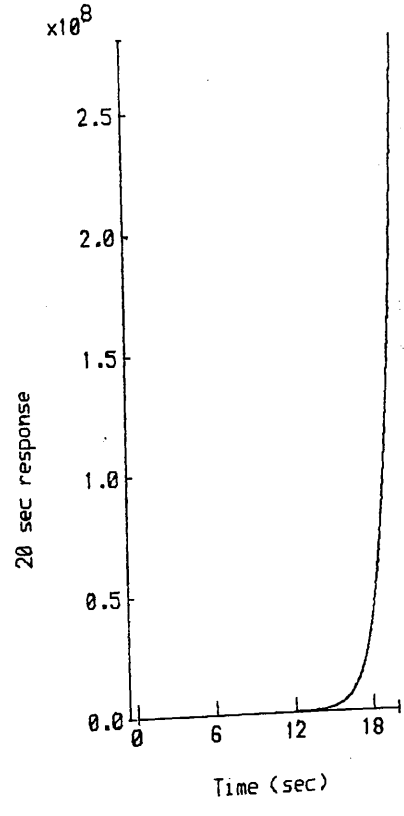


Figure 3.7

The response produced by the input energy constrained optimal input for a 20 second test record (simple first order system, a=+1, b=1)



3.2.3 A Simulation Study of Input Energy Constrained Optimal Inputs

In order to gain an insight into the characteristics of D− optimal inputs when subjected to an input energy constraint, such inputs were designed for the following simple first order system.

$$dx(t)/dt = a\,x(t) + b\,u(t) \qquad (3.28)$$

$$y(t) = x(t), \qquad z(t) = y(t) + \eta(t)$$

where,

$\eta(t)$ is a Gaussian process with zero mean and unity variance.

a,b are the relevant parameters of the system

For a stable system (a=−1, b=1), optimal inputs were designed for a variety of lengths of test record from 2 seconds to 20 seconds. The corresponding values of |D| are shown in figure 3.2, while the optimal input for a 20 second test record is shown in figure 3.3.

It can be seen that |D| is large for the short test records, but decreases rapidly as the record length increases and more information is obtained about the system. However, once sufficiently long test records are used, the rate of improvement in |D| becomes much smaller, and for a 20 second test record |D| is within 3% of that for an infinitely long record.

Turning to the optimal input for a 20 second test record, this is essentially a sine wave of frequency approximately 0.6 rads/sec, and compares with the sine wave of 0.582 rad/sec which is the optimal input for an infinite test record. A Bode plot of the gain and phase of the system is shown in figure 3.4. It can be seen that at 0.582 rads/sec the gain is still high, giving a good signal− to− noise ratio in the system response. In addition, at this frequency the phase has started to roll off, giving information about the position of the pole of the system. Hence, the optimal inputs appear to be intuitively reasonable.

Moreover, it can be shown that the input for an infinite test record contains between q/2p and q(q+1)/2 frequencies, where q is the number of parameters, and p is the number of system output variables [3]. Hence, for the current system the input should have between 1 and 3 frequencies (q=2, p=1). The input obtained is therefore relatively efficient, since it contains only one frequency.

Figures 3.5, 3.6, and 3.7 show the results for an unstable system (a=1, b=1). These exhibit significantly different characteristics from those for the stable case. Firstly, as the length of the test record is increased, |D| decreases exponentially, and does not appear to converge to a final value as in the stable system.

It can also be seen that the optimal input is essentially an impulse, and this produces a divergent system response. As time goes to infinity, this response will become infinite, and so the signal–to–noise ratio will also be infinite. In this situation, the system parameters can be identified exactly, giving a |D| of zero. Hence, once again the optimal inputs are intuitively reasonable, and the behaviour of |D| shown in figure 3.5 can be understood.

The results obtained for the unstable system serve to highlight several of the shortcomings of the input energy constraint that was used. Firstly, while the energy of the input is constrained, this energy may be concentrated at a particular point, giving a large input amplitude at that point. Secondly, and more importantly perhaps, no constraints are placed on the response of the system. Hence, the response may also be of large amplitude.

A simple input energy constraint such as that used is therefore unsuitable for rotorcraft applications, where the system is unstable and it is essential that the response is of limited amplitude in order to prevent it becoming non–linear.

In previous work on fixed–wing and rotorcraft input design, either an input energy constraint has been used, or an input amplitude constraint. No constraints were placed on the system response. As shown for the unstable first order system above, when there are no constraints on the response, optimal inputs tend to produce a large excitation of the system in order to obtain more information, and to give an improved signal–to–noise ratio. The disappointing results, discussed in Chapter one, that have been obtained in previous flight trials of such optimal inputs may therefore be due to these inputs producing responses of such large amplitude that non–linearities have been present.

3.3 OPTIMAL INPUTS WITH OUTPUT AND INPUT ENERGY CONSTRAINTS

3.3.1 Frequency–Domain Designs

In rotorcraft applications, if an input produces too great an excitation of the system, then the response will be of large amplitude, and hence non–linear. It is therefore necessary to ensure that any inputs designed are constrained to produce system responses of low amplitude. The simplest such constraint is to place a restriction on the energy of the system response. This can be achieved by a straightforward extension of the input energy constraint discussed in section 3.2 [3].

Taking the frequency–domain case first of all, assuming a stable system and sufficiently long test records, the input design problem with combined output and input energy constraints is the same terms as that in section 3.2.1, but with the set U defined as follows.

$$U = \left\{ u : \int_0^\infty (u^T(t)\, u(t) + y^T(t)\, Q\, y(t))\, dt = 1 \right\} \tag{3.29}$$

$$= \left\{ u : \int_{-\infty}^{\infty} (I + G^*(\omega)\, Q\, G(\omega))\, S_{uu}(\omega)\, d\omega = 1 \right\} \tag{3.30}$$

Q is a suitable weighting matrix

Now, define a new input, u' such that,

$$S_{u'u'}(\omega) = (I + G^{*}(\omega)\ Q\ G(\omega))\ S_{uu}(\omega) \qquad (3.31)$$

and express M in terms of u', giving,

$$M = \int_{-\infty}^{\infty} F'^{*}(\omega)\ R^{-1}\ F'(\omega)\ S_{u'u'}(\omega)\ d\omega \qquad (3.32)$$

where,

$$F'(\omega) = F(\omega)\ (I + G^{*}(\omega)\ Q\ G(\omega))^{-\frac{1}{2}} \qquad (3.33)$$

The optimisation problem can be re-expressed in terms of u' and F' as,

$$\underset{u' \epsilon U'}{\text{minimise}}\ |D|$$

where,

$$U' = \left\{ u : \int_{-\infty}^{\infty} S_{u'u'}(\omega)\ d\omega = 1 \right\}_J$$

This can be solved using the frequency-domain algorithm given in section 3.2.1. The input, u, can then be obtained again by using the relation,

$$S_{uu}(\omega) = (I + G^{*}(\omega)\ Q\ G(\omega))^{-1}\ S_{u'u'}(\omega) \qquad (3.34)$$

3.3.2 Time-Domain Designs

As in the case of the simple input energy constraint, the time-domain algorithms for the combined output and input energy constraint permit the use of both stable and unstable systems, and any length of test record.

The only published time–domain algorithm involving output and input energy constraints is that of Mehra [3]. However, this algorithm produces a randomised experiment design which is not suitable for use with practical systems. A new algorithm will therefore be presented, which is an extension of that described in section 3.2.2.

The time–domain input design problem subject to output and input energy constraints is the same as that in section 3,2.2, except that the set U is defined as follows.

$$U = \left\{ u : \int_0^T (u^T(t)\, u(t) + y^T(t)\, Q\, y(t))\, dt = 1 \right\} \tag{3.35}$$

where Q is a suitable weighting matrix.

Consider an input u_{n+1} formed from the combination of inputs $u_0 \epsilon U$ and $u_n \epsilon U$, according to the expression,

$$u_{n+1}(t) = \alpha\, u_0(t) + \beta\, u_n(t) \tag{3.36}$$

For $u_{n+1} \epsilon U$, require,

$$\alpha^2 + \beta^2 + 2\alpha\beta\gamma_s = 1 \tag{3.37}$$

$$\text{where,} \quad \gamma_s = \int_0^T (u_0^T(t)\, u_n(t) + y_0^T(t)\, Q\, y_n(t))\, dt \tag{3.38}$$

The information matrix, M_{n+1} of the input u_{n+1} is given by equation (3.23), and,

$$\left. \frac{d \log |D_{n+1}|}{d\alpha} \right|_{\alpha=0} = -2 \left[Tr(M_n^{-1} M_{0n}) - \gamma_s q \right] \tag{3.39}$$

Hence, it is possible to maximise $Tr\,(M_n^{-1} M_{0n}) - \gamma_s q$ to give the following input design algorithm which will converge to the global optimum.

1. Start with any input, $u_1 \epsilon U$ which has a non–singular information matrix, M_1. Let n=1.

2. Find the input $u_0 \epsilon U$ which maximises $\varphi(u_0)$,

where $\varphi(u_0) = Tr\ (M_n^{-1} M_{0n}) - \gamma_s q$

3. Update u_{n+1} to,

$$u_{n+1}(t) = \alpha\ u_0(t) + \beta\ u_n(t)$$

where,

$\alpha^2 + \beta^2 + 2\alpha\beta\gamma_s = 1$

α, β chosen to minimise $|D_{n+1}|$

4. If $|D_n| - |D_{n+1}| < \epsilon\ |D_n|$ for some specified ϵ, then stop, else $n = n+1$; goto step 2.

Once again, most of the computing time required by this algorithm is used in step 2.

3.3.3 A Simulation Study of Output and Input Energy Constrained Optimal Inputs

As in the case of the input energy constraint discussed in section 3.2, combined output and input energy constrained D−optimal inputs were designed for a simple first order system. The system was as follows :

$$dx(t)/dt = a\ x(t) + b\ u(t) \qquad (3.40)$$

$$y(t) = x(t), \qquad z(t) = y(t) + \eta(t)$$

where,

$\eta(t)$ is a Gaussian process with zero mean and unity variance.

a,b are the relevant parameters of the system

Figure 3.8

|D| versus the length of test record used for input and output energy constrained (Q=1) optimal inputs (simple first order system, a=−1, b=1)

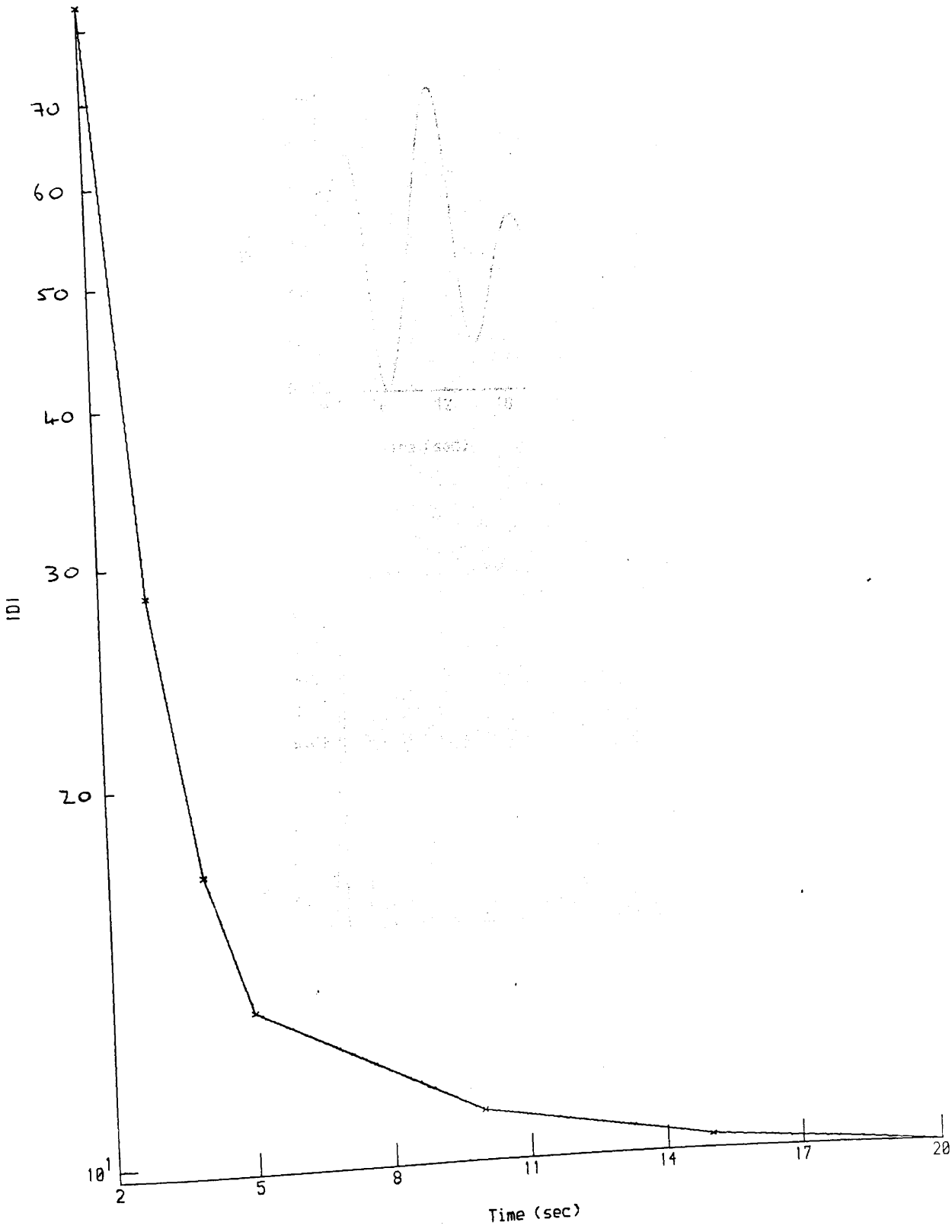
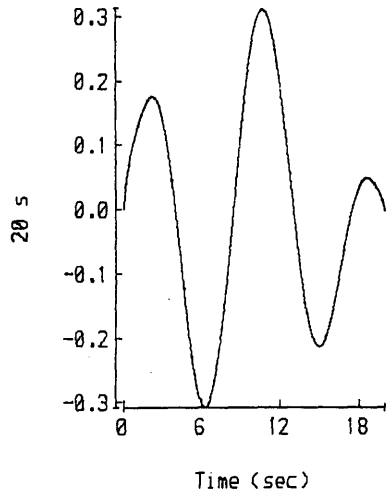


Figure 3.9

The time history and auto-spectrum of the input and output energy constrained (Q=1) optimal input for a 20 second test record (simple first order system, a=−1, b=1)



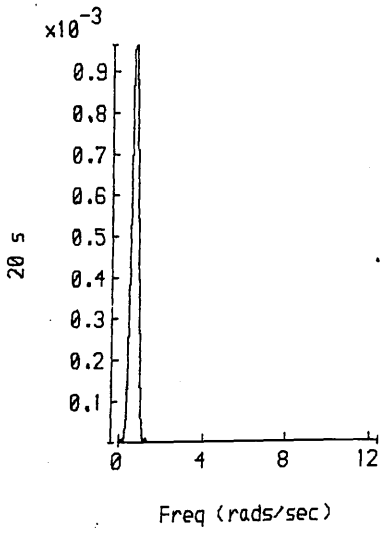


Figure 3.10

|D| versus the length of test record used for input and output energy constrained (Q=1) optimal inputs (simple first order system, a=+1, b=1)

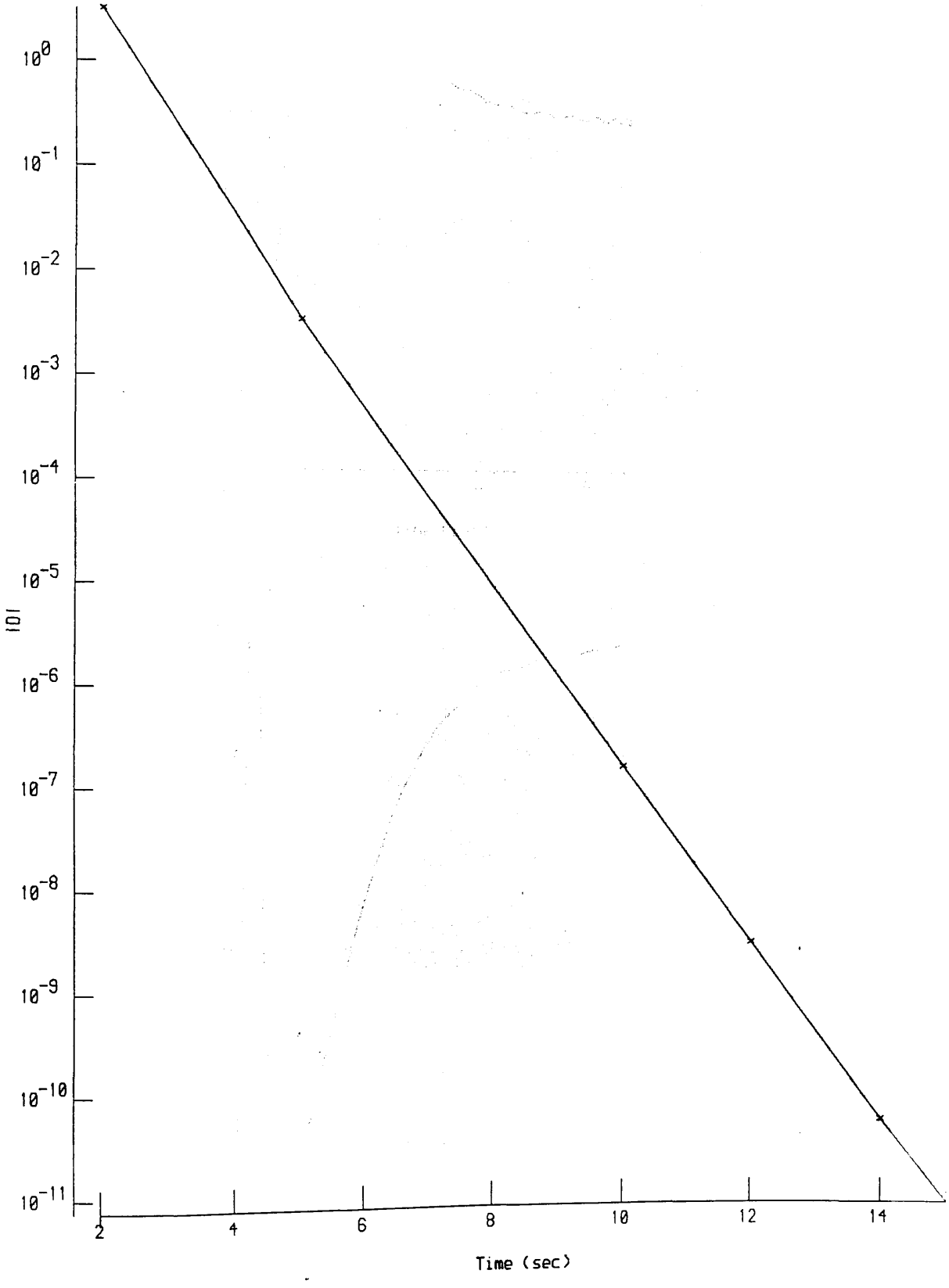
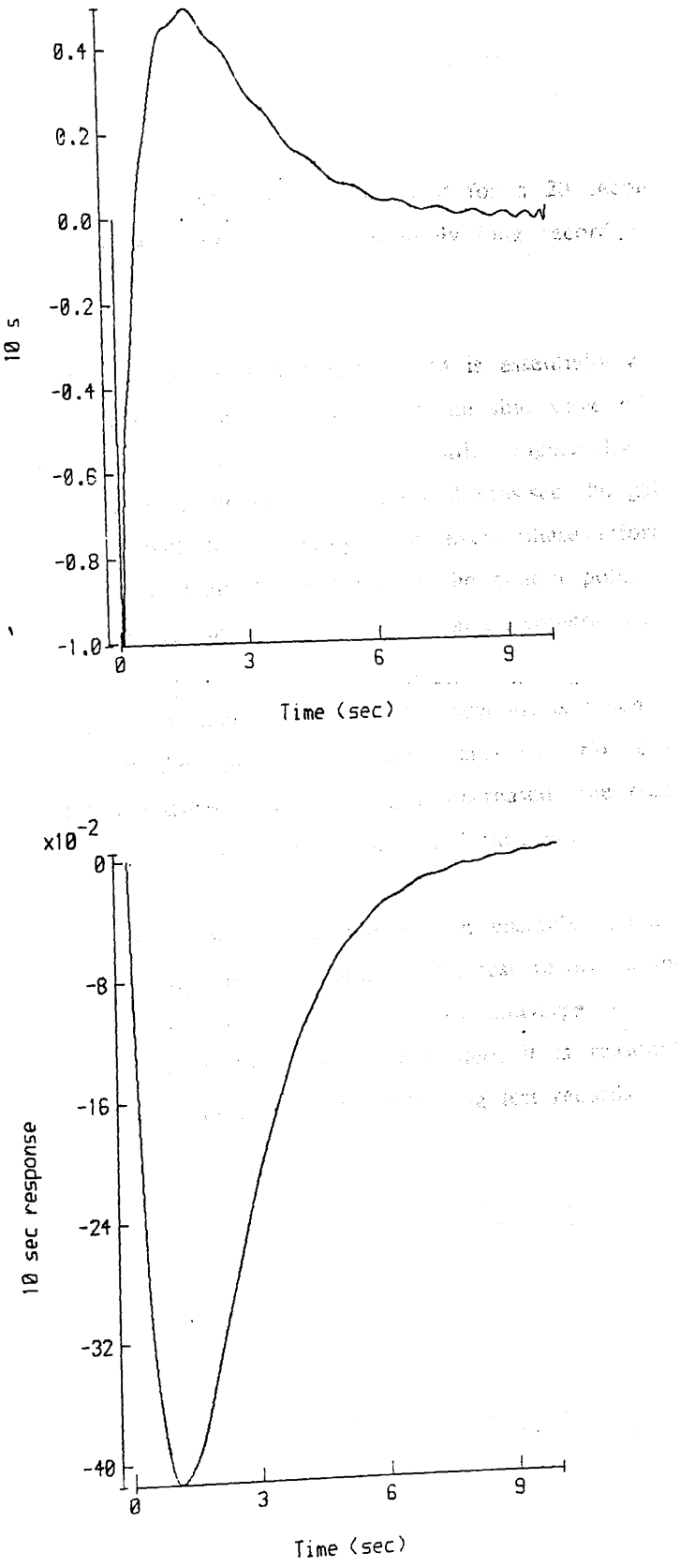


Figure 3.11

The input and output energy constrained (Q=1) optimal input for a 10 second test record together with the response produced by this input (simple first order system, a=+1, b=1)



For a stable system ($a=-1$, $b=1$), optimal inputs were designed with weighting $Q=1$ for a variety of lengths of test record from 2 seconds to 20 seconds. The corresponding values of $|D|$ are shown in figure 3.8, while the optimal input for a 20 second test record is shown in figure 3.9.

It can be seen once again that $|D|$ is large for the short test records, but decreases rapidly as the record length increases and more information is obtained about the system. However, once sufficiently long test records are used, the rate of improvement in $|D|$ becomes much smaller, and for a 20 second test record $|D|$ is 28.63, compared with 29.08 for an infinitely long record i.e. they agree to within 2%.

The optimal input for a 20 second test record is essentially a sine wave of frequency around 0.7 rads/sec, and compares with the sine wave of 0.673 rad/sec which is the optimal input for an infinite test record. Figure 3.4 shows a Bode plot of the gain and phase of the system. At 0.673 rads/sec the gain is reduced, giving a system response with lower energy. However, phase information is still present to give information about the postition of the system pole. This appears to be an intuitively reasonable input design for the given constraints.

As the weighting, Q, is increased, and more account is taken of the output energy, the frequency of the optimal input also increases. For example, $Q=10$ gives an input of 0.877 rads/sec. While if Q is decreased, the frequency of the optimal input also decreases, e.g. $Q=0.1$ gives 0.592 rads/sec.

Figures 3.10, and 3.11 show the results for an unstable system ($a=1$, $b=1$) and weighting $Q=1$. Firstly, as the length of the test record is increased, $|D|$ decreases exponentially, and does not appear to converge to a final value. However, since the energy of the response is limited, it is reasonable to expect that $|D|$ will eventually converge for sufficiently long test records.

The optimal input is a form of doublet, producing a response containing a large peak. Such a peak is clearly sensitive to the system dynamics, while also giving a good signal– to– noise ratio, and so should lead to accurate parameter estimates. However, this demonstrates that while the energy of the inputs and outputs is constrained, this energy may be concentrated at a particular point, giving a large input or output amplitude at that point. An output and input energy constraint such as that used therefore appears to be unsuitable for rotorcraft applications, where the system is unstable and it is essential that the response is of limited amplitude in order to prevent it becoming non– linear.

3.4 OUTPUT AMPLITUDE CONSTRAINED OPTIMAL INPUTS

3.4.1 Time– Domain Designs

In order to ensure that the response of the system concerned is prevented from being too large, constraints can be placed to restrict the maximum amplitude of response permitted. Such output amplitude constraints will guarantee responses with the desired magnitudes, unlike the energy constraints considered previously. The set U of valid inputs is now defined as follows :

$$U = \left\{ u : |y(t)| \leqslant L_y, \; 0 \leqslant t \leqslant T \right\} \tag{3.41}$$

where,

y(t) is the noise– free system response

T is the length of test record used.

Consider an input u_{n+1} formed from inputs $u_0 \epsilon U$ and $u_n \epsilon U$, according to the following expression

$$u_{n+1}(t) = \alpha \, u_0(t) + \beta \, u_n(t) \tag{3.42}$$

Then $u_{n+1} \epsilon U$ is satisfied by,

$$\alpha + \beta = K \tag{3.43}$$

where $K \geqslant 1$ is chosen such that $|x_{n+1}(t)| \leqslant L_y, \; 0 \leqslant t \leqslant T$.

Hence,

$$d\beta/d\alpha = -1 \tag{3.44}$$

This gives,

$$\left.\frac{d\log|D_{n+1}|}{d\alpha}\right|_{\alpha=0} = -2\left[\mathrm{Tr}\,(M_n^{-1}M_{on}) - q\right] \tag{3.45}$$

Maximising $\mathrm{Tr}\,(M_n^{-1}M_{on})$ will ensure that for u_n sub-optimal, and α sufficiently small, $|D_{n+1}| < |D_n|$. This gives the following input design algorithm.

1. Start with any input, $u_1 \epsilon U$ which has a non-singular information matrix, M_1. Let $n=1$.

2. Find the input $u_0 \epsilon U$ which maximises $\varphi(u_0)$,

   where $\varphi(u_0) = \mathrm{Tr}\,(M_n^{-1}M_{on})$

3. Update u_{n+1} to,

   $$u_{n+1}(t) = \alpha\, u_0(t) + \beta\, u_n(t)$$

   where,

   $\alpha + \beta = K$, K chosen such that $u_{n+1} \epsilon U$

   α, β chosen to minimise $|D_{n+1}|$

4. If $|D_n| - |D_{n+1}| < \varepsilon\, |D_n|$ for some specified ε, then stop, else $n=n+1$; goto step 2.

While this algorithm appears as straightforward as the algorithms described previously for the energy constraints, in practice the amplitude constraint is more complex to implement in step 2 (see Appendix B). Moreover, since amplitude constraints are essentially discontinuous in nature, the resulting optimisation problem can be prone to ill-conditioning.

Figure 3.12

The output amplitude constrained (L_y=0.2) optimal inputs plus responses for test records of 2, 10 and 20 seconds (simple first order system, a=−1, b=1)

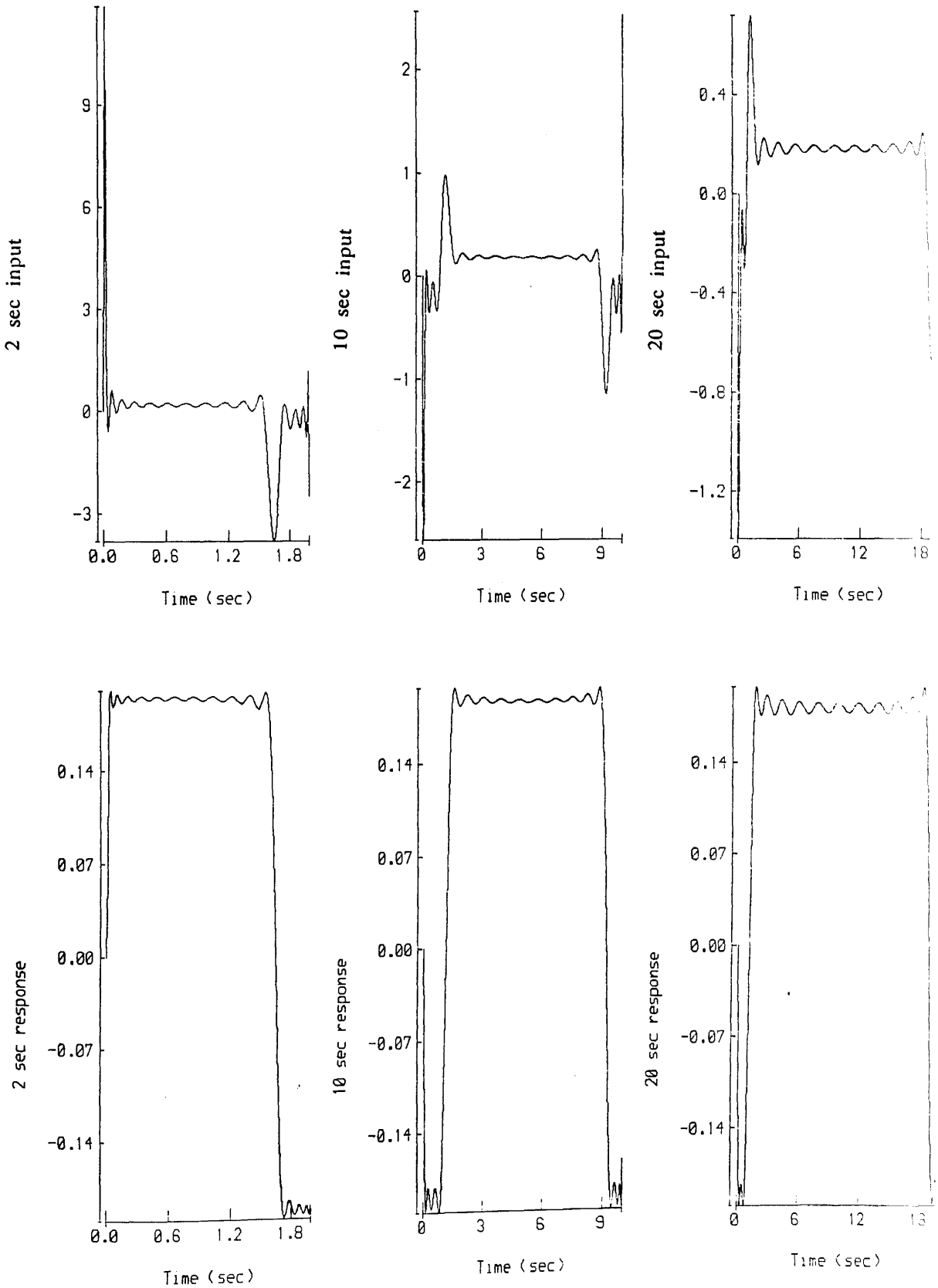


Figure 3.13

The sensitivities, dy(t)/ da and dy(t)/ db, for the output amplitude constrained (L_y=0.2) optimal input for a 2 second test record (simple first order system, a=−1, b=1)

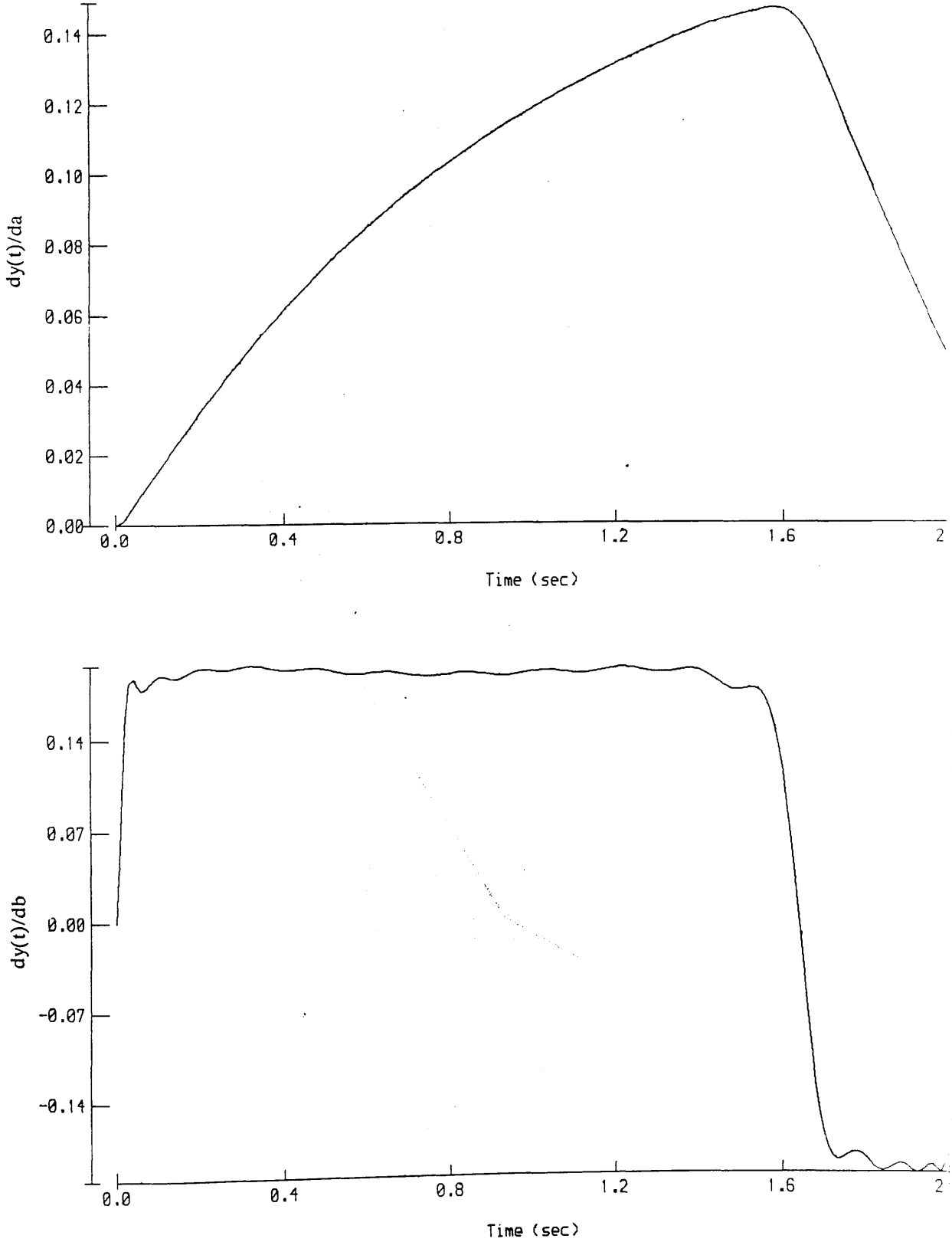


Figure 3.14

|D| versus the length of test record for output amplitude constrained (L_y=0.2) optimal inputs (simple first order system, a=−1, b=1)

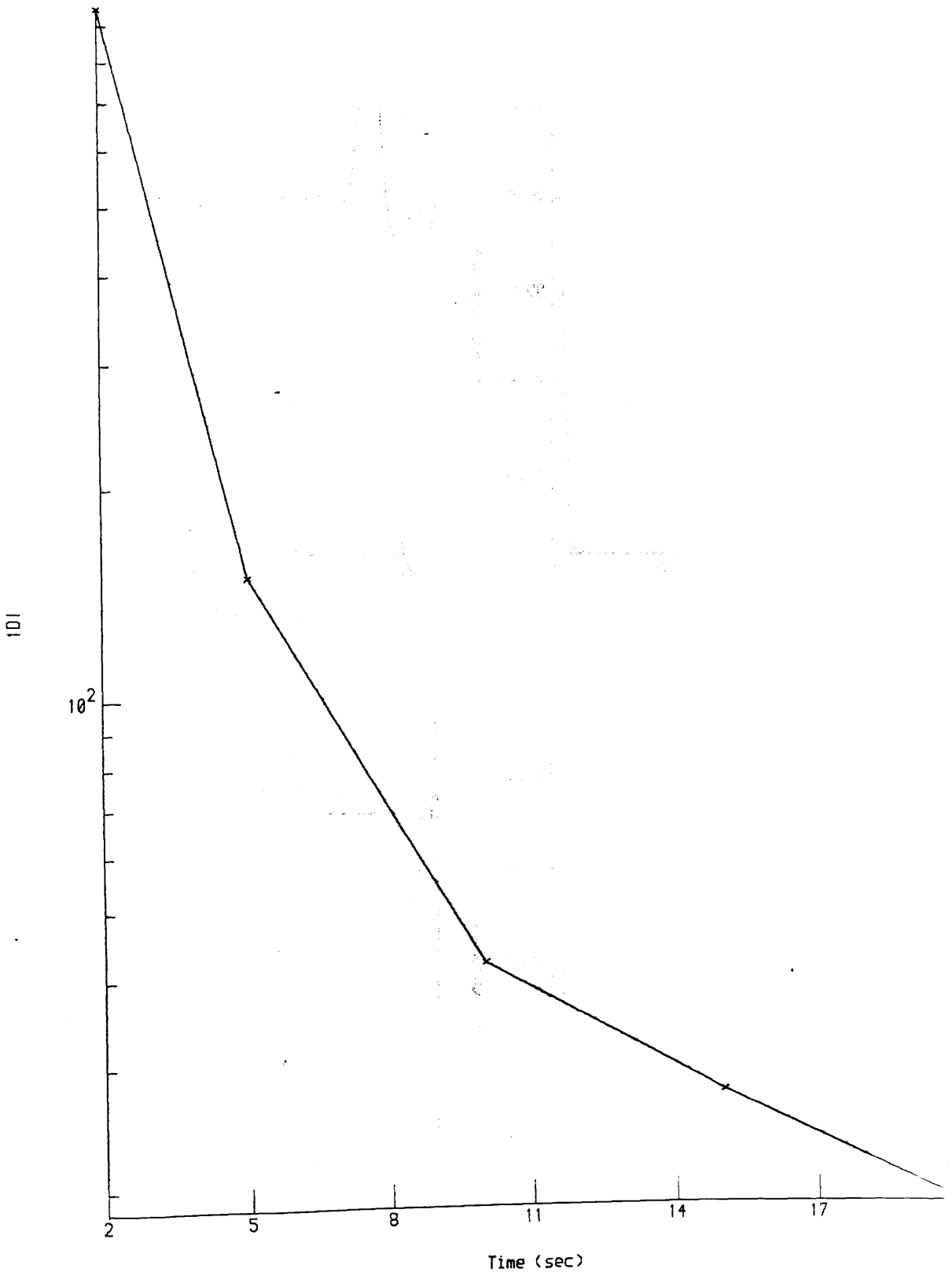


Figure 3.15

The output amplitude constrained (L_y=0.2) optimal inputs plus responses for test records of 2 and 5 seconds (simple first order system, a=+1, b=1)

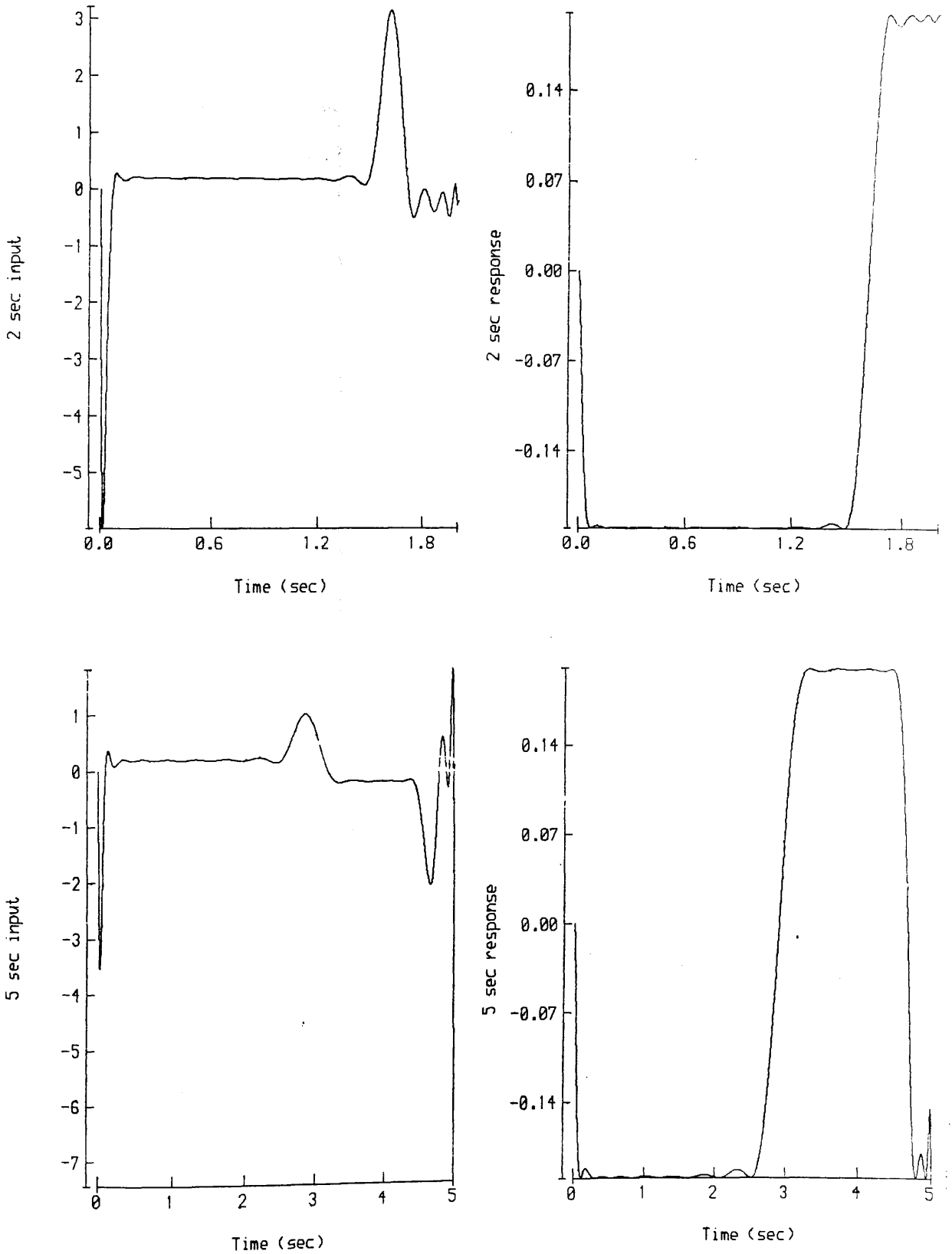


Figure 3.16

A bang–bang response of length T, with one transition at time T_1.

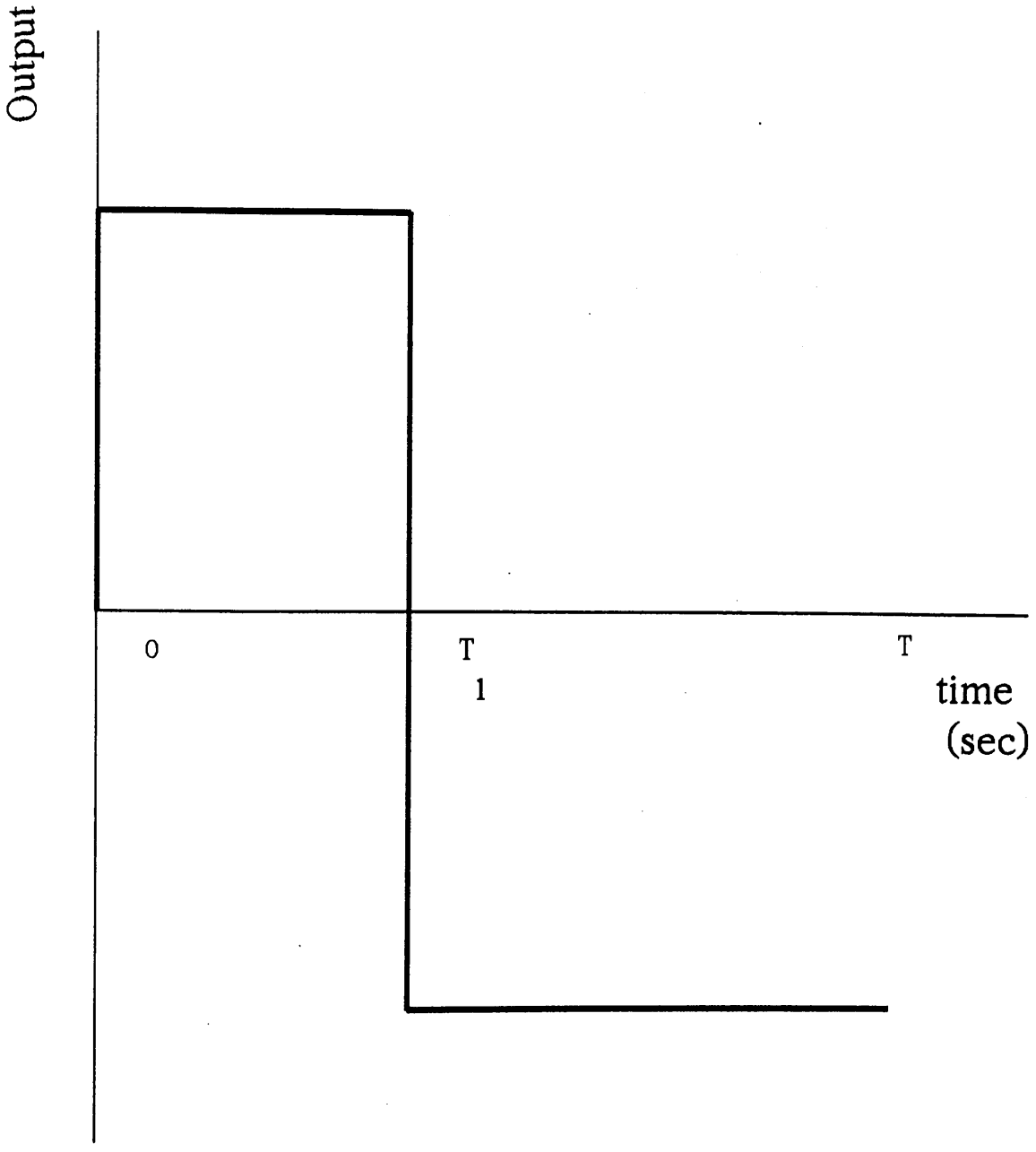
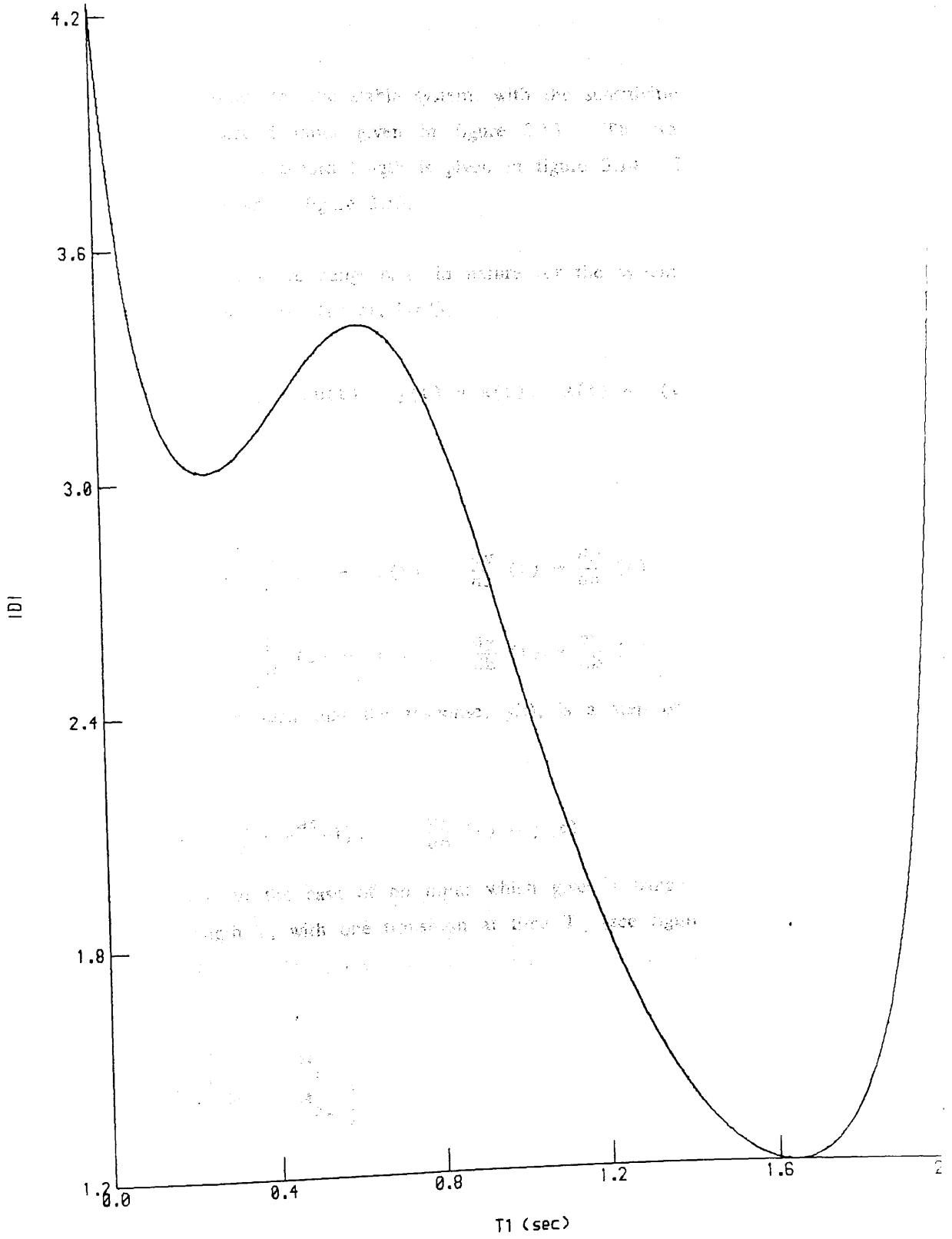


Figure 3.17

|D| versus T_1, for a bang–bang response of length T, with one transition at time T_1 (simple first order system, a=−1, b=1). (T = 2 sec)



3.4.2 A Simulation Study of Output Amplitude Constrained Optimal Inputs

In order to evaluate the suitability of output amplitude constraints for rotorcraft applications, D− optimal inputs were designed for the same simple scalar system that was used to study the energy constraints previously (see equation (3.28)). For both stable (a= − 1, b= 1) and unstable (a= 1, b= 1) systems, inputs were designed with responses constrained to be less than 0.2 units in amplitude. Figure 3.12 shows the optimal inputs and corresponding responses for test records of 2, 10, and 20 seconds for the stable system, with the sensitivities dx(t)/da and dx(t)/db for the 2 second input given in figure 3.13. The variation of the optimal |D| with the test record length is given in figure 3.14. Results for the unstable system are shown in figure 3.15.

Firstly, the responses are bang− bang in nature for the system used. Recall that the system was as follows (a= ±1, b= 1),

$$\dot{x}(t) = ax(t) + bu(t), \quad y(t) = x(t), \quad z(t) = y(t) + \eta(t) \tag{3.46}$$

Hence,

$$\frac{d\dot{x}}{da}(t) = a\frac{dx}{da}(t) + x(t), \quad \frac{dy}{da}(t) = \frac{dx}{da}(t) \tag{3.47}$$

$$\frac{d\dot{x}}{db}(t) = a\frac{dx}{db}(t) + u(t), \quad \frac{dy}{db}(t) = \frac{dx}{db}(t) \tag{3.48}$$

Consider an input such that the response, y(t), is a step of amplitude, h. Then,

$$\frac{dy}{da}(t) = \frac{h}{a}\left(e^{at}-1\right), \qquad \frac{dy}{db}(t) = y(t) \tag{3.49}$$

Now apply this to the case of an input which gives a bang− bang response of amplitude h, length T, with one transition at time T_1 (see figure 3.16). .For the stable system (a= − 1, b= 1) the information matrix, M, is as follows.

$$M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} \tag{3.50}$$

where,

$$M_{11} = \int_0^T \left(\frac{dy}{da}\right)^2 (t)\, dt$$

$$= \int_0^{T_1} h^2(1 - e^{-t})^2 \, dt + \int_{T_1}^{T} \left[e^{-(t-T_1)}(h_1 + h) - h \right]^2 dt \quad (3.51)$$

$$h_1 = (1 - e^{-T_1})\, h \quad (3.52)$$

$$M_{12} = M_{21} = \int_0^T \frac{dy}{da}(t) \frac{dy}{db}(t) \, dt$$

$$= \int_0^{T_1} h^2(1 - e^{-t}) \, dt + \int_{T_1}^{T} -h \left[e^{-(t-T_1)}(h_1 + h) - h \right] dt \quad (3.53)$$

$$M_{22} = \int_0^T \left(\frac{dy}{db}\right)^2 (t) \, dt = h^2 T \quad (3.54)$$

and,

$$|D| = \frac{1}{|M|} = \frac{1}{M_{11} M_{22} - M_{12} M_{21}} \quad (3.55)$$

$$\Rightarrow \frac{d|D|}{dT_1} = -\frac{1}{|M|^2} \left[\frac{dM_{11}}{dT_1} M_{22} - 2 \frac{dM_{12}}{dT_1} \right] \quad (3.56)$$

$$= 0 \quad \text{for maxima/minima}$$

For h = 0.2 units, T = 2 seconds, a line search algorithm was used to solve this for T_1. A value of 1.65 seconds was obtained for T_1, which is in agreement with the results given in figure 3.13. A plot of |D| versus T_1 is shown in figure 3.17, and clearly shows the minima at 1.65 seconds.

Using these results, insight may be gained into the factors affecting output amplitude constrained D−optimal inputs. It can be seen from equation (3.55) that $|M|$ is maximised, and hence $|D|$ minimised, by having M_{11} and M_{22} large, and M_{12} small. M_{11} and M_{22} measure the energy of dy(t)/da and dy(t)/db respectively, while M_{12} measures the correlation between dy(t)/da and dy(t)/db. A good input will therefore attempt to maximise the energy of the parameter sensitivities, dy(t)/da and dy(t)/db, while minimising their correlation.

In the example above, x(t) = dy(t)/db, and so a bang−bang response maximises the energy of dy(t)/db. The time of the transition, T_1, is then chosen to maximise the energy of dy(t)/da while minimising the correlation with dy(t)/db (see figure 3.13 for dy(t)/da and dy(t)/db when $T_1 = 1.65$ seconds).

Finally, figure 3.14 shows that $|D|$ does not appear to converge to a final value as the test record length is increased. This may be attributed to the energy of the responses increasing as the test record is made longer, since only the amplitude is constrained.

It is proposed that the output amplitude constraint produces inputs suitable for rotorcraft applications, since it guarantees responses of limited amplitude. If a suitable amplitude is chosen, then the response will be linear, as required.

3.5 RESPONSE ROBUST OPTIMAL INPUTS

3.5.1 Time Domain Designs

For a given model, the output amplitude constraint discussed in section 3.4 can be used to ensure a linear response. However, since the model of interest is known to be inaccurate, it is important that the responses are robust, and remain linear when used with the real system. The following constraint is suggested for this purpose.

$$\left|\frac{dy}{d\theta_i}(t)\right| \leqslant R_{y_i} \qquad i = 1, 2, \ldots, q \qquad (3.57)$$

where,

q is the number of parameters to be identified

θ_i is the i^{th} parameter

R_{yi} is the robustness limit required.

This constraint restricts the sensitivity of the response to changes in the model parameters. Using this constraint to give robustness, and the output amplitude constraint for linearity, leads to an input design problem as follows,

$$\min_{u \epsilon U} \; |D| \tag{3.58}$$

where,

$$U = \left\{ u : |y(t)| \leqslant L_y \; ; \; \left|\frac{dy}{d\theta_i}(t)\right| \leqslant R_{y_i} \right\} \tag{3.59}$$

with $i = 1, 2, \ldots q; \quad 0 \leqslant t \leqslant T$

Since the robustness constraint can be expressed as an amplitude constraint, the input design algorithm given in section 3.4 for the output amplitude constraint can also be used for this robust input design problem. The only change in the algorithm is to now use the set U given in equation (3.59). This change is relatively straightforward to implement with the author's software (see Appendix B). However, the resulting optimisation problem is significantly more difficult to solve, due to the large number of non−linear constraints now required.

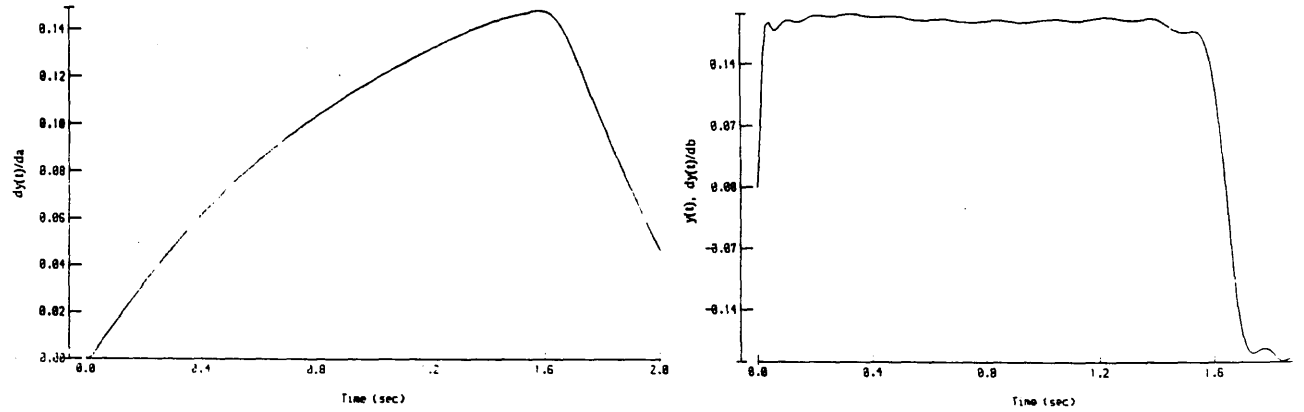
3.5.2 A Simulation Study of Response Robust Optimal Inputs

The simple scalar system (equation (3.28)) used to study previous constraints was also used to investigate the response robustness constraint given in (3.57). A response amplitude constraint of 0.2 units was used throughout with test records of 2 seconds duration. Figures 3.18 and 3.19 show the results obtained.

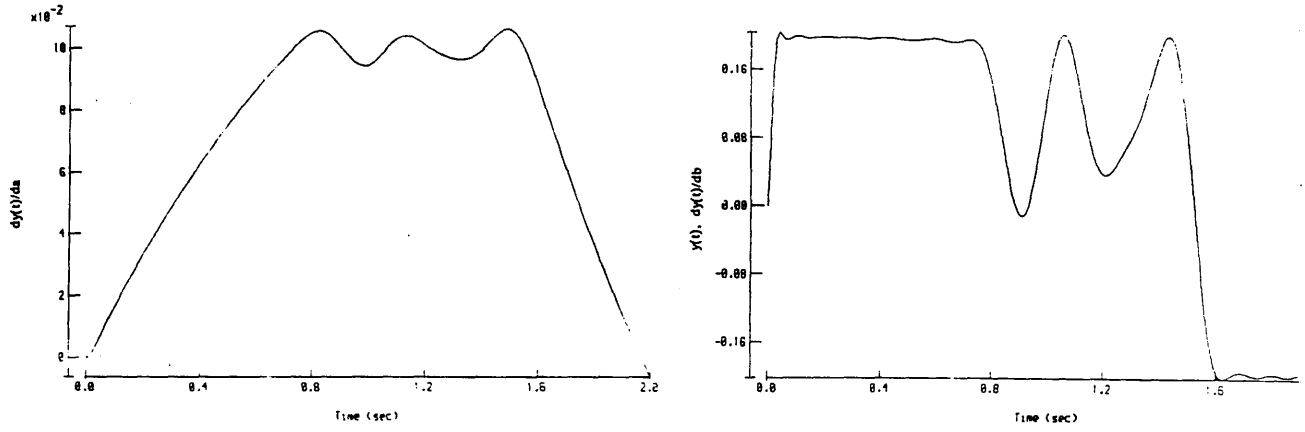
Figure 3.18

The response, y(t), and sensitivities, dy(t)/da and dy(t)/db, for output amplitude constrained (L_y=0.2) response robust (R_{ya}=0.2, 0.1, 0.05; R_{yb}=0.2) optimal inputs (simple first order system, a=−1, b=1)

R_{ya}= 0.2



R_{ya}= 0.1



R_{ya}= 0.05

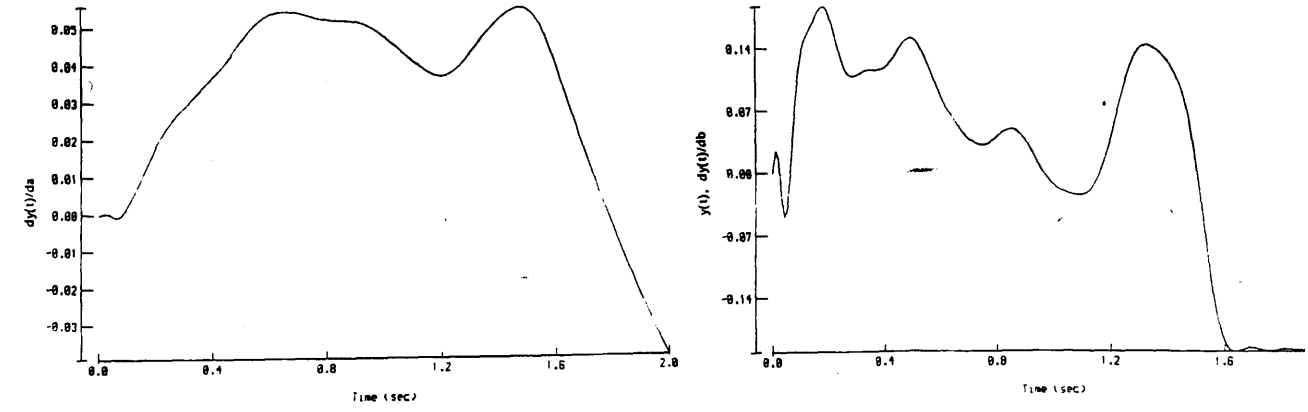
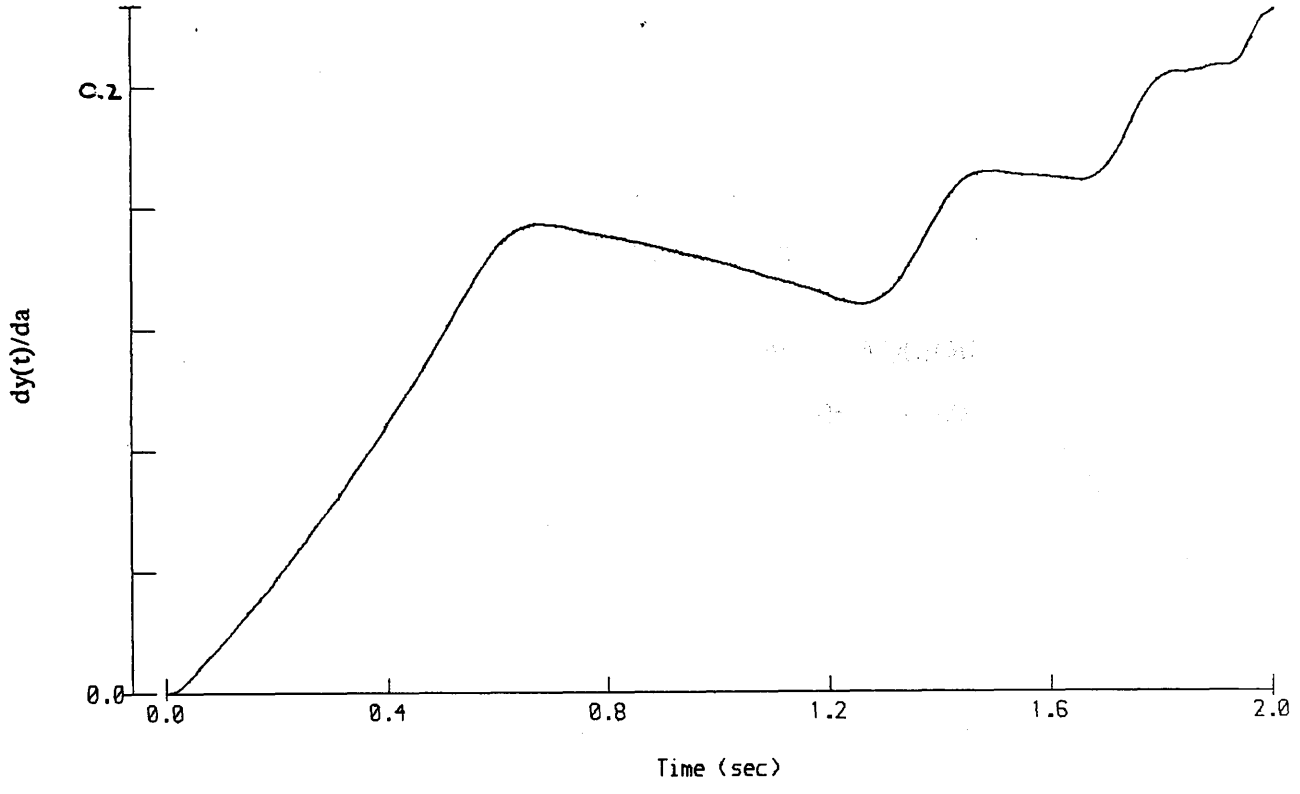


Figure 3.19

The response, y(t), and sensitivities, dy(t)/da and dy(t)/db, for output amplitude constrained (L_y=0.2) response robust (R_{ya}=0.2; R_{yb}=0.2) optimal inputs (simple first order system, a=+1, b=1)



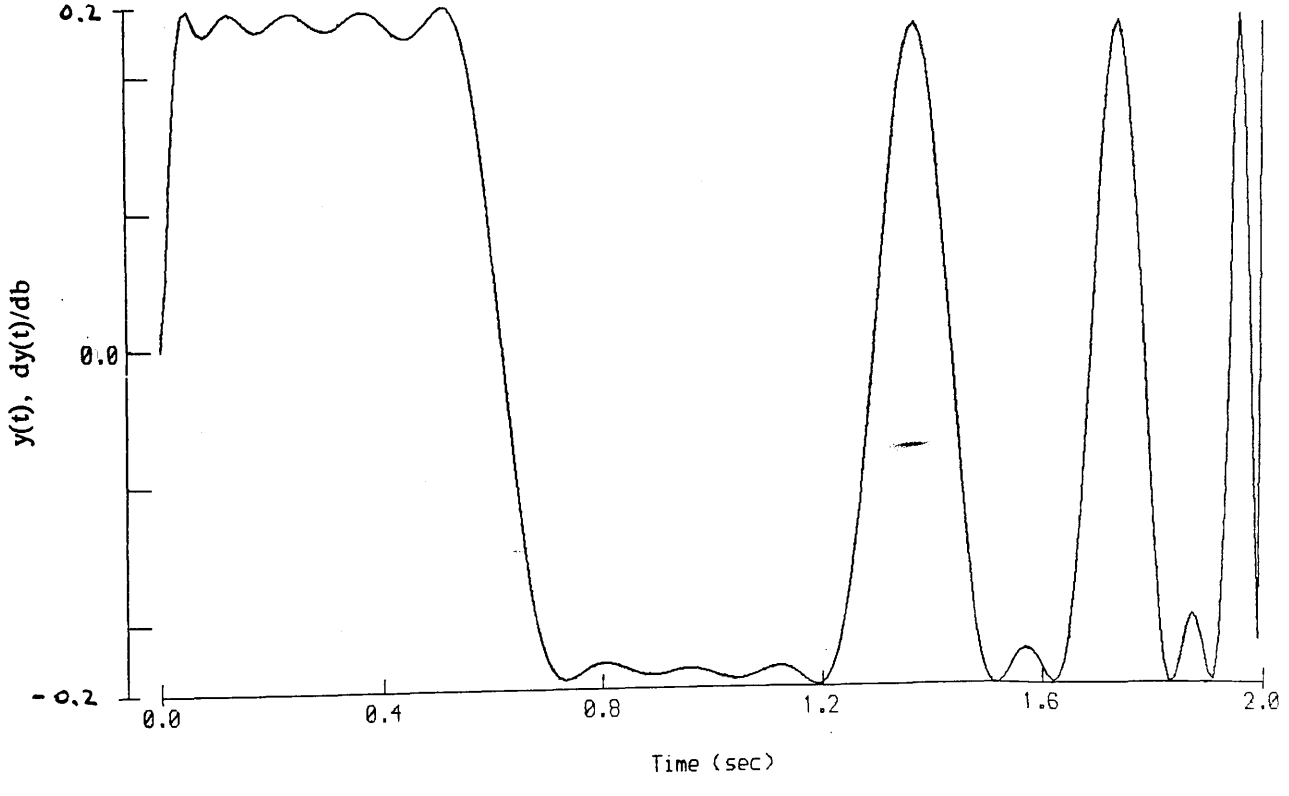
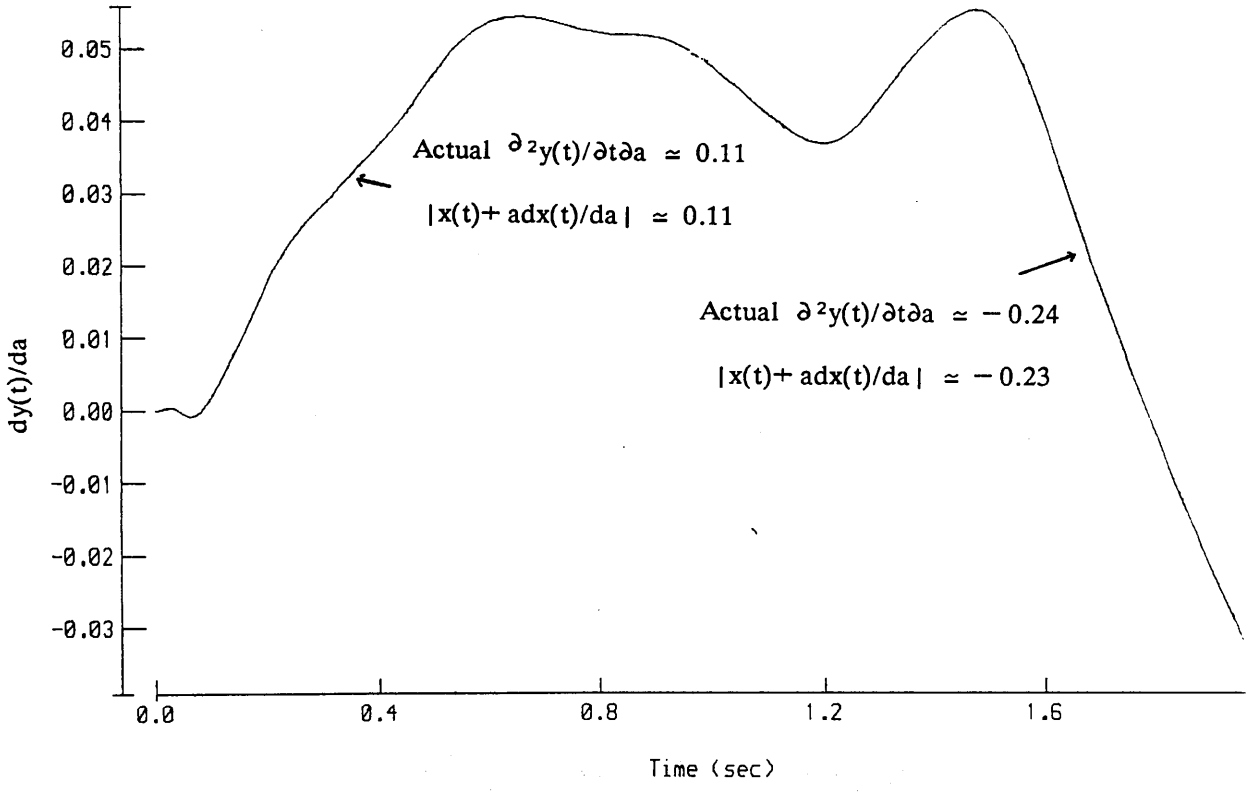
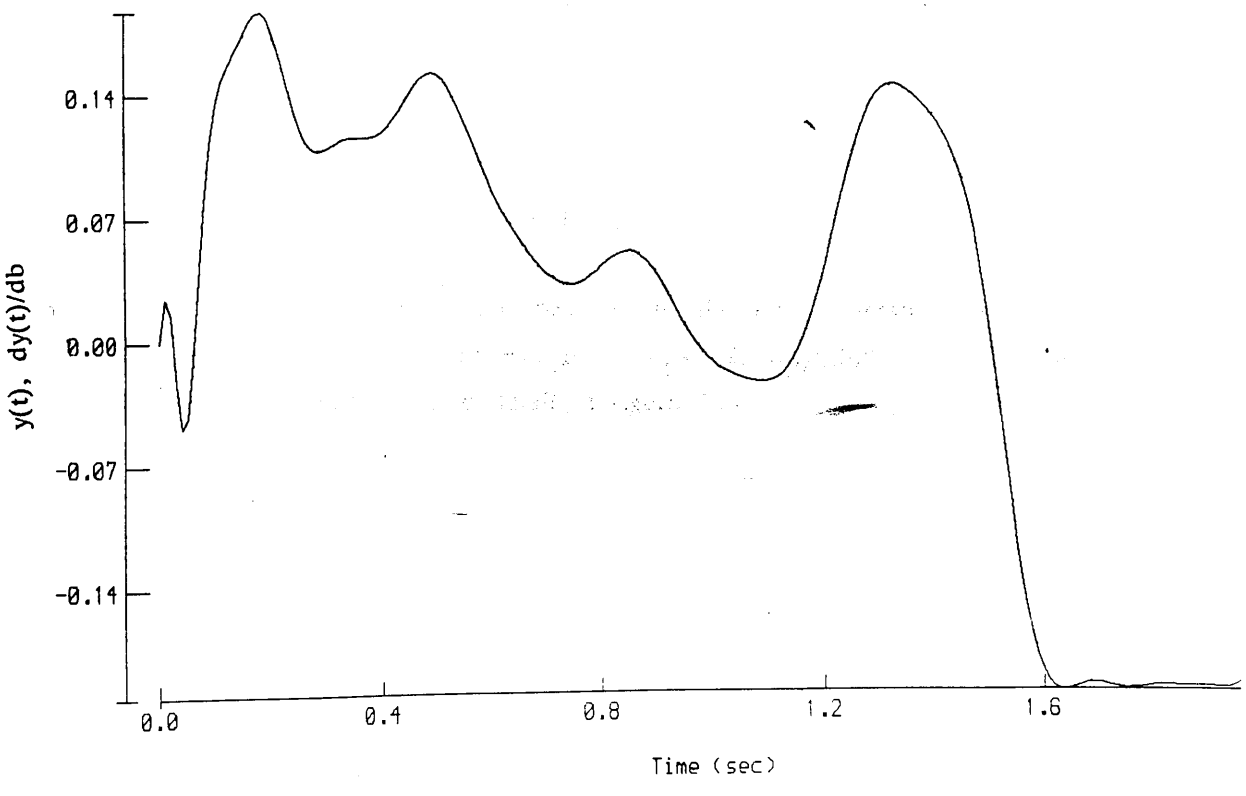


Figure 3.20

The response, y(t), and sensitivities, dy(t)/da and dy(t)/db, for the output amplitude constrained (L_y=0.2) response robust (R_{ya}=0.05; R_{yb}=0.2) optimal input (simple first order system, a=−1, b=1)





Recall that $dy(t)/db = y(t)$ in the simple scalar system. $dy(t)/db$ is therefore constrained by the output amplitude constraint, and so attention was concentrated on $dy(t)/da$. From the results, it can be seen that as R_{ya} is reduced, the sensitivity $dy(t)/da$ changes to remain within the constraint. The response robustness constraint also interacts with the output amplitude constraint to restrict $\partial^2 x(t)/\partial t \partial a$. To see this, consider the output–error model, as follows.

$$\frac{dx}{dt}(t) = A\,x(t) + B\,u(t), \qquad y(t) = C\,x(t) \tag{3.60}$$

For a parameter, θ_A, which is in the matrix A,

$$\frac{\partial^2 x}{\partial t\,\partial \theta_A}(t) = \frac{dA}{d\theta_A}x(t) + A\frac{dx}{d\theta_A}(t) \tag{3.61}$$

Since $|x(t)|$ and $|dx(t)/d\theta_A|$ are constrained, $|\partial^2 x(t)/\partial t \partial \theta_A|$ is also constrained. In particular, for the simple scalar system used, this gives,

$$|x(t)| = |y(t)| \leqslant L_y \tag{3.62}$$

$$\left|\frac{dx}{da}(t)\right| = \left|\frac{dy}{da}(t)\right| \leqslant R_{y_a} \tag{3.63}$$

and,

$$\left|\frac{\partial^2 x}{\partial t\,\partial \theta_A}(t)\right| = \left|\frac{\partial^2 y}{\partial t\,\partial \theta_A}(t)\right| = \left|x(t) + a\frac{dx}{da}(t)\right| \tag{3.64}$$

$$\leqslant |x(t)| + \left|a\frac{dx}{da}(t)\right| \tag{3.65}$$

i.e.

$$\left|\frac{\partial^2 x}{\partial t\,\partial \theta_A}(t)\right| \leqslant L_y + |a|\,R_{y_a} \tag{3.66}$$

The effect of this constraint can be seen in the results obtained using this system. In particular, the case $a=-1$, $b=1$, $L_y=0.2$, $R_{ya}=0.05$, $R_{Di}\to\infty$ shown in figure 3.18 can be seen in more detail in figure 3.20.

Now consider a parameter, θ_B, which is in the matrix B.

$$\frac{\partial^2 x}{\partial t\, \partial \theta_B}(t) = A \frac{dx}{d\theta_B}(t) + \frac{dB}{d\theta_B} u(t) \qquad (3.67)$$

However, unlike the case of a a parameter in A, this equation does not place an amplitude constraint on $\partial^2 x(t)/\partial t \partial \theta_B$, since the input, u(t), is not amplitude constrained. Nevertheless, $dx(t)/d\theta_B$ is coupled to x(t), and hence the response robustness and output amplitude constraints affect each other, as follows.

Rewrite (3.60) and (3.67) in state space form to give,

$$\begin{bmatrix} \frac{dx}{dt}(t) \\ \frac{\partial^2 x}{\partial t\, \partial \theta_B}(t) \end{bmatrix} = \begin{bmatrix} A & 0 \\ 0 & A \end{bmatrix} \begin{bmatrix} x(t) \\ \frac{dx}{d\theta_B}(t) \end{bmatrix} + \begin{bmatrix} B \\ \frac{dB}{d\theta_B} \end{bmatrix} u(t) \qquad (3.68)$$

The controllability matrix [9] for this system is,

$$\begin{bmatrix} B & AB \\ \frac{dB}{d\theta_B} & A\frac{dB}{d\theta_B} \end{bmatrix}$$

The columns of this matrix are clearly linearly dependant i.e. it is not full rank. Hence, x(t) and $dx(t)/d\theta_B$ cannot be controlled independantly. This is easily seen in the case of the simple scalar system.

$$\begin{bmatrix} \frac{dx}{dt}(t) \\ \frac{\partial^2 x}{\partial t\, \partial b}(t) \end{bmatrix} = \begin{bmatrix} a & 0 \\ 0 & a \end{bmatrix} \begin{bmatrix} x(t) \\ \frac{dx}{db}(t) \end{bmatrix} + \begin{bmatrix} b \\ 1 \end{bmatrix} u(t) \qquad (3.69) .$$

If b=1, then x(t) = dx(t)/db, and more generally, x(t) = b dx(t)/db.

Such coupling may also be present between x(t) and $dx(t)/d\theta_A$. Combining (3.60) and (3.61) gives,

$$\begin{bmatrix} \frac{dx}{dt}(t) \\ \frac{\partial^2 x}{\partial t\, \partial \theta_A}(t) \end{bmatrix} = \begin{bmatrix} A & 0 \\ \frac{dA}{d\theta_A} & A \end{bmatrix} \begin{bmatrix} x(t) \\ \frac{dx}{d\theta_A}(t) \end{bmatrix} + \begin{bmatrix} B \\ 0 \end{bmatrix} u(t) \qquad (3.70)$$

The controllability matrix is,

$$\begin{bmatrix} B & AB \\ 0 & \frac{dA}{d\theta_A} B \end{bmatrix}$$

If the matrix A has dimension greater than one, then $dA/d\theta_A$ will contain rows and columns of zeroes. In this case, the controllability matrix will once again have less than full rank.

However, for the simple scalar system, this controllability matrix is,

$$\begin{bmatrix} b & ab \\ 0 & b \end{bmatrix}$$

This has determinant equal to b^2, which is non-zero for $b \neq 0$. Hence, x(t) and dx(t)/da are controllable for $b \neq 0$.

When choosing values for L_y and R_{yi} it is important that the couplings between the constraints be taken into account. Otherwise, the constraints will be in conflict, giving a very ill-conditioned optimisation problem.

Finally, considering the response robustness conflict itself, it can be seen that this produces only 'short-range' robustness. For the model used to design the input, $dy(t)/d\theta$ may be small. However, there is no guarantee that when perturbed from this model $dy(t)/d\theta$ will remain small. Hence, this response robustness constraint is relatively limited.

Unfortunately, wider–range constraints, which continue to hold when perturbed to a certain extent from the designed model, are difficult to implement. For example, constraining the higher derivatives of $dy(t)/d\theta$, as well as $dy(t)/d\theta$ itself, will produce a wider–range constraint. However, differentiation is numerically unstable, leading to accuracy problems. Moreover, generating a large number of derivatives is extremely expensive in computing time. It is felt that the simple constraint used provides a practical compromise between no robustness and these more complex robustness constraints.

3.6 PARAMETER ROBUST OPTIMAL INPUTS

3.6.1 Time Domain Designs

In addition to obtaining robust responses, it is also important that the parameter estimates continue to be of high quality when an input is used with the real system. To give robust parameter estimates, it is possible to limit the sensitivity of $|D|$ to changes in the parameters, using the following constraint.

$$\frac{\theta_i}{|D|}\frac{d|D|}{d\theta_i} \leqslant R_{D_i} \qquad i = 1, 2, \ldots q \qquad (3.71)$$

The set of allowable inputs is then given by,

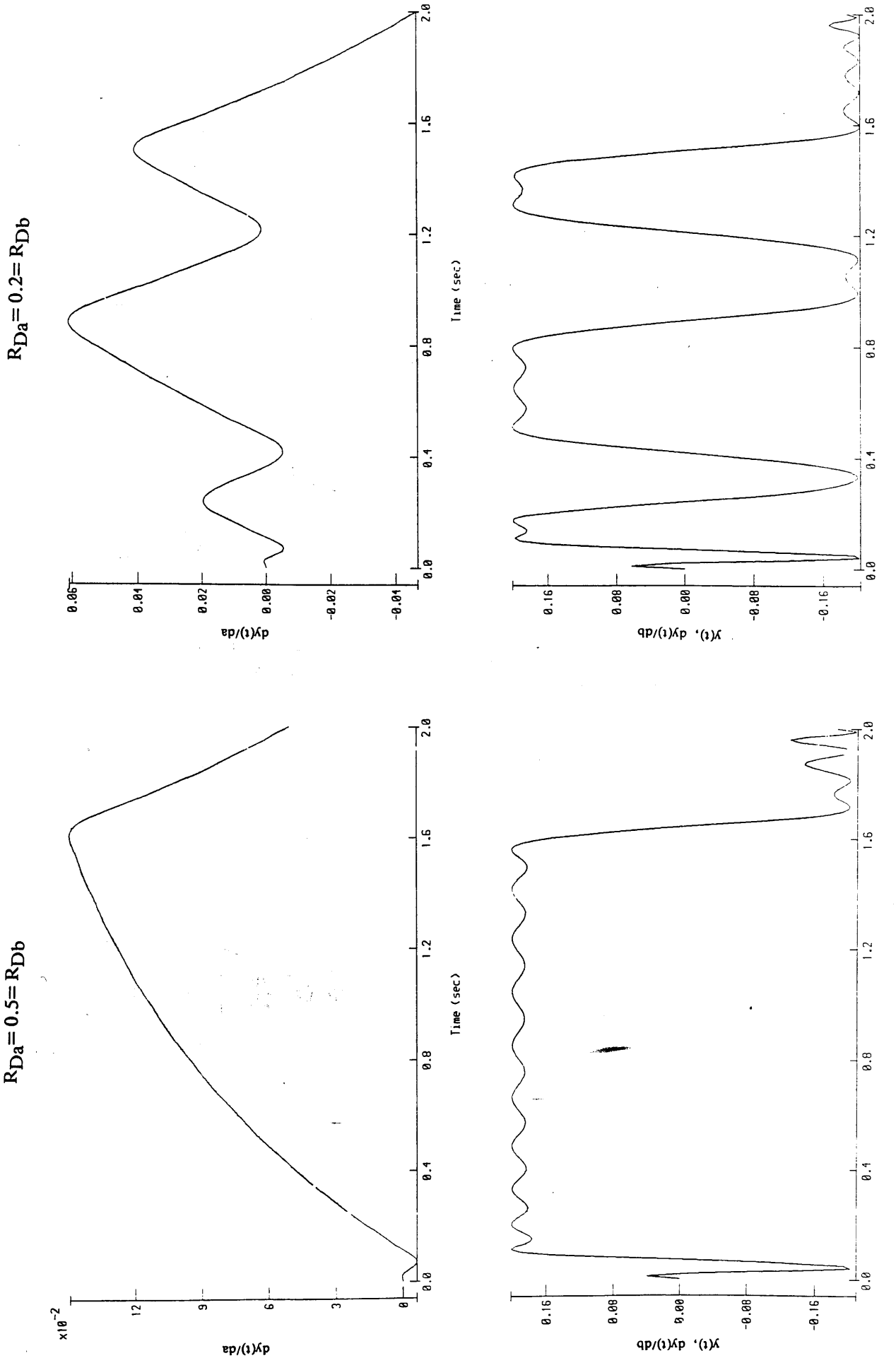
$$U = \left\{ u : |y(t)| \leqslant L_y \; ; \; \frac{\theta_i}{|D|}\frac{d|D|}{d\theta_i} \leqslant R_{D_i} \right\} \qquad (3.72)$$

with $i = 1, 2, \ldots q;\ 0 \leqslant t \leqslant T$

Unfortunately, the parameter robustness constraint in (3.71) is very non–linear. When combining two inputs $u_n \epsilon U$ and $u_0 \epsilon U$ to give a new input $u_{n+1} = \alpha u_0 + \beta u_n$, no simple relationship between α and β has been found to ensure that $u_{n+1} \epsilon U$. It has therefore not been possible to use the special form of input design algorithm developed for the previous constraints. Instead, a general–purpose optimisation algorithm was used, but at the cost of poorer performance (see Appendix B).

Figure 3.21

The response, y(t), and sensitivities, dy(t)/da and dy(t)/db, for output amplitude constrained (L_y=0.2) parameter robust (R_{Da}=0.5=R_{Db} and R_{Da}=0.2=R_{Db}) optimal inputs (simple first order system, a=−1, b=1)



3.6.2 A Simulation Study of Parameter Robust Optimal Inputs

The results of a study of the parameter robustness constraint in equation (3.71) are shown in figure 3.21. The simple scalar system described previously was used, with a response amplitude constraint of 0.2 units, and test records of 2 seconds duration. It can be seen that while y(t) and dy(t)/db are bang–bang in nature, dy(t)/da is not.

The parameter robustness constraint may be written as follows,

$$\frac{\theta}{|D|}\frac{d|D|}{d\theta} = -\frac{\theta}{|M|}\frac{d|M|}{d\theta} = -\theta\ \mathrm{Tr}\left(M^{-1}\frac{dM}{d\theta}\right) \qquad (3.73)$$

For the scalar system used, the information matrix, M, is a 2x2 matrix, i.e.

$$M = \begin{bmatrix} M_{11} & M_{12} \\ M_{12} & M_{22} \end{bmatrix} \qquad (3.74)$$

$$\Rightarrow -\frac{\theta}{|M|}\frac{d|M|}{d\theta}$$

$$= -\theta\ \mathrm{Tr}\left[\frac{1}{M_{11}M_{22} - M_{12}^2}\begin{bmatrix} M_{11} & M_{12} \\ M_{12} & M_{22} \end{bmatrix}\begin{bmatrix} \frac{dM_{11}}{d\theta} & \frac{dM_{12}}{d\theta} \\ \frac{dM_{12}}{d\theta} & \frac{dM_{22}}{d\theta} \end{bmatrix}\right]$$

$$= \frac{-\theta}{M_{11}M_{22} - M_{12}^2}\left[M_{22}\frac{dM_{11}}{d\theta} - 2M_{12}\frac{dM_{12}}{d\theta} + M_{11}\frac{dM_{22}}{d\theta}\right] \qquad (3.75)$$

and,

$$M_{11} = \int_0^T \left(\frac{dy}{da}\right)^2(t)\,dt \qquad (3.76)$$

$$M_{12} = \int_0^T \frac{dy}{da}(t)\,\frac{dy}{db}(t)\,dt \qquad (3.77)$$

$$M_{22} = \int_0^T \left(\frac{dy}{db}\right)^2(t)\,dt \qquad (3.78)$$

Hence, the parameter robustness constraint is essentially constraining the energy of the sensitivities, and the correlation between them.

3.7 COMBINED PARAMETER ROBUST AND RESPONSE ROBUST OPTIMAL INPUTS

To gain the advantages of both parameter robustness and response robustness, both of these constraints may be applied simultaneously when designing an input. To ensure linearity, the output amplitude constraint is also needed. This gives the following set, U, of valid inputs.

$$U = \left\{ u : |y(t)| \leqslant L_y \; ; \; \left|\frac{dy}{d\theta_i}(t)\right| \leqslant R_{y_i} \; ; \; \frac{\theta_i}{|D|}\frac{d|D|}{d\theta_i} \leqslant R_{D_i} \right\} \qquad (3.79)$$

with $i = 1, 2, \ldots q;\ 0 \leqslant t \leqslant T$

Unfortunately, implementing this set of constraints was found to require excessive amounts of computing time (typically around a week of CPU time on a DEC VAX 11/750). The simultaneous use of response robustness and parameter robustness impractical at present, until significantly more efficient software is developed, or more computing power is available. It is therefore necessary to decide which of these two constraints should be used. However, this choice depends on the application being considered.

3.8 CONCLUSIONS

The results obtained have emphasised the importance of using appropriate constraints during the input design process. If care is not taken, then inputs may be produced which are unsuitable for practical use. It is suggested that this may account for some of the poor results reported in the literature when optimal inputs have been used in practice for aerospace applications (see section 1.5.2).

In order to ensure linear responses, the output amplitude constraint may be used. This constraint guarantees response of limited amplitude. If a small enough amplitude is chosen, then the responses will be linear.

Robustness is also an important factor in obtaining a successful input. While both response robustness and parameter robustness constraints were studied, it was found that it is impractical at present to use both constraints simultaneously. It is therefore necessary to decide which of these constraints should be used. This choice depends on the application under consideration. However, in the helicopter case, if the response becomes non-linear, then it is not suitable for identification purposes. It is therefore essential that linear responses are obtained. In the present work, it is thus suggested that response robustness is the more important.

To conclude, the following set, U, of permissible inputs is advocated for rotorcraft applications :

$$U = \left\{ u : |y(t)| \leqslant L_y \; ; \; \left|\frac{dy}{d\theta_i}(t)\right| \leqslant R_{y_i} \right\} \qquad (3.80)$$

with $i = 1, 2, \ldots q; \quad 0 \leqslant t \leqslant T$

REFERENCES FOR CHAPTER THREE

1. Pronzato, L.; Walter, E. 'Robust experiment design via stochastic approximation' Math Bio Vol 75, pp103–120, 1985

2. Pronzato, L.; Walter, E. 'Robust experiment design via maximin optimisation' Math Bio Vol 89, pp161–176, 1988

3. Mehra, R.K. 'Optimal input signals for parameter estimation in dynamic systems – survey and new results' IEEE Trans Aut Control, Vol AC–19, 1974, pp753–768

4. Mehra, R.K. 'Frequency–domain synthesis of optimal inputs for linear system parameter estimation' Trans ASME (J. Dynamic Systems, Meas, and Control) Vol 98, 1976, pp130–138

5. Fedorov, V.V. 'Theory of optimal experiments' Academic Press, London 1972

6. Atwood, C.L 'Sequences converging to D–optimal designs of experiments' Annals of Stats, Vol 1, 1973, pp342–352

7. Zarrop, M.B. 'Sequential generation of D–optimal input designs for linear dynamic systems' J. Opt Theory Apps, Vol 35, 1981, pp277–291

8. Gupta, N.; Hall, W. 'Input design for identification of aircraft stability and control derivatives' NASA CR–2493, 1975

9. Richards, R.J. 'An introduction to dynamics and control' Longman, New York 1979

CHAPTER FOUR

DESIGN OF OPTIMAL LYNX INPUT

4.1 INTRODUCTION

The results obtained in Chapter 3, concerning suitable constraints for identification inputs, were applied to the design of optimal rotorcraft inputs. Attention was restricted to identifying the pitching moment equation parameters: M_u, M_w, M_q, and $M_{\eta_{1S}}$. In addition, the 5^{th} order model derived from output–error identification in section 2.4.3 was used as the basis for the input design process. Finally, the covariance matrix, R, of the noise on the outputs was estimated from flight test data to be as follows :

$$R = \begin{bmatrix} 0.0025 & & & & & \\ & 0.0085 & & & 0 & \\ & & 0.00001 & & & \\ & & & 0.00025 & & \\ & 0 & & & & \\ & & & & & 0.000036 \end{bmatrix}$$

Two inputs were studied : one subject only to an output amplitude constraint, and the other subject to both output amplitude and response robustness constraints. The amplitude constraints on the outputs were as follows :

longitudinal velocity, u	...	5 m/s (15 ft/sec)
vertical velocity, w	...	3.3 m/s (10 ft/sec)
pitch rate, q	...	0.04 rads/sec
pitch angle, θ'	...	0.1 rads
roll rate, p	...	0.03 rads/sec

The amplitudes of the outputs, with the exception of roll rate, p, were chosen on the basis of previous flight data, known from its coherence functions to be linear. Roll rate, however, was kept to a lower amplitude than that required simply for linearity. This was to allow for errors in the modelling of this state, and was used to reinforce the response robustness constraints since the 5^{th} order model used contains only a limited representation of the rotorcraft lateral dynamics. The accuracy of the modelling of roll rate is discussed further in section 2.4.3.

It was not possible to use the coherence to check the linearity of the responses obtained using these contraints, since only a linear identified model was available. Nevertheless, experience suggests that the amplitudes used are reasonable.

For response robustness, the relative sensitivities, $|\theta/x \; dx/d\theta|$, were constrained to be around 1%. Taking the parameters, θ, from the model used, and the output amplitude constraints as the values for $|x|$, the absolute sensitivities, $|dx/d\theta| = \{|x|/|\theta|\} \, |\theta/x \; dx/d\theta|$, are given in table 4.1.

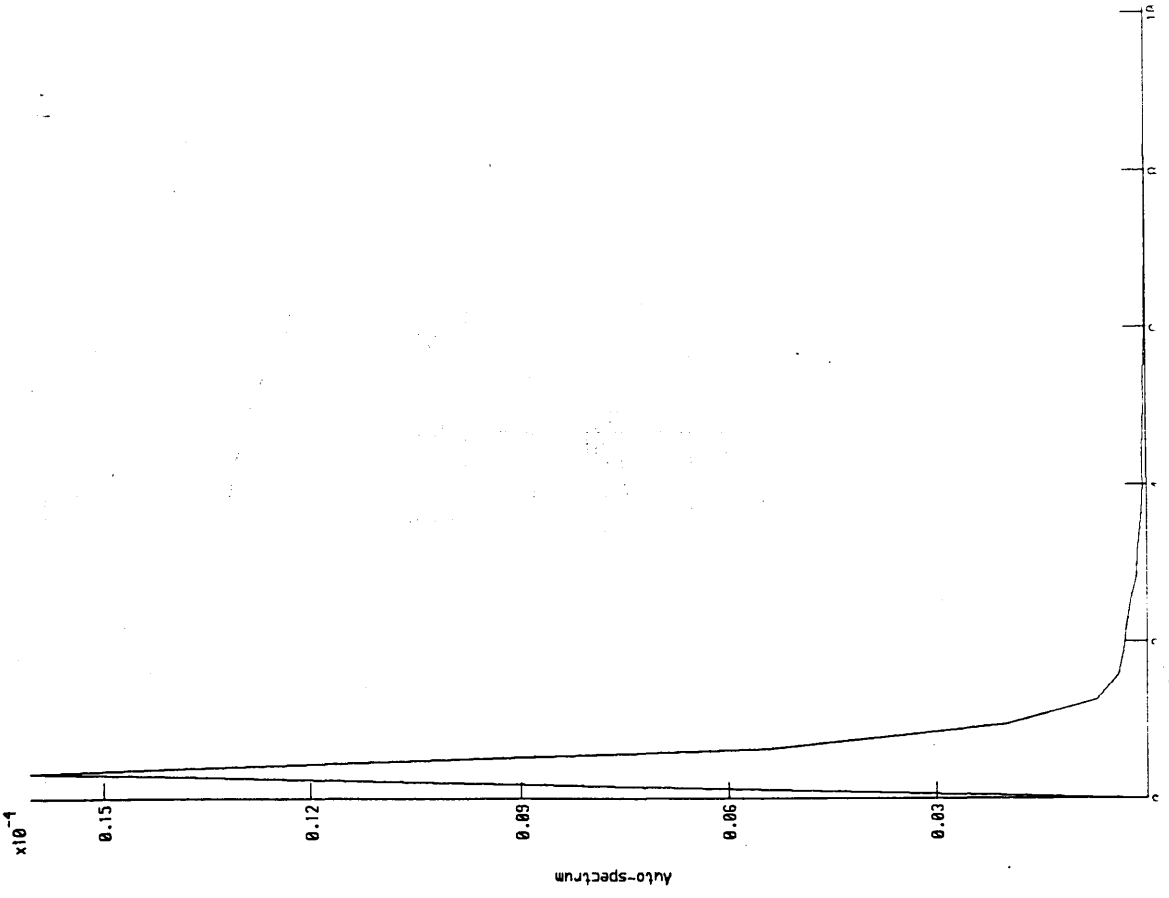
Sensitivity of outputs to changes in the model parameters	u	w	q	θ'	p
M_u	97	65	0.26	0.65	0.194
M_w	441	294	1.17	2.94	0.882
M_q	0.22	0.14	0.6E-3	0.14E-2	0.44E-3
$M_{\eta_{1s}}$	0.11E-1	0.73E-2	0.2E-4	0.7E-4	0.2E-4

Table 4.1 – *Response robustness constraints on $|dx/d\theta|$, used with the robust Lynx optimal input.*

Finally, a 20 second test record length was used. This corresponded with the typical lengths of records in flight trials using multi–step inputs, although it is shorter than that obtained with the double–doublet input in Chapter 2. Moreover, it produced a more manageable input design problem than if a longer record length had been used.

Figure 4.1

The time history and auto-spectrum of the output amplitude constrained optimal input for the Lynx helicopter at 80 knots level flight. Details of the model and the constraints used are given in the text.



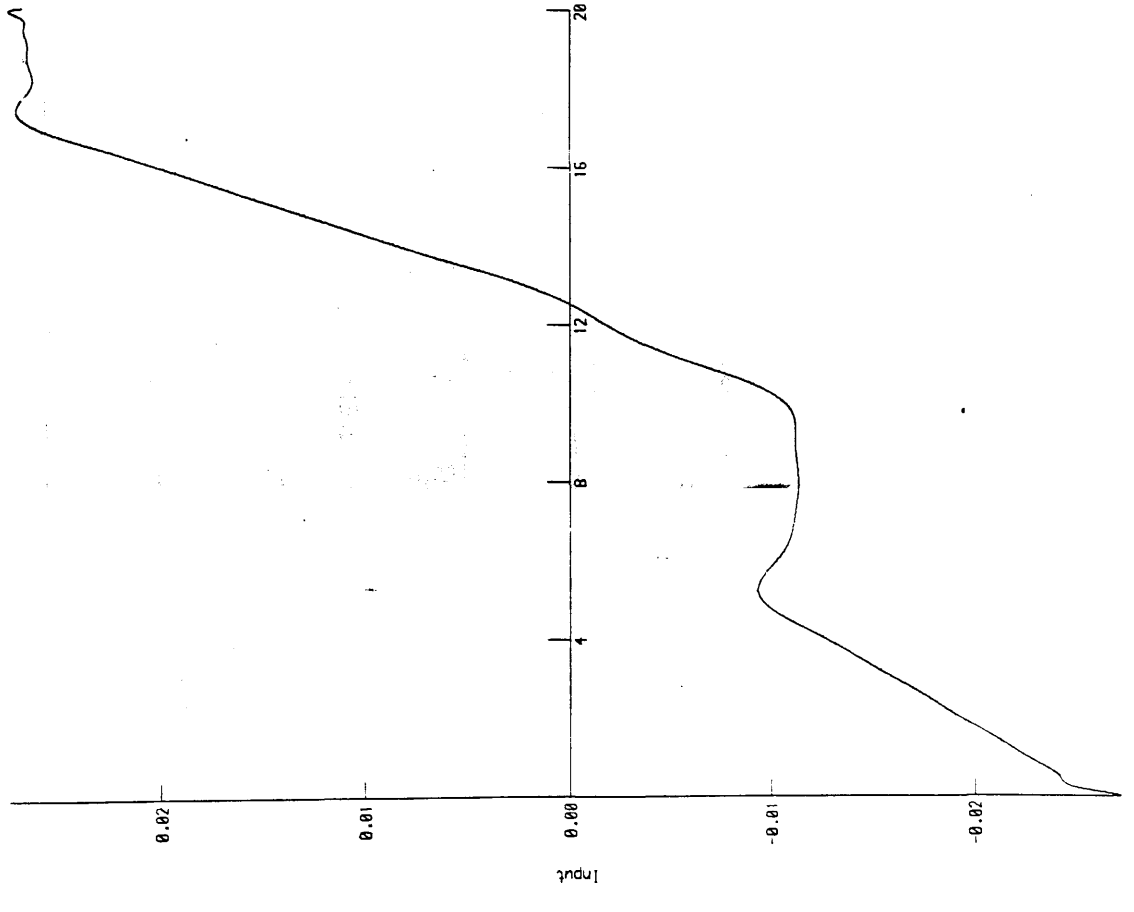


Figure 4.2

The responses produced by the output amplitude constrained optimal input for the Lynx helicopter at 80 knots level flight. Details of the model and the constraints used are given in the text.

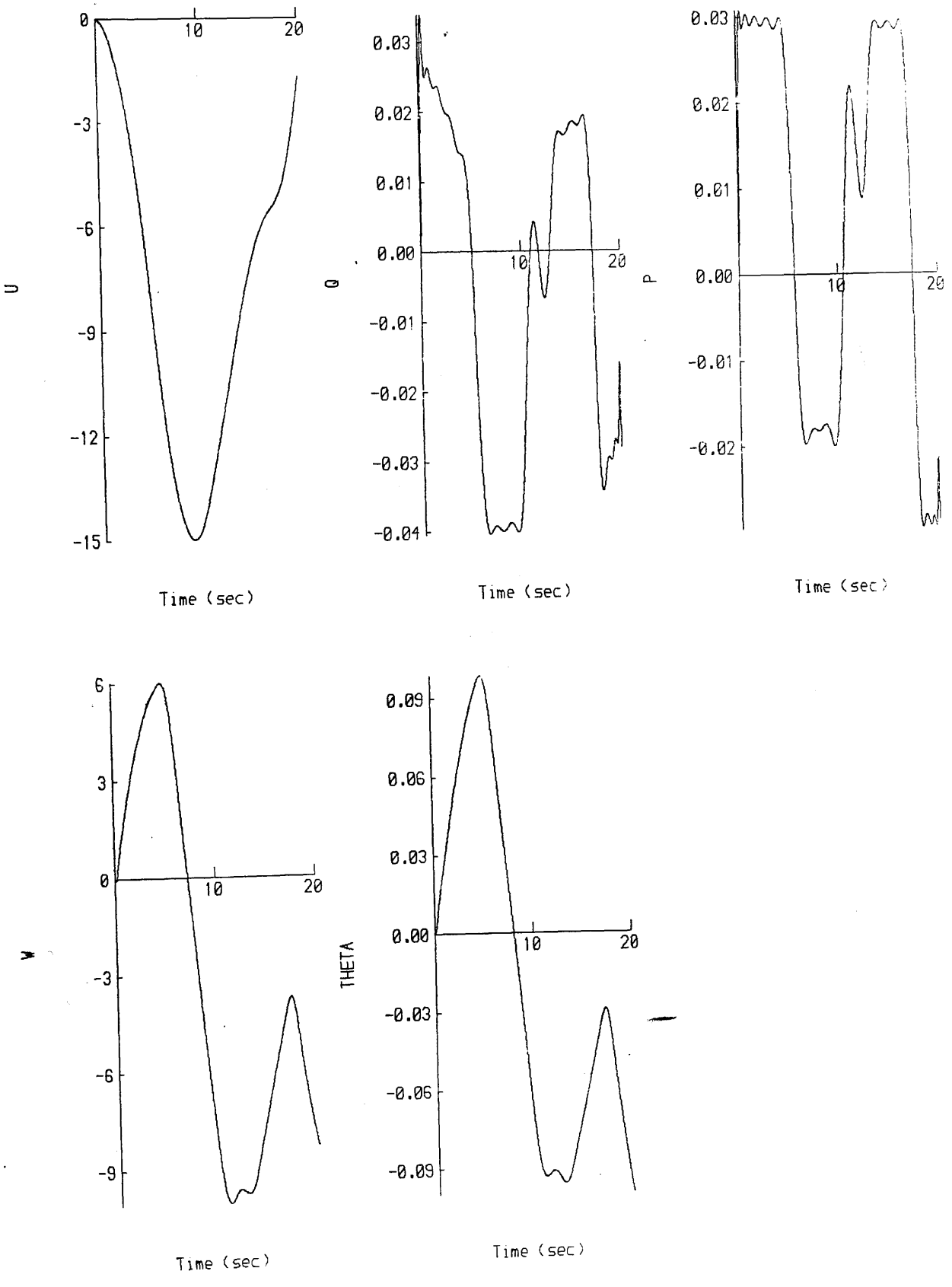


Figure 4.3

The auto–spectra of the responses produced by the output amplitude constrained optimal input for the Lynx helicopter at 80 knots level flight. Details of the model and the constraints used are given in the text.

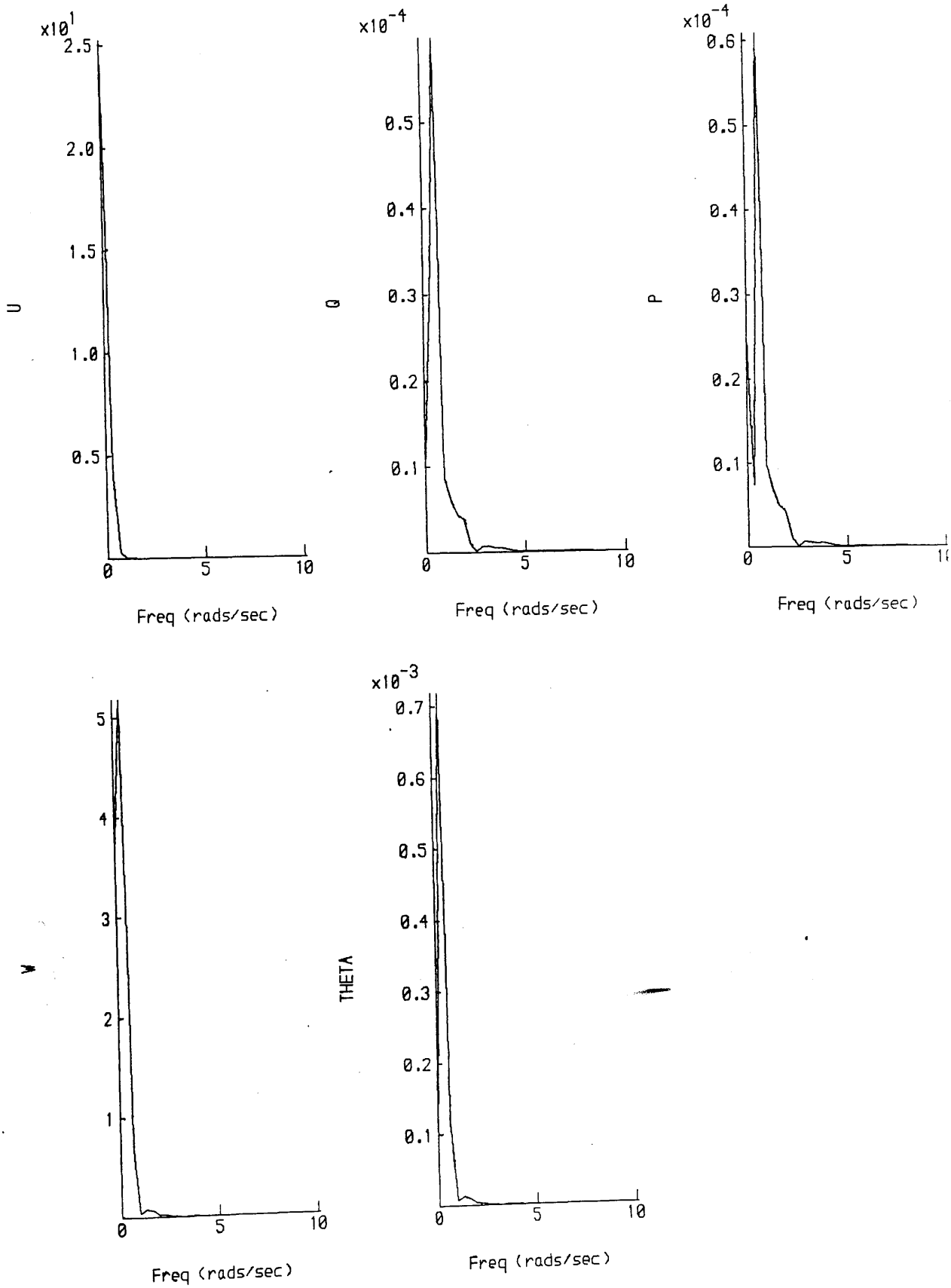


Figure 4.4 (a)

The sensitivities to the parameters of the pitching moment equation of the responses to the output amplitude constrained optimal input for the Lynx helicopter at 80 knots level flight.

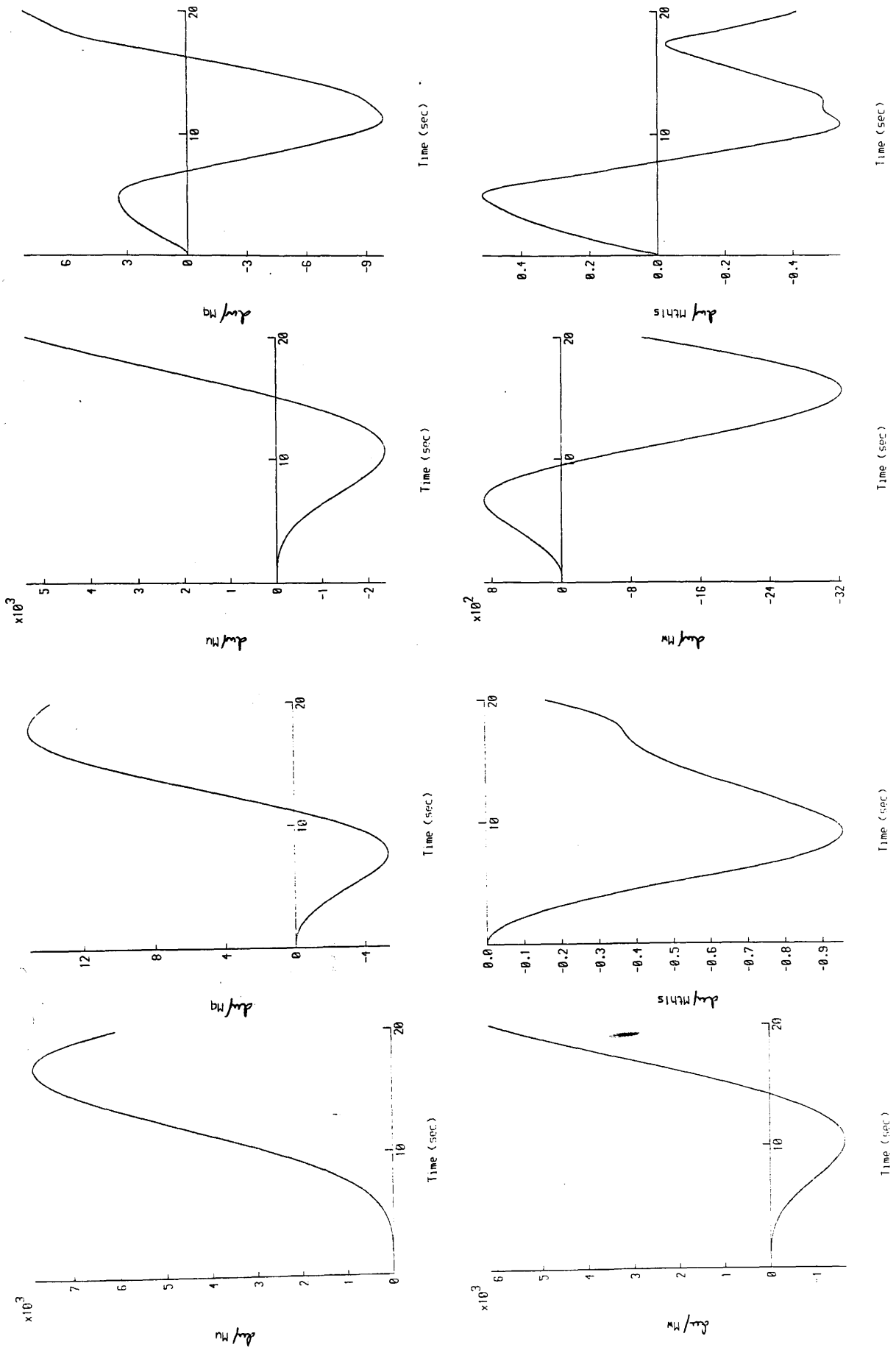


Figure 4.4 (b)

The sensitivities to the parameters of the pitching moment equation of the responses to the output amplitude constrained optimal input for the Lynx helicopter at 80 knots level flight.

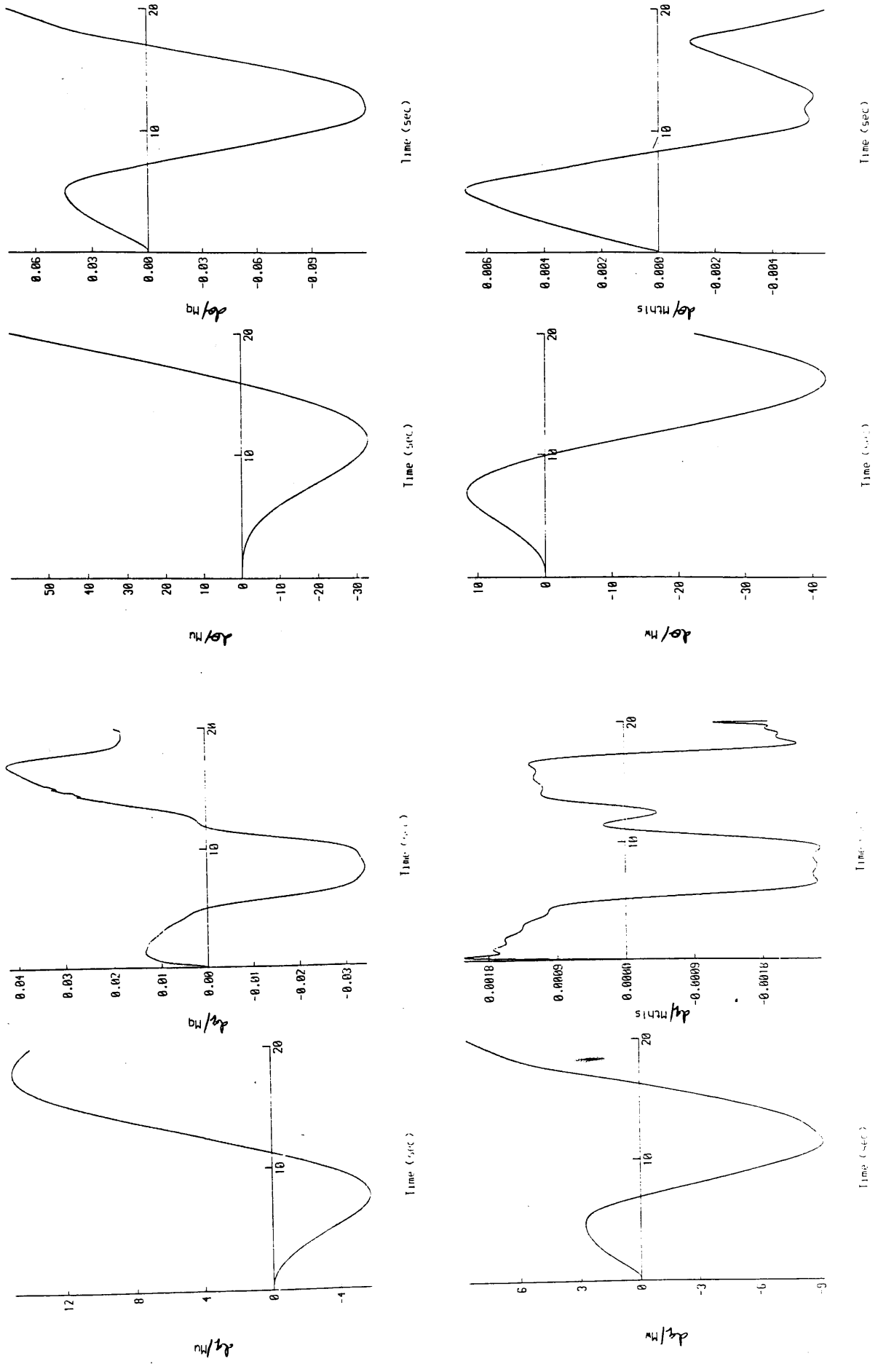


Figure 4.4 (c)

The sensitivities to the parameters of the pitching moment equation of the responses to the output amplitude constrained optimal input for the Lynx helicopter at 80 knots level flight.

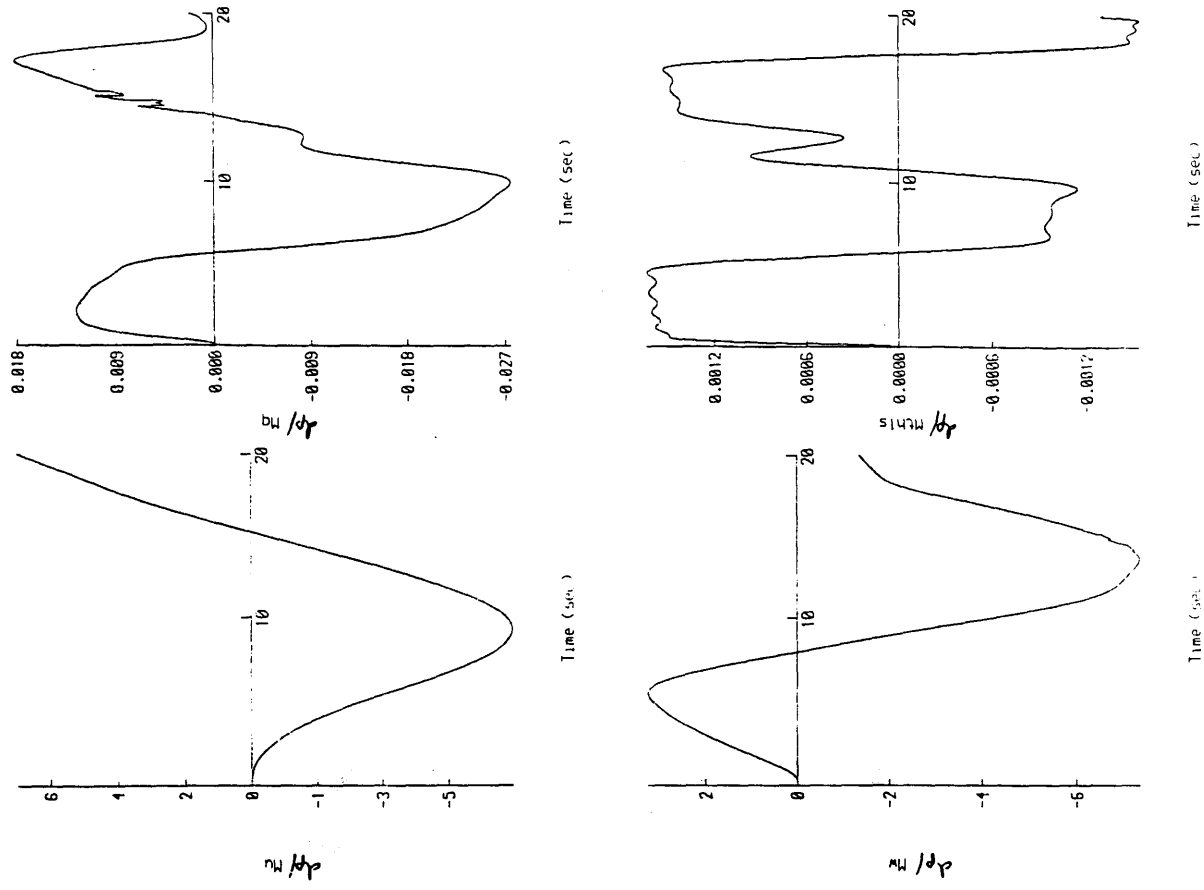


Figure 4.5 (a)

The sensitivities to the parameters of the pitching moment equation of the responses to the double–doublet input for the Lynx helicopter at 80 knots level flight.

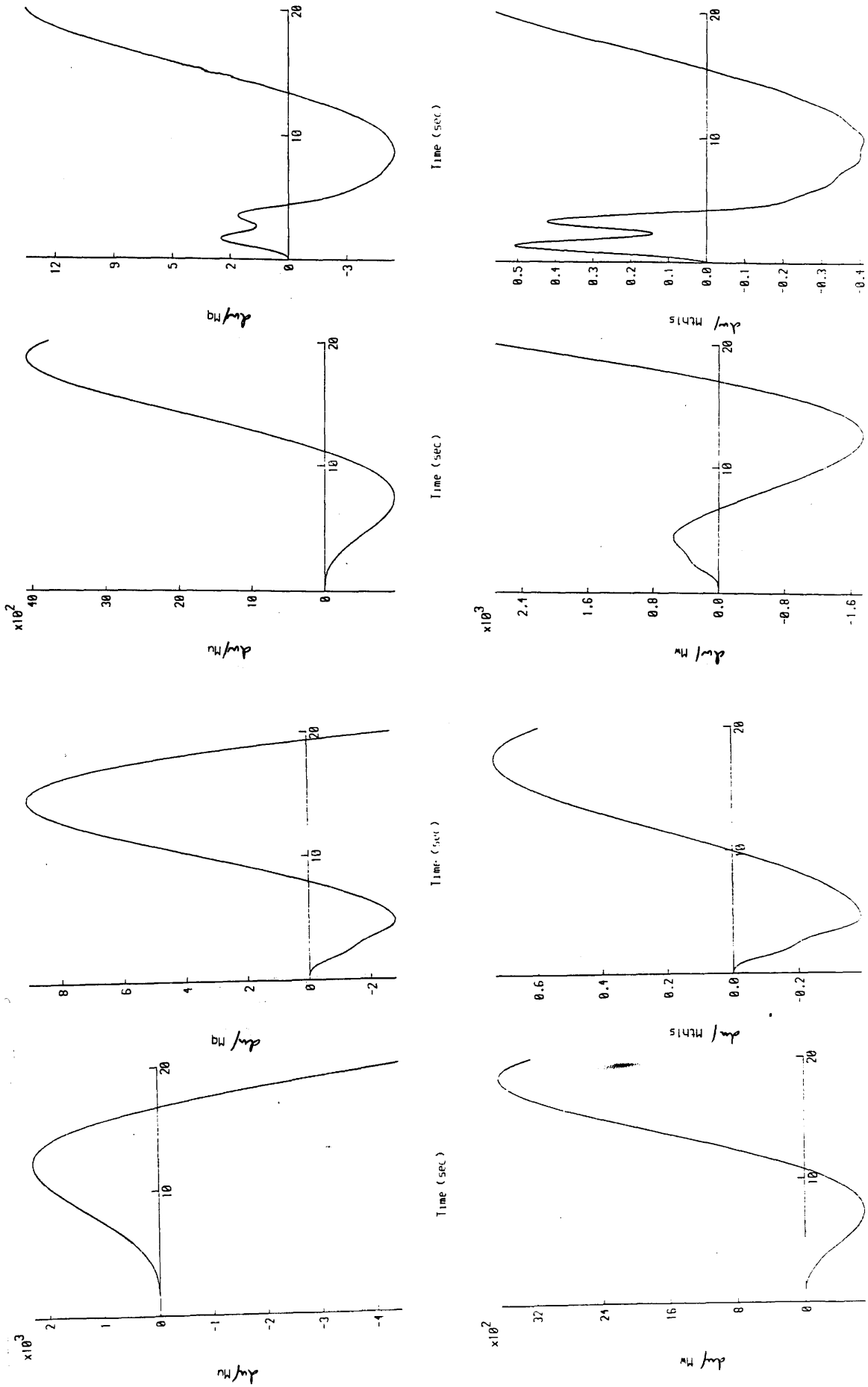


Figure 4.5 (b)

The sensitivities to the parameters of the pitching moment equation of the responses to the double–doublet input for the Lynx helicopter at 80 knots level flight.

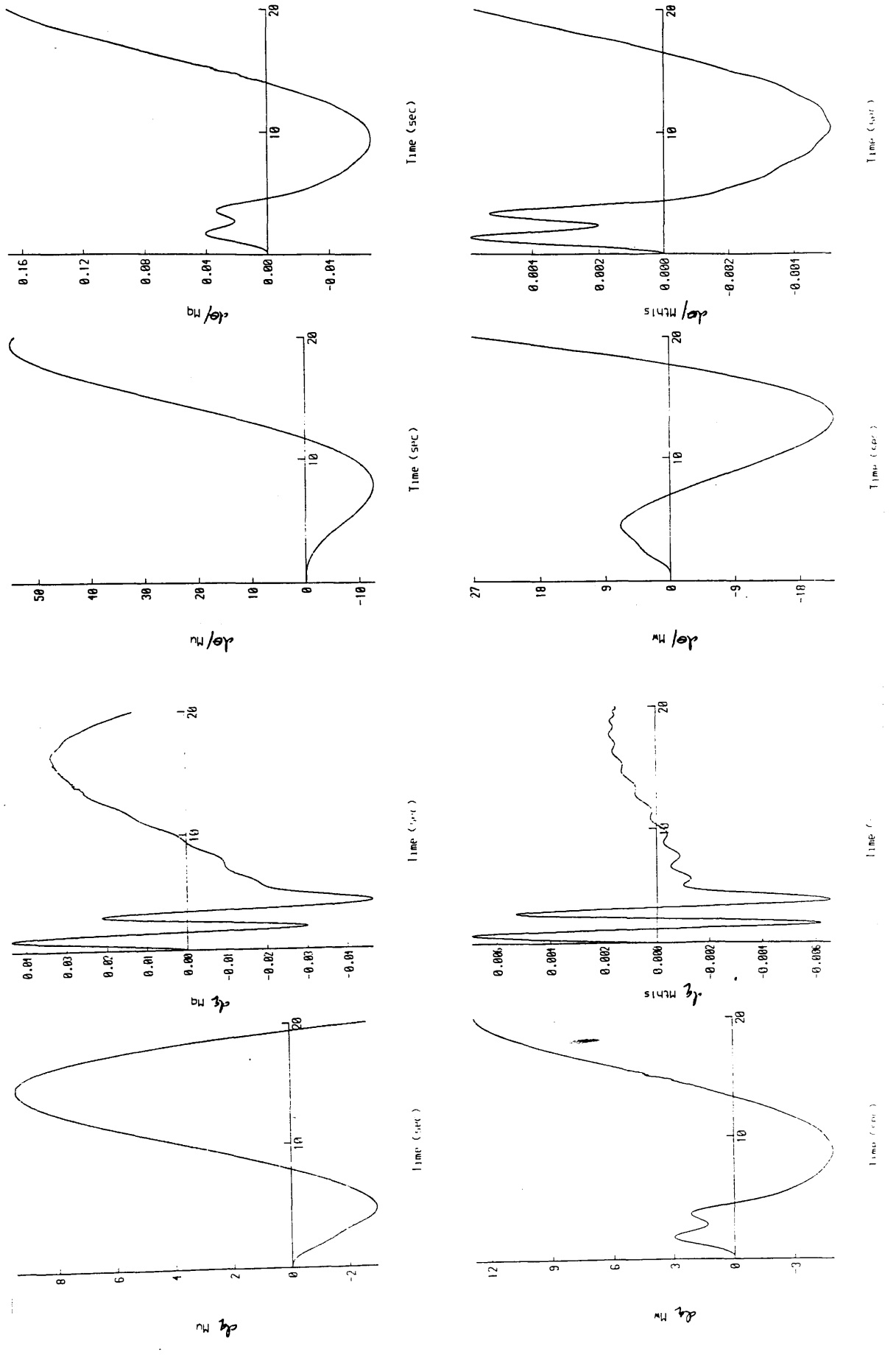
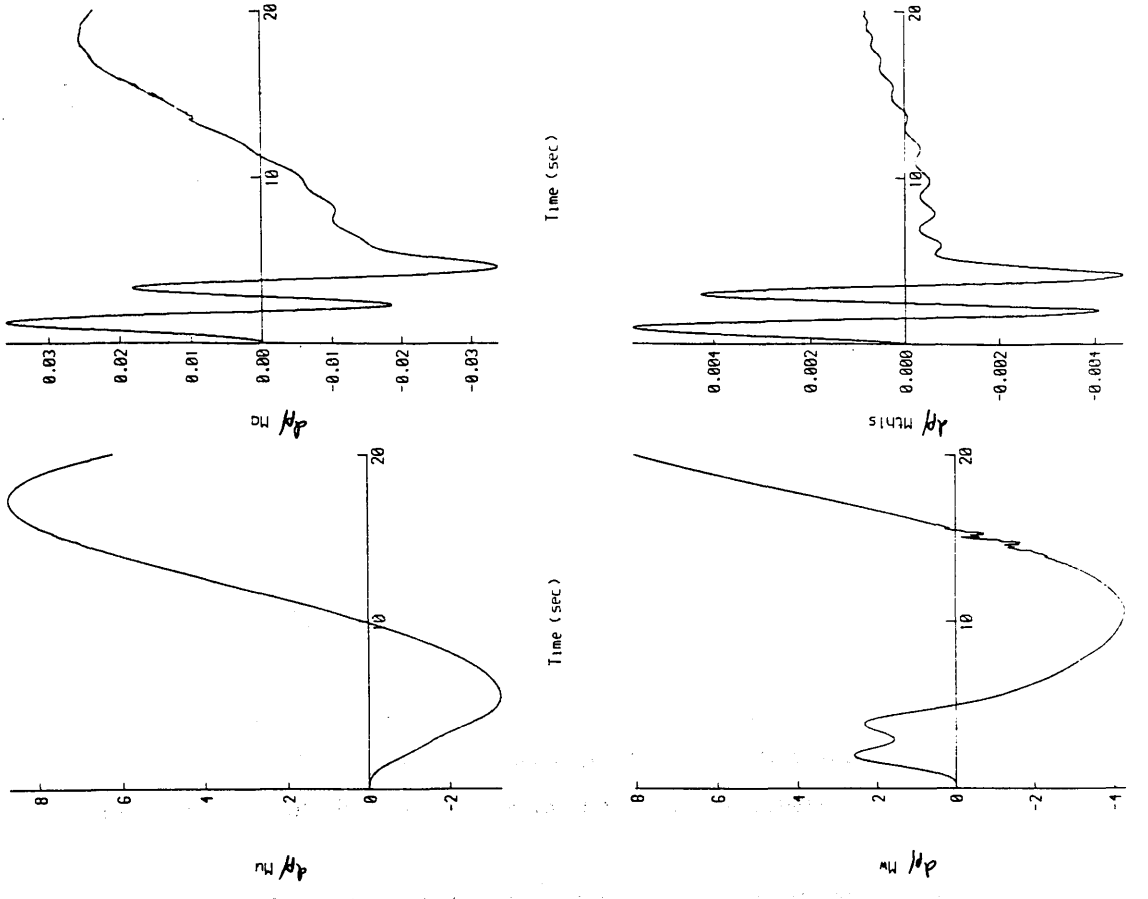


Figure 4.5 (c)

The sensitivities to the parameters of the pitching moment equation of the responses to the double–doublet input for the Lynx helicopter at 80 knots level flight.



While this is acceptable for the present work, deciding on a suitable length of test record is non–trivial in general. Perhaps the simplest approach is to increase the length until parameter estimates are obtained which have an acceptable level of variance. As shown in Chapter 3, for unstable system such as a helicopter, |D| does not converge to a final value. Hence, it is always possible to obtain a given variance if sufficiently long test records are used.

However, the response of the helicopter to external disturbances, such as gusts, can build up to significant level over a long test record. Related to this, the longer the test record, the more difficult it also is to obtain an input with acceptable response robustness. A degree of engineering judgement is therefore required when selecting what is a suitable length of test record to use.

4.2 OUTPUT AMPLITUDE CONSTRAINED OPTIMAL INPUT

The D–optimal, output amplitude constrained, Lynx input is shown in figure 4.1. Figures 4.2 and 4.3 give the responses produced by this input, and figure 4.4 shows the sensitivity of these responses to the model parameters.

It can be seen that the responses are amplitude constrained, as required, and exhibit a tendency for bang–bang behaviour. However, due to the need to constrain several outputs, and the couplings between these outputs, the responses are not truly bang–bang. This is in contrast to the simpler cases studied in Chapter 3.

An example of this coupling can be seen between the pitch angle, θ', and the pitch rate, q, as follows.

$$q(t) = d\theta'(t)/dt$$

$$\Rightarrow \quad \theta'(t) = \int_0^t q(t) \, dt + \theta'(0)$$

A 20 second test record is used, and q is constrained to be less than 0.04 rad/sec, while θ' is less than 0.1 rads. If q is at its maximum amplitude, then it is possible to calculate how long it will take before θ' reaches its maximum, if it starts at zero. This is done as follows,

$$q(t) = 0.04, \qquad \theta'(0) = 0$$

$$\Rightarrow \qquad \theta'(t_0) = \int_0^{t_0} 0.04 \; dt = 0.04 \; t_0$$

But, $\theta'(t_0) = 0.1$, since θ' is at its maximum at time, t_0.

i.e. $t_0 = 0.1/0.04 = 2.5$ seconds

Hence, the pitch rate, q, cannot remain at its maximum for longer than 2.5 seconds if θ' starts at zero, or 5 seconds if θ' starts at its minimum amplitude, -0.1 rads. It is therefore not possible for both q and θ' to be bang–bang.

Considering the auto–spectra of the responses (see figure 4.3), it can be seen that the pitch and roll rates contain frequency components up to around 5 rads/sec. In contrast, the longitudinal and vertical velocities have little power above 2 rads/sec. This difference in the bandwidth of the responses is reasonable, since the rates involve faster dynamics than the velocities, e.g. the fast pitch mode has an eigenvalue of -3.25, whereas the phugoid eigenvalues are $0.0447 \pm j \; 0.233$.

Now, in Chapter 2 it was concluded that, for the Lynx, the rotor dynamics were mainly present at frequencies above about 7 rads/sec. Hence, since the responses to the optimal input contain little power above this frequency, it can also be concluded that the optimal input will not excite the rotor dynamics significantly. Once again, this is reasonable, since the parameters which the optimal input has been designed to identify are not concerned with the rotor dynamics.

For this optimal input, the determinant of the dispersion matrix, |D|, was calculated to be 0.170. The corresponding values of |D| for other inputs are shown in table 4.2.

The 1221 and the doublet both give determinants around 150. In contrast, the double–doublet gives a significantly lower value, 3.95. This is in agreement with the results given in Chapter 2, where the double–doublet was found to be far superior to the doublet and 1221 inputs.

Input	|D|
Optimal Lynx output constrained input	0.170
Double-Doublet	3.95
1221	159.7
Doublet	136.2

Table 4.2 – *|D| for various inputs, calculated using 5^{th} order identified Lynx model*

It can be seen that the optimal input value of 0.170 for |D| is an order of magnitude better than that of the double–doublet. The sensitivities of the double–doublet responses to changes in the model parameters are shown in figure 4.5. These are of comparable magnitude to the sensitivities for the output amplitude constrained optimal input. Hence, the improvement in |D| using the optimal input has been achieved without increasing the amplitude of the responses, or reducing their robustness. This is an encouraging result, and demonstrates the potential for obtaining improved parameter estimates by using more carefully designed inputs.

4.3 OUTPUT AMPLITUDE CONSTRAINED AND RESPONSE ROBUST OPTIMAL INPUT

Unfortunately, attempts to design an optimal Lynx input with both output amplitude constraints and response robustness were unsuccessful. The input design algorithm used is described in Chapter 3 and Appendix B. A central component of this algorithm is the NAG general–purpose optimisation routine E04UCF. This routine performs non–linear optimisation by repeatedly linearising the required non–linear functions, and solving the resulting linear optimisation problems.

However, it was found that the response robustness constraints for the Lynx helicopter case studied are extremely non–linear, and have derivatives of large magnitude. Severe ill–conditioning therefore results when it is attempted to treat the non–linear optimisation as a series of linear optimisations. This ill–conditioning was found to lead to the constraints being grossly violated, and repeatedly gave rise to numeric overflow.

It is therefore the opinion of the author that an optimisation routine is required which deals directly with the non–linear nature of the problem, rather than linearising it. For example, in [1], a development of the Simplex method is given which will optimise a non–linear function subject to non–linear constraints, without the use of linearisation. The development of stable and efficient numerical techniques is held to be of great importance for the future, if optimal inputs with realistic constraints are to gain widespread use.

REFERENCES FOR CHAPTER FOUR

1. Rao, S.S. 'Optimisation theory and applications' Wiley, New Delhi 1987

CHAPTER FIVE

CONCLUSIONS

CONCLUSIONS

The test input applied to a helicopter, or any other system, for the purpose of system identification can have a substantial effect on the parameter estimates obtained. It is therefore important that an appropriate input is chosen, and in particular inputs must take account of the requirements, and restrictions, of the application.

In the rotorcraft case studied, the principal aim of the identification has been the development and validation of flight mechanics models. Due to practical restrictions, it is currently only feasible to use a linearised model. It is therefore essential that the input produces a linear response. Moreover, despite the unstable nature of a helicopter, if good estimates of the system parameters are to be obtained this response must be of reasonable duration, and must contain enough information about the system. With these considerations, several approaches to the design of system identification test inputs have been studied and evaluated.

Firstly, a straightforward method has been developed for the design of multi–step inputs. This method is based in the frequency–domain, and involves tailoring the auto–spectra of the inputs to give long, linear test records, and parameter estimates with reasonably low variances. In flight trials using the Lynx helicopter at RAE (Bedford), the double–doublet input, designed with this method, has been found to be a significant improvement over more traditional inputs.

Using the data from the flight trials of the double–doublet, both equation–error and output–error identification has been carried out. Several discrepancies were found between the theoretical and identified models. In particular, the unstable phugoid mode of the Lynx appeared to be poorly reproduced by the theoretical model. More work is required to clarify this. Numerical difficulties were encountered during the output–error identification. These were attributed to ill–conditioning resulting from the use of an unstable system. Little work appears to have been published on the particular difficulties involved in applying identification algorithms to unstable systems, and it is an area that is in need of further investigation.

The experience gained from these flight trials served to highlight the importance of robustness to the success of an input. Several types of robustness have been noted. Firstly, when inputs are applied manually, they must be insensitive to errors in amplitude and timing introduced at this stage. Secondly, the model used to design the inputs is likely to be inaccurate, otherwise the system identification would be unneccessary. Inputs must be able to tolerate these inaccuracies, and still give linear responses and parameter estimates of high quality.

In addition to designing multi-step inputs, the design of optimal inputs has also been investigated. In particular, constraints have been developed which are suitable for ensuring that the optimal inputs produce linear responses, and are robust. Conventional energy constraints were found to be of little use for these purposes. It is suggested that previous unsatisfactory results obtained by other authors using optimal inputs in aerospace applications may have been a consequence of using these inappropriate energy constraints.

Algorithms have been developed for the design of optimal inputs with a variety of constraints, and simulation studies have been made to gain an understanding of the effect of these constraints on the form of the inputs. These simulation studies were found to give a valuable insight into the characteristics of optimal test inputs.

With the constraints obtained from this work, an optimal input has been designed for use with the Lynx helicopter. This input is as robust as the double-doublet multi-step input designed for the Lynx, and yet is predicted to give significantly improved parameter estimates. Unfortunately, due to the Lynx at RAE (Bedford) being unavailable at the time of writing, no flight trials have been performed using this input. Attempts to design an optimal Lynx input with still greater robustness failed due to numerical problems in the software used. These problems are non-trivial, and further work is required if they are to be overcome.

Comparing the multi–step inputs and the optimal inputs, whereas the multi–step inputs are restricted to a sequence of steps, the optimal inputs can have a completely general form. Moreover, while the multi–steps only give reasonably low parameter variances, the optimal inputs are designed to strictly minimise the variances. Finally, the multi–steps are straightforward to design. In contrast, the design of the optimal inputs involves a more complex process, which perhaps leads to inputs that are more difficult to understand intuitively.

System identification is, in general, an iterative process. Initially, only a poor description of the system may be available. Hence, only crude inputs can be designed, giving parameter estimates with relatively large variances. These can be used, however, to give an improved model of the system, which can then be used to design improved inputs. These in turn give more accurate parameter estimates, and a further improvement in the model, and so on. The multi–step inputs and the optimal inputs developed in this work complement each other, and it is proposed that both can be used to advantage within this iterative framework.

In the initial identification, it is suggested that the multi–steps are a more appropriate type of input. Multi–step inputs are largely designed manually. There is therefore scope for the inclusion of information about the system from a wide variety of sources, in both qualitative and quantitative form. Since only a limited numerical model of the system may be available in the early stages of the identification, the ability to incorporate any extra information is important. Moreover, when the model is very inaccurate, then inputs need to have a large degree of robustness. However, as noted in Chapter 3, it is difficult in practice to produce optimal inputs with wide–range robustness. In such situations, multi–step inputs are therefore often superior at present to optimal inputs. It is proposed that optimal inputs are more suitable later in the identification process, when a fuller, and more accurate, model has typically been obtained.

Considering the future, it is important that the optimal inputs undergo flight trials, in order to confirm the simulation results presented. Work is also required to develop more stable and efficient numerical software for the design of optimal inputs. The present author's software is largely a research tool, and so has sacrificed efficiency for flexibility.

In the longer term, it is suggested that the greatest scope for further improvements in identification test inputs lies in the use of multi–axis inputs. For example, a conventional rotorcraft has four pilot controls, yet at present an input is applied to only one of these at a time. By using all four simultaneously, more information can be gained about the system in a shorter time. This must be balanced against the extra complexity of designing such inputs, and hence the greater time required for the input design stage of the identification.

APPENDIX A – THE ORDINARY COHERENCE FUNCTION

A.1 Introduction

In rotorcraft identification, a linear model is used. However, this is obtained by linearising the non–linear HELISTAB model about a particular flight condition. If the system response departs too far from this flight condition, then it becomes non–linear, and the linearised model is invalid.

By definition, the coherence function is a measure of the linearity of the system relating two signals. Hence, the coherence may be used to determine when the response is linear or nonlinear, and therefore whether a linearised model is valid.

A.2 Theoretical Background

The ordinary coherence function relating two transient signals x(t) and y(t) is defined in [1] as,

$$|\gamma_{xy}(\omega)|^2 = \frac{|S_{xy}(\omega)|^2}{S_x(\omega)\ S_y(\omega)} \qquad (A.1)$$

where,

$|\gamma_{xy}(\omega)|^2$ is the coherence between signals x(t) and y(t).
$S_{xy}(\omega)$ is the cross–spectrum of x(t) and y(t).
$S_x(\omega)$ is the auto–spectrum of x(t).
$S_y(\omega)$ is the auto–spectrum of y(t).

The coherence can be interpreted as a measure of the power in the signal y(t) which is due to a linear relation with the signal x(t) [1], as follows.

Consider the following general stationary system relating two signals x(t) and y(t),

$$y(t) = F(x(t)) \qquad (A.2)$$

The relation between x(t) and y(t) can be split into a linear part and a non– linear part. Translated to the frequency– domain, this gives,

$$Y(\omega) = H(\omega)\, X(\omega) + N_x(\omega) \qquad (A.3)$$

where,

$X(\omega)$ is the Fourier transform of the system output y(t).

$Y(\omega)$ is the Fourier transform of the input signal x(t).

$H(\omega)$ is the transfer function of the system i.e. the linear part of the relation between $X(\omega)$ and $Y(\omega)$.

$N_x(\omega)$ is the Fourier transform of the component of y(t) produced non– linearly from x(t).

Then,

$$S_y(\omega) = |H(\omega)|^2\, S_x(\omega) + S_n(\omega) \qquad (A.4)$$

i.e total linear power in signal y = linear power from signal x + non-linear power from signal x (A.5)

The ordinary coherence function can be shown to be,

$$|\gamma_{xy}(\omega)|^2 = \frac{|H(\omega)|^2\, S_x(\omega)}{S_y(\omega)} \qquad (A.6)$$

$$= \frac{\text{linear power from signal x}}{\text{total power in signal y}} \qquad (A.7)$$

From this expression, it can be seen that the coherence must lie between zero and unity. A coherence of unity means that the system relating x(t) and y(t) is purely linear. While a coherence of less than unity means that there is a non– linear component in this system.

In order to estimate the ordinary coherence, it is first necessary to estimate the auto– and cross– spectra of signals x(t) and y(t). An estimate of the coherence can then be obtained using equation (A.1) above.

By definition [1],

$$S_x(\omega) = E[\ |X(\omega)|^2\ ] \qquad (A.8)$$

$$S_y(\omega) = E[\ |Y(\omega)|^2\ ] \qquad (A.9)$$

$$S_{xy}(\omega) = E[\ X(\omega)\ Y(\omega)^*\ ] \qquad (A.10)$$

where,

* superscript denotes the complex conjugate.

E[ ] denotes averaging over an infinite number of frequency spectra.

Take the signals x(t) and y(t), and sample them to give the discrete signals x(i) and y(i),

$$x(i) = x((i-1)\Delta t) \qquad i = 1, 2, \ldots N \qquad (A.11)$$

$$y(i) = y((i-1)\Delta t) \qquad (A.12)$$

where,

Δt is the sampling interval used.

N is the number of samples.

Given x(i) and y(i), the Fast Fourier Transform can then be used to obtain discrete estimates of $X(\omega)$ and $Y(\omega)$. Let these discrete estimates be denoted by X(k) and Y(k). It is important that some form of windowing is now used to reduce side-lobe leakage due to signal truncation. If this is not done, the coherence estimates obtained may be severely biased [2]. It is suggested that the GEO window [3] be used for this purpose (see section A.3 below).

Using these windowed X(k) and Y(k), initial estimates for the auto- and cross-spectra can be obtained by omitting the averaging in equations (A.8) - (A.10) above. Averaging is then performed in the frequency-domain to give final estimates of the auto- and cross-spectra. Frequency averaging can be expressed [3] as,

$$\tilde{S}_x(k) = 1/n \sum_{i=-n/2}^{n/2} \hat{S}_x(k+i) \qquad k = \frac{n}{2}, \ \frac{3n}{2}, \ \ldots \qquad (A.13)$$

where,

$\hat{S}_x(k)$ is the discrete initial auto-spectrum estimate.

$\tilde{S}_x(k)$ is the averaged discrete auto-spectrum estimate.

n is the averaging interval i.e. the number of samples from the initial spectrum averaged together at each point of the final spectrum.

An estimate of the coherence can now be obtained using equation (A.1).

However, since in practice only a finite averaging interval, n can be used in equation (A.13), the averaging is not perfect, and a bias is introduced into the coherence estimate as a result. Theoretically, this bias is given by [4, 5, 6, 7] as,

$$\text{Bias, } B(|\gamma_{xy}(\omega)|^2) \approx \frac{1}{n}\left[1-|\gamma_{xy}(\omega)|^2\right]^2\left[1+2\,\frac{|\gamma_{xy}(\omega)|^2}{n}\right] \qquad \text{(A.14)}$$

where,

$|\gamma_{xy}(\omega)|^2$ is the true coherence.

Hence, an approximate correction for this bias can be obtained using,

$$\text{Approx. Bias} = B(|\hat{\gamma}_{xy}(k)|^2) \qquad \text{(A.15)}$$

$$|\tilde{\gamma}_{xy}(k)|^2 = |\hat{\gamma}_{xy}(k)|^2 - \text{Approx. Bias} \qquad \text{(A.16)}$$

where,

$|\hat{\gamma}_{xy}(k)|^2$ is the discrete initial estimate of the coherence

$|\tilde{\gamma}_{xy}(k)|^2$ is the discrete corrected coherence estimate

This corrected estimate was taken as the final coherence estimate.

However, other biases can also affect the coherence estimate. For example, misalignment bias and correlation bias. Misalignment bias [2, 8] is due to time delays being present in the system relating the two signals x(t) and y(t). Such delays result in a part of y(t) corresponding to a part of x(t) which is outside that available time record. Hence, the coherence estimates are too low.

Theoretically, misalignment bias is given by the expression,

$$\text{Bias} \approx -\frac{|\tau|}{T} |\gamma_{xy}(\omega)|^2 \qquad \text{(A.17)}$$

where,

τ is the time delay.
T is the length of the time history.
$|\gamma_{xy}(\omega)|^2$ is the true coherence.

Correlation bias occurs when the ordinary coherence function is used with systems that have more than one input. If the coherence is taken with respect to one input, the other inputs will appear to produce non-linearities and so give a low coherence. However, if the inputs are correlated, the coherence will be higher than if they were uncorrelated. In particular, if the inputs are linearly related, then the coherence will behave as if only one input was present.

This bias is of particular interest in situations where the initial conditions of the system are not zero. The initial conditions act as extra inputs to the system. However, these extra inputs are in the form of Dirac δ-functions. Hence, they are linearly related to the true system input, and so can be ignored.

A.3 A Simulation Study of the Ordinary Coherence Function

Given the large number of biases which may affect the coherence function, it was considered important that the behaviour of the coherence calculations used be properly characterised. In [5] Carter mentions a method for generating Gaussian signals with known coherence. These were therefore used for an initial study of the coherence calculations.

Consider two signals x(t) and y(t), as follows,

$$x(t) = a(t) + K\, b(t) \tag{A.18}$$

$$y(t) = b(t) + K\, a(t) \tag{A.19}$$

where,

K is some constant.

a(t) and b(t) are uncorrelated Gaussian signals with identical auto– spectra.

Then,

$$\begin{aligned} S_x(\omega) &= S_a(\omega) + K^2\, S_b(\omega) \\ &= (1 + K^2)\, S_a(\omega) \quad \text{since} \quad S_a(\omega) = S_b(\omega) \end{aligned} \tag{A.20}$$

$$S_y(\omega) = (1 + K^2)\, S_a(\omega) \tag{A.21}$$

and,

$$\begin{aligned} S_{xy}(\omega) &= K\, S_a(\omega) + K\, S_b(\omega) \\ &= 2K\, S_a(\omega) \end{aligned} \tag{A.22}$$

This gives,

$$\begin{aligned} |\gamma_{xy}(\omega)|^2 &= \frac{4K^2\, S_a(\omega)^2}{(1 + K^2)^2\, S_a(\omega)^2} \\ &= \frac{4K^2}{(1 + K^2)^2} \end{aligned} \tag{A.23}$$

Figure A.1

The estimated coherence between two Gaussian signals with a known coherence of 0.8 (frequency averaging interval of 3 used)

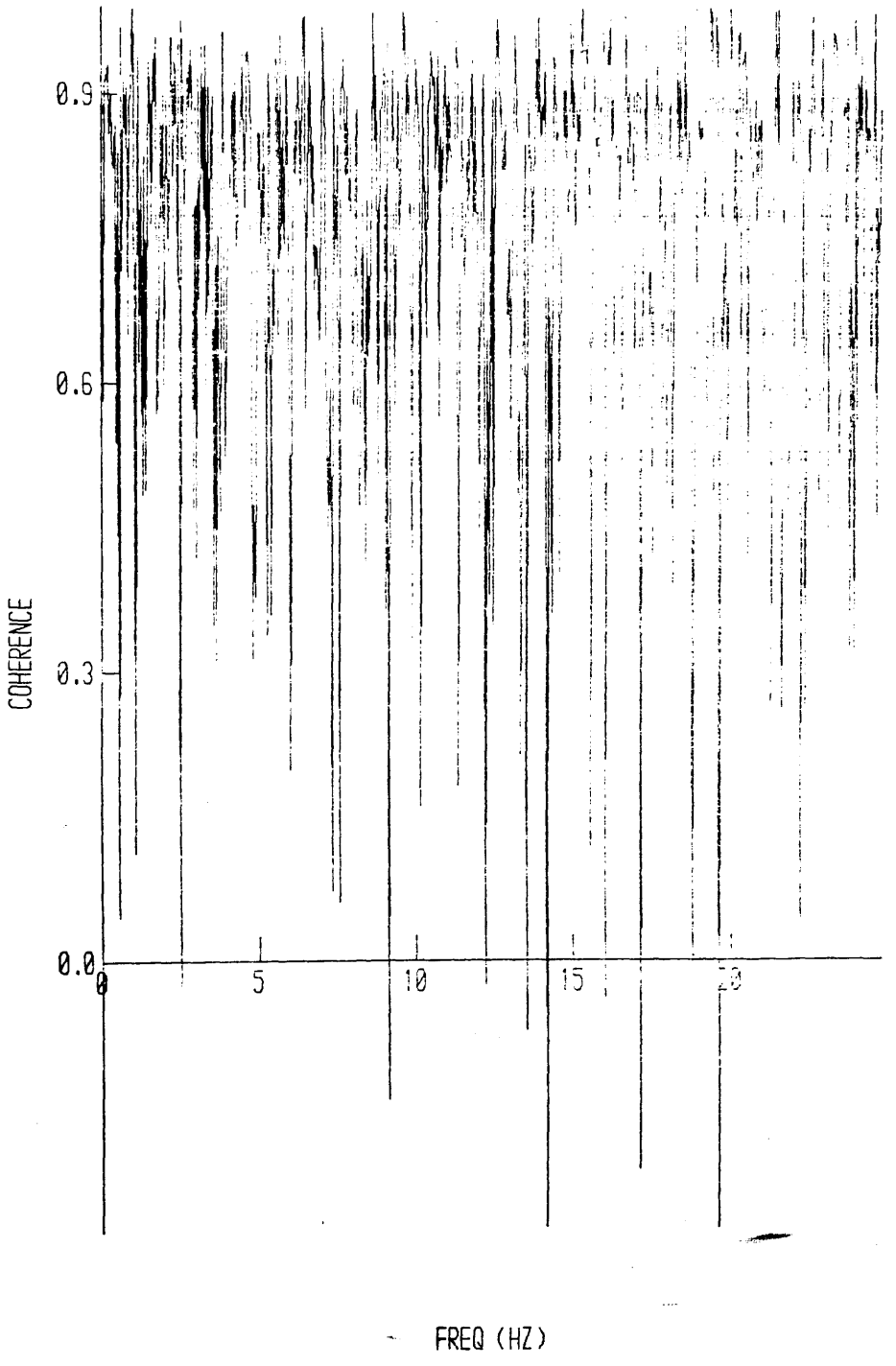
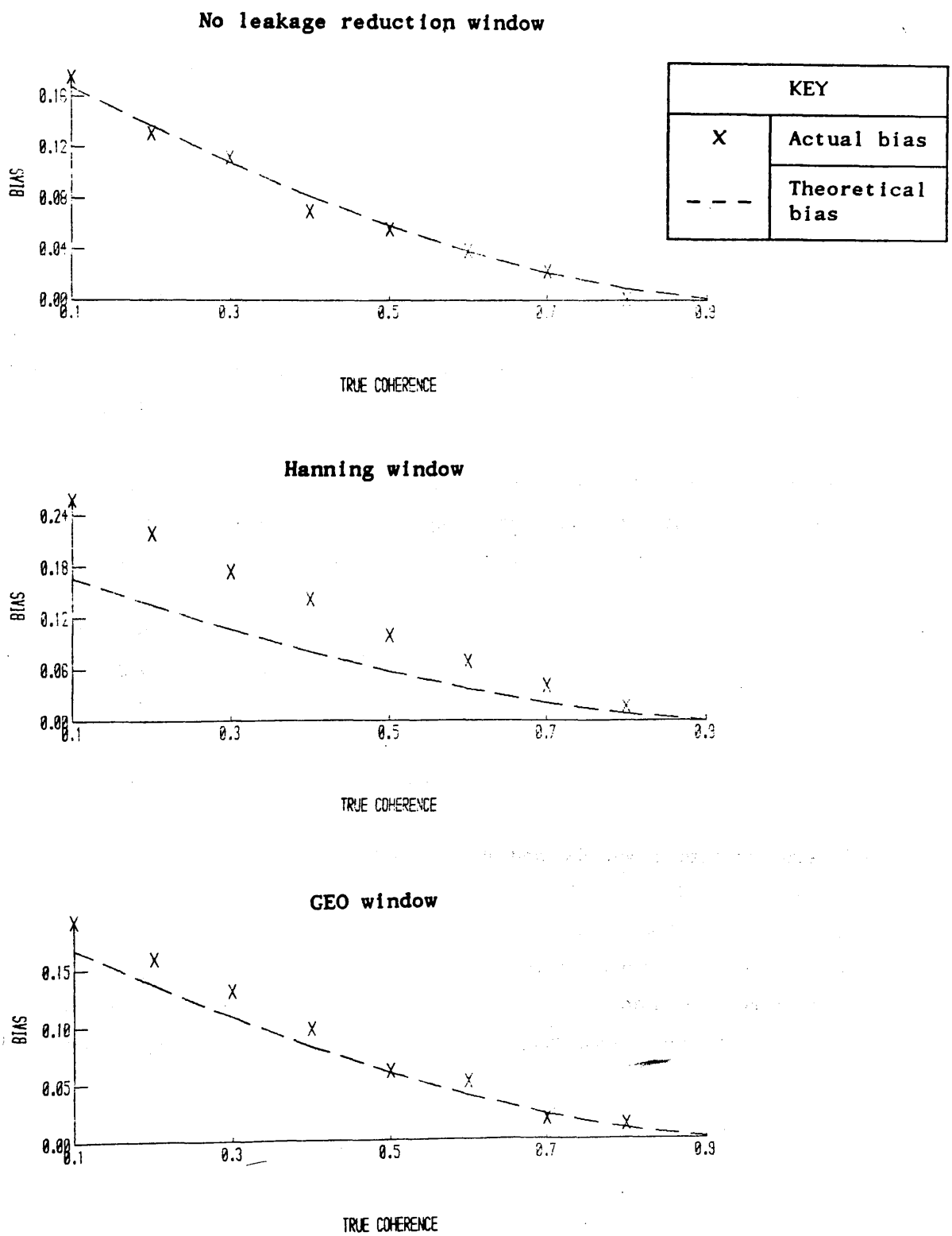


Figure A.2

Theoretical and actual bias of estimated coherence for no leakage reduction window, a Hanning window, and a GEO window. Frequency averaging interval of 5 used, and actual bias obtained by taking the mean of 1000 coherence estimates)



Rearranging such that $|\gamma_{xy}(\omega)|^2 = 0$ when $K = 0$ gives,

$$K = \frac{1 - \sqrt{(1 - |\gamma_{xy}(\omega)|^2)}}{|\gamma_{xy}(\omega)|} \qquad (A.24)$$

Hence, by controlling the value of K it is possible to control the coherence between signals x(t) and y(t).

The coherence calculations were found to produce estimates of the coherence between x(t) and y(t) which were in good agreement with equation (A.24) above (e.g. see figure A.1).

Now, for these Gaussian test signals the true coherence is constant over all frequencies. Hence, the coherence calculations estimate the coherence at N discrete frequencies, and so produce N estimates of the true coherence. Therefore, by averaging these estimates together, the variance and bias of the estimates can be calculated.

Using this method, the bias of the coherence estimates, without any correction using equation (A.16), was compared with the predicted theoretical bias given by equation (A.14). This comparison was performed for,

a) No leakage reduction window,
b) A Hanning window, and
c) A GEO window.

See figure A.2 for the results obtained.

It can be seen that when no leakage reduction window is used the bias of the estimates agrees extremely well with the theoretical bias. However, when a Hanning window is used, the bias of the estimates is considerably higher than in theory, although the bias still exhibits characteristics similar to those of the theoretical bias. The GEO window appears to offer a compromise between these two cases : it gives a bias which is only slightly higher than in theory.

It is suggested that these results are due to the Hanning and GEO windows effectively reducing the frequency averaging interval used in calculating the coherence estimates, and so producing a larger bias than when no window is involved. If this is true, then the Hanning window produces a larger reduction in the effective averaging interval than does the GEO window, since the Hanning window produces a larger bias.

A result by Carter [2] was used in order to study further the issues associated with leakage reduction windows. Carter investigated the coherence of a linear second order digital filter,

$$y_n = A\, y_{n-1} + B\, y_{n-2} + C\, x_n \qquad (A.25)$$

where,

$A = 1.973$

$B = -0.98202$

$C = 0.0087$

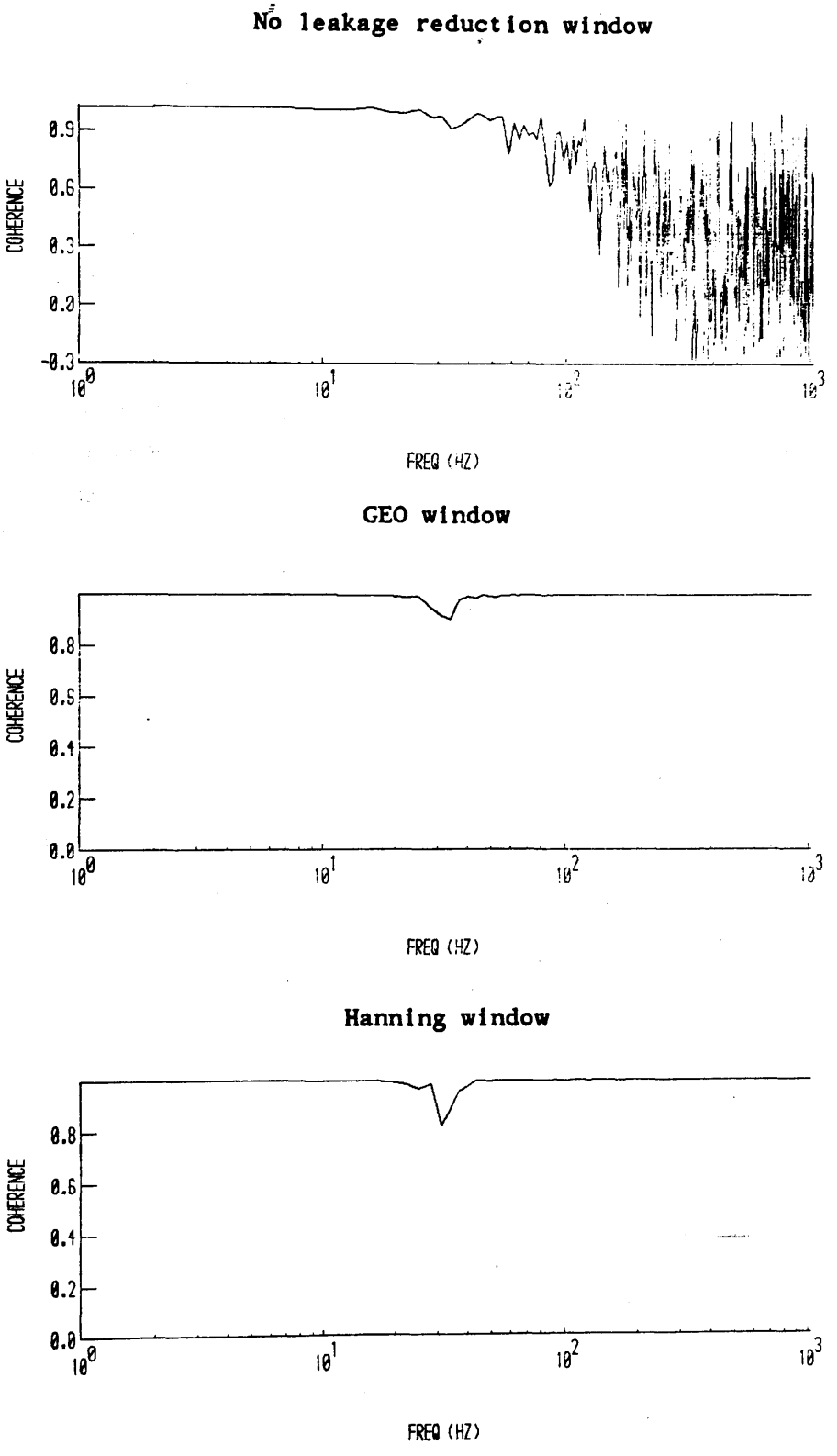
y_n is the n^{th} output sample from the filter.

x_n is the n^{th} input sample to the filter. The input is Gaussian white noise, with a sampling interval of 4.88281×10^{-4} sec and 2048 samples.

Carter found that when no leakage reduction window was used, the coherence was significantly underestimated at frequencies above 100 Hz. It was also underestimated at frequencies around 20 Hz. When Carter used a Hanning window, this resulted in the coherence above 100 Hz being correctly estimated. The low coherence around 20 Hz still remained, and this was shown to be due to misalignment bias. Hence, it appears that the estimated coherence can be grossly underestimated if no leakage reduction window is used.

Figure A.3

Estimated coherence for Carter's second order digital filter using various leakage reduction windows, and a frequency averaging interval of 3



In order to examine this, the author's coherence calculation routines were used to estimate the coherence for Carter's filter, with,

a) No leakage reduction window,
b) A Hanning window, and,
c) A GEO window.

See figure A.3 for the results obtained.

It can be seen that these results agree with Carter. For no window, the coherence is underestimated above 100 Hz. Using either a Hanning or a GEO window solves this problem.

Hence, a leakage reduction window must be used when calculating the coherence function. Since the GEO window produces a smaller averaging bias than the Hanning window, it is suggested that the GEO window is suitable for this purpose.

REFERENCES FOR APPENDIX A

1. Bendat, J.; Piersol, A. 'Engineering applications of correlation and spectral analysis', Wiley, Toronto 1980

2. Carter, G.; Knapp, C. 'Coherence and its estimation via the partitioned modified Chirp−Z transform', IEEE Trans ASSP, Vol ASSP−23, pp257−264, 1975

3. Otnes, R.; Enochson, L. 'Applied time series analysis', Wiley, London 1978

4. Nuttall, A.; Carter, G. 'Bias of the estimate of magnitude−squared coherence', IEEE Trans ASSP, Vol ASSP−24, pp582−583, 1976

5. Carter, G.; Knapp, C., Nuttall, A. 'Estimation of the magnitude−squared coherence function via overlapped fast Fourier transform processing', IEEE Trans Acoust, Vol AU−21, pp337−344, 1973

6. Benignus, V. 'Estimation of the coherence spectrum and its confidence interval using the fast Fourier transform', IEEE Trans Acoust, Vol AU−17, pp145−150, 1969

7. Carter, G.; Knapp, C.; Nuttall, A. 'Statistics of the estimate of the magnitude−coherence function', IEEE Trans Acoust, Vol AU−21, pp388−389, 1973

8. Carter, G. 'Bias in magnitude−squared coherence estimation due to misalignment', IEEE Trans ASSP, Vol ASSP−28, pp97−99, 1980

APPENDIX B – NUMERICAL METHODS

B.1 Introduction

The theoretical details of the input design algorithms used are given in Chapter 3. In this appendix, the practical implementation of these algorithms using numerical software is discussed. Firstly, an important consideration in practice is the efficiency of the software used. Since complete time histories are being manipulated, the computing time required is generally very high. Hence, if an algorithm is inefficient then it may not be possible to obtain results within a realistic time. In addition, it is necessary to consider the conditioning of the optimisation used. Ill conditioning can prevent the input design algorithms from converging, or, in less severe cases, can lead to an increase in the amount of computing time required. The questions of efficiency and conditioning therefore dominate much of the discussion in this appendix.

Attention is directed mainly at the time–domain algorithms, since these present a more diffcult problem than the relatively straightforward frequency–domain algorithms. The time–domain algorithms used were, with the exception of those involving the parameter robustness constraint, of the following general form:

1. Choose suitable initial input, $u_1(t)$. Let $n = 1$.

2. Find input, $u_0(t)$ to maximise $\varphi(u_0)$, subject to any constraints required,

   where,

   $\varphi(u_0)$ depends on the constraints being used (see Chapter 3)

3. Generate new input, $u_{n+1}(t)$ using,

   $$u_{n+1}(t) = \alpha\, u_0(t) + \beta\, u_n(t)$$

   where,

   α, β are chosen such that $|D_{n+1}|$ is minimised, and
   $u_{n+1}(t)$ meets any required constraints.

4. If $u_{n+1}(t)$ is not optimal, goto step 2 and repeat.

Step 2 of this algorithm is the most complex, and absorbs much of the computing time required by the algorithm. Particular attention was therefore given to efficiently implementing this step. A Rayleigh– Ritz approach was used, which involved expanding the input in a series of orthogonal functions, as follows,

$$u(t) = \sum_{i=0}^{N} a_i \, u_i(t) \qquad (B.1)$$

where,

a_i are constants.

$u_i(t)$, $i = 0, 1, \ldots N$, are a suitable set of orthogonal functions.

The number, N, of orthogonal functions was selected by the user. For this value of N, the coefficients, a_i, were then chosen by the software to maximise $\varphi(u)$ subject to the constraints being used.

This optimisation was implemented using the general– purpose NAG routine E04UCF [1], which will minimise/maximise an arbitary smooth function subject to linear, and/or smooth non– linear, constraints. The objective and constraint functions, plus their derivatives, are supplied by user subroutines. A more efficient approach would have been to develop individual optimisation routines, each tailored to deal with a particular set of constraints. However, it was felt that the flexibility resulting from the use of a general– purpose routine more than compensated for the loss in efficiency. With this flexibility it was possible to study a wide variety of constraints quickly and with relative ease.

For the algorithms involving the parameter robustness constraint, the input was once again expanded in an orthogonal series as in equation (B.1). Routine E04UCF was then used to directly minimise $|D|$ subject to the required constraints. Unfortunately, this approach was found to be significantly slower and less well conditioned than the more specialised algorithms used with the other types of constraint studied.

B.2 Calculation of $\varphi(u_0)$

B.2.1 Introduction

The optimisation routine used to perform step 2 of the input design algorithm typically requires several hundred evaluations of the function $\varphi(u_0)$ and its derivatives $d\varphi(u_0)/da_i^0$, where a_i^0 are the coefficients of $u_0(t)$ in equation (B.1). It is therefore important that these calculations are performed efficiently.

Expressions for $\varphi(u_0)$ for each of the input design algorithms studied are given in Chapter 3. In each case, $\varphi(u_0)$ involves the term $Tr(M_n^{-1}M_{0n})$, and it is this term which is the most time-consuming to calculate. From equation (2.4), the information matrix is as follows,

$$M_n = \int_0^T \left[\frac{dy_n}{d\theta}(t) \right]^T R^{-1} \frac{dy_n}{d\theta}(t) \quad dt \qquad (B.2)$$

While from equation (3.24),

$$M_{0n} = \int_0^T \left[\frac{dy_0}{d\theta}(t) \right]^T R^{-1} \frac{dy_n}{d\theta}(t) \quad dt \qquad (B.3)$$

Now,

$$u_n(t) = \sum_{i=0}^{N} a_i^n u_i(t), \quad u_0(t) = \sum_{i=0}^{N} a_i^0 u_i(t) \qquad (B.4)$$

and therefore,

$$y_n(t) = \sum_{i=0}^{N} a_i^n y_i(t), \quad y_0(t) = \sum_{i=0}^{N} a_i^0 y_i(t) \qquad (B.5)$$

where,

$y_i(t)$ is the system response to input $u_i(t)$.

Hence,

$$M_n = \int_0^T \left[\sum_{i=0}^{N} a_i^n \frac{dy_i}{d\theta}(t) \right]^T R^{-1} \sum_{i=0}^{N} a_i^n \frac{dy_i}{d\theta}(t) \quad dt$$

$$= \sum_{i=0}^{N} \sum_{j=0}^{N} a_i^n a_j^n M_{ij} \qquad (B.6)$$

and,

$$M_{on} = \sum_{i=0}^{N} \sum_{j=0}^{N} a_i^0 a_j^n M_{ij} \qquad (B.7)$$

where,

$$M_{ij} = \int_0^T \left[\frac{dy_i}{d\theta}(t) \right]^T R^{-1} \frac{dy_j}{d\theta}(t) \quad dt = M_{ji}^T \qquad (B.8)$$

Moreover,

$$\frac{dM_n}{da_k^n} = 2 \sum_{i=0}^{N} a_i^n M_{ki} , \qquad \frac{dM_{on}}{da_k^0} = \sum_{i=0}^{N} a_i^n M_{ki} \qquad (B.9)$$

By precalculating the M_{ij}, then M_n, M_{on}, and their derivatives can be calculated by a simple summation, requiring little computing time.

B.2.2 Calculation of the Sensitivity Functions

Since little time is now needed to calculate M_n, M_{on}, and their derivatives, much of the computing time is spent precalculating the M_{ij}. The time taken to calculate the M_{ij} is dominated by the calculation of the sensitivities, $dy_i(t)/d\theta$. An efficient method for obtaining the sensitivities is therefore necessary.

Originally, the sensitivities were calculated using a transition matrix approach. It is well known, e.g. [2], that a standard state–space model may be solved to give the states, x(t), using a convolution integral, as follows :

$$x(t) = \Phi(t)\, x(0) + \int_0^t \Phi(t-\tau)\, B\, u(\tau)\, d\tau \qquad (B.10)$$

where,

$\Phi(t) = e^{At}$, the state transition matrix.

This can be straightforwardly rearranged into the discrete recursive form,

$$x(i\Delta t) = \Phi(\Delta t)\, x(\{i-1\}\Delta t) + \int_{\{i-1\}\Delta t}^{i\Delta t} \Phi(i\Delta t-\tau)\, B\, u(\tau)\, d\tau \qquad (B.11)$$

where,

Δt is the sampling interval.

Differentiating with respect to θ gives,

$$\frac{dx(i\Delta t)}{d\theta} = \frac{dA}{d\theta}\,\Delta t\,\Phi(\Delta t)\, x(\{i-1\}\Delta t) + \Phi(\Delta t)\frac{dx(\{i-1\}\Delta t)}{d\theta} + \int_{\{i-1\}\Delta t}^{i\Delta t} \left\{ \frac{dA}{d\theta}\{i\Delta t-\tau\}\,\Phi(i\Delta t-\tau)\, B\, u(\tau) + \Phi(i\Delta t-\tau)\,\frac{dB}{d\theta}\, u(\tau) \right\} d\tau \qquad (B.12)$$

It should be noted that since θ is simply a vector containing elements of A and B, $dA/d\theta$ and $dB/d\theta$ are constants.

Now recall that,

$$y(t) = C\, x(t) \qquad (B.13)$$

Hence,

$$dy(t)/d\theta = C\ dx(t)/d\theta \qquad (B.14)$$

Equations (B.11), (B.12), and (B.14) provide a means of calculating $\partial y(t)/\partial\theta$, given $\Phi(t)$. Unfortunately, calculating the transition matrix, $\Phi(t)$ is prone to ill-conditioning. A variety of methods for calculating $\Phi(t)$ have been published [3-6], but one of the most stable and accurate of these is that described by Ward [7]. This uses a diagonal Pade approximation to $\Phi(t)$ [8,9] combined with repeated squaring. A particular attraction of this method is that it produces an estimate of the accuracy of the result obtained, and so it is possible to output a warning if the accuracy falls too low, giving improved reliability. A version of Ward's method, modified to take account of the need to calculate $\Phi(t)$ at several values of t, was therefore used for the present work.

However, while the algorithm described above for calculating $dy(t)/d\theta$ has extremely good accuracy, it is relatively slow. In the current application, it is essential that the algorithms used are sufficiently fast if results are to be obtained in realistic time scales. Hence, this slow initial algorithm was later replaced by a significantly more efficient, but less accurate [10], method for calculating the sensitivities.

It can be shown [11,12] that the states, x(t), and sensitivities, $dy(t)/d\theta$ of any linear system with zero initial conditions may be expressed as follows.

$$\bar{x}(t) = G \int_0^t \bar{a}(t-\tau)^* u(\tau)\, d\tau \qquad (B.15)$$

where,

$$\bar{x}(t) = \begin{bmatrix} x(t) \\ \frac{dx}{d\theta_1}(t) \\ \vdots \\ \frac{dx}{d\theta_q}(t) \end{bmatrix}, \qquad \bar{x}(0) = 0 \qquad (B.16)$$

$u(t)$ = system input vector, r = number of inputs

θ_i = i^{th} parameter, $i = 1, 2, \ldots q$

$G = [\, G_1 \; G_2 \; \ldots \; G_r \,]$

$$G_j = \begin{bmatrix} b_j & Ab_j & \cdots & A^{2n-1}b_j \\ \frac{d(b_j)}{d\theta_1} & \frac{d(Ab_j)}{d\theta_1} & \cdots & \frac{d(A^{2n-1}b_j)}{d\theta_1} \\ \vdots & \vdots & & \vdots \\ \frac{d(b_j)}{d\theta_q} & \frac{d(Ab_j)}{d\theta_q} & \cdots & \frac{d(A^{2n-1}b_j)}{d\theta_q} \end{bmatrix} \tag{B.17}$$

b_j = j^{th} column of control matrix, B

$$\bar{a}(t) = \begin{bmatrix} a_1(t) \\ a_2(t) \\ \vdots \\ a_{2n}(t) \end{bmatrix} = \Lambda^{-1}\,\bar{\beta}(t) \tag{B.18}$$

$$\Lambda = \begin{bmatrix} \bar{\lambda}_1^T \\ \frac{d}{d\lambda_1}\bar{\lambda}_1^T \\ \vdots \\ \frac{d^{2n_1-1}}{d\lambda_1^{2n_1-1}}\bar{\lambda}_1^T \\ \vdots \\ \frac{d^{2n_m-1}}{d\lambda_m^{2n_m-1}}\bar{\lambda}_m^T \end{bmatrix}, \quad \bar{\lambda}_k = \begin{bmatrix} 1 \\ \lambda_k \\ \vdots \\ \lambda_k^{2n-1} \end{bmatrix} \qquad \text{(B.19)}$$

λ_k = k^{th} eigenvalue of A, k = 1, 2, ... m

n_k = multiplicity of k^{th} eigenvalue

n = number of state variables

$$\bar{\beta}(t) = \begin{bmatrix} e^{\lambda_1 t} \\ \vdots \\ t^{2n_1-1} e^{\lambda_1 t} \\ \vdots \\ e^{\lambda_m t} \\ \vdots \\ t^{2n_m-1} e^{\lambda_m t} \end{bmatrix} \qquad \text{(B.20)}$$

and,

$$\int_0^t \bar{a}\ (t-\tau)^*\ u(\tau)\ d\tau \;=\; \begin{bmatrix} \int_0^t a(t-\tau)\ u_1(\tau)\ d\tau \\ \vdots \\ \int_0^t a(t-\tau)\ u_r(\tau)\ d\tau \end{bmatrix} \qquad \text{(B.21)}$$

$$= \begin{bmatrix} \Lambda^{-1} & 0 & \cdot & \cdot & 0 \\ 0 & \Lambda^{-1} & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & \Lambda^{-1} \end{bmatrix} \int_0^t \bar{\beta}\ (t-\tau)^*\ u(\tau)\ d\tau \qquad \text{(B.22)}$$

In this method, G and Λ are straightforward to calculate. It then remains to obtain,

$$\int_0^t \bar{\beta}\ (t-\tau)^*\ u(\tau)\ d\tau \;=\; \begin{bmatrix} \int_0^t \beta(t-\tau)\ u_1(\tau)\ d\tau \\ \vdots \\ \int_0^t \beta(t-\tau)\ u_r(\tau)\ d\tau \end{bmatrix} \qquad \text{(B.23)}$$

In the present work, interest has been confined to situations where only a single input is used i.e. r=1. Moreover, the eigenvalues are all of multiplicity one i.e. $n_k=1$, k=1, 2, ... n. This gives,

$$\int_0^t \overline{\beta}\ (t-\tau)^*\ u(\tau)\ d\tau \ = \ \int_0^t \begin{bmatrix} e^{\lambda_1(t-\tau)} \\ (t-\tau)e^{\lambda_1(t-\tau)} \\ . \\ . \\ (t-\tau)e^{\lambda_n(t-\tau)} \end{bmatrix} u(\tau)\ ds \qquad \text{(B.24)}$$

i.e it is necessary to calculate,

$$\int_0^t e^{\lambda_k(t-\tau)}u(\tau)\ d\tau \quad \text{and} \quad \int_0^t (t-\tau)e^{\lambda_k(t-\tau)}u(\tau)\ d\tau$$

where k = 1, 2, ...n

These 2n convolutions may be expressed as follows,

$$e^{\lambda_k t}\ \int_0^t e^{-\lambda_k \tau}u(\tau)\ d\tau \qquad \text{(B.25)}$$

and,

$$te^{\lambda_k t}\ \int_0^t e^{-\lambda_k \tau}u(\tau)\ d\tau \ - \ e^{\lambda_k t}\ \int_0^t \tau e^{-\lambda_k \tau}u(\tau)\ ds \qquad \text{(B.26)}$$

This involves only 2n quadratures, and so is extremely efficient when compared to the transition matrix approach described previously. Unfortunately, when these quadratures are implemented on a computer, the terms $e^{\lambda_k t}$ and $e^{-\lambda_k t}$ are prone to numeric overflow and underflow. However, this problem is easily overcome by re-arranging the quadratures as follows,

$$h_k(t+\Delta t) \ = \ e^{\lambda_k \Delta t}\ \left[h_k(t) \ + \ \int_t^{t+\Delta t} e^{\lambda_k(t-\tau)}u(\tau)\ d\tau \ \right] \qquad \text{(B.27)}$$

$$g_k(t+\Delta t) = e^{\lambda_k \Delta t} \left[g_k(t) + \Delta t\ h_k(t) + \int_t^{t+\Delta t} (t+\Delta t-\tau) e^{\lambda_k (t-\tau)} u(\tau)\ d\tau \right] \tag{B.28}$$

where,

$$h_k(t) = \int_0^t e^{\lambda_k (t-\tau)} u(\tau)\ d\tau \tag{B.29}$$

$$g_k(t) = \int_0^t (t-\tau)\ e^{\lambda_k (t-\tau)} u(\tau)\ d\tau \tag{B.30}$$

$$h_k(0) = 0 = g_k(0) \tag{B.31}$$

This form was found to give good results for stable systems, i.e. $\lambda_k<0$. However, in unstable cases accuracy problems were encountered. These were overcome by taking advantage of some of the special features of the present application. Recall that it is required to calculate $dy_i(t)/d\theta$ i.e. the sensitivities for inputs $u_{i(t)}$, i=1, 2, ... N. The inputs, $u_i(t)$, are based on Chebyshev polynomials (see section B.3 below), and this can be used to give improved accuracy.

Chebyshev polynomials are related by the following recurrence formula [13],

$$T_i(x) = 2x\ T_{i-1}(x) - T_{i-2}(x)$$

where,

$T_i(x)$ = Chebyshev polynomial of order i

$T_0(x) = 1,\ T_1(x) = x$

$-1 \leqslant x \leqslant 1$

This gives the following recurrence relations between the quadratures,

$$h_k^i(t) = \int_0^t e^{\lambda_k (t-\tau)} T_i(x(\tau))\ d\tau$$

$$= 2\int_0^t x(\tau)\ e^{\lambda_k(t-\tau)} T_{i-1}(x(\tau))\ d\tau - h_k^{i-2}(t) \qquad (B.33)$$

$$g_k^i(t) = \int_0^t (t-\tau)\ e^{\lambda_k(t-\tau)} T_i(x(\tau))\ d\tau$$

$$= t\ h_k^i(t) - \int_0^t \tau\ e^{\lambda_k(t-\tau)} T_i(x(\tau))\ d\tau \qquad (B.34)$$

where,

$$x(s) = (2s-T)/T \qquad (B.35)$$

T = duration of input

Now,

$$\int_0^t \tau\ e^{\lambda_k(t-\tau)} T_i(x(\tau))\ d\tau = \frac{T}{2}\int_0^t x(\tau) e^{\lambda_k(t-\tau)} T_i(x(\tau))\ d\tau + \frac{T}{2}\ h_k^i(t) \qquad (B.36)$$

Using integration by parts,

$$f_k^i(t) = \int_0^t x(\tau) e^{\lambda_k(t-\tau)} T_i(x(\tau))\ d\tau$$

$$= \left[\frac{e^{\lambda_k(t-\tau)}}{-\lambda_k}\ x(\tau)\ T_i(x(\tau)) \right]_0^t$$

$$- \int_0^t \frac{e^{\lambda_k(t-\tau)}}{-\lambda_k} \left[\frac{dx}{d\tau}(\tau)\ T_i(x(\tau)) + x(\tau)\ \frac{dT_i}{dx}(x(\tau)) \frac{dx}{d\tau}(\tau) \right] d\tau \qquad (B.37)$$

$$= \frac{x(t)T_i(x(t))}{-\lambda_k} + \frac{e^{\lambda_k t}}{\lambda_k} x(0)T_i(x(0)) + \frac{1}{\lambda_k}\ \frac{2}{T}\ h_k^i(t)$$

$$+ \frac{1}{\lambda_k}\ \frac{2}{T} \int_0^t e^{\lambda_k(t-\tau)} x(\tau) \frac{dT_i}{dx}(x(\tau))\ d\tau \qquad (B.38)$$

Since dT_i/dx can be expressed in terms of a Chebyshev series of order $i-1$, $f_k^i(t)$ may be evaluated given $h_k^i(t)$, $h_k^{i-1}(t)$, ... $h_k^0(t)$. While from equations (B.33) and (B.34), $h_k^i(t)$ may be evaluated given $h_k^{i-2}(t)$ and $f_k^{i-1}(t)$, and $g_k^i(t)$ evaluated given $h_k^i(t)$ and $f_k^i(t)$. This allows $h_k^i(t)$ and $g_k^i(t)$ to be obtained recurrently.

However, in practice this approach was found to be prone to inaccuracies when polynomials of high order were used. A solution to the problem of accuracy was finally obtained by using the following method.

Chebyshev polynomials may be expressed as power series in x. Moreover, since $|x| \leqslant 1$, these power series are well conditioned, and are not dominated by the higher order terms. Using this power series form, the following integrals need to be evaluated.

$$I_k^i(t) = \int_0^t e^{\lambda_k(t-\tau)} x^i(\tau)\, d\tau, \qquad i=0,\ 1,\ 2,\ \ldots\ N+1 \quad (B.39)$$

This can be achieved using the following reverse recurrence relation,

$$I_k^{i-1}(t) = \frac{I_k^i(t) - \left[\frac{e^{\lambda_k(t-\tau)}}{-\lambda k} x^i(\tau) \right]_0^t}{\frac{i}{\lambda_k} \frac{dx}{d\tau}(\tau)} \qquad (B.40)$$

Note that this reverse relation is numerically stable, whereas the forward version of the relation is not. The first value of the reverse recurrence relation, I_k^{N+1}, was obtained using numerical integration, with the integration step chosen to give a result to machine accuracy. Subsequent values were then obtained using (B.40). This was found to give extremely accurate results, even for very high order (e.g 50^{th} order) polynomials, while still giving good efficiency.

B.3 Choice of Basis Functions

In equation (B.1) above, the input u(t) is expanded in a series of orthogonal functions, $u_i(t)$. Various sets of orthogonal functions are commonly used in such expansions. These include,

1. Walsh series
2. Power series
3. Cosine series
4. Chebyshev series

Each of these was considered, in order to decide which was most suitable for the present application.

Taking Walsh functions [14] first of all, these consist of a sequence of steps, and this simple form facilitates their use. However, steps cannot be realised in real systems, resulting in difficulties when using inputs based on Walsh series. Moreover, the discontinuity at a step leads to power at high frequencies, which is known to be undesirable for helicopter inputs.

In contrast, power series involve smooth functions and do not introduce discontinuities. However, power series are very sensitive to the coefficients, a_i, of the higher order terms. Consider the following :

$$u(t) = \sum_{i=0}^{N} a_i t^i \qquad \Rightarrow \qquad \frac{du(t)}{da_i} = t^i \qquad (B.41)$$

Hence, if for example, t=10 sec, then $du/da_1 = 10$, while $du/da_{10} = 10^{10}$. This large spread in the values of du/da_i results in severe ill conditioning.

Cosine series also do not introduce discontinuities. Moreover, the derivatives, du/da_i are bounded, as follows,

$$u(t) = \sum_{i=0}^{N} a_i \cos(i+\tfrac{1}{2})\frac{\pi t}{T} \quad \Rightarrow \quad \frac{du(t)}{da_i} = \cos(i+\tfrac{1}{2})\frac{\pi t}{T} \qquad (B.42)$$

Hence,

$$-1 \leqslant \frac{du(t)}{da_i} \leqslant 1 \qquad (B.43)$$

where, T is the duration of the input being used.

Finally, Chebyshev series are closely related to cosine series, and share may of their properties. Chebyshev polynomials do not introduce discontinuities, and also have bounded derivatives, as follows,

$$u(t) = \sum_{i=0}^{N} a_i T_i(x(t)) \Rightarrow \frac{du(t)}{da_i} = T_i(x(t)) \qquad (B.44)$$

where,

$x(t) = (2t - T)/T$, i.e. $-1 \leqslant x(t) \leqslant 1$

T = duration of input

$T_i(x(t))$ is the Chebyshev polynomial of order i.

Hence,

$$-1 \leqslant \frac{du(t)}{da_i} \leqslant 1 \qquad (B.45)$$

Both cosine and Chebyshev series appear to be reasonable basis functions for the current application, although it was felt that cosine series were more suitable for problems involving periodicity, whereas the present problem uses aperiodic inputs.

However, difficulties were encountered when using these series with unstable systems. In the unstable case, the system response to the input components, $u_i(t)$, may be divergent, even when the response to the full input, $u(t)$, is convergent. As a result, constraints placed on the system responses (see Chapter 3) tend to produce an ill-conditioned optimisation problem. For high order Chebyshev and cosine series, this ill conditioning was sufficient to prevent convergence of the input design algorithms.

In equations (B.15) – (B.24), it is shown that the system response may be considered from a model standpoint. This allows the unstable sections of the system to be isolated. Inputs, $u_i(t)$, can then be designed which will stabilise these sections.

Consider a system with a single unstable mode, having eigenvalue λ_1, say. Then the components of the system response due to this mode may be expressed as follows (see equation (B.24)),

$$h^i(t) = \int_0^t e^{\lambda_1(t-\tau)} u_i(\tau)\, d\tau \qquad (B.46)$$

and,

$$g^i(t) = \int_0^t (t-\tau)\, e^{\lambda_1(t-\tau)} u_i(\tau)\, d\tau \qquad (B.47)$$

Taking Laplace transforms gives,

$$L\{h^i(t)\} = H^i(s) = \frac{u_i(s)}{s - \lambda_1} \qquad (B.48)$$

$$L\{g^i(t)\} = G^i(s) = \frac{u_i(s)}{(s - \lambda_1)^2} \qquad (B.49)$$

Choosing $g^i(t)$ to be a suitable function, $\sigma_i(t)$, gives,

$$U_i(s) = (s - \lambda_1)^2 \Sigma_i(s) \qquad (B.50)$$

$$= s^2\Sigma_i(s) - 2\lambda_1 s\Sigma_i(s) + \lambda_1^2\Sigma_i(s) \qquad (B.51)$$

where $\Sigma_i(s) = L\{\sigma_i(t)\}$

If $\sigma_i(0) = 0$, and $d\sigma_i(0)/dt = 0$, then,

$$u_i(t) = \frac{d^2\sigma_i}{dt^2}(t) - 2\lambda_1 \frac{d\sigma_i}{dt}(t) + \lambda_1^2\sigma_i(t) \qquad (B.52)$$

Moreover,

$$H^i(s) = (s - \lambda_1)\Sigma_i(s) \tag{B.53}$$

$$\Rightarrow \quad h^i(t) = \frac{d\sigma_i}{dt}(t) - \lambda_1\sigma_i(t) \tag{B.54}$$

Hence, if $\sigma_i(t)$, $d\sigma_i(t)/dt$, and $d^2\sigma_i(t)/dt^2$ are not divergent, then neither are $h^i(t)$, $g^i(t)$, and $u_i(t)$.

In the present work, Chebyshev polynomials were taken as the basis for $\sigma_i(t)$, using the following,

$$\sigma_i(t) = T_i(x(t)) - T_i(x(0)) - \frac{dT_i}{dt}(x(0))\, t \tag{B.55}$$

Giving $\sigma_i(0) = 0$, $d\sigma_i(0)/dt = 0$, and $\sigma_i(t)$, $d\sigma_i(t)/dt$, $d^2\sigma_i(t)/dt^2$ not divergent.

This technique was found to give significantly improved conditioning in the input design algorithms. It is easily extended to the more general case of multiple, complex eigenvalues.

Figure B.1

The function, f(y), used as the basis for the smooth, continuous amplitude constraint

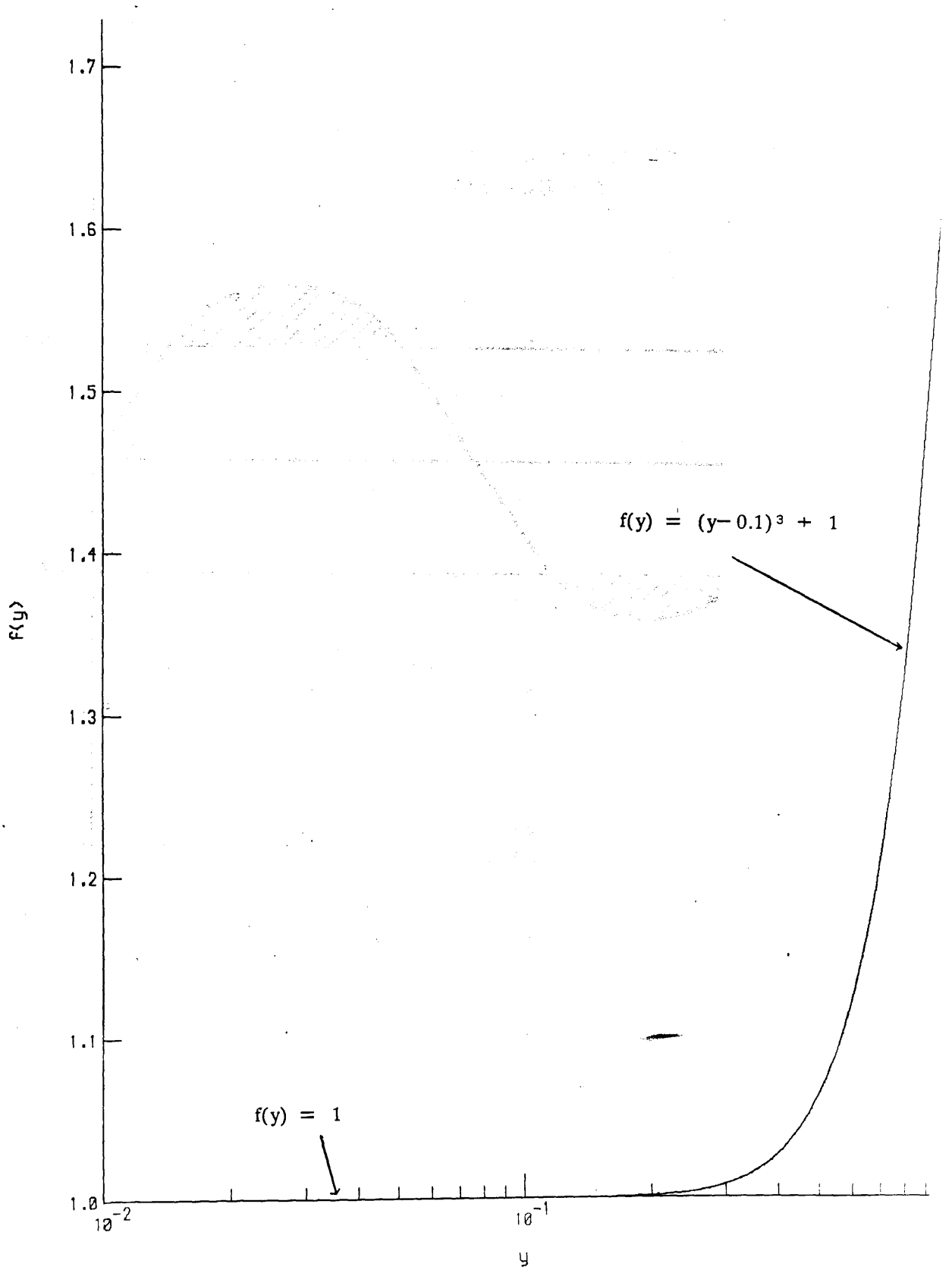
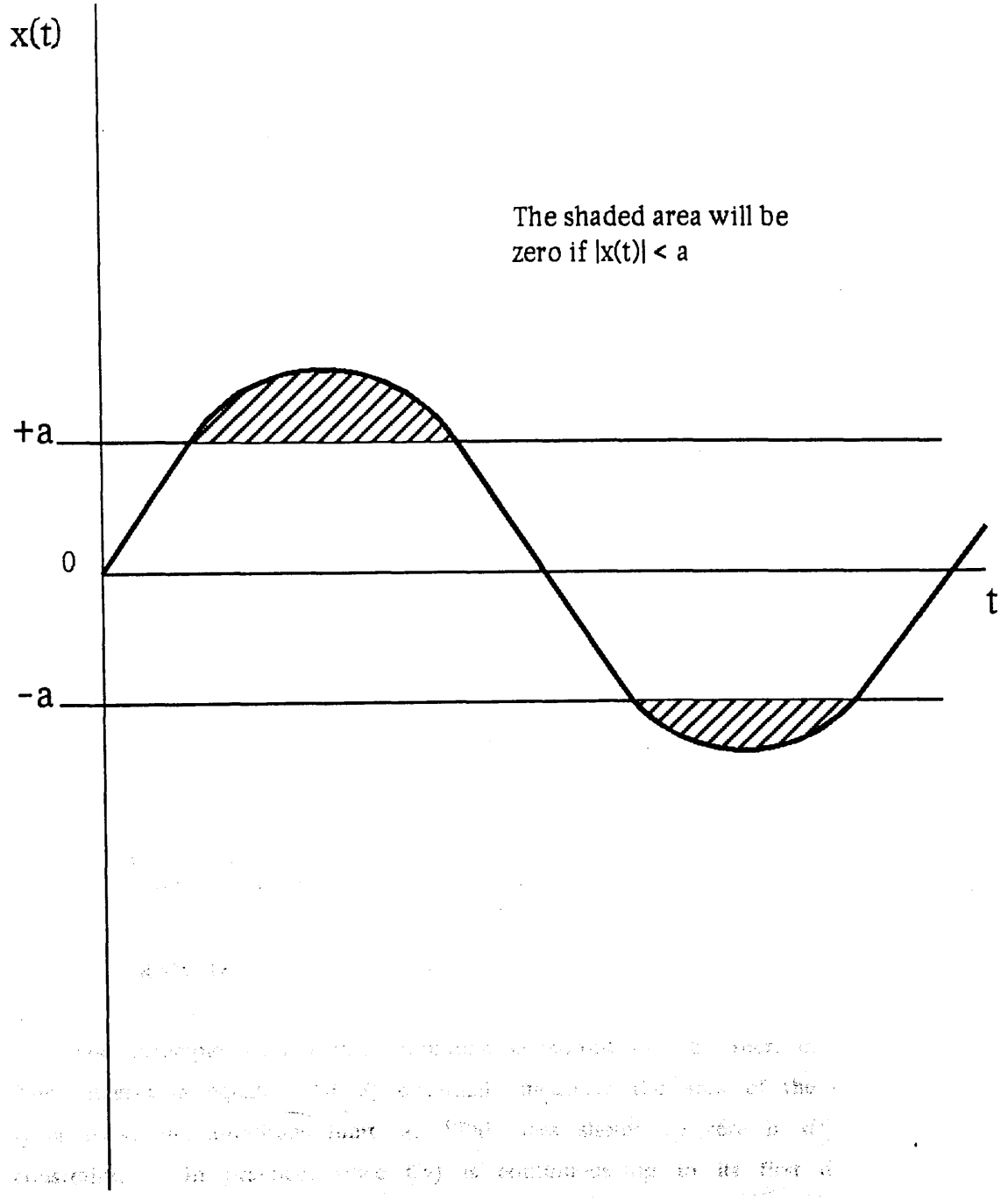


Figure B.2

An outline of the arrangement used in the smooth, continuous amplitude constraint. For a function, x(t), constrained to lie between +a and −a, then the area marked should be zero.



B.4 Implementation of Amplitude Constraints

Two amplitude constraints are used in the input design algorithms studied : the output amplitude constraint, and the response robustness constraint. Since amplitude constraints are essentially discontinuous in nature, they can lead to optimisation problems which are prone to ill–conditioning. In order to improve the conditioning, and give a simple, efficient constraint, the following technique was used.

The function, f(y), shown in figure B.1, formed the basis of the amplitude constraints used. A Chebyshev series was used to approximate this function. It is defined on $0 \leqslant y \leqslant 1$, and is continuous up to, and including, its first derivative. Using f(y), an amplitude constraint, $|x(t)| \leqslant a$, $0 \leqslant t \leqslant T$, may be implemented as follows.

1. Let $w(t) = \dfrac{x^2(t)}{9\,a^2}$ (B.56)

2. Let $y(t) = \dfrac{w(t)}{w(t) + 1}$ (B.57)

   y(t) is the bilinear transform of w(t), giving $0 \leqslant y(t) \leqslant 1$.

   w(t) is such that $y(t) = 0.1$ when $|x(t)| = a$. From figure B.1, 0.1 is the break–point of f(y).

3. Then $\dfrac{1}{T}\int_0^T f(y(t))\,dt = 1$ (B.58)

   gives $|x(t)| \leqslant a$, $0 \leqslant t \leqslant T$

The principle behind this amplitude constraint can be seen in figure B.2. The integral in equation (B.58) essentially measures the area of the curve, x(t), lying above the amplitude limit, a. This area should be zero if x(t) meets the constraint. In practice, since f(y) is continuous up to its first derivative in order to improve the conditioning of the optimisation problem. Hence, it does not have sufficient resolution to show small violations of the amplitude constraint. However, in the present application this presented no problems, and the improved conditioning resulting from this approach was found to be extremely valuable.

REFERENCES FOR APPENDIX B

1. Numerical Algorithms Group 'NAG FORTRAN Library Manual, Mark 12', Numerical Algorithms Group, Oxford 1987

2. Richards, R.J. 'An introduction to dynamics and control' Longman, New York 1979

3. Moler, C.; Van Loan, C. 'Nineteen dubious ways to compute the exponential of a matrix' SIAM Review, Vol 20, pp801–837, 1978

4. Bickart, T.A. 'Matrix exponential : approximation by truncated power series' Proc IEEE Vol 56, pp872–873, 1968

5. Levis, A.H. 'Some computational aspects of the matrix exponential' IEEE Trans Aut Control Vol AU–14, pp410–411, 1969

6. Singhal, K.; Vlach, J. 'Computation of time–domain response by numerical inversion of the laplace transform' J. Franklin Inst, Vol 299, pp109–126, 1975

7. Ward, R.C. 'Numerical computation of the matrix exponential with accuracy estimate' SIAM J. Num Anal, Vol 14, pp600–610, 1977

8. Zakian, V. 'Rational approximants to the matrix exponential' Electronics Letters, Vol 6, pp814–815, 1970

9. Fair, W.; Luke, Y.L. 'Pade approximations to the operator exponential' Num Math, Vol 14, pp379–382, 1970

10. Markussen, C.R.; Distefano, J.J. 'Evaluation of four methods for computing parameter sensitivities in dynamic system models' Math Bio, Vol 61, pp135–148, 1982

11. Reid, G.; Maybeck, P.; Asher, R.; Dillow, J. 'An algebraic representation of parameter sensitivity in linear time–invariant systems' J. Franklin Inst, Vol 301, pp121–141, 1976

12. Reid, G. 'Computation of optimal inputs for parameter identification with Walsh or other orthogonal basis functions' Paper THRY 5–2 presented at the Joint Automatic Control Conference, 1976

13. Conte, S.D.; de Boor, C. 'Elementary numerical analysis' McGraw–Hill, Singapore 1980

14. Walsh, J.L. 'A closed set of normal orthogonal functions' Am J. Math, Vol 45, pp5–24, 1923