

UNCLASSIFIED

AD

419695

DEFENSE DOCUMENTATION CENTER

FOR

SCIENTIFIC AND TECHNICAL INFORMATION

CAMERON STATION, ALEXANDRIA, VIRGINIA



UNCLASSIFIED

NOTICE: When government or other drawings, specifications or other data are used for any purpose other than in connection with a definitely related government procurement operation, the U. S. Government thereby incurs no responsibility, nor any obligation whatsoever; and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use or sell any patented invention that may in any way be related thereto.

AFOSR 4669

64-5

419695

CATALOGED BY DDC
AS AD No. 419695

August 1963



USCEC Rept 98-101
EE-27

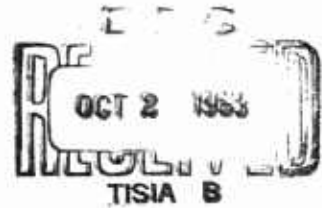
UNIVERSITY OF SOUTHERN CALIFORNIA
SCHOOL OF ENGINEERING

Final Report

RESEARCH IN COMPUTER DECISION PROCESS

Gregory O. Young
Robert McGhee

AF-AFOSR 62-69



ELECTRICAL ENGINEERING DEPARTMENT



August 1963

USCEC Report 98-101
EE-27

FINAL REPORT

RESEARCH IN COMPUTER DECISION PROCESS

Gregory O. Young, Principal Investigator
Robert McGhee

ELECTRICAL ENGINEERING DEPARTMENT
UNIVERSITY OF SOUTHERN CALIFORNIA
LOS ANGELES 7, CALIFORNIA

AF-AFOSR-62-69

Prepared
for

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
OFFICE OF AEROSPACE RESEARCH
UNITED STATES AIR FORCE
WASHINGTON 25, D.C.

ABSTRACT

This report summarizes the research accomplished under Grant AF-AFCSR-62-69. It consists of two separate parts representing two distinct investigations carried out within the scope of the grant. The first part of the report is concerned with optimization criteria in systems design; the second deals with the identification problem for nonlinear systems. These two topics correspond to the abstracts included in the Summary Report, USCEC 98-201, EE-24 submitted in April, 1963.

Research in both of the areas described herein is continuing under the renewal Grant AF-AFOSR-75-63. Mr. R. B. McGhee has completed a dissertation based upon further development of Part II of this report. This dissertation is presently being prepared for journal publication. As further significant results are obtained, they will also be submitted for publication in the appropriate technical journals.

TABLE OF CONTENTS

	Page
ABSTRACT	ii
LIST OF ILLUSTRATIONS	iv
LIST OF TABLES	iv
PART I. OPTIMIZATION CRITERIA	1
1. INTRODUCTION	1
2. SYSTEMS EVALUATION	5
3. THE INFORMATION RATE CRITERION	7
4. LINEAR INFORMATION THEORY	9
5. EFFECT OF FILTERING ON ENTROPY	14
PART II. ESTIMATION AND OPTIMIZATION OF NONLINEAR DYNAMIC SYSTEM PARAMETERS BY REGRESSION ANALYSIS METHODS	36
1. INTRODUCTION	36
2. LINEAR REGRESSION ANALYSIS	39
3. NONLINEAR REGRESSION ANALYSIS	46
4. STATISTICAL PROPERTIES OF NONLINEAR LEAST SQUARES PARAMETERS ESTIMATES	67
5. AN AUTOMATIC COMPUTATIONAL ALGORITHM PARAMETER IDENTIFICATION BY NONLINEAR REGRESSION ANALYSIS	75
6. EXPERIMENTAL RESULTS	82
7. SUMMARY AND CONCLUSIONS	92
8. REFERENCES	93

LIST OF ILLUSTRATIONS

PART I.

Figure		Page
1	Communications System.	9
2	General System.	9
3	Signal and Noise Spectrum for Simple System.	21
4	Typical Servo Block Diagram.	23
5	Simplified Servo Block Diagram.	24

PART II.

1	Flow Chart for Binary Search.	78
2	Flow Diagram for Parameter Estimation by Nonlinear Regression Analysis.	79
3	Response Curve Fitted to Noisy Data by the Computer.	83

LIST OF TABLES

Table		
1	Results of Gauss-Newton Iteration Using Exact Data.	84
2	Sequence of Parameter Values Obtained Using Gauss-Newton Iteration Stabilized by Binary Scale Factor Adjustment.	86
3	Parameter Estimates Obtained from Noisy Data.	89
4	Effect of Length of Record on Estimation Accuracy.	91

PART I.
OPTIMIZATION CRITERIA
1. INTRODUCTION

There is a fundamental difference between information theory and the theory of control. In information theory, it is desired to operate on an output signal so that as much information about the desired input signal is obtained as possible. In control, the output is often asked to reproduce some desired signal, usually the input, as closely as possible. If systems are optimized according to these two criteria, the results are in general not the same. The difficulty in formulating a suitable optimization criterion for systems is mainly due to the wide range of purposes for which systems are used. It would seem hardly reasonable, therefore, to expect to define a unique optimization criterion or measure of quality; at best one can only devise criteria which are satisfactory over a wide range of applications. The most general description of the purpose of a system is perhaps that it should give information about the input (or other desired signal). If the information desired is simply a reproduction of the input, or other desired signal, then we have a problem in control. Sometimes, however, it may be required to extract all of the information that is implicit in the output without regard to the possible complexity of the interpretation process. This type of requirement is particularly important when it is difficult or impossible to repeat an experiment. Such cases are common in practice, and under these circumstances it is a prime necessity that the information gained be maximum whether or not the output exactly resembles

the input. In such a case, the quality of the output is more important than that it exactly resemble the input. In such a case, the output must be interpreted or reconstructed in some way. Herein lies the difficulty of instrumenting an appropriate interpretive system. In other words, somehow the information must be made useful to the observer. Such a procedure is a type of decoding.

One of the purposes of this research is to consider various optimization criteria so as to determine which is best for the constrained system under consideration. Once a system has been optimized, it is very desirable to be able to evaluate it by comparing it to the theoretically optimum system. In other words, it should be possible to determine the optimum system performance without regard to any specific system configuration. Once this is done, a precise system evaluation can then be made. Such a procedure also has the valuable property that additional research on the proposed system can be made to yield the maximum return for the minimum cost. In other words, the effort can be adjusted so as to take advantage of the law of diminishing returns. If the proposed system is far from this theoretical optimum, additional effort is indicated, whereas if the proposed system is near to the theoretical optimum, the additional effort may not be worthwhile. Furthermore, the systems evaluation techniques to be proposed will be useful in the comparison of systems on an absolute basis, so that if more than one system is under consideration, the optimum system can be appropriately chosen.

The most common optimization criterion used in system design, particularly feed back systems design, is minimization of mean square (or integral square) error. The disadvantages of such a criterion are well known, and include: the requirement of long time averaging, so that the system is essentially optimized only for steady state performances; excessive emphasis of large amplitude but short time deviations of the output signal from the mean of the desired signal, as well as other disadvantages. The principal advantage of this criterion is that it generally leads to mathematically tractable expressions, where as related criteria such as the minimization of absolute error do not. When a system is subject to stochastic disturbances such as noise, random load disturbances, etc., a commonly used criterion of goodness is the maximization of signal-to-noise ratio at certain points in the system. Other arbitrary criteria which have been used include specification of some minimum likelihood of loss of track in tracking systems, minimum probability of exceeding thresholds on certain variables, certain statistical detection criteria, etc. Let us consider some of the disadvantages of these arbitrary criteria. A disadvantage of the maximization of signal-to-noise ratio criterion and related criteria is that no constraint is put upon the amount of time required to perform the desired maximization. For example, the signal-to-noise ratio can be made to approach infinity by averaging (that is, integrating) over an infinite period of time. In the vast majority of engineering systems, however, it is required that the system perform optimally in a specified period of time, or, more commonly, that the system

reach its optimum performance in the minimum possible time. Thus the question of a system bandwidth, response time, etc., is equally as important as maximization of signal-to-noise ratio. The same disadvantage applies to any criterion such as minimization of mean square error which has no time constraint included in it. A disadvantage of setting thresholding limits and other arbitrary criteria which put bounds upon system behavior is that they are arbitrary, and therefore do not necessarily lead to an optimum system in the sense that all errors are minimized. As an example, consider the typical specifications and criteria set upon antenna system design. It is common to specify somewhat arbitrarily the beamwidth, sidelobe level, antenna gain, resolution capability, and other properties of the system. All these specifications are actually somewhat crude attempts to maximize the rate at which desired information is being acquired by the antenna and processed by the system. A common requirement is that the beamwidth be as narrow as possible for the specified aperture. In a searching radar, if there is a specified maximum search time, a narrow beam antenna must search at a high angular rate in order to get adequate coverage during the specified frame time. It could very well be that the rate at which information is being acquired would be improved by using a wider beam antenna and searching at a slower rate. As another example, it is commonly required that the sidelobes of an antenna be small compared to the main beam. In fact, it is often stated that the ideal antenna should have no sidelobes at all. In the case of ground mapping, however, information is gained through the sidelobes as well as through the main beam, and if this information could be so processed as to be made useful, the sidelobes might

actually be desirable.

It is therefore concluded that the optimum optimization criterion for any system is the maximization of the information rate at the desired output. This is a most general statement, but it is valid, since it is always possible in principle to define just what information is desired from the system. It then follows that that system which maximizes the desired information rate will be the optimum one. If the maximization of information rate is to be used as an optimization criterion, the optimization must take place subject to the constraints which exist on the system. The most common of such constraints is system cost. There are many types of cost functions: dollar cost is just one of them, although most cost functions ultimately have some sort of economic basis. Typical cost functions include size, weight, power consumption, complexity of equipment, etc. Reliability is also of great practical importance. In general, the system must be optimized subject to some upper bound on cost and lower bound on reliability. Since reliability is related to complexity of equipment, it can also often be specified as a cost function. Thus, the proposed optimization criterion will be the maximization of information rate subject to some maximum cost constraint where the cost constraint may include any or all of the preceding specified functions.

2. SYSTEMS EVALUATION

Note that the proposed criterion subject to constraints is well suited to the comparison of or evaluation of systems. Clearly, two systems can be readily compared as to their relative merit, or a single system can be

compared to the theoretical optimum. The criterion can also be used directly for system optimization, except that when the system is so optimized, the information acquired is not necessarily in its most usable form. The example of the information available in the sidelobes of an antenna system illustrates this possibility. Also, the determination of the theoretical optimum for a certain class of systems does not necessarily mean that such a system can be instrumented. The classical example of problems of this sort occurs in the problem of coding. Shannon has shown that even in a noisy communication channel, it is possible to encode a message source in such a way that the information rate will approach the channel capacity of the system with the probability of error being negligibly small. Thus one need only determine the channel capacity of the system exclusive of the source to determine the optimum rate at which information can theoretically be transmitted. However, the theorem says that it is "possible to encode" so as to reach this theoretical maximum; it does not specify what this optimum code should be. As a matter of fact, a practical code which even approaches the optimum has not as yet been devised.

The aforementioned difficulties do not affect the utility of the criterion for system evaluation and comparison. When the criterion is used for system optimization, the requirement of maximum utility can be imposed as a constraint on the system. For example, a constraint which is commonly imposed on most criteria is that of physical realizability, since a system which is optimized without such a constraint may end up by being not physically realizable and hence not useful. Another way around this

difficulty is to specify the form of the system to be used apriori, leaving only certain parameters (such as the coefficients in the system differential equation) undetermined. The system is then optimized by maximizing the information rate with respect to the variable parameters subject to whatever constraints are necessary. This approach also has the advantage of initially specifying the order of the system differential equation and hence the complexity of the system.

3. THE INFORMATION RATE CRITERION

Information content or entropy is a concept that deals with freedom of choice. The greater the number of choices available, the greater the entropy. In the case of a signal which can assume only certain discrete levels, the entropy per sample point increases with the number of discrete levels possible. Indeed, if the probability of each level occurring is equally likely, then the entropy is simply the logarithm of the number of possible levels. If it were not for the disturbing influence of noise, a continuous signal would be capable of being quantized into an infinite number of distinguishable levels, even if the overall range in signal amplitude were finite. If the minimum distinguishable difference in levels is set, and the dynamic range is finite, an upper bound on the entropy is set. The maximum entropy is, therefore, a function of the minimum distinguishable interval where the latter is a function of the signal-to-noise ratio. It is very difficult to set any absolute criterion for resolution. For example, consider the resolution of two targets by an antenna. Targets

commonly are said to be resolved if they are a half-beamwidth (or more) apart. But this designation is purely arbitrary and again depends on the signal-to-noise ratio. Such a separation may be wholly inadequate for low signal-to-noise ratios; conversely targets much closer than this angular separation may be resolved at high signal-to-noise ratios. Rather than attempt to specify arbitrarily some minimum distinguishable interval or difference in levels, a far more sophisticated approach would be to specify information content in terms of the probability distributions of signal and noise in such a manner that the useful total entropy or information rate is, in effect, a function of signal-to-noise ratio. The total entropy is related to not only the signal-to-noise ratio but also the bandwidth and time duration of the message. This can be seen from Shannon Sampling theorem which states that the number of samples in the message is proportional to the bandwidth-time product. The information rate is the total entropy divided by the time duration of the message (provided the message is quite long). Consequently the rate is proportional to the bandwidth of the signal and the signal-to-noise ratio. There is no absolute level of distinguishability in such a definition. The information content and rate simply improve as the signal-to-noise ratio improves and/or the bandwidth increases. The classical Shannon communication system is represented in Figure 1. In this system the nature of the source is assumed to be known completely (that is, its statistics are known)*. The encoder includes the transmitter, the modulator, and whatever actual encoding scheme is

* It is true that when a specific message is transmitted its exact form is known, however, one cannot design a system for that message alone, thus, a priori the nature of any message transmitted by a known source can at best be known statistically.

used preceding transmission. The link is the actual transmission channel between transmitter and receiver. The decoder performs the inverse operation to that of the encoder and includes the receiver and whatever demodulating and decoding apparatus is necessary. For the systems presently under consideration, an appropriate representation will be Figure 2.

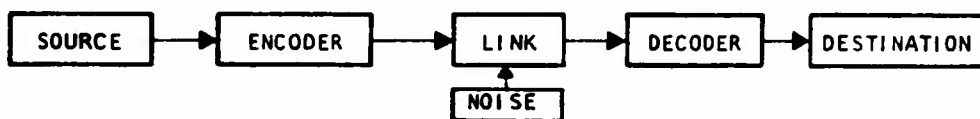


Figure 1. Communications System

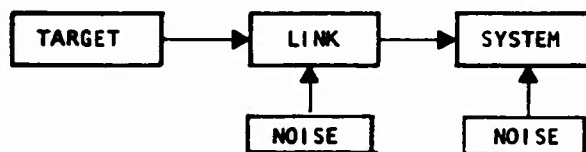


Figure 2. General System

In general, signal and noise may appear on the input of a system, where the link is essentially an adder, but, different from the classical communications system in Figure 1, the noise may also be added in the system itself as shown.

5. LINEAR INFORMATION THEORY

The information content, or entropy, of a random time function $x(t)$ with a (continuous) amplitude probability distribution $p(x)$ is by definition

$$H(x) = - \int_{-\infty}^{\infty} p(x) \ln p(x) dx \quad (1)$$

The entropy $H(x)$ is the information contained in a sample of x at time t . For example, the entropy of a Gaussian distribution with variance σ^2 and mean zero is given by

$$H = \frac{1}{2} \ln (2\pi e\sigma^2) \quad (2)$$

This, incidently, is the maximum entropy per sample of all continuous distributions with a given variance. The information contained in the total function $x(t)$, which is assumed to be of duration T , is found from the joint entropy function

$$H(x_1, x_2, \dots, x_n) = - \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} p(x_1, x_2, \dots, x_n) \ln p(x_1, x_2, \dots, x_n) \prod_{i=1}^n dx_i \quad (3)$$

where the subscripts denote the instants of time at which the samples are taken. If the time function has a finite frequency bandwidth B , $2TB$ samples will completely represent the time function. It is important to recognize that a function of finite duration cannot have a Fourier transform which extends over a finite domain, so that the sampling theorem

$$x(t) = \sum_{n=-TB}^{TB} x(nt_0) \frac{\sin (2Bt - n)}{(2Bt - n)} \quad (4)$$

$$t_0 = \frac{1}{2B} \quad (5)$$

is only an approximation to the true $x(t)$. This approximation is very close when the function and its transform fall off without abrupt discontinuities (e.g., the Gaussian function). Equation (4) therefore represents a

• Actually $2TB + 1$ samples are needed, but $2TB + 1 \approx 2TB$ for $TB \gg 1$.

function whose principal spectral energy lies in an audio band of width B . Note that if $x(t)$ is of duration T , approximately $\frac{T}{t_0} = 2BT$ samples are needed to completely define the function. If the samples $x(t_i) = x_i$ are statistically independent, Equation (3) becomes

$$\begin{aligned}
 H_{\text{total}} = H(x_1, x_2, \dots, x_n) &= - \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \ln \left[\prod_{j=1}^n p(x_j) \right] \prod_{i=1}^n p(x_i) dx_i \\
 &= \sum_{j=1}^n \int_{-\infty}^{\infty} p(x_j) \ln p(x_j) dx_j = \sum_{j=1}^n H(x_j) \quad (6)
 \end{aligned}$$

In this case, the total entropy can be found by adding the entropies at each sample point. If $p(x)$ is stationary in time, the entropy at each sample point is the same, and the total entropy is simply the entropy per sample times the total number of sample points. The successive samples will be uncorrelated if the spectrum of the signal is flat over the specified bandwidth. Zero correlation implies statistical independence for most distributions of interest, such as the Gaussian distribution.

Information rate is defined as

$$R = \lim_{T \rightarrow \infty} \frac{H_{\text{total}}}{T} \quad (7)$$

where H_{total} is the total entropy of the signal $x(t)$ of duration T . If the successive samples are independent and the process is stationary, then, from Equation (6)

$$R = \lim_{T \rightarrow \infty} \frac{1}{T} 2TB H(x) = 2B H(x) \quad (8)$$

Under these circumstances, the information rate can be thought of as the entropy per sample times the number of samples per second.

The "useful entropy" at the receiver is the received entropy $H(y)$ less the received conditional entropy $H(y|x)$ where the distribution of received signals y depends on the transmitted signal x . By definition,

$$H(y|x) = - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x,y) \ln p(y|x) dy dx \quad (9)$$

where $p(x,y)$ is the joint distribution of x and y and $p(y|x)$ is the conditional distribution of y given x . The useful entropy is then

$$H_u = H(y) - H(y|x) \quad (10)$$

It can be shown that $H(y|x) \rightarrow 0$, or $H_u \rightarrow H(y)$ when x is linearly related to y and that $H(y|x) \rightarrow H(y)$ or $H_u \rightarrow 0$ when x and y are independent. y will differ from x because of noise and distortion, so that the useful entropy decreases with increasing noise. The concepts of useful entropy and its time rate of change, called "information rate", are very important. In this report, the system is said to be optimized when these quantities are maximized on the output. It can be shown that $H(y|x) = H(N)$ if x is the sum of signal and noise ($x = S + N$) and the signal and noise are independent and that $H(y) = H(x)$ to within an additive constant if y and x are linearly related. If x consists of the sum of independent signal S and noise N then the useful entropy at the receiver is the signal-plus-noise entropy less the noise entropy, or

$$H_u = H_{S+N} - H_N \quad (11)$$

The useful information rate at the receiver is then

$$R = 2B(H_{S+N} - H_N) \quad (12)$$

The maximum rate, or "channel capacity" for a given average transmitter power P_S and noise power P_N is achieved when both the signal and noise are Gaussian and independent. From Equation (2), Equation (12) becomes

$$\begin{aligned} C = R_{\max} &= 2B \frac{1}{2} \left\{ \ln [2\pi e(P_S + P_N)] - \ln (2\pi e P_N) \right\} \\ &= B \ln \left(1 + \frac{P_S}{P_N} \right) \end{aligned} \quad (13)$$

where "C" is the "channel capacity". Equation (13) is one of the most widely used (and misused) formulas in information theory. It is strictly valid only when: (1) the signal is Gaussian, (2) the noise is Gaussian and additive to the signal, (3) the noise and signal are statistically independent, (4) all processes are stationary, and (5) the signal, noise, and signal-plus-noise spectra are flat and limited to band B. Equation (13) illustrates one of the advantages of the maximization of information rate over the maximization of signal-to-noise ratio and related criteria. The rate or channel capacity is clearly proportional to the signal-to-noise ratio but it is also proportional to the modulation bandwidth B. Thus, maximizing the information rate maximizes the signal-to-noise ratio subject to a restraint on system modulation bandwidth or response time.

Equation (13) is an example of the calculation of information rate. In the work which follows, more complicated expressions will have to be evaluated. The more complicated expressions arise because the restrictive conditions mentioned following Equation (13) are no longer satisfied. One of the principal complicating effects when information theory is applied to system design is that of filtering.

5. EFFECT OF FILTERING ON ENTROPY

Even if the time samples are independent at the input to filter, they will in general not be independent at the output, since filtering produces a non-flat output spectrum. Only in the special case where the filter attenuation characteristic is flat will the output time samples remain independent. The joint entropy on the output is, therefore, no longer the sum of the marginal entropies, and the entropy per degree of freedom is difficult to find. It is possible to find the entropy for a particular time sample, and if the output process is stationary (which it will be only in the steady state) the entropy at each sample is the same. The total entropy is not the sum of the entropies at each sample point, however, since the successive sample values are not independent. As an example of such a calculation, suppose $x(t)$ is Gaussian with mean zero, and spectral density and $u(f)$ is the unit step function

$$G_x(f) = w_0 [u(f+B) - u(f-B)] \quad (14)$$

Let $x(t)$ be the input to a linear filter whose transfer function is $T(f)$. The output $y(t)$ will also be Gaussian since a linear operation

on a Gaussian variable yields a Gaussian variable. The output spectral density is

$$G_y(f) = G_x(f) |T(f)|^2 \quad (15)$$

and the output variance is

$$\sigma_y^2 = \int_0^{\infty} G_y(f) df = \int_0^{\infty} G_x(f) |T(f)|^2 df = w_0 \int_0^B |T(f)|^2 df \quad (16)$$

The variance of the input is

$$\sigma_x^2 = \int_0^{\infty} G_x(f) df = w_0 \int_0^B df = w_0 B \quad (17)$$

The difference between the input and output entropies is

$$H(y) - H(x) = \frac{1}{2} \ln \left(\frac{\sigma_y^2}{\sigma_x^2} \right) \quad (18)$$

Substituting Equations (16) and (17) in Equation (18),

$$H(y) = H(x) + \frac{1}{2} \ln \left[\frac{1}{B} \int_0^B |T(f)|^2 df \right] \quad (19)$$

Equation (19) is the entropy for a particular time sample. It is not the average entropy per sample except for special cases and thus will not yield the total entropy when multiplied by the number of samples. To find the average entropy per sample for an arbitrary filter, it is more convenient to consider frequency rather than time sampling. The transformation between frequency and time sampling points is orthogonal, hence the

Jacobian of the transformation has the absolute value of unity. Since the difference between the entropies of two variables related by a linear transformation is simply the logarithm of the absolute value of the Jacobian, it follows that the joint entropy and/or total entropy of the function is the same whether expressed in terms of frequency or time samples. The entropy per degree of freedom after frequency filtering can, therefore, be calculated from the effect of frequency filters on frequency rather than time samples. This greatly simplifies the calculation of the output entropy, since the relationship between input (x) and output (y) frequency samples is the simple product

$$Y(j\omega) = T(j\omega) X(j\omega) \quad (20)$$

The frequency samples at the filter output are independent since multiplication of the input variables by a nonrandom function does not affect their independence. If the frequency samples are stationary in frequency as well as independent, the total entropy can be found by simply multiplying the average entropy per sample by the total number of samples. Although $X(j\omega)$ may be "stationary in frequency," $Y(j\omega)$ is not in general. The term "stationary in frequency" means that the probability distribution of amplitudes in the frequency domain is not a function of the origin in frequency. Equation (26) can be used to generate a set of linear equations relating Y_i and X_i . The successive samples are independent, so that the total or joint entropy is the sum of the entropies, and the entropy per degree of freedom is this sum divided by $2TB$.

It can be shown that the joint entropy after filtering is related to the joint entropy before filtering by

$$H(Y_1, Y_2, \dots, Y_{TB}) = H(X_1, X_2, \dots, X_{TB}) + \sum_{i=1}^{TB} \ln |T(f_i)|^2 \quad (21)$$

Note that the upper limit on the sum is TB rather than $2TB$ since of the $2TB$ independent samples needed to define the function, TB are associated with the real part and TB with the imaginary part. Since the samples are independent, the average entropy per sample is

$$\begin{aligned} H(Y) &= \frac{H(Y_1, Y_2, \dots, Y_{TB})}{2TB} = \sum_{i=1}^{TB} \frac{H(Y_i)}{2TB} \\ &= \sum_{i=1}^{TB} \frac{H(X_i)}{2TB} + \frac{1}{2TB} \sum_{i=1}^{TB} \ln |T(f_i)|^2 \end{aligned} \quad (22)$$

If $TB \gg 1$, the frequency samples are equally and closely spaced in the band B . Let

$$\Delta f = \frac{1}{T} \quad (23)$$

$$\begin{aligned} H(Y) &= (H(X) + \frac{1}{2TB} T \sum_{i=1}^{TB} \ln |T(f_i)|^2 \Delta f) \\ &\approx H(X) + \frac{1}{2B} \int_0^B \ln |T(f)|^2 df \end{aligned} \quad (24)$$

Equation (24) actually represents the average entropy per degree of freedom on the output, where a "degree of freedom" refers to either a time sample or a frequency sample, since the average entropy per sample is the same for both types of samples. It is interesting to compare Equation (24) with Equation (19). In general, these two results are different, since the log of the integral is not equal to the integral of the log. In the special case where the filter characteristic is constant across the band B the two are the same, however.

Note that the successive input time samples are independent if the power spectral density of $x(t)$ is constant over B. In other words,

$$G_x(f) = \frac{2}{T} \left| \mathcal{F}_T [x(t)] \right|^2 = w_0 [u(f+B) - u(f-B)] \quad (25)$$

where w_0 is a constant. This condition is sufficient to insure that samples spaced $t_0 = \frac{1}{2B}$ apart be uncorrelated. Likewise, the successive input frequency samples are independent if the power spectral density of $X(f)$ is constant over T. In other words,

$$G_x(t) = \frac{1}{B} \left| \mathcal{F}_B^{-1} [X(f)] \right|^2 = w'_0 [u(t+T) - u(t-T)] \quad (26)$$

If Equation (26) is satisfied by $x(t)$, the frequency samples spaced $f_0 = \frac{1}{T}$ apart are uncorrelated.

Let us now consider the case where the average power for $X(f)$ per unit time interval is constant but where the amplitudes of the frequency spectra of both signal and noise are non-flat. Since successive frequency samples

are independent in this case, but the process is not stationary in frequency, the total entropy will be the sum of the entropies contributed by infinitesimal frequency bands over the bandwidth B . If a non-flat Gaussian signal $S(f)$ is added to independent non-flat Gaussian noise $N(f)$, the useful entropy at the output of the adder at a particular frequency $[X(f_i) = S(f_i) + N(f_i)]$ will be

$$H_{ui} = H(X_i) - H(X_i | S_i) = \frac{1}{2} \ln \left(\frac{\sigma_{X_i}^2}{\sigma_{X_i | S_i}^2} \right) \quad (27)$$

where

$$\sigma_{X_i | S_i}^2 = \sigma_{N_i}^2 = G_n(f_i) \frac{T}{2} \quad (28)$$

$$\sigma_{X_i}^2 = G_x(f_i) \frac{T}{2} = [G_s(f_i) + G_n(f_i)] \frac{T}{2}$$

where the capital letters refer to frequency samples and the lower case letters to time samples. $G_x(f)$, $G_n(f)$, and $G_y(f)$ are the frequency spectral densities for the time functions $x(t) = \text{input}$, $n(t) = \text{noise}$, and $y(t) = \text{output}$, respectively. Equation (27) becomes

$$H_u(X_i) = \frac{1}{2} \ln \left(1 + \frac{G_s(f_i)}{G_n(f_i)} \right) \quad (29)$$

To get the total useful entropy, the sum must be taken over the total number of frequency samples, which is $2BT$. Thus,

$$H_u \text{ total} = H_{ut} = \sum_{i=-BT}^{BT} \frac{1}{2} \ln \left(1 + \frac{G_x(f_i)}{G_n(f_i)} \right) \quad (30)$$

Now let T approach infinity and Δf approach 0 in such a way that

$$T\Delta f = 1 \quad (31)$$

Now that we have an infinite number of samples of infinitesimal separation, the sum will approach an integral as follows:

$$\begin{aligned} H_{ut} &= \frac{T}{2} \sum_{i=-BT}^{BT} \ln \left(1 + \frac{G_s(f_i)}{G_n(f_i)} \right) \Delta f \\ &= \frac{T}{2} \int_{-B}^B \ln \left(1 + \frac{G_s(f)}{G_n(f)} \right) df \end{aligned} \quad (32)$$

The average entropy per frequency sample is then

$$H_u = \frac{H_{ut}}{2TB} = \frac{1}{4B} \int_{-B}^B \ln \left(1 + \frac{G_s(f)}{G_n(f)} \right) df \quad (33)$$

Observe that if the signal and noise spectral densities are constants given by

$$G_s(f) = S_o$$

$$G_n(f) = N_o \quad (34)$$

then equations (32) and (33) reduce to, respectively,

$$H_{ut} = \frac{T}{2} \int_{-B}^B \ln \left(1 + \frac{S_o}{N_o} \right) df = BT \ln \left(1 + \frac{S_o}{N_o} \right) \quad (35)$$

$$H_u = \frac{1}{2} \ln \left(1 + \frac{S_o}{N_o} \right) \quad (36)$$

Equations (35) and (36) are the results which would have been obtained had the spectra been flat to begin with. Now consider the situation depicted in Figure 3

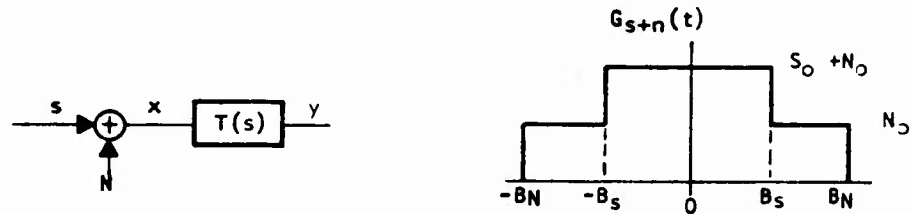


Figure 3. Signal and Noise Spectrum for Simple System

where $T(s)$ is a linear filter. Let the signal, S , be Gaussian with mean 0 and spectrum S_0 over a bandwidth from $-B_S$ to B_S and let the noise, N , be Gaussian with spectral density N_0 over the bandwidth $-B_N$ to B_N , where

$$|B_S| \leq |B_N| \quad (37)$$

The total entropy at x will be

$$\begin{aligned} H_{ut}(x) &= H_t(S+N) - H_t(N) \\ &= B_S T \ln [2\pi e(\sigma_S^2 + \sigma_N^2)] + (B_N - B_S) T \ln [2\pi e \sigma_N^2] \\ &\quad - B_N T \ln [2\pi e \sigma_N^2] = B_S T \ln \left(1 + \frac{\sigma_S^2}{\sigma_N^2} \right) \end{aligned} \quad (38)$$

where

σ_S^2 = variance of a signal frequency sample

σ_N^2 = variance of a noise frequency sample

Notice that the useful entropy at x is exactly the same as if $B_n = B_s$. Thus, so long as Equation (37) is satisfied, the useful information contributed by the noise alone is 0. This is in keeping with the concept that the noise contributes no useful information. Now linear filtering is a linear transformation on the frequency samples, and the useful entropy is invariant under linear transformation*. According to Equation (38), the total entropy is independent of the noise bandwidth so long as it is greater than or equal to the signal bandwidth. Thus, if maximization of entropy is to be used as the optimization criterion to determine the bandwidth of the transfer function $T(s)$ (assumed to be a square filter), there is no constraint on its bandwidth except that it be greater than or equal to the spectral width of S . This conclusion is a direct contradiction to the conclusion that would be drawn from the mean square error criterion, which says that the transfer function spectral width should be exactly equal to B_s so as to maximize the signal-to-noise ratio at y . This is a very good illustration of the difference between control and information theoretic concepts. The control argument says make the average signal-to-noise ratio per frequency sample as great as possible and maximize the overall power signal-to-noise ratio on the output.

* F.M. Reza, An Introduction to Information Theory, McGraw-Hill, 1961, p. 277.

The information theoretic approach says that it is unnecessary to do this -- that all that is required is that the total information be passed. Thus, although the average signal-to-noise ratio per frequency sample is smaller the larger B_N gets relative to B_S , the larger the number of samples, so that the total average information remains the same. The fact is that the information is in a more usable form if $B_S = B_N$. But the information is indeed available if condition 37 is satisfied. It also specifies what that maximum available useful information is. It will be shown in later reports that the condition of noise at the output of the filter does cause an interaction between the transfer function $T(s)$ and the output, or receiver, noise. Under these circumstances, the useful entropy is indeed a function of the transfer function $T(s)$.

Information theoretic concepts may be applied to the optimization of feedback and servo system design. A general servo system block diagram appears as

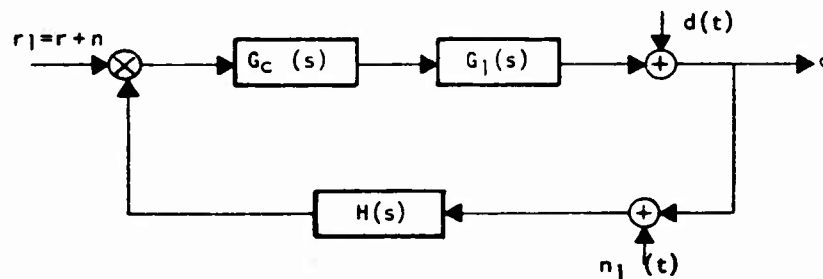


Figure 4. Typical Servo Block Diagram

where $d(t)$ is load disturbance, $n_1(t)$ is measurement noise or internal amplifier noise, n is noise at the input to the servo and r is the desired input signal. $r_1 = r + n$ is the total input signal. $c(t)$ is the servo output. Internal sources of noise such as $d(t)$ and $n_1(t)$ affect the output information rate at c in such a way that the servo transfer function is related to $d(t)$ and $n_1(t)$. The output entropy or rate is dependent on the servo transfer function in this case, and the system may be optimized by maximizing the output rate as a function of the variable parameters in the servo transfer function. The open loop transfer functions are $G_c(s)$, the series compensator, $G_1(s)$, the plant, and $H(s)$, the shunt compensator. In the case where $d(t)$ and $n_1(t)$ are 0, however, the output entropy of c is independent of the servo function and is simply equal to the input entropy at r_1 . If the servo is to be optimized with signal and noise only on the input and no constraints on the system, the optimization by the mean square error criterion is the Wiener problem. Let us replace the servo in Figure 4 by the simpler structure in Figure 5.

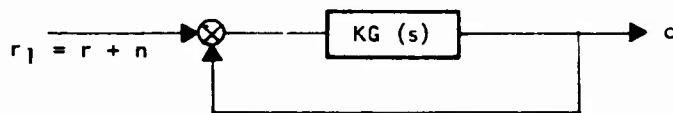


Figure 5. Simplified Servo Block Diagram

The error is defined as

$$\epsilon = r - c = r - \frac{KG r_1}{1 + KG} = r - F r_1 \quad (39)$$

where

$$F = \frac{KG}{1+KG} \quad (40)$$

$$r_1 = r + n \quad (41)$$

The spectral density for the complex variable s is defined as

$$\phi(s) = \lim_{T \rightarrow \infty} \frac{2}{T} \overline{F(s)F'(s)} \quad (42)$$

where $F'(s) = F(-s)$ and where the bar or overscore denotes an ensemble average. The corresponding spectral density in the real frequency variable ω is defined as

$$G(f) = \lim_{T \rightarrow \infty} \frac{2}{T} \overline{|F(j\omega)|^2} \quad \text{where } -\frac{T}{2} \leq t \leq \frac{T}{2} \quad (43)$$

From Equation (39), the spectral density for ϵ is

$$\phi_{\epsilon\epsilon} = \phi_{rr} + FF'\phi_{r_1r_1} - \phi_{r_1r}F' - \phi_{rr_1}F \quad (44)$$

The optimum closed loop transfer function can be found by satisfying the following equation

$$\frac{\partial \phi_{\epsilon\epsilon}}{\partial F'} = X_1 = F\phi_{r_1r_1} - \phi_{r_1r} \quad (45)$$

where X_1 is analytic in the left half plane and on the imaginary axis.

Let

$$ZZ' = \phi_{r_1r_1} = \phi_{rr} + \phi_{nn} + \phi_{rn} + \phi_{nr} \quad (46)$$

and

$$\phi_{r_1 r} = \phi_{rr} + \phi_{nr} \quad (47)$$

then

$$FZZ' - \phi_{r_1 r} = X_1 \quad (48)$$

F by definition must have poles and zeroes only in the left half plane or on the imaginary axis since it is to be a stable closed loop transfer function. Z by definition has left half plane poles and zeroes only, so that Z' has right half plane poles and zeroes only. $\phi_{r_1 r}$ may have both left and right half plane poles and zeroes, but may be expanded into a partial fraction expansion such that

$$\phi_{r_1 r} = (\phi_{r_1 r})_{++} + (\phi_{r_1 r})_{--} \quad (49)$$

where

$(\phi_{r_1 r})_{++}$ has poles in the LHP only

and

$(\phi_{r_1 r})_{--}$ has poles in the RHP only

Rewriting Equation (48) so that terms having left half plane poles only are on the left half side of the equation and terms having right half plane poles only are on the right half side of the equation, we have

$$FZ - \left(\frac{\phi_{r_1 r}}{Z'}\right)_{++} = X_1 + \left(\frac{\phi_{r_1 r}}{Z'}\right)_{--} \quad (50)$$

The only way that both sides of Equation (50) can be equal to each other is if they are equal to a constant and it turns out that constant must be equal to zero. It therefore follows that the solution for F is

$$F = \frac{1}{Z} \left(\frac{\phi_{r_1 r}}{Z'} \right)_{++} \quad (51)$$

Equation (51) yields the optimum closed loop transfer function for minimizing the mean square error, and when this function in Equation (51) is substituted back into Equation (44), the least mean square error results. The condition for physical realizability of the transfer function has been inherently included in this development.

Unfortunately, the foregoing procedure does not seem to be feasible for the technique of maximization of information rate with respect to the variable parameters of the system. For one thing, the variable parameters of the system are unknown in the Wiener problem and are specified only after the optimization has taken place. In the second place, the information rate criterion involves the bandwidth of the system, in this case, the closed loop bandwidth of the servo, which is specified only if the form of the transfer function is known, but is unspecified in the Wiener problem. It therefore follows that the optimization must take place after the closed loop transfer function and the open loop transfer function have been specified as to form. This procedure has the advantage that the complexity of the system can be specified by specifying the order of the differential equation describing the system before the optimization takes place. From Equation (39), the mean square error can then be written as

$$e = \frac{KG}{1+KG} (r + n) = \frac{r}{1+KG} - \frac{KGn}{1+KG} = Ar - Fn \quad (52)$$

$$\therefore G_e(f) = |A(j\omega)|^2 G_{rr}(f) + |F(j\omega)|^2 G_{nn}(f) \quad (53)$$

since the signal and noise are independent. The mean square error is then

$$e^2 = \int_0^{\infty} G_e(f) df \quad (54)$$

and this mean square error is then minimized with respect to the variable parameters of the system by taking partial derivatives of the mean square error with respect to those parameters and setting them equal to zero. This procedure yields the optimum set of parameters in the mean square sense.

For the Wiener problem, the total useful entropy and useful rate on the output of the servo are the same as that on the input, since the closed loop servo is simply the equivalent of a linear transfer function. Therefore, this procedure does not yield any way of optimizing the closed loop transfer function in an information theoretic sense. It will now be proposed to define the minimization of the useful entropy of the error as the optimization criterion rather than the maximization of the useful entropy of the output in order to get an information theoretic criterion which is closer in both form and concept to the mean square error criterion. The minimization of the information content or rate of the error criterion is, to the best of our knowledge, completely new. The following argument shows that it can be justified on fundamental grounds. The proposed criterion is

similar to the mean square error criterion and leads to similar results when the probability distribution is Gaussian. The criterion, which is the minimization of the useful entropy

$$H_u(\epsilon) = H(\epsilon) - H(\epsilon|r) = H(r) - H(r|\epsilon) \quad (55)$$

is particularly useful when used with nonlinear systems, since a non-linearity results in a non-Gaussian distribution even when the input is a Gaussian distribution. A probabilistic criterion, such as that in Equation (55), takes into account differences between the signal and the noise in moments higher than the second, and thereby produces improved discrimination over that which would be provided by the mean square error criterion alone. The error is defined as

$$\epsilon = r - c \quad (56)$$

Ideally the error should go to zero so that $r = c$. The correlation between the error and the signal input (all means assumed zero) is then

$$\overline{\epsilon r} = \overline{r^2} - \overline{rc} \quad (57)$$

In the ideal situation where $\epsilon = 0$,

$$\overline{\epsilon r} = \overline{r^2} - \overline{r^2} = 0 \quad (58)$$

Thus if ϵ and r are uncorrelated and the distribution is such that they are independent,

$$H(\epsilon|r) = H(\epsilon) \quad (59)$$

and

$$H_u = H(\epsilon) - H(\epsilon|r) = H(\epsilon) - H(\epsilon) = 0 \quad (60)$$

Equation (60) leads one to the logical conclusion that the system will be optimized if the useful entropy of the error is minimized. It is also interesting to observe that the servo error must be simply the noise on the input in the ideal situation when ϵ in Equation (56) is zero. The "servo error" is defined as

$$e = r_1 - c = r + n - c \quad (61)$$

Ideally, $\epsilon = r - c = 0$ and

$$e = n \quad (62)$$

The basic conclusions in Equations (56) through (62) do not take into account the physical realizability constraints. When this constraint is taken into account, the mean square error criterion yields a Wiener filter. When a servo is optimized in the Wiener sense, the servo error is white noise^{*}. This conclusion is illustrated in the following example (refer to Figure 5): Let $\phi_{rr} = \frac{A^2}{\omega^2 + a^2}$ and $\phi_{nn} = 1$

* R. E. Kalman, A New Approach to Linear Filtering and Prediction Problems, Trans. of the ASME, March 1960, page 42, item m.

Then

$$\phi_{r_1 r_1} = ZZ' = \frac{\omega^2 + a^2 + A^2}{\omega^2 + a^2} = \frac{\omega^2 + b^2}{\omega^2 + a^2} \text{ where } b^2 = a^2 + A^2$$

$$Z = \frac{s+b}{s+a}$$

$$F = \frac{1}{Z} \left(\frac{\phi_{rr}}{Z'} \right)_{++} = \frac{s+a}{s+b} \left(\frac{A^2(-s+a)}{(-s+a)(s+b)(-s+b)} \right)_{++} = \frac{A^2}{(a+b)(s+b)} = 1 - \frac{1}{1+Y}$$

$$\text{But } e = \frac{r_1}{1+Y} = r_1(1-F) \text{ and } \phi_{ee} = \phi_{r_1 r_1} (1-F)(1-F') = \phi_{r_1 r_1} (1-F-F'+FF')$$

$$\text{where } 1-F = \frac{(a+b)(s+b) - A^2}{(a+b)(s+b)} = \frac{s+b - \frac{A^2}{a+b}}{s+b}$$

$$\phi_{ee} = \phi_{r_1 r_1} |1-F|^2 = \frac{\omega^2 + b^2}{\omega^2 + a^2} \frac{\omega^2 + (b - \frac{A^2}{a+b})^2}{\omega^2 + b^2} = \frac{\omega^2 + (b - \frac{A^2}{a+b})^2}{\omega^2 + a^2}$$

$$\text{But } b - \frac{A^2}{a+b} = \frac{ba + b^2 - A^2}{a+b} = \frac{ba + a^2}{a+b} = a$$

$$\therefore \phi_{ee} = 1 = \phi_{nn}$$

or the servo error spectrum is white noise, and in this case, has the same spectrum as the input noise.

Note that Equation (59) implies that

$$e \perp\!\!\!\perp r \text{ or } r-c \perp\!\!\!\perp r \text{ (the symbol } \perp\!\!\!\perp \text{ means "is independent of")}$$

$$\therefore \overline{er_1} = \overline{[(r-c)+n] [r+n]} = \overline{n(r+n)} = \overline{n r} + \overline{n^2} = \overline{n^2} \quad (63)$$

Thus, the correlation function relating the input signal plus noise and the servo error at the same instant of time is found to be simply the mean noise

power. This agrees with the conclusion in Equation (62). Equation (63) suggests that the criterion in Equation (55) will lead to a result similar to that found for Wiener optimization when the physical realizability constraint is included. A necessary and sufficient condition that the error be minimum in the Wiener sense is that the signal and the error be orthogonal to (i.e., independent of) each other*. But, as has been pointed out, this is exactly the condition asked for in the proposed criterion, that is, that

$$H(r|\epsilon) = H(\epsilon|r) = 0 \quad (64)$$

Returning to the problem of minimizing the information content in the error defined in Equation (52), let $G_\epsilon(f)$ be the spectral density of the error and $G_{\epsilon|r}(f)$ be the spectral density of the error conditioned on r . Then from Equation (32), the total useful entropy is

$$H_{ut} = \frac{T}{2} \int_{-B}^B \ln \left(\frac{G_\epsilon(f)}{G_{\epsilon|r}(f)} \right) df \quad (65)$$

providing the input signal and noise are Gaussian and independent and the system is linear. Since r is independent of n , the conditional mean of ϵ given r is

$$\bar{\epsilon}|r = Ar \quad (66)$$

$$\therefore \epsilon - \bar{\epsilon}|r = -Fn \quad (67)$$

and

$$G_{\epsilon|r}(f) = |F(j\omega)|^2 G_{nn}(f) \quad (68)$$

* *ibid*

where the mean of the noise is assumed to be zero. The total entropy in Equation (57), where $G_{\epsilon}(s)$ is defined in Equation (53) and $G_{\epsilon|r}(f)$ is defined in Equation (68), is then maximized with respect to the variable parameters in the transfer functions by taking the appropriate partial derivatives.

It is illustrative to illustrate the mean square error approach and the maximization of the useful entropy of error approach with a simple example. Let $G(s)$ be 1 in Figure 5 so that we are trying to optimize a simple positioning servo. K is then the parameter to be varied to yield the optimum. First consider the mean square error approach.

$$F = \frac{K}{1+K} \quad A = \frac{1}{1+K} \quad (69)$$

$$\overline{\epsilon^2} = A^2 \overline{r^2} + F^2 \overline{n^2} = \frac{\overline{r^2} + K^2 \overline{n^2}}{(1+K)^2} \quad (70)$$

$$\frac{\partial \overline{\epsilon^2}}{\partial K} = \frac{(1+K)^2 \cdot 2Kn^2 - (\overline{r^2} + K^2 \overline{n^2}) 2(1+K)}{(1+K)^4} = 0$$

or

$$(1+K) Kn^2 = \overline{r^2} + K^2 \overline{n^2}$$

$$\therefore K = \frac{\overline{r^2}}{\overline{n^2}} \quad (71)$$

According to Equation (71), the mean square error is minimized when the servo gain is equal to the power signal to noise ratio. Let us now proceed to do this same problem by the minimization of total information or information rate of the error criterion.

$$\sigma_e^2 = \bar{\epsilon}^2 = A^2 \bar{r}^2 + F^2 \bar{n}^2 \quad (72)$$

$$\sigma_{e|r}^2 = \overline{(\epsilon - \epsilon|r)^2} = F^2 \bar{n}^2 \quad (73)$$

$$\therefore H_{ut} = BT \ln \left(1 + \frac{\sigma_e^2}{\sigma_{e|r}^2} \right) = BT \ln \left(1 + \frac{A^2 \bar{r}^2}{F^2 \bar{n}^2} \right) \quad (74)$$

Let

$$Y^2 = \frac{\bar{r}^2}{\bar{n}^2} \quad (75)$$

The information rate is then

$$R = \frac{H_{ut}}{T} = B \ln \left(1 + \frac{Y^2}{K^2} \right) \quad (76)$$

Assume that Equation (69) holds only over the servo bandwidth and that the closed loop servo bandwidth increases linearly with gain K . This is very nearly the case in practice. The rate in Equation (76) then becomes

$$R = f_0 K \ln \left(1 + \frac{Y^2}{K^2} \right) \text{ where } B = f_0 K \quad (77)$$

$$\frac{\partial R}{\partial K} = \frac{f_0 K}{1 + \frac{Y^2}{K^2}} \left(-\frac{2Y^2}{K^3} \right) + f_0 \ln \left(1 + \frac{Y^2}{K^2} \right) = 0$$

or

$$\ln \left(1 + \frac{Y^2}{K^2} \right) = \frac{2Y^2}{K^2 + Y^2}$$

$$\text{For } K \text{ large, } \ln \left(1 + \frac{Y^2}{K^2} \right) \approx \frac{Y^2}{K^2} = \frac{2Y^2}{K^2 + Y^2}$$

or

$$K = Y = \sqrt{\frac{\bar{r}^2}{\bar{n}^2}} \quad (78)$$

Notice the close similarity between solutions in Equations (71) and (78). This example illustrates the similarity between the two approaches and suggests that the two approaches can be expected to yield similar results. In the forthcoming period, research will continue along lines of investigation into various optimization criteria. In particular, the approaches investigated during the past period will be considered further. Both the criteria of maximization of information rate at the output and the minimization of information rate at the error will be considered for more complicated systems, particularly those involving noise entering at points other than at the input. So far only the unconstrained or Wiener case has been considered. Extension of the information rate maximization criterion to constrained cases will also be considered. Ultimately it is hoped to develop an extensive optimization of information rate theory analagous to that presently existing for the mean square error criterion.

PART II.

ESTIMATION AND OPTIMIZATION OF NONLINEAR DYNAMIC SYSTEM PARAMETERS BY REGRESSION ANALYSIS METHODS

1. INTRODUCTION

Modern theories of control are generally based upon an assumption that an accurate quantitative model for the object being controlled is available. There are many practical situations in which this is not the case. In such circumstances, it becomes necessary to devise a computational procedure which will permit the inference of a mathematical description for the controlled object from records of input and output. This problem has been attacked with considerable vigor and substantial success by a large number of investigators for the important but restrictive class of systems possessing the properties of linearity and time invariance [1, 2, 3, 4]¹.

In comparison to the linear problem, little progress has been made toward a practical solution of the more general problem of "identification" of nonlinear dynamic systems². While a very broad theory for nonlinear systems has been advanced by N. Wiener, certain rather serious computational difficulties inherent in his formulation have retarded the application of this theory to real physical systems [6, 7, 8]. The difficulties encountered in attempting to make use of the Wiener theory seem to arise from the very generality of his approach. Wiener assumes at the outset that a state of complete ignorance exists concerning the nature of the object to be identified. This is rarely the case when real physical devices are under consideration. In a great many situations, basic physical theories permit the construction

¹In this document, superscript numbers refer to footnotes, while bracketed numbers correspond to the list of references collected at the end of the text.

²The term, "identification", appears to have been coined by L. A. Zadeh to describe processes which seek to infer a mathematic description for a dynamic system from input and output data [5].

of a parametric model of finite dimensionality for the system under investigation and it is only the numerical values of parameters which remain unknown.

A simple example of a parametric model is provided by the equation for the motion of a damped pendulum. Straightforward summation of moments yields the equation

$$\ddot{\theta} = -Mgl \sin \theta - B\dot{\theta} \quad (1)$$

which can be normalized to

$$c_2 \ddot{\theta} + c_1 \dot{\theta} + \sin \theta = 0 \quad (2)$$

The determination of c_2 and c_1 along with two initial conditions completely identifies this system³. A more intriguing example for which a parametric model exists can be obtained by considering the equations describing the atmospheric re-entry of a ballistic vehicle. In this case, nonlinear trajectory prediction becomes possible if the important vehicle parameters can be obtained in a sufficiently short period of time. Other situations in which parametric models exist are not hard to imagine.

When an appropriate parametric model for an unknown system has been established, then, if the system is to be identified by a process of parameter estimation, it becomes necessary to state an explicit criterion for the adjustment of the model to match the data. A criterion which has been widely utilized in linear problems is the integral-squared error function

³ In the discussion to follow, a set of such coefficients and initial conditions will be called a "parameter vector" and will be denoted by the symbol, c .

evaluated over some interval of operation. Specifically, let $y(t; \vec{c})$ represent the theoretical or model response for a specified set of parameters, \vec{c} . Then if the true parameter vector is indicated by \vec{c}_0 , the observed response will generally consist of the theoretical response plus an error term; i.e., if $y_0(t)$ is the experimentally observed system output, then

$$y_0(t) = y(t; \vec{c}_0) + \epsilon(t) \quad (3)$$

where $\epsilon(t)$ represents an error function. This error may be due either to measurement error, noise internally generated in the system under study, imperfections in the assumed model, or possibly a combination of these and other effects. In any event, the integral squared error function is given by

$$\phi(\vec{c}; y_0) = \int_0^T [y_0(t) - y(t; \vec{c})]^2 dt \quad (4)$$

A "least squares" estimate of the true parameter vector, \vec{c}_0 , can be obtained by finding a vector, \vec{c}_e , such that

$$\min_{\vec{c}} \phi(\vec{c}; y_0) = \phi(\vec{c}_e; y_0) \quad (5)$$

Estimation of nonlinear system parameters according to this criterion forms the basis for the research to be reported in the following discussion. The inference of mathematical models from experimental data based on such estimation will be referred to as the "parameter space method".

Most of the remainder of this study is concerned with the development of an explicit computational technique for the achievement of the desired least squares estimation of nonlinear system parameters. The particular procedure

to be described is derived from the methods employed in classical statistics for the estimation of linear system parameters; i.e., from the subject of "linear regression analysis" [9]. However, before continuing the discussion of computational procedures, it is worthwhile to note that the problem of parameter estimation is very closely related to the problem of parameter optimization. If, for example, the desired response of a network and the form (possibly nonlinear) of the network are both specified in advance, then adjustment of the network parameters to obtain an approximation to the desired response which is "best" in a mean square sense is essentially the same problem as the estimation of unknown system parameters [10]. While the discussion to follow will use terminology consistent with parameter estimation, the results obtained translate very simply into parameter optimization problems.

2. LINEAR REGRESSION ANALYSIS

2.1 Least Squares Estimation and Normal Equations

In the preceding introduction, a parameter space characterization of dynamic systems was proposed. It was suggested that a very general method for estimating unknown parameter vectors could be based upon the minimization of an integral-squared error function defined over the parameter space. In classical statistics, a very similar problem is treated under the heading of "regression analysis". In regression analysis it is assumed that a finite number of measured values of a variable, say $y(t_1; \bar{c}_0)$, are available. It is further assumed that there are random errors included in the experimental values for y ; i.e., the variable actually measured is

$$y_{o_i} = y(t_i; \bar{c}_o) + \epsilon_i \quad i = 1, 2, 3 \dots N \quad (6)$$

The objective of regression analysis is then to determine a vector, \bar{c}_e , which estimates the true parameter vector, \bar{c}_o , by making use of the theoretical response function, $y(t; \bar{c})$. This estimation is conventionally accomplished by the minimization of sum squared error; i.e., if

$$G(\bar{c}; \bar{y}_o) = \sum_{i=1}^N [y_{o_i} - y(t_i; \bar{c})]^2 \quad (7)$$

then

$$\min_{\bar{c}} G(\bar{c}; \bar{y}_o) = G(\bar{c}_e; \bar{y}_o) \quad (8)$$

where \bar{y}_o is an N dimensional vector composed of the N time samples of $y_o(t)$.

The minimization of integral squared error discussed in the above introduction may be cast in the form of regression analysis by approximating the integral by a sum. That is,

$$\begin{aligned} \phi(\bar{c}; y_o) &= \int_0^T [y_o(t) - y(t)]^2 dt \\ &\approx \sum_{i=1}^N [y_o(t_i) - y(t_i)]^2 \Delta t \\ &= \frac{T}{N} \sum_{i=1}^N [y_{o_i} - y_i]^2 \end{aligned} \quad (9)$$

Since the N time samples of either y_o or y may be regarded as N dimensional column vectors, then, except for a scale factor, ϕ is just

an inner product. Because minimization of ϕ is unaffected by a constant scale factor, it will be assumed in all that follows that $T/N = 1$. This being the case, if an error column vector is defined by

$$\bar{e} = \bar{y}_0 - \bar{y} \quad (10)$$

then

$$\phi(\bar{c}; \bar{y}_0) \approx \bar{e}' \bar{e} \quad (11)$$

where the prime denotes a transpose⁴. In the remainder of this investigation, ϕ will be redefined so that equation (11) is exact; i.e., least squares estimation will be based upon the sum-squared error rather than the integral-squared error function.

When the theoretical response function, $y(t; \bar{c})$, depends linearly upon the parameter vector, \bar{c} , the minimization of sum-squared error is particularly simple. In such circumstances, the N dimensional vector of samples of y can be written

$$\bar{y}_0 = A \bar{c}_0 + \bar{\epsilon} \quad (12)$$

where A is an $N \times M$ matrix of coefficients and $\bar{\epsilon}$ is a random error vector⁵. The function ϕ may be written in this case as

$$\begin{aligned} \phi(\bar{c}; \bar{y}_0) &= (\bar{y}_0 - A\bar{c})' (\bar{y}_0 - A\bar{c}) \\ &= \bar{e}' \bar{e} \end{aligned} \quad (13)$$

⁴ Note that there are two sources of error which contribute to ϕ . There is, first of all, the difference between \bar{y} and \bar{y}_0 which results from the fact $\bar{c} \neq \bar{c}_0$. Secondly, the random error, $\bar{\epsilon}$, causes ϕ to be greater than zero even when $\bar{c} = \bar{c}_0$.

⁵ The elements of A are usually known or measured values of independent variables or functions of independent variables. A constant term may be included in this equation by choosing one column of A such that every element is equal to one.

which is a positive definite quadratic form in the coordinates of the trial parameter vector, \vec{c} . The unique minimum value of ϕ may be found, therefore, by equating to zero all of the partial derivatives of ϕ with respect to the components of \vec{c} ; i.e.

$$\begin{aligned}\phi &= \vec{y}_0' \vec{y}_0 - \vec{y}_0' A \vec{c} - (A \vec{c})' \vec{y}_0 + (A \vec{c})' (A \vec{c}) \\ &= \vec{y}_0' \vec{y}_0 - 2 \vec{c}' A' \vec{y}_0 + \vec{c}' A' A \vec{c}\end{aligned}\quad (14)$$

so

$$\frac{\partial \phi}{\partial \vec{c}} = \vec{\nabla} \phi = -2A' \vec{y}_0 + 2A' A \vec{c}\quad (15)$$

and

$$\left. \frac{\partial \phi}{\partial \vec{c}} \right|_{\vec{c} = \vec{c}_e} = -2A' \vec{y}_0 + 2A' A \vec{c}_e = \vec{0}\quad (16)$$

This equation reduces to the standard "normal equation" for least squares estimation [9]:

$$A' A \vec{c}_e = A' \vec{y}_0\quad (17)$$

Assuming A is of full rank, this relationship may be inverted to give

$$\vec{c}_e = [A' A]^{-1} A' \vec{y}_0 = S^{-1} A' \vec{y}_0\quad (18)$$

The parameter vector provided by equation (18) is thus the "least squares estimate" of the true parameter vector, \vec{c}_0 .⁶

⁶ If the number of time samples of $y(t)$ is exactly equal to the number of unknown parameters, then A is a square matrix. In this case, $[A' A]^{-1} A' = A^{-1}$ so $\vec{c}_e = A^{-1} \vec{y}_0$; i.e. least squares analysis reduces to simple simultaneous solution of linear equations.

There are several reasons why equation (18) provides a very desirable estimate of \vec{c}_0 . To begin with, since the matrix A is an array of known coefficients (possibly values of known time functions), \vec{c}_0 is computed by means of a simple linear transformation applied to the data vector, \vec{y}_0 . The linearity of this relationship results solely from the fact that the criterion function, ϕ , is quadratic in the parameter vector, \vec{c} . If any other criterion were chosen, differentiation would lead to a nonlinear relationship between \vec{c}_0 and the data vector. So far as the computational aspects of parameter estimation are concerned, this is the strongest reason for choosing a squared error criterion.

From a statistical point of view, equation (18) is a optimum estimator in two senses. First of all, providing only that the random measurement error, $\vec{\epsilon}$, has the properties

$$\begin{aligned} E(\vec{\epsilon}) &= \vec{0} \\ E(\vec{\epsilon} \vec{\epsilon}') &= \sigma^2 I \end{aligned} \quad (19)$$

where I is a unit matrix, it follows from the Gauss-Markoff theorem that equation (18) provides the minimum variance unbiased linear estimator of the true parameter vector, \vec{c} [9]. Secondly, if in addition to possessing these properties, $\vec{\epsilon}$ is also a Gaussian random variable, then the value for \vec{c} specified by equation (18) is, moreover, a maximum likelihood estimate of \vec{c}_0 .

2.2 An Example of Parameter Estimation by Linear Regression Analysis

Suppose that an oscillating system is known to be governed by the equation

$$\ddot{y} + \omega^2 y = 0 \quad (20)$$

The parameter vector for this equation is

$$\vec{c} = \begin{pmatrix} \omega \\ y(0) \\ \dot{y}(0) \end{pmatrix} \quad (21)$$

The solution to equation (20) is given by

$$y(t; \vec{c}) = y(0) \cos \omega t + \frac{\dot{y}(0)}{\omega} \sin \omega t \quad (22)$$

While this equation is nonlinear in ω , it is linear in the unknown initial conditions. If additional information exists so that it is known that $\omega = \omega_0$, then the coefficient matrix, A , appearing in equation (12) is given by

$$A = \begin{pmatrix} \cos \omega_0 t_1 & \frac{1}{\omega_0} \sin \omega_0 t_1 \\ \cos \omega_0 t_2 & \frac{1}{\omega_0} \sin \omega_0 t_2 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cos \omega_0 t_n & \frac{1}{\omega_0} \sin \omega_0 t_n \end{pmatrix} \quad (23)$$

and the parameter vector \vec{c} is reduced to

$$\vec{c} = \begin{pmatrix} y(0) \\ \dot{y}(0) \end{pmatrix} \quad (24)$$

Since the sine and cosine functions are linearly independent, A is of full rank (unless all rows of A are identical) and equation (18) may be used to compute estimates of \vec{c}_0 from sample values for y .

2.3 Orthogonalization in Linear Regression Analysis

When the columns of A are chosen to be orthogonal to each other, then the matrix

$$S = A' A \quad (25)$$

is a diagonal matrix.

$$S = \begin{pmatrix} S_{11} & 0 & 0 & \dots & 0 \\ 0 & S_{22} & 0 & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \dots & S_{NN} \end{pmatrix} \quad (26)$$

In such circumstances, equation (17) consists of an uncoupled set of linear equations. If the columns of A are further adjusted by a change of scale so that they are orthonormal, then the normal equation is itself a solution for \vec{c}_e since S becomes a unit matrix and equation (17) therefore reduces to

$$S\vec{c}_e = I\vec{c}_e = \vec{c}_e = A'\vec{y}_0 \quad (27)$$

Inversion of S is thereby avoided. This may be of considerable practical value when the dimensionality of \vec{c} (and consequently, of S) is large. If A is thought of as a row vector with orthonormal column vector elements

$$A = (\vec{A}_1, \vec{A}_2, \dots, \vec{A}_M) \quad (28)$$

then, from equation (27) the k^{th} component of the vector \vec{c}_0 is given

$$c_{0k} = \vec{A}_k' \vec{y}_0 = (\vec{A}_k, \vec{y}_0) \quad (29)$$

That is, the components of \vec{c}_0 are just scalar products of the columns of A with the observed output variable.

The example matrix specified by equation (23) is easily orthogonalized by a proper choice of sampling times, t_1 . Unfortunately, most coefficient matrices are not orthogonalized so simply. However, when modern digital computers are used to solve the normal equations, the time required to invert the matrix S is not likely to be significant until the dimensionality of \vec{c} becomes quite large⁷.

3. NONLINEAR REGRESSION ANALYSIS

3.1 Nonlinear Analysis and the Need for Iterative Methods

While linear regression analysis produces estimators with many desirable characteristics, the assumption of linearity with respect to all parameters is

⁷ Exactly what constitutes a "large" matrix will depend upon the particular computer used. However, matrices up to tenth order will certainly be handled with ease by almost any computer.

seldom supportable in practical circumstances. Even in extremely simple systems such as the one exemplified by equation (20), it will usually be found that the dependence of solutions on some parameters will be very nonlinear. Thus, for example, if it is assumed in equation (20) that ω is also to be estimated from data, equation (22) shows that the connection between ω and y can scarcely be approximated by a linear relationship (except for very small variations in ω). This illustration is typical rather than exceptional. Consequently, it is necessary to consider minimization of error functions which are not quadratic in the parameter space.

The location of minima for arbitrary surfaces by automatic computational procedures is a subject which has received considerable attention in recent years. The mathematical treatment of such techniques is usually labeled "mathematical programming" [11]. In the language of mathematical programming, the sum-squared error function of regression analysis is called a "criterion" or "objective" function [11,12].

In problems treated by mathematical programming, it is usually required that the minimization of criterion functions be carried out in some bounded subspace of the parameter space which is specified by a set of constraint equations. When both the criterion function and the constraint equations are linear, the minimum is unique and may be located in a straightforward way by application of the techniques of "linear programming". Related methods have been developed for quadratic criterion functions [11].

For general nonlinear problems, only iterative methods exist. Furthermore, there may well be more than one minimum associated with a nonlinear criterion function so that it becomes necessary to examine the local minima to find the lowest or absolute minimum. The research reported in this document deals entirely with the problem of determination of local minima for nonlinear criterion functions in unconstrained parameter spaces. Subsequent reports will deal with the solution of computational problems arising from the consideration of multiple minima and constraint equations.

3.2 Parameter Estimation by Iterative Linear Regression Analysis

It has been suggested, occasionally, that nonlinear parameter estimation problems may be solved by iterative application of linear least squares regression analysis [10,13,14]. This approach differs from the more general techniques of nonlinear programming in the respect that it is restricted to situations where the criterion function is a sum-squared error function computed from data and the response of an assumed parametric model. When such a model exists, then assuming that the model response, $y(t;\vec{c})$, possesses first partial derivatives with respect to each component of the parameter vector, \vec{c} , these derivatives may be used to fit an approximating quadratic surface to $\phi(\vec{c};\vec{y}_0)$. Let X be the $N \times M$ matrix of partial derivatives given by ⁸

⁸ The elements of X have sometimes been referred to as "parameter influence coefficients" [15]. A computational method for obtaining these coefficients will be presented later in this discussion.

$$X = \begin{pmatrix} \frac{\partial y(t_1)}{\partial c_1} & \frac{\partial y(t_1)}{\partial c_2} & \dots & \frac{\partial y(t_1)}{\partial c_M} \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \frac{\partial y(t_N)}{\partial c_1} & \dots & \dots & \frac{\partial y(t_N)}{\partial c_M} \end{pmatrix} \quad \bar{c} = \bar{c}_1 \quad (30)$$

Then an approximation to the response vector $\bar{y}(\bar{c}_1 + \Delta\bar{c})$ can be constructed by truncating the Taylor series for \bar{y} to obtain

$$\hat{\bar{y}}(\bar{c}_1 + \Delta\bar{c}) = \bar{y}(\bar{c}_1) + X\Delta\bar{c} \quad (31)$$

Associated with this approximation to \bar{y} is an approximate criterion function:

$$\hat{\beta}(\Delta\bar{c}; \bar{c}_1, \bar{y}_0) = (\bar{y}_0 - \hat{\bar{y}})'(\bar{y}_0 - \hat{\bar{y}}) \quad (32)$$

Utilizing equation (10), this may be written as

$$\begin{aligned} \hat{\beta}(\Delta\bar{c}) &= (\bar{e} - X\Delta\bar{c})'(\bar{e} - X\Delta\bar{c}) \\ &= \bar{e}'\bar{e} - \bar{e}'X\Delta\bar{c} - \Delta\bar{c}'X'\bar{e} + \Delta\bar{c}'X'X\Delta\bar{c} \\ &= \hat{\beta}(\bar{c}_1; \bar{y}_0) - 2\Delta\bar{c}'X'\bar{e} + \Delta\bar{c}'S\Delta\bar{c} \end{aligned} \quad (33)$$

This expression is a quadratic form in $\bar{\Delta c}$ analogous to equation (14).

The associated normal equation for the minimizing value of $\bar{\Delta c}$ is obtained by differentiating with respect to $\bar{\Delta c}$ and equating the derivative to zero:

$$\left. \frac{\partial \mathcal{J}}{\partial \bar{\Delta c}} \right|_{\bar{\Delta c} = \bar{\Delta c}_1} = -2X'e + 2S\bar{\Delta c}_1 = \bar{0} \quad (34)$$

The minimizing value of $\bar{\Delta c}$ is then given by

$$\bar{\Delta c}_1 = S^{-1}X'e \quad (35)$$

This result may be written in a more meaningful form by noting that

$$\left. \mathcal{J}(c_1) = \frac{\partial \mathcal{J}}{\partial \bar{c}} \right|_{\bar{c} = c_1} = \frac{\partial}{\partial \bar{c}} e'e = 2\left(\frac{\partial e}{\partial \bar{c}}\right)'e = -2X'e \quad (36)$$

so that

$$\bar{\Delta c}_1 = -\frac{1}{2} S^{-1} \mathcal{J}'(\bar{c}_1) = \bar{\beta}_1 \quad (37)$$

The value for $\bar{\Delta c}$ provided by equation (37) can be used to obtain a new value for \bar{c}

$$\bar{c}_2 = \bar{c}_1 + \bar{\Delta c}_1 \quad (38)$$

which may then be used as a basis for further iteration. This procedure has been called "Gauss - Newton" iteration [14]. As in linear regression analysis, the normal equation for iteration (equation (34)) is linear

(and therefore solvable by matrix inversion) only because the criterion function ϕ is quadratic in $\bar{\Delta c}$.

Since the validity of equation (37) rests upon an assumption of linearity, there is no reason, a priori, to suppose that iterative application of this result to the estimation of the parameters of a nonlinear system will produce a convergent sequence of estimating vectors. Indeed, while this method has been used with some success in certain problems [10,13], it may well fail to converge in other circumstances. Necessary and sufficient conditions for convergence and other aspects of stability are discussed later in this report.

3.3 Determination of Solution Partial Derivatives

In order to make use of the techniques of nonlinear regression analysis, it is necessary to compute partial derivatives of the solution to the assumed system differential equation with respect to each unknown parameter. These derivatives can be obtained approximately by perturbing the parameters one at a time and approximating derivatives by finite differences; i.e.

$$x_{1j} = \frac{\partial y_1}{\partial c_j} \approx \frac{y(t_1; c_1, c_2, \dots, c_j + \Delta c_j, \dots, c_M) - y(t_1; c_1, c_2, \dots, c_j, \dots, c_M)}{\Delta c_j} \quad (39)$$

Precise determination of derivatives in this manner is limited by round-off error in digital computation and by noise in analog computation. These difficulties may be avoided by differentiating the assumed differential equation with respect to the parameters to obtain the "parameter influence

equations" [15]. The solution to these differential equations then yields the desired partial derivatives as functions of time.

The system governed by equation (20) provides a convenient basis for the illustration of the parameter influence method.

Let

$$x_1(t) = \frac{\partial y(t)}{\partial c_1} = \frac{\partial y(t)}{\partial \omega} \quad (40)$$

Then, differentiating equation (20) with respect to ω :

$$\frac{\partial^3 y}{\partial \omega \partial t^2} + 2\omega y + \omega^2 \frac{\partial y}{\partial \omega} = 0 \quad (41)$$

Assuming continuity of the appropriate partial derivatives, the order of differentiation can be reversed yielding:

$$\frac{\partial}{\partial t^2} \frac{\partial y}{\partial \omega} + \omega^2 \frac{\partial y}{\partial \omega} = -2\omega y \quad (42)$$

This equation may be written

$$\ddot{x}_1 + \omega^2 x_1 = -2\omega y \quad (43)$$

which is a linear inhomogeneous equation in x_1 . The initial conditions for this equation are obtained from

$$x_1(0) = \frac{\partial y(0)}{\partial \omega} = 0 \quad (44)$$

$$\dot{x}_1(0) = \frac{\partial}{\partial t} \frac{\partial y(0)}{\partial \omega} = 0$$

Equations (43) and (44) taken together uniquely specify $x_1(t)$.

Ordinarily, the forcing function y in equation (43) would be obtained by simultaneous computer solution of equations (43) and (20).

However, in this particular case, y can be obtained analytically and is given by equation (22). Thus, the equation to be solved for x_1 is

$$\ddot{x}_1 + \omega^2 x_1 = -2\omega y(0) \cos \omega t - 2\dot{y}(0) \sin \omega t \quad (45)$$

The solution to this differential equation with the initial conditions given by equation (44) is

$$x_1(t) = - \left[t y(0) + \frac{\dot{y}(0)}{\omega} \right] \sin \omega t + t \frac{\dot{y}(0)}{\omega} \cos \omega t \quad (46)$$

as may be verified by substitution. Since equation (46) represents the partial derivatives of the solution with respect to the parameter c_1 at all values of time, the coefficients x_{i1} of the matrix X are obtained from the relation

$$x_{i1} = x_1(t_1) \quad i = 1, 2, \dots, N \quad (47)$$

The validity of equation (46) may be ascertained by direct differentiation of the system response function, equation (20). This, of course, represents the conventional method for obtaining partial derivatives of an analytic function. However, while this operation is easier for a human being than solving a differential equation, quite the converse is true of an electronic computer. The purpose of the parameter influence equations therefore, is to provide a method for determining partial derivatives which is well suited to computer rather than human implementation. Further illustrations of parameter influence equations and their computational solution are furnished by the examples provided later in this report.

3.4 Stability and Convergence Properties of Gauss-Newton Iteration

3.4.1 Local Stability in the Absence of Measurement Error

An ideal iteration scheme for fitting theoretical response curves to data should not only be convergent, but should also have the property that as a minimizing value for the parameter vector is approached, the ratio of the computed parameter error to the true error tends to unity for each parameter. That is, if \vec{c}_0 is vector such that $\phi(\vec{c}_0)$ is locally minimum, and $\Delta\vec{c}_i$ is the parameter change vector computed at the i^{th} stage of iteration, then it is desirable that

$$\lim_{i \rightarrow \infty} \frac{\Delta\vec{c}_i}{\vec{c}_i - \vec{c}_0} = -I \quad (48)$$

where I is a unit matrix. An iteration procedure which has this property will be termed "asymptotically efficient"; obviously it is also locally convergent. An equivalent statement to equation (48) is

$$\left. \frac{\partial \Delta\vec{c}}{\partial \vec{c}} \right|_{\vec{c} = \vec{c}_0} = -I \quad (49)$$

When experimental conditions are such that the response function, y_0 , can be measured with entirely negligible error and, moreover, the parametric model for the process producing the data points is exact, then \vec{c}_0 will itself be a minimizing value for ϕ . Moreover, in such circumstances

$$\vec{e}(\vec{c}_0) = \vec{y}_0 - \vec{y}(\vec{c}_0) = \vec{0} \quad (50)$$

so

$$\phi(\vec{c}_0; \vec{y}_0) = \vec{e}'\vec{e} = 0 \quad (51)$$

It turns out that, under these conditions, Gauss-Newton iteration is not only stable at $\vec{c} = \vec{c}_0$, but it is also asymptotically efficient. More precisely, if $\vec{\beta}_1$ denotes the parameter change vector computed at $\vec{c} = \vec{c}_1$ by equation (37), then the following theorem states a set of conditions sufficient to ensure the asymptotic efficiency of the iteration:

Theorem 1:

Suppose that the sum squared error function, $\phi(\vec{c})$, takes on the value zero at $\vec{c} = \vec{c}_0$. Then, if $y(t; \vec{c})$ and its first partial derivatives possess a uniformly convergent Taylor series in an ϵ -neighborhood of \vec{c}_0 , it follows that

$$\left. \frac{\partial \phi}{\partial \vec{c}} \right|_{\vec{c} = \vec{c}_0} = -I \quad (52)$$

Proof:

Let $\delta \vec{c}$ denote the true parameter error; i.e.

$$\delta \vec{c} = \vec{c} - \vec{c}_0 \quad (53)$$

Then for $|\delta \vec{c}| < \epsilon$,

$$\vec{y}(\vec{c}) = \vec{y}_0 + X\delta \vec{c} + \frac{1}{2} \sum_{i,j} \left. \frac{\partial^2 \vec{y}}{\partial c_i \partial c_j} \right|_{\vec{c} = \vec{c}_0} \delta c_i \delta c_j + O(\delta \vec{c}^3) \quad (54)$$

The notation $O(\delta c^3)$ indicates a remainder consisting of cubic and higher terms in δc . Now, from the definition of φ , equation (11)

$$\varphi = (\bar{y}_0 - \bar{y})'(\bar{y}_0 - \bar{y})$$

so substituting from equation (54)

$$\varphi = \delta c' X' X \delta c + O(\delta c^3) \quad (55)$$

Differentiating this expression,

$$\bar{\varphi} = 2X' X \delta c + O(\delta c^2) \quad (56)$$

Now from equation (37)

$$\bar{\beta}_1 = -\frac{1}{2} [X_1' X_1]^{-1} \bar{\varphi}(\bar{c}_1) \quad (57)$$

so

$$\bar{\beta}_1 = -[X_1' X_1]^{-1} X_1' X \delta c_1 - \frac{1}{2} [X_1' X_1]^{-1} O(\delta c_1^2) \quad (58)$$

However, under the continuity and full rank assumptions regarding X ,

$$X_1 = X + O(\delta c) \quad (59)$$

so equation (58) reduces to

$$\bar{\beta} = -[X' X]^{-1} X' X \delta c + O(\delta c^2) \quad (60)$$

and consequently

$$\left. \frac{\partial \bar{\beta}}{\partial \bar{c}} \right|_{\bar{c} = \bar{c}_0} = -[X' X]^{-1} X' X = -I \quad (61)$$

3.4.2 Necessary and Sufficient Conditions for Stability in the General Case

In real problems of system identification it is not to be expected that $\phi(\vec{c})$ can be forced to zero for any \vec{c} . Therefore, it is important to determine necessary and sufficient conditions for the convergence of the Gauss-Newton procedure in the general case

$$\min_{\vec{c}} \phi(\vec{c}; \vec{y}_0) > 0 \quad (62)$$

Toward this end, let \vec{c}_e be a value of \vec{c} which locally minimizes ϕ and let \vec{y}_e be the response vector associated with this parameter vector. Then a matrix D , of second partial derivatives may be defined at $\vec{c} = \vec{c}_e$ as⁹

$$D = \left[\frac{\partial^2 \phi}{\partial c_i \partial c_j} (\vec{y}_0 - \vec{y}_e)' \vec{y} \right]_{\vec{c} = \vec{c}_e} \quad (63)$$

Another matrix, Q , may be defined from D as

$$Q = [X'X]^{-1}_{\vec{c} = \vec{c}_e} D = S^{-1}D \quad (64)$$

Utilizing this definition, the following theorem provides the desired conditions for local stability of the Gauss-Newton procedure.

Theorem 2:

Suppose that $y(t; \vec{c})$ and its first partial derivatives possess a uniformly convergent Taylor series in an ϵ -neighborhood of \vec{c}_e in parameter space. Assume also that the matrix X is of full rank in this neighborhood.

⁹ The elements of this matrix may be obtained by solving the appropriate parameter influence equations.

Then if \bar{c}_i denotes the trial parameter vector at the i^{th} stage of iteration and the initial parameter vector, \bar{c}_1 , is chosen so that

$$|\bar{c}_1 - \bar{c}_e| < \epsilon_0, \quad 0 < \epsilon_0 < \epsilon \quad (65)$$

it follows that there exists an ϵ_0 for which

$$\lim_{i \rightarrow \infty} \bar{c}_i = \bar{c}_e \quad (66)$$

for any c_1 satisfying equation (65) if and only if all of the eigenvalues of the matrix Q are less than one in absolute value.

Proof:

As before, let $\delta\bar{c}$ stand for the true parameter error measured from the minimizing vector

$$\delta\bar{c}_i = \bar{c}_i - \bar{c}_e \quad (67)$$

Then an equivalent statement to equation (66) is

$$\lim_{i \rightarrow \infty} \delta\bar{c}_i = \bar{0} \quad (68)$$

Now for $|\delta\bar{c}| < \epsilon$

$$\bar{y}(\bar{c}) = \bar{y}_e + X\delta\bar{c} + \frac{1}{2} \sum_{i,j} \frac{\partial^2 \bar{y}}{\partial c_i \partial c_j} \Big|_{\bar{c} = \bar{c}_e} \delta c_i \delta c_j + O(\delta\bar{c}^3) \quad (69)$$

so, making use of this expansion

$$\begin{aligned}
\phi &= (\bar{y}_0 - \bar{y})' (\bar{y}_0 - \bar{y}) \\
&= (\bar{y}_0 - \bar{y}_e)' (\bar{y}_0 - \bar{y}_e) - 2(\bar{y}_0 - \bar{y}_e)' X \delta c \\
&= (\bar{y}_0 - \bar{y}_e)' \sum_{i,j} \frac{\partial^2 \bar{y}}{\partial c_i \partial c_j} \delta c_i \delta c_j \\
&\quad + \delta c' X' X \delta c + o(\delta c^3)
\end{aligned} \tag{70}$$

Referring to equation (36) and noting also that $(\bar{y}_0 - \bar{y}_e)'$ is a constant vector which may be taken inside of the differentiation operation, the expression for ϕ reduces to

$$\phi = \phi(\bar{c}_e) + \nabla \phi(\bar{c}_e)' \delta c + \delta c' (S - D) \delta c + o(\delta c^3) \tag{71}$$

Since \bar{c}_e is, by definition, a minimum,

$$\nabla \phi(\bar{c}_e) = 0 \tag{72}$$

and ϕ simplifies still further to

$$\phi = \phi(\bar{c}_e) + \delta c' (S - D) \delta c + o(\delta c^3) \tag{73}$$

Now, utilizing the Gauss-Newton iteration equation

$$\bar{b}_1 = -\frac{1}{2} S_1^{-1} \nabla \phi \tag{74}$$

where S_1 denotes an evaluation of the matrix S at \bar{c}_1 . Because of the assumptions regarding y and X , this equation may be rewritten

$$\bar{b}_1 = -\frac{1}{2} S^{-1} \nabla \phi + o(\delta c) \nabla \phi \tag{75}$$

From equation (73)

$$\mathcal{P} = 2(S - D) \bar{\delta}c + O(\bar{\delta}c^2) \quad (76)$$

so

$$\begin{aligned} \bar{\beta}_1 &= -S^{-1} (S - D) \bar{\delta}c + O(\bar{\delta}c^2) \\ &= (-I + S^{-1}D) \bar{\delta}c + O(\bar{\delta}c^2) \end{aligned} \quad (77)$$

Now

$$\bar{c}_{i+1} = \bar{c}_e + \bar{\delta}c_{i+1} = \bar{c}_i + \bar{\beta}_1 = \bar{c}_e + \bar{\delta}c_i + \bar{\beta}_1 \quad (78)$$

so

$$\bar{\delta}c_{i+1} = \bar{\delta}c_i + \bar{\beta}_1 \quad (79)$$

$$\begin{aligned} &= \bar{\delta}c_i - I \bar{\delta}c_i + S^{-1}D \bar{\delta}c_i + O(\bar{\delta}c_i^2) \\ &= S^{-1}D \bar{\delta}c_i + O(\bar{\delta}c_i^2) = Q \bar{\delta}c_i + O(\bar{\delta}c_i^2) \end{aligned} \quad (80)$$

By choosing ϵ_0 sufficiently small, the quadratic term in this expression may be made as small as desired in comparison to the linear term. It follows, therefore, that for $|\bar{\delta}c_1| < \epsilon_0$

$$\bar{\delta}c_i = Q^{i-1} \bar{\delta}c_1 + O(\bar{\delta}c_{i-1}^2) \quad (81)$$

and consequently

$$\lim_{i \rightarrow \infty} \bar{\delta}c_i = \lim_{i \rightarrow \infty} Q^{i-1} \bar{\delta}c_1 + \lim_{i \rightarrow \infty} O(\bar{\delta}c_{i-1}^2) = 0 \quad (82)$$

if and only if

$$\lim_{i \rightarrow \infty} Q^i = [0] \quad (83)$$

This situation in turn prevails if and only if all eigenvalues of Q are less than one in absolute value. This completes the proof.

The following additional inferences may be made from this theorem:

Corollary I:

If $D = [0]$, then the Gauss-Newton procedure not only converges, but it is also asymptotically efficient. This condition is satisfied, for example, when the response function depends linearly upon the parameters. It is also satisfied when y_e corresponds exactly to y_0 ; i.e. when no measurement error is present and $\bar{c}_e = \bar{c}_0$. Thus, Theorem 1 is a special case of Theorem 2.

Corollary II:

If the second derivatives of \bar{y} are computed and

$$D_i = (\bar{y}_0 - \bar{y}(c_i))' \frac{\partial^2 \bar{y}}{\partial c_i \partial c_j} \Big|_{\bar{c} = \bar{c}_i} \quad (84)$$

then providing the second partials of \bar{y} are continuous in $|\delta \bar{c}| < \epsilon$, the adjusted Gauss-Newton iteration equation given by

$$\Delta \bar{c}_i = \bar{\alpha}_i = (S_i - D_i)^{-1} \bar{y}'(c_i) \quad (85)$$

will always converge locally and will be asymptotically efficient.

3.4.3 Convergence with Scale Factor Adjustment

When a criterion function and its partial derivatives are continuous in an ϵ -neighborhood of a point, \vec{c}_1 , in parameter space, then if $\vec{\nabla}\phi(\vec{c}_1)$ differs from zero it is always possible to reduce $\phi(\vec{c})$ by proceeding in parameter space along the negative gradient for a sufficiently small distance. Equation (37) shows that Gauss-Newton iteration is based upon a linearly transformed step along the gradient. This linear transformation will, in general, induce both a magnitude and an angle change in $\vec{\nabla}\phi$ and (as shown by the preceding theorem) may well fail to produce a convergent estimator of the true parameter vector, \vec{c}_0 . It is relevant, therefore, to consider the possibility of stabilizing Gauss-Newton iteration in such circumstances by reducing $\vec{\beta}$ by an appropriate scale factor whenever the basic procedure fails to converge. The crucial question involved in the implementation of such a scheme relates to the angle which exists between the gradient vector and the Gauss-Newton correction vector, $\vec{\beta}$. Specifically, since the gradient vector is normal to the contour lines of ϕ , motion along $\vec{\beta}$ for a sufficiently small distance will result in a decrease in ϕ providing only that the angle between $\vec{\beta}$ and $\vec{\nabla}\phi$ is obtuse; i.e., providing that the scalar product of $\vec{\beta}$ and $\vec{\nabla}\phi$ is negative. The following lemma expresses this statement more precisely and shows that the angle condition is always satisfied.

Lemma 1:

Consider a region, R , in parameter space such that the matrix X is of full rank everywhere in R . Assume, furthermore, that about every

point in R there exists an ϵ -neighborhood in which a uniformly convergent Taylor series may be constructed for the criterion function, $\phi(\vec{c})$. Then for any point in R such that $\vec{\phi}(\vec{c}) \neq \vec{0}$

$$\left. \frac{d}{d\eta} \phi(\vec{c} + \eta\vec{\beta}) \right|_{\eta=0} = \vec{\phi}'\vec{\beta} < 0 \quad (86)$$

Proof:

Locally, ϕ is represented at any point in R by

$$\phi(\vec{c} + \eta\vec{\beta}) = \phi(\vec{c}) + \vec{\phi}'\eta\vec{\beta} + o(\eta^2) \quad (87)$$

so

$$\frac{d\phi}{d\eta} = \vec{\phi}'\vec{\beta} + o(\eta) \quad (88)$$

and

$$\left. \frac{d\phi}{d\eta} \right|_{\eta=0} = \vec{\phi}'\vec{\beta} = \vec{\beta}'\vec{\phi} \quad (89)$$

Now from equation (37)

$$\vec{\phi} = -2S\vec{\beta} = -2X'X\vec{\beta} \quad (90)$$

so

$$\vec{\beta}'\vec{\phi} = -2\vec{\beta}'X'X\vec{\beta} = -2(X\vec{\beta})'(X\vec{\beta}) = -2|X\vec{\beta}|^2 \leq 0 \quad (91)$$

However, since X is of full rank, $|X\vec{\beta}| \neq 0$ unless $\vec{\beta} = \vec{0}$. Furthermore, since $\vec{\phi} \neq \vec{0}$, equation (90) shows that $\vec{\beta} \neq \vec{0}$ so the inequality in equation

(91) must be strict; i.e.

$$\vec{\beta}' \vec{\nabla} \phi = -2 |\vec{x}\vec{\beta}|^2 < 0 \quad (92)$$

This completes the proof.

Lemma 1 provides the basis for the following basic convergence theorem for the modified Gauss-Newton procedure¹⁰.

Theorem 3:

Assume that the conditions of Lemma 1 are satisfied. Suppose that the region R is in addition convex and has the further properties that the criterion function, $\phi(\vec{c})$, is concave in R and possesses a minimum at an interior point, \vec{c}_e , of R . Let ϕ^* be the infimum of ϕ assumed on the boundary of R and let \vec{c}_1 denote the parameter vector obtained at the i^{th} stage of iteration. Then for any initial point, \vec{c}_1 , interior to R such that

$$\phi(\vec{c}_1) < \phi^* \quad (93)$$

it is possible to choose a sequence of points

$$\vec{c}_{i+1} = \vec{c} + \eta_i \vec{\beta}_i, \quad 0 < \eta_i < 1 \quad (94)$$

such that

$$\lim_{i \rightarrow \infty} \vec{c}_i = \vec{c}_e \quad (95)$$

¹⁰ A related theorem involving more restrictions on R has been proved by H. O. Hartley [14]. Hartley refers to the Gauss-Newton iteration procedure with scale factor adjustment as the "modified Gauss-Newton Method".

Proof:

From equation (87), ϕ is represented locally at every point in R by

$$\phi(\bar{c} + \eta\bar{\beta}) = \phi(\bar{c}) + \eta\bar{\beta}'\bar{\phi} + O(\eta^2) \quad (96)$$

Now due to the concavity of ϕ , if $\bar{c}_i \neq \bar{c}_e$, then $\bar{\phi} \neq 0$ and

$$\bar{\beta}'\bar{\phi}(\bar{c}_i) = k_i < 0 \quad (97)$$

so there exists an η_i at every stage of iteration such that when \bar{c}_{i+1} is chosen according to equation (94), then \bar{c}_{i+1} is interior to R and

$$\phi(\bar{c}_{i+1}) = \phi(\bar{c}_i + \eta_i\bar{\beta}_i) < \phi(\bar{c}_i) \quad (98)$$

Since ϕ is bounded from below by $\phi(\bar{c}_e)$ and $\phi(\bar{c}_i)$ forms a monotone decreasing sequence for the \bar{c}_i so chosen, it follows that

$$\lim_{i \rightarrow \infty} \phi(\bar{c}_i) = \phi(\bar{c}_e) \quad (99)$$

The concavity of ϕ in R then guarantees that the sequence in parameter space also converges; i.e.

$$\lim_{i \rightarrow \infty} \bar{c}_i = \bar{c}_e \quad (100)$$

3.4.4 Discussion and Conclusions

Theorem 1 can be utilized in circumstances where no real physical measurements are to be made. It is applicable, for example, in situations where a differential equation is to be determined whose solution is a given

analytic or computed function [10]. Since this theorem relates to local stability only, it may be necessary to employ a scale factor reducing algorithm in the initial stages of an iteration in order to achieve a convergent sequence of parameter vectors. The existence of such an algorithm is guaranteed by Theorem 3. It does not appear that Theorem 1 is of any significance in parameter estimation problems involving real physical measurements since errors are inevitably present in such data.

Theorem 2 provides necessary and sufficient conditions for the Gauss-Newton iteration to be locally stable without scale factor adjustment in the case where the minimum of ϕ is greater than zero. As a practical computational matter, it is unlikely that any eigenvalues of the matrix Q will be close to unity. More likely, either all values will be considerably less than one or else at least one will be considerably larger than one. Thus one would expect that unmodified Gauss-Newton iteration will either converge very rapidly in the later stages of iteration or else it will sharply diverge. In any event, computation of D for the purpose of establishing convergence does not seem worthwhile. The number of differential equations to be solved to obtain the elements of D increases alarmingly with an increase in the dimensionality of \vec{c} . It is much easier to simply compute $\phi(\vec{c}_i + \vec{\beta}_i)$ at each step to see if an improvement results than to investigate stability by determining D and its eigenvalues.

The above remarks also apply to the modified Gauss-Newton equation of Corollary II. A simple search along the vector $\vec{\beta}$ may be carried out with less labor than is involved in evaluating D at each stage.

The net conclusions to be drawn from the stability theorems developed in this section may be stated in terms of the following computation policy:

1. If the Gauss-Newton parameter change vector, $\vec{\beta}_i$, provides a reduction in the sum-squared error functions, $\phi(\vec{c}; \vec{y}_0)$, at the i -th stage of iteration, use $\vec{c}_{i+1} = \vec{c}_i + \vec{\beta}_i$.
2. If this is not so, then determine \vec{c}_{i+1} by searching along $\vec{\beta}_i$ for a locally minimum value of ϕ .

The implementation of this policy in terms of a complete computational algorithm for unconstrained nonlinear regression analysis is presented later in this report.

4. STATISTICAL PROPERTIES OF NONLINEAR LEAST SQUARES PARAMETER ESTIMATES

4.1 Linearized Estimation of the Variability of Parameter Estimates

The preceding investigation of the stability properties of iterated linear least squares regression analysis shows that it is possible to devise a computational algorithm which will ensure convergence of a sequence of parameter estimating vectors to a vector, \vec{c}_0 , which locally minimizes the sum-squared error function, $\phi(\vec{c})$. This vector may fail to accurately estimate the true parameter vector, \vec{c}_0 , for two reasons. First of all, it is possible that the iteration process will converge to a local minimum of $\phi(\vec{c})$ which is not the minimum associated with $\vec{c} = \vec{c}_0$. The detection

and correction of this type of error is a very difficult problem whose treatment has been reserved for a later report. Secondly, even if the computational procedure converges to the proper minimum of $\phi(\vec{c})$, measurement and other errors will prevent computer generated solutions from fitting the data exactly. Consequently, there will remain a residual statistical error in the estimation of the unknown system parameters. Estimation of the magnitude of this type of error by analytic procedures is extremely difficult when the systems involved are nonlinear and the errors are large. Quite often, error estimates can be obtained only by recourse to statistical methods involving many repetitions of the same physical experiment. However, if it can be assumed that \vec{c}_e is quite close to \vec{c}_0 , then an asymptotic error estimate may be obtained by linearizing the relationship between measurement errors and estimation errors.

In order to obtain the desired error estimate, assume that $\vec{y}(\vec{c}_e)$ can be represented by a uniformly convergent Taylor series expansion about \vec{c}_0 . Denote the difference between the estimated and true parameter vector by $\delta\vec{c}$; i.e.

$$\delta\vec{c} = \vec{c}_e - \vec{c}_0 \quad (101)$$

Then

$$\vec{y}(\vec{c}_e) = \vec{y}(\vec{c}_0) + X(\vec{c}_0) \delta\vec{c} + \frac{1}{2} \sum_{i,j} \left. \frac{\partial^2 \vec{y}}{\partial c_i \partial c_j} \right|_{\vec{c} = \vec{c}_0} \delta c_i \delta c_j + O(\delta c^3) \quad (102)$$

Now, the data vector, \bar{y}_0 , is given by equation (3) as $\bar{y}_0 = \bar{y}(\bar{c}_0) + \bar{\varepsilon}$ so the residual error vector, \bar{e} , may be written

$$\bar{e} = \bar{y}_0 - \bar{y} = \bar{\varepsilon} - X\delta\bar{c} - \frac{1}{2} \sum_{i,j} \frac{\partial^2 \bar{y}}{\partial c_i \partial c_j} \delta c_i \delta c_j + O(\delta\bar{c}^3) \quad (103)$$

and

$$\begin{aligned} \phi &= \bar{e}'\bar{e} = \bar{\varepsilon}'\bar{\varepsilon} - 2\delta\bar{c}'X'\bar{\varepsilon} + \delta\bar{c}'X'X\delta\bar{c} \\ &\quad - \sum_{i,j} \left(\frac{\partial^2}{\partial c_i \partial c_j} \bar{\varepsilon}'\bar{y} \right) \delta c_i \delta c_j + O(\delta\bar{c}^3) \\ &= \bar{\varepsilon}'\bar{\varepsilon} - 2\delta\bar{c}'X'\bar{\varepsilon} + \delta\bar{c}'(S-D)\delta\bar{c} + O(\delta\bar{c}^3) \end{aligned} \quad (104)$$

where S and D have the same significance as before. Assuming that term-wise differentiation of the series for ϕ is permissible,

$$\frac{d\phi}{d\bar{c}} = \frac{d\phi}{d(\delta\bar{c})} = -2X'\bar{\varepsilon} + 2(S-D)\delta\bar{c} + O(\delta\bar{c}^2) \quad (105)$$

Since \bar{c}_e is a minimizing value for ϕ , this expression must equal zero and therefore, to a first order of approximation

$$X'\bar{\varepsilon} = (S - D)\delta\bar{c} \quad (106)$$

or

$$\delta\bar{c} = (S - D)^{-1} X'\bar{\varepsilon}$$

Typically, ε_i will be a random variable with zero mean. In such circumstances

$$E(\delta\vec{c}) = (S-D)^{-1} X' E(\vec{\varepsilon}) = (S-D)^{-1} X' \vec{0} = \vec{0} \quad (107)$$

and the covariance matrix for $\delta\vec{c}$ is given by

$$\begin{aligned} [\text{cov } \delta\vec{c}] &= E\{\delta\vec{c} \delta\vec{c}'\} = E\{(S-D)^{-1} X' \vec{\varepsilon} [(S-D)^{-1} X' \vec{\varepsilon}]'\} \\ &= (S-D)^{-1} X' [\text{cov } \vec{\varepsilon}] X (S-D)^{-1} \end{aligned} \quad (108)$$

It is a common occurrence in experimental situations to find that

$$[\text{cov } \vec{\varepsilon}] = \sigma^2 I \quad (109)$$

When this is the case

$$\begin{aligned} [\text{cov } \delta\vec{c}] &= \sigma^2 (S-D)^{-1} X' X (S-D)^{-1} \\ &= \sigma^2 (S-D)^{-1} S (S-D)^{-1} \end{aligned} \quad (110)$$

If, furthermore, $D = [0]$, (as in linear regression analysis, for example), then the covariance matrix reduces still further to

$$[\text{cov } \delta\vec{c}] = \sigma^2 S^{-1} S S^{-1} = \sigma^2 S^{-1} \quad (111)$$

which is a standard result in linear statistical theory [16].

4.2 Relationship of Length of Record to Parameter Estimation Accuracy

It is a general characteristic of estimation problems that accuracy may be increased by considering more data points. When the estimation problem involves determination of dynamic system parameters, this means that it should be possible to increase the reliability of parameter estimates

by observing the system for a longer period of time. In situations where the system under observation is known to be linear in the unknown parameters, it is possible to obtain quantitative relationships linking estimation accuracy to the amount of data available in a relatively straightforward fashion [17]. However, when the system is nonlinear, determination of such relationships becomes very difficult.

While the relationship between observation time and estimation accuracy may very well be unattainable by analytic methods in the general nonlinear case, under certain simplifying assumptions, an asymptotic error theory can be constructed. In order to arrive at such a theory, it is necessary to modify the parameter estimation equation slightly to avoid the possibility of the S matrix becoming singular in the limit. The estimation formula given by equation (37) may be rewritten as

$$\bar{\beta} = \frac{1}{2} NS^{-1} \bar{\gamma} \frac{\sigma}{N} = \frac{1}{2} \left(\frac{1}{N} S\right)^{-1} \bar{\gamma} \frac{\sigma}{N} \quad (112)$$

and in a similar fashion, equation (110) can be expressed as

$$[\text{cov } \delta \bar{c}] = \frac{\sigma^2}{N} \left(\frac{S}{N} - \frac{D}{N}\right)^{-1} \frac{S}{N} \left(\frac{S}{N} - \frac{D}{N}\right)^{-1} \quad (113)$$

With the covariance matrix for $\delta \bar{c}$ written in this form, it becomes possible to consider its behavior as the number of samples, N , approaches infinity.

Assume for the moment that the matrices S/N and D/N have limiting values as N approaches infinity. Specifically let

$$\lim_{N \rightarrow \infty} \frac{S}{N} = V$$

$$\lim_{N \rightarrow \infty} \frac{D}{N} = U \quad (114)$$

Then, providing that the limiting operations can be interchanged with the algebraic operations, it follows that

$$\lim_{N \rightarrow \infty} N[\text{cov } \delta \vec{c}] = \sigma^2 [V - U]^{-1} V [V - U]^{-1} \quad (115)$$

This equation reveals that the magnitude of the elements of the covariance matrix vary inversely with the number of samples; this is a familiar fact in linear statistics [18].

The existence of the assumed limit matrices may be related to the behavior of the solution partial derivatives. Referring to the definition of the matrix of partial derivatives, X , (equation (30)) it can be seen that this matrix may be written as a row vector

$$X = (\vec{X}_1, \vec{X}_2, \dots, \vec{X}_M) \quad (116)$$

whose elements are the column vectors, \vec{X}_i . Each \vec{X}_i is in turn given by

$$\vec{X}_i = \begin{pmatrix} \frac{\partial y(t_1)}{\partial c_i} \\ \frac{\partial y(t_2)}{\partial c_i} \\ \vdots \\ \frac{\partial y(t_N)}{\partial c_i} \end{pmatrix} \quad (117)$$

Since the matrix S is defined as $S = X'X$, this notation permits the elements, S_{ij} , of S to be written as

$$S_{ij} = \bar{X}_i' \bar{X}_j \quad (118)$$

That is, the elements of S are just the scalar products of the appropriate columns of the matrix X . The elements of the normalized matrix, V , are in turn given by

$$v_{ij} = \lim_{N \rightarrow \infty} \frac{\bar{X}_i' \bar{X}_j}{N} \quad (119)$$

When $i = j$, this expression reduces to

$$v_{ii} = \lim_{N \rightarrow \infty} \sum_{k=1}^N \frac{X_{ki}^2}{N} = \lambda_{ii} \quad (120)$$

i.e. v_{ii} is just the mean square value of the partial derivative $\partial y / \partial c_i$. Consequently

$$\lambda_{ii} \geq 0 \quad (121)$$

when it exists. Moreover, unless the partial derivative, $\partial y / \partial c_i$, tends toward zero as t becomes arbitrarily large, the inequality is strict; i.e.

$$\lambda_{ii} > 0 \quad (122)$$

When $i \neq j$, the situation is somewhat different. Since the partial derivatives involved may be either positive or negative, the terms in the sum

$$v_{ij} = \lim_{N \rightarrow \infty} \sum_{k=1}^N \frac{X_{ki} X_{kj}}{N} = \lambda_{ij} \quad (123)$$

may have either positive or negative signs and nothing can be said about the sign of λ_{ij} . It is, in fact, quite possible that λ_{ij} may equal zero even though X_i and X_j do not individually tend to zero.

The asymptotic behavior of the matrix D may likewise be related to the characteristics of certain vectors. Referring to equation (63)

$$D = \left[\frac{\partial^2}{\partial c_i \partial c_j} (\bar{y}_o - \bar{y}_e)' \bar{y} \right]_{\bar{c} = \bar{c}_e}$$

The vector $\bar{y}_o - \bar{y}_e$ is the residual error between the data and the best fit generated in the estimation process. It is therefore a constant vector which may be factored out. Making use of this fact, the elements of the vector U may be written

$$U_{ij} = \lim_{N \rightarrow \infty} \sum_{k=1}^N \frac{1}{N} \left\{ [y_o(t_k) - y_e(t_k)] \frac{\partial^2 y(t_k)}{\partial c_i \partial c_j} \right\} \quad (124)$$

The convergence of the matrix D/N to the limit matrix U is thus contingent upon the existence of an average cross product between the two terms of equation (124). As in the case of the off diagonal elements of the V matrix, every element of U may be positive, negative, or zero.

There is one circumstance of particular interest in relation to the above results. If it should happen that the vectors \bar{X}_i behave as if they were samples of uncorrelated random variables, then, in the limit, S/N must become the diagonal matrix

$$V = \begin{pmatrix} \lambda_{11} & \cdot & \cdot & \cdot & \cdot & 0 \\ \cdot & \lambda_{22} & & & & \cdot \\ \cdot & & \cdot & & & \cdot \\ \cdot & & & \cdot & & \cdot \\ \cdot & & & & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot & \lambda_{MM} \end{pmatrix} \quad (125)$$

If the vectors $\bar{y}_0 - \bar{y}_e$ and $\partial^2 \bar{y} / \partial c_i \partial c_j$ are likewise uncorrelated for every i and j , then U must tend to a zero matrix. When this occurs, equation (115) reduces to

$$\lim_{N \rightarrow \infty} N[\text{cov } \delta \bar{c}] = \sigma^2 V^{-1} \quad (126)$$

or

$$E\{\delta c_i^2\} = \frac{\sigma^2}{N \lambda_{ii}} \quad (127)$$

and

$$E\{\delta c_i \delta c_j\} = 0, \quad i \neq j \quad (128)$$

5. AN AUTOMATIC COMPUTATIONAL ALGORITHM FOR PARAMETER IDENTIFICATION BY NONLINEAR REGRESSION ANALYSIS

5.1 Stopping Rules

To provide a fully automatic basis for the determination of minima of a criterion function, it is necessary to provide a stopping rule for the iteration sequence, $\{\bar{c}_k\}$. That is, the iteration cycle must be subjected to a supervisory procedure which determines that further refinement of the estimate of the true parameter vector, \bar{c}_0 , is not justified. Such a stopping rule may be based, for example, on a percentage change in either the parameter space or the criterion function itself. Let

$$d_c = \frac{\frac{\bar{\Delta c}_i}{\bar{c}_i} - \frac{\bar{\Delta c}_i}{\bar{c}_i}}{\frac{\bar{\Delta c}_i}{\bar{c}_i} + \frac{\bar{\Delta c}_i}{\bar{c}_i}} \quad (129)$$

and

$$d_\phi = \frac{\phi(\bar{c}_i) - \phi(\bar{c}_{i+1})}{\phi(\bar{c}_i)} \quad (130)$$

be the normalized differences between two successive computations measured in terms of the parameter estimates and criterion function respectively.

Then an appropriate stopping rule is:

1. Stop if $d_c < \epsilon_c$
2. Stop if $d_\phi < \epsilon_\phi$

The choice of the values of ϵ_c and ϵ_ϕ must be made in advance by the human experimenter. Since d_c and d_ϕ may behave quite differently, the initial values for these two quantities may require adjustment during the course of an experiment. While this particular stopping rule is somewhat arbitrary, it is essential in every case that some criterion for halting computation be provided since otherwise the computer will continue to calculate new values of \bar{c} indefinitely even though the practical value of greater precision may be entirely negligible.

5.2 Stabilization of the Iteration by Binary Search

In the discussion of the significance of the convergence theorems, a specific computational policy was proposed. This policy involved stabilization of the iteration procedure by superimposing a scale factor search routine onto the basic Gauss-Newton iteration scheme. A very simple type of search may be obtained by simply reducing the parameter change vector by a factor of two repeatedly until a value of the criterion function is obtained such that:

1. ϕ is less than the starting value.
2. ϕ is locally minimized (with respect to the binary search along $\bar{\beta}$).

Figure 1 is a flow diagram for a computer program to achieve the desired stabilization. This diagram also includes a "stability" test for the function $y(t; \bar{c})$ for the purpose of assuring that the magnitude of the computed variables does not exceed the range of the computer arithmetic registers. The equations written in the boxes of Figure 1 represent difference or "up-dating" relationships as is conventional in such diagrams. The sequence of operations is always to be taken from top to bottom in each box.

It is certainly possible to devise more sophisticated search procedures than the one shown on Figure 1. More efficient techniques are currently being developed and will provide the basis for future reports. However, binary search has been found to be entirely satisfactory for the purpose of experimentally verifying the theoretical results obtained in this study.

5.3 Flow Diagram for the Complete Algorithm

All of the preceding discussion is summarized in the flow diagram displayed as Figure 2. This diagram has been coded in the Fortran compiling language and utilized to produce the experimental results presented later in this report. The numbers appearing at various points of Figures 1 and 2 refer to statement numbers from the Fortran program. While the program itself has not been included as part of this document, copies of both the Fortran decks and a listing of these decks are available upon request.

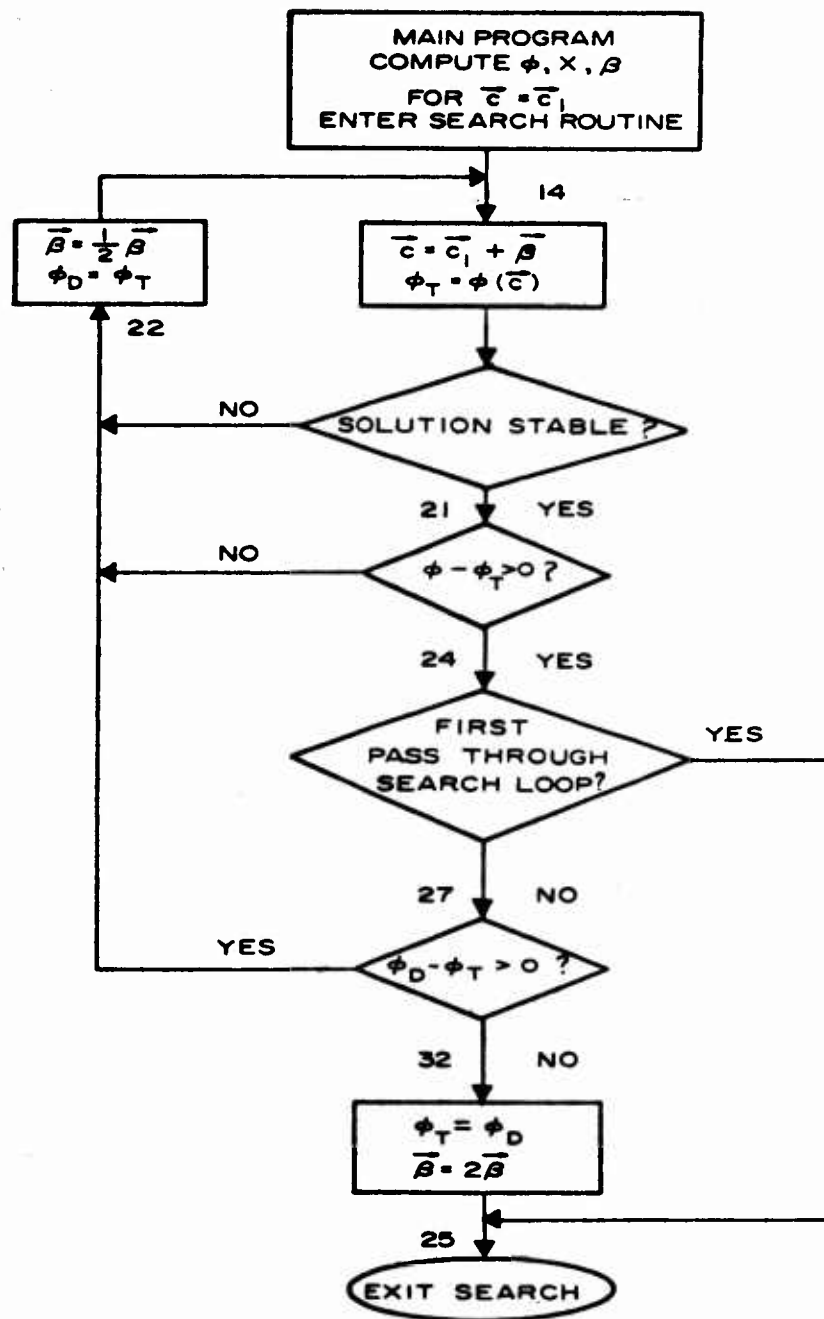


FIGURE 1 FLOW CHART FOR BINARY SEARCH .

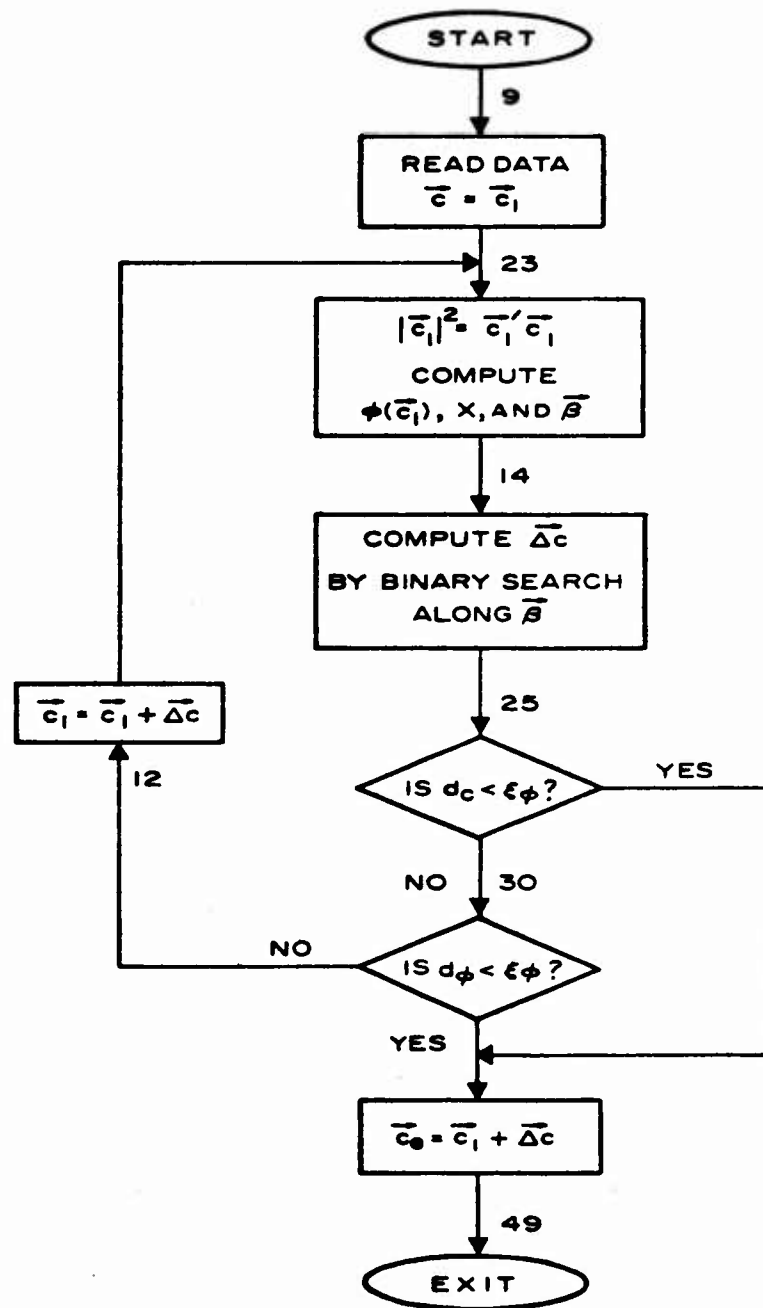


FIGURE 2 FLOW DIAGRAM FOR PARAMETER ESTIMATION BY NONLINEAR REGRESSION ANALYSIS.

5.4 Equations for the Inference of Pendulum Parameters

The experimental part of this study is concerned with the estimation of parameters for the pendulum equation (2)

$$c_2 \ddot{\theta} + c_1 \dot{\theta} + \sin \theta = 0$$

In conformity with the notation established for regression analysis, this may be rewritten as

$$c_2 \ddot{y} + c_1 \dot{y} + \sin y = 0 \quad (131)$$

The parameter vector, \vec{c} , is therefore given by

$$\vec{c} = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ y(0) \\ \dot{y}(0) \end{pmatrix} \quad (132)$$

and elements of the parameter influence matrix, X , consist of the partial derivatives

$$x_{ij} = \frac{\partial y(t_i)}{\partial c_j}$$

These partial derivatives are obtainable from the appropriate parameter influence equations. For example, differentiating equation (131) with respect to c_1 yields

$$\frac{\partial}{\partial c_1} (c_2 \ddot{y} + c_1 \dot{y} + \sin y) = \frac{\partial}{\partial c_1} 0 = 0 \quad (133)$$

or

$$c_2 \frac{\partial^3 y}{\partial c_1 \partial t^2} + c_1 \frac{\partial^2 y}{\partial c_1 \partial t} + \frac{\partial y}{\partial t} + \cos y \frac{\partial y}{\partial c_1} = 0 \quad (134)$$

Assuming that the order of differentiation may be interchanged, this last equation may be written

$$c_2 \ddot{x}_1 + c_1 \dot{x}_1 + x_1 \cos y = -\dot{y} \quad (135)$$

Repeating these steps with respect to each component of \vec{c} results in the set of equations

$$c_2 \ddot{x}_2 + c_1 \dot{x}_2 + x_2 \cos y = -\dot{y} \quad (136)$$

$$c_2 \ddot{x}_3 + c_1 \dot{x}_3 + x_3 \cos y = 0 \quad (137)$$

$$c_2 \ddot{x}_4 + c_1 \dot{x}_4 + x_4 \cos y = 0 \quad (138)$$

The initial conditions necessary for the solution of these equations are given by

$$x_1(0) = \frac{\partial y}{\partial c_1}(0) = 0 ; \dot{x}_1(0) = \frac{\partial}{\partial t} \frac{\partial y}{\partial c_1}(0) = \frac{\partial \dot{y}}{\partial c_1}(0) = 0 \quad (139)$$

$$x_2(0) = \frac{\partial y}{\partial c_2}(0) = 0 ; \dot{x}_2(0) = \frac{\partial \dot{y}}{\partial c_2}(0) = 0 \quad (140)$$

$$x_3(0) = \frac{\partial y}{\partial y(0)}(0) = 1 ; \dot{x}_3(0) = \frac{\partial \dot{y}}{\partial y(0)}(0) = 0 \quad (141)$$

$$x_4(0) = \frac{\partial y}{\partial \dot{y}(0)}(0) = 0 ; \dot{x}_4(0) = \frac{\partial \dot{y}}{\partial \dot{y}(0)}(0) = 1 \quad (142)$$

This completes the set of equations required for regression analysis. The matrix X may be calculated for an initial parameter vector estimate, $\vec{c} = \vec{c}_1$, by simultaneously solving equation (131) with equations (135)

thru (138) subject to the above initial conditions. Quantitative results obtained by carrying out this procedure are presented in the next section of this report.

6. EXPERIMENTAL RESULTS

6.1 Terminal Convergence Without Measurement Noise

In order to illustrate the behavior predicted by Theorem 1, the program illustrated in Figure 2 was executed using noise-free data generated by the computer. The data was produced by numerical solution of the pendulum equation using the parameter vector

$$\vec{c}_0 = \begin{pmatrix} .5 \\ 1.0 \\ 1.569 \\ .0 \end{pmatrix} \quad (143)$$

The resulting response function, $y(t; \vec{c}_0)$, is shown on Figure 3 as the curve labeled "true response". The data vector, \vec{y}_0 , required for parameter estimation was obtained by using the ten equally spaced samples of $y(t; \vec{c}_0)$ shown on the figure as "accurate data points".

The initial guess for the parameter vector used to start the iteration was taken as

$$\vec{c}_1 = \begin{pmatrix} .51 \\ 1.01 \\ 1.57 \\ .01 \end{pmatrix} \quad (144)$$

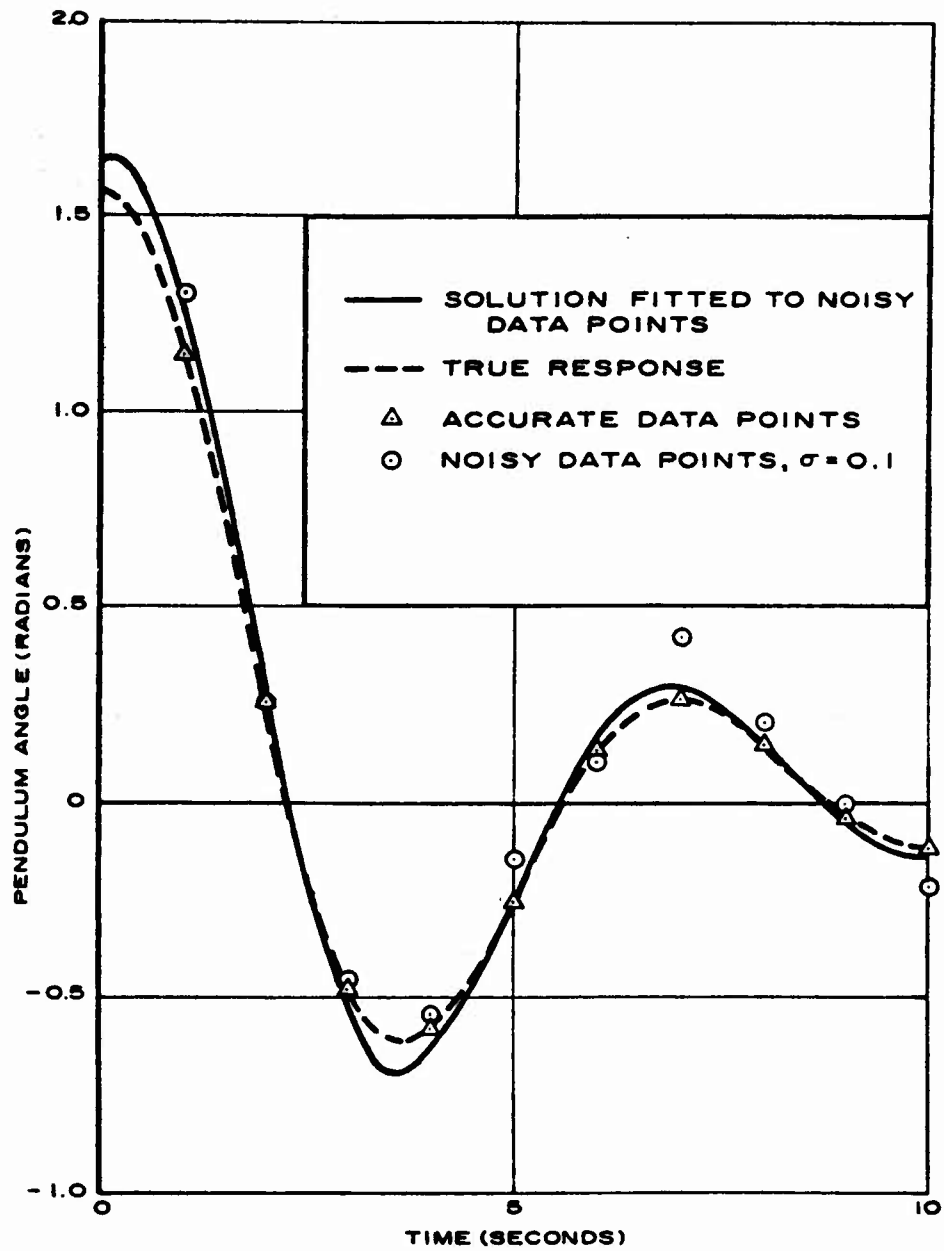


FIGURE 3 RESPONSE CURVE FITTED TO NOISY DATA BY THE COMPUTER.

These values were deliberately chosen to be very close to the true parameters so that only the terminal convergence properties of unmodified Gauss-Newton iteration would be displayed. The stopping criterion used for this experiment was

$$\epsilon_{\phi} = \epsilon_c = 10^{-6}$$

This very stringent criterion was satisfied after only two unmodified Gauss-Newton iteration cycles. The sum-squared error was reduced from a value of 1.85×10^{-3} to a value of 1.09×10^{-13} by the two steps taken in parameter space. Table 1 shows the sequence of values generated by the computer for the parameter vector, \vec{c} . These results strongly support the conclusions of Theorem 1.

TABLE 1
RESULTS OF GAUSS-NEWTON ITERATION USING EXACT DATA

Variable	Starting Value	First Iteration	Second Iteration	True Value
c_1	0.51000000	0.50013945	0.50000021	0.50000000
c_2	1.0100000	0.99966252	1.0000002	1.0000000
c_3	1.5700000	1.5690003	1.5690004	1.5690000
c_4	0.01000000	0.00007233	0.00000068	0.00000000
ϕ	1.853×10^{-3}	4.428×10^{-7}	1.086×10^{-13}	0

6.2 Stabilization of Gauss-Newton Iteration by Binary Scale Factor Adjustment

According to Lemma 1, binary scale factor adjustment should always produce a monotonically decreasing sequence of values for the sum-squared error function, ϕ . To demonstrate this property, a computer experiment was conducted utilizing the modified Gauss-Newton iteration technique illustrated in Figure 2. The "accurate data" vector computed to produce Table 1 was also used in this experiment. However, the parameter vector selected as the starting point for the iteration was intentionally chosen to be quite far from the true parameter vector so that searching would be required.

Table 2 summarizes the results obtained from 60 iteration cycles. The variable N appearing in this table refers to the number of times the computer reduced the parameter change vector, $\vec{\beta}$, by a factor of two while searching for a local minimum within each cycle. Large values for N indicate that the unmodified Gauss-Newton procedure is very unstable at that point.

This example agrees with theory in the respect that a lower value for ϕ was obtained at every stage. It also shows that while convergence does occur in the sequence of values for ϕ , successive values for \vec{c} may vary quite erratically. It must be emphasized however, that the starting value for \vec{c} which produced this data was grossly in error and was deliberately chosen to illustrate this type of behavior. Later examples will show that the modified Gauss-Newton procedure works quite

TABLE 2
 SEQUENCE OF PARAMETER VALUES OBTAINED USING GAUSS-NEWTON
 ITERATION STABILIZED BY BINARY SCALE FACTOR ADJUSTMENT

Iteration Number	c_1	c_2	c_3	c_4	ϕ	N
0	1.000	2.000	3.0000	1.000	360.3	-
1	0.093	- 5.188	4.582	1.241	212.4	3
2	0.129	- 9.877	3.643	- 1.141	163.8	3
3	0.329	- 73.41	0.4556	- 0.1078	2.422	0
4	- 0.948	- 19.96	0.4733	- 0.1223	1.819	8
5	- 0.449	- 16.53	0.4784	- 0.1269	1.816	10
6	- 0.016	- 13.79	0.4834	- 0.1313	1.813	10
7	0.192	- 12.61	0.4858	- 0.1335	1.811	11
8	0.386	- 11.55	0.4882	- 0.1356	1.809	11
9	0.568	- 10.57	0.4905	- 0.1377	1.807	11
10	6.249	18.55	0.5641	- 0.2018	1.619	6
11	11.91	31.69	0.6667	- 0.3057	1.563	6
12	18.09	44.74	0.7209	- 0.3642	1.541	7
13	23.51	55.87	0.7490	- 0.3955	1.529	8
14	31.49	72.08	0.7776	- 0.4280	1.519	8
15	37.71	84.61	0.7922	- 0.4448	1.513	9
20	105.1	219.0	0.8517	- 0.5152	1.494	10
25	243.0	493.1	0.8747	- 0.5430	1.488	11
30	534.6	1072.	0.8853	- 0.5559	1.485	13
40	1772.	3531.	0.8909	- 0.5628	1.483	14
60	1213.	2237.	0.9330	- 0.6143	1.421	14
True Values	0.500	1.000	1.5690	0.0000	0.0000	-

well, even with noisy data, when more reasonable starting points are provided.

It was not determined whether or not the sequence of values listed in Table 2 would eventually converge to the true parameter values or to some secondary minimum in parameter space. Since the parameter estimation procedure was unconstrained in this experiment, the possibility even exists that the sequence in \vec{c} might grow without bound and fail to converge altogether even though ϕ were to approach a limiting value. The experiment was discontinued after 60 iterations because the cost of computer time seemed to exceed the value of the information which would probably be obtained from further computer runs.

The poor terminal behavior of the iteration procedure in this case is explained in part by a consideration of the angle existing between the Gauss-Newton parameter change vector, $\vec{\beta}$, and the gradient of ϕ . At iteration number 60, the values obtained for these two vectors were

$$\vec{\beta} = \begin{pmatrix} 1.108 \times 10^6 \\ 2.033 \times 10^6 \\ 4.027 \\ -5.074 \end{pmatrix} \quad \vec{\nabla}\phi = \begin{pmatrix} -1.159 \times 10^{-3} \\ 6.286 \times 10^{-4} \\ -1.849 \\ -2.389 \end{pmatrix}$$

The cosine of the angle existing between these vectors may be obtained by normalizing the scalar produce; i.e.

$$\cos \theta = \frac{\vec{\beta} \cdot \vec{\nabla}\phi}{|\vec{\beta}| |\vec{\nabla}\phi|} \approx -2.2 \times 10^{-7} \quad (145)$$

This result shows that the angle between the contour lines of ϕ and the $\vec{\beta}$ vector is only about 2×10^{-7} radians; i.e., $\vec{\beta}$ points "down-hill" only very slightly. As a consequence, ϕ is reduced only when $\vec{\beta}$ is diminished by a very large scale factor (2^{13} for the optimum step in this case).

6.3 Convergence with Measurement Noise.

To simulate measurement errors, ten uncorrelated samples of a unit variance, zero mean Gaussian random process were generated by the digital computer. These errors were then scaled to several different levels and added to the data points of Figure 3. The resulting noisy response data is plotted on this figure for a scale factor, σ , equal to one-tenth¹¹; Figure 3 also shows the response fitted to these data points by the computer.

Table 3 summarizes the results obtained by using different values of σ in this experiment. The variable M in the table stands for the number cycles of iteration required to satisfy the error criterion

$$\epsilon_{\phi} = \epsilon_c = 10^{-6}$$

The starting vector used in each case was as given by equation (144). The iteration converged without scale factor adjustment for values of σ less than $\sigma = .4$. For $\sigma = .4$ and $.5$, the modified iteration procedure converged but the computer sometimes found it necessary to

¹¹ Since the basic process sampled by the computer possesses a unit variance, the scale factor, σ , is just the standard deviation of the scaled process.

reduce $\bar{\beta}$ by a factor of two. The drastic scale factor adjustment encountered in the computation leading to Table 2 did not occur in this experiment since the starting parameter vector was reasonably close to the final minimizing vector. The apparently anomalous behavior of the variable M for $\sigma = .4$ and $.5$ is caused by the fact that the parameter adjusting steps are more effective when the search algorithm is evoked.

TABLE 3

PARAMETER ESTIMATES OBTAINED FROM NOISY DATA

σ	c_1	c_2	c_3	c_4	ϕ	M
.0	.50000021	1.0000002	1.5690004	.00000068	1.085×10^{-13}	2
.1	.4682	.9542	1.653	.06502	.0568	6
.2	.4502	.9265	1.782	.05652	.2290	11
.3	.4387	.9064	1.945	-.0092	.5180	23
.4	.4319	.8905	2.143	-.1371	.9240	11
.5	.4289	.8766	2.380	-.3369	1.4470	10
True Values	.50000000	1.00000000	1.56900000	.00000000	0	-

This experiment shows that very large errors may exist in data points without upsetting the convergence of the modified Gauss-Newton procedure when the iteration is started sufficiently close to the final value.

6.4 Effect of Length of Record on Estimation Accuracy

The linearized error theory developed in section 4.1 requires that the matrix of second partial derivatives, D , be computed. As has been pointed out, this is a formidable task in all but the very simplest of situations. However, as the previous experiment shows, the convergence of the unmodified Gauss-Newton procedure turns out to be quite rapid even in the presence of modest amounts of noise ($\sigma \leq .1$). This being the case, the results of Theorem 1 and 2 seem to indicate that the matrix D is probably quite insignificant in comparison to the regression matrix, S . When this is in fact true, the linearized estimate of the variability of parameter estimates is given by equation (111)

$$[\text{cov } \delta \vec{c}] = \sigma^2 S^{-1}$$

To check the applicability of equation (111), a total of thirty statistical experiments were conducted. A data vector was generated for each experiment by producing a Gaussian error vector from a random number generation routine and adding this to the standard correct solution to the system equation. The parameter vector used to generate the basic solution was

$$\vec{c}_0 = \begin{pmatrix} .05 \\ 1.00 \\ 1.00 \\ .0 \end{pmatrix} \quad (146)$$

The thirty computer runs were broken into three groups of ten each with the data points collected over 10 seconds of simulated operation in the first group, 30 seconds in the second group and 100 seconds in the third group. Samples were taken at one second intervals in every case and uncorrelated Gaussian noise with $\sigma = .1$ was added to each set of data points.

Table 4 summarizes the results of this experiment. The theoretical standard deviations appearing in this table were computed using the approximate formula given by equation (111). The experimental standard deviations represent the observed r.m.s. deviation of the sample values from the true value for each parameter.

TABLE 4
EFFECT OF LENGTH OF RECORD ON ESTIMATION ACCURACY

Parameter	Number of Response Samples	True Value	Average Experimental Value	Theoretical Standard Deviation	Experimental Standard Deviation
c_1	10	0.0500	0.0534	.0355	.0407
c_2	10	1.0000	1.0095	.0391	.0268
c_3	10	1.0000	1.0422	.1181	.1424
c_4	10	0.0000	0.0529	.1028	.1138
c_1	30	0.0500	0.0516	.00868	.00350
c_2	30	1.0000	0.9995	.01028	.00741
c_3	30	1.0000	1.0071	.0674	.0469
c_4	30	0.0000	0.0174	.0633	.0340
c_1	100	0.0500	0.0493	.00375	.00279
c_2	100	1.0000	0.9995	.00440	.00378
c_3	100	1.0000	0.9777	.0487	.0436
c_4	100	0.0000	0.0183	.0520	.0453

The agreement between the theoretical and observed quantities in this table appears to support the analysis carried out in this report. The Gauss-Newton iteration procedure converged without searching for every one of the thirty computer runs. The rms error in parameter estimation seems to decrease roughly in proportion to the square root of the number of samples, as predicted by theory. The experimental standard deviations are well predicted by equation (111) in every case. This result lends credence to the hypothesis that the D matrix may commonly be ignored in computing rough error estimates in situations where the unmodified Gauss-Newton procedure converges.

7. SUMMARY AND CONCLUSIONS

This study shows that the methods of linear regression analysis may be extended to cover the estimation of parameters for nonlinear dynamic systems. The minimization of the sum squared error criterion function may be accomplished by treating the nonlinear problem as a succession of linear problems. The local linearization required can be achieved by solving the auxiliary parameter influence differential equations. The necessary and sufficient conditions for the convergence of this process involve the eigenvalues of a particular matrix of first and second order partial derivatives of the nonlinear system response with respect to the unknown parameters.

A complete parameter estimation algorithm possessing universal stability may be constructed by adding an automatic scale factor

adjusting feature to the iterated linear estimation procedure. This algorithm produces not only parameter estimates, but also an approximate evaluation of the accuracy of the estimates. The results obtained by utilizing this method in experiments involving the response of a pendulum are entirely in agreement with theory and establish iterated linear regression analysis as a practical tool for the identification of non-linear dynamic systems.

Certain problems remain unresolved. While it always converges, the modified Gauss-Newton or iterated linear regression method is really effective only when initial parameter estimates are not grossly in error. Alternative procedures are needed for situations in which good initial estimates are not available. The problem of constrained parameter estimation has not been treated. The possible existence of multiple minima of the criterion function has likewise been ignored. All of these problems will be made the subject of further research and reports will be submitted whenever significant progress has been made.

REFERENCES

1. Fuchs, A. M., "A Bibliography of the Frequency Response Method as Applied to Automatic Feedback Control Systems", Trans. of the ASME, Vol. 76, pp. 1185-1194 (November, 1954).
2. Goodman, T. P., and Reswick, J. B., "Determination of System Characteristics from Normal Operating Records", Trans. of the ASME, Vol. 78, pp. 259-271 (February, 1956).
3. Kalman, R. E., "Design of a Self Optimizing Control System", Trans. of the ASME, Vol. 80, pp. 468-478 (February, 1958).

4. Levin, M. J., "Optimum Estimation of Impulse Response in the Presence of Noise", IRE Transactions on Circuit Theory, Vol. CT-7, No. 1, pp. 50-56 (March, 1960).
5. Zadeh, L. A., "On the Identification Problem", IRE Transactions on Circuit Theory, Vol. CT-3, No. 4, pp. 277-281 (December, 1956).
6. Wiener, N., Nonlinear Problems in Random Theory, John Wiley and Sons, Inc., New York, N. Y., (1958).
7. Bose, A. G., A Theory of Nonlinear Systems, TR 309, Massachusetts Institute of Technology, Cambridge, Mass. (May 15, 1956).
8. McGhee, R. B., "A Survey of Methods for Identifying Dynamic Systems", USCEC Report No. 97-101, Part I, University of Southern California, Los Angeles, California (February, 1963).
9. Graybill, F. A., An Introduction to Linear Statistical Models, McGraw-Hill Book Co., Inc., New York, N. Y., 1961, chapter 6.
10. Aaron, R. M., "The Use of Least Squares in System Design", IRE Transactions on Circuit Theory, Vol. CT-3, No. 4, pp. 224-231 (Dec. 1956).
11. Vajda, S., Mathematical Programming, Addison-Wesley Publishing Co., Reading, Massachusetts, (1961).
12. Bellman, R., Dynamic Programming, Princeton University Press, Princeton, N. J., (1957).
13. Williams, E. J., Regression Analysis, John Wiley and Sons, Inc., New York, N. Y., (1959).
14. Hartley, H. O., "The Modified Gauss-Newton Method for the Fitting of Nonlinear Regression Functions by Least Squares", Technometrics, Vol. 3, No. 2, pp. 269-280 (May, 1961).
15. Meissinger, H. F., "The Use of Parameter Influence Coefficients in Computer Analysis of Dynamic Systems", Proceedings of the Western Joint Computer Conference, 1960, pp. 181-192 (May, 1960).
16. Scheffe, H., Analysis of Variance, John Wiley and Sons, Inc., New York, N. Y., p. 12 (1959).
17. Kerr, R. B. and Surber, W. H., "Precision of Impulse - Response Identification Based on Short, Normal Operating Records", IRE Transactions on Automatic Control, Vol. AC-6, No. 2, pp. 173-182 (May, 1961).
18. Hoel, P. G., Introduction to Mathematical Statistics, John Wiley and Sons, Inc., New York, N. Y., (1962).