A SURVEY OF GENERALIZED LEAST SQUARES ESTIMATION AND ITS RELATION TO COMMUNICATION SYSTEM DESIGN

by

Altaf-ur-Rashid

B.S., University of Karachi, 1974

A MASTER'S REPORT

submitted in partial fulfillment of the

requirement for the degree

MASTER OF SCIENCE

Department of Electrical Engineering

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1976

Approved by:

Donald R. Hummels
Major Professor

TABLE OF CONTENTS

Chapter | Page

I. INTRODUCTION 2

Objective of the Report 2
Statement of the Problem 3
Organization of the Report 4

II. HISTORICAL BACKGROUND 7

III. AN OVERVIEW OF THE MAJOR REFERENCES AND SIGNIFICANT RESULTS 12

Summary of Major References 12
Classical Least Squares Method 14
Least Squares Estimate with a Priori Data 18
Swerling's Method 21
Battin's Method 22

IV. APPLICATION TO COMMUNICATION 27

Amplitude Modulation 34
Phase Modulation 41

V. CONCLUSIONS 46

BIBLIOGRAPHY 48

CHAPTER I

I. INTRODUCTION

The advent of the space age and satellite communication systems has placed increased emphasis on the study of estimation in communication systems. The subject of estimation is a vast one, so attention will be devoted only to the particular problem of least squares estimation. Emphasis will be placed on an approach which is actually quite similar to that originally proposed by Gauss. Least squares theory not only provides useful solutions to certain specific estimation problems, but it also has applications to a large number of other problems of both statistical and deterministic types. The deterministic form will be emphasized in this report. In many cases the statistical results are special cases of the results reviewed in this report.

A. OBJECTIVE OF THE REPORT

The object of this report is to collect some important results regarding least squares estimation and apply them to various communication Problems such as the derivation of optimum receiver structures. It is shown that the method of least squares and certain types of error theorems derived therefrom can be defined without using any statistical concepts. The report summarizes a number of substantive results which are insufficiently well known or in some cases about which there has been a great deal of confusion, e.g.

(1) Many important procedures such as recursive parameter estimation do not require any statistical concepts or assumptions.

(2) Even when a problem is statistically formulated there need be no difference in the treatment of the problem where the parameters are stochastic as opposed to where the parameters are non-stochastic or deterministic. All problems in which the parameters are stochastic processes can be considered a subset of the set of problems of estimating a vector of non-stochastic parameters giving identical results. This is a major argument of Swerling [1].

B. STATEMENT OF THE PROBLEM

The general estimation problem to be considered can be formulated as follows: Let

r be an observation vector.

x be the parameter vector for which an estimate $\hat{x}$ is to be made.

$s(x)$ be the noise free observation such that:

$$r = s(x) + n \qquad 1.1$$

where n is a vector representing the error in the observation. Throughout most of the report, vectors will be represented by lower case letters and capital letters will be used for matrices except for Q which is always a scalar. In cases where confusion is likely to occur between vectors and scalars, lower case greek letters are used for scalars.

If $s(x)$ depends linearly on the parameter x to be estimated, then equation 1.1 becomes the "case of linear dependance" or the "linear case." If x is regarded as a stochastic process, then it will have a prescribed covariance matrix C_x, but it can still be treated by the same methods as a non-stochastic process.

C. ORGANIZATION OF THE REPORT

The balance of the report is organized along the following lines. First the early history of least squares estimation from Gauss to N. Wiener (1795-1942) and more recent developments in the field of estimation by Blackman (1964) and P. Swerling (1971) are outlined in Chapter II.

Chapter III contains the overview and significant results obtained from Swerling [1],[3] and Blackman [2]. Throughout, the least squares estimate is defined as estimate derived by the method of "least squares" as given in section III-B. This should not be confused with the minimum mean square error (MMSE) estimate which is the one for which the statistical mean square error is minimum among all estimates of a given parameter. In many cases, however, they are the same.

In section III-C, least square estimates obtained with the benefit of a priori data are considered. The a priori data is taken as a set of a priori estimates x_a which could come from an estimate based on some previous set of observational data, could be the mean values of an a priori joint statistical distribution or any set of numbers which could be regarded as estimates of x_a.

Swerling's method of estimation using least squares is described in Section III-D. This recursive method avoids the computation of a large $D(x_o)$ matrix and uses updated estimates and covariance matrices.

The inversion of matrices can be avoided if the new data is introduced one at a time. This is described in Battin's method of estimation in section III-E. It is basically a special case of Swerling's method, the main feature being that it uses scalar observations to avoid need for matrix inversions.

Chapter IV considers the applications of least squares estimation to communication problems. Block diagrams are used to illustrate the results for amplitude and phase modulation schemes.

Chapter V concludes the report.

CHAPTER II

II. HISTORICAL BACKGROUND

Estimation can be defined as the process of making a decision or judgment, concerning the approximate value of a certain undefined variable when the decision is weighted or influenced by all available information. From the earliest times, people have been concerned with making estimates and predictions. An early force in the development of estimation theory was provided by astronomical studies in which planet and comet motions were studied. The method of least squares was invented to solve the problems concerning the revolution of heavenly bodies. It was separately and almost simultaneously, postulated by two men [5]. Legendre was the first to put forward his ideas in 1806. He was closely followed by Gauss who presented a paper in 1809, in which he derived the method from fundamental principles. Gauss also claimed that in 1795 he used the method of least squares to solve a problem concerning the orbit determination of minor planets.

Gauss showed the generality of the least squares estimator for treating scientific data. The following quotation [5] from Theoria Motus illustrates the understanding Gauss had of this method as well as its application beyond the field of dynamical astronomy. "The most probable value of the desired parameters will be that in which the sum of the squares of the difference between the actual observed and computed values multiplied by numbers that measure the degree of precision, is a minimum." It is interesting to note that Gauss tried to minimize other even powers [4th, 6th powers etc.] of the sum of the difference between observed data and the corresponding true values. Soon after the publications of "Theoria Motus" the method of least squares estimation was quickly adopted as a standard technique for the determination of orbit parameters. Many astronomers since Gauss have used the method of least squares for solving the problems relating to astronomical observations. The significant names are Bessel, Laplace, Poisson. More recently Swerling and Blackman have applied the method to satellite orbit determination.

Perhaps the next major development in estimation theory after least squares, was the introduction of "Method of moments," given by K. Pearson [5]. This method is not widely used, the main disadvantage being that the estimates obtained are not the best from the viewpoint of computational efficiency. The contributions of R. A. Fisher provide the basis for much of modern estimation theory. He showed that the method of maximum likelihood was usually superior to the method of moments. Gauss had felt that the maximum likelihood estimator would be inferior to least square estimation but in many cases they are the same.

A frequent problem in communication systems relates to the random noise signal added during transmission. Early attempts made to reduce the unwanted noise by means of filters were often unsatisfactory because of the lack of a theory that could be used to synthesize the required filters. An alternate approach to the study of information transmission in presence of noise is generally attributed to N. Wiener (1942). Two important contributions were made by Wiener were:

(a) Estimation theory could be applied to synthesize an electrical filter that would provide the best separation of a desired signal in presence of undesired noise.

(b) Wiener emphasized treating signals and noise as stochastic processes, rather than viewing them in terms of their frequency spectra. Wiener theory appears to be basically a least squares estimation process, but the similarities to Gauss's method soon disappear, because Wiener made use of data parameters input in the form of a stochastic process. He derived an optimum estimator which was capable of making the best separation between a desired signal and undesired noise. He required that the estimator be a linear physically realizable filter.

Since the middle fifties several methods of orbit refinement were developed for determining the orbits of artificial satellites and spacecraft. Until then, one of the most frequently used methods was the classical method, the so called "differential correction" method, (which is more or less the method of least squares developed by Gauss). However, this technique was not very helpful for large quantities of observational data. Till 1955, this method was good enough, but after the development of artificial satellites the need for alternative methods for orbit refinement began to be felt. The first definite proposal for an alternate method was given by P. Swerling [8] in 1958. A little different method developed by Blackman [2] in 1958 was used with Telestar I. A similar but improved version of this method was put forward by A. J. Claus and R. H. Battin [9] [10] in 1962. All the methods basically use the background of least squares.

It was not till 1966 that P. Swerling [3] gave a new application of the least squares method to the estimation of a signal or signal parameters, in the presence of noise. In 1971, another paper was presented by Swerling on the subject of state estimation. This paper was an elaboration of Gauss's method of least squares, which differs in viewpoint some what from the more conventional development, that is, the development from the viewpoint of Wiener's linear filter theory. It was shown by Swerling that the results obtained from linear filter theory are special cases of results that are obtained from a generalized least squares (gls) method.

CHAPTER III

III. AN OVERVIEW OF MAJOR REFERENCES AND SIGNIFICANT RESULTS

This chapter deals with an overview of major references and extracts the significant results from each. The definitions and the mathematical approach of the important methods are presented in the following sections.

A. SUMMARY OF MAJOR REFERENCES

The significant results of interest are obtained from three major references [1] [2] [3]. The references [1] [3] by P. Swerling deal with the general topic of estimation, whereas [2] describes the special methods developed for orbit refinement but which actually have wider application.

In his analysis Swerling uses a weighting matrix which may or may not be the inverse covariance matrix of the error (noise). The classical method described by Gauss had a diagonal weighting matrix. Swerling allowed the matrix to be non-daigonal and termed the resulting procedure a "generalized least squares" (gls) method. He shows [1] that if the weighting matrix equals the inverse covariance matrix of the error (noise), then the components of the estimates are minimum mean square error (MMSE) among all linear estimates of the unknown parameter also, Swerling further shows that if the errors have joint Gaussian statistics, the components of the estimate are then maximum likelihood (ML) estimates.

Swerling also demonstrates how a priori data can be used to improve gls estimates when available. The a priori data may be in the form of estimates based on a previous set of observations, the mean value of an a priori joint statistical distribution or any set of numbers which may be regarded as estimates. It is shown that the a priori data may be considered as additional observations which effect the estimate. If in this case the weighting matrix equals the inverse covariance matrix of the error, and in addition the a priori statistics of the observation error and the statistics of the actual error are jointly Gaussian, then the estimate obtained becomes the maximum a posteriori (MAP) estimate and is (MMSE) minimum mean square error among all estimates. It was also shown by Swerling [1] that the results obtained by linear filtering and prediction theory are similar to the results obtained by the generalized least squares method if the stochastic parameters used in linear filtering and prediction are used to form the weighting matrix for the gls approach.

The other major reference [2] describes several methods for orbit refinement, which were specifically developed for use with artificial satellites and spacecraft. In addition to these methods, the paper also describes the classical method. It also compares the relative advantages and disadvantages for practical system applications.

Battin's method is also considered which is essentially the improved version of Swerling's method. Battin's method uses scalar quantities to avoid the inversion of matrices and hence decreases the computational time.

Mathematical explanation for the methods described are given in the following sections.

B. CLASSICAL LEAST SQUARES METHOD

The "classical least squares" procedure for obtaining estimate $\hat{x}$ of the parameter x is as follows:

Let:

$$Q = [r - s(x)]^T \; C_n^{-1} \; [r - s(x)]. \qquad 3.1$$

where: $x^T = (x_1 . \; . \; . x_n)$.

C_n^{-1} is a positive definite symmetric matrix that is not necessarily the inverse covariance matrix of the observation error but which might be if some statistical information is available. Then by definition a procedure in which the estimate $\hat{x}$ is obtained by minimizing Q with respect to x is a least squares procedure. For the classical least squares method, the matrix C_n^{-1} is a diagonal matrix. For the generalized least squares procedure as defined by Swerling [1] a non-diagonal matrix is used. Except for this, the development is the same.

Let x_o be an initially assumed value of x, close to the true value. Then to the first order term in $(x-x_o)$, a Taylor series expansion of s(x) is given by:

$$s(x) = s(x_o) + \left[\frac{\partial s(x_o)}{\partial (x_o)}\right]^T (x-x_o) + \text{higher order terms.} \qquad 3.2$$

The derivative matrix is defined by:

$$\left[\frac{\partial s(x_o)}{\partial (x_o)}\right]^T = \left[\frac{\partial s(x)}{\partial (x)}\right]^T_{x=x_o} = D^T(x_o) \qquad 3.3$$

where:

$$D^T(x) = \begin{bmatrix} \frac{\partial s_1(x)}{\partial x_1} & \frac{\partial s_1(x)}{\partial x_2} & ------ & \frac{\partial s_1(x)}{\partial x_n} \\ \frac{\partial s_2(x)}{\partial x_1} & & & | \\ | & & & | \\ \frac{\partial s_m(x)}{\partial x_1} & & & \frac{\partial s_m(x)}{\partial x_n} \end{bmatrix} \qquad 3.4$$

Then:

$$s(x) \simeq s(x_o) + D^T(x_o)\ (x-x_o) \qquad 3.5$$

and Equation 3.1 becomes: 3.6

$$Q = [\ r - s(x_o) - D^T(x_o)\ (x-x_o)\ ]^T\ C_n^{-1}\ [r - s(x_o) - D^T(x_o)\ (x-x_o)]$$

Now observe that if a scalar Q is defined by a function of this form.

$$Q = f^T(x)\ A\ f(x) \qquad 3.7$$

where A is a symetric matrix, then

$$\frac{\partial Q}{\partial x} = \frac{\partial f^T(x)}{\partial x}\ A\ f(x) + f^T(x)\ A\ \frac{\partial f(x)}{\partial x} \qquad 3.8$$

but

$$g^T(x)\ [A \cdot f(x)] = [\{A\ f(x)\}^T\ g(x)]^T \qquad 3.9$$

$$= [f^T(x)\ A^T\ g(x)]^T \qquad 3.10$$

Because of the symmetric matrix A

$$g^T(x)\ [A\ f(x)] = f^T(x)\ A\ g(x) \qquad 3.11$$

Hence Equation 3.7 can be written as:

$$\frac{\partial Q}{\partial x} = 2\ \frac{\partial f^T(x)}{\partial x}\ A\ f(x) \qquad 3.12$$

Applying this result to Equation 3.6 yields

$$\frac{\partial Q}{\partial x} = -2\, D(x_o)\, C_n^{-1}\, [\, r - s(x_o) - D^T(x_o)\, (x-x_o)] \qquad 3.13$$

Now Q is minimum if:

$$0 = -2\, D(x_o)\, C_n^{-1}\, [r - s(x_o) - D^T(x_o)\, (x-x_o)] \qquad 3.14$$

$$0 = D(x_o)\, C_n^{-1}\, [r - s(x_o) - D^T(x_o)\, (x-x_o)] \qquad 3.15$$

$$D(x_o)\, C_n^{-1}\, [r - s(x_o)] = D(x_o)\, C_n^{-1}\, D^T(x_o)\, (x-x_o) \qquad 3.16$$

If $D^T(x_o)$ is non singular and if the value of x which satifies Equation 3.16 is $\hat{x}$ then:

$$\hat{x} = x_o + [D(x_o)\, C_n^{-1}\, D^T(x_o)]^{-1}\, D(x_o)\, C_n^{-1}\, [r - s(x_o)] \qquad 3.17$$

The procedure for computing estimates is to substitute the value of $\hat{x}$ back in Equation 3.17 for x_o in order to obtain another $\hat{x}$. This process is iterated until $\hat{x}$ has essentially converged. The final value of $\hat{x}$ is the least squares estimate of the parameter vector x. Equation 3.17 can be further written in a more compact form as:

$$\hat{x} = x_o + B^{-1}(x_o)P \qquad 3.18$$

where

$$B(x_o) = [\, D(x_o)\quad C_n^{-1}\quad D^T(x_o)\,] \qquad 3.19$$

$$P = [\, D(x_o)\quad C_n^{-1}]\, [\, r - s(x_o)\,] \qquad 3.20$$

In linear case this can be simplified somewhat because $s(x) = D^T(x)$.

In the linear case 3.17 is exact and B and D are independent of x, a simpler formula for calculating $\hat{x}$ is obtained as follows:

$$\hat{x} = B^{-1}\, D\, C_n^{-1}\, [r - D^T x_o\,] + x_o \qquad 3.21$$

$$\hat{x} = B^{-1} D C_n^{-1} D^T x_o + B^{-1} D C_n^{-1} [r - D^T x_o] \qquad 3.22$$

$$\hat{x} = B^{-1} D C_n^{-1} D^T x_o + B^{-1} D C_n^{-1} r - B^{-1} D C_n^{-1} D^T x_o \qquad 3.23$$

$$\hat{x} = B^{-1} D C_n^{-1} r. \qquad 3.24$$

The above equation 3.24 is true whether or not errors are regarded as statistical variables. In this case it is not necessary to assume the initial value x_o or iterate the solution to find $\hat{x}$.

The average or mean of the error is found by rearranging 3.24 and forming:

$$E[\hat{x}-x] = E[B^{-1} D C_n^{-1} r - x] \qquad 3.25$$

$$= E[B^{-1} D C_n^{-1} (D^T x + n) - x]$$

$$= E[B^{-1} D C_n^{-1} \cdot n] \qquad 3.26$$

$$= E[n] \cdot B^{-1} D C_n^{-1} \qquad 3.27$$

$$= 0 \qquad 3.28$$

Since n is a zero mean. The covariance matrix of the error can be found from

$$E[(\hat{x}-x)(\hat{x}-x)^T] = E[(B^{-1} D^T C_n^{-1} n)(B^{-1} D^T C_n^{-1} n)^T] \qquad 3.29$$

$$= B^{-1} D^T C_n^{-1} E[n\, n^T] [B^{-1} D^T C_n^{-1}]^T$$

Since $C_n = E[n\, n^T]$ 3.30

Therefore

$$E[(\hat{x}-x)(\hat{x}-x)^T] = B^{-1} D^T C_n^{-1} C_n B^{-1} D C_n^{-1} \qquad 3.31$$

$$= B^{-1} (D^T C_n^{-1} D) C_n B^{-1} C_n^{-1}$$

$$= B^{-1}$$

Therefore, for the linear case:

$$E\,[(\hat{x}-x)\;(\hat{x}-x)^T] = B^{-1} \qquad 3.32$$

The matrix B^{-1} is the covariance matrix of the estimation error. The computation of $\hat{x}$ are minimum mean square error among all linear estimates of x [1]. In practice, C_n^{-1} may or may not be the inverse covariance matrix. The actual choice of C_n^{-1} is just a trade off between the estimation accuracy, computational speed and simplicity [1].

C. LEAST SQUARES ESTIMATE WITH A PRIORI DATA

In this section generalized least squares estimation is extended to incorporate an a priori estimate in addition to the observational data.

Here it is supposed that in addition to the observational data r there are also available a set of a priori estimates x_a. These a priori data might arise in several different ways:

(1) Estimates based on some previous set of observational data.

(2) Mean value of an a priori joint statistical distribution.

(3) A set of numbers which may be regarded as estimates of x_a.

The generalized least squares estimation procedure which consists of observational data r and an a priori estimates x_a can be defined by the quadratic equation given below:

$$Q = [r - s(x)]^T \; C_n^{-1} \; [r - s(x)] + [x_a - x]^T \; C_{x_a}^{-1} \; [x_a - x] \qquad 3.33$$

This equation is minimized with respect to the components of x in order to get the estimate. In the above equation:

C_n^{-1} is a positive definite symmetric matrix which may not be the inverse covariance matrix of the observation error (noise).

$C_{x_a}^{-1}$ is a positive definite symmetric matrix for the a priori data.

Therefore, as before $s(x) \approx s(x_o) + D^T(x_o)\ (x-x_o)$.

$$Q = [r - s(x_o) - D^T(x_o)\ (x-x_o)]^T\ C_n^{-1}\ [r - s(x_o) - D^T(x_o)\ (x-x_0)] + [x_a-x]^T\ C_{x_a}^{-1}\ [x_a-x] \qquad 3.34$$

$$\frac{dQ}{dx} = -2\ D\ (x_o)\ C_n^{-1}\ [r - s(x_o) - D^T(x_o)\ (x-x_o)] + (-2)\ C_{x_a}^{-1}\ (x_a-x)$$

If $\hat{x}$ is a solution the above equation then:

$$C_n^{-1}\ D\ (x_o)\ [r - s(x_o) - D^T(x_o)\ (\hat{x}-x_o)] + C_{x_a}^{-1}\ (x_a-\hat{x}) = 0.$$

$$D(x_o)\ C_n^{-1}\ [r - s(x_o)] - D(x_o)\ C_n^{-1}\ D^T(x_o)\ (\hat{x}-x_o) + C_{x_a}^{-1}\ (x_a-\hat{x}) = 0.$$

$$[r - s(x_o)]\ C_n^{-1}\ D\ (x_o) + D(x_o)\ C_n^{-1}\ D^T(x_o)\ (x_o) + C_{x_a}^{-1}\ (x_a) = D(x_o)\ C_n^{-1} \cdot \hat{x}\ \ D^T(x_o) + C_{x_a}^{-1}\ \hat{x}$$

Solving for x yields:

$$\hat{x} = [D^T(x_o)\ C_n^{-1}\ D\ (x_o) + C_{x_a}^{-1}]^{-1}\ \ [C_n^{-1}\ D\ (x_o)\ [r - s(x_o)] + D(x_o)\ C_n^{-1}\ \ D^T(x_o)\ \ (x_o) + C_{x_a}^{-1}\ \ x_a]$$

If $\quad C = [D^T(x_o)\ C_n^{-1}\ D\ (x_o) + C_{x_a}^{-1}]^{-1} \qquad 3.35$

$$\hat{x} = C\ [D(x_o)\ C_n^{-1}\ \{r - s(x_o)\} + D(x_o)\ C_n^{-1}\ \ D^T(x_o)\ (x_o) + C_{x_a}^{-1}\ \ x_a] \qquad 3.36$$

This value of $\hat{x}$ is substituted back in equation 3.34 to obtain another estimate. The process is iterated until $\hat{x}$ is essentially converged. The final value of $\hat{x}$ is the generalized least squares estimate of the parameter.

In general it is logical to set $x_o = x_a$ i.e. the initial guess is taken as the a priori data estimate. This leads to:

$$Q = [r - s(x_o) - D^T(x_a)\ (x-x_a)]^T\ C_n^{-1}[r - s(x_o) - D^T(x_a)\ (x-x_a)] + [x_a-x]^T\ D_{x_a}^{-1}\ [x_a-x] \qquad 3.37$$

Letting $x_o = x_a$ in equation 3.36 the estimate $\hat{x}$ obtained is given as:

$$\hat{x} = C[D(x_a)\ C_n^{-1}\ \{r - s(x_a)\} + D(x_a)\ C_n^{-1} \cdot D^T(x_a)\ (x_a) + C_{x_a}^{-1}\ x_a] \qquad 3.38$$

$$\hat{x} = C[D(x_a)\ C_n^{-1}\ \{r - s(x_a)\}] + C[D(x_a)\ D^T(x_a)\ C_n^{-1} x_a + C_{x_a}^{-1}\ x_a]$$

$$\hat{x} = C[D(x_a)\ C_n^{-1}\ \{r - s(x_a)\}] + C\quad C^{-1}\quad x_a$$

$$\hat{x} = x_a + C[D(x_a)\ C_n^{-1}\ \{r - s(x_a)\}] \qquad 3.39$$

It is clear that the a priori estimate x_a effects the least square estimate in exactly the same way as a set of n additional observations. It follows that the following statements hold good for the n additional observations.

1. The observed values are x_a.
2. The observation errors are $x_a - x$.

The general equation $r = s(x) + n$ equals

$$x_a = x + (x_a - x) \qquad 3.40$$

$(x_a - x)$ are considered to have a priori statistics with zero mean and are to be statistically un-correlated with actual observation errors. In addition to the estimate obtained in equation 3.36 the a priori statistics of $(x - x_a)$ and the statistics of error are jointly Gaussain, then $\hat{x}$ becomes the maximum a posteriori (MAP) estimate and minimum mean square error (MMSE) among all the estimates.

D. SWERLING'S METHOD OF LEAST SQUARES

The main feature of Swerling's method is that computation involving lar matrices can be avoided by calculating a sequence of estimates. The procedure is as follows.

Let $\hat{x}$ be an estimate of the element at time t_1 and let $C_{\hat{x}}$ be its covariance matrix. Also let r be the new vector observation at time t_2 having covariance matrix C_n. To obtain the least squares estimate of the elements at time t_2, the estimate $\hat{x}$ and its covariance matrix $C_{\hat{x}}$ are first updated, i.e., extrapolated to time t_2. If $\hat{x}_1$ is the result of updating $\hat{x}$, the updated matrix $C_{\hat{x}_1}$ is given by:

$$C_{\hat{x}_1} = MC_{\hat{x}}M^T \qquad 3.41$$

where

$$M = \frac{\partial \hat{x}_1}{\partial \hat{x}} \qquad 3.42$$

Then assuming that the errors in the new data are not correlated with the errors in the old data, the quadratic form to be minimized is:

$$Q = (x-\hat{x}_1)^T C_{\hat{x}_1}^{-1} (x-\hat{x}_1) + [r - s(x)]^T C_n^{-1} [r - s(x)] \qquad 3.43$$

where

$$s(x) = s(\hat{x}_1) + D^T(\hat{x}_1) (x-\hat{x}_1) \qquad 3.44$$

and

$$D^T(\hat{x}_1) = \frac{\partial s(\hat{x}_1)}{\partial \hat{x}_1} \qquad 3.45$$

Then to second order terms Q is given by

$$Q = (x-\hat{x}_1) C_{\hat{x}_1}^{-1} (x-\hat{x}_1) + [r - s(\hat{x}_1) - D^T(\hat{x}_1) (x-\hat{x}_1)]^T \cdot$$

$$C_n^{-1} [r - s(\hat{x}_1) - D^T(\hat{x}_1) (x-\hat{x}_1)] \qquad 3.46$$

The minimization is the same as in case of classical method with a priori data, hence the result can be written directly.

If $\hat{x}$ satisfies the above equation then

$$\hat{x} = \hat{x}_1 + C_{\hat{x}} \gamma \qquad 3.47$$

where

$$C_{\hat{x}} = [C_{\hat{x}_1}^{-1} + D(\hat{x}_1) C_n^{-1} D^T(\hat{x}_1)]^{-1} \qquad 3.48$$

$$\gamma = C_n^{-1} D (\hat{x}_1) [\gamma - s(\hat{x}_1)] \qquad 3.49$$

Equations 3.41 and 3.42, together with 3.47, 3.48 and 3.49 constitute Swerling's procedure for calculating an up-dated estimate $\hat{x}$ using a previous estimate as a starting point.

E. BATTIN'S METHOD FOR SCALAR OBSERVATIONS

This method is basically a special case of Swerling's method. Its main feature is that, by introducing new data one at a time it avoids the inversion of matrices and saves computation time.

The estimate obtained from Swerling's method is:

$$\hat{x} = \hat{x}_1 + C_{\hat{x}} \quad D(\hat{x}_1) C_n^{-1} [r - s(\hat{x}_1)] \tag{3.51}$$

Let the observation r be a scalar $r = \ell$. The vector $s(\hat{x}_1)$ also becomes a scalar and the matrix $D(\hat{x}_1)$ becomes a vector and is written as $d(\hat{x}_1)$. Since ℓ is a scalar, the matrix C_n is also a scalar and equation 3.51 can be further reduced to

$$\hat{x} = \hat{x}_1 + C_{\hat{x}} \, d(\hat{x}_1) \frac{1}{\sigma_n^2} [\ell - s(\hat{x}_1)] \tag{3.52}$$

where

$$C_n = \sigma_n^2 \tag{3.53}$$

and

$$C_{\hat{x}} = [d(\hat{x}_1) \frac{1}{\sigma_n^2} d^T(\hat{x}_1) + C_{\hat{x}_1}^{-1}]^{-1} \tag{3.54}$$

Substituting

$$\hat{x} = \hat{x}_1 + \frac{1}{\sigma_n^2} [\frac{1}{\sigma_n^2} d(\hat{x}_1) \, d^T(\hat{x}_1) + C_{\hat{x}_1}^{-1}]^{-1} \, d(\hat{x}_1) \quad [\ell - s(\hat{x}_1)]$$

Rearranging yields

$$\hat{x} = \hat{x}_1 + [d(\hat{x}_1) \, d^T(\hat{x}_1) + \sigma_n^2 \, C_{\hat{x}_1}^{-1}]^{-1} \, d(\hat{x}_1) \, [\ell - s(\hat{x}_1)] \tag{3.55}$$

The troublesome aspect of using this result for calculating $\hat{x}$ is the inverse operation. Suppose that a scalar α is defined by

$$\alpha = d^T(\hat{x}_1) \, C_{\hat{x}_1} \, d(\hat{x}_1) + \sigma_n^2 \tag{3.56}$$

Since α is a scalar, it can be written as

$$d(\hat{x}_1) = d(\hat{x}_1) \, [d^T(\hat{x}_1) \, C_{\hat{x}_1} \, d(\hat{x}_1) + \sigma_n^2] \, \frac{1}{\alpha}$$

Rearranging

$$d(\hat{x}_1) = \frac{1}{\alpha} [d(\hat{x}_1) \, d^T(\hat{x}_1) \, C_{\hat{x}_1} \, d(\hat{x}_1) + \sigma_n^2 \, C_{\hat{x}_1}^{-1} \, C_{\hat{x}_1} \, d(\hat{x}_1)]$$

Finally

$$d(\hat{x}_1) = \frac{1}{\alpha}\,[d(\hat{x}_1)\; d^T(\hat{x}_1) + \sigma_n^{\;2}\; C_{\hat{x}_1}^{\;-1}]\; C_{\hat{x}_1}\; d(\hat{x}_1) \qquad 3.57$$

Substituting Equation 3.57 in 3.55 results

$$\hat{x} = \hat{x}_1 + [d(\hat{x}_1)\; d^T(\hat{x}_1) + \sigma_n^{\;2}\; C_{\hat{x}_1}^{\;-1}]^{-1}\, \frac{1}{\alpha}\,[d(\hat{x}_1)\; d^T(\hat{x}_1) + \sigma_n^{\;2}\; C_{\hat{x}_1}^{\;-1}]$$

$$C_{\hat{x}_1}\; d(\hat{x}_1) \quad [\ell - s(\hat{x}_1)] \qquad 3.58$$

Making the obvious cancelations leads to

$$\hat{x} = \hat{x}_1 + \frac{1}{\alpha}\, C_{\hat{x}_1}\; d(\hat{x}_1)\; [\ell - s(\hat{x}_1)] \qquad 3.59$$

where $C_{\hat{x}_1}$ is an update of the matrix $C_{\hat{x}}$. Since Equation 3.54 can be written as

$$C_{\hat{x}} = \sigma_n^{\;2}\; [d(\hat{x}_1)\; d^T(\hat{x}_1) + \sigma_n^{\;2}\; C_{\hat{x}_1}^{\;-1}]^{-1} \qquad 3.60$$

Following the steps given by Blackman [2], this can be manipulated to eliminate the implied inversion operation of the matrices. Subtracting $C_{\hat{x}_1}$ from both sides of Equation 3.60 gives

$$C_{\hat{x}} - C_{\hat{x}_1} = \sigma_n^{\;2}\; [d(\hat{x}_1)\; d^T(\hat{x}_1) + \sigma_n^{\;2}\; C_{\hat{x}_1}^{\;-1}]^{-1} - C_{\hat{x}_1} \qquad 3.61$$

Multiplying both sides to clear the inverse yields

$$[d(\hat{x}_1)\; d^T(\hat{x}_1) + \sigma_n^{\;2}\; C_{\hat{x}_1}^{\;-1}]\; (C_{\hat{x}} - C_{\hat{x}_1}) = \sigma_n^{\;2} - C_{\hat{x}_1} \qquad 3.62$$

$$[d(\hat{x}_1)\; d^T(\hat{x}_1) + \sigma_n^{\;2} C_{\hat{x}_1}^{\;-1}]$$

Replacing $d(\hat{x}_1)$ by Equation 3.57 on the right hand side of Equation 3.62 gives

$$C_{\hat{x}} = C_{\hat{x}_1} - \frac{1}{\alpha}\, C_{\hat{x}_1}\; d(\hat{x}_1)\; d^T(\hat{x}_1)\; C_{\hat{x}_1} \qquad 3.63$$

Summarizing:

$$\hat{x} = \hat{x}_1 + \frac{1}{\alpha}\, C_{\hat{x}_1}\; d(\hat{x}_1)\; [\,\ell - s(\hat{x}_1)]$$

$$\alpha = d^T(\hat{x}_1)\, C_{\hat{x}_1}\, d(\hat{x}_1) + \sigma_n^{\,2}$$

$$C_{\hat{x}} = C_{\hat{x}_1} - \frac{1}{\alpha} C_{\hat{x}_1}\, d(\hat{x}_1)\, d^T(\hat{x}_1)\, C_{\hat{x}_1}$$

$$C_{\hat{x}_1} = M\, C_{\hat{x}}\, M^T$$

$$M = \frac{\partial \hat{x}_1}{\partial \hat{x}}$$

$$d(x) = \frac{\partial s(x)}{\partial x}$$

The preceding defines a procedure for calculating an estimate $\hat{x}$ without the need for matrix inversion. It uses scalar quantities and because there are no inversions the computational speed is faster than Swerling's Method.

CHAPTER IV

IV. THE COMMUNICATION APPLICATIONS

A problem of interest to communication theorists is that of finding an optimum receiver for various modulation schemes. In this chapter a general method is first considered, then it is applied to double side band amplitude modulation (DSB-AM) and phase modulation (PM) schemes.

A. GENERALIZED LEAST SQUARE CRITERION FOR THE CONTINUOUS CASE

The block diagram for the model assumed is shown in Figure 1.

Let

$r(t)$ be the observed signal.

$n(t)$ be the error (noise) in the observation.

$s[t,x(t)]$ be the transmitted signal; a no memory function of the message $x(t)$.

The output of the channel and receiver input is given by

$$r(t) = s\{t,x(t)\} + n(t) \qquad 4.1$$

In the search for optimum receivers, the term optimum will be taken to mean the receiver with output $\hat{x}(t)$ where $\hat{x}(t)$ is the value of $x(t)$ which minimizes the generalized least squares criterion.

$$Q = \int_o^t [r(\alpha) - s\{\alpha,x(\alpha)\}]^2 \; q_n(t,\alpha)\, d\alpha + \int_o^t \{x(\alpha) - x_a(\alpha)\}^2 \cdot q_x(t,\alpha) d\alpha \qquad 4.2$$

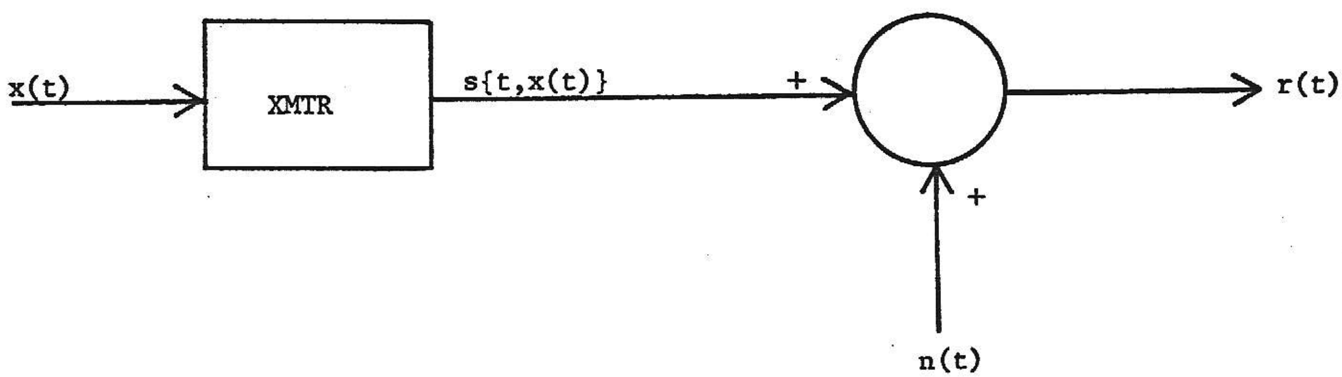

Figure 1. Block Diagram of the Assumed Model

$q_n(t,\alpha)$ and $q_x(t,\alpha)$ are taken as arbitrary weighting functions which may or may not be the inverse covariance matrices of the error $n(t)$ and message $x(t)$ respectively. Differentiating and setting the result equal to zero.

$$\left.\frac{dQ}{dx}\right|_{x=\hat{x}} = \int_0^t 2[r(\alpha) - s\{\alpha,\hat{x}(\alpha)\}] \left[- \frac{\partial s\{\alpha,\hat{x}(\alpha)\}}{\partial d\hat{x}}\right] q_n(t,\alpha)\, d\alpha + \int_0^t 2[\hat{x}(\alpha) - x_a(\alpha)]\ (1)\ q_x(t,\alpha)\, d\alpha = 0$$

A more convenient form is

$$\int_0^t [r(\alpha) - s\{\alpha,\hat{x}(\alpha)\}] \left[\frac{\partial s\{\alpha,\hat{x}(\alpha)\}}{\partial \hat{x}}\right] q_n(t,\alpha)\, d\alpha = \int_0^t [\hat{x}(\alpha) - x_a(\alpha)]\, q_x(t,\alpha)\, d\alpha \qquad 4.3$$

Block diagrams representing Equation 4.3 are shown in Figure 2 and Figure 3, where the integrals have been interpreted as linear filtering operations with time varying impulse responses $q_x(t,\alpha)$ and $q_n(t,\alpha)$. Assuming that these filters have inverses e.g. $q_x^{-1}(t,\alpha)$ for $q_x(t,\alpha)$ as in Figure 4 and $q_n^{-1}(t,\alpha)$ for the filter $q_n(t,\alpha)$. The two systems of Figure 2 and Figure 4 can be combined as shown in Figure 5 which is a complete block diagram for the system. Applications to amplitude and phase modulation are considered next.

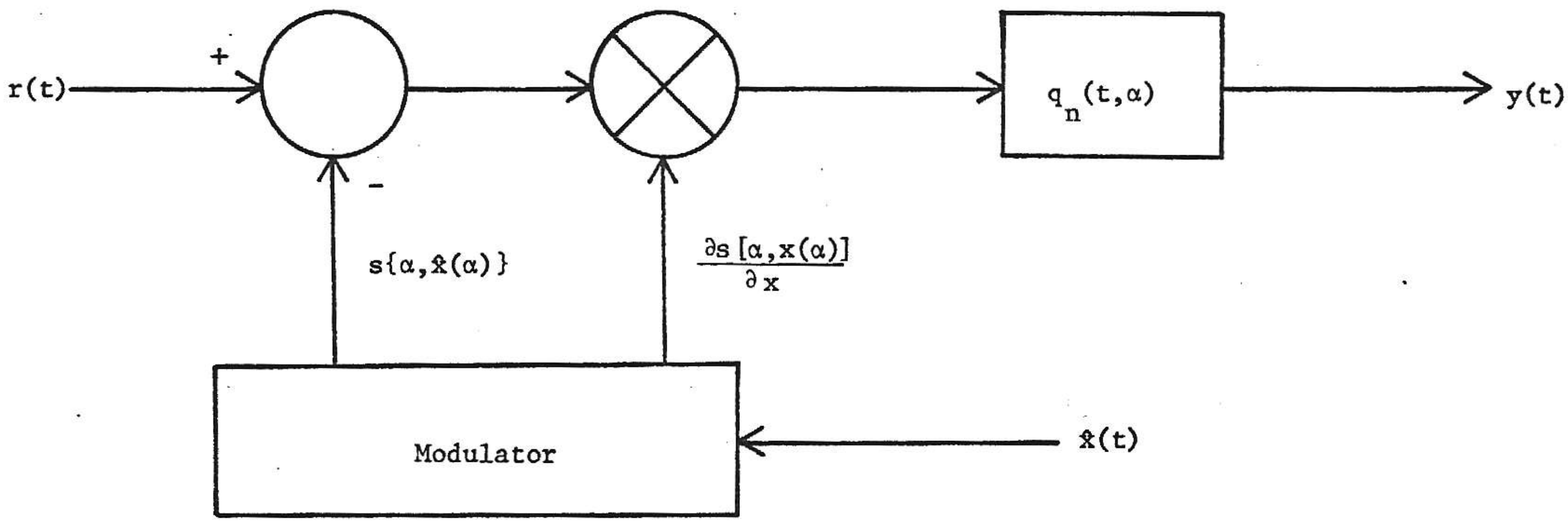

Figure 2. Representing Left Side of Equation 4.3

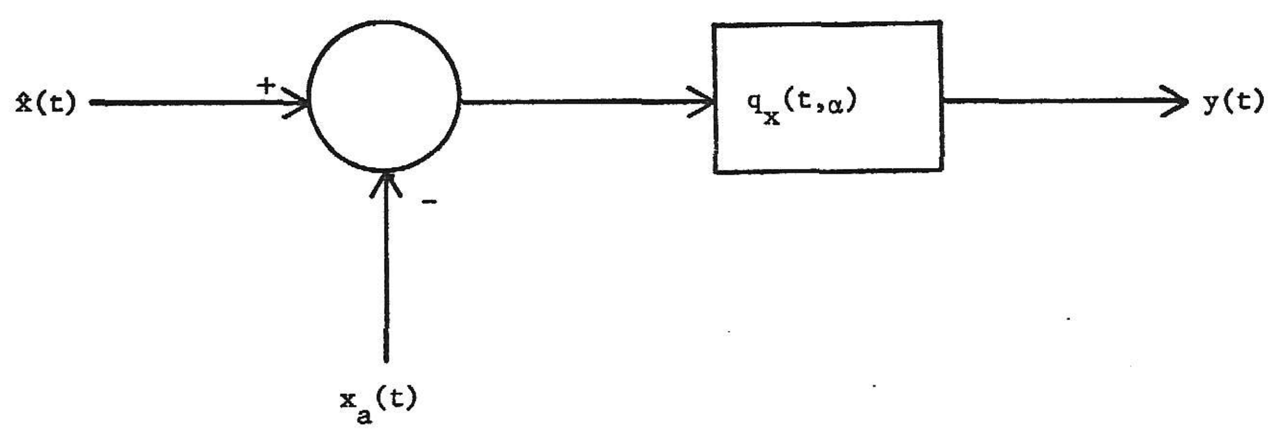

Figure 3. Representing Right Side of Equation 4.3

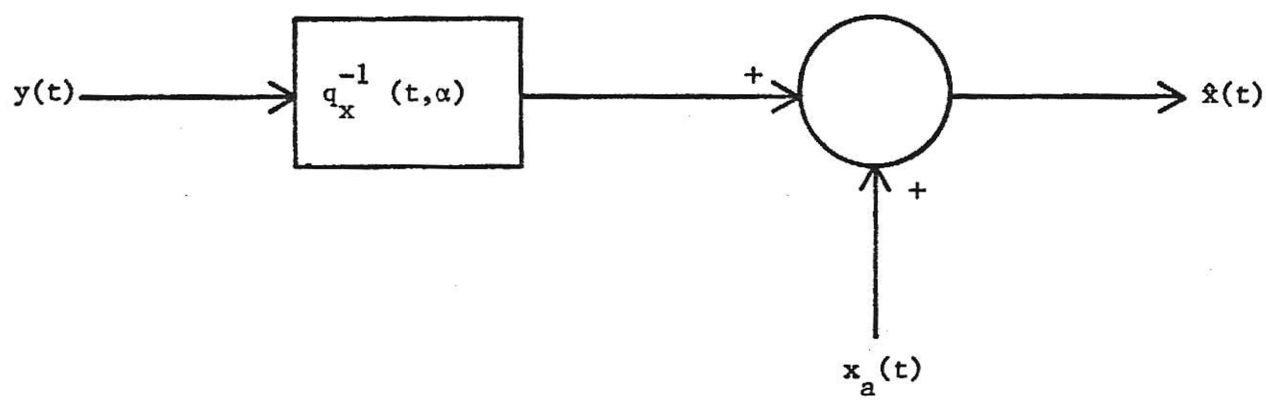

Figure 4. Block Diagram Showing Inverse of $q_x(t,\alpha)$

ILLEGIBLE

THE FOLLOWING DOCUMENT (S) IS ILLEGIBLE DUE TO THE PRINTING ON THE ORIGINAL BEING CUT OFF

ILLEGIBLE

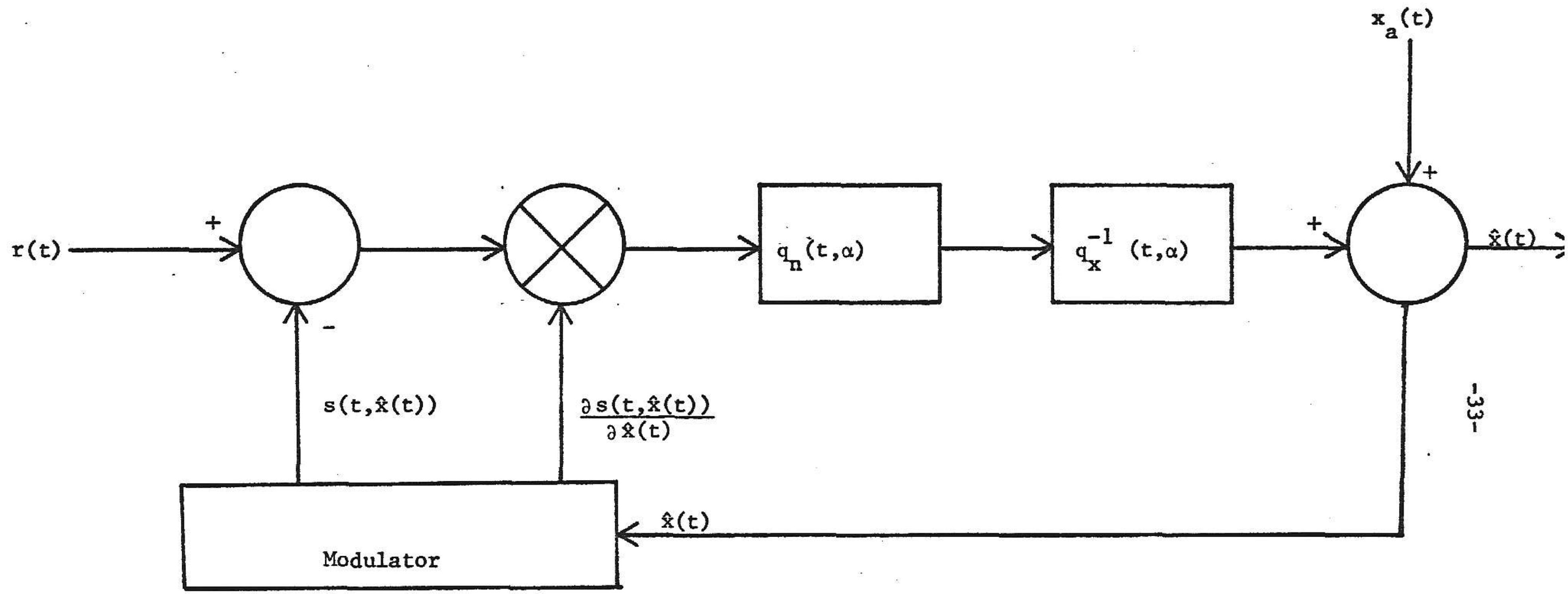

Figure 5. Complete System

B. AMPLITUDE MODULATION

One of the most important and widely used schemes of linear modulation is amplitude modulation (AM). One common form, double side-band AM, is mathematically represented as

$$s\,[t,x(t)] \;=\; A\;x(t)\;\mathrm{Cos}\;w_o t \qquad 4.3$$

In this case,

$$\frac{\partial s\,[t,x(t)]}{\partial x(t)} \;=\; A\;\mathrm{COs}\;w_o t$$

Assuming that the message x(t) is a zero mean signal so that the a priori estimate x_a can be set to zero, then Figure 6 shows the block diagram for obtaining estimate of the signal x(t) when using amplitude modulation. In Figure 6 the terms r(t) and A $\hat{x}(t)$ Cox $w_o t$ do not contribute anything to the estimate, hence this output of the balanced modulator is removed as shown in the improved version of Figure 7. Also the output from the two multipliers of Figure 7 can be written as $\frac{A^2}{2} + \frac{A^2}{2}$ Cos $2w_o t$. Only the low pass term $\frac{A^2}{2}$ will produce an output and hence the other double frequency term can be discarded. These changes are shown in Figure 8.

Assuming that the weighting functions are dependent only on the time difference so that $q_n(t,\alpha) = q_n(t-\alpha)$, the filters will no longer be time varying. Then a frequency domain analysis is useful. In Figure 9, the filters are represented by their transfer functions $H_x(f)$ and $H_n(f)$ respectively, where $H_n(f)$ and $H_x(f)$ are given by

$$H_n(f) \;=\; \int_{-\infty}^{t} q_n(\zeta)\; e^{-j2\pi f\zeta}\; d\zeta$$

$$H_x(f) \;=\; \int_{-\infty}^{\infty} q_x^{-1}\;(\zeta)\; e^{-j2\pi f\zeta}\; d\zeta$$

A final block diagram for amplitude modulation scheme is shown in Figure 10 where all filters have been combined.

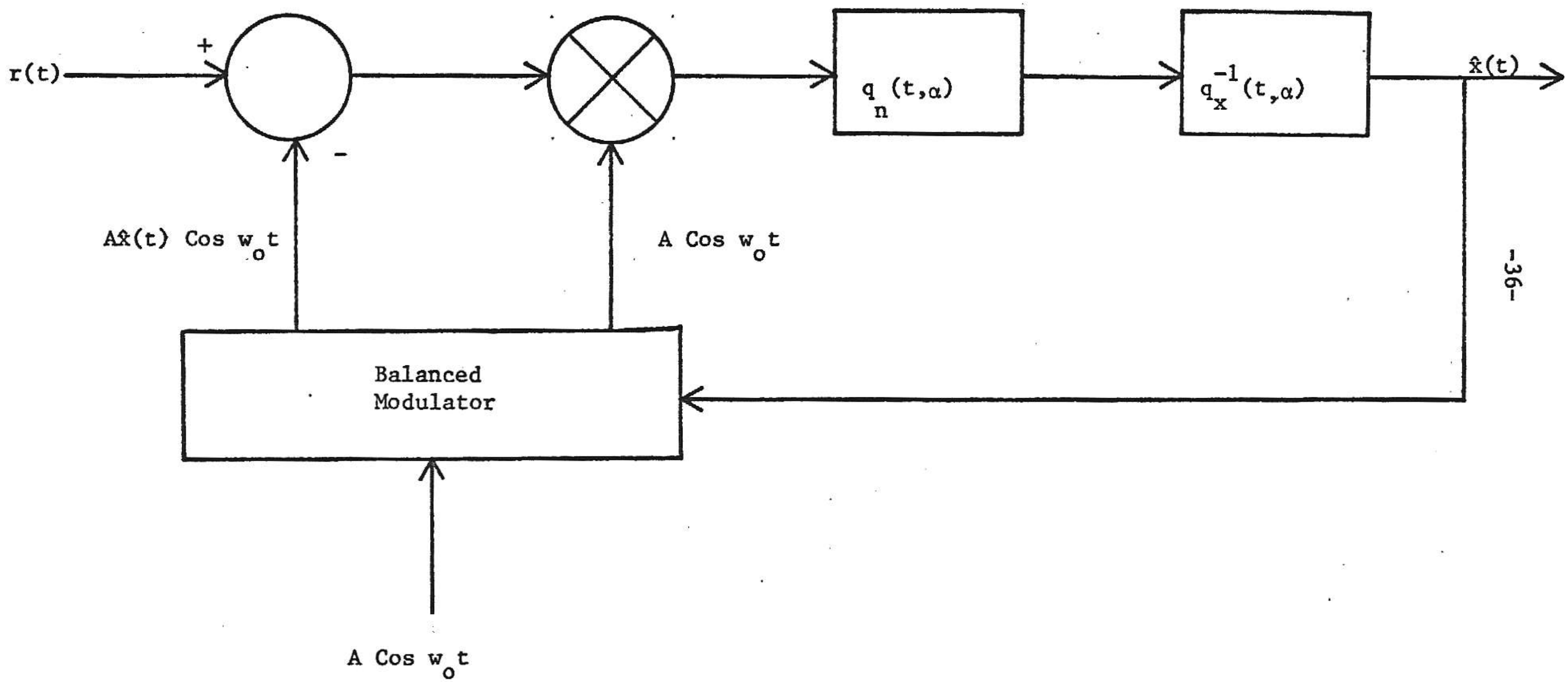

Figure 6. Physical Model for DSB-AM

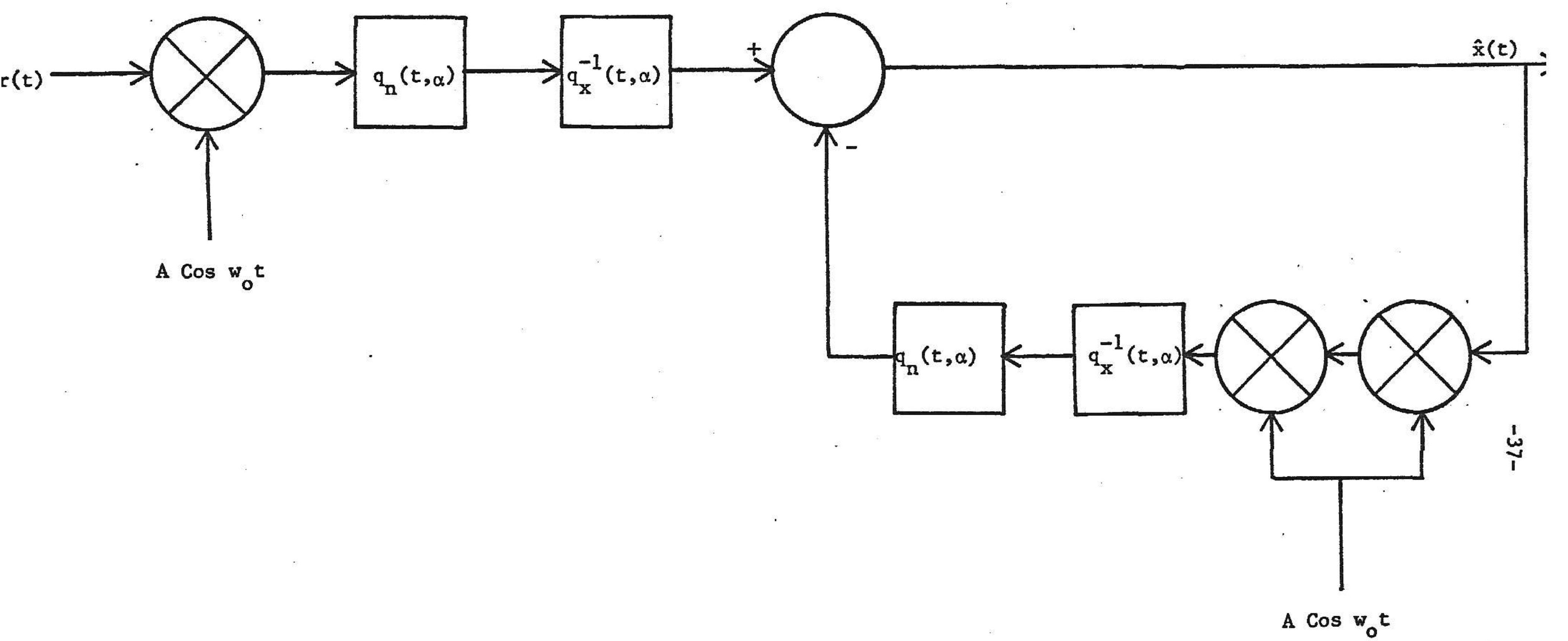

Figure 7. Modified Version of Figure 6

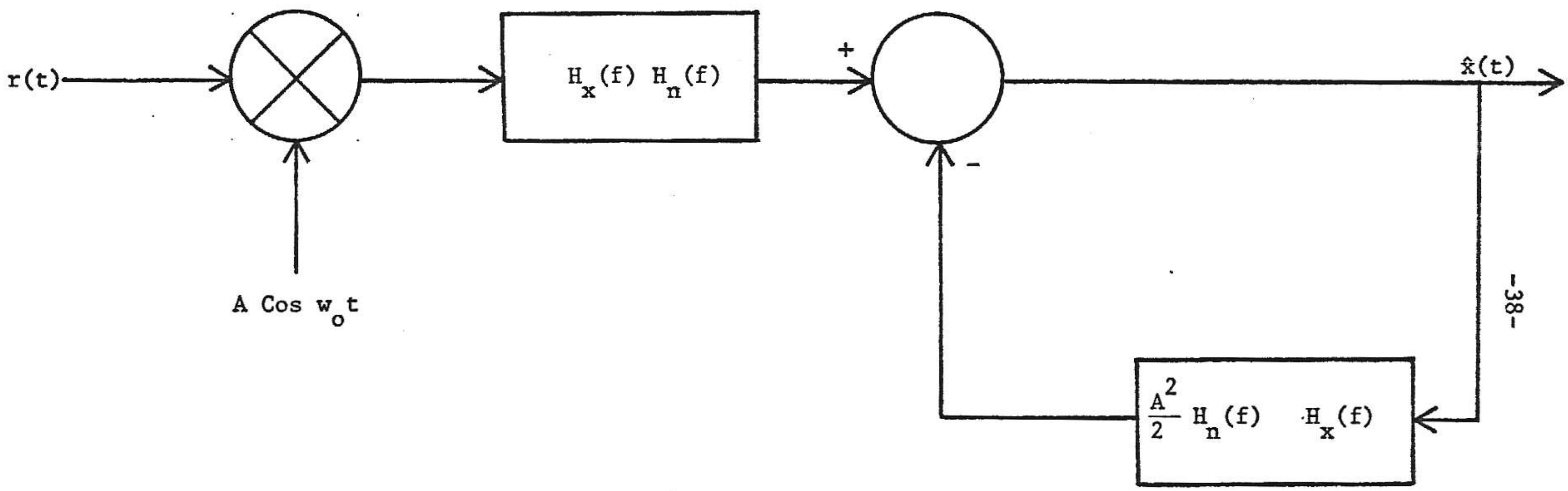

Figure 8. Simplified Version of Figure 7

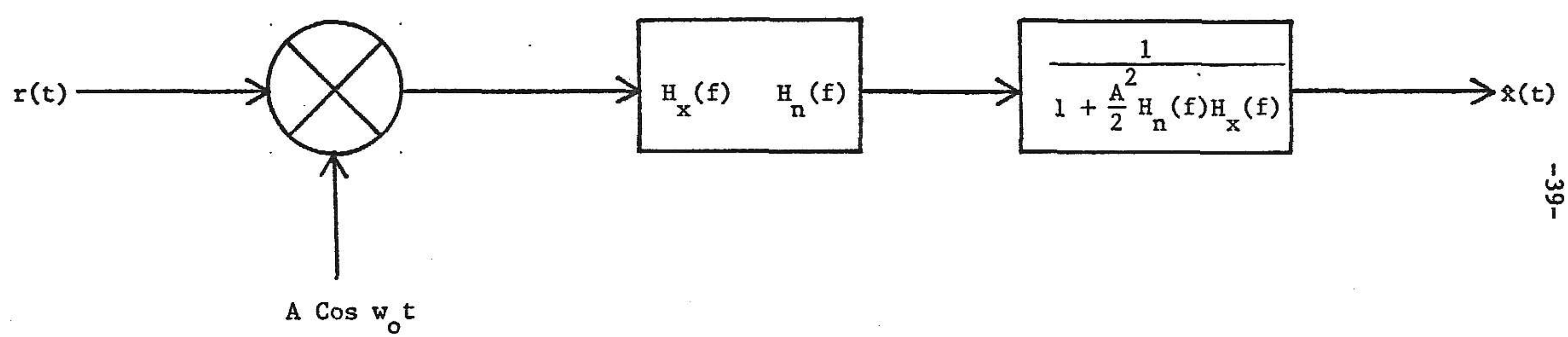

Figure 9. Further Simplification to Figure 8

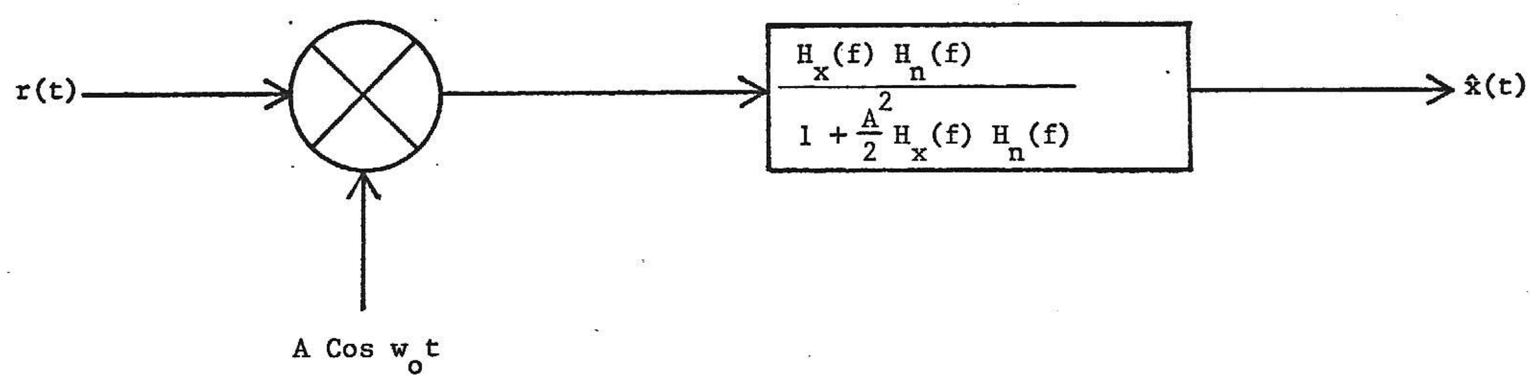

Figure 10. Final Block Diagram

C. PHASE MODULATION

Phase modulation (PM) is another scheme widely used by communication theorists. Phase modulation is an example of angle modulation which is very similar to frequency modulation. Let

$$s[t\ x(t)] = A \operatorname{Cos} [w_o t + k_m\ x(t)] \qquad 4.5$$

Then

$$\frac{\partial s[t,x(t)]}{\partial x(t)} = k_m\ A\ \sin [w_o t + k_m\ s(t)] \qquad 4.6$$

where

$x(t)$ = message signal for which an estimate $\hat{x}(t)$ is desired.

A = peak amplitude of the input signal.

and

k_m = gain constant of the modulator.

The block diagrams of the optimum PM reciever based on Equation 4.3 are shown in Figures 11 and 12. Here as before, the a priori data are taken equal to zero, i.e., $x(t)$ is assumed to be a zero mean signal. The modification of Figure 11 to Figure 12 is justified by noting that the received signal $r(t)$ is given by

$$r(t) = A \operatorname{Cos} [w_o t + k_m x(t)] + n(t) \qquad 4.7$$

where

$n(t)$ = noise (error).

The output of the modulator $A \operatorname{Cos} [w_o t + k_m \hat{x}(t)]$ when subtracted from $r(t)$ and subsequently multiplied by $\sin [w_o t + k_m \hat{x}(t)]$ produces only double frequency terms. Since these will not be passed by the low pass filter, they may be discarded and the differencing operation may be deleted as it makes no useful contribution to the estimation process.

An important aspect of the two examples involving AM and PM is that the receiver structures derived are essentially the same as the optimum receivers derived by others using a statistical approach, for example see VanTrees [13]. The significant feature of these derivations is that they result from a gls approach which does not require any statistical foundation.

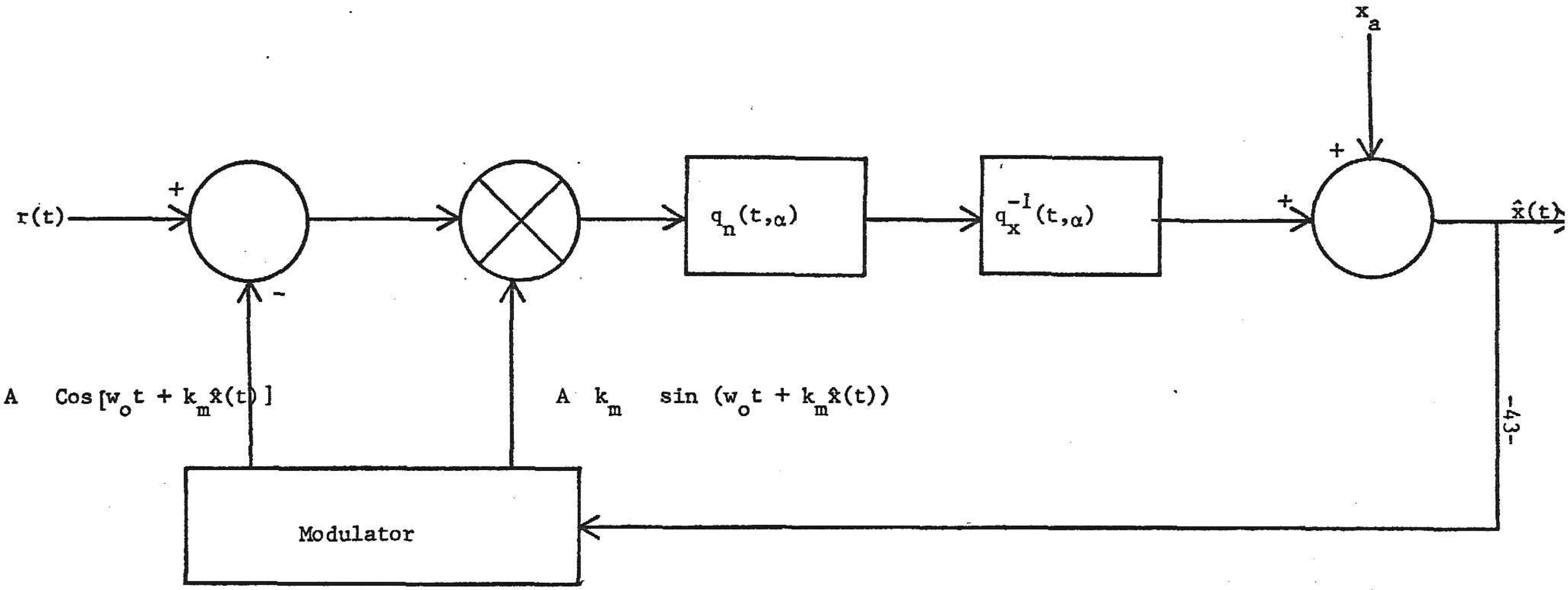

Figure 11. Block Diagram for Phase Modulation Scheme

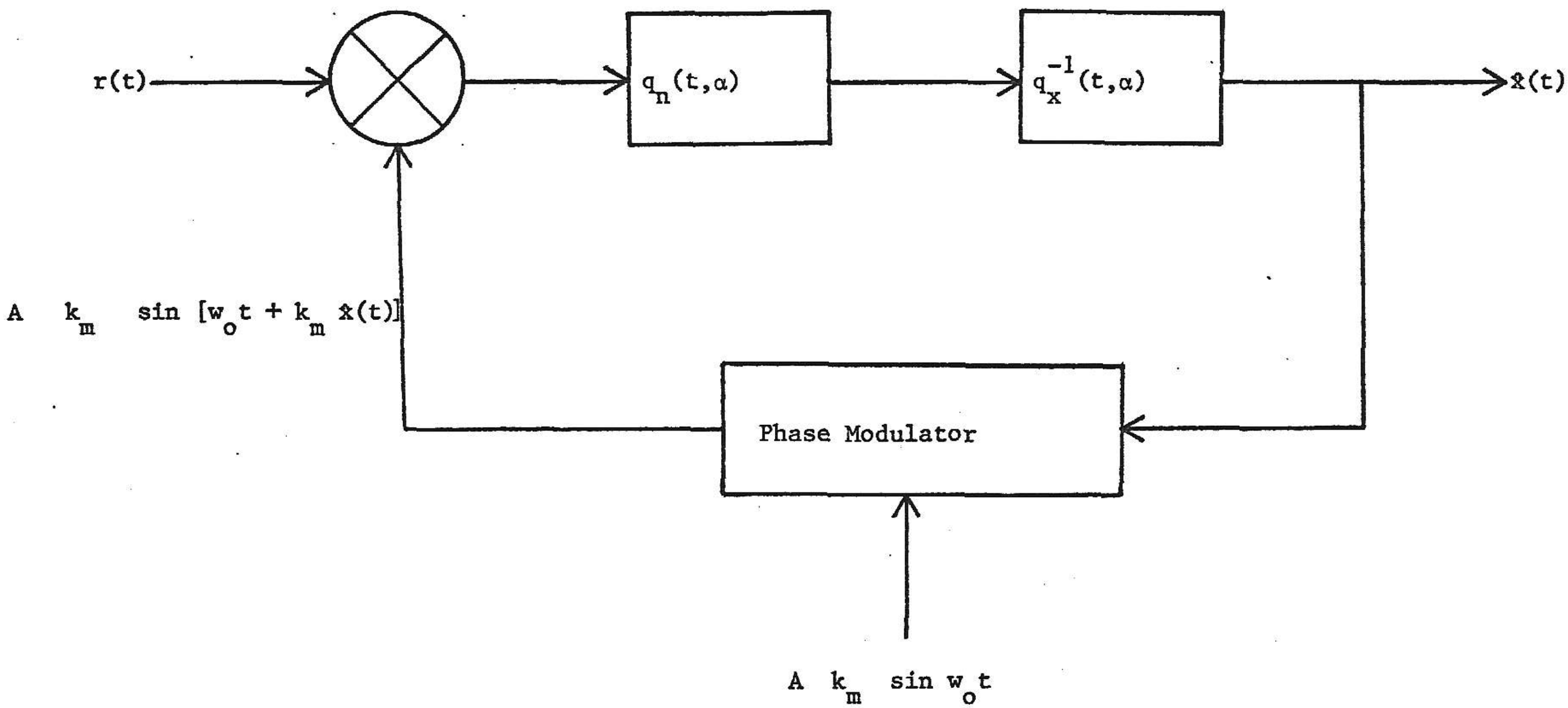

Figure 12. Modified Version of Figure 11

CHAPTER V

V. CONCLUSIONS

In the preceding pages a survey of important methods used for finding estimates of unknown parameters using the least squares technique have been described. Estimation of parameters with and without a priori data have been considered. Swerling's method which uses an update of the estimate and its covariance matrix is also considered. It is shown that Battin's method which uses scalar observations does not involve any inversion of matrices and consequently which the computational time required by this method is less. It is also shown that the receiver structures derived are essentially the same as the optimum receivers derived by using a statistical approach, whereas the results from the gls approach do not require any statistics.

BIBLIOGRAPHY

BIBLIOGRAPHY

1. Swerling, P. "Modern State Estimation Methods from the Viewpoint of the Method of Least Squares," IEEE Transaction on Automatic Control, (December 1971), page 707-719.

2. Blackman, R. B. "Methods of Orbit Refinement," The Bell Systems Technical Journal, (May 1964), page 885-909.

3. Swerling, P. "Topics in Generalized Least Squares Signal Estimation," J. Siam Applied Math, (September 1966), page 998-1031.

4. Kailath, Thomas. "A View of Three Decades of Linear Filtering Theory," IEEE Transaction on Information Theory, (March 1974), page 146-181.

5. Deutsch, R. "Estimation Theory," Prentice-Hall, Inc. (1965).

6. Swerling, P. "Parameter Estimation Accuracy Formulas," IEEE Transaction on Information Theory, (October, 1964), page 302-314.

7. Sorenson, H. W. "Least Squares Estimation From Gauss to Kalman," IEEE Spectrum, (July 1970), page 63-68.

8. Swerling, P. "A Proposed Stage Wise Differential Correction Procedure for Satellite Tracking and Prediction," Rand Corporation Report, (January 1958), page 1292.

9. Claus, A. J. Private Communication.

10. Battin, R. H. "A Statistical Optimizing Navigation Procedure for Space Flight," Journal ARA, 32, (November 1962), page 1681-1696.

11. VanTrees, Harry L. "Detection, Estimation and Modulation Theory," Part I. New York: John Wiley & Sons, Inc., 1968.

12. Blackman, R. B. Unpublished work.

13. VanTrees, Harry L. "Analog Communication Over Randomly-Time-Varying Channels," IEEE Transaction on Information Theory, (January 1966), page 51-63.

ACKNOWLEDGEMENTS

I am indebted to Dr. Donald R. Hummels for his valuable guidance and kind supervision throughout the course of this work. It is only due to the constant encouragement of my advisor that I have been able to achieve my goal.

I am also thankful to the Head of the Department, Dr. W. W. Koepsel for providing financial assistance during my stay in this school.

A SURVEY OF GENERALIZED LEAST SQUARES ESTIMATION AND ITS RELATION TO COMMUNICATION SYSTEM DESIGN

by

Altaf-ur-Rashid

B.S., University of Karachi, 1974

AN ABSTRACT OF A MASTER'S REPORT

submitted in partial fulfillment of the

requirement for the degree

MASTER OF SCIENCE

Department of Electrical Engineering

KANSAS STATE UNIVERSITY

Manhattan, Kansas

1976

A general approach due to Swerling is presented for the problem of estimation where the parameters are non-stochastic or "deterministic." It is shown that an alternate method which does not require statistics is the generalized least squares method. The development is presented as an elaboration of Gauss's method of least squares, and consequently differs from the more conventional development, which is based on statistical methods. All the methods considered, use vector parameters except Battin's method, in which scalar quantities are used. Battin's method also suggests that if the new data are introduced one at a time, the inversion of matrices can be avoided. Because of this the computational time is reduced and accurate results are more easily obtained. The results obtained are applied to communication problems for finding the optimum receivers for various modulation schemes, e.g. amplitude and phase modulation. It is observed that the resulting receivers structures are the same as those obtained by statistical methods when statistical properties are assumed for the noise and message. An advantage of the approach is discussed here, is that no statistical assumptions are required to specify the optimum receiver.