

PARAMETER ESTIMATION OF LINEAR SYSTEMS IN THE PRESENCE OF PROCESS NOISE USING THE MAXIMUM LIKELIHOOD METHOD

1. INTRODUCTION

This is the third Data Item dealing with methods for the parameter estimation of linear systems.

The first Data Item, No. 87039 (Reference 1), introduced the parameter estimation process with methods based on the least-squares principle in which the system model was assumed free from process noise. The methods were essentially divided into direct, non-iterative methods in which the system equations were solved in terms of both input and output measurements, and iterative methods that solved the system equations in terms of only input measurements. The former methods result in a decoupling of those equations containing different parameters so that each parameter is found separately. The latter methods result in a solution of all the parameters as a set and provide a measure of the correlation between pairs of parameters. A disadvantage of these latter methods is that biasing of the parameter estimates occurs when each of the measured state variables contains differing levels of noise.

The second Data Item, No. 88011 (Reference 2), presented the Maximum Likelihood method for the parameter estimation of linear systems whose models were assumed free from process noise. The method was iterative, like the latter class of method of Item No. 87039, but could also take account of differing levels of noise on each of the various instruments.

This Item presents the parameter estimation procedure for linear systems in the presence of process noise in which all measurements are corrupted by measurement noise. The Maximum Likelihood method is used to find optimised values of the parameters in terms of the solutions of the state equations. The Kalman filter (see Reference 3) is used to estimate the time varying state solutions of a system driven by a combination of known (deterministic) inputs and random inputs (process noise).

The parameter estimation process is introduced by considering a first-order system involving a single unknown parameter. The system is assumed driven by both a deterministic input and a random input. Direct measurements of the state variable of the system are available and are assumed corrupted by measurement noise. It is shown that the unknowns of the system to be found are the unknown in the state equation and in addition the variances of both the process noise and the measurement noise. An example of an aircraft executing a pure rolling manoeuvre illustrates the estimation procedure for a single first-order system.

The generalised parameter estimation process for a multiparameter system of many degrees-of-freedom is also described.

2. NOTATION

Any coherent system of units may be used.

$\underline{A}$	plant or stability matrix
a	coefficient of state variable (used for system approximated by single first-order differential equation)
$\underline{B}$	control or input matrix

$\underline{C}, \underline{D}, \underline{E}$	matrices associated with observation equation
$E\{\dots\}$	expected value of $\{\dots\}$
$\underline{F}, \underline{G}$	matrices associated with alternative form of observation equation
$\underline{H}$	system (or process) noise matrix
h	number of unknown parameters
I	identity matrix
J	Objective Function
k	Kalman gain
$L\{\dots\}$	Likelihood Function of $\{\dots\}$
l	number of observation variables
m	number of coefficients associated with input quantities
N	total number of discrete time points
n	number of coefficients associated with state variables
$\underline{P}$	covariance matrix of state variables
p	number of white noise inputs
$p\{\dots\}$	probability density of $\{\dots\}$
$p\{y \gamma\}$	probability density of y given γ
$\underline{Q}$	covariance matrix associated with system noise
$\underline{R}$	covariance matrix associated with measurement noise
$\underline{S}$	covariance matrix associated with difference quantity
t, t''	time
u	input variable
v	measurement noise
w	system (or process) noise
x	state variable
y	observation variable
z	measured quantity

γ	unknown parameter
$\underline{\gamma}$	row vector containing unknown parameters
Δ	denotes increment
δ_{ij}	Kronecker delta function, $= \begin{matrix} 1 & i = j \\ 0 & i \neq j \end{matrix}$
$\delta(t - t'')$	Dirac delta function, $= \begin{matrix} 1 & t = t'' \\ 0 & t \neq t'' \end{matrix}$
ε	difference between measured quantity and model observation variable
$\underline{\Lambda}, \underline{\Phi}, \underline{\Psi}$	matrices associated with discrete form of state equation
τ	time interval, $t_{i+1} - t_i$
σ	standard deviation

Subscripts

f	denotes forecast
I	denotes initial estimate
i	denotes discrete time points
j, k	arbitrary integers
max	maximum
0	initial condition
opt	optimised
v	relates to measurement noise
w	relates to system noise
s	denotes solution
ε	relates to difference quantity

Superscripts

T	denotes transpose
$'$	denotes an uprated value

A bar below a quantity denotes matrix or vector quantity.

One dot above a quantity denotes first derivative of that quantity with respect to time.

3. SINGLE FIRST-ORDER SYSTEM

3.1 Problem Formulation

The parameter estimation method for systems including process, or system, noise is illustrated by considering the following first-order linear differential equation involving a single unknown parameter:

$$\dot{x}(t) = ax(t) + u(t) + w(t), \quad (3.1)$$

where $x(t)$ is the state variable of the system, a is the unknown parameter to be found, $u(t)$ is the known (deterministic) input to the system and $w(t)$ is the system (or process) noise. The random system noise is assumed to have the following mean and variance values that are assumed independent of time

$$E\{w(t)\} = 0, \text{ for all } t \quad (3.2a)$$

and $E\{w(t)^2\} = \sigma_w^2, \quad (3.0c)$

where σ_w^2 is another unknown parameter to be found.

The system state response, *i.e.* the solution of Equation (3.1), is the addition of the deterministic response due to the known input, $u(t)$, and the random response due to the system noise, $w(t)$. The state variable, $x(t)$, is consequently a random variable with an uncertainty expressed by its variance which is denoted by $\sigma^2(t)$. Assuming that $u(t)$ is zero for $t < 0$, then with the aid of Equation (3.2a) the initial state of the system may be expressed by the following mean and variance values

$$E\{x(0)\} = x_0 \quad (3.4a)$$

and $E\{[x(0) - x_0]^2\} = \sigma^2(0), \quad (3.0e)$

where the terms x_0 and $\sigma^2(0)$ are assumed known.

It is assumed, for simplicity, that the observation variable is identified with the state variable and so in discrete form the observation equation is

$$y_i = x_i, \quad i = 1, 2, \dots, N, \quad (3.6)$$

where i denotes the discrete values of t_i at which the measurements are taken. The observation variable, y_i , it should be recognised, is also a random variable with mean and variance values identical to x_i .

Actual measurements, z_i , of the system are related to the observation variable by

$$z_i = y_i + v_i, \quad (3.7)$$

where v_i represents the discrete form of the measurement noise with mean and variance values given by

$$E\{v_i\} = 0, \text{ for all } i \quad (3.8a)$$

and $E\{v_i^2\} = \sigma_v^2, \quad (3.0i)$

where σ_v^2 , which is independent of time, is a further unknown parameter to be found. It should be recognised, once again, that the measurements, z_i , are random variables containing uncertainties due to

both the process noise and the measurement noise. The system noise and the measurement noise are assumed to be independent and so

$$E\{w(t)v(t'')\} = 0, \text{ for all } t \text{ and } t''.$$

Measurements of the input, u_i , are denoted by $z(u)_i$, $i = 1, \dots, N$, and are assumed in this method to be measurement error free. Since the measurements are in discrete form it is convenient to express the state equation in discrete form which becomes

$$x_{i+1} = (a\tau + 1)x_i + \tau u_i + \tau w_i, \quad (3.10)$$

where τ is the time interval which is equal to $t_{i+1} - t_i$.

The unknown parameter set, denoted by $\underline{\gamma}$, is then a row vector with elements given by

$$\underline{\gamma} = [a, \sigma_w^2, \sigma_v^2]. \quad (3.11)$$

The objective of this method is to find the values of the parameters contained in $\underline{\gamma}$ that minimise the difference between the set of measurements and a set of equivalent observation variables. A difference quantity, ε_i , may be written as

$$\varepsilon_i = z_i - (y_f)_i \equiv z_i - (x_f)_i, \quad (3.12a)$$

where the subscript f denotes the forecast solution of x_i in terms of $z(u)_i$ for given values of the parameters. The difference function, ε_i , is minimised using the Maximum Likelihood method (Section 3.2) and the forecast state solutions, $(x_f)_i$, are obtained with the use of a Kalman filter (Section 3.3). A general flow chart of this Maximum Likelihood estimation process is shown in Sketch 3.1 where that part pertaining to the Kalman filter is highlighted.

When ε_i has been minimised, *i.e.* it has reached its optimum value, then

$$(\varepsilon_{opt})_i = z_i - \{(x_f)_{opt}\}_i, \quad (3.0m)$$

where $\{(x_f)_{opt}\}_i$ is interpreted as the forecast state solutions when $\underline{\gamma} \rightarrow \underline{\gamma}_{opt}$, where

$$\underline{\gamma}_{opt} = [a_{opt}, (\sigma_w^2)_{opt}, (\sigma_v^2)_{opt}]. \quad (3.14)$$

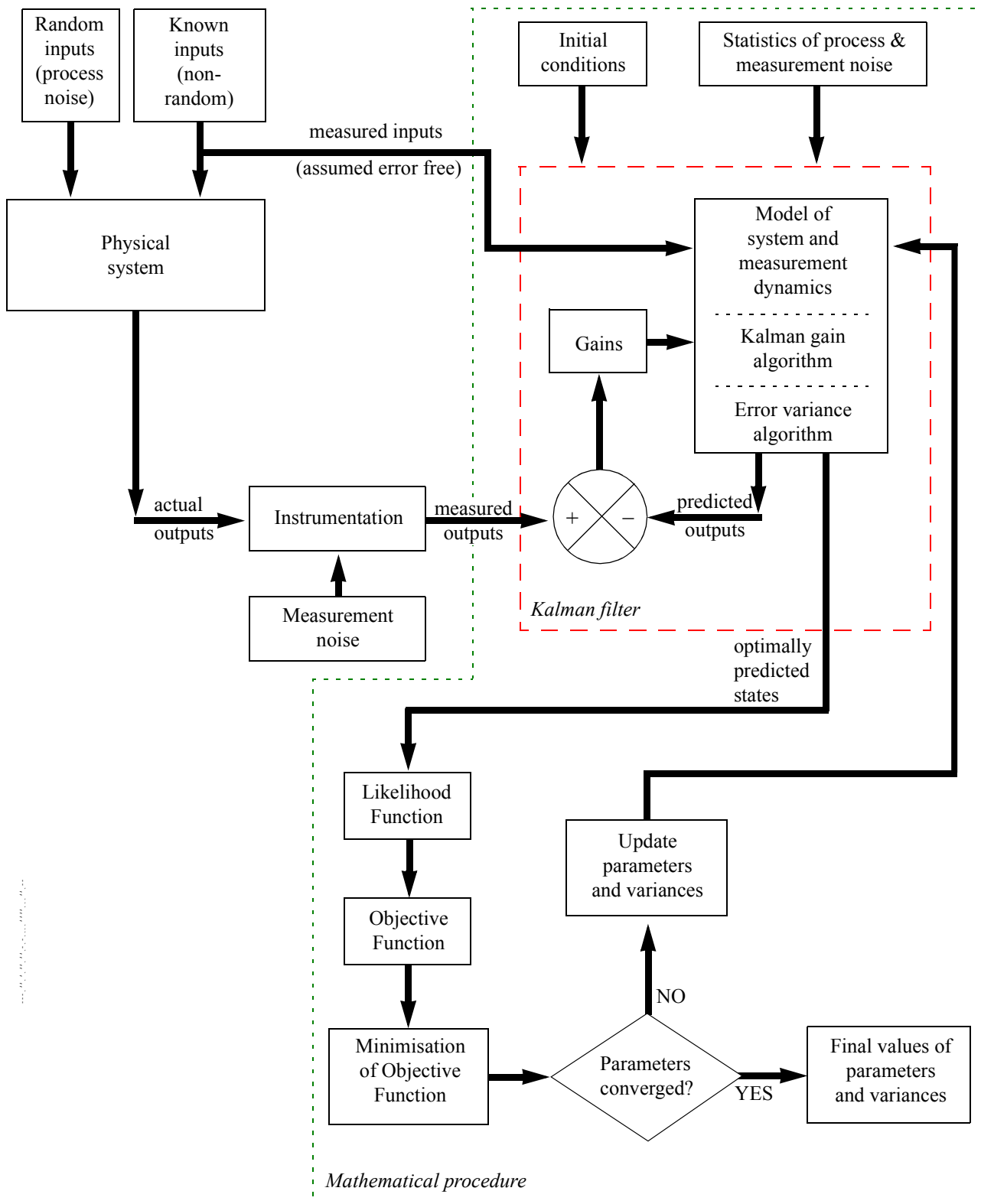
The uncertainty associated with the difference quantity, ε_i , may be found by the following analysis.

Squaring both sides of Equation (3.12a) and with the use of Equations (3.6) and (3.7) the following expression is obtained

$$\varepsilon_i^2 = [x_i + v_i - (x_f)_i]^2.$$

The expected value of this equation gives

$$E\{\varepsilon_i^2\} = E\{[x_i + v_i - (x_f)_i]^2\}.$$



Sketch 3.1 General flow chart of Maximum Likelihood estimation process for systems including process noise

Assuming that $[x_i - (x_f)_i]$ and v_i are uncorrelated random variables, this last expression reduces to

$$E\{\varepsilon_i^2\} = E\{[x_i - (x_f)_i]^2\} + E\{v_i^2\}.$$

Since x_i is a true value of the state variable at i , then $[x_i - (x_f)_i]$ is the error in the estimation of the forecast state variable and so

$$(\sigma_\varepsilon^2)_i = (\sigma_f^2)_i + (\sigma_v^2), \quad (3.15)$$

where $(\sigma_\varepsilon^2)_i$ and $(\sigma_f^2)_i$ are the variances of the difference term and the forecast state solution, respectively. This equation shows that the variance of ε_i is the sum of the variances of the forecast state solution and the measurement noise.

3.2 Maximum Likelihood Method

The Maximum Likelihood method is based on finding the set of optimised parameters of a system that maximises a Likelihood Function. At this optimised condition the Maximum Likelihood Function is the joint probability density function of the observations.

The Maximum Likelihood method* makes use of the assumption that the difference quantity, ε_i , is white and has a Gaussian probability distribution with a mean level of zero. It is further assumed that when an optimised solution has been obtained, *i.e.* when the unknown parameters have their optimised values, then the difference quantity, $(\varepsilon_{opt})_i$, is a sample of the sum of the measurement noise, v_i , and the error in the forecast state variable, $(x_f)_i$. Both v_i and $(x_f)_i$ are assumed Gaussian random variables with means of zero. Consequently the mean and variance of this difference quantity may be expressed as

$$E\{(\varepsilon_{opt})_i\} = \frac{1}{N} \sum_{i=1}^N (\varepsilon_{opt})_i \rightarrow 0 \quad (3.16a)$$

and

$$E\{(\varepsilon_{opt})_i^2\} = \frac{1}{N} \sum_{i=1}^N (\varepsilon_{opt})_i^2 \rightarrow \{(\sigma_\varepsilon^2)_{opt}\}_i = \{(\sigma_f^2)_{opt}\}_i + (\sigma_v^2)_{opt}. \quad (3.1q)$$

At the optimised condition for each error, $(\varepsilon_{opt})_i$, the probability density function is given by

$$p\{(\varepsilon_{opt})_i | \gamma_{opt}\} = \frac{1}{[2\pi\{(\sigma_\varepsilon^2)_{opt}\}_i]^{1/2}} \exp\left[-\frac{(\varepsilon_{opt})_i^2}{2\{(\sigma_\varepsilon^2)_{opt}\}_i}\right]. \quad (3.18)$$

The joint probability density function for the sequence of errors, $(\varepsilon_{opt})_i$, $i = 1, \dots, N$, assuming all the errors are uncorrelated, is given by

$$p\{|\} = \prod_{i=1}^N \left[\frac{1}{[2\pi\{(\sigma_\varepsilon^2)_{opt}\}_i]^{1/2}} \exp\left[-\frac{(\varepsilon_{opt})_i^2}{2\{(\sigma_\varepsilon^2)_{opt}\}_i}\right] \right], \quad (3.19a)$$

* For description of the Maximum Likelihood estimator see Section 3.3 of Data Item No. 88011 (Reference 2).

and

$$p\{\underline{\gamma}\} = \prod_{i=1}^N \left[\frac{1}{[2\pi\{(\sigma_{\varepsilon}^2)_{opt}\}_i]^{1/2}} \right] \exp \left[- \sum_{i=1}^N \frac{(\varepsilon_{opt})_i^2}{2\{(\sigma_{\varepsilon}^2)_{opt}\}_i} \right]. \quad (3.1t)$$

The Maximum Likelihood Function is

$$L_{max}\{\underline{\gamma}_{opt}\} = p\{\underline{\gamma}\}, \quad (3.21)$$

where $p\{\underline{\gamma}\}$ is given by Equation (3.1t).

A Likelihood Function may now be formulated, based on the form of Equations (3.1t) and (3.21), as

$$L\{\underline{\gamma}\} = \left[\prod_{i=1}^N \frac{1}{[2\pi(\sigma_{\varepsilon}^2)_i]^{1/2}} \right] \exp \left[- \sum_{i=1}^N \frac{\varepsilon_i^2}{2(\sigma_{\varepsilon}^2)_i} \right] \quad (3.22)$$

and as the optimisation proceeds $L\{\underline{\gamma}\}$ is maximised and each variable approaches its optimised value, *i.e.*

$$\begin{aligned} \underline{\gamma} &\rightarrow \underline{\gamma}_{opt}, \\ \text{or} \quad a &\rightarrow a_{opt}, \\ \sigma_w^2 &\rightarrow (\sigma_w^2)_{opt} \\ \text{and} \quad \sigma_v^2 &\rightarrow (\sigma_v^2)_{opt}. \end{aligned}$$

In the Maximum Likelihood method it is usual to define the Objective Function, J (often referred to as the Cost Function), as the negative logarithm of the Likelihood Function and so

$$J = -\ln \{L\{\underline{\gamma}\}\}, \quad (3.23a)$$

or

$$J = \sum_{i=1}^N \left[\frac{\varepsilon_i^2}{2(\sigma_{\varepsilon}^2)_i} + \frac{1}{2} \ln \{(\sigma_{\varepsilon}^2)_i\} \right] + \text{constant}. \quad (3.1x)$$

This equation may be rewritten using Equation (3.12a) as

$$J = \sum_{i=1}^N \left[\frac{[z_i - (x_f)_i]^2}{2(\sigma_{\varepsilon}^2)_i} + \frac{1}{2} \ln \{(\sigma_{\varepsilon}^2)_i\} \right] + \text{constant}. \quad (3.1y)$$

The maximum value of the Likelihood Function, L , occurs when the Objective Function, J , is a minimum.

The Objective Function, J , is seen to be a function of the two independent unknowns, $(\sigma_{\varepsilon}^2)_i$ and $(x_f)_i$, or $(\sigma_{\varepsilon}^2)_i$ and the parameters a , σ_w^2 and σ_v^2 .

The minimum value of J is at the optimised condition and so

$$\left\{ \frac{\partial J}{\partial a} \right\}_{opt} = \left\{ \frac{\partial J}{\partial \sigma_w^2} \right\}_{opt} = \left\{ \frac{\partial J}{\partial \sigma_v^2} \right\}_{opt} = 0. \quad (3.26)$$

These three equations give

$$\sum_{i=1}^N \left[\frac{[z_i - (x_f)_i]^2}{2(\sigma_\varepsilon^2)_i^2} \left[\frac{\partial(\sigma_\varepsilon^2)_i}{\partial \gamma_j} \right] + \frac{[z_i - (x_f)_i]}{(\sigma_\varepsilon^2)_i} \left[\frac{\partial(x_f)_i}{\partial \gamma_j} \right] - \frac{1}{2(\sigma_\varepsilon^2)_i} \left[\frac{\partial(\sigma_\varepsilon^2)_i}{\partial \gamma_j} \right] \right] \gamma_j = (\gamma_{opt})_j, \quad j = 1, 2, 3. \quad (3.27)$$

where $\gamma_1 = a$, $\gamma_2 = \sigma_w^2$ and $\gamma_3 = \sigma_v^2$.

To solve the nonlinear equations, Equations (3.27), it is necessary to adopt an iterative approach. The procedure that will be used is a second-order optimisation procedure known as the Modified Newton-Raphson procedure or Newton-Gauss procedure. This procedure approximates the Objective Function by the first three terms of its Taylor series expansion about an initial estimate of the parameters, say a_I , $(\sigma_w^2)_I$ and $(\sigma_v^2)_I$. Expanding J about these initial values gives the following equations:

$$J[a, (\sigma_w^2)_I, (\sigma_v^2)_I] = J[a_I, (\sigma_w^2)_I, (\sigma_v^2)_I] + (a - a_I) \left(\frac{\partial J}{\partial a} \right)_I + \frac{(a - a_I)^2}{2} \left(\frac{\partial^2 J}{\partial a^2} \right)_I, \quad (3.28a)$$

$$J[a_I, \sigma_w^2, (\sigma_v^2)_I] = J[a_I, (\sigma_w^2)_I, (\sigma_v^2)_I] + [\sigma_w^2 - (\sigma_w^2)_I] \left(\frac{\partial J}{\partial \sigma_w^2} \right)_I + \frac{[\sigma_w^2 - (\sigma_w^2)_I]^2}{2} \left(\frac{\partial^2 J}{\partial (\sigma_w^2)^2} \right)_I \quad (3.1ac)$$

and

$$J[a_I, (\sigma_w^2)_I, \sigma_v^2] = J[a_I, (\sigma_w^2)_I, (\sigma_v^2)_I] + [\sigma_v^2 - (\sigma_v^2)_I] \left(\frac{\partial J}{\partial \sigma_v^2} \right)_I + \frac{[\sigma_v^2 - (\sigma_v^2)_I]^2}{2} \left(\frac{\partial^2 J}{\partial (\sigma_v^2)^2} \right)_I, \quad (3.1ad)$$

where a , σ_w^2 and σ_v^2 are related to increments in the parameters, Δa , $\Delta \sigma_w^2$ and $\Delta \sigma_v^2$, by

$$a = a_I + \Delta a, \quad (3.31a)$$

$$\sigma_w^2 = (\sigma_w^2)_I + \Delta \sigma_w^2 \quad (3.1af)$$

and

$$\sigma_v^2 = (\sigma_v^2)_I + \Delta \sigma_v^2. \quad (3.1ag)$$

Equations (3.28a) to (3.1ad) are now differentiated with respect to the unknown parameters, a , σ_w^2 and σ_v^2 , respectively, to give, upon rearrangement and use of Equations (3.31a) to (3.1ag),