General Disclaimer

One or more of the Following Statements may affect this Document

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.

- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.

- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.

- This document is paginated as submitted by the original source.

- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.

Produced by the NASA Center for Aerospace Information (CASI)

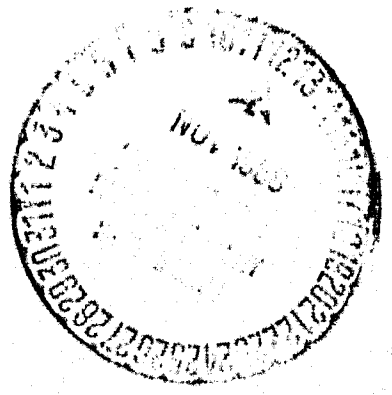
Clemson University

Department of Mathematics

LEAST SQUARES, MINIMUM VARIANCE AND KALMAN ESTIMATION APPLIED TO A MATHEMATICAL MODEL

December 1967

N69-25230
(ACCESSION NUMBER)
59
(PAGES)
CR 92431
(NASA CR OR TMX OR AD NUMBER)
(THRU)
(CODE)
19
(CATEGORY)
FACILITY FORM 602



Clemson University

Department of Mathematics

LEAST SQUARES, MINIMUM VARIANCE AND KALMAN ESTIMATION APPLIED TO A MATHEMATICAL MODEL

December 1967

Final Report on Contract NAS 8-11259

Advanced Studies Office

Astrionics Laboratory

Marshall Space Flight Center

National Aeronautics and Space Administration

Huntsville, Alabama

CLEMSON UNIVERSITY INVESTIGATORS

S. M. Lukawecki

B. J. Prochaska

Table of Contents

Abstract iv

Summary. v

Illustrations. vi

Tables vii

List of Symbols by Chapters. viii

I. Introduction 1

II. Least Squares and Minimum Variance Unbiased Estimators 6

III. Application of L. S. and M. V. U. Estimators to Trajectory Estimation . . 17

IV. The Mathematical Model 19

V. Simulation Details and Results 30

VI. Summary and Future 42

Appendix I: Basic Statistics. 44

Appendix II: Basic Linear Algebra 45

Bibliography 47

Abstract

The investigating members of the Mathematics Department of Clemson University, under an extension of NASA Contract 8-11259, investigated an application of a least squares estimator, a minimum variance unbiased estimator, and Kalman's filter to a linear mathematical model. The model considered had an exact transition matrix. By using approximations to the known matrix, comparisons were made of the various estimators.

Summary

A review of the concepts of least squares (L. S.) and minimum variance unbiased (M. V. U.) estimators as well as the Kalman filter is given. The L. S. and M. V. U. estimators are then derived for a multivariate linear model.

The L. S. and M. V. U. estimators and the Kalman filter are applied to the problem of trajector estimation.

A mathematical model is constructed to illustrate these three estimators. The model is that of a particle in force free flight on a parabolic curve in the x-y plane. With this model the state vector is estimated; that is, the position and velocity of the particle is estimated at a final time. Additive noise in the observations is employed.

An exact transition matrix is obtained for the linear model. Approximate transition matrices are derived using Taylor series expansions in order to compare the three estimation schemes.

The dynamical system is particularized to generate a sequence of simulation runs. There are eight computer runs for each estimation scheme due to alternate selections of transition matrices and variations in the observation schedule.

The data is analyzed and presented in tabular form. The results exhibited in these tables indicate for the model employed the Kalman filter gave the "best" results.

Illustrations

Figure		Page
1	Flow Chart for the L. S. Estimator Given in Equation (3.3).	27
2	Flow Chart for the M. V. U. Estimator Given in Equation (3.5)	28
3	Flow Chart for the Solution of the Model Given in Equations (1.11) and (1.12).	29
4	Trace of Model's Actual Trajectory.	31

Tables

Table		Page
1	Φ ; $\Delta t = 0.25$; $t_f = 16.0$	34
2	Φ_1; $\Delta t = 0.25$; $t_f = 16.0$	35
3	Φ_2; $\Delta t = 0.25$; $t_f = 16.0$	36
4	Φ_5; $\Delta t = 0.25$; $t_f = 16.0$	37
5	Φ ; $\Delta t = 0.50$; $t_f = 16.25$	38
6	Φ_1; $\Delta t = 0.50$; $t_f = 16.25$	39
7	Φ_2; $\Delta t = 0.50$; $t_f = 16.25$	40
8	Φ_5; $\Delta t = 0.50$; $t_f = 16.25$	41

List of Symbols by Chapters

Introduction

X_1	random variables
E	expectation
λ	constant
$\hat{\lambda}$	estimate of λ in L. S. estimator
$\Phi(X_1, \cdots, X_n)$	function
$L(\Phi, \lambda)$	loss function
Φ	estimate of λ in M. V. U. estimator
t	time in general; present time
x(t)	state n-vector
$\Phi(t + 1; t)$	n × n transition matrix
u(t)	input n-vector
y(t)	output p-vector
M(t)	n × p matrix
Q(t)	covariance matrix of u(t)
$x^*(t + 1\|t)$	estimate of x(t + 1)
$\hat{x}^*(t_n)$	updated estimate of $x(t_n)$
$\hat{x}(t_n)$	estimate of $x(t_n)$
$P(t_n)$	covariance matrix of the error in the estimate of $\hat{x}(t_n)$
$\hat{y}(t_n)$	estimate of $y(t_n)$
$P^*(t_n)$	covariance matrix of the error in estimation at time t_n
$\hat{x}(t_0)$	estimate of $x(t_0)$

Least Squares and Minimum Variance Unbiased Estimators

q_i	variables
∇_q	partial differential operator vector
x	vector
x_j	jth component of x
Q	quadratic form
A	vector
Φ	symmetric positive definite matrix
f	function
y_i	vectors
m	vector
$A^{[n\nu \times q]}$	$n\nu \times q$ matrix
b	matrix
$\hat{b}$	least squares linear unbiased estimate of b
ψ	covariance matrix of error in estimation of b
$\hat{b}_\Phi$	weighted least squares linear estimate of b
X_i	random variable p-vector
μ	mean
W_i	weighing matrices
I	identity matrix
$\tilde{X}$	unbiased estimator of μ
Φ_i	covariance matrices
$\tilde{P}$	covariance matrix of $\tilde{X}$
$\hat{W}_i$	estimate of W_i
$\hat{X}$	unbiased estimator of μ
$\hat{P}$	covariance matrix of $\hat{X}$
Φ	diag $\{\Phi_1, \cdots, \Phi_n\}$

I^*	$(I, I, \cdots, I)$; p × np matrix
W	$(W_1, W_2, \cdots, W_n)$; p × np matrix
F, G	matrices
$\bar{X}$	least square estimator of μ
$\bar{P}$	covariance matrix of $\bar{X}$

Application of L. S. and M. V. U. Estimators to Trajectory Estimation

t_i	time at ith observation
t_f	final time
$\hat{x}(t_i)$	unbiased estimator of the state vector $x(t_i)$
$Cov(\hat{x}(t_i))$	covariance matrix of $\hat{x}(t_i)$
Φ	transition matrix
$\hat{x}_L(t_f)$	L. S. estimator of $x(t_f)$
$Cov(\hat{x}_L(t_f))$	covariance matrix of $\hat{x}_L(t_f)$
$\hat{x}_M(t_f)$	M. V. U. estimator of $x(t_f)$
$Cov(\hat{x}_M(t_f))$	covariance matrix of $\hat{x}_M(t_f)$

The Mathematical Model

$F(t)$	matrix
$A(t)$	matrix
$\Phi_k(t_2; t_1)$	approximation of $\Phi(t_2; t_1)$
$S(t_2; t_1)$	series expansion of an element of the Φ matrix
$R(t_2; t_1)$	series expansion of an element of the Φ matrix
$S_k(t_2; t_1)$	k terms of $S(t_2; t_1)$
$R_k(t_2; t_1)$	k terms of $R(t_2; t_1)$

Simulation Details and Results

$Q(t)$	covariance matrix of noise
t_0	initial time
$P(t_0)$	covariance matrix of error
$\hat{x}(t_0)$	initial estimate of $x(t_0)$
Δt	time increment
$xA(t_f)$	actual value of $\hat{x}(t)$ at $t = t_f$

Chapter I: Introduction

Considerable investigation in the past few years concerning space navigation problems has focused on three estimation techniques; least square estimator, minimum variance unbiased estimator, and statistical filter theory developed by R. E. Kalman.

The least square technique [1], one of the oldest methods of estimation, goes back to Gauss who first developed this technique. Generally it is used for estimation of parameters in a linear model.

Let $X_1, X_2, \cdots, X_n$ be independent random variables such that $E(X_i) = \lambda_1 + \lambda_2 y_i$ where y_i is a known constant and λ_1 and λ_2 are unknown constants. Consider the quadratic function

$$w(X_1, X_2, \cdots, X_n, \lambda_1, \lambda_2) = \sum_{i=1}^{n} [X_i - \lambda_1 - \lambda_2 y_i]^2. \tag{1.1}$$

Let $\hat{\lambda}_1$ and $\hat{\lambda}_2$ be the estimates of λ_1 and λ_2 respectively which minimize (1.1), then $\hat{\lambda}_1$ and $\hat{\lambda}_2$ are called least squares estimates of λ_1 and λ_2. The so-called normal equations

$$\frac{\partial w(X_1, X_2, \cdots, X_n, \lambda_1, \lambda_2)}{\partial \lambda_1} = -2 \sum_{i=1}^{n} [X_i - \lambda_1 - \lambda_2 y_i] = 0,$$

$$\text{or} \quad \sum_{i=1}^{n} X_i = n\lambda_1 + n\lambda_2 \sum_{i=1}^{n} y_i, \tag{1.2}$$

$$\text{and} \quad \frac{\partial w(X_1, \cdots, X_n, \lambda_1, \lambda_2)}{\partial \lambda_2} = -2 \sum_{i=1}^{n} [X_i - \lambda_1 - \lambda_2 y_i] y_i = 0,$$

$$\text{or} \quad \sum_{i=1}^{n} y_i X_i = \lambda_1 \sum_{i=1}^{n} y_i + \lambda_2 \sum_{i=1}^{n} y_i^2, \tag{1.3}$$

can be solved for λ_1 and λ_2. One can thus obtain

$$\hat{\lambda}_2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(y_i - \bar{Y})}{\sum_{i=1}^{n} (y_i - \bar{Y})^2} \tag{1.4}$$

and $$\hat{\lambda}_1 = \bar{X} - \hat{\lambda}_2 \bar{Y}. \tag{1.5}$$

To define an unbiased estimate, let $X_1, X_2, \cdots, X_n$ be independent random variables. A function of $X_1, X_2, \cdots, X_n$, $\Phi(X_1, X_2, \cdots, X_n)$, is said to be an unbiased estimator of λ if

$$E[\Phi(X_1, X_2, \cdots, X_n)] = \lambda. \tag{1.6}$$

If

$$E[\Phi(X_1, X_2, \cdots, X_n)] = \lambda + b, \quad b \neq 0 \tag{1.7}$$

then $\Phi(X_1, X_2, \cdots, X_n)$ is said to be a biased estimate of λ with expected bias b.

Intuitively, a loss function assigns a penalty to the error made in estimation. Consider the loss function $L(\Phi, \lambda) = (\Phi - \lambda)^2$. Assume the value of λ is unknown. This unknown value of λ can be estimated by Φ; if a value of Φ is close enough to λ the corresponding loss will be small, if it is not, then the loss will be large. Thus, one is interested in finding Φ such that

$$E[(\Phi - \lambda)^2] \leq E[(\Phi' - \lambda)^2] \quad \text{for every } \Phi'. \tag{1.8}$$

Then Φ has minimum mean square error and it is called a minimum mean square error estimate.

Now in addition, assume that Φ and Φ' are unbiased estimators of λ then (1.8) implies that Φ has minimum variance and hence it is a minimum variance unbiased estimator.

Kalman's statistical filter is a particular formulation of a solution to the Weiner problem. Kalman [2] expresses the underlying situation as a dynamic model

$$x(t + 1) = \Phi(t + 1; t)x(t) + u(t), \tag{1.9}$$

$$y(t) = M(t)x(t), \tag{1.10}$$

where $u(t)$ is an independent normally distributed random process of n-vectors with zero mean and covariance matrix $Q(t)$, $x(t)$ is an n-vector, $y(t)$ is a p-vector ($p \leq n$), and $\Phi(t + 1; t)$, $M(t)$ are $n \times n$ and $p \times n$ matrices, respectively, with elements as non-random functions of time. The problem is to find an estimate, $x^*(t + 1|t)$ of $x(t + 1)$ which minimizes the expected loss; given the observed values $y(t_0)$, $\cdots$, $y(t)$ that are assumed to be linearly independent.

The assumed normal distribution hypothesis implies that the optimal estimate would be the same as if the optimal estimate was restricted to a linear function of the observed random variables and a loss function $L(\varepsilon) = \varepsilon^2$, where ε is the difference between the quantity to be estimated and the estimator.

A dynamic model that is more applicable to trajectory analysis is Solloway's [3]

$$x(t_{n+1}) = \Phi(t_{n+1}; t_n)x(t_n), \tag{1.11}$$

$$y(t_n) = M(t_n)x(t_n) + u(t_n), \tag{1.12}$$

where $u(t_n)$ is an independent random process of p-vectors with zero mean and known covariance matrix, $Q(t_n)$, i.e.,

$$Q(t_n) = E[u(t_n)u'(t_n)]$$

$$E[u(t_n)u'(t_n)] = 0, \quad n \neq m$$

$$E[u(t_n)] = 0 \quad \text{for all } n.$$

The prime notation on a matrix is used to indicate the transpose. The problem is then to find an updated estimate $\hat{x}^*(t_n)$ of the state $x(t_n)$ by using the observation $y(t_n)$ and an estimate $\hat{x}(t_n)$ based on the observations $y(t_0)$, $y(t_1)$, $\cdots$, $y(t_{n-1})$. The covariance matrix $P(t_n)$ of the error in estimate $\hat{x}(t_n)$ is assumed to be known. The criteria used in establishing the updated estimate $\hat{x}^*(t_n)$ is to minimize the trace of the covariance matrix

$$E[\{x(t_n) - \hat{x}^*(t_n)\}\{x(t_n) - \hat{x}^*(t_n)\}'].$$

The solution of Solloway's model, using R. E. Kalman's result, can be represented by

$$\hat{x}^*(t_n) = \hat{x}(t_n) + P(t_n)M'(t_n)(M(t_n)P(t_n)M'(t_n) + Q(t_n))^{-1}(y(t_n) - \hat{y}(t_n)) \tag{1.13}$$

where

$$\hat{y}(t_n) = M(t_n)\hat{x}(t_n), \tag{1.14}$$

$$\hat{x}(t_n) = \Phi(t_n; t_{n-1})\hat{x}^*(t_{n-1}), \tag{1.15}$$

$$P(t_n) = \Phi(t_n; t_{n-1})P^*(t_{n-1})\Phi'(t_n; t_{n-1}). \tag{1.16}$$

The covariance matrix, $P^*(t_n)$, of the error in the estimate is

$$P^*(t_n) = P(t_n) - P(t_n)M'(t_n)(M(t_n)P(t_n)M'(t_n) + Q(t_n))^{-1}M(t_n)P(t_n). \tag{1.17}$$

An initial estimate $\hat{x}(t_0)$ and the covariance matrix $P(t_0)$ of the error in the estimate is required to initiate the iterative scheme.

The major difficulty in applications is that the state transition matrix Φ is not known analytically. Since any application of the previously mentioned estimation schemes assumes that the transition matrix is the exact transition matrix for the given dynamic system, the accuracy of the estimator is as good as the approximation of the transition matrix. A major task of this study is to compare the accuracy of these estimation schemes when approximate transition matrices are used.

Chapter II: Least Squares and Minimum Variance Unbiased Estimators

Some aspects of linear least squares procedures and minimum variance unbiased estimator procedures need to be considered in more generality and detail than in Chapter I.

Matrices whose elements are functions of several variables will be encountered and in particular interest will be centered on matrices which result from considering partial derivatives of the elements of the matrices. Let $q_1, \cdots, q_n$ be n independent variables. Define the matrix of partial derivative operators formally as the vector,

$$\nabla_q^{[n\times 1]} = \begin{bmatrix} \frac{\partial}{\partial q_1} \\ \vdots \\ \frac{\partial}{\partial q_n} \end{bmatrix}^{[n\times 1]} .$$

The vector ∇_q operates only on the transpose of column vectors. Let x be the vector whose j^{th} component is $x_j = x_j(q_1, \cdots, q_n)$. Then

$$\nabla_q^{[n\times 1]}{}_x{}_{,[1\times p]} = \begin{bmatrix} \frac{\partial}{\partial q_1} \\ \vdots \\ \frac{\partial}{\partial q_n} \end{bmatrix} [x_1, \cdots, x_p]$$

$$= \begin{bmatrix} \frac{\partial x_1}{\partial q_1} & \frac{\partial x_2}{\partial q_1} & \cdots & \frac{\partial x_p}{\partial q_1} \\ \vdots & & & \\ \frac{\partial x_1}{\partial q_n} & \frac{\partial x_2}{\partial q_n} & \cdots & \frac{\partial x_p}{\partial q_n} \end{bmatrix} .$$

Usually the operator ∇_q is applied to quadratic forms (i.e., 1×1 matrices). The following theorem will be useful.

Theorem 2.1. Let $Q = A'\Phi A$ where A is a vector whose elements are functions of the independent variables $q_1, \cdots, q_n$ and Φ is a symmetric positive definite matrix. If the partial derivatives exist, then

$$\nabla_q Q = 2(\nabla_q A')\Phi A.$$

The proof of this theorem may be found in reference [4].

The following theorem from calculus concerning maxima and minima of functions is stated without proof.

Theorem 2.2. If $f(q_1, q_2, \cdots, q_n)$ is a function of n independent variables and if all partials $\partial f/\partial q_i$ are continuous, then f attains its maxima and minima only at the points where

$$\frac{\partial f}{\partial q_1} = \frac{\partial f}{\partial q_2} = \cdots = \frac{\partial f}{\partial q_n} = 0 .$$

In least square linear estimation one proceeds as follows.

Let $y_1, \cdots, y_n$ be a series of observation vectors where $y_i = (y_{i1}, \cdots, y_{i\nu})$ and

$$y' = [y_1' \cdots y_n'] .$$

Assume that $nv \geq q$ and

$$y^{[nv\times 1]} = m^{[nv\times 1]} + x^{[nv\times 1]} \qquad (2.1)$$

$$= \text{mean} + \text{noise},$$

and where $E[y] = m$ and $m = A^{[nv\times q]}b^{[q\times 1]}$.

A is a known matrix of non-random quantities and it is of full rank. The b_i's are unknown constants. [The assumption that A is of full rank implies that A'A is non-singular and hence $(A'A)^{-1}$ exists.] If A is not of full rank then the methods of generalized inverses may be used. The assumption that $Ey = m$ is equivalent to the assumption that $Ex = 0$.

The least squares linear unbiased estimate of b, is defined to be the vector $\hat{b}$ which minimizes the quadratic function,

$$R(b) = (y - Ab)'(y - Ab) \ . \qquad (2.2)$$

Now applying ∇_b to R(b) and setting the resulting vector of partials equal to zero, the critical value(s) will be obtained as stated in Theorem (2.2). It follows from Theorem (2.1) that,

$$\nabla_b R(b) = -2A'y + 2A'Ab$$

which yields

$$\hat{b} = (A'A)^{-1}A'y \ .$$

It is easy to show then that $\hat{b}$ is actually a minimum and that the covariance matrix of the error, $b - \hat{b}$, is given by

$$\psi = (A'A)^{-1}A'E(x\ x')A(A'A)^{-1} \ .$$

In weighted least squares linear estimations, assume that the observations are of the form of equation (2.1). Instead of using equation (2.2), the following quadratic function is used;

$$R(b) = (y - Ab)'\Phi^{-1}(y - Ab) \tag{2.3}$$

where Φ is a positive definite symmetric matrix. The weighted least squares linear estimate of b, denoted $\hat{b}_\Phi$, is the vector which minimizes equation (2.3).

One may again use ∇_b to determine $\hat{b}_\Phi$ as:

$$\hat{b}_\Phi = (A'\Phi^{-1}A)^{-1}A'\Phi^{-1}y \tag{2.4}$$

and furthermore show that R(b) in equation (2.3) is minimum when we use $\hat{b}_\Phi$ and $\Phi = E(x\ x')$, (i.e.,the "best" weighing matrix is the covariance matrix of the "noise"). Under these conditions the covariance matrix of the error, $b - \hat{b}$, is

$$\psi = (A'\ E(x\ x')^{-1}A)^{-1}\ .$$

Attention is now turned to the problem of a minimum variance unbiased estimator.

Theorem 2.3. Let $X_1, \cdots, X_n$ be n p-vector valued random variables with mean vector μ. Let $W_1, \cdots, W_n$ be n weighing matrices with dimensions $p \times p$. If $\Sigma W_i = I$, then $\tilde{X} = \Sigma W_i X_i$ is an unbiased estimator of μ.

Proof: $E\tilde{X} = E(\Sigma W_i X_i)$

$= \Sigma W_i EX_i$

$= \Sigma W_i \mu$

$= I\mu$

$= \mu$

i.e., $\tilde{X}$ is unbiased as an estimator of μ.

Theorem 2.4. If $X_1, \cdots, X_n$ are n uncorrelated p-vector valued random variables with mean vector μ and covariance matrices $\Phi_1, \cdots, \Phi_n$ and $W_1, \cdots, W_n$ are n $p \times p$ weighing matrices, then the covariance matrix of $\tilde{X} = \Sigma W_i X_i$, denoted by $\tilde{P}$ is given by $\tilde{P} = \Sigma W_i \Phi_i W_i'$.

Proof: By definition

$$\tilde{P} = E[(\tilde{X} - E\tilde{X})(\tilde{X} - E\tilde{X})'].$$

$$(\tilde{X} - E\tilde{X})(\tilde{X} - E\tilde{X})' = (\Sigma W_i X_i - \Sigma W_i EX_i)(\Sigma W_i X_i - \Sigma W_i EX_i)'$$

$$= (\Sigma W_i (X_i - EX_i))(\Sigma W_i (X_i - EX_i))'$$

$$= (\Sigma W_i (X_i - EX_i))(\Sigma (X_i - EX_i)' W_i')$$

$$= \Sigma W_i (X_i - EX_i)(X_i - EX_i)' W_i' + \sum_{i \neq j} \sum W_i (X_i - EX_i)(X_j - EX_j)' W_j'.$$

Hence

$$\tilde{P} = E(\tilde{X} - E\tilde{X})(\tilde{X} - E\tilde{X})'$$

$$= \Sigma W_i E(X_i - EX_i)(X_i - EX_i)' W_i' + \sum_{i \neq j} \sum W_i E(X_i - EX_i)(X_j - EX_j)' W_j'.$$

By definition, $\Phi_i = E(X_i - EX_i)(X_i - EX_i)'$. Since X_i and X_j are uncorrelated, $i \neq j$

$$E(X_i - EX_i)(X_j - EX_j)' = 0.$$

Hence

$$\tilde{P} = \Sigma W_i \Phi_i W_i' .$$

Corollary 2.1. If $X_1, \cdots, X_n$ are n uncorrelated p-vector valued random variables with mean vector μ and covariance matrices $\Phi_1, \cdots, \Phi_n$ where each is non-singular and $\hat{W}_1, \cdots, \hat{W}_n$ are defined by $\hat{W}_i = [\Sigma \Phi_j^{-1}]^{-1} \Phi_i^{-1}$, then $\hat{X} = \Sigma \hat{W}_i X_i$ is an unbiased estimator of μ and the covariance matrix of $\hat{X}$, denoted by $\hat{P}$, is given by

$$\hat{P} = [\Sigma \Phi_j^{-1}]^{-1}.$$

Proof: $[\Sigma \Phi_j^{-1}]^{-1}$ is well defined since each Φ_i is nonsingular and positive definite (i.e., the rank of $\hat{P}$ is p).

Since
$$\Sigma \hat{W}_i = \Sigma_i [\Sigma_j \Phi_j^{-1}]^{-1} \Phi_i^{-1}$$

$$= [\Sigma_j \Phi_j^{-1}]^{-1} [\Sigma_i \Phi_i^{-1}]$$

$$= I ,$$

$\hat{X}$ is an unbiased estimator of μ. (Theorem 2.3).

By Theorem 2.4,

$$\hat{P} = \Sigma \hat{W}_i \Phi_i \hat{W}_i'$$

$$= \Sigma_i \{\Sigma_j \Phi_j^{-1})^{-1}\} \Phi_i^{-1} \Phi_i (\Phi_i^{-1})' \{(\Sigma_j \Phi_j^{-1})^{-1}\}'$$

$$= \{(\Sigma_j \Phi_j^{-1})^{-1}\} \{\Sigma_i (\Phi_i^{-1})'\} \{(\Sigma_j \Phi_j^{-1})'\}^{-1}$$

$$= (\Sigma_j \Phi_j^{-1})^{-1} I = (\Sigma_j \Phi_j^{-1})^{-1}.$$

Before proof of the next theorem is presented, some new notation is introduced.

Let

$$\Phi = \begin{bmatrix} \Phi_1 & 0 & \cdots & 0 \\ 0 & \Phi_2 & & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \cdots & \Phi_n \end{bmatrix}^{[np\times np]}$$

and

$$I^* = (I\ I \cdots I)^{[p\times np]}$$

where I is the p × p identity matrix.

It is easily seen that

$$\Phi^{-1} = \begin{bmatrix} \Phi_1^{-1} & 0 & \cdots & 0 \\ 0 & \Phi_2^{-1} & & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \cdots & \Phi_n^{-1} \end{bmatrix}.$$

Let $W = (W_1, W_2, \cdots, W_n)^{[p\times np]}$

and $\hat{W} = (\hat{W}_1, \hat{W}_2, \cdots, \hat{W}_n)^{[p\times np]}$.

Now Theorem 2.3 states that if $\Sigma W_i = I$, then $\tilde{X}$ is unbiased. But

$$\Sigma W_i = WI^{*\prime}.$$

Notice also that in Theorem 2.3 and Theorem 2.4

$$\hat{P} = [\Sigma \Phi_j^{-1}]^{-1}$$

$$= (I^*\Phi^{-1}I^{*\prime})^{-1}$$

and $$\tilde{P} = W\Phi W'.$$

Recall also that if A is a positive definite symmetric matrix then A may be expressed as

$$A = A^{\frac{1}{2}}A^{\frac{1}{2}}$$

where $A^{\frac{1}{2}}$ is non-singular and $A^{\frac{1}{2}}$ inverse is denoted by $A^{-\frac{1}{2}}$. Also $A^{-\frac{1}{2}}A^{-\frac{1}{2}} = A^{-1}$ and $A^{\frac{1}{2}}$ is symmetric.

Corollary 2.2. If $\tilde{X} = \Sigma W_i X_i$ is an unbiased estimator of μ and each Φ_i is nonsingular then $WI^{*\prime} = I$ and $\tilde{P}$ is nonsingular.

Proof: Since $\tilde{X}$ is unbiased

$$\begin{aligned} E(\tilde{X} - \mu) &= E(\Sigma W_i X_i - \mu) \\ &= \Sigma W_i EX_i - \mu \\ &= \Sigma W_i \mu - \mu \\ &= (\Sigma W_i - I)\mu \equiv 0 \quad \text{for all } \mu. \end{aligned}$$

Hence $\Sigma W_i - I = 0$

and $\Sigma W_i = WI^{*\prime} = I.$

Since Φ is positive definite and symmetric

$$\Phi = \Phi^{\frac{1}{2}}\Phi^{\frac{1}{2}}$$

where $\Phi^{\frac{1}{2}}$ is nonsingular.

The rank of $\Phi^{\frac{1}{2}}W'$ equals the rank of W' which is the same as the rank of W.

Since $WI^{*\prime} = I$ and the rank $W \leq p$ we have

$$\begin{aligned} p = \text{rank } I &= \text{rank } WI^{*\prime} \\ &\leq \min(\text{rank } W, \text{rank } I^{*\prime}) \\ &\leq \text{rank } W \leq p. \end{aligned}$$

Hence rank W = p and rank $\Phi^{\frac{1}{2}}W'$ = p and thus

$$\tilde{P} = W\Phi W'$$
$$= W\Phi^{\frac{1}{2}}\Phi^{\frac{1}{2}}W'$$
$$= (W\Phi^{\frac{1}{2}})(W\Phi^{\frac{1}{2}})'$$

and has rank p.

The following theorem is stated without proof [5, p. 62-63].

Theorem 2.5. If F and G are m × g matrices where m $\geq$ g and rank F and rank G is g then,

$$F'F \geq (G'F)'(G'G)^{-1}(G'F)$$

where a positive definite matrix A is less than or equal to a positive definite matrix B denoted A $\leq$ B if and only if B - A is non-negative definite.

Theorem 2.6. Let $X_1, \cdots, X_n$ be random vectors as in Corollary 2.1. $\hat{X}$ is the minimum variance unbiased estimator of μ in the class of unbiased estimators of the form $\tilde{X} = \Sigma W_i X_i$ (i.e., $\hat{P} \leq \tilde{P}$ for all choices of $\tilde{X}$).

Proof: Let $\tilde{X} = \Sigma W_i X_i$. Since $\tilde{X}$ is unbiased

$$WI^{*\prime} = I = (I^{*}W')'.$$

In Theorem 2.5, let $F' = W\Phi^{\frac{1}{2}} = (\Phi^{\frac{1}{2}}W')'$ and $G = \Phi^{-\frac{1}{2}}I^{*\prime}$ $= (I^{*}\Phi^{-\frac{1}{2}})'$.

F is np × p and G is np × p and each is of rank p and np $\geq$ p for all n. Since

$$F'F = \tilde{P}$$

and

$$G'G = \tilde{P}^{-1}$$

and

$$\text{rank } (F'F) = \text{rank } F$$

and rank $(G'G)$ = rank G ,

thus both G and F are of rank p.

But the inequality in the previous theorem yields:

$$(F'F) = (W\Phi^{\frac{1}{2}}\Phi^{\frac{1}{2}}W') = W\Phi W' = \tilde{P}$$
$$\geq (I^*\Phi^{-\frac{1}{2}}\Phi^{\frac{1}{2}}W')'(I^*\Phi^{-\frac{1}{2}\ -\frac{1}{2}}I^{*\prime})^{-1}(I^*\Phi^{-\frac{1}{2}}\Phi^{\frac{1}{2}}W')$$
$$= (I^*W')'(I^*\Phi^{-1}I^{*\prime})^{-1}(I^*W')$$
$$= (I^*\Phi^{-1}I^{*\prime})^{-1} = \hat{P} .$$

The minimum variance unbiased estimator of μ may be expressed as

$$\hat{X} = [I^*\Phi^{-1}I^{*\prime}]^{-1}I^*\Phi^{-1}X.$$

The following theorem is a well-known result.

Theorem 2.7. Let $X_1, \cdots, X_n$ be random p-vectors which are uncorrelated and have mean vector μ. The least square estimator of μ is given by

$$\overline{X} = \frac{1}{n}\Sigma X_i.$$

$\overline{X}$ is unbiased and has covariance matrix $\overline{P}$ given by

$$\overline{P} = (I^*I^{*\prime})^{-1}I^*\Phi I^{*\prime}(I^*I^{*\prime})^{-1}$$
$$= (\frac{1}{n}\ I)I^*\Phi I^{*\prime}(\frac{1}{n}\ I)$$
$$= \frac{1}{n^2}\Sigma\Phi_i .$$

From the theory of weighted least squares the following result is obtained.

Theorem 2.8. If $X_1, \cdots, X_n$ are n uncorrelated p-vector valued random variable with mean vector μ and non-singular covariance matrices $\Phi_1, \cdots, \Phi_n$ respectively, then the weighted least square estimator for μ which has the smallest covariance matrix is $\hat{X}$.

Hence the linear minimum variance unbiased estimator and the weighted least squares estimator coincide.

Chapter III: Application of L. S. and M. V. U. Estimators to Trajector Estimation

The least square (L. S.) estimator, the minimum variance unbiased (M. V. U.) estimator, and the Kalman filter will be applied to a parabolic trajectory. It is known that if the system defining the trajectory is known, that is, the transition matrix is given in closed form, then the M. V. U. estimator and Kalman's filter are equivalent [5]. Thus, for large samples one would expect that under conditions for which all three methods of estimation are applicable the estimators would yield essentially the same results. If the transition matrix associated with the trajectory is approximated, then an analysis of the estimation results using the three methods will be useful.

For this analysis the following assumptions are made:

(1) Assume that at each observation time t_i one is able to obtain independent unbiased estimates of the state vector with a known covariance matrix.

(2) Assume that the transition matrix is known and has the standard properties [6].

The problem then considered is to estimate the state vector by the three methods mentioned previously at some later time t_f.

Let $\hat{x}(t_i)$, $i = 0, 1, 2, \cdots, n$, where $t_n \leq t_f$, be an unbiased estimator of the state vector $x(t_i)$ with nonsingular covariance matrix $\mathrm{Cov}(\hat{x}(t_i))$. Since Φ is a transition matrix with nonrandom elements then it can be shown that

$$\hat{x}^{(1)}(t_f) = \Phi(t_f;\ t_1)\hat{x}(t_1) \tag{3.1}$$

is an unbiased estimator of the state vector at time t_f, $x(t_f)$. The covariance matrix of $x^{(1)}(t_f)$ is given by

$$\Phi(t_f;\ t_1)\mathrm{Cov}(\hat{x}(t_1))\Phi'(t_f;\ t_1). \tag{3.2}$$

From Theorem 2.7 it follows that the L. S. estimator is

$$\hat{x}_L(t_f) = \frac{1}{n+1}\sum_{i=0}^{n}\Phi(t_f;\ t_1)\hat{x}(t_1). \tag{3.3}$$

It also is seen that the covariance matrix of $\hat{x}_L(t_f)$ is

$$\mathrm{Cov}(\hat{x}_L(t_f)) = (n+1)^{-2}\sum_{i=0}^{n}\Phi(t_f;\ t_1)\mathrm{Cov}(\hat{x}(t_1))\Phi'(t_f;\ t_1). \tag{3.4}$$

Theorem 2.6 yields the M. V. U. estimator of $x(t_f)$

$$\hat{x}_M(t_f) = \left[\sum_{i=0}^{n}[\Phi(t_f;\ t_1)\mathrm{Cov}(\hat{x}(t_1))\Phi'(t_f;\ t_1)]^{-1}\right]^{-1}$$

$$\times \sum_{i=0}^{n}[\Phi(t_f;\ t_1)\mathrm{Cov}(\hat{x}(t_1))\Phi'(t_f;\ t_1)]^{-1}\hat{x}^{(i)}(t_f), \tag{3.5}$$

where $\hat{x}^{(i)}(t_f)$ is given by equation (3.1). The covariance matrix of $\hat{x}_M(t_f)$ is

$$\mathrm{Cov}(\hat{x}_M(t_f)) = \left[\sum_{i=0}^{n}[\Phi(t_f;\ t_1)\mathrm{Cov}(\hat{x}(t_1))\Phi'(t_f;\ t_1)]^{-1}\right]^{-1}. \tag{3.6}$$

It must be emphasized that these equations are valid only when the transition matrix Φ is exactly known. Computer oriented flow charts of these two estimators (L. S. and M. V. U.) as well as one of Kalman's filter will appear in Chapter IV.

Chapter IV: The Mathematical Model

The basic mathematical model is that of a particle in force free flight on a parabolic curve in the x-y plane. With this model, the estimation techniques are used to estimate the position and velocity of the particle (i.e., the state of the dynamic system).

The following standard notation will be needed;

$$\dot{x} = dx/dt,$$
$$\dot{y} = dy/dt,$$
$$\ddot{x} = d^2x/dt^2,$$
$$\ddot{y} = d^2y/dt^2,$$
$$y^{(n)} = d^ny/dt^n. \qquad n = 1, 2, 3, \cdots$$

The basic parametric equations which describe the motion of the particle are chosen to be:

$$\begin{aligned} x(t) &= 100 + at, \\ y(t) &= 100 + b\sqrt{t}, \end{aligned} \qquad t > 0 \qquad (4.1)$$

where a and b may be considered as orbital parameters and the independent variable t is considered as time. From equation (4.1),

$$\begin{aligned} \dot{x} &= a, \\ \dot{y} &= b/(2\sqrt{t}), \\ \ddot{x} &= 0, \\ \ddot{y} &= -b/(4t^{3/2}) = -y/(2t). \end{aligned} \qquad (4.2)$$

Choosing the components of the system to be the quantities $x_1(t) = x(t)$, $x_2(t) = y(t)$, $x_3(t) = \dot{x}(t)$, and $x_4(t) = \dot{y}(t)$; equations (4.1) and (4.2) produce a force free linear, time variant dynamic system

$$\dot{x}(t) = F(t)\, x(t) \tag{4.3}$$

where

$$x(t) = (x_1(t),\ x_2(t),\ x_3(t),\ x_4(t))' = (x(t),\ y(t),\ \dot{x}(t),\ \dot{y}(t))',$$

$$\dot{x}(t) = (\dot{x}_1(t),\ \dot{x}_2(t),\ \dot{x}_3(t),\ \dot{x}_4(t))' = (\dot{x}(t),\ \dot{y}(t),\ \ddot{x}(t),\ \ddot{y}(t))',$$

and

$$F(t) = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1/(2t) \end{bmatrix}. \tag{4.4}$$

It is interesting to note that F(t) does not depend on the orbital parameters a and b.

The next problem is to determine the state transition matrix, that is the matrix, $\Phi(t;\ t_0)$, that relates the state vector at time t_0 to the state vector at time t. More precisely one needs to determine a matrix $\Phi(t;\ t_0)$ such that $x(t) = \Phi(t;\ t_0)\ x(t_0)$.

This problem is fundamental in the theory of dynamic systems and can generally be solved in several ways, some more appropriate than others. The initial step in determining the transition matrix is to solve the system of differential equations given in equation (4.3). In many cases an approximate solution must be obtained by numerical methods; but, in this example, an analytic solution is easily obtained.

The general solution of equation (4.3) is

$$x(t) = A(t)c, \quad \text{for all } t > 0 \tag{4.5}$$

where

$$A(t) = \begin{bmatrix} t & 1 & 0 & 0 \\ 0 & 0 & 2t^{\frac{1}{2}} & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & t^{-\frac{1}{2}} & 0 \end{bmatrix} \tag{4.6}$$

and $c' = (c_1, c_2, c_3, c_4)$, a vector of arbitrary constants.

Given an initial condition $x(t_0)$ for the state vector, the technique is to solve for the arbitrary constant c and thus, eliminate c in equation (4.5).

Since A(t) is non-singular for $t > 0$, the inverse of A(t) exists, i.e.,

$$A^{-1}(t) = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & -t & 0 \\ 0 & 0 & 0 & t^{\frac{1}{2}} \\ 0 & 1 & 0 & -2t \end{bmatrix}. \tag{4.7}$$

In equation (4.5) allow t to be equal to t_0 and multiply on the left by $A^{-1}(t_0)$ to obtain

$$c = A^{-1}(t_0)x(t_0). \tag{4.8}$$

Substituting equation (4.8) into equation (4.5) gives

$$x(t) = A(t)A^{-1}(t_0)x(t_0). \tag{4.9}$$

Equation (4.9) shows that $A(t)A^{-1}(t_0)$ is a transition matrix.

Let

$$\Phi(t;\ t_0) = A(t)A^{-1}(t_0)$$

$$= \begin{bmatrix} 1 & 0 & t-t_0 & 0 \\ 0 & 1 & 0 & 2(t\ \ t_0)^{\frac{1}{2}}-2t_0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & (t_0/t)^{\frac{1}{2}} \end{bmatrix}. \qquad (4.10)$$

Thus, $x(t) = \Phi(t;\ t_0)x(t_0)$. (4.11)

At this stage the transition matrix given in equation (4.10) is limited to relating the state vector at t_0 to the state vector at time t. In using the transition matrix it is desirable to have $\Phi(t_2;\ t_1)$, $t_2 \geq t_1 > 0$, that is, a matrix satisfying the relationship

$$x(t_2) = \Phi(t_2;\ t_1)x(t_1).$$

Next the expression for $\Phi(t_2;\ t_1)$ is derived. By equation (4.11)

$$x(t_2) = \Phi(t_2;\ t_0)x(t_0)$$

and $$x(t_1) = \Phi(t_1;\ t_0)x(t_0).$$

Since $\Phi^{-1}(t;\ t_0)$ exists

$$x(t_0) = \Phi^{-1}(t_1;\ t_0)x(t_1).$$

Substitution gives

$$x(t_2) = \Phi(t_2;\ t_0)\Phi^{-1}(t_1;\ t_0)x(t_1)$$
$$= \Phi(t_2;\ t_0)\Phi(t_0;\ t_1)x(t_1).$$

Furthermore,

$$\Phi(t_2;\, t_1) = \Phi(t_2;\, t_0)\Phi(t_0;\, t_1)$$

$$= A(t_2)A^{-1}(t_0)A(t_0)A^{-1}(t_1)$$

$$= A(t_2)A^{-1}(t_1)$$

$$= \begin{bmatrix} 1 & 0 & t_2-t_1 & 0 \\ 0 & 1 & 0 & 2(t_1t_2)^{\frac{1}{2}}-2t_1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & (t_1/t_2)^{\frac{1}{2}} \end{bmatrix}. \qquad (4.12)$$

Reviewing, the particle in force free flight is represented by the set of differential equations given in equation (4.3), where F(t) is given in equation (4.4). The exact state transition matrix for any two times t_2 and t_1 is given in equation (4.12).

For comparison purposes $\Phi_k(t_2;\, t_1)$, i.e., approximations of the state transition matrix, are derived for integer values of k. The two main tools used in this development are Taylor series expansions and a recursive relation between $y^{(n)}(t)$ and $\dot{y}(t)$.

Notice in equation (4.2), $\ddot{x}(t) = 0$ and consequently, all the higher derivatives of x(t) with respect to the independent variable t will be zero. This is not true for y(t).

In equation (4.2)

$$y^{(2)}(t) = -y^{(1)}(t)/(2t)$$

$$= (-\tfrac{1}{2})^{2-1}(1)t^{-(2-1)}y^{(1)}(t). \qquad (4.13)$$

It follows from equation (4.2) that

$$y^{(3)}(t) = (3/2)(-\tfrac{1}{4})t^{-5/2}$$

$$= (-\tfrac{1}{2})^{3-1}(3)(1)t^{-(3-1)}y^{(1)}(t).$$

By mathematical induction, it follows that for $n \geq 2$

$$y^{(n)}(t) = (-\tfrac{1}{2})^{n-1}(2n-3)(2n-5) \cdot \cdot \cdot (3)(1)t^{-(n-1)}y^{(1)}(t). \quad (4.14)$$

Equation (4.14) represents a recursive formula for the n^{th} derivative of y(t) with respect to the independent variable in terms of $y^{(1)}(t)$. This formula will be used in the Taylor expansion of y(t).

One form of the Taylor series expansion is:

$$f(x + h) = \sum_{n=0}^{\infty} h^n f^{(n)}(x)/n! \quad .$$

Since y(t) satisfies the hypothesis of Taylor's Theorem, for $h = (t_2 - t_1)$

$$y(t_2) = y(t_1 + h)$$

$$= \sum_{n=0}^{\infty} (t_2 - t_1)^n y^{(n)}(t_1)/n! \quad . \quad (4.15)$$

Substituting equation (4.14) into equation (4.15) gives

$$y(t_2) = y(t_1) + (t_2 - t_1)y^{(1)}(t_1) + \sum_{n=2}^{\infty}(t_2 - t_1)^n(-1)^{n-1}2^{-2n+1}$$

$$\binom{2n}{n}(2n - 1)^{-1}t_1^{-(n-1)} \quad (4.16)$$

For convenience equation (4.16) will be written as

$$y(t_2) = y(t_1) + (t_2 - t_1)y^{(1)}(t_1) + y^{(1)}(t_1)S(t_2; t_1), \quad (4.17)$$

where

$$S(t_2; t_1) = t_1 \sum_{n=2}^{\infty}\left(\frac{t_2 - t_1}{t_1}\right)^n (-1)^{n-1}2^{-2n+1}\binom{2n}{n}(2n-1)^{-1} \quad . \quad (4.18)$$

It is easy to verify by D'Alembert's ratio test that this series is absolutely convergent for $t_2 < 2t_1$ and $t_1 < t_2$.

Analogous to equation (4.17), the relationship in equation (4.14) gives a Taylor series expansion

$$\dot{y}(t_2) = \dot{y}(t_1)R(t_2;\ t_1) \tag{4.19}$$

where

$$R(t_2;\ t_1) = 1 + \sum_{n=1}^{\infty}\left(\frac{t_2 - t_1}{t_1}\right)^n (-1)^{n_2-2n}\binom{2n}{n} \ . \tag{4.20}$$

$R(t_2;\ t_1)$ is absolutely convergent for $t_2 < 2t_1$ and $t_1 < t_2$.

Define for $k \geq 1$

$$S_k(t_2;\ t_1) = t_1 \sum_{n=1}^{k}\left(\frac{t_2 - t_1}{t_1}\right)^n (-1)^{n-1_2-2n+1}\binom{2n}{n}(2n-1)^{-1} \tag{4.21}$$

and

$$R_k(t_2;\ t_1) = 1 + \sum_{n=1}^{k}\left(\frac{t_2 - t_1}{t_1}\right)^n (-1)^{n_2-2n}\binom{2n}{n} \ . \tag{4.22}$$

One can approximate $y(t_2)$ to any degree of numerical accuracy by equations (4.15) and (4.21) with an appropriate choice of k. Likewise $\dot{y}(t_2)$ can be approximated as close as desired.

The approximate transition matrix is defined as

$$\Phi_k(t_2;\ t_1) = \begin{bmatrix} 1 & 0 & t_2-t_1 & 0 \\ 0 & 1 & 0 & S_k(t_2;t_1) \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & R_k(t_2;t_1) \end{bmatrix} . \tag{4.23}$$

It should be noted that the approximate Φ matrix $\Phi_k(t_2; t_1)$ does not have all of the properties of the exact transition matrix. For instance, in the approximate matrix,

$$\Phi_k(t_2; t_0) \neq \Phi_k(t_2; t_1)\Phi_k(t_1; t_0),$$

and also; due to the approximation the transition matrices Φ_k may be singular.

Figures 1, 2, and 3 are computer orientated flow charts for the L. S. estimator, M. V. U. estimator, and C. B. Solloway's solution of the Kalman problem, respectively.

The estimates $\hat{x}(t_1)$ in Figures 1 and 2 are generated by adding to the actual states at time t_i normal (Gaussian) random numbers with zero means. The variance of the random number added to the jth components of $u(t_i)$ is given by the jth diagonal elements in $Q(t_i)$. The "observations" $y(t_n)$ in Figure 3 are generated the same way.

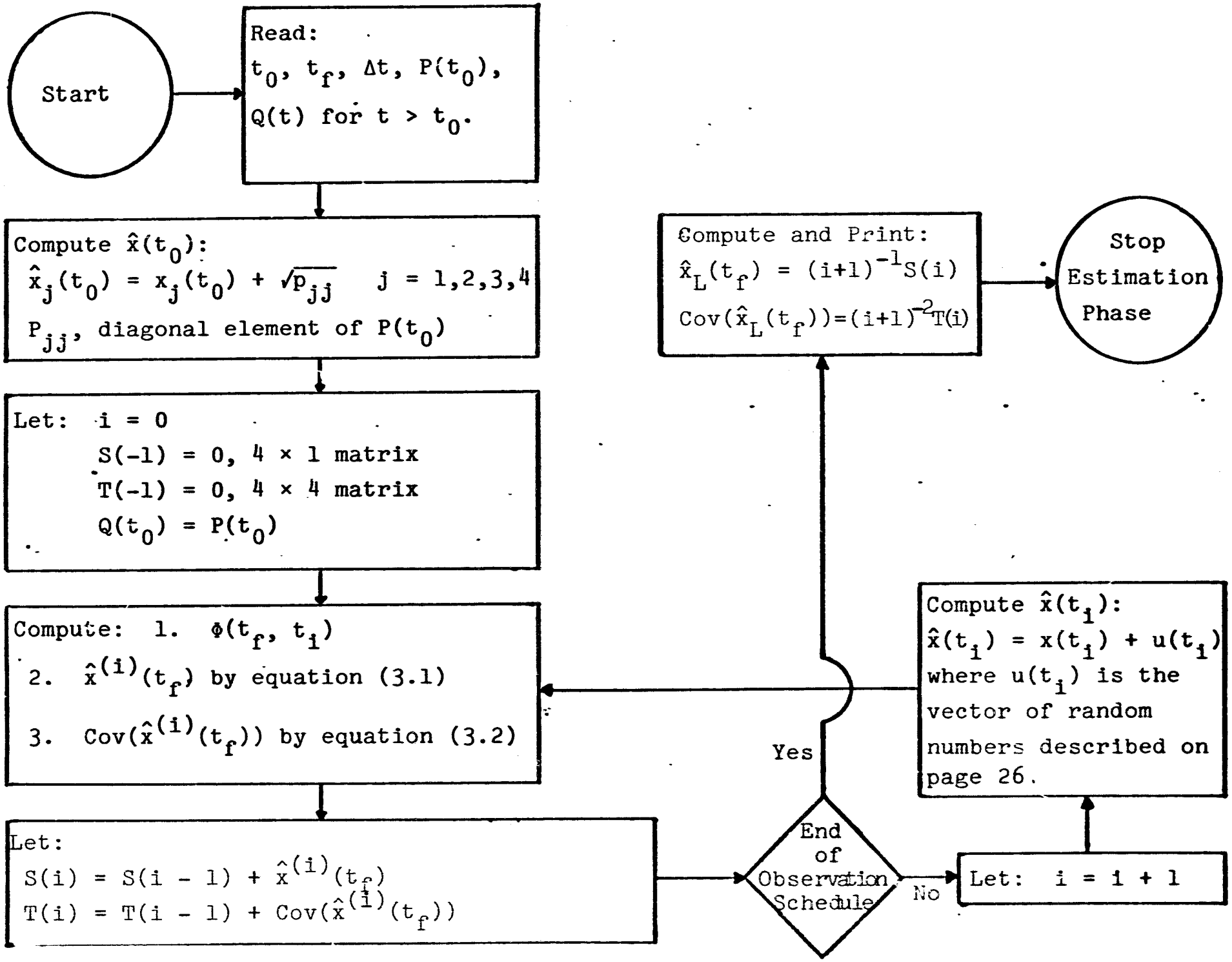


Figure 1: Flow Chart for the L. S. Estimation Given in Equation (3.3).

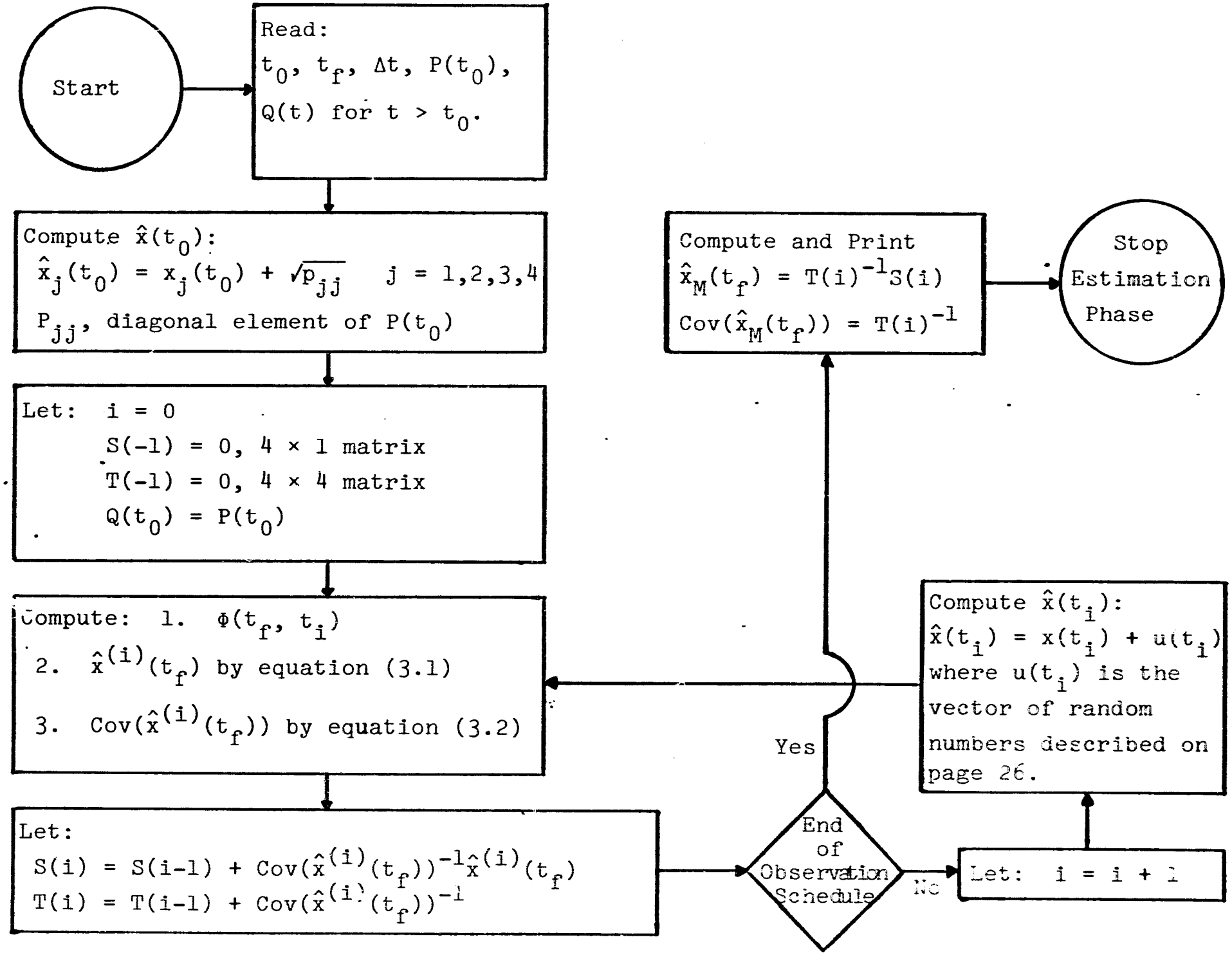


Figure 2: Flow Chart for the M. V. U. Estimator Given in Equation (3.5).

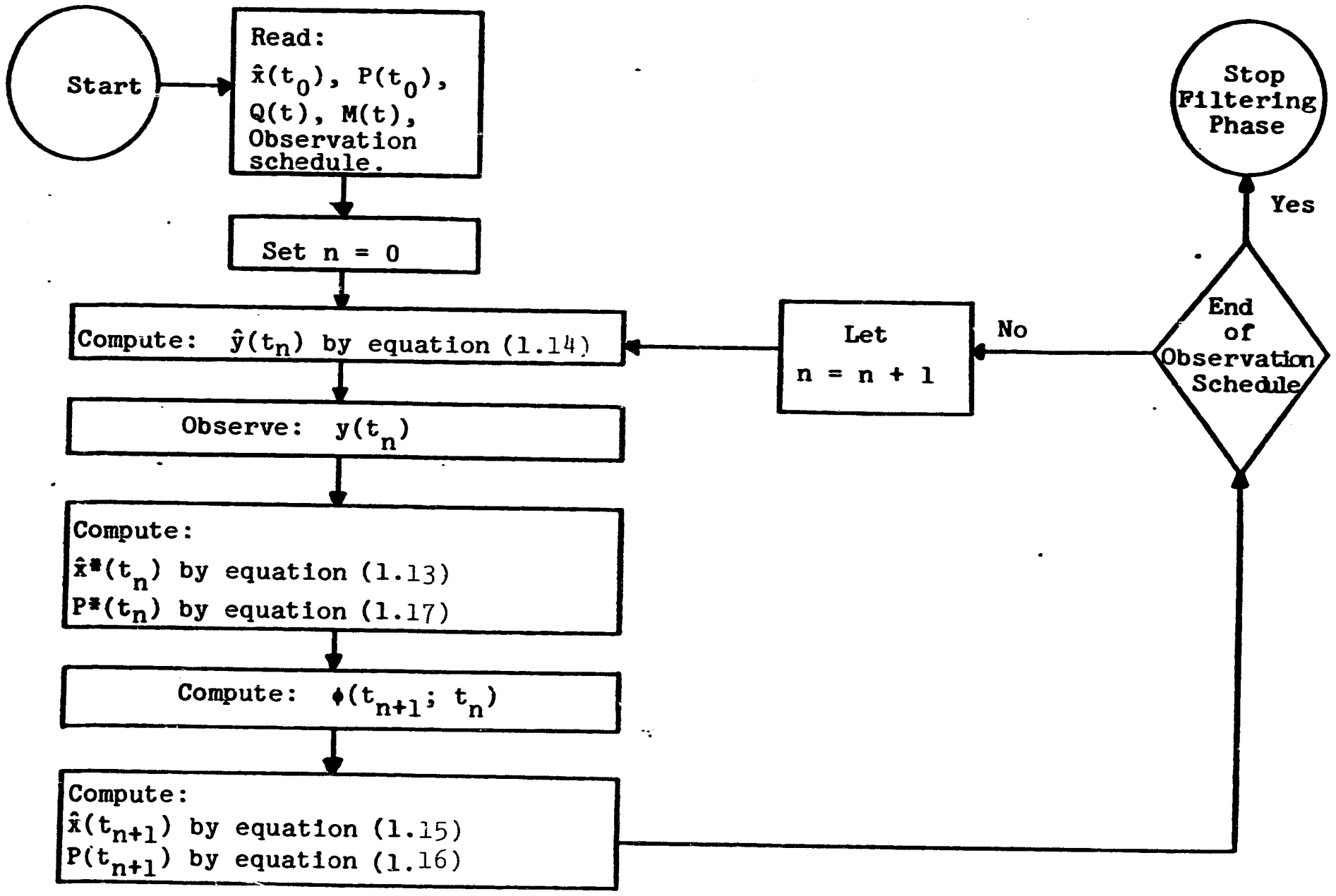


Figure 3. Flow Chart for the Solution of the Model Given in Equations (1.11) and (1.12).

Chapter V: Simulation Details and Results

The dynamical system stated in the previous section is now particularized to generate a sequence of simulation runs. Since the parametric equations given by (4.1) describes a family of trajectories depending on the parameters a and b; a particular selection of a = -25 and b = 100 is specified. This choice determines a reasonable time, t = 4, in which the trajectory crosses the vertical axis. The sketch of this actual trajectory is given in Figure 4.

Now that we have a particular dynamical system the simulation requires observations of this system with noise contaminations. These observations are generated by altering the actual states at time t with additive error. This noise is prescribed by previous requirements to be independent Gaussian with zero mean.

In the simulation program, the covariance matrix for noise Q(t) is

$$Q(t) = \text{diag}\{1.0,\ 1.0,\ .81,\ .81\}. \tag{5.1}$$

The initial time t_0 is taken to be .25 and time t_f is taken to be 16 or 16.25 depending on the step size. In general the matrix Q(t) could be considered as a function of time t but this simulation treats the noise as time invariant.

The original estimate is assumed to have certain statistical properties. In this simulation the covariance matrix, $P(t_0)$, of the error is

$$P(t_0) = \text{diag}\{1.0,\ 1.0,\ 1.0,\ 1.0\}\ . \tag{5.2}$$

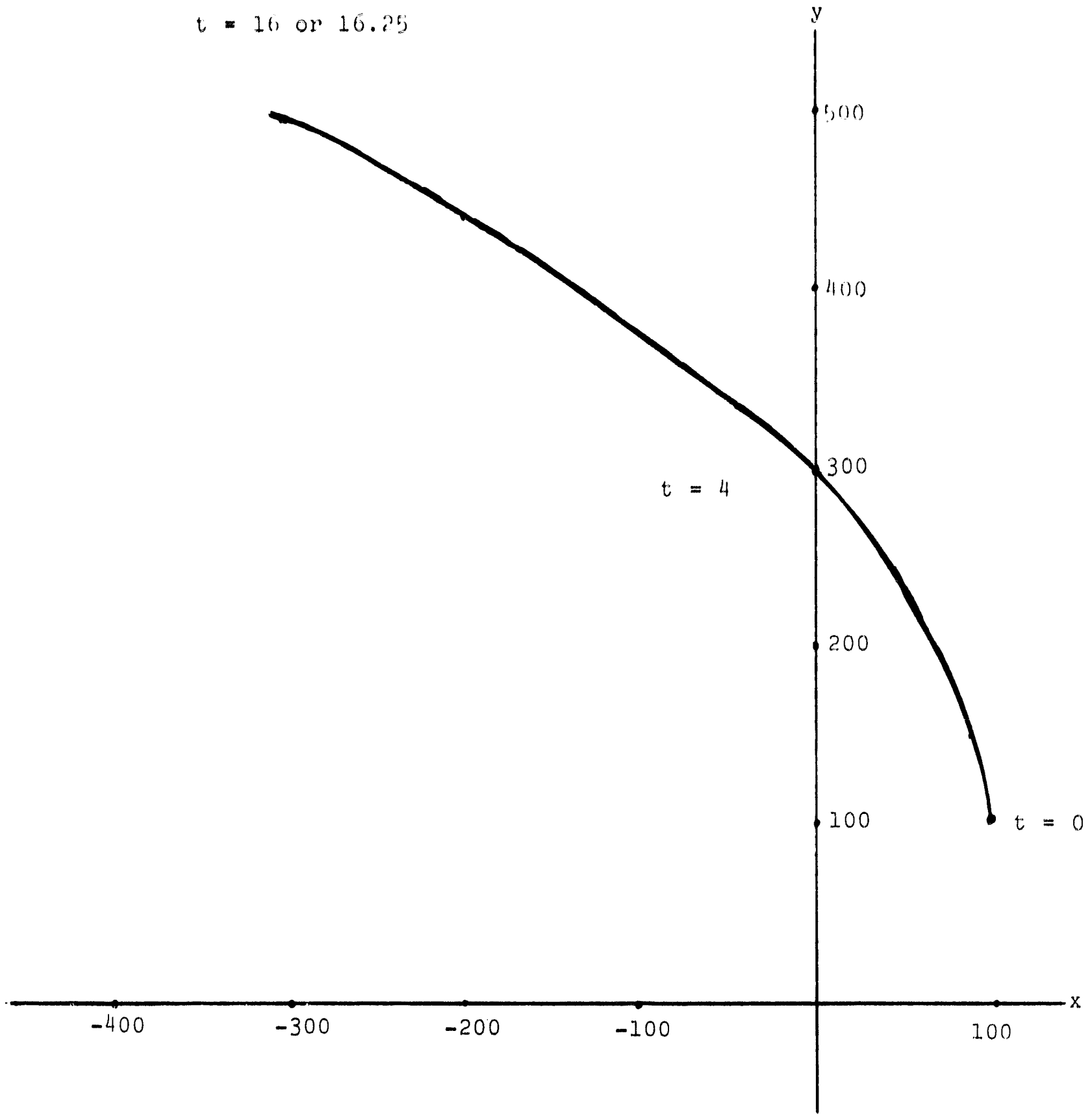


Figure 4. Trace of Model's Actual Trajectory.

The initiation of the computer program for the Kalman filter requires the initial estimate, denoted by $\hat{x}(t_0)$, of the state vector at time t_0. In this simulation $\hat{x}(t_0)$ was taken to be the actual value of the state at time t_0 plus one standard deviation. This standard deviation is obtained from the diagonal elements of the P matrices given by (5.2). The initial estimate vector is

$$\hat{x}(t_0) = \begin{bmatrix} 94.75 \\ 151.0 \\ -24 \\ 101 \end{bmatrix} . \qquad (5.3)$$

To compare the Kalman filter with the L. S. and M. V. U. estimators the mapping matrix M needs to be the identity matrix. In general a more elaborate observation structure for the Kalman filter may be treated.

For the observation schedule variations were introduced in order that we could observe the effects of different estimation techniques. These simulations contrast large step size and small step size.

The large steps were generated by adding a time increment $\Delta t = .5$ to the previous time to give a sequence of observations with large time increments, i.e., $t_n = t_{n-1} + .5$ which implies $t_f = 16.25$. The small steps were generated by adding a time increment $\Delta t = .25$, i.e., $t_n = t_{n-1} + .25$ which implies $t_f = 16$.

The major analysis in this simulation is the comparison of the different estimation schemes using various approximations of the state transition matrix Φ. The results generated by the approximate Φ's for each estimation scheme were contrasted with the results of the exact Φ. The approximate transition matrices selected were Φ_1, Φ_2, and Φ_5, given by (4.23).

This simulation was not a Monte Carlo-type study. All of the results obtained utilized the same set of random numbers as the additive error in the observation vector.

These variations give eight computer runs for each estimation scheme. There are four alternate selections of the Φ matrices, i.e., Φ, Φ_1, Φ_2, and Φ_5; and two variations in the observation schedule, i.e., large time increments and small time increments. The complete set of twenty-four possible combinations were investigated. The programming and computing were performed at MSFC, Huntsville, Alabama.

The results of these simulations are given in Tables 1, 2, 3, 4, 5, 6, 7, and 8. In these tables $xA(t_f)$ denotes the actual state of the dynamical system. An analysis of these results is given in Chapter 6.

Table 1; Φ; $\Delta t = 0.25$; $t_f = 16.0$.

	$x_A(t_f)$	$\hat{x}_L(t_f)$	$\hat{x}_M(t_f)$	$\hat{x}^*(t_f)$
x	-300.00	-298.04	-310.49	-299.54
y	500.00	501.42	500.39	500.53
$\dot{x}$	- 25.00	- 24.88	- 28.36	- 24.96
$\dot{y}$	12.50	12.57	12.49	12.53

diag $\mathrm{Cov}(\hat{x}_L(t_f)) = (1.08, 0.45, 0.01, 0.006)$

diag $\mathrm{Cov}(\hat{x}_M(t_f)) = (-0.64, 0.01, -0.06, -0.001)$

diag $\mathrm{Cov}(\hat{x}^*(t_f)) = (0.06, 0.04, 0.007, 0.002)$

Table 2; Φ_1; $\Delta t = 0.25$; $t_f = 16.0$.

	$xA(t_f)$	$\hat{x}_L(t_f)$	$\hat{x}_M(t_f)$	$\hat{x}^*(t_f)$
x	-300.00	-298.04	-299.60	-299.54
y	500.00	602.31	371.46	491.48
$\dot{x}$	- 25.00	- 24.88	- 24.98	- 24.96
$\dot{y}$	12.50	- 84.92	- 5.35	11.38

diag Cov $(\hat{x}_L(t_f))$ = (1.08, 1.08, 0.01, 0.31)

diag Cov $(\hat{x}_M(t_f))$ = (0.06, 0.03, 0.001, .001)

diag Cov $(\hat{x}^*(t_f))$ = (0.06, 0.04, 0.001, 0.0002)

Table 3; Φ_2; $\Delta t = 0.25$; $t_f = 16.0$.

	$xA(t_f)$	$\hat{x}_L(t_f)$	$\hat{x}_M(t_f)$	$\hat{x}^*(t_f)$
x	-300.00	-298.03	-299.49	-299.54
y	500.00	-202.84	369.07	502.06
$\dot{x}$	- 25.00	- 24.88	- 24.95	- 24.96
$\dot{y}$	12.50	2,995.35	- 4.78	12.71

diag $Cov(\hat{x}_L(t_f))$ = (1.08, 16.98, 0.01, 548.82)

diag $Cov(\hat{x}_M(t_f))$ = (0.06, 0.04, 0.001, 0.002)

diag $Cov(\hat{x}^*(t_f))$ = (0.06, 0.04, 0.001, 0.0002)

Table 4; Φ_5; $\Delta t = 0.25$; $t_f = 16.0$.

	$xA(t_f)$	$\hat{x}_L(t_f)$	$\hat{x}_M(t_f)$	$\hat{x}^*(t_f)$
x	-300.00	-298.04	-298.82	-299.54
y	500.00	760.43	157.79	493.10
$\dot{x}$	- 25.00	- 24.88	- 24.80	- 24.96
$\dot{y}$	12.50	-3684.24	- 23.49	11.56

diag $Cov(\hat{x}_L(t_f))$ = (1.08, 4.78, 0.01, 1279.99)

diag $Cov(\hat{x}_M(t_f))$ = (0.04, -0.04, -0.001, -0.0003)

diag $Cov(\hat{x}^*(t_f))$ = (0.06, 0.04, 0.001, 0.0002)

Table 5; Φ; $\Delta t = 0.50$; $t_f = 16.25$.

	$xA(t_f)$	$\hat{x}_L(t_f)$	$\hat{x}_M(t_f)$	$\hat{x}^*(t_f)$
x	-306.25	-303.78	-305.27	-305.81
y	503.11	504.62	503.38	503.16
$\dot{x}$	- 25.00	- 24.69	- 24.85	- 24.95
$\dot{y}$	12.40	12.54	12.40	12.41

diag $Cov(\hat{x}_L(t_f)) = (2.20, 0.88, 0.02, 0.01)$

diag $Cov(\hat{x}_M(t_f)) = (0.16, 0.14, 0.004, 0.003)$

diag $Cov(\hat{x}^*(t_f)) = (0.11, 0.08, 0.001, 0.0004)$

Table 6; Φ_1; $\Delta t = 0.50$; $t_f = 16.25$.

	$xA(t_f)$	$\hat{x}_L(t_f)$	$\hat{x}_M(t_f)$	$\hat{x}^*(t_f)$
x	-306.25	-303.78	-306.78	-305.81
y	503.11	617.43	310.01	369.20
$\dot{x}$	- 25.00	- 24.69	- 25.00	- 24.95
$\dot{y}$	12.40	-120.60	5.43	0.00

diag $Cov(\hat{x}_L(t_f)) = (2.20, 2.20, 0.02, 0.99)$

diag $Cov\ (\hat{x}_M(t_f)) = (0.13, 0.02, 0.001, -0.0005)$

diag $Cov(\hat{x}^*(t_f)) = (0.11, 0.03, 0.001, 0.000)$

Table 7; Φ_2; $\Delta t = 0.50$; $t_f = 16.25$.

	$xA(t_f)$	$\hat{x}_L(t_f)$	$\hat{x}_M(t_f)$	$\hat{x}^*(t_f)$
x	-306.25	-303.78	-305.59	-305.81
y	503.11	-492.60	382.70	508.65
$\dot{x}$	- 25.00	- 24.69	- 24.91	- 24.95
$\dot{y}$	12.40	5,009.43	- 3.17	13.10

diag $Cov(\hat{x}_L(t_f))$ = (2.20, 57.31, 0.02, 2099.21)

diag $Cov(\hat{x}_M(t_f))$ = (0.12, 0.09, 0.002, 0.005)

diag $Cov(\hat{x}^*(t_f))$ = (0.11, 0.08, 0.001, 0.0004)

Table 8; Φ_5; $\Delta t = 0.50$; $t_f = 16.25$.

	$xA(t_f)$	$\hat{x}_L(t_f)$	$\hat{x}_M(t_f)$	$\hat{x}^*(t_f)$
x	-306.25	-303.78	-385.70	-305.81
y	503.11	969.21	-1728.49	421.99
$\dot{x}$	- 25.00	- 24.69	- 32.02	- 24.95
$\dot{y}$	12.40	-7631.99	- 83.49	4.43

diag $Cov(\hat{x}_L(t_f)) = (2.20, 18.86, 0.02, 5684.46)$

diag $Cov(\hat{x}_M(t_f)) = (0.23, -1.98, 0.002, -0.005)$

diag $Cov(\hat{x}^*(t_f)) = (0.11, 0.04, 0.001, 0.0001)$

Chapter VI: Summary and Future

It is a well known fact that diagonal elements of covariance matrices, that is, the variances are non-negative. When approximate transition matrices are used, it is not surprising that the covariance matrix of an estimator yields negative variances. This situation arises due to the fact that the approximate transition matrix may be singular at some time steps. For example, in Table 6 the estimate of $\dot{y}$ is zero and the variance of the estimate is zero. This resulted from the fact $\Phi_1(0.25; 0.25)$ is singular.

In Table 1 which uses the exact transition matrix with time step $\Delta t = .25$ the covariance matrix for the M. V. U. estimator exhibits negative components. The reason for this occurrence at this time is unknown. In Table 5 which gives the results for the exact transition matrix with time step $\Delta t = .50$ the covariance matrix for the M. V. U. estimator is well behaved. Thus further investigation is needed.

The most important conclusion drawn from this study is that for this linear model the Kalman filter is the best estimation scheme for estimating the actual state vector. In particular in Table 8; Φ_5; $\Delta t = .50$; $t_f = 16.25$ note that $\hat{x}_L(t_f)$ and $\hat{x}_M(t_f)$ are very poor estimates while $\hat{x}^*(t_f)$ was closer to the actual state vector.

Another conclusion is that the covariance matrix for the L. S. estimator reflected the fact that the estimations were poor. This is expected when the exact transition matrix is used but somewhat surprising for approximate transition matrices.

As it was previously mentioned this was not a Monte Carlo study. The same set of random variables were used for each estimation scheme. Thus, a future task could be to compare the estimators for many different sets of random variables.

Another future task would be to apply this study to a nonlinear model.

Appendix I: BASIC STATISTIC

Throughout this appendix the following notations will be used.

E	mathematical expectation.
X	(with or without subscripts) a vector valued random variable.
V	covariance matrix.
A, B	matrices of constants.

FACT 1: If $X_1, X_2, \ldots, X_n$ are vector valued random variables of the same dimension, then

$$E[\sum_1^n X_i] = \sum_1^n EX_i .$$

FACT 2: If X is a m × 1 vector valued random variable and A and B are matrices of constants with dimensions n × m and m × 1 respectively, then

$$E[AX + B] = AEX + B.$$

FACT 3: If V is a covariance matrix of a non-degenerate random vector X, then V is positive definite and symmetric.

NOTE: $V = E(X-EX)(X-EX)'$ where ' denotes the transpose.

FACT 4: Let X be a p × 1 random vector with mean vector μ and nonsingular covariance matrix V. If B is a p × p nonsingular matrix, then Y = BX has mean vector Bμ and covariance matrix BVB'.

Appendix II: BASIC LINEAR ALGEBRA

Throughout this appendix the following notations will be used.

A,B,V	(with or without subscripts) matrices.
A'	transpose of matrix A.
A^{-1}	inverse of matrix A. (whenever A^{-1} is used it is assumed that A is a square matrix and has an inverse.)
0	a matrix of zeroes.
I	identity matrix.
x	vector.

FACT 1: $(A + B)' = A' + B'$

FACT 2: $(AB)' = B'A'$

FACT 3: $(A')^{-1} = (A^{-1})'$

FACT 4: If $A = (A_1, A_2, \ldots, A_n)$ then

$$A' = \begin{bmatrix} A_1' \\ A_2' \\ \vdots \\ A_n' \end{bmatrix}$$

where A_i is $m \times m$ for each $i = 1,..,n$.

FACT 5: If $Ax = x$ for all vectors x, then $A = I$, the identity matrix.

FACT 6: If A is positive definite symmetric, then there exists a matrix denoted by $A^{\frac{1}{2}}$ which is nonsingular and symmetric such that $A = A^{\frac{1}{2}} A^{\frac{1}{2}}$. The inverse of $A^{\frac{1}{2}}$ is denoted by $A^{-\frac{1}{2}}$ and $A^{-\frac{1}{2}} A^{-\frac{1}{2}} = A^{-1}$.

FACT 7: If $V_1, V_2, \ldots, V_n$ are positive definite symmetric matrices of dimension $p \times p$, then

$$V^{-1} = \begin{bmatrix} V_1^{-1} & 0 & \cdots 0 \\ 0 & V_2^{-1} & \cdots 0 \\ \vdots & \vdots & \\ 0 & 0 & \cdots V_n^{-1} \end{bmatrix}$$

and

$$V^{\frac{1}{2}} = \begin{bmatrix} V_1^{\frac{1}{2}} & 0 & \cdots 0 \\ 0 & V_2^{\frac{1}{2}} & \cdots 0 \\ \vdots & \vdots & \\ 0 & 0 & \cdots V_n^{\frac{1}{2}} \end{bmatrix},$$

where 0 is a p × p matrix of zeroes.

FACT 8: The rank of matrix A is equal to the rank of A'. The rank of A equals the rank of AA' and the rank of A'A.

FACT 9: If A is a n × p matrix with rank p where $p \leq n$, the rank of A'A is p and A'A is positive definite.

FACT 10: The rank of AB is less than or equal to the minimum of the rank of A and the rank of B.

FACT 11: If A is nonsingular, then the rank of AB is the rank of B.

Bibliography

1. Wasan, M. T.: Parameter Estimation, Volume 1 (Lecture Notes). Queen's Papers in Pure and Applied Mathematics No. 9, Queens University, Canada, 1967.

2. Kalman, R. E.: A New Approach to Linear Filtering and Prediction Problems. Trans. ASME, Ser. D, Jour. Basic Engr., Vol. 82, No. 1, Mar. 1960, pp. 34-45.

3. Solloway, C. B.: Dynamic Filtering. Jet Propulsion Laboratory Space Programs Summary No. 37-21, Pasadena, California, March, 1963, pp. 7-10.

4. Graybill, F. A.: An Introduction to Linear Statistical Models, Volume 1. McGraw-Hill Book Company, 1961.

5. Deutsch, R.: Estimation Theory, Prentice Hall, Inc., 1965.

6. Aucoin, C. V., Kenelly, J. W., Lukawecki, S. M., and Prochaska, B. J.: Aspects of Statistical Filtering.