TECHNICAL REPORT

R-135

APPLICATION OF STATISTICAL FILTER THEORY TO THE OPTIMAL ESTIMATION OF POSITION AND VELOCITY ON BOARD A CIRCUMLUNAR VEHICLE

By Gerald L. Smith, Stanley F. Schmidt,
and Leonard A. McGee

Ames Research Center
Moffett Field, Calif.

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
WASHINGTON 1962

NA

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

TECHNICAL REPORT R-135

APPLICATION OF STATISTICAL FILTER THEORY TO THE OPTIMAL ESTIMATION OF POSITION AND VELOCITY ON BOARD A CIRCUMLUNAR VEHICLE

By Gerald L. Smith, Stanley F. Schmidt,
and Leonard A. McGee

ABSTRACT

Statistical filter theory is employed to develop a method for determining the best possible estimate of the position and velocity of a space vehicle in the midcourse phase of flight. Results of a computer simulation are given to illustrate the performance attainable. An on-board system is visualized in which the source of information is an arbitrary sequence of observations of space angles, corrupted by measurement errors. The scheme is in effect a dynamical time-varying filter, implemented by a digital computer, which processes the incoming data to compute an up-to-date optimal estimate of position and velocity.

TABLE OF CONTENTS

SUMMARY 1
INTRODUCTION 2
SYMBOLS 3
  Notation Conventions 5
  Subscripts 5
ANALYSIS 6
  Problem Formulation 6
THEORY OF THE OPTIMAL FILTER 9
  Simplification for a Special Case 12
IMPLEMENTATION OF THE ESTIMATION EQUATIONS 14
RESULTS OF SIMULATION STUDY 16
  Description of the Simulation Program 16
  Performance of System for Different Observation Error Magnitudes . 18
  Effect of Varying the Observation Rate 20
  The Effect of Bias Errors 21
COMPARISON WITH SOME OTHER TRAJECTORY ESTIMATION METHODS 21
  Bayes Estimate 21
  A Minimum Data Method 23
  A Least Squares Method 24
CONCLUDING REMARKS 25
APPENDIX A.- THE EQUATIONS OF MOTION 27
APPENDIX B.- THE LINEAR PERTURBATION EQUATION 29
APPENDIX C.- RELATIONS BETWEEN SPACE ANGLES AND POSITION 32
APPENDIX D.- SIMPLIFICATION OF THE ESTIMATION EQUATIONS FOR
  UNCORRELATED INSTRUMENT ERRORS 35
APPENDIX E.- COMPUTATION OF THE TRANSITION MATRIX 38
APPENDIX F.- RECURSION RELATIONS FOR A BAYES ESTIMATION PROCEDURE . 40
REFERENCES 47
FIGURES 49

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

TECHNICAL REPORT R-135

APPLICATION OF STATISTICAL FILTER THEORY TO THE OPTIMAL ESTIMATION OF POSITION AND VELOCITY ON BOARD A CIRCUMLUNAR VEHICLE

By Gerald L. Smith, Stanley F. Schmidt, and Leonard A. McGee

SUMMARY

Concepts from statistical filter theory are applied to the problem of in-flight determination of the position and velocity of a space vehicle for the purposes of midcourse guidance. The source of information is assumed to be an arbitrary sequence of measurements of any desired set of "observables" (e.g., space angles), the measurements being corrupted by additive errors so that the position and velocity are never known perfectly. A state transition approach is employed which leads naturally to a computational scheme that is readily implemented by a digital computer. The scheme can be regarded as a dynamical time-varying filter which weights the incoming observations in an optimal sense for use in producing an up-to-date optimal estimate of position and velocity.

The advantages of the scheme are (1) it provides the best possible estimate (minimum error) based upon ensemble statistics of injection conditions and measurement errors; (2) it is extremely versatile, not requiring adherence to a predetermined observation schedule or reference trajectory; and (3) the computations are sufficiently simple to be practical in an on-board computer.

A digital computer simulation of the proposed system is employed to demonstrate the feasibility of an all on-board system and to illustrate the performance attainable in a hypothetical physical situation.

INTRODUCTION

One of the problem areas in research relating to space flights is that of midcourse guidance. For the manned circumlunar mission used as an example in this report, the midcourse phase is defined as all of the flight after boost and before re-entry into the earth's atmosphere for landing. Studies of trajectories suitable for such a mission indicate that small errors at injection produce such large errors later along the trajectory (for instance, near the moon) that guidance is generally necessary in the midcourse phase to insure the success of the mission. The problem then is to design a system that will perform this function to some specified degree of accuracy with a minimum expenditure of fuel. The mating of the midcourse guidance system with other aspects of the complete system is also an important part of the design problem, but will not be considered in this report.

The design of the guidance system is a closed-loop control problem, the aspects of which may be described as follows. First, it is necessary to determine by means of data obtained from imperfect sensors (that is, instruments whose measurements are subject to errors) as good an estimate as possible of the position and velocity of the vehicle. This can be called trajectory determination since the position and velocity vectors at any time uniquely determine the trajectory in a free-fall situation. Then, on the basis of the best estimate of the trajectory, end-point conditions must be predicted (e.g., what would the estimated perilune and perigee be if no corrective action were taken). Next, a guidance law must be used which would make possible the calculation of desired corrective action to change the estimated end-point conditions to correspond to those desired. Finally, the indicated control action must be implemented by applying thrust. To close the loop, the applied thrust, acting through the kinematics and geometry, influences the observables which constitute the input to the sensors.

In this paper will be described results of studies of the trajectory estimation portion of the control system problem. The remainder of the problem, with application to a specific (hypothetical) manned circumlunar mission, is treated in another NASA paper (ref. 1).

The description of the problem given above is seen to be rather general. Specifics are required to begin a solution and these may be stated as a set of ground rules, or conditions, as follows:

First, it is assumed that corrective maneuvers in midcourse will be applied not continuously, but intermittently and impulsively. The justification for this assumption is one of practicability; that is, propulsive devices presently at the highest state of development are relatively high-thrust rocket engines. To be employed most efficiently in a situation where relatively small corrective maneuvers are envisioned, such rocket engines must be turned on only briefly and at widely separated time intervals. The result of such a mode of operation is that most of the time the control loop is not closed, and trajectory estimation can be treated separately from the remainder of the guidance problem except during the brief periods of control action.

The second condition is that the trajectory estimation system must constitute a completely on-board operation; that is, observational data will be obtained by on-board sensors (e.g., optical devices for measuring space angles) and all trajectory calculations will be performed by an on-board computer. The justification for this requirement is that in a manned mission an on-board system provides added safety for the crew of the vehicle by eliminating dependence upon the earth-vehicle communication link. This does not mean that the mission will be totally dependent upon the on-board system, of course, and the question of whether or not the on-board system will be the primary system need not be considered at this time. The most significant consequence of this condition is that the on-board system must be in itself accurate enough to satisfy mission objectives and at the same time simple so that it can be reliable and light in weight.

The problem is to find the best estimate of the trajectory from a sequence of imperfect observations of certain arbitrary space angles made repetitiously in any pattern deemed desirable. This is basically a filtering problem and is attacked in the report by means of statistical filter theory. First, the theoretical development of the optimal trajectory estimation system is given. Then it is shown how such a scheme might be implemented in an actual space vehicle. Finally, the results of a simulation study are presented to illustrate the potential usefulness of such a system in an on-board navigation scheme.

SYMBOLS

Lower case English letters are used for vectors (column matrices), except for r, v, p; upper case letters generally denote multiple-column matrices.

D submatrix in M related to n

F perturbation matrix

H	submatrix in M relating m to x
I	unit matrix
K,K*	weighting matrix in optimal filter
m	message, Hx
M	matrix relating y to x*
n	observational error vector
$\hat{p}$	magnitude of the predicted deviation from reference at perilune
$\tilde{p}$	magnitude of the error in prediction of position at perilune
P	covariance matrix of $\tilde{x}$
Q	covariance matrix of observational error, n
r	magnitude of position deviation from reference
$\tilde{r}$	magnitude of error in estimating position
R	covariance matrix of deviation from reference, x
R_o	radius of earth
t	time
u_n	white noise
v	magnitude of velocity deviation from reference
$\tilde{v}$	magnitude of error in estimating velocity
x	deviation from reference
$\hat{x}$	estimate of x
$\tilde{x}$	error in estimating x, $x-\hat{x}$
x*	generalized state variable, including x and n
X,Y,Z	position coordinates in geocentric reference frame
y	observation or measurement of m
α_e	declination angle of earth as seen from vehicle

β_e	right ascension angle of earth as seen from vehicle
γ_e	half the earth-subtended angle as seen from vehicle
Δ	indicates "increments of" as in Δt, $\Delta\Lambda$
σ_n	standard deviation of a single observational error
σ_{x_i}	standard deviation of random variable $x_i(t_o)$
φ_{ij}	element of Φ
Φ, Φ^*	transition matrix

Notation Conventions

$()^{-1}$	inverse of matrix ()
$()^T$	transpose of matrix ()
$()_i$	submatrix of ()
E[]	expected value of [], sometimes used without the brackets

Subscripts

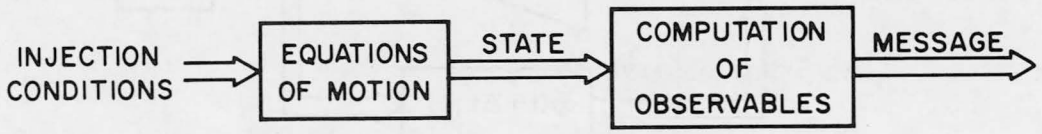
1,2,...,6	numbered state variables
o	at injection
e	earth
k	at kth observation
m	moon
n	related to observation errors, n
s	sun

ANALYSIS

Problem Formulation

The approach employed in solving the problem of trajectory determination is a specialization of some concepts of statistical filter theory proposed by R. E. Kalman (ref. 2). This approach utilizes the state transition formulation which in the present problem may be summarized as follows: Given the equations of motion for the space vehicle, as developed in appendix A, the trajectory of the vehicle can be specified uniquely from a knowledge at any time of the three components of the vehicle's position vector and three components of its velocity vector in an orthogonal reference frame. These six variables are defined as the state vector which is a continuous time function generated by integration of the equations of motion with appropriate initial conditions. Because the initial conditions are not known precisely, the present state is also not known, and it is the function of the trajectory determination system to estimate the state on the basis of observations made by on-board instruments. The system is then regarded as a multidimensional filter, its input being a time sequence of observations of variables related to the state, corrupted by additive errors. Its output is the estimate of the state at present time, and the filter is to be designed to make this an optimal estimate in the sense of minimizing some function of the estimation error.

In the filter design, it is convenient to think of the input to the trajectory determination system as composed of a "message" plus "noise." The message in this case is a set of observables (e.g., space angles) which are a consequence of the physical situation as defined by the state. The message-generating process can then be represented by an integration of the equations of motion to obtain the state, followed by computation of the observables, as illustrated in sketch (a).



Sketch (a).- Message-generating process.

This message process is, of course, nonlinear. To employ techniques from linear theory, suppose that the nonlinear equations of motion are linearized. The procedure used is described in appendix B and is, in effect, a Taylor series expansion about a reference trajectory, retaining only the first-order terms in the expansion. The state can then be

expressed in the form of a deviation state vector, $x(t)$, which is the solution of a set of linear time-varying differential equations. In standard matrix form these equations may be written as

$$\dot{x}(t) = F(t)x(t) \tag{1}$$

which is called the perturbation equation. Alternatively, equation (1) can be written in the form

$$x(t + \Delta t) = \Phi(t + \Delta t;\ t)x(t) \tag{2}$$

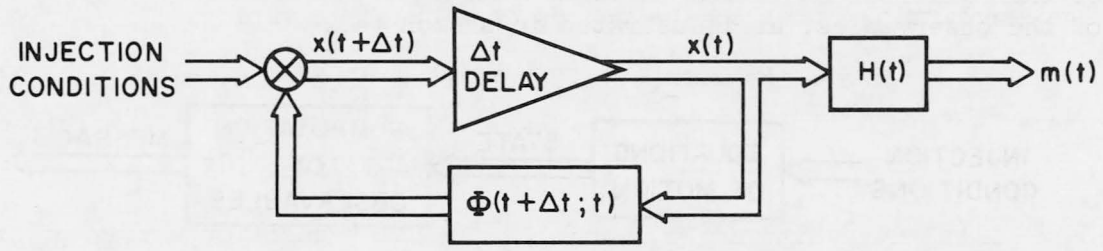
where state transition concepts are emphasized. The matrix $\Phi(t + \Delta t;\ t)$ is the transition matrix associated with the linearized equation of motion and describes how the state changes from time t to time $t + \Delta t$. It has the properties that $\Phi(t;\ t)$ is the unit matrix, I, and $\Phi(t_2;\ t_1) = \Phi^{-1}(t_1;\ t_2)$.

Note that no forcing function appears in equations (1) and (2). This is because a free-fall condition, with negligible disturbing forces, is postulated for the vehicle trajectory.

In the linearization process, the equations that relate the observables to the state variables are also linearized as shown in appendix C. The message is then expressed in terms of deviations from a reference, and is linearly related to the deviation state:

$$m(t) = H(t)x(t) \tag{3}$$

The message process can now be represented by the linear system illustrated in sketch (b).



Sketch (b).- Linearized message process.

The actual injection conditions, of course, are not known, but it is assumed that they can be described probabilistically at least up to second-order statistics. Thus, the injection conditions are regarded as a vector-valued random variable. When expressed in terms of deviation from expected (or ideal) injection conditions, this random variable has

zero mean. If the linear model of the message process is assumed to be valid, the state and the message are then also random variables with zero mean. The second-order statistics assumed are the covariance matrix of injection errors defined as follows:

$$\mathrm{cov}[x_o, x_o] = E[x_o x_o^T] = P_o$$

The covariance matrix of trajectory deviations is then given, by means of equation (2), as

$$\begin{aligned}\mathrm{cov}[x(t), x(t)] &= E[x(t)x^T(t)] \\ &= \Phi(t; t_o)P_o\Phi^T(t; t_o)\end{aligned}$$

Furthermore, by use of equation (3),

$$\mathrm{cov}[m(t), m(t)] = H(t)\Phi(t; t_o)P_o\Phi^T(t; t_o)H^T(t)$$

Thus, the statistics of the message are expressed in terms of the statistics of injection conditions and the linear model of the message-generating process.

A treatment similar to that accorded the message statistics is employed for the instrument errors. However, without assuming a particular instrumentation system, it is not possible to be as specific as in the case of the message process. Nevertheless, the assumption, which is standard in engineering applications of statistical filter theory, is that these errors can be represented as the output, $n(t)$, of a dynamic system excited by an independent (vector-valued) Gaussian random process, $u_n(t)$. The instrument errors are regarded collectively as a vector, $n(t)$, having as many components as the individual sources of error considered. The error, $n(t)$, then can be represented by the equation

$$n(t + \Delta t) = \Phi_n(t + \Delta t; t)n(t) + u'(t + \Delta t, t) \qquad (4)$$

where

$$u'(t + \Delta t, t) = \int_t^{t+\Delta t} \Phi_n(t + \Delta t; \tau)u_n(\tau)\,d\tau$$

Here $u_n(t)$ is "white noise," and Φ_n is the transition matrix of the error process. The statistical properties of $n(t)$ are expressed in terms of Φ_n and the covariance matrices,

$$N_o = \mathrm{cov}[n(t_o), n(t_o)] = E[n(t_o)n^T(t_o)]$$

$$Q_n(t) = \mathrm{cov}[u_n(t), u_n(t)] = E[u_n(t)u_n^T(t)]$$

The mean values of $n(t_o)$ and $u_n(t)$, hence also of $n(t)$, are assumed zero. (This represents no loss in generality since nonzero means would normally be calibrated out of the on-board instruments.)

The linear equations (2) and (4) may now be combined in the form

$$x^*(t + \Delta t) = \Phi^*(t + \Delta t;\ t)x^*(t) + u(t + \Delta t,\ t) \tag{5}$$

where x^* is a generalized state vector including both the vehicle state and instrument error components, and Φ^* includes both Φ and Φ_n. The statistical properties of x^* are thus expressed in terms of Φ^* and the covariance matrices P_o, N_o, and $Q^*(t + \Delta t;\ t)$, where Q^* is defined as

$$\left.\begin{aligned} Q^* &= \begin{bmatrix} 0 & 0 \\ 0 & Q' \end{bmatrix} \\ Q'(t + \Delta t,\ t) &= \int_t^{t+\Delta t} \Phi_n(t + \Delta t;\ \tau)Q_n(\tau)\Phi_n^T(t + \Delta t;\ \tau)\,d\tau \end{aligned}\right\} \tag{6}$$

To complete the formulation of the problem, it is necessary to obtain an expression for the observations by appropriately combining equations (3) and (4). If additive instrument errors are assumed, the observations will be the sum of two random processes, which constitute another random process having values only at times t_k when observations are made. Thus, the observation, termed $y(t_k)$, is a linear combination of the generalized state variables, as follows:

$$\begin{aligned} y(t_k) &= m(t_k) + D(t_k)n(t_k) \\ &= H(t_k)x(t_k) + D(t_k)n(t_k) \\ &= M(t_k)x^*(t_k) \end{aligned} \tag{7}$$

It is noted that since x^* has zero mean, so does y. Since each observation may be perturbed by several different sources of noise (such as calibration errors, readout errors, tracking errors, etc.) the matrix D is provided to combine the various noise sources appropriately into a single observation error. Thus n may be of larger (or smaller) dimension than m.

THEORY OF THE OPTIMAL FILTER

We now come to the principal problem, which may be stated as follows: Given a set of observed values of the message, $y(t_o)$, $y(t_1)$, . . . $y(t_k)$, find a "best" estimate $\hat{x}^*(t)$ of $x^*(t)$, where the judgment of what is "best" (i.e., the optimality criterion) is yet to be specified.

To arrive at a reasonable optimality criterion, it is natural to assign some penalty (i.e., loss) for incorrect estimates. If the error in estimate is defined as $\tilde{x}^*(t) = x^*(t) - \hat{x}^*(t)$, it is clear that this

loss must be positive for every nonzero value of $\tilde{x}^*(t)$. Among the more obvious of such loss functions is the scalar product $\tilde{x}^{*T}\tilde{x}^*$. This is the vector equivalent of the familiar squared-error criterion of one-dimensional filter problems. Although it is possible to raise some arguments about the appropriateness of this criterion, the general form seems correct because it implies minimization of fuel requirements for midcourse corrections. The implication of fuel minimization is explained by noting that the error in determination of the proper velocity correction at any time t is for all practical purposes a linear function of $\tilde{x}^*(t)$. Thus, minimizing $\tilde{x}^{*T}(t)\tilde{x}^*(t)$ also minimizes fuel waste for a correction made at time t. The mathematical convenience of this criterion is persuasive, and furthermore it gives an estimation formula that is correct for a somewhat more general criterion (refs. 2 and 3).

It is next assumed that the optimal estimate will be restricted to be a linear function of the observations; that is, the best estimate of the form

$$\hat{x}^* = \sum_{i=1}^{k} A_i y_i \tag{8}$$

is to be obtained.[1] Although this assumption is in keeping with the linear philosophy employed throughout this development, it may be objected that this is too restrictive. However, it should be realized that for an on-board system the utmost simplicity is important, and a linear estimating procedure probably is the simplest. Another way of stating this argument is to say that there is some loss or penalty associated with computer complexity, and if this were incorporated in the optimality criterion the expected result should be to favor a linear system.

Under the above assumptions it is clear the optimal estimator may be regarded as a linear filter whose input is the actually occurring sequence of observations. The next step in the development of the theory is to view this filter from the state transition point of view; that is, the estimation computations are to be considered as proceeding in real time, utilizing only the previous estimate and the latest observations at any one time. With the assumption that at the time of the kth observation the estimate based on the $k-1$ previous observations has been computed, it is readily deduced that the new estimate based on k observations must be of the linear form

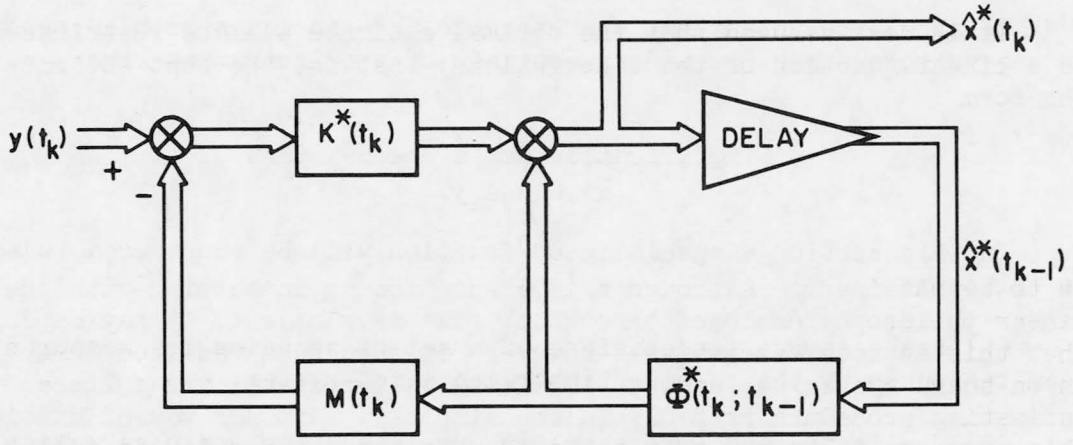
$$\hat{x}^*(t_k) = \Phi^*(t_k; t_{k-1})\hat{x}^*(t_{k-1}) + K^*(t_k)[y(t_k) - M(t_k)\Phi^*(t_k; t_{k-1})\hat{x}^*(t_{k-1})] \tag{9}$$

---

[1] If $x(t)$, $n(t)$ are Gaussian, the unrestricted optimal estimate is of the linear form (8). In other words, only when the system inputs are not Gaussian can the estimate be improved by a nonlinear estimator. Thus, the restriction to linear estimation could be replaced by the assumption of Gaussian inputs if so desired. (See refs. 2 and 3.)

It is noted here that the quantity $\Phi^*(t_k; t_{k-1})\hat{x}^*(t_{k-1})$ is the estimate of $\hat{x}^*(t_k)$ based on the first k-1 observations. The quantity in brackets is then the difference between the kth observation, $y(t_k)$, and the estimated value of the vector observable at time t_k. The matrix, $K^*(t_k)$ weights the residual (quantity in brackets) to produce an increment to be added to the estimate. Thus, the form of the estimation equation is perfectly natural, for if the current observation should happen to agree perfectly with the estimated observable, the fact that an observation took place should have no effect on the estimate.

Equation (9) can be represented in block diagram form as shown in sketch (c), where it is seen that realization of the optimal filter



Sketch (c)

requires only a model of the $x^*(t)$ generating system and the weighting matrix $K^*(t_k)$. Thus, the optimal properties of the filter depend upon the proper selection of $K^*(t_k)$.

To obtain an equation whereby $K^*(t_k)$ may be computed, the principle of orthogonal projection in a multidimensional space is employed. Kalman shows (ref. 2) that for the norm-squared error criterion and linear filter restriction, the optimal estimate $\hat{x}^*$ is the orthogonal projection of x^* upon a linear manifold, or vector space, formed by the set of all linear combinations of the random variables in the set of observations, $y(t_1), \ldots, y(t_k)$. The result is that the error in estimate, $\tilde{x}^*(t_k)$ is orthogonal to the estimate $\hat{x}^*(t_k)$. This principle is used in reference 2 to show that the weighting matrix is given by the expression

$$K^*(t_k) = P^*(t_k)M^T(t_k)[M(t_k)P^*(t_k)M^T(t_k)]^{-1} \qquad (10)$$

where $P^*(t_k)$ is the covariance matrix of the estimation error at time t_k based on the previous k-1 observations. The matrix P^* is in turn given by the recursion relation

$$P^*(t_{k+1}) = \Phi(t_{k+1};\, t_k)[P^*(t_k) - K^*(t_k)M(t_k)P^*(t_k)]\Phi^T(t_{k+1};\, t_k) + Q^*(t_{k+1},\, t_k) \qquad (11)$$

Thus, the computation of the optimal estimate is seen to be a straightforward step-by-step procedure. To begin the computations some starting values of $\hat{x}^*$ and P^* are required. These might be, for instance, an initial estimate of injection conditions and noise, and the covariance matrices N_o and P_o, as obtained from some source such as the boost guidance system.

It should be noted that the estimate $\hat{x}^*$ includes both an estimate of the vehicle state, x, and an estimate of the instrument error, n.

Simplification for a Special Case

In this section a specialized situation will be considered, described as follows:

(1) Each observation consists of a set of space angles measured simultaneously by the on-board instrumentation system.

(2) The instrument errors are uncorrelated from one observation time to the next (although there may be correlation between the errors in the separate components of an individual observation).

The first assumption above is entirely arbitrary, introduced so that the discussion of system performance can be more specific. The second assumption is not unrealistic in that observations on board the vehicle are likely to be well separated in time, particularly if the number of observations required for navigation is not large.

It is readily shown with the second assumption that the implementation of the estimate of instrument errors can be omitted from the optimal filter and the computations thereby markedly simplified. The development of this simplification is given in appendix D. The major effect is to reduce the order of the matrices involved in the calculations, since only $\hat{x}$, the covariance matrix of $\tilde{x}$ (designated P), and the weighting matrix associated with $\hat{x}$ (designated K) need be computed. The equations to be solved are:

$$\hat{x}(t_k) = \Phi(t_k;\ t_{k-1})\hat{x}(t_{k-1}) + K(t_k)\left[y(t_k) - H(t_k)\Phi(t_k;\ t_{k-1})\hat{x}(t_{k-1})\right] \tag{12}$$

$$K(t_k) = P(t_k)H^T(t_k)\left[H(t_k)P(t_k)H^T(t_k) + Q(t_k)\right]^{-1} \tag{13}$$

$$P(t_{k+1}) = \Phi(t_{k+1};\ t_k)\left[P(t_k) - K(t_k)H(t_k)P(t_k)\right]\Phi^T(t_{k+1};\ t_k) \tag{14}$$

The $Q(t_k)$ is presumed known a priori as described in appendix D. It is noted that only the model of the message process and not of the instrument error process is required in the optimal filter.

For purposes of computation which will be explained later, it is convenient to represent equation (14) in the form of two operations:

$$P(t_{k+1}) = \Phi(t_{k+1};\ t_k)P'(t_k)\Phi^T(t_{k+1};\ t_k) \tag{15}$$

$$P'(t_k) = P(t_k) - K(t_k)H(t_k)P(t_k) \tag{16}$$

It is noted that equation (16) represents the manner in which P changes at time t_k as a result of the information contained in the observation, and equation (15) represents the way in which it changes as a result of transition along the trajectory.

These equations will now be specialized to the case of observations consisting of three angles. These angles may be, for instance, the subtended angle of the earth and the right ascension and declination of the earth center as viewed from the vehicle, as assumed in the simulation studies presented later. Since the angles depend only on vehicle position and not on velocity, the H matrix may be in this case partitioned in the form

$$H = \begin{bmatrix} H_1 & 0 \end{bmatrix}$$

where H_1 is a 3X3 matrix of partial derivatives of the three angles with respect to the X, Y, Z coordinates of vehicle position, and 0 is a 3X3 null matrix. If P is likewise partitioned in the form

$$P = \begin{bmatrix} P_1 & P_2^T \\ P_2 & P_3 \end{bmatrix}$$

where the submatrices are all 3X3's, equation (13) can be written in the form

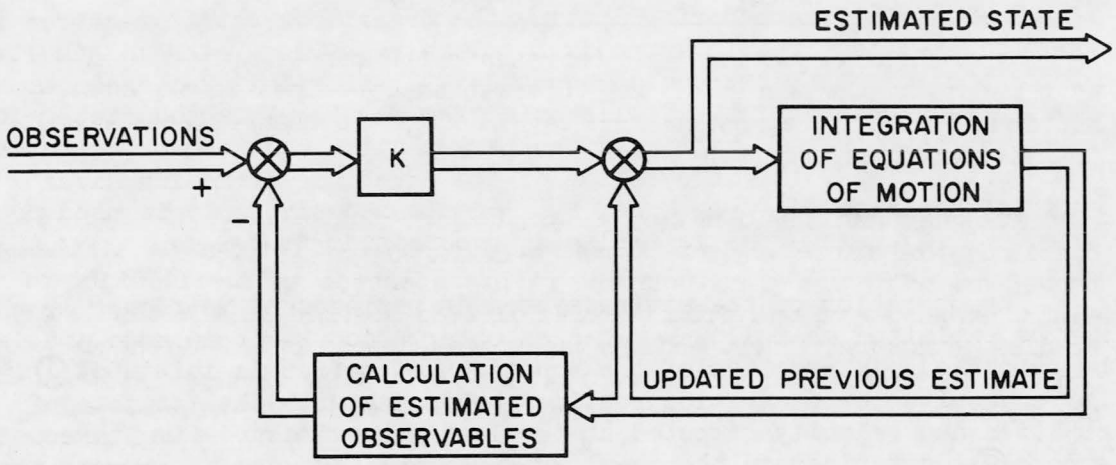
$$K = \begin{bmatrix} P_1 \\ P_2 \end{bmatrix} H_1^T\left[H_1 P_1 H_1^T + Q\right]^{-1} \tag{17}$$

The computation of K is thus seen to be relatively simple, involving only the inversion and multiplication of 3X3 matrices.

IMPLEMENTATION OF THE ESTIMATION EQUATIONS

Some comments regarding the implementation of the optimal filter are desirable to clarify the manner in which the calculations might be performed in an on-board digital computer. The simplified system described in the last section will be considered.

First of all, it is clear from equation (12) that the model of the m(t) generating system (the message process) employed in the optimal filter need not be the linearized version but could be a more accurate (generally nonlinear) representation - in other words, the equations of motion themselves. The computations would then not be in terms of deviation quantities but the original variables. A block diagram representation of the computation would then appear as shown in sketch (d). The input is the actual observation from which the latest estimate of the



Sketch (d)

vector observable is subtracted. The difference is then multiplied by the matrix K, just as before, to increment the estimate. The new estimate then is a new set of starting conditions used for integration of the model equations until the time of the next observation.

Computation of P and hence K still requires the linearization approach because of the manner in which the Φ, H, and Q matrices appear in equations (13) and (14). Since Q represents the statistics of the errors in the measuring instruments, which are presumably known a priori, it seems reasonable that this would be a stored matrix. Likewise, Φ and H could theoretically be stored since they represent the equations of motion linearized about the reference trajectory which is known before

launch. However, in practice it might be awkward to store these in an easily utilized form because of the arbitrariness in the times at which observations are made. Thus, it appears simpler to arrange to compute these matrices in the vehicle. It is seen that these computations need be performed only at times observations are made, the values at other times having no meaning for the problem.

Another reason for computing H and Φ in the vehicle is that this makes it possible to linearize around the estimated rather than the reference trajectory. This is clearly the correct procedure since P has to do with the difference between the estimate and the true state, and the estimate is on the average closer to the true state than is the reference. Errors arising from the linearization assumptions are thereby minimized. It is noted that if this procedure is used, very large deviations from the reference (such as would occur in an abort situation for instance) would not jeopardize the accuracy of the trajectory determination scheme as long as the estimate itself was always reasonably good.

One possible method of computing the transition matrix required in equation (14) is given in appendix E. This requires a sixfold simultaneous integration of the perturbation equation (1), starting over again at the time of each observation. To linearize around the estimated state, it is necessary that the F matrix of the perturbation equation be computed from $\hat{x}(t)$. Thus, the integration of the equations of motion gives $\hat{x}(t)$ in the interval $t_{k-1} < t < t_k$. Simultaneously, $\hat{x}(t)$ is used to compute $F(t)$, and $F(t)$ is employed to obtain $\Phi(t; t_{k-1})$.

The operation of the entire system is represented in figure 1, which may be explained as follows: When an observation has been made and is to be processed, integration of the equations of motion is initiated (1),[2] (by any suitable integration routine) beginning with the estimate of position and velocity computed at the last observation. Simultaneously, scheme (2), described in the previous paragraph, is used to compute the transition matrix from the last observation (3). This process continues until computer time equals the time of the observation. The integration is then stopped, $P(t_k)$ is computed (4) from equation (15), and $H(t_k)$ is computed (5) from the last estimate of the state. The matrix $K(t_k)$ is then computed (6). The estimated angles are also computed (7) from the last estimate of the state just prior to the observation. These are subtracted (8) from the actually observed angles, the difference is multiplied by $K(t_k)$ (9), and the estimate is incremented (10) by this amount to produce the new estimate of the state. The P' matrix is then computed (11) from equation (16), reflecting the change in P due to the observation, and P' is stored. The delay unit (12) represents the storage of $P'(t_k)$ until the time of the next observation when it is needed to compute $P(t_{k+1})$.

---

[2]Circled numbers identify computer operations illustrated in figure 1.

The computation cycle having thus been completed, the computer simply waits, either in standby or off condition, until the next observation is to be processed.

In regard to computer speed requirements for the foregoing computations, it is, of course, necessary that the time required to complete the computation cycle be less, on the average, than the time interval between successive observations. If relatively few observations are necessary for satisfactory navigation, this consideration is of little consequence since times on the order of minutes or even hours would be available for each cycle. What is more critical is the fact that a long time delay between making an observation and obtaining the improved estimate would likely be undesirable for a number of reasons. Even so, a relatively modest computer speed would appear adequate. It may be noted that with widely spaced observations the computer would actually be off most of the time, resulting in a considerable saving in power consumption as compared to running continuously.

RESULTS OF SIMULATION STUDY

Description of the Simulation Program

In this section the results of a digital computer study are presented to illustrate the performance of a trajectory determination scheme of the type which has been described. A lunar circumnavigation mission is assumed, with a nominal trajectory such that the vehicle achieves a perilune altitude of 4766 km in 3.28 days of flight and returns to a vacuum perigee altitude of 72 km. This trajectory is shown in figure 2. A diagonal covariance matrix of injection errors is assumed:

$$P(t_o) = \begin{bmatrix} \sigma_{x_1}^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_{x_2}^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_{x_3}^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_{x_4}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_{x_5}^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_{x_6}^2 \end{bmatrix}$$

The observation schedule assumed is described as follows: Observations of three angles - earth-subtended angle and two angles which describe the direction of the vehicle-earth line - are made with optical instruments on board the vehicle. The physical situation is described in appendix C. The first measurement is made 1/2 hour after injection, and subsequent observations are spaced at 6-minute intervals until a total of 20 observations has been completed.

It is assumed that additive noise having zero mean and a diagonal covariance matrix,

$$Q(t) = \begin{bmatrix} \sigma_n^2 & 0 & 0 \\ 0 & \sigma_n^2 & 0 \\ 0 & 0 & \sigma_n^2 \end{bmatrix}$$

contaminates these measurements. The noise in each angular measurement has the same Gaussian distribution which does not vary with time. Also, there is no correlation between the noise samples at different observation times.

The average, or ensemble, performance of the system can be seen to be given by the variance equations, (15) and (16). The solution of these equations obviously depends upon the initial condition $P(t_o)$, the matrix parameters Q, H, and Φ, and the spacing of observations. The choice of nominal trajectory, starting time, and observation schedule, as described above, essentially removes H and Φ as parameters. Thus, we are concerned with determining the effects of varying $P(t_o)$ and Q. From equation (13) it can be seen that the time constant or rate of decay of $P(t)$ depends upon the relative magnitudes of Q and $P(t_o)$; that is, if Q and $P(t_o)$ were both increased by the same scale factor the shape of the $P(t)$ curve would be unchanged. Its magnitude, however, would be increased in proportion to the increase in $P(t_o)$. Thus, a change in $P(t_o)$ is equivalent to a change in Q except for scale factor. Therefore, to determine the nature of the variation due to varying $P(t_o)$ and Q we need vary only one of these. Here we choose to vary the instrumentation noise, Q.

The problem as stated was programmed for a digital computer in the form shown in figure 1, with the addition of a computation of the covariance matrix of deviations from the nominal trajectory,

$$R(t) = \Phi(t;\, t_o)P(t_o)\Phi^T(t;\, t_o) \qquad (18)$$

and a computation of the estimate and error in estimate of position at nominal perilune using linear prediction. Together these computations make possible an assessment of the average, or statistical, performance of the system. At the same time, with specific randomly selected injection errors and noise as inputs, each computer run gives a specific member of the ensemble of actual trajectories and estimates thereof.

For convenience in presentation, the data are given only in terms of the magnitudes of position and velocity deviations. Thus, the actual trajectory resulting from the specific injection conditions employed is represented by the quantities

$$\left.\begin{aligned} r &= \sqrt{x_1^2 + x_2^2 + x_3^2} \\ v &= \sqrt{x_4^2 + x_5^2 + x_6^2} \end{aligned}\right\} \qquad (19)$$

which are the magnitudes of the deviations in position and velocity from the nominal trajectory. Similarly, the error in estimate $\tilde{x}$ for a specific run is given in terms of its position and velocity components:

$$\left.\begin{aligned} \tilde{r} &= \sqrt{\tilde{x}_1^2 + \tilde{x}_2^2 + \tilde{x}_3^2} \\ \tilde{v} &= \sqrt{\tilde{x}_4^2 + \tilde{x}_5^2 + \tilde{x}_6^2} \end{aligned}\right\} \qquad (20)$$

where $\tilde{x}_i = x_i - \hat{x}_i$. Likewise, in presenting ensemble results, we plot rms position and velocity deviations. From the six terms in the principal diagonal of the R matrix we obtain

$$\left.\begin{aligned} r_{rms} &= \sqrt{Ex_1^2 + Ex_2^2 + Ex_3^2} \\ v_{rms} &= \sqrt{Ex_4^2 + Ex_5^2 + Ex_6^2} \end{aligned}\right\} \qquad (21)$$

and similarly, from the P matrix we obtain the rms position and velocity estimation errors:

$$\left.\begin{aligned} \tilde{r}_{rms} &= \sqrt{E\tilde{x}_1^2 + E\tilde{x}_2^2 + E\tilde{x}_3^2} \\ \tilde{v}_{rms} &= \sqrt{E\tilde{x}_4^2 + E\tilde{x}_5^2 + E\tilde{x}_6^2} \end{aligned}\right\} \qquad (22)$$

A Fortran program designed for use on the IBM 704 digital computer was written to perform the computations described above. The storage space used by the program (including provisions for a number of other computational features not used in the present study) is about 13,000 words. Computation time on the IBM 704 is roughly 15 minutes for the 2-1/2-hour flights simulated in the study.

Performance of System for Different Observation Error Magnitudes

To illustrate the effect of varying the magnitude of the observation errors as defined by Q, four computer runs were obtained simulating an observation routine consisting of a total of 20 observations spaced at 6-minute intervals beginning 1/2 hour after injection.

The statistical descriptions of the inputs were specified as follows: (1) For Q, σ_n was taken to be 5, 20, 50, and 200 sec arc, respectively, in each of the four runs. (2) For $P(t_o)$, σ_{x_1}, σ_{x_2}, and σ_{x_3} were taken to be 1 km, and σ_{x_4}, σ_{x_5}, and σ_{x_6} were taken to be 1 m/sec, the same for all runs.

The specific injection conditions and observation errors were chosen randomly in accordance with the assumed statistics as follows: (1) Observation errors were generated by a Gaussian random number computation and scaled for use in each run according to the σ_n employed. The time histories of the sample employed are shown in figure 3, and were the same for all runs except for scale factor. (2) The injection errors were selected at random from a table of Gaussian random numbers. The same values were used for each run:

$$x_1(t_o) = 0.495 \text{ km}$$

$$x_2(t_o) = -0.886 \text{ km}$$

$$x_3(t_o) = -1.001 \text{ km}$$

$$x_4(t_o) = 0.281 \text{ m/sec}$$

$$x_5(t_o) = 1.999 \text{ m/sec}$$

$$x_6(t_o) = 0.194 \text{ m/sec}$$

Figure 4 shows, for the four different observation error levels, the errors in estimating position and velocity as a function of time. The irregularities in the estimate as a function of time are due to the noise and are of the same character for each of the runs because the same noise sample (except for scale) was used in each case. The actual deviations from the nominal trajectory are also shown in figure 4 for comparison with the estimation errors. It should be noted that if no observations were made, or equivalently if the system were designed for infinite measurement errors, the "best estimate" would be the nominal trajectory itself, and the error in estimate would be the actual deviations from nominal, given by the quantities r and v. The differences between r and $\tilde{r}$, v and $\tilde{v}$, thus represent the improvement in knowledge of the trajectory due to the observations.

Figure 5 shows several time histories of the system performance with the same injection errors in each case, but with different noise time histories, σ_n = 20 sec arc. A run made with no noise is also shown for comparison, to give an indication of how much the estimate is perturbed by the noise.

The first portions of the $\tilde{r}$ and $\tilde{v}$ curves represent the errors in estimate prior to the first observation. As previously explained, these are simply the deviations from the nominal trajectory due to injection errors, which are the same for all runs. It is seen that the estimation error is generally larger when noise is present than when it is not. This is a result of presenting the estimation errors in terms of the magnitude quantities $\tilde{r}$ and $\tilde{v}$. If the error components were shown separately (e.g., $\tilde{x}_1$, $\tilde{x}_2$, etc.), the no-noise results would tend to represent the average of the several noise time histories.

Figure 6 shows the ensemble average estimation errors, obtained from the P matrix. The average deviations from the nominal trajectory, r_{rms} and v_{rms}, obtained from the R matrix, are also plotted for comparison. The average improvement in knowledge of the trajectory is represented by the differences between r_{rms} and $\tilde{r}_{rms}$, v_{rms} and $\tilde{v}_{rms}$, which are seen to be always positive and greatest when the noise is the least. Comparison of figures 4, 5, and 6 indicates that the individual performances shown in figures 4 and 5 are reasonable members of the ensembles shown in figure 6.

The performance of the system in predicting the error at perilune is shown in figure 7 for the four different noise levels. The actual estimate of the deviation from reference, designated $\hat{p}$, and the mean-square error in the estimate, $\tilde{p}_{rms}$, are plotted as a function of time. The irregular character of the $\hat{p}$ curves is again due to the noise. It is seen that $\hat{p}$ tends to build up in a roughly exponential manner toward the actual miss (4528 km), the time constant being greater for the larger noise magnitude, reflecting the poorer confidence in the measurements which exists when the noise is large.

In figure 8 a different method of presenting the estimation error data is employed to illustrate the fact that the estimate is better in some directions than in others. This fact is obscured in the plots of figure 6 but is significant in determining the character of the information that the assumed set of observations contributes to the estimate. It should be noted that the elements of the P matrix describe the shape, size, and orientation of a time-varying estimation error ellipsoid. This ellipsoid may be considered as centered on the current estimate of the trajectory. The intersection of this ellipsoid with the equatorial plane is shown in figure 8 for the case of σ_n = 20 sec arc. The equivalent ellipse obtained from the R matrix, which illustrates the statistics of deviations from the reference trajectory, is also shown for comparison. One interesting point to be noted is that the major axis of the error ellipse tends to be oriented along the vehicle-earth line. The implication is that the specified observations give less information regarding the position along this line than in any other direction. Since the distance from the earth is obtained principally from the measurement of subtended earth angle, this indicates that the information available from this angle is relatively poor.

Effect of Varying the Observation Rate

If the number of observations made during a given period of time is increased, an improvement in the knowledge of the trajectory should be expected. To illustrate this effect, a run was made using the same initial conditions and noise as in a previous run with σ_n = 20 sec arc, with observations starting at the same time but spaced 3 minutes

apart. Thus, during the same 2-hour period 39 observations were completed. Figure 9 compares the performance for this case with that for 20 observations at 6-minute intervals. It is seen that the increased observation rate produces a reduction of roughly 20 percent in $\tilde{r}_{rms}$ and about 17 percent in $\tilde{v}_{rms}$.

The Effect of Bias Errors

The assumption that the instrumentation noise samples at different observation times are statistically independent is reasonable for certain sources of noise. However, it is expected that generally there will exist in any instrumentation scheme additional errors which are systematic in nature (e.g., telescope boresight errors), and thus are definitely not independent from one observation to another. Such errors might also be termed bias errors since they tend to remain the same over a long period of time. To give some insight into the manner in which these errors affect the performance of the system, which is not optimized for such errors, three additional computer runs were made, each with a constant bias error of +5 sec arc added to one of the three angles being measured in addition to the random noise (σ_n = 20 sec arc) previously assumed. The results of these runs are shown in figure 10, together with the corresponding no-bias run repeated from figure 3 for comparison. It is seen that bias error on α_e and β_e (the declination and right ascension of the earth, respectively) has virtually no effect on $\tilde{r}$ and $\tilde{v}$ but that the same bias on γ_e (the earth-subtended angle) produces a substantial effect, increasing $\tilde{r}$ and $\tilde{v}$ by roughly a third. The conclusion to be drawn is that the system is quite a bit more sensitive to bias errors on γ_e than on α_e and β_e, although this cannot be stated as a general conclusion since there is an obvious dependence upon the geometry of the particular situation simulated. Nevertheless, it can be stated that bias errors of certain types should be investigated carefully in the design of a guidance system. If necessary, such bias errors can be treated as additional state variables and estimated along with the others. The theory presented is sufficiently general to allow this, provided suitable statistical descriptions of the errors can be supplied.

COMPARISON WITH SOME OTHER TRAJECTORY ESTIMATION METHODS

Bayes Estimate

The most widely recognized trajectory estimation technique, and one which is in actual use by Jet Propulsion Laboratory (ref. 4) and others,

is based on Bayes estimation.[3] This is an optimal method which, since it utilizes the same optimality criterion as in the filter theory approach, one should suspect should give the same results. Such is not immediately apparent in viewing the two formulations.

The Bayes estimation approach is developed from decision theory concepts, the estimate being based upon maximizing a multivariate a posteriori probability density function. The filter theory approach uses the idea of orthogonal projection in a multidimensional space (ref. 2). A careful study of these ideas shows that they are basically the same. The two approaches differ then only (1) in the manner of introducing such matters as linearizing a basically nonlinear process[4] and assuming Gaussian distributions, and (2) in the form of the estimation equations. The Bayes estimation equations are generally expressed in a form such that the estimate $\hat{x}(t_k)$ is obtained by operating on the entire set of k observations at once, whereas in the filter formulation the estimation procedure is a sequential operation on the observations taken one at a time in the order of their occurrence. One way of demonstrating the equivalence of the two approaches then is to develop recursion relations for the Bayes estimation equations so that the mode of operation is the same as that employed in the filter approach. This has been done in appendix F for the restricted case of uncorrelated observations, where it is shown that the equations are the same and hence the methods equivalent. A rigorous general treatment is not attempted, the purpose being primarily to verify what common sense already indicates, namely, that there cannot be two different optimal methods having the same optimality criterion and basic assumptions.

The two methods then should not be called by different names - both are Bayes estimation. The equations whereby the estimate is obtained by operating upon the entire set of observations comprise the closed-form solution of the Bayes estimation problem. In contrast, the equations developed from filter theory are the difference equations corresponding to the closed-form solution; the estimate in this case is obtained by a process analogous to solution of differential equations by numerical integration.

The question of which of these methods of solution is superior is not easily answered since it depends on the application intended. A point in favor of using the closed solution is that if the solution at

---

[3]This method has been called maximum likelihood estimation by J. P. L., although technically speaking it should be termed "Bayes estimation." The distinction in the present context is that the Bayes estimate utilizes the a priori statistics of injection errors and the maximum likelihood estimate does not. See references 5, 6, and 7 for more detailed definitions.

[4]The difference in the manner in which linearization is employed in the two methods should result in a slight numerical difference in the answers obtained. However, this is a practical rather than a theoretical consideration, the two methods still being fundamentally identical.

only one point is desired, only a single calculation is required, although it must be noted that this calculation may be far more complex than each of the many computations involved in using the difference equation. This is particularly true if the observations are correlated, for then the closed-form solution involves the inversion of the very large Q matrix, which is quite apt to be ill-conditioned. Of course, it is assumed that Q is known a priori and the difficult inversion process need not be performed in the on-board computer; that is, the Q^{-1} matrix could be precomputed and stored. However, if the solution is to be obtained at a number of points, for instance, at each observation time as in the present problem, it is seen that Q^{-1} is different for each calculation and a very large storage would be necessary. In this case, the difference equation form of solution appears to be the natural approach since the number of calculations required is the same for both methods (i.e., one calculation for each observation), but the calculations are much less complex. This conclusion applies particularly to the case where correlated observation errors are considered. However, even with uncorrelated observations the difference equation solution is apt to be simpler because the closed-form solution always involves the inversion of a 6×6 matrix (eq. (F15)), whereas the difference equations involve inversion of a matrix which is of order equal only to the number of measurement components comprising an observation (see eq. (13)). It should be further noted that if the measurements comprising an observation are themselves uncorrelated, each measurement could be treated as an independent observation (even though they occur at the same time); the matrix to be inverted is in this case a 1×1, and the inversion is trivial so that the ultimate in calculation simplicity is realized.

A Minimum Data Method

Although the trajectory determination technique presented is optimal, the question naturally arises as to how much better it is than other nonoptimal schemes which may be simpler and therefore have an advantage from an implementation point of view. No definitive answer to the question will be attempted here because the number of possible nonoptimal systems is myriad. However, some idea of the trade-off considerations involved can be obtained by a cursory examination of two alternative schemes. The first of these is a minimum data method wherein determination of the trajectory is based on two observations, each consisting of three angles. The second (given in the next section) is a least squares data smoothing method described in reference 8.

The two-observation method can be described as follows. A measurement is made of the earth-subtended angle and direction of the vehicle-earth line at time t_1, which is taken to be 1/2 hour after injection to coincide with the time observations begin in the optimal system. Then during the subsequent 2-hour period another similar observation is made.

On the basis of the two observations, an estimate of the position and velocity of the vehicle can be computed from purely geometrical relationships if the times of the observations are known. The estimation error due to the instrumentation noise can also be computed. In figure 11 this error is plotted as a function of the time of the second observation for noise magnitude σ_n = 1 sec arc in each angle. The position estimation error, figure 11(a), depends only on the second observation and increases with time, whereas the velocity estimation error, figure 11(b), generally decreases as the spacing increases as should be expected. The slight upturn in the velocity error curve at the end of the observation period indicates that there is a distinct optimum spacing. For the situation assumed, this spacing is about 0.055 day (1.32 hours). The increased error for larger spacing is due principally to the rapidly increasing error in position determination. For σ_n = 20 sec arc the errors in position and velocity estimation are about five times as great as for the optimal filter in the interval between 2 and 2-1/2 hours after injection. It should be noted that this error is almost as great as the expected deviation from the reference trajectory for the assumed magnitude of injection errors. Thus, under these conditions, this type of calculation adds very little to the knowledge of the trajectory and is not competitive with the optimal system. Of course, since the two-observation calculations are so simple, a digital computer is not required as it is for the optimal system. Thus, such a technique might be seriously considered as a standby which could be used during certain periods of the flight in case of a failure of the computer.

A Least Squares Method

The next method to be considered, taken from reference 8, is a least squares curve-fitting scheme. The idea here is to smooth the observations themselves (i.e., the measured angles), using a method of least squares. After a number of observations the angles would then be known with greater accuracy than they would be from any single observation, and could then be employed to obtain a navigation fix as in the minimum data method described above. When this idea is applied to the problem of estimating angles from the data obtained in a series of uniformly spaced observations, the precision of the angle estimates is given approximately by the following formula, taken from reference 8:

$$\sigma_m = \frac{\sigma_n(d + 1)}{\sqrt{k}} \tag{23}$$

where

σ_m standard deviation of the estimation error

σ_n standard deviation of instrumentation error

d order of polynomial that fits the true angle time history sufficiently accurately

k number of observations

From the estimated angles an estimate of the position of the vehicle can be computed. When σ_m is known for each of the three smoothed angle measurements, the precision of the estimate of vehicle position can be calculated readily if it is assumed that the estimated angles are independent random variables. This precision is plotted as a function of time in figure 11(a) with σ_m assumed to equal 1 second for each of the three angles. At 2-1/2 hours after injection (the time of the last estimate made by the optimal system), the precision is seen to be about 1.5 km per second of arc. At this time the optimal system with 20 observations, σ_n = 20 sec, gives a precision of about 4.8 km. Thus, to give the same precision, the least-squares method must result in a σ_m of 3.2 seconds of arc. If we assume that d = 1 is adequate to fit the angle time histories (a higher d gives a poorer estimate), it is seen from equation (23) that 156 observations are required to achieve the same accuracy as the optimal system obtains with 20 observations. A similar calculation for σ_n = 200 sec arc shows that about 1300 observations are necessary to match the optimal system performance.

A large number of observations is required with this technique because the method is not optimal, in large part because it does not make use of available statistical information regarding injection conditions. The large number of measurements also implies a close spacing of observation times which tends to invalidate the assumption of independent observations. Additional disadvantages of the technique are that it involves certain approximations that introduce errors the magnitudes of which are difficult to determine, and further, that the technique is not very flexible. Nevertheless, the computations required are perhaps simpler than those of Bayes estimation, although the processing of so much data is not necessarily an easy task.

An analysis similar to that given above could be applied to the determination of vehicle velocity with results similar to those obtained for position estimation. The details of such an analysis are of limited interest here and will not be developed.

CONCLUDING REMARKS

No attempt has been made here to present the theory of optimal filtering in a particularly sophisticated form. The idea has been primarily to describe the application of the theory in a manner readily understood by system design engineers. It is of particular interest that, unlike many applications of optimization theory, here the theory is

actually embodied in the system design rather than simply establishing a criterion for the assessment of the performance of systems designed in some other way. Thus, theoretical optimal performance is actually attainable (as long as the basic assumptions are not violated). However, it should be borne in mind that certain approximations might still be quite fruitful in simplifying the system design, and such matters deserve further consideration.

Some of the problems remaining in the practical design of a system utilizing the theory have already been mentioned. These include:

(1) Design of a digital computer to implement the computations

(2) Design of an instrumentation system and observation schedule

(3) Detailed consideration of the true nature of injection errors and instrumentation errors (including biases)

(4) Integration of the system into a complete guidance system, including such possible operating modes as abort, lunar orbiting and/or landing, rendezvous, and re-entry

Although the trajectory determination scheme has been described here in terms of guidance of a circumlunar vehicle, it is apparent that the same scheme is also applicable, perhaps with some practical modifications, to guidance problems for near-earth satellites and interplanetary vehicles. It also is not restricted to on-board applications, but could form the basis of an earth-based tracking system.

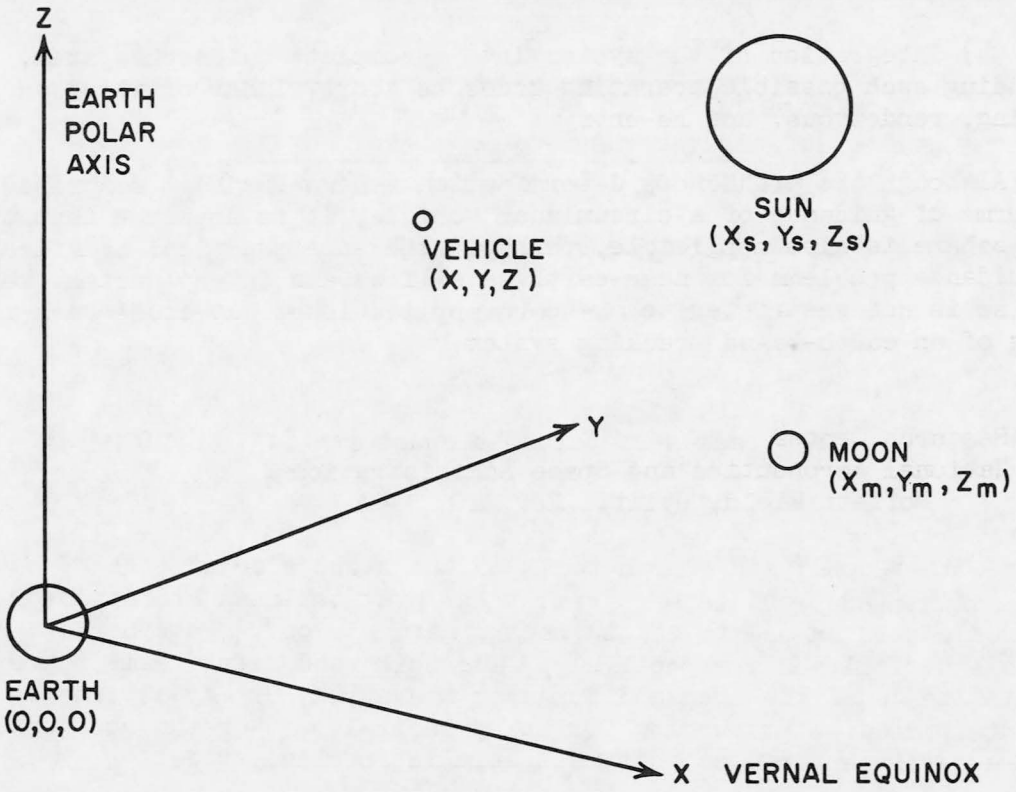
Ames Research Center
National Aeronautics and Space Administration
Moffett Field, Calif., Nov. 20, 1961

APPENDIX A

THE EQUATIONS OF MOTION

The equations of motion are derived on the basis of including the gravitational effects on the vehicle of the earth (including the second harmonic term of the earth's oblateness) and a spherical and homogeneous moon and sun.

The coordinate system chosen is that of a nonrotating Cartesian geocentric frame. The Z axis lies along the earth's polar axis, positive to the north. The X and Y axes are in the equatorial plane with the positive X axis in the direction of the first point of Aries and the Y axis oriented so as to produce a right-handed orthogonal system. A diagram of this coordinate system is given in the accompanying sketch.



The equations of motion expressed in the coordinate system described are as follows (see, e.g., ref. 9):

$$\ddot{X} = -\frac{\mu_e X}{r^3}\left[1 + J\left(\frac{a}{r}\right)^2\left(1 - 5\frac{Z^2}{r^2}\right)\right] - \frac{\mu_m(X - X_m)}{\Delta_m^3} - \frac{\mu_m X_m}{r_m^3} - \frac{\mu_s(X - X_s)}{\Delta_s^3} - \frac{\mu_s X_s}{r_s^3} \tag{A1}$$

$$\ddot{Y} = -\frac{\mu_e Y}{r^3}\left[1 + J\left(\frac{a}{r}\right)^2\left(1 - 5\frac{Z^2}{r^2}\right)\right] - \frac{\mu_m(Y - Y_m)}{\Delta_m^3} - \frac{\mu_m Y_m}{r_m^3} - \frac{\mu_s(Y - Y_s)}{\Delta_s^3} - \frac{\mu_s Y_s}{r_s^3} \tag{A2}$$

$$\ddot{Z} = -\frac{\mu_e Z}{r^3}\left[1 + J\left(\frac{a}{r}\right)^2\left(3 - 5\frac{Z^2}{r^2}\right)\right] - \frac{\mu_m(Z - Z_m)}{\Delta_m^3} - \frac{\mu_m Z_m}{r_m^3} - \frac{\mu_s(Z - Z_s)}{\Delta_s^3} - \frac{\mu_s Z_s}{r_s^3} \tag{A3}$$

where

$$r = \sqrt{X^2 + Y^2 + Z^2}$$

$$r_m = \sqrt{X_m^2 + Y_m^2 + Z_m^2}$$

$$r_s = \sqrt{X_s^2 + Y_s^2 + Z_s^2}$$

$$\Delta_m = \sqrt{(X - X_m)^2 + (Y - Y_m)^2 + (Z - Z_m)^2}$$

$$\Delta_s = \sqrt{(X - X_s)^2 + (Y - Y_s)^2 + (Z - Z_s)^2}$$

$$\mu_e = 3.986135 \times 10^{14} \text{ m}^3/\text{sec}^2$$

$$\mu_m = 4.89820 \times 10^{12} \text{ m}^3/\text{sec}^2$$

$$\mu_s = 1.3253 \times 10^{20} \text{ m}^3/\text{sec}^2$$

$$a = \text{radius of earth at equator} = 6.37826 \times 10^6 \text{m}$$

$$J = 1.6246 \times 10^{-3}$$

The first, second, and fourth terms on the right side of each of equations (A1), (A2), and (A3) represent the gravitational attraction upon the vehicle of an oblate earth (second harmonic only), a spherical moon, and a spherical sun, respectively. The third and fifth terms represent the influence of the moon and sun upon the earth, or may alternatively be interpreted as accounting for the principal part of the acceleration of the earth-centered coordinate system in inertial space.

APPENDIX B

THE LINEAR PERTURBATION EQUATION

The equations of motion, (A1) to (A3), are of the form:

$$\left.\begin{aligned} \ddot{X} &= f_1(X, Y, Z) \\ \ddot{Y} &= f_2(X, Y, Z) \\ \ddot{Z} &= f_3(X, Y, Z) \end{aligned}\right\} \qquad \text{(B1)}$$

To linearize these equations, we expand each in a Taylor series about a reference position, X_R, Y_R, Z_R, for example,

$$\ddot{X} = f_1(X_R, Y_R, Z_R) + \frac{\partial f_1}{\partial X}(X - X_R) + \frac{\partial f_1}{\partial Y}(Y - Y_R) + \frac{\partial f_1}{\partial Z}(Z - Z_R)$$

$$+ \text{higher order terms} \qquad \text{(B2)}$$

and similarly for the $\ddot{Y}$ and $\ddot{Z}$ equations. Here it is understood that the partial derivations are evaluated at the reference position. If the higher order terms are dropped (a reasonable approximation when the difference quantities $X - X_R$, etc., are small), the equations are linear in the difference quantities.

It is convenient to describe the state of this system of dynamical equations in terms of the difference quantities, remembering that X, Y, Z, X_R, Y_R, Z_R are all functions of time (i.e., the Taylor's series expansion is performed at each point in time, using a reference trajectory to specify the X_R, Y_R, Z_R quantities). Thus, the state is a six vector:

$$\{x\} = \begin{Bmatrix} x_1 \\ x_2 \\ x_3 \\ \cdot \\ \cdot \\ x_6 \end{Bmatrix}$$

where

$$\left.\begin{aligned} x_1 &= X - X_R \\ x_2 &= Y - Y_R \\ x_3 &= Z - Z_R \\ x_4 &= \dot{X} - \dot{X}_R = \dot{x}_1 \\ x_5 &= \dot{Y} - \dot{Y}_R = \dot{x}_2 \\ x_6 &= \dot{Z} - \dot{Z}_R = \dot{x}_3 \end{aligned}\right\} \qquad (B3)$$

Equations (B1) are then of the form

$$\left.\begin{aligned} \ddot{x}_1 &= \frac{\partial f_1}{\partial X} x_1 + \frac{\partial f_1}{\partial Y} x_2 + \frac{\partial f_1}{\partial Z} x_3 \\ \ddot{x}_2 &= \frac{\partial f_2}{\partial X} x_1 + \frac{\partial f_2}{\partial Y} x_2 + \frac{\partial f_2}{\partial Z} x_3 \\ \ddot{x}_3 &= \frac{\partial f_3}{\partial X} x_1 + \frac{\partial f_3}{\partial Y} x_2 + \frac{\partial f_3}{\partial Z} x_3 \end{aligned}\right\} \qquad (B4)$$

In the so-called standard form, equations (B4) appear as

$$\left.\begin{aligned} \dot{x}_1 &= x_4 \\ \dot{x}_2 &= x_5 \\ \dot{x}_3 &= x_6 \\ \dot{x}_4 &= \frac{\partial f_1}{\partial X} x_1 + \frac{\partial f_1}{\partial Y} x_2 + \frac{\partial f_1}{\partial Z} x_3 \\ \dot{x}_5 &= \frac{\partial f_2}{\partial X} x_1 + \frac{\partial f_2}{\partial Y} x_2 + \frac{\partial f_2}{\partial Z} x_3 \\ \dot{x}_6 &= \frac{\partial f_3}{\partial X} x_1 + \frac{\partial f_3}{\partial Y} x_2 + \frac{\partial f_3}{\partial Z} x_3 \end{aligned}\right\} \qquad (B5)$$

or in short-hand notation

$$\dot{x}(t) = F(t)x(t) \qquad (B6)$$

where $x(t)$ is a six vector and $F(t)$ is a (time-varying) 6×6 matrix defined as follows:

$$F = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ \frac{\partial f_1}{\partial X} & \frac{\partial f_1}{\partial Y} & \frac{\partial f_1}{\partial Z} & 0 & 0 & 0 \\ \frac{\partial f_2}{\partial X} & \frac{\partial f_2}{\partial Y} & \frac{\partial f_2}{\partial Z} & 0 & 0 & 0 \\ \frac{\partial f_3}{\partial X} & \frac{\partial f_3}{\partial Y} & \frac{\partial f_3}{\partial Z} & 0 & 0 & 0 \end{bmatrix} \qquad (B7)$$

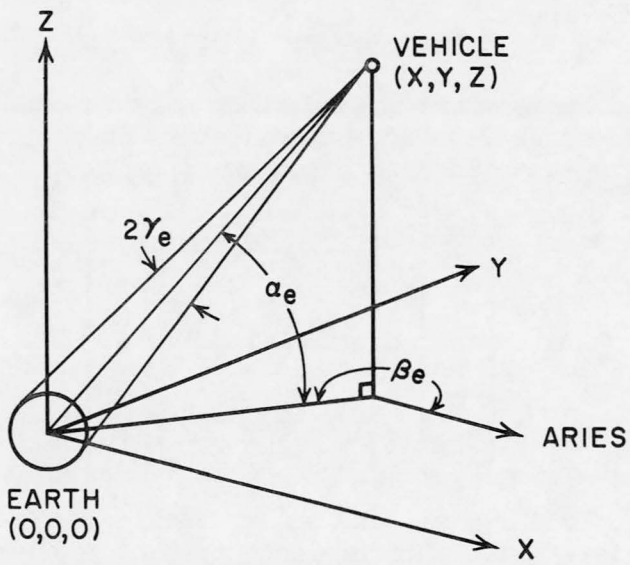
If the partials in F are evaluated along the reference trajectory, they are merely functions of time once this reference has been selected. If the reference for which the partials are evaluated is the estimated trajectory, then the partials must be computed as the flight progresses since the estimate is not known a priori and is subject to change as each observation is made.

APPENDIX C

RELATIONS BETWEEN SPACE ANGLES AND POSITION

The system under study employs optical on-board instrumentation capable of measuring the directions in space of lines of sight from the vehicle to selected celestial bodies. Within this restriction there is no attempt made here to optimize the choice of angles to be measured. A particular plausible set of angles is selected simply to permit examination of the behavior of the trajectory determination system. This is necessary because the relationships between the location of the vehicle and the angles measured are an integral part of the system. Derivations of these relations are developed in this appendix for the chosen set of angles.

Because the observational period studied here is such that the vehicle is relatively close to the earth, it is natural to conceive of an instrumentation system that involves looking at the earth. Such a system could conceivably provide simultaneous measurements of the direction of the vehicle-earth line of sight and the subtended earth angle. The geometry of this situation is illustrated in the accompanying sketch



where the direction of the line of sight is specified by the angles α_e and β_e, and the subtended earth angle is $2\gamma_e$. The angle β_e is assumed to be measured clockwise from Aries (the X axis), and α_e is taken to be positive if the vehicle is below the equatorial plane (i.e., $Z < 0$).

There are many possible instrumentation arrangements that could provide measurements (either directly or indirectly) of α_e, β_e, and γ_e,

the fundamental difference being in the way the inaccuracies in the measurements (i.e., the noise) enter the system. To avoid involvement in the details of specifying a particular instrumentation scheme, we simply assume, without any considerations of practicality, that the three angles are measured independently, with the same kind of Gaussian errors in each angle. The noise covariance matrix is therefore diagonal.

The equations which relate the angles α_e, β_e, γ_e to vehicle positions are readily derived from the geometry. They are:

$$\left.\begin{aligned} \alpha_e &= -\sin^{-1}\left(\frac{Z}{R}\right) \\ \beta_e &= \begin{cases} \sin^{-1} - \dfrac{Y}{(X^2 + Y^2)^{1/2}} \\ \cos^{-1} - \dfrac{X}{(X^2 + Y^2)^{1/2}} \end{cases} \\ \gamma_e &= \sin^{-1}\left(\frac{R_o}{R}\right) \end{aligned}\right\} \qquad \text{(C1)}$$

where R_o = radius of earth

$$R = (X^2 + Y^2 + Z^2)^{1/2}$$

The linear perturbation form of equations (C1) is obtained by Taylor's series expansion about a reference trajectory in the same manner as described for the equations of motion in appendix A. Thus,

$$\begin{Bmatrix} \delta\alpha_e \\ \delta\beta_e \\ \delta\gamma_e \end{Bmatrix} = \begin{bmatrix} \frac{\partial\alpha_e}{\partial X} & \frac{\partial\alpha_e}{\partial Y} & \frac{\partial\alpha_e}{\partial Z} \\ \frac{\partial\beta_e}{\partial X} & \frac{\partial\beta_e}{\partial Y} & \frac{\partial\beta_e}{\partial Z} \\ \frac{\partial\gamma_e}{\partial X} & \frac{\partial\gamma_e}{\partial Y} & \frac{\partial\gamma_e}{\partial Z} \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \\ x_3 \end{Bmatrix} \qquad \text{(C2)}$$

where $\delta\alpha_e$, $\delta\beta_e$, $\delta\gamma_e$, x_1, x_2, and x_3 are deviations from the angles and positions associated with the reference trajectory, and the partials are evaluated along this trajectory. The 3×3 matrix of partial derivatives is the H_1 matrix of the body of the report and is, of course, time-varying. Explicit expressions for the partials are obtained by partial differentiation of equations (C1) and are tabulated in the following table:

Quantity (q)	$\frac{\partial q}{\partial X}$	$\frac{\partial q}{\partial Y}$	$\frac{\partial q}{\partial Z}$
α_e	$\frac{XZ}{R^2\sqrt{X^2 + Y^2}}$	$\frac{YZ}{R^2\sqrt{X^2 + Y^2}}$	$\frac{Z^2 - R^2}{R^2\sqrt{X^2 + Y^2}}$
β_e	$\frac{-Y}{X^2 + Y^2}$	$\frac{X}{X^2 + Y^2}$	0
γ_e	$\frac{-R_oX}{R^2\sqrt{R^2 - R_o^2}}$	$\frac{-R_oY}{R^2\sqrt{R^2 - R_o^2}}$	$\frac{-R_oZ}{R^2\sqrt{R^2 - R_o^2}}$

$R^2 = X^2 + Y^2 + Z^2$

R_o = radius of earth

It may be noted parenthetically that the three angles α_e, β_e, γ_e, are sufficient for a position fix. Although the computation of such a fix is not necessary in the data reduction method employed in this report, it is used in some navigation techniques. Equations for calculating the fix are obtained by inverting equations (C1):

$$\left.\begin{aligned} X &= -\frac{R_o\cos\alpha_e\cos\beta_e}{\sin\gamma_e} \\ Y &= -\frac{R_o\cos\alpha_e\sin\beta_e}{\sin\gamma_e} \\ X &= -\frac{R_o\sin\alpha_e}{\sin\gamma_e} \end{aligned}\right\} \qquad (C3)$$

APPENDIX D

SIMPLIFICATION OF THE ESTIMATION EQUATIONS FOR UNCORRELATED INSTRUMENT ERRORS

In the situation of uncorrelated instrument errors, it is readily shown that the estimate $\hat{n}$ is not required. Suppose the estimate $\hat{x}^*(t_{k-1})$ is represented in partitioned form

$$\hat{x}^*(t_{k-1}) = \begin{Bmatrix} \hat{x}(t_{k-1}) \\ \hat{n}(t_{k-1}) \end{Bmatrix} \tag{D1}$$

The operation $\Phi^*(t_k; t_{k-1})\hat{x}^*(t_{k-1})$ indicated in equation (9) can thus be represented in the form

$$\begin{bmatrix} \Phi & 0 \\ 0 & \Phi_n \end{bmatrix} \begin{Bmatrix} \hat{x} \\ \hat{n} \end{Bmatrix} = \begin{Bmatrix} \Phi\hat{x} \\ \Phi_n\hat{n} \end{Bmatrix} \tag{D2}$$

where arguments have been omitted for simplicity. Now, since for uncorrelated instrument errors $\Phi_n(t_k; t_{k-1}) \equiv 0$, it is apparent that the updated estimate of n is simply zero and there is no point in implementing this portion of the estimation calculations.

The same sort of analysis applies to the computation of the P^* and K^* matrices. If the matrix in braces in equation (11) is denoted as $P^{*\prime}$, this equation can be rewritten

$$P^* = \Phi P^{*\prime} \Phi^T + Q^* \tag{D3}$$

where $P^{*\prime}$ can be written in the partitioned form

$$P^{*\prime} = \begin{bmatrix} P_s & P_{sn} \\ P_{sn}^T & P_n \end{bmatrix} \tag{D4}$$

The subscripts s and n refer, respectively, to errors in the estimate of x and n, and sn refers to correlation between these errors. Note that any covariance matrix, such as $P^{*\prime}$, is symmetric, and therefore the off-diagonal submatrices can be expressed as transposes of each other. Equation (D3) can then be written

$$P^* = \begin{bmatrix} \Phi & 0 \\ 0 & \Phi_n \end{bmatrix} \begin{bmatrix} P_s & P_{sn} \\ P_{sn}^T & P_n \end{bmatrix} \begin{bmatrix} \Phi^T & 0 \\ 0 & \Phi_n^T \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & Q' \end{bmatrix}$$

$$= \begin{bmatrix} \Phi P_s \Phi^T & \Phi P_{sn} \Phi_n^T \\ \Phi_n P_{sn}^T \Phi^T & \Phi_n P_n \Phi_n^T + Q' \end{bmatrix} \tag{D5}$$

Obviously, if $\Phi_n = 0$, this reduces to

$$P^* = \begin{bmatrix} \Phi P_s \Phi^T & 0 \\ 0 & Q' \end{bmatrix} \tag{D6}$$

The parts of P^* which refer to the errors in estimating x and n are thus uncorrelated, and may be computed separately. The matrix Q' is computed from equation (6), where it is seen that for uncorrelated errors the lower limit of the integral can be replaced by $-\infty$. Thus, in this case Q' is a function only of t, and is designated $Q(t)$. Since $Q(t)$ may be assumed to be known a priori this computation may be omitted.

The computation can be further simplified when the computation of P_s is examined in detail. In partitioned form, the operation MP^*M^T can be written

$$MP^*M^T = [H \quad D] \begin{bmatrix} P & 0 \\ 0 & Q \end{bmatrix} \begin{bmatrix} H^T \\ D^T \end{bmatrix} = [HPH^T + DQD^T] = B \tag{D7}$$

where

$$P = \Phi P_s \Phi^T \tag{D8}$$

The operation to compute $P^{*\prime}$ (the matrix in braces in eq. (11)) is then:

$$P^{*\prime} = P^* - P^* M^T B^{-1} M P^*$$

$$= \begin{bmatrix} P & 0 \\ 0 & Q \end{bmatrix} - \begin{bmatrix} PH^T B^{-1} HP & PH^T B^{-1} DQ \\ QD^T B^{-1} HP & QD^T B^{-1} DQ \end{bmatrix} \tag{D9}$$

From a comparison of equations (D4) and (D9), it is clear that the submatrix P_s is computed by the relation $P - PH^T B^{-1} HP$. Substituting this into equation (D8) then gives the recursion relation for P as:

$$P(t_{k+1}) = \Phi(t_{k+1};\, t_k)[P(t_k) - P(t_k)\, H^T(t_k) B^{-1} H(t_k) P(t_k)] \Phi^T(t_{k+1};\, t_k) \tag{D10}$$

where

$$B = [H(t_k)P(t_k)H^T(t_k) + Q(t_k)]$$

Here a further simplification has been introduced; because Q appears in the equations only in the form DQD^T, no generality is lost by redefining Q to be DQD^T.

The K^* matrix may be written

$$K^* = P^*M^TB^{-1} = \begin{bmatrix} P & 0 \\ 0 & Q \end{bmatrix} \begin{bmatrix} H^T \\ D^T \end{bmatrix} B^{-1} = \begin{bmatrix} PH^TB^{-1} \\ QD^TB^{-1} \end{bmatrix} \tag{D11}$$

Only the upper part of this has to do with incrementing the estimate of x. Designating this portion of K^* as K, we then have

$$K = PH^TB^{-1} \tag{D12}$$

Following the same procedure, the operation $M\Phi^*\hat{x}^*$ in equation (9) is reduced to $H\Phi\hat{x}$.

With the simplifications given above, the optimal estimation equations become

$$\hat{x}(t_k) = \Phi(t_k; t_{k-1})\hat{x}(t_{k-1}) + K(t_k)[y(t_k) - H(t_k)\Phi(t_k; t_{k-1})\hat{x}(t_{k-1})] \tag{D13}$$

$$K(t_k) = P(t_k)H^T(t_k)[H(t_k)P(t_k)H^T(t_k) + Q(t_k)]^{-1} \tag{D14}$$

$$P(t_{k+1}) = \Phi(t_{k+1}; t_k)[P(t_k) - K(t_k)H(t_k)P(t_k)]\Phi^T(t_{k+1}; t_k) \tag{D15}$$

APPENDIX E

COMPUTATION OF THE TRANSITION MATRIX

All solutions of the linear differential equation (B6) can be written in the form

$$x(t) = \Phi(t;\, t_o)x(t_o) \tag{E1}$$

where $x(t_o)$ is a vector initial condition at time t_o, $x(t)$ is the vector state variable at time t, and $\Phi(t;\, t_o)$ is the transition matrix which relates the two. As was shown in the main body of this report, we are principally interested in the transition from one observation to the next. Since the transition matrix depends on the particular trajectory the vehicle is on, and since the observation times are arbitrary, it is desirable that Φ be continuously calculated in the vehicle.

It is seen in equation (E1) that if an initial condition of unity is put on x_1 at time t_o with all other components of x set equal to zero, then the ensuing time history of x is the first column of the $\Phi(t;\, t_o)$ matrix. This is demonstrated as follows. If we define

$$\Phi(t;\, t_o) = \begin{bmatrix} \varphi_{11} & \varphi_{12} & \cdots & \varphi_{16} \\ \varphi_{21} & \varphi_{22} & \cdots & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \varphi_{61} & \cdot & \cdots & \varphi_{66} \end{bmatrix} \tag{E2}$$

where each element φ_{ij} is a function of t and t_o, then we see that

$$\begin{Bmatrix} \varphi_{11} \\ \varphi_{21} \\ \varphi_{31} \\ \varphi_{41} \\ \varphi_{51} \\ \varphi_{61} \end{Bmatrix} = \Phi(t;\, t_o) \begin{Bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{Bmatrix} \tag{E3}$$

This is equivalent to introducing the proper initial conditions into the perturbation equations and integrating numerically, a procedure readily implemented by means of a digital computer. If this is done simultaneously for six sets of perturbation equations, with each set having a unit initial condition on one of the components of x and zero on all the others, all six columns of the transition matrix can be generated continuously. The transition matrix from the latest observation is always available if the initial conditions are reset after each observation.

APPENDIX F

RECURSION RELATIONS FOR A BAYES ESTIMATION PROCEDURE

To formulate Bayes estimation equations, we begin with the same assumptions employed in the development of the optimal filter theory. These are:

(1) The equations of motion are linearized with respect to a reference trajectory, and the state and observables are described in terms of deviations from the reference. Thus, the state is given at time t_k by

$$x_k = \Phi_k x_o \tag{F1}$$

where Φ_k is the 6×6 transition matrix from time t_o to time t_k. The set of deviation observables at time t_k is given by

$$m_k = H_k x_k \tag{F2}$$

where H_k is the matrix of partial derivatives of the observables with respect to the state variables at time t_k.

(2) An observation is made at time t_k with additive error uncorrelated with errors at successive observation times:

$$y_k = H_k x_k + n_k \tag{F3}$$

Here, if y_k is a j vector (i.e., j components in the observation) and x_k is a 6 vector, H_k is a j×6 matrix.

A series of k observations of the type described might then be represented in the form

$$y = H\Phi x_o + n \tag{F4}$$

which may be partitioned

$$\begin{Bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_k \end{Bmatrix} = \begin{bmatrix} H_1\Phi_1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & H_2\Phi_2 & & & & \\ \cdot & & \cdot & & & \\ \cdot & & & \cdot & & \\ \cdot & & & & \cdot & \\ 0 & & & & & H_k\Phi_k \end{bmatrix} \begin{Bmatrix} x_o \end{Bmatrix} + \begin{Bmatrix} n_1 \\ n_2 \\ \cdot \\ \cdot \\ \cdot \\ n_k \end{Bmatrix} \tag{F5}$$

If y_i and n_i are j vectors, and x_o is a 6 vector, the $H_i\Phi_i$ are $j\times 6$ matrices. When Gaussian injection errors and noise are assumed, with zero means, the statistical description of the random vectors x_o and n is given by the covariance matrices

$$\text{cov}[x_o, x_o] = P_o$$

$$\text{cov}[n, n] = Q$$

If the n_i are uncorrelated (i.e., observations uncorrelated), Q may be partitioned in the form

$$Q = \begin{bmatrix} Q_1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & Q_2 & & & & \\ \cdot & & \cdot & & & \\ \cdot & & & \cdot & & \\ \cdot & & & & \cdot & \\ 0 & & & & & Q_k \end{bmatrix}$$

Suppose it is desired to estimate the injection state, x_o, from the series of observations, y. When the problem is considered probabilistically, it is evident that all the information about the initial condition, x_o, conveyed by the observations, y, is contained in the so-called a posteriori probability density function, written as $p(x_o|y)$. Now, from decision theory it can be shown (e.g., ref. 5) that to minimize a mean-square-error loss function (or more correctly, a quadratic loss function in the case of a multidimensional estimation problem) the estimate of x_o, which may be called $\hat{x}_o$, is the mean of the a posteriori random variable $(x_o|y)$. This is called a Bayes estimate in decision theory.

To obtain an analytical expression for $p(x_o|y)$, Bayes theorem is employed:

$$p(x_o|y) = \frac{p(y|x_o)p(x_o)}{p(y)} \tag{F6}$$

Since the noise and injection errors are assumed Gaussian and independent, the density functions in equation (F6) can be written (e.g., see ref. 7, p. 13):

$$p(y|x_o) = p(n) = p(y - H\Phi x_o)$$

$$= \frac{1}{[(2\pi)^{jk}|Q|]^{1/2}} \exp\left[-\frac{1}{2}(y^T - x_o^T \Phi^T H^T)Q^{-1}(y - H\Phi x_o)\right] \tag{F7}$$

$$p(x_o) = \frac{1}{[(2\pi)^6 |P_o|]^{1/2}} \exp\left[-\frac{1}{2} x_o^T P_o^{-1} x_o\right] \tag{F8}$$

$$p(y) = \frac{1}{[(2\pi)^{jk} |Q + \Gamma|]^{1/2}} \exp\left[-\frac{1}{2} y^T (Q + \Gamma)^{-1} y\right] \tag{F9}$$

where

$\Gamma = \text{cov}[H\Phi x, H\Phi x] = H\Phi P_o \Phi^T H^T$

j = number of measurements in an observation

k = number of observations

Thus, the a posteriori density function is

$$p(x_o|y) = \frac{|Q + \Gamma|^{1/2}}{(2\pi)^3 |P_o|^{1/2} |Q|^{1/2}} \exp\left[-\frac{1}{2} x_o^T P_o^{-1} x_o - \frac{1}{2} (y^T - x_o^T \Phi^T H^T) Q^{-1} (y - H\Phi x_o) + \frac{1}{2} y^T (Q + \Gamma)^{-1} y\right] \tag{F10}$$

which, after some manipulation, becomes

$$p(x_o|y) = \frac{|Q + \Gamma|^{1/2}}{(2\pi)^3 |P_o|^{1/2} |Q|^{1/2}} \exp\left[-\frac{1}{2} (x_o^T - y^T Q^{-1} H\Phi\Lambda)\Lambda^{-1} (x_o - \Lambda\Phi^T H^T Q^{-1} y)\right] \tag{F11}$$

where

$$\Lambda = (P_o^{-1} + \Phi^T H^T Q^{-1} H\Phi)^{-1}$$

Clearly, the mean of the random variable $(x_o|y)$ is

$$\hat{x}_o = \Lambda\Phi^T H^T Q^{-1} y \tag{F12}$$

Also, the covariance matrix of the error in estimate, $\tilde{x}_o = x_o - \hat{x}_o$, is Λ.

If the estimate of the state at time t_k is desired, this is given by

$$\hat{x}_k = \Phi_k \Lambda \Phi^T H^T Q^{-1} y \tag{F13}$$

where Φ_k is the 6×6 transition matrix from t_o to t_k. The covariance matrix of the error in estimate is

$$\Lambda'_k = \Phi_k \Lambda \Phi_k^T \tag{F14}$$

These explicit formulas for the estimate $\hat{x}_k$ and the statistics of the error in estimate involve the multiplication and inversion of matrices of rather high order when there are a large number of observations. Some simplification in this regard is possible under the assumption of independent observations if equations (F13) and (F14) are written in terms of the partitioned parts of Q, H, and Φ:

$$\hat{x}_k = \Phi_k \left(P_o^{-1} + \sum_{i=1}^{k} \Phi_i^T H_i^T Q_i^{-1} H_i \Phi_i \right)^{-1} \sum_{i=1}^{k} \Phi_i^T H_i^T Q_i^{-1} y_i \qquad \text{(F15)}$$

$$\Lambda_k' = \Phi_k \left(P_o^{-1} + \sum_{i=1}^{k} \Phi_i^T H_i^T Q_i^{-1} H_i \Phi_i \right)^{-1} \Phi_k^T \qquad \text{(F16)}$$

Now if it is supposed that the estimation calculations are to be performed on a step-by-step basis - that is, the state estimate is "improved" at the time each new observation is obtained - it is clear that the previous estimate is always available. For instance, at time t_k when the kth observation is made, the estimate $\hat{x}_{k-1}$ of the state at t_{k-1} has already been computed on the basis of the first k-1 observations:

$$\hat{x}_{k-1} = \Phi_{k-1} \left(P_o^{-1} + \sum_{i=1}^{k-1} \Phi_i^T H_i^T Q_i^{-1} H_i \Phi_i \right)^{-1} \sum_{i=1}^{k-1} \Phi_i^T H_i^T Q_i^{-1} y_i \qquad \text{(F17)}$$

Also, the estimate of x_k based on k-1 observations is obtained simply by updating $\hat{x}_{k-1}$ to time t_k:

$$\Phi_k \Phi_{k-1}^{-1} \hat{x}_{k-1} = \Phi_k \left(P_o^{-1} + \sum_{i=1}^{k-1} \Phi_i^T H_i^T Q_i^{-1} H_i \Phi_i \right)^{-1} \sum_{i=1}^{k-1} \Phi_i^T H_i^T Q_i^{-1} y_i \qquad \text{(F18)}$$

where $\Phi_k \Phi_{k-1}^{-1}$ is simply the transition matrix from time t_{k-1} to t_k. Next, it is seen that equations (F15) and (F18) may be rewritten as:

$$\left(P_o^{-1} + \sum_{i=1}^{k} \Phi_i^T H_i^T Q_i^{-1} H_i \Phi_i \right) \Phi_k^{-1} \hat{x}_k = \sum_{i=1}^{k} \Phi_i^T H_i^T Q_i^{-1} y_i \qquad \text{(F19)}$$

$$\left(P_o^{-1} + \sum_{i=1}^{k-1} \Phi_i^T H_i^T Q_i^{-1} H_i \Phi_i \right) \Phi_{k-1}^{-1} \hat{x}_{k-1} = \sum_{i=1}^{k-1} \Phi_i^T H_i^T Q_i^{-1} y_i \qquad \text{(F20)}$$

Subtracting equation (F20) from (F19) gives

$$\left(P_o^{-1} + \sum_{i=1}^{k} \Phi_i^T H_i^T Q_i^{-1} H_i \Phi_i\right)\left(\Phi_k^{-1} \hat{x}_k - \Phi_{k-1}^{-1} \hat{x}_{k-1}\right)$$

$$+ \Phi_k^T H_k^T Q_k^{-1} H_k \Phi_k \Phi_{k-1}^{-1} \hat{x}_{k-1} = \Phi_k^T H_k^T Q_k^{-1} y_k$$

or

$$\hat{x}_k = \Phi_k \Phi_{k-1}^{-1} \hat{x}_{k-1} + \Lambda_k' H_k^T Q_k^{-1} (y_k - H_k \Phi_k \Phi_{k-1}^{-1} \hat{x}_{k-1}) \qquad \text{(F21)}$$

It is seen that this equation would be identical to equation (12) in the text, developed from filter theory, if

$$\Lambda_k' H_k^T Q_k^{-1} = K(t_k) = P(t_k) H_k^T [H_k P(t_k) H_k^T + Q_k]^{-1}$$

That this relationship is true will be proved here in two steps: First, it will be shown that Λ_k', the covariance matrix of Bayes estimation errors, is equal to the $P'(t_k)$ covariance matrix; this is proved by showing that Λ_k' and $P'(t_k)$ satisfy the same recursion equations. Second, it is necessary to prove the matrix relationship

$$\Lambda_k' H_k^T Q_k^{-1} = \Lambda_k H_k^T [H_k \Lambda_k H_k^T + Q_k]^{-1}$$

where $\Lambda_k' = \Lambda_k + (\Delta\Lambda)_k$, and $(\Delta\Lambda)_k$ is the change in the covariance matrix of estimation errors which occurs when the observation is made. (In other words, two covariance matrices at time t_k are considered: Λ_k based on $k-1$ observations, and Λ_k' based on k observations.)

To develop a recursion formula for Λ_k', it is noted that the change in this matrix due to the kth observation is given by

$$(\Delta\Lambda)_k = \Lambda_k' - \Lambda_k = \Phi_k \left[\left(P_o^{-1} + \sum_{i=1}^{k} \Psi_i\right)^{-1} - \left(P_o^{-1} + \sum_{1}^{k-1} \Psi_i\right)^{-1}\right] \Phi_k^T \qquad \text{(F22)}$$

where

$$\Lambda_k = (\Phi_k \Phi_{k-1}^{-1}) \Lambda_{k-1}' (\Phi_k \Phi_{k-1}^{-1})^T \qquad \text{(F23)}$$

is the covariance matrix before the information from the kth observation has been included, and $\Psi_i = \Phi_i^T H_i^T Q_i^{-1} H_i \Phi_i$. Equation (F22) can be rearranged to give

$$\left(P_o^{-1} + \sum_1^k \Psi_i\right) \Phi_k^{-1} (\Delta\Lambda)_k \Phi_k^{T-1} \left(P_o^{-1} + \sum_1^{k-1} \Psi_i\right) = -\Psi_k$$

$$\left(\Phi_{k-1}^T \Lambda_{k-1}^{'-1} \Phi_{k-1} + \Psi_k\right) \Phi_k^{-1} (\Delta\Lambda)_k \Phi_k^{T-1} \left(\Phi_{k-1}^T \Lambda_{k-1}^{'-1} \Phi_{k-1}\right) = -\Psi_k$$

$$(\Delta\Lambda)_k = -\Phi_k \left[\left(\Phi_{k-1}^T \Lambda_{k-1}^{'-1} \Phi_{k-1} + \Psi_k\right)^{-1} \Psi_k \left(\Phi_{k-1}^{-1} \Lambda_{k-1}^{'} \Phi_{k-1}^{T-1}\right)\right] \Phi_k^T$$

$$= -\left(\Phi_k^{T-1} \Phi_{k-1}^T \Lambda_{k-1}^{'-1} \Phi_{k-1} \Phi_k^{-1} + H_k^T Q_k^{-1} H_k\right)^{-1} H_k^T Q_k^{-1} H_k \left(\Phi_k^{T-1} \Phi_{k-1}^T \Lambda_{k-1}^{'-1} \Phi_{k-1} \Phi_k^{-1}\right)^{-1}$$

$$= -\Lambda_k \left(I + H_k^T Q_k^{-1} H_k \Lambda_k\right)^{-1} H_k^T Q_k^{-1} H_k \Lambda_k \qquad \text{(F24)}$$

Now the matrix identity

$$H_k^T Q_k^{-1} H_k \equiv H_k^T Q_k^{-1} \left(H_k \Lambda_k H_k^T + Q_k\right) \left(H_k \Lambda_k H_k^T + Q_k\right)^{-1} H_k$$

$$= \left(H_k^T Q_k^{-1} H_k \Lambda_k + I\right) H_k^T \left(H_k \Lambda_k H_k^T + Q_k\right)^{-1} H_k$$

or

$$\left(H_k^T Q_k^{-1} H_k \Lambda_k + I\right)^{-1} H_k^T Q_k^{-1} H_k = H_k^T \left(H_k \Lambda_k H_k^T + Q_k\right)^{-1} H_k \qquad \text{(F25)}$$

is substituted into equation (F24) to obtain

$$(\Delta\Lambda)_k = -\Lambda_k H_k^T \left(H_k \Lambda_k H_k^T + Q_k\right)^{-1} H_k \Lambda_k$$

Thus,

$$\Lambda_k' = \Lambda_k - \Lambda_k H_k^T \left(H_k \Lambda_k H_k^T + Q_k\right)^{-1} H_k \Lambda_k \qquad \text{(F26)}$$

The recursion formulas (F23) and (F26) for Λ_k are seen to be identical to the equivalent equations, (15) and (16), for the $P(t_k)$ matrix. Therefore, $P(t_k)$ is exactly the same as Λ_k.

The second portion of the proof is developed as follows: Expression (F26) is substituted into the weighting function $\Lambda_k' H_k^T Q_k^{-1}$, with the result that

$$\Lambda_k' H_k^T Q_k^{-1} = \left[\Lambda_k - \Lambda_k H_k^T \left(H_k \Lambda_k H_k^T + Q_k\right)^{-1} H_k \Lambda_k\right] H_k^T Q_k^{-1}$$

$$= \Lambda_k H_k^T \left[Q_k^{-1} - \left(H_k \Lambda_k H_k^T + Q_k\right)^{-1} H_k \Lambda_k H_k^T Q_k^{-1}\right]$$

$$= \Lambda_k H_k^T \left[\left(H_k \Lambda_k H_k^T + Q_k\right)^{-1} \left(H_k \Lambda_k H_k^T + Q_k\right) Q_k^{-1}\right.$$

$$\left. - \left(H_k \Lambda_k H_k^T + Q_k\right)^{-1} H_k \Lambda_k H_k^T Q_k^{-1}\right]$$

$$= \Lambda_k H_k^T \left[\left(H_k \Lambda_k H_k^T + Q_k\right)^{-1} \left(H_k \Lambda_k H_k^T Q_k^{-1} + I\right.\right.$$

$$\left.\left. - H_k \Lambda_k H_k^T Q_k^{-1}\right)\right]$$

$$= \Lambda_k H_k^T \left(H_k \Lambda_k H_k^T + Q_k\right)^{-1} \qquad \text{(F27)}$$

This completes the proof. Thus, the Bayes estimate is seen to be identical to the estimate obtained by the filter theory approach.

REFERENCES

1. McLean, John D., Schmidt, Stanley F., and McGee, Leonard A.: Optimal Filtering and Linear Prediction Applied to a Midcourse Navigation System for the Circumlunar Mission. NASA TN D-1208, 1962.

2. Kalman, R. E.: A New Approach to Linear Filtering and Prediction Problems. Jour. of Basic Engr., vol. 82, no. 1, March 1960, pp. 35-50. (Also pub. as: RIAS, Inc., Monograph 60-11; Office of Scientific Research TN 59-268; and ASME Paper 59-IRD-11.)

3. Kalman, R. E., and Bucy, R. S.: New Results in Linear Filtering and Prediction Theory. Jour. of Basic Engr., vol. 83, no. 1, March 1961, pp. 95-108. (Also pub. as: RIAS, Inc., Monograph 61-8; and ASME Paper 60-JAC-12.)

4. Carr, Russell E., and Hudson, R. Henry: Tracking and Orbit-Determination Program of the Jet Propulsion Laboratory. TR 32-7, Jet Propulsion Lab., Pasadena, Calif., Feb. 22-26, 1960.

5. Abramson, N. M.: Seminar on Decision Theory and Its Applications to Communication Theory. RJ-MR-8, IBM, San Jose, Calif., Jan. 1958.

6. Blackwell, David, and Girshick, M. A.: Theory of Games and Statistical Decisions. John Wiley and Sons, Inc., N. Y., 1954.

7. Shapiro, I. I.: The Prediction of Ballistic Missile Trajectories From Radar Observations. McGraw-Hill Book Co., N. Y., 1957.

8. Proschan, Frank: Precision of Missile Least Squares Position and Velocity Estimates. Tech. MEMO EDL-M263, Electronic Defense Lab., Inc., Mountain View, Calif., Feb. 29, 1960.

9. Baker, Robert M. L., Jr., and Makemson, Maud W.: An Introduction to Astrodynamics. Academic Press, Inc., N. Y. and London, 1960.

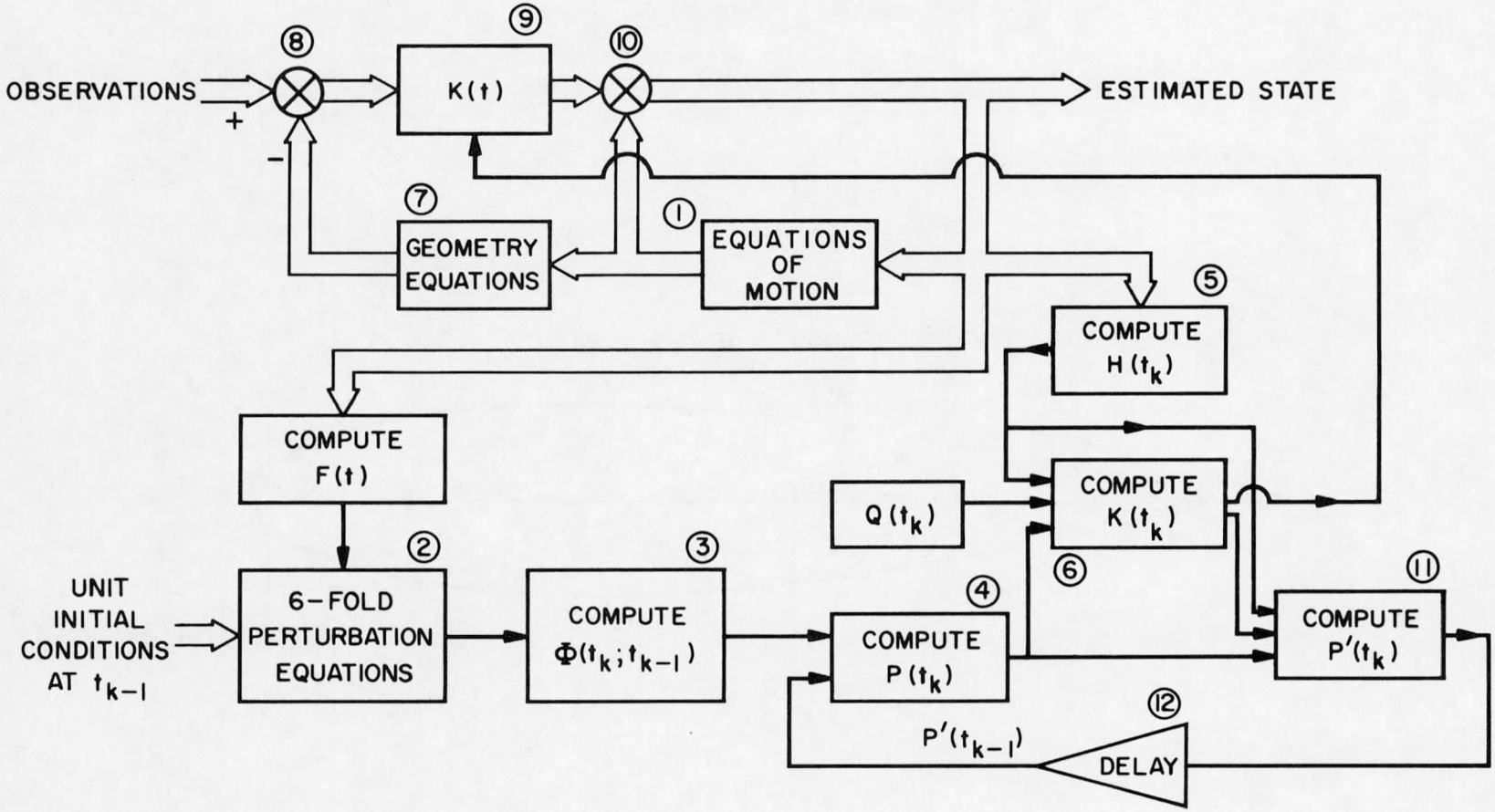


Figure 1.- Block diagram of trajectory estimation system.

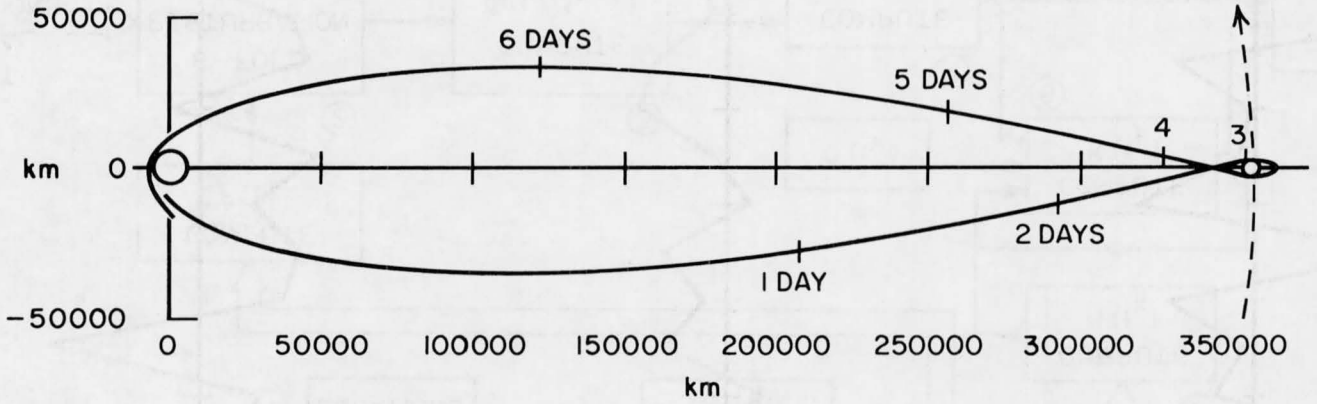


Figure 2.- Reference trajectory for circumlunar mission.

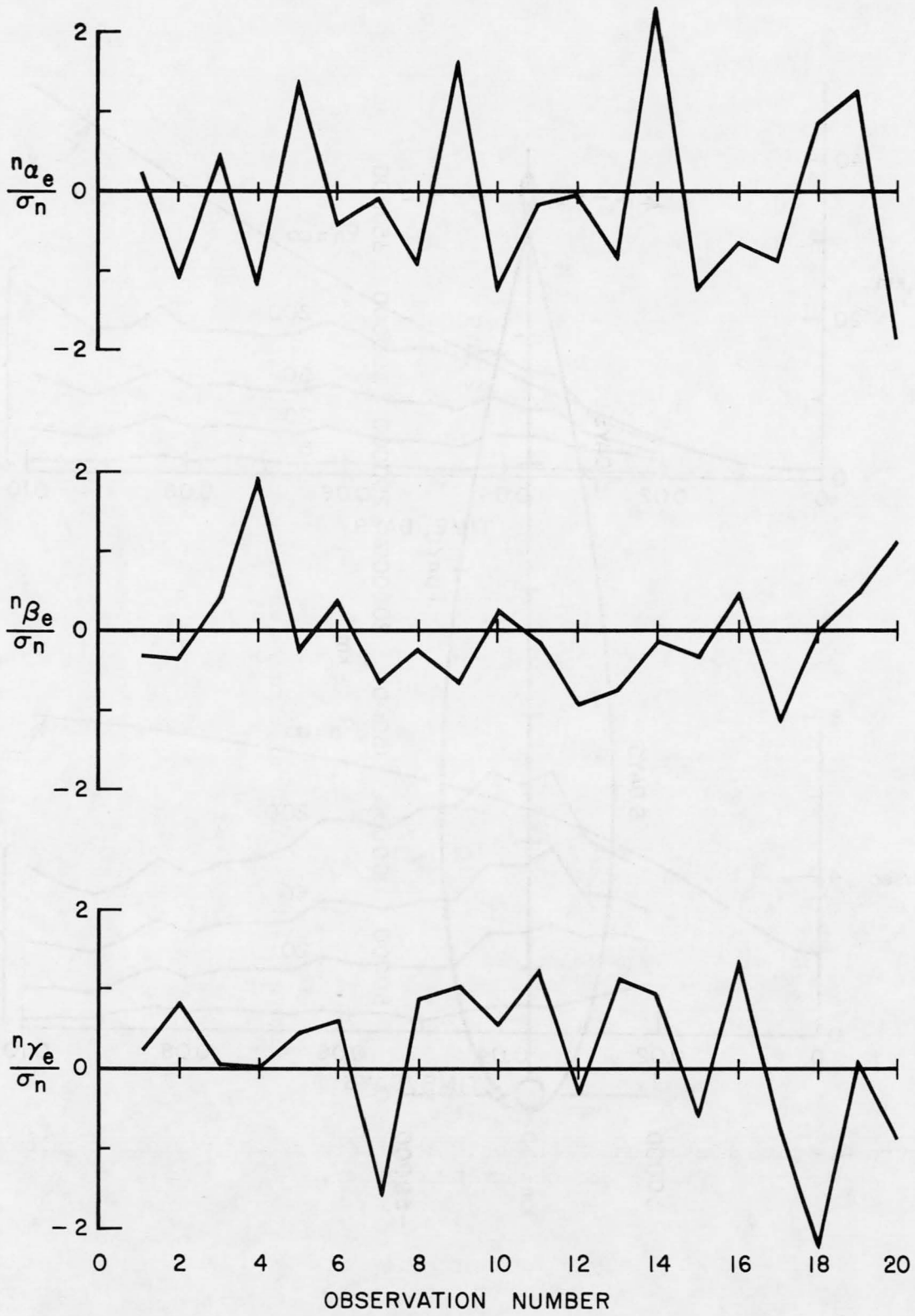


Figure 3.- Time histories of observation errors assumed.

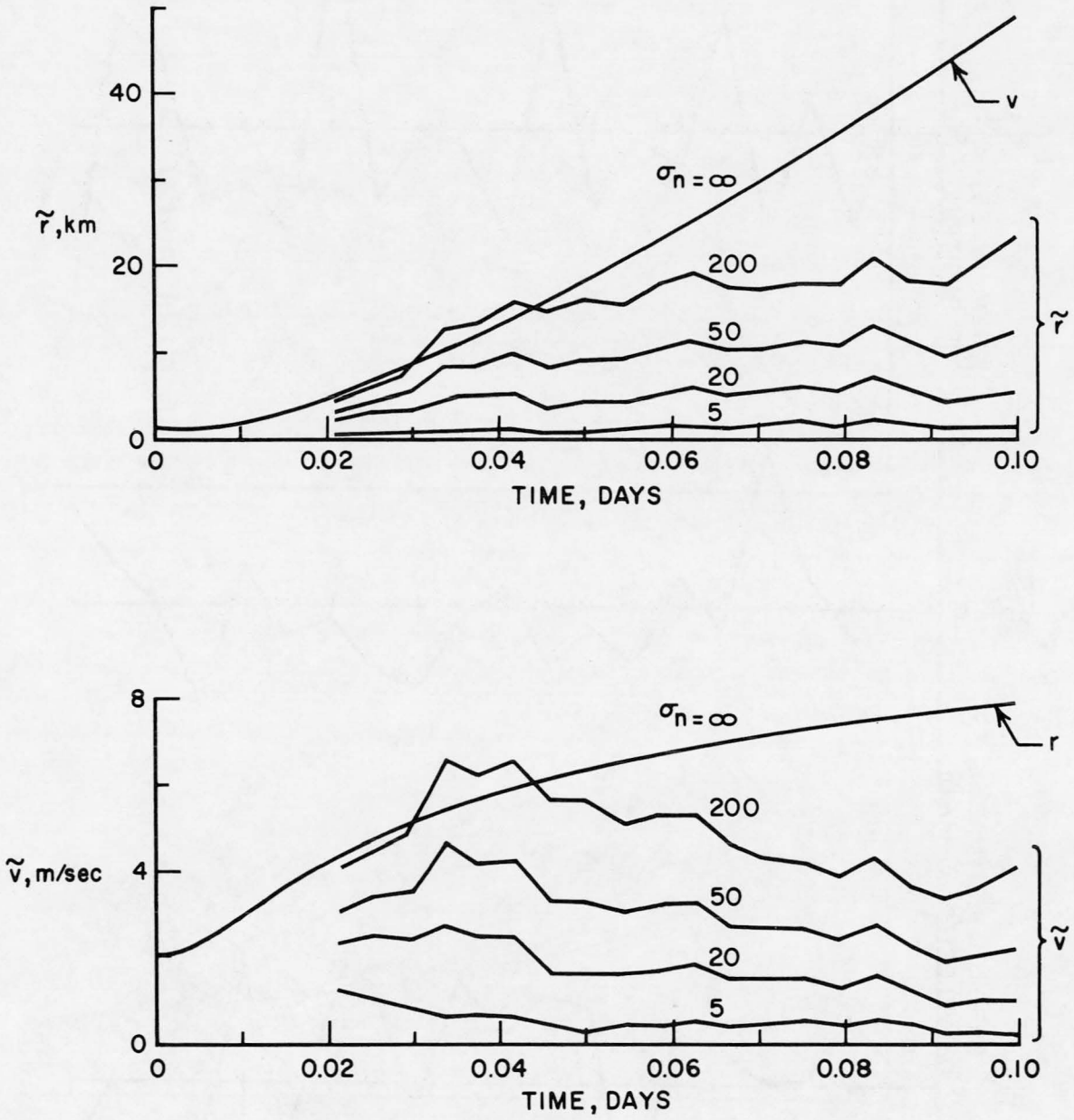


Figure 4.- Time histories of the estimation errors for various magnitudes of observation error.

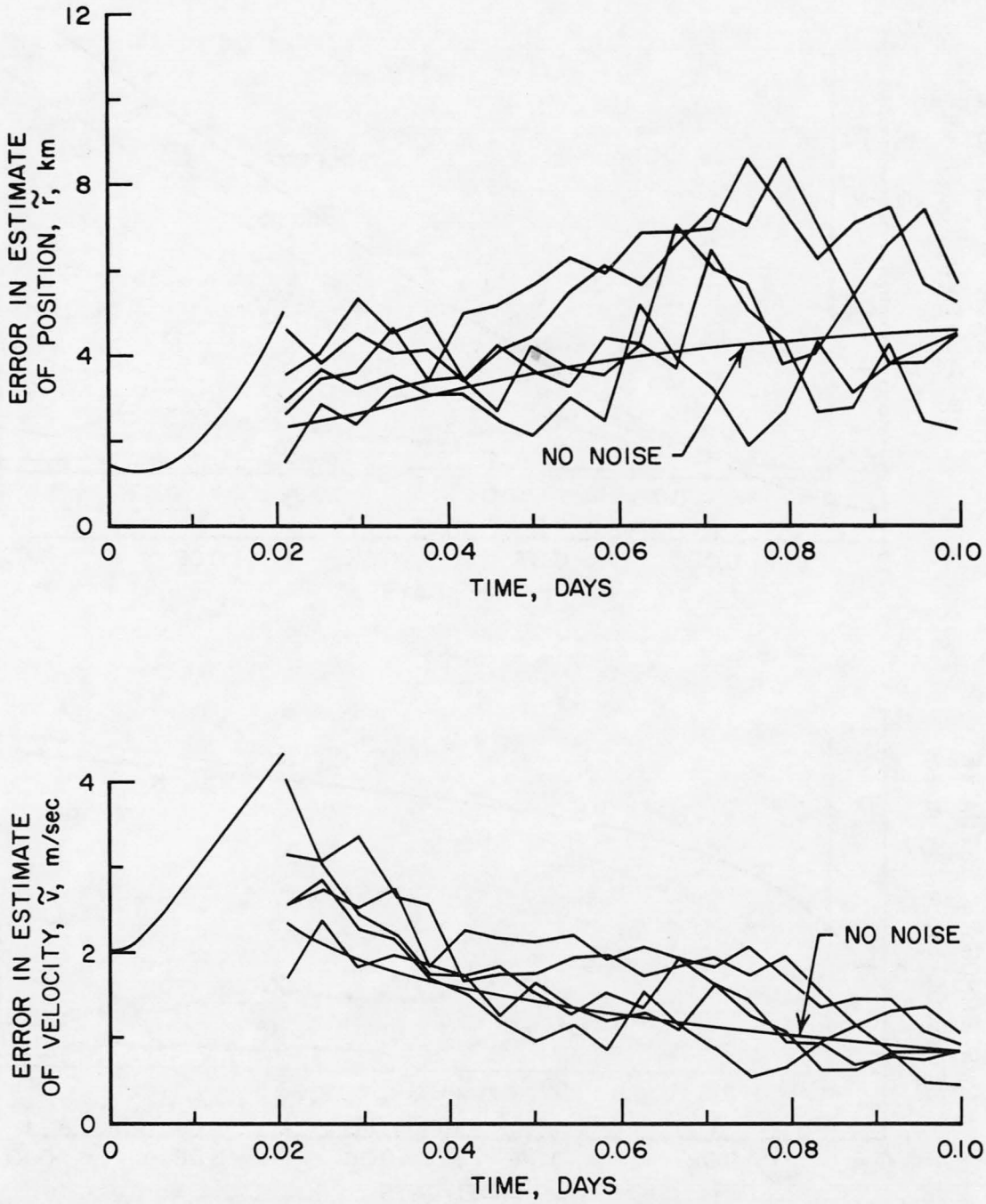


Figure 5.- Time histories of the estimation errors for various observation error samples; σ_n = 20 sec arc.

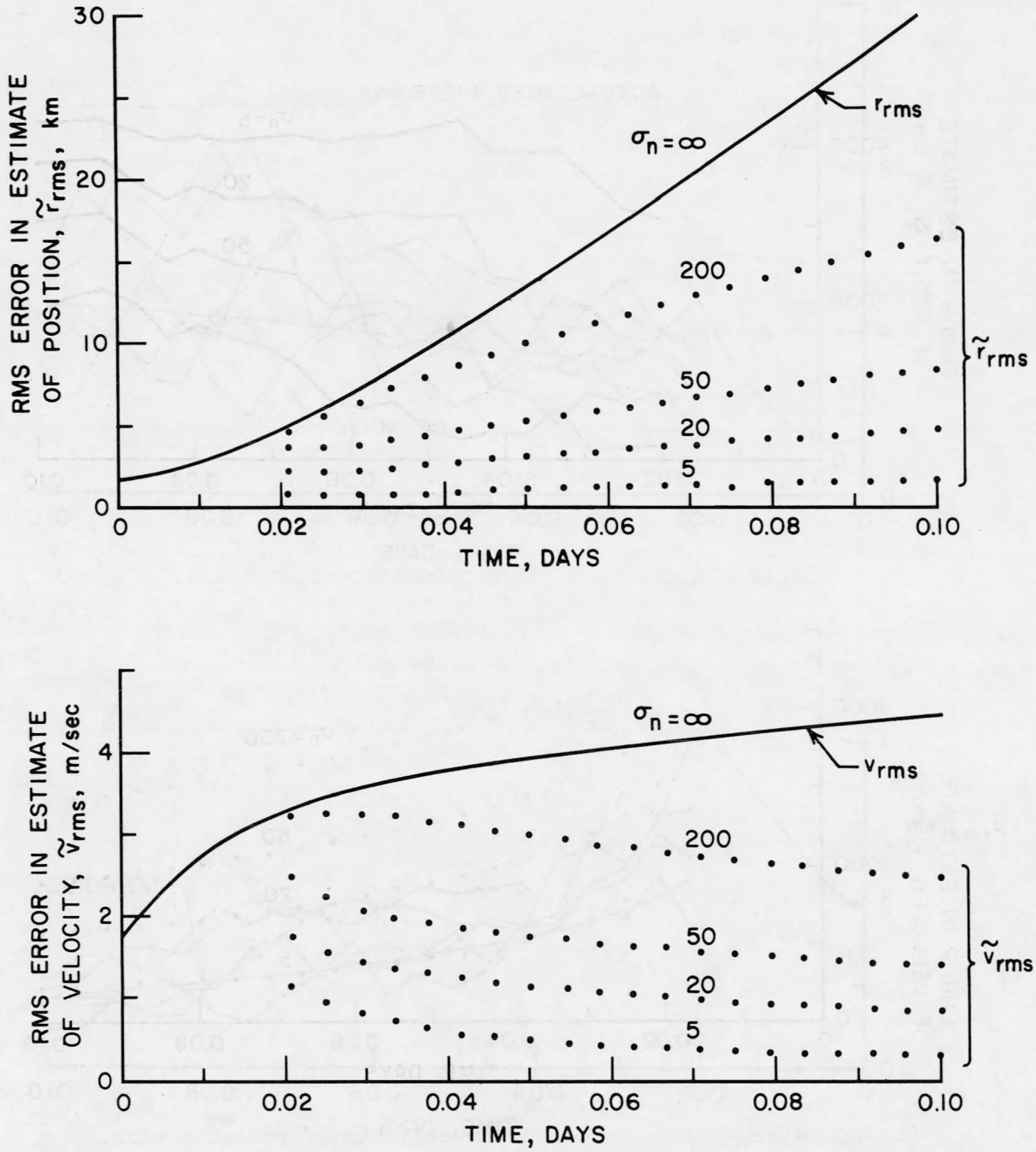
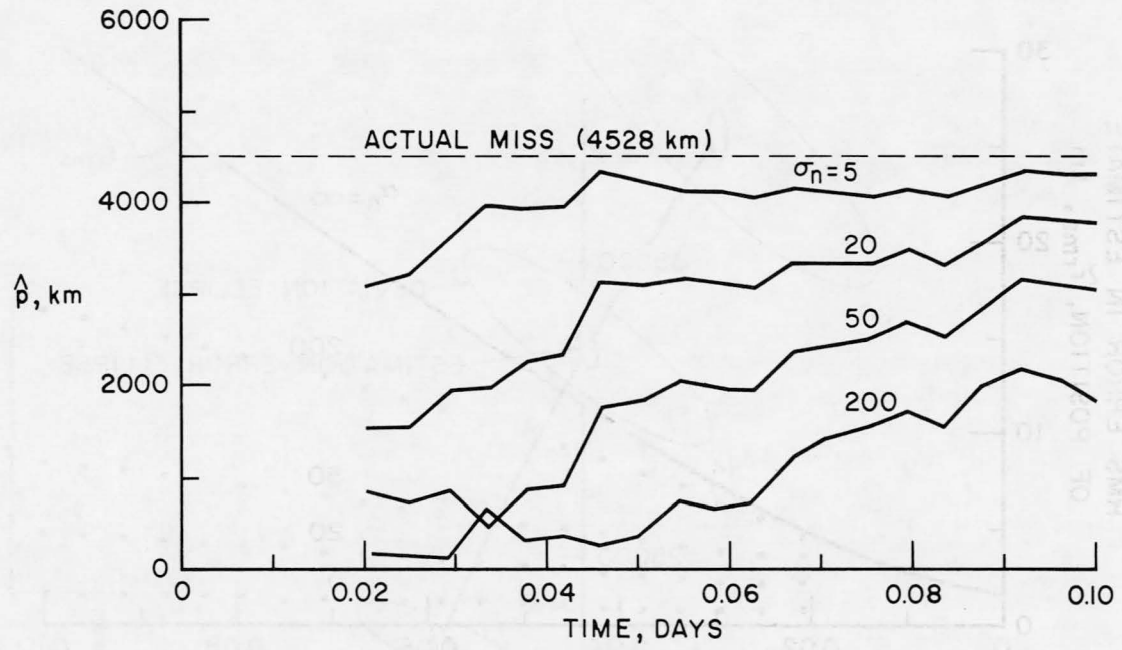
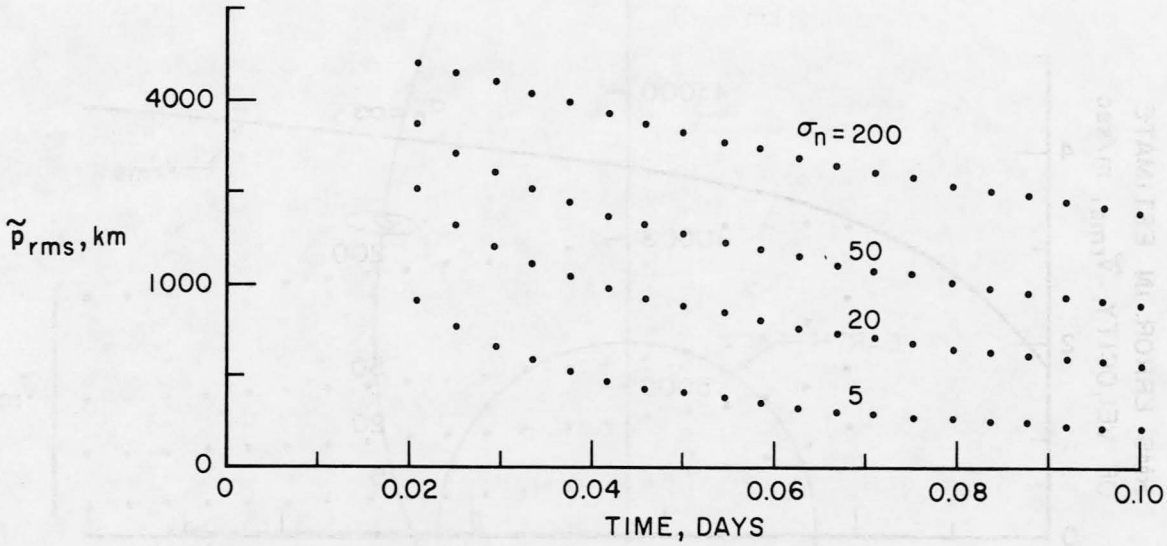


Figure 6.- Root-mean-square estimation errors for various standard deviations of observation errors.



(a) Time history of the estimated perilune miss.



(b) Root-mean-square error in the estimate of perilune miss.

Figure 7.- Performance of the system in estimating the miss at nominal perilune for various magnitudes of observation errors.

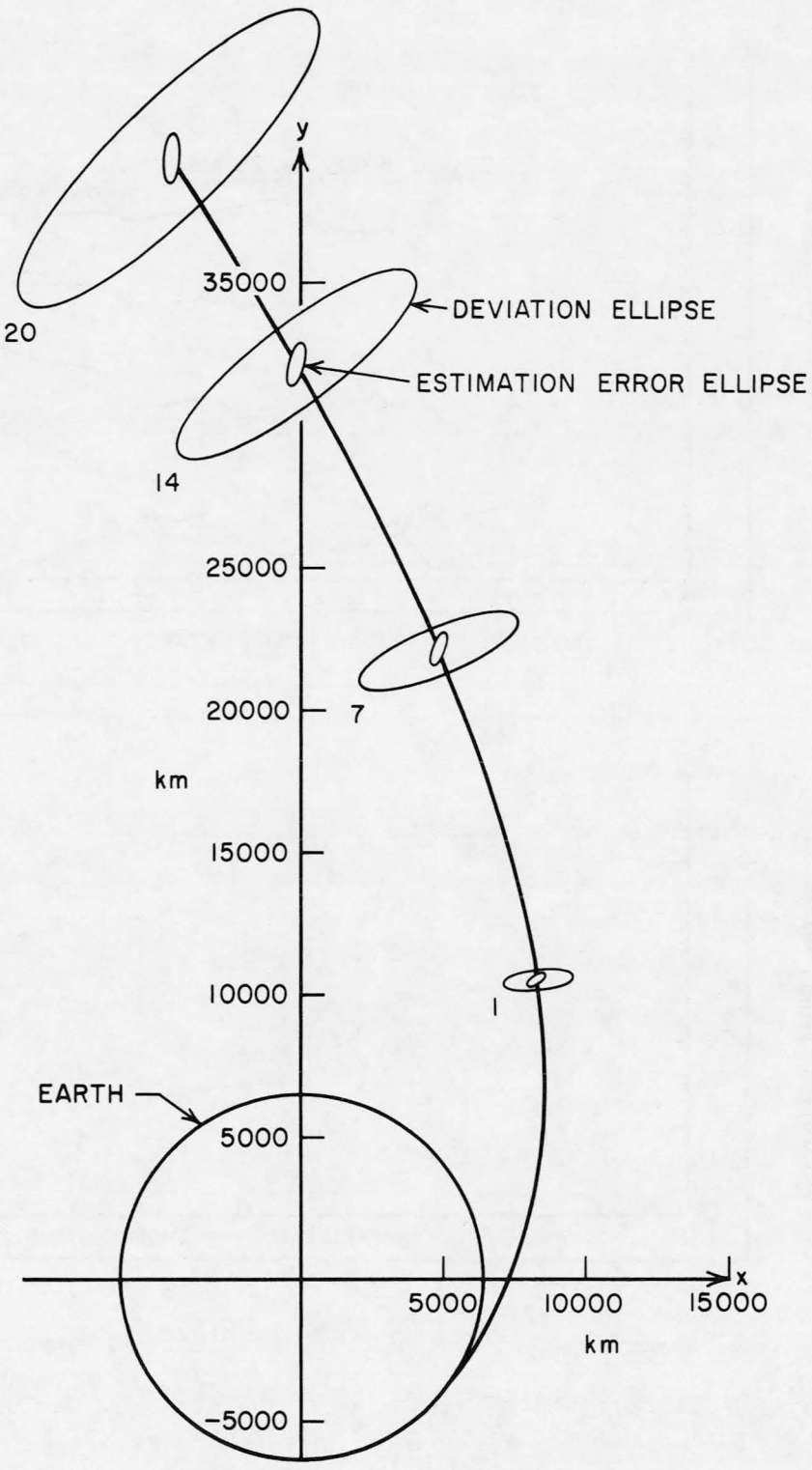


Figure 8.- Deviation and estimation error ellipses in the equatorial plane; σ_n = 20 sec arc.

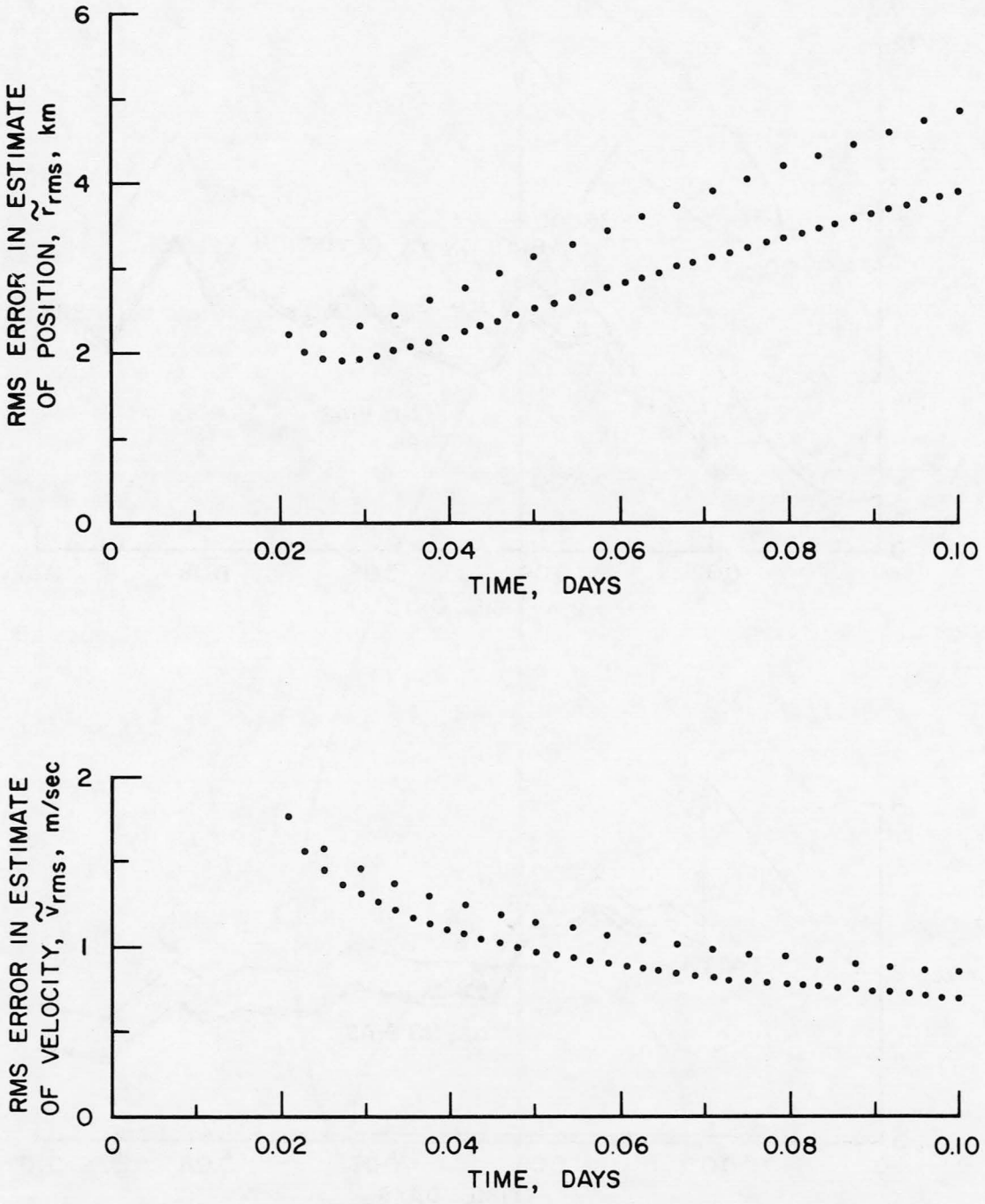


Figure 9.- Root-mean-square estimation errors for two different observation rates.

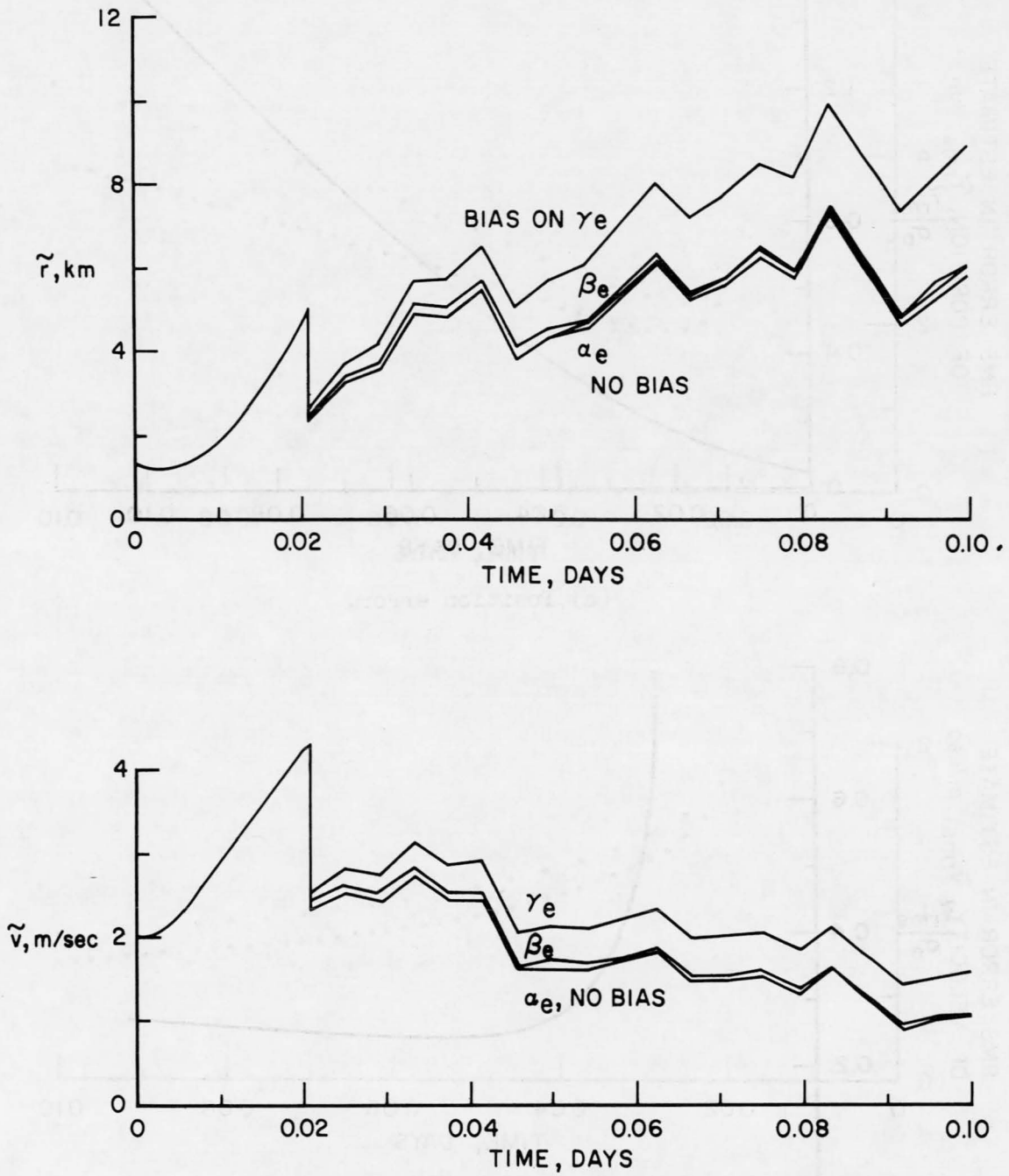
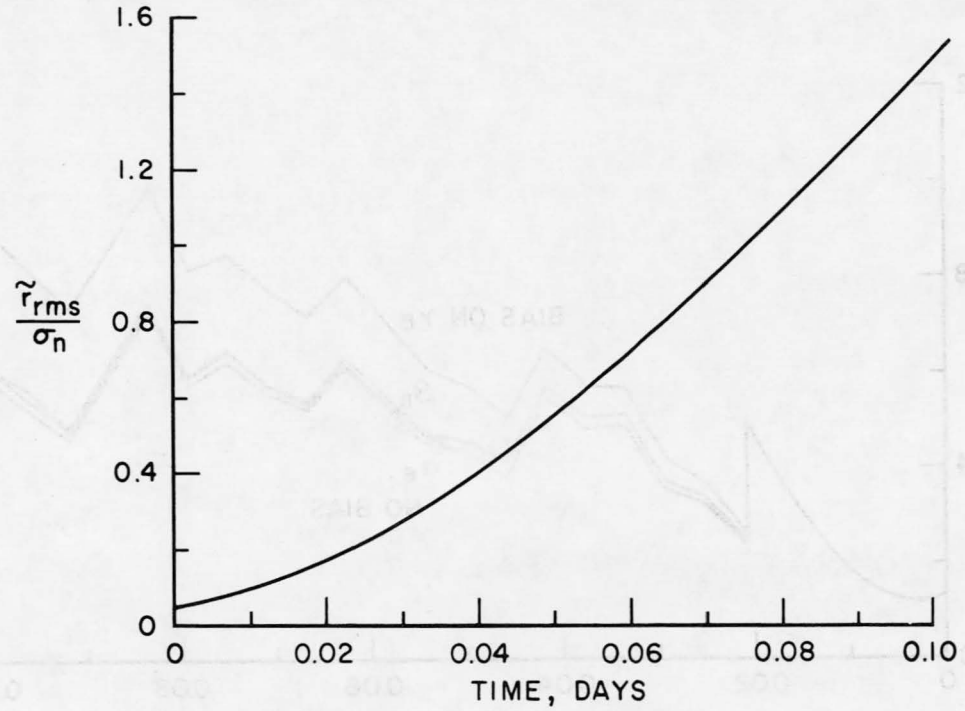
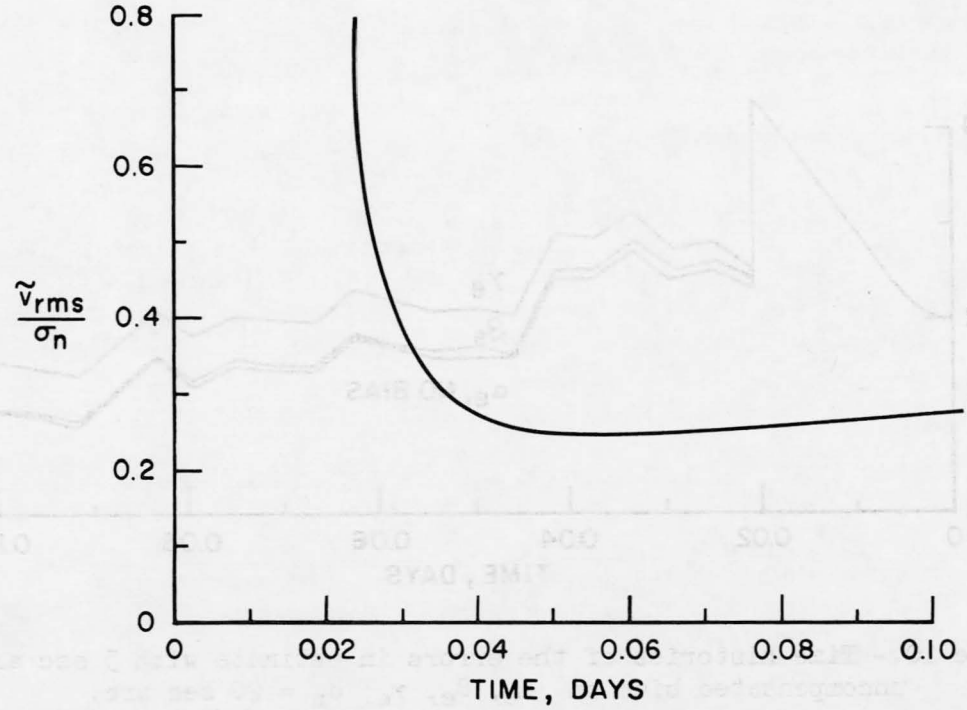


Figure 10.- Time histories of the errors in estimate with 5 sec arc uncompensated bias on α_e, β_e, γ_e; σ_n = 20 sec arc.



(a) Position error.



(b) Velocity error.

Figure 11.- Root-mean-square estimation errors for a minimum data method of trajectory estimation.

NASA-Langley, 1962 A-550