U. S. AIR FORCE

PROJECT RAND

RESEARCH MEMORANDUM

FIRST-ORDER ERROR PROPAGATION IN A STAGEWISE SMOOTHING PROCEDURE FOR SATELLITE OBSERVATIONS

Peter Swerling

RM-2329

June 15, 1959

Assigned to ______________________

The RAND Corporation
1700 MAIN ST. • SANTA MONICA • CALIFORNIA

SUMMARY

Suppose that there are a number of observed quantities, the values of which would, in the absence of observation error, be determined by the observation times plus a finite number of quantities called elements. The observations are to be smoothed in such a way as to obtain estimates of the elements. The immediate motivation for this arises in the study of estimations of satellite orbits, in which case the elements might be the position and velocity components at a particular time instant, or alternatively the semi-major axis, eccentricity, inclination, etc.; the applicability of the results is, however, not limited just to satellite tracking.

The smoothing methods considered are variations of the method, often used in astronomy, of minimizing a quadratic form in the residuals; in some contexts, these smoothing methods may be considered to be maximum likelihood estimation methods.

Among the methods considered are certain stagewise methods in which new observations are smoothed with "current" element estimates to give improved estimates. These stagewise methods are advantageous in some respects in that they require the storage, at any time, of only a fixed number of quantities depending on the number of elements and not on the total number of observations to be processed.

A complete first-order error analysis is given for all the smoothing methods considered; that is, the dependence of the errors in the element estimates on the observation errors is established, for sufficiently small values of the latter. It is shown that the stagewise methods may be defined in such a way as to yield estimates having the same first-order dependence

on observation errors as the estimates resulting from simultaneous smoothing of all available observations.

The aforementioned results are used to establish certain properties of the error statistics of element estimates as functions of the observation error statistics.

The elements are at first considered to be constants; later, they are considered to be slowly varying functions of time.

The present report deals mainly with the theory of first-order error propagation in the smoothing methods considered. Later reports are planned in which the results will be applied to specific cases--particularly, to the evaluation of various satellite tracking networks.

CONTENTS

I. INTRODUCTION

The subject to be discussed is that of smoothing of observational data in cases where the observations, in the absence of observational error, would all be determined by the time of observation plus a finite number of parameters, called elements. The objective is to estimate the elements.

The immediate motivation for this arises in studies of estimation of earth-satellite orbits from observational data. In this case, if, for example, the force field were known exactly, one could regard the elements as the position and velocity components at a particular instant t_o ; if the field were a central inverse square field, the elements could alternatively be the six elements of a Keplerian elliptic orbit. (It should be mentioned, however, that satellite orbit prediction is only one of the possible applications of the results.)

In satellite tracking and prediction, it is desired to produce ephemerides--i.e., predictions of the future position as a function of time--as well as to make various other types of decisions and predictions. As new observations become available, one can improve the accuracy of these predictions. Ideally, one would like to store all previous observations of the object, and combine these in some optimum way to yield the desired predictions or decisions. The optimum method for processing the available data would be based on analysis of the error statistics for the individual observations, and on the functional dependence of future predicted quantities on the previous observations.

In satellite tracking one is dealing with a situation in which there may be a large number of observations of varying degrees of accuracy, as well as large numbers of tracked objects. Also, methods of orbit prediction (even in the absence of observation errors) are subject to various sources of error, such as

(a) uncertainty in the forces acting on the body (earth's gravitational and magnetic fields, atmospheric resistance);

(b) cumulative errors in solving the equations of motion.

Two features would be desirable in a tracking and prediction method:

(1) The data processing load per tracked object should not exceed a certain maximum, regardless of how many observations are available to be processed. On the other hand, the prediction method should not throw available observations away.

(2) The method should be adaptable to situations in which the underlying prediction functions are subject to the above-mentioned uncertainties.

The stagewise procedures described below are motivated by these considerations.

The particular methods of data smoothing to be discussed are variations of the classical method of minimizing a quadratic form in the residuals (in practice this is usually a weighted sum of squares of residuals). After defining this method, the first-order dependence of the errors in the resulting element estimates on the observation errors is established.

We then go on to describe a stagewise procedure for processing the observational data, in which the element estimates at each stage are smoothed in a particular way with some additional observations. This is in essence a type of differential correction procedure. It is shown that the errors in the resulting element estimates, after stagewise smoothing of a given set of observations, have the same first-order dependence on the observation errors as would the errors in the estimates obtained by simultaneous processing of the same total set of observations.

Some statistical properties of errors in the element estimates are derived for the case in which the observation errors are regarded as statistical variables and in which the matrix of the quadratic form to be minimized is the inverse covariance matrix of the observation errors. (For Gaussian error statistics, this would result in a maximum-likelihood method of estimation.)

The elements are at first regarded as constants, and later the treatment is extended to the case in which the elements are regarded as time-dependent.

We suppose there are N observed quantities F^{μ}, $\mu = 1, \cdots, N$, each F^{μ} being a real scalar. Also we assume that n real constants x_i, $i = 1, \cdots, n$ exist such that in the absence of observation error, all N observed quantities would satisfy the relations

$$F^{\mu} = f^{\mu} (x_1, \cdots, x_n, t_{\mu}) . \tag{1}$$

The vector $x = (x_1, \cdots, x_n)$ will be called the 'element vector,' and its components the 'elements;' t_{μ} is the time at which the μ^{th} observation is taken.

When observation errors are present, the F^{μ} are given by

$$F^{\mu} = f^{\mu} (x_1, \cdots, x_n, t_{\mu}) + \mathcal{E}^{\mu} . \tag{2}$$

For purposes of illustration, consider the observation of a satellite following a Keplerian orbit. The elements might then be taken to be the eccentricity, semi-major axis, inclination, etc; the quantities F^{μ} might be observations of such things as range, azimuth, elevation, or range rate from particular observation sites; the f^μ would be the functions describing the dependence of these observed quantities on the elements and time; and the $\mathcal{E}^\mu$ would be the observation errors.

Many of the formulas below will involve the functions f^μ and their partial derivatives $\frac{\partial f^\mu}{\partial x_i}$. Practical application of these formulas would be possible both for cases where analytic expressions are known for the f^μ and their partial derivatives and for cases where these functions must be evaluated by numerical integration, as well as for cases where some of the functions have known analytic expressions and others must be determined by numerical methods.

Henceforth it will be supposed that $N \geq n$. The problem to be considered is the estimation of the elements by means of smoothing, in some sense, of the observations.

A classical smoothing method is as follows:
writing $P = (P_1, \cdots, P_n)$ for an estimate of the element vector, and writing $f(x,t)$ for $f(x_1, \cdots, x_n, t)$, $f(P,t)$ for $f(P_1, \cdots, P_n, t)$, and so forth, the method consists of minimizing with respect to $P_1, \cdots, P_n$ the quadratic form

$$Q = \sum_{\mu,\nu = 1}^{N} \eta_{\mu\nu} \left[F^\mu - f^\mu (P,t_\mu)\right] \left[F^\nu - f^\nu (P,t_\nu)\right] \qquad (3)$$

where $(\eta_{\mu\nu})$ is a symmetric, positive definite matrix. Thus, the method consists in minimizing a positive definite quadratic form in the residuals.

Differentiating Q with respect to P_i and setting the results equal to zero, we find that the minimizing estimates P_i must satisfy

$$\sum_{\mu,\nu=1}^{N} \frac{\partial f^{\nu}}{\partial x_i}(P,t_\nu)\left[F^{\mu} - f^{\mu}(P,t_\mu)\right]\eta_{\mu\nu} = 0 \qquad (i = 1, \cdots, n) \; . \tag{4}$$

It is clear that if $\varepsilon^{\mu} = 0$, all μ, then $P = x$ is a solution of (4). The first question to be investigated is that of the first-order propagation of errors--i.e., the first-order dependence of $P - x$ on the errors ε^{μ}. It will be assumed that the functions f^{μ} are sufficiently well-behaved for the following operations to be valid. We may write

$$f^{\mu}(P,t_\mu) = f^{\mu}(x,t_\mu) + \sum_{j=1}^{n} \frac{\partial f^{\mu}}{\partial x_j}(x,t_\mu)(P_j - x_j) + \cdots \tag{5}$$

$$\frac{\partial f^{\mu}}{\partial x_i}(P,t_\mu) = \frac{\partial f^{\mu}}{\partial x_i}(x,t_\mu) + \sum_{j=1}^{n} \frac{\partial^2 f^{\mu}}{\partial x_i \, \partial x_j}(x,t_\mu)(P_j - x_j) + \cdots \tag{6}$$

Neglecting all terms of higher order in $P - x$, and substituting (5) and (6) into (4), we obtain

$$\sum_{\mu,\nu=1}^{N} \eta_{\mu\nu}\left[\frac{\partial f^{\nu}}{\partial x_i}(x,t_\nu) + \sum_{j=1}^{n} \frac{\partial^2 f^{\nu}}{\partial x_i \, \partial x_j}(x,t_\nu)(P_j - x_j)\right] \text{(cont. on next page)} \tag{7}$$

(7) Continued...

$$X \left[F^{\mu} - f^{\mu}(x, t_{\mu}) - \sum_{j=1}^{n} \frac{\partial f^{\mu}}{\partial x_j}(x, t_{\mu})(P_j - x_j) \right] = 0$$

It is also clear that for sufficiently small $\mathcal{E}^{\mu}$ and $P_j - x_j$, the term involving second derivatives of f^{ν} may be neglected. The result may conveniently be expressed as follows: let

$$a_i^{\mu}(x, t_{\mu}) = \frac{\partial f^{\mu}}{\partial x_i}(x, t_{\mu}) \tag{8}$$

$$\rho_i(x) = \sum_{\mu, \nu = 1}^{N} \eta_{\mu\nu}\, a_i^{\nu}(x, t_{\nu}) \left[F^{\mu} - f^{\mu}(x, t_{\mu}) \right] \tag{9}$$

$$\rho(x) = \left\{ \rho_1(x), \cdots, \rho_n(x) \right\} \tag{10}$$

$$B_{ij}(x) = \sum_{\mu, \nu = 1}^{N} \eta_{\mu\nu}\, a_i^{\nu}(x, t_{\nu})\, a_j^{\mu}(x, t_{\mu}) \tag{11}$$

$$B(x) = \left\{ B_{ij}(x) \right\} \tag{12}$$

Then, assuming $B(x)$ to be non-singular,

$$P - x = \left[B(x) \right]^{-1} \rho(x) \tag{13}$$

Eq. (13) expresses the first-order dependence of $P - x$ on the errors ε^{μ}. This can be expressed equivalently:

$$P_i - x_i = \sum_{\mu=1}^{N} \Gamma_i^{\mu}(x) \left[F^{\mu} - f^{\mu}(x, t_{\mu})\right] \tag{14}$$

$$\Gamma_i^{\mu}(x) = \sum_{j=1}^{n} \sum_{\nu=1}^{N} \left[B(x)\right]_{ij}^{-1} \eta_{\mu\nu}\, a_j^{\nu}(x, t_{\nu}) \quad . \tag{15}$$

A special case of this is as follows: suppose one has already an estimated element vector $p = (p_1, \cdots, p_n)$ together with K new observations. One can form a new estimate vector P by smoothing the original estimate vector p with the new observations in the following manner: P is determined by minimizing, with respect to $P_1, \cdots, P_n$, the quadratic form

$$Q = \sum_{\mu,\nu=1}^{n} \eta_{\mu\nu} (p_{\mu} - P_{\mu})(p_{\nu} - P_{\nu}) \tag{16}$$

$$+ \sum_{\mu,\nu=n+1}^{n+K} \eta_{\mu\nu} \left[F^{\mu} - f^{\mu}(P, t_{\mu})\right] \left[F^{\nu} - f^{\nu}(P, t_{\nu})\right] \quad .$$

In this case,

$$f^{\mu}(x,t_{\mu}) = x_{\mu} \qquad , \mu = 1, \cdots, n \tag{17}$$

$$a_i^{\mu}(x,t_{\mu}) = \delta_{i\mu} \qquad i = 1, \cdots, n \quad \mu = 1, \cdots, n$$

$$F^{\mu} = p_{\mu} \qquad \mu = 1, \cdots, n \quad .$$

Also

$$\rho_i(x) = \sum_{j=1}^{n} \eta_{ij}(p_j - x_j) + \sum_{\mu,\nu=n+1}^{n+K} \eta_{\mu\nu}\left[F^{\mu} - f^{\mu}(x,t_{\mu})\right] a_i^{\nu}(x,t_{\nu}) \qquad (i = 1, \cdots, n) \tag{18}$$

and

$$B_{ij}(x) = \eta_{ij} + \sum_{\mu,\nu=n+1}^{n+K} \eta_{\mu\nu}\, a_i^{\nu}(x,t_{\nu})\, a_j^{\mu}(x,t_{\mu}) \qquad (i,j = 1, \cdots, n) \quad . \tag{19}$$

If we also define $r^{\mu}(p,t_{\mu})$ for $\mu = n + 1, \cdots, n + K$ by

$$r^{\mu}(p,t_{\mu}) = F^{\mu} - f^{\mu}(p,t_{\mu}) \qquad \mu = n + 1, \cdots, n + K \tag{20}$$

then $\rho_i(x)$ becomes, to first order,

$$\rho_i(x) = \sum_{j=1}^{n} B_{ij}(x)(p_j - x_j) \tag{21}$$

$$+ \sum_{\mu,\nu=n+1}^{n+K} \eta_{\mu\nu}\, a_i^{\nu}(x,t_\nu)\, r^{\mu}(p,t_\mu) \quad .$$

Consequently, (13) reduces to

$$P - x = p - x + \left[B(x)\right]^{-1} \rho^*(x) \tag{22}$$

where

$$\rho_i^*(x) = \sum_{\mu,\nu=n+1}^{n+K} \eta_{\mu\nu}\, a_i^{\nu}(x,t_\nu)\, r^{\mu}(p,t_\mu) \quad . \tag{23}$$

Eq. (22) may be used as the basis for a first-order differential correction to p, given the additional observations F^{μ}, $\mu = n + 1, \cdots, n + K$. First rewrite (22) as

$$P - p = \left[B(x)\right]^{-1} \rho^*(x) \quad . \tag{24}$$

Then, to first order, one can also write

$$P - p = \left[B(p)\right]^{-1} \rho^*(p) \tag{25}$$

Since all the quantities on the right of (25) are known, (25) gives the required first-order correction to p.

Now define sequences of matrices $B^{(s)}$ and $B^{(s)}(x)$ as follows:*

$$B_{ij}^{(s)} = B_{ij}^{(s-1)} + \sum_{\mu,\nu = M_{s-1}+1}^{M_s} \eta_{\mu\nu}^{(s)} a_i^{\nu}(P^{(s)},t_{\nu})\, a_j^{\mu}(P^{(s)},t_{\mu}) \tag{29}$$

$$B_{ij}^{(o)} = 0$$

$$B_{ij}^{(s)}(x) = B_{ij}^{(s-1)}(x) + \sum_{\mu,\nu = M_{s-1}+1}^{M_s} \eta_{\mu\nu}^{(s)} a_i^{\nu}(x,t_{\nu})\, a_j^{\mu}(x,t_{\mu})$$

$$B_{ij}^{(o)}(x) = 0 \quad .$$

For $s > 1$, the s^{th} element estimates $\left\{P_i^{(s)}\right\}$ are obtained by minimizing, with respect to $P^{(s)}$, the quadratic form

$$Q^{(s)} = \sum_{i,j=1}^{n} B_{ij}^{(s-1)} \left[P_i^{(s-1)} - P_i^{(s)}\right] \left[P_j^{(s-1)} - P_j^{(s)}\right] \tag{30}$$

$$+ \sum_{\mu,\nu = M_{s-1}+1}^{M_s} \eta_{\mu\nu}^{(s)} \left[F^{\mu} - f^{\mu}(P^{(s)},t_{\mu})\right] \left[F^{\nu} - f^{\nu}(P^{(s)},t_{\nu})\right] \quad .$$

Let us also define $\rho_i^{(s)}(x)$, $s = 1,2, \cdots, S$, by

*Alternative definitions for the matrices $B^{(s)}$ are possible; see the remark at the end of this Section.

$$\rho_i^{(s)}(x) = \sum_{j=1}^{n} B_{ij}^{(s-1)}(x) \left[P_j^{(s-1)} - x_j\right] \tag{31}$$

$$+ \sum_{\substack{\mu,\nu = \\ M_{s-1}+1}}^{M_s} \eta_{\mu\nu}^{(s)} a_i^{\nu}(x,t_\nu) \left[F^\mu - f^\mu(x,t_\mu)\right] .$$

Then, using (13), (18), and (19), it can be shown that to first order

$$P^{(s)} - x = \left[B^{(s)}(x)\right]^{-1} \rho^{(s)}(x) . \tag{32}$$

It is also not hard to show by an inductive argument that

$$P^{(S)} - x = \left[B^{(S)}(x)\right]^{-1} \rho^{(S)}(x) = \left[B(x)\right]^{-1} \rho(x) \tag{33}$$

where $B(x)$ and $\rho(x)$ are defined as in (9), (10), (11), and (12). This is proved by showing that $B^{(S)}(x) = B(x)$, and $\rho^{(S)}(x) = \rho(x)$.

Thus, the first-order dependence of the errors in the estimates $\left\{P_i^{(S)}\right\}$ is the same as that for the $\left\{P_i\right\}$ obtained by processing all N observations at once by minimizing Q as defined by (3). Another way of stating this is to say that, to first order, the estimate obtained by processing the N observations by the stagewise procedure just described is the same as that obtained by minimizing Q as defined by (3). (However, the range of magnitudes of ϵ^μ for which the first-order expressions give good approximations to the estimation errors is not necessarily the same for the stagewise method as for the non-stagewise method.)

A stagewise smoothing procedure may be advantageous in certain situations. For example, suppose that observations are coming in at some average rate; in the stagewise procedure, it is not necessary to store all previous observations and $N \times N$ matrices $(\eta_{\mu\nu})$ with N increasing. It is at any time necessary to store only the 'current' element estimates $\left\{P_i^{(s)}\right\}$ and the matrix $B^{(s)}$, that is, $\frac{n}{2}(n + 3)$ quantities (taking into account the symmetry of $B^{(s)}$).

It can be seen that the matrices $B^{(s)}$ play the role of estimates of the matrices $B^{(s)}(x)$. Thus, the particular method of defining the sequence $\left\{B^{(s)}\right\}$ above is not the only one possible. For example, one could define

$$\tilde{B}_{ij}^{(s)} = B_{ij}^{(s)}, \quad s = 1$$

$$= \tilde{B}_{ij}^{(s-1)} + \sum_{\mu,\nu = M_{s-1}+1}^{M_s} \eta_{\mu\nu}^{(s)} \; a_i^{\nu}(P^{(s-1)}, t_\nu) \; a_j^{\mu}(P^{(s-1)}, t_\mu), \quad s > 1 \qquad (29')$$

and define the stagewise process using the matrices $\tilde{B}^{(s)}$ instead of $B^{(s)}$. Comparing with (29), we see that the main difference is that $\tilde{B}^{(s)}$ can be computed, for $s > 1$, before one computes $P^{(s)}$.

This lends itself conveniently to first-order determination, for $s > 1$, of $P^{(s)} - P^{(s-1)}$ by means of Eqs. (20), (23), and (24). One would write $p = P^{(s-1)}$; $P = P^{(s)}$; and $P^{(s)} - P^{(s-1)} = \left[\tilde{B}^{(s)}\right]^{-1} \rho^{*(s)}(P^{(s-1)})$. (The vector $\rho^{*(s)}$ would be defined by obvious modifications of (20) and (23).)

III. APPLICATION WHEN THE ELEMENTS ARE FUNCTIONS OF TIME

Suppose the elements $x_1, \cdots, x_n$ are functions of time: $x_i = x_i(t)$. Also suppose that error-free observations are given by

$$F^{\mu} = g^{\mu}\left[x(t_{\mu}),\ t_{\mu}\right] \quad . \tag{34}$$

Suppose the manner in which $x(t)$ depends on its values at some t_o is known:

$$x_i(t) = \mathscr{F}_i\left[x(t_o),\ t_o,\ t\right] \quad . \tag{35}$$

Then we may obtain the case considered in Sections I and II by defining the elements x_i in the formulas of those Sections to be $x_i = x_i(t_o)$ and by defining

$$f^{\mu}\left[x,t_{\mu}\right] = g^{\mu}\left\{\mathscr{F}_1\left[x,t_o,t_{\mu}\right],\ \cdots,\ \mathscr{F}_n\left[x,t_o,t_{\mu}\right],t_{\mu}\right\} \tag{36}$$

(where $x = x(t_o)$).

IV. MODIFICATION WHEN THE FUNCTIONS f^{μ} ARE IMPERFECTLY KNOWN

Suppose the observations are given by

$$F^{\mu} = h^{\mu}\left[x, t_{\mu}\right] + \varepsilon^{\mu} \tag{37}$$

(where ε^{μ} are observation errors),

but that the estimated elements $\left\{P_i\right\}$ are obtained by minimizing Q as defined by (3), with the functions f^{μ} (which differ from h^{μ}).

So long as $F^{\mu} - f^{\mu}$ are sufficiently small, Eqs. (9)-(13) still describe the dependence of $P - x$ on $F^{\mu} - f^{\mu}$. This dependence was expressed

$$P_i - x_i = \sum_{\mu=1}^{N} \Gamma_i^{\mu}(x)\left[F^{\mu} - f^{\mu}(x, t_{\mu})\right] \tag{14}$$

where:

$$\Gamma_i^{\mu}(x) = \sum_{j=1}^{n} \sum_{\nu=1}^{N} \left[B(x)\right]_{ij}^{-1} \eta_{\mu\nu}\, a_j^{\nu}(x, t_{\nu}) \; . \tag{15}$$

The observation errors ε^{μ} are now given by

$$\varepsilon^{\mu} = F^{\mu} - h^{\mu}(x, t_{\mu}). \tag{38}$$

Therefore

$$F^{\mu} - f^{\mu}(x, t_{\mu}) = \varepsilon^{\mu} + h^{\mu}(x, t_{\mu}) - f^{\mu}(x, t_{\mu}) \tag{39}$$

Thus, (14) becomes

$$P_i - x_i = \sum_{\mu=1}^{N} \Gamma_i^{\mu}(x)\,\mathcal{E}^{\mu} + \sum_{\mu=1}^{N} \Gamma_i^{\mu}(x)\left[h^{\mu}(x,t_{\mu}) - f^{\mu}(x,t_{\mu})\right]. \quad (40)$$

V. STATISTICS OF PROPAGATED ERRORS

Equation (13) may be used in an obvious manner to determine the means and covariance matrix of $\left\{P_i - x_i\right\}$ as functions of the means and covariance matrix of $\left\{\varepsilon^{\mu}\right\}$.

We shall deal here with a special case, namely, one in which the ensemble means and covariance matrix of $\left\{\varepsilon^{\mu}\right\}$ are known and in which the matrix $\left\{\eta_{\mu\nu}\right\}$ of the quadratic form Q has a special relation to the covariance matrix of $\left\{\varepsilon^{\mu}\right\}$.

Since the ensemble means of $\left\{\varepsilon^{\mu}\right\}$ are assumed known, it is no loss of generality to assume they are zero. In this case the covariance matrix is (denoting ensemble means by E( ))

$$\phi_{\mu\nu} = E(\varepsilon^{\mu}\varepsilon^{\nu}) . \tag{41}$$

It will now be assumed that the matrix $(\eta_{\mu\nu})$ in (3) is

$$(\eta_{\mu\nu}) = (\phi_{\mu\nu})^{-1} \quad \text{(matrix inverse)} \tag{42}$$

If $\left\{\varepsilon^{\mu}\right\}$ were to have a Gaussian probability distribution (and if $f^{\mu} \equiv h^{\mu}$, all μ) then the resulting method of obtaining P would be the maximum likelihood method.

The covariance matrix of $\left\{P_i - x_i\right\}$ assumes a particularly simple form in this case. For generality, we will deal with the case described in Section IV, in which f^{μ}, the functions used in the quadratic form Q, may differ from h^{μ}.

The means of $\{P_i - x_i\}$ are

$$E(P_i - x_i) = \sum_{\mu=1}^{N} \Gamma_i^{\mu}(x) \left[h^{\mu}(x,t_{\mu}) - f^{\mu}(x,t_{\mu})\right] . \tag{43}$$

The covariance matrix of $\{P_i - x_i\}$ is readily established to be

$$E\left\{P_i - x_i - E(P_i - x_i)\right\}\left\{P_j - x_j - E(P_j - x_j)\right\} = \left[B(x)\right]^{-1}_{ij} . \tag{44}$$

This holds whether the observations are processed all together or by a stagewise procedure as described in Section II. In the latter case, of course, it must be assumed that the covariance matrix of $\{\mathcal{E}^{\mu}\}$ can be broken up into a diagonal array of covariance matrices corresponding to the different stages. It is, in fact, quite easy to establish that $B^{(s)}(x)$ is the inverse covariance matrix of $\left\{P_i^{(s)} - x_i\right\}$:

$$E\left\{P_i^{(s)} - x_i - E(P_i^{(s)} - x_i)\right\}\left\{P_j^{(s)} - x_j - E(P_j^{(s)} - x_j)\right\} = \left[B^{(s)}(x)\right]^{-1}_{ij} \tag{45}$$

This throws some further light on the stagewise method of Section II. The matrices $B^{(s-1)}$ occurring in the quadratic forms $Q^{(s)}$ are seen to be estimates of the inverse covariance matrices $B^{(s-1)}(x)$ of the element estimates $P^{(s-1)}$. Thus, if the error statistics were Gaussian, this procedure would consist at each stage of a maximum likelihood smoothing of the previous element estimates with the new observations.

The above formulas may be used to determine the rate at which the covariance matrix of the errors $P_i^{(s)} - x_i$ decreases as additional observations are processed. In fact, this information is contained in (44) and (45).

As a special case, consider the case where all observation errors are mutually uncorrelated. (If this is not true originally, it can be made true by means of linear transformations.) In this case, the matrices ϕ and η are diagonal.

If we now regard $P^{(s)}$ as the element estimate vector resulting from the processing of the first s observations, we have the inverse covariance matrix for $\left\{P_i^{(s)} - x_i\right\}$ given by

$$B_{ij}^{(s)}(x) = \sum_{\mu=1}^{s} \eta_{\mu\mu}\, a_i^{\mu}(x,t_{\mu})\, a_j^{\mu}(x,t_{\mu}) \quad . \tag{46}$$

(This holds whether the s observations were processed all together or stagewise.)

It is also quite easy to verify that, for $s \geqq n$,

$$\left[B^{(s)}(x)\right]_{ij}^{-1} = \left[B^{(s-1)}(x)\right]_{ij}^{-1} - d_i^{(s)}(x)\, d_j^{(s)}(x) \tag{47}$$

where

$$d_i^{(s)}(x) = \frac{\sqrt{\eta_{ss}} \displaystyle\sum_{k=1}^{n} a_k^{s}(x,t_s) \left[B^{(s-1)}(x)\right]_{ik}^{-1}}{\sqrt{1 + \eta_{ss} \displaystyle\sum_{j,k=1}^{n} \left[B^{(s-1)}(x)\right]_{jk}^{-1} a_k^{s}(x,t_s)\, a_j^{s}(x,t_s)}} \tag{48}$$

VI. MODIFIED STAGEWISE PROCEDURE FOR TIME-VARYING ELEMENTS

In this section the elements will be considered functions of time, $x_i = x_i(t)$. It will be assumed that the elements vary much more slowly with respect to time than do the functions describing observations.

We suppose that the matrix $(\eta_{\mu\nu})$ can be written (as in Section II, Eq. (26)) as a diagonal array of matrices $\eta^{(s)}$ $s = 1, 2, \cdots$; the further assumption is made that the times t_μ, $\mu = M_{s-1} + 1, \cdots, M_s$, are sufficiently close together that they may be regarded as equal insofar as variation of the elements is concerned. This will be expressed

$$x_i(t_\mu) = x_i(T_s) \tag{49}$$

$$\mu = M_{s-1} + 1, \cdots, M_s .$$

Henceforth, the time parameter occurring in the argument of an element or element estimate will be written T.

It will be supposed that error-free observations are given by

$$F^\mu = g^\mu \left[x(T_s), t_\mu\right] \tag{50}$$

$$\mu = M_{s-1} + 1, \cdots, M_s$$

$$s = 1, 2, \cdots$$

and that the variation of the elements with respect to time is given by

$$x_i(T_s) = \mathscr{F}_i \left[x(T_r), T_r, T_s\right] . \tag{51}$$

Now consider the following stagewise procedure for processing the observations (the main reason for which is its adaptability to cases in which the prediction functions f^{μ} or $\mathcal{F}_i$ are imperfectly known):

The element estimate vector $P^{+}(T_1)$ is obtained by minimizing with respect to $P^{+}(T_1)$ the quadratic form

$$Q^{(1)} = \sum_{\mu,\nu=1}^{N_1} \eta_{\mu\nu}^{(1)} \left\{F^{\mu} - g^{\mu}\left[P^{+}(T_1), t_{\mu}\right]\right\}\left\{F^{\nu} - g^{\nu}\left[P^{+}(T_1), t_{\nu}\right]\right\} . \tag{52}$$

For $T_s \leqq T < T_{s+1}$, $s = 1,2,\cdots,$

$$P_i(T) = \mathcal{F}_i \left[P^{+}(T_s), T_s, T\right] . \tag{53}$$

Also,

$$P_i^{-}(T_{s+1}) = \mathcal{F}_i \left[P^{+}(T_s), T_s, T_{s+1}\right] . \tag{54}$$

The quantities $P^{+}(T_s)$ will be defined for $s > 1$ below. To do this, we define sequences of matrices $B^{(s)}$, $B_{+}^{(s)}$, $\psi^{(s)}$, $\psi_{+}^{(s)}$ as follows:

$$B_{+ij}^{(1)} = \sum_{\mu,\nu=1}^{N_1} \eta_{\mu\nu}^{(1)}\, b_i^{\nu}\left[P^{+}(T_1), t_{\nu}\right] b_j^{\mu}\left[P^{+}(T_1), t_{\mu}\right] \tag{55}$$

$$b_i^{\nu}\left[x, t_{\nu}\right] = \frac{\partial g^{\nu}}{\partial x_i}\left[x, t_{\nu}\right] \tag{56}$$

$$\psi_+^{(s)} = \left[B_+^{(s)}\right]^{-1}, \quad s = 1,2,\cdots \tag{57}$$

$$\psi^{(s+1)} = \sum_{k,\ell=1}^{n} a_{ik}\left[P^+(T_s),T_s,T_{s+1}\right]\; a_{j\ell}\left[P^+(T_s),T_s,T_{s+1}\right]\; \psi_{+k\ell}^{(s)} \tag{58}$$

$$a_{ij}\left[x,T_s,T\right] = \frac{\partial \mathcal{F}_i}{\partial x_j}\left[x,T_s,T\right] \tag{59}$$

$$B^{(s)} = 0, \; s = 1 \tag{60}$$

$$= \left[\psi^{(s)}\right]^{-1}, \; s > 1$$

$$B_{+ij}^{(s)} = B_{ij}^{(s)} + \sum_{\mu,\nu = M_{s-1}+1}^{M_s} \eta_{\mu\nu}^{(s)}\, b_i^{\nu}\left[P^+(T_s),t_\nu\right] b_j^{\mu}\left[P^+(T_s),t_\mu\right] \tag{61}$$

Then $P^+(T_s)$ is obtained for $s > 1$ by minimizing, with respect to $P^+(T_s)$, the quadratic form

$$Q^{(s)} = \sum_{i,j=1}^{n} B_{ij}^{(s)}\left\{P_i^-(T_s) - P_i^+(T_s)\right\}\left\{P_j^-(T_s) - P_j^+(T_s)\right\} \tag{62}$$

$$+ \sum_{\mu,\nu = M_{s-1}+1}^{M_s} \eta_{\mu\nu}^{(s)}\left\{F^{\mu} - g^{\mu}\left[P^+(T_s),t_\mu\right]\right\}\left\{F^{\nu} - g^{\nu}\left[P^+(T_s),t_\nu\right]\right\} .$$

Scrutiny of Eqs. (52) - (62) reveals that a stagewise smoothing procedure has been completely defined. Now, in order to give the first-order error equations for the resulting estimates, define the following matrices:

$$B_{+ij}\left[T_1\right] = \sum_{\mu,\nu=1}^{N_1} \eta_{\mu\nu}^{(1)}\, b_i^{\nu}\left[x(T_1),t_\nu\right]\, b_j^{\mu}\left[x(T_1),t_\mu\right] \tag{63}$$

(where $b_i^{\nu}(x,t_\nu)$ is defined as in (56));

$$\psi_+\left[T_s\right] = \left\{B_+\left[T_s\right]\right\}^{-1} \tag{64}$$

for $T_s < T \leqq T_{s+1}$,

$$\psi_{ij}\left[T\right] = \sum_{k,\ell=1}^{n} a_{ik}\left[x(T_s),T_s,T\right] a_{j\ell}\left[x(T_s),T_s,T\right] \psi_{+k\ell}\left[T_s\right] \tag{65}$$

(where $a_{ij}(x,T_s,T)$ is defined as in (59));

$$B\left[T\right] = \left\{\psi\left[T\right]\right\}^{-1}, \quad T_s < T \leqq T_{s+1} \tag{66}$$

and

$$B_{+ij}\left[T_s\right] = B_{ij}\left[T_s\right] + \sum_{\mu,\nu=M_{s-1}+1}^{M_s} \eta_{\mu\nu}^{(s)}\, b_i^{\nu}\left[x(T_s),t_\nu\right]\, b_j^{\mu}\left[x(T_s),t_\mu\right]. \tag{67}$$

Also define

$$\rho_i\left[T_s\right] = \sum_{j=1}^{n} B_{ij}\left[T_s\right]\left[P_j^-(T_s) - x_j(T_s)\right] \tag{68}$$

$$+ \sum_{\substack{\mu,\nu = \\ M_{s-1}+1}}^{M_s} \eta_{\mu\nu}^{(s)}\, b_i^{\nu}\left[x(T_s),t_\nu\right]\left\{F^\mu - g^\mu\left[x(T_s),t_\mu\right]\right\} .$$

Then, the first-order dependence of $P^+(T_s) - x(T_s)$ on the observation errors is:

$$P^+(T_s) - x(T_s) = \left\{B_+\left[T_s\right]\right\}^{-1} \rho\left[T_s\right] . \tag{69}$$

Also, for $T_s < T < T_{s+1}$, the first-order dependence of $P(T) - x(T)$ is

$$P_i(T) - x_i(T) = \sum_{j=1}^{n} a_{ij}\left[x(T_s),T_s,T\right]\left[P_j^+(T_s) - x_j(T_s)\right] . \tag{70}$$

Further, suppose that $(\eta_{\mu\nu})$ represents the inverse covariance matrix of $\left\{F^\mu - g^\mu\left[x(t_\mu),t_\mu\right]\right\}$. Then, $B_+\left[T_s\right]$ is the inverse covariance matrix of $\left\{P_i^+(T_s) - x_i(T_s)\right\}$, and $B\left[T\right]$ is the inverse covariance matrix of $\left\{P_i(T) - x_i(T)\right\}$, $T_s < T < T_{s+1}$.

In Section III, an alternative method was described for dealing with time-varying elements. That method involved a reduction to the case of constant elements.

Choose any time T, $T_s \leqq T < T_{s+1}$, and regard T as now fixed. Define a constant element vector x by: $x = x(T)$. Let $P^*(T)$ represent the estimate of $x = x(T)$ obtained by the method described in Section III, based on the first M_s observations.

Then, it can be verified that, to first order, $P^*(T) = P(T)$, where $P(T)$ is defined by (53). That is, the stagewise procedure described in (52) - (62) yields an estimate $P(T)$ for $x(T)$ having the same first-order dependence on $F^\mu - g^\mu \left[x(t_\mu), t_\mu\right]$, $\mu = 1, \cdots, M_s$, as the estimate derived by the method of Section III.

The stagewise procedure described above actually goes through the following steps: $P^+(T_s)$ represents an estimate of $x(T_s)$, based on the first M_s observations; for $T_s < T < T_{s+1}$, $P(T)$ is obtained by (53), that is, simply by prediction from $P^+(T_s)$ according to the functional relation by which the elements are known to vary; $P^-(T_{s+1})$ represents the estimate of $x(T_{s+1})$ just before the observations F^μ, $\mu = M_s + 1, \cdots, M_{s+1}$ are processed.

The particular method chosen above to define the matrices $B_+^{(s)}$ --i.e., by (61)--is not the only one possible. One could, for example, define matrices $\tilde{B}_+^{(s)}$ for $s > 1$ by using (61) with $P^-(T_s)$ instead of $P^+(T_s)$ in the arguments of b_i^ν and b_j^μ. Then one could compute $\tilde{B}_+^{(s)}$, for $s > 1$, before computing $P^+(T_s)$.

This would be convenient for purposes of using (20), (23), and (24) for a first-order determination of $P^+(T_s) - P^-(T_s)$: one would put $p = P^-(T_s)$; $P = P^+(T_s)$; $P^+(T_s) - P^-(T_s) = \left[\tilde{B}_+^{(s)}\right]^{-1} \rho^{*(s)}\left[P^-(T_s)\right]$, with $\rho^{*(s)}$ defined by the appropriate modification of (20) and (23).

In practical applications, for example, in satellite observations where perturbation forces are imperfectly known, the functions $\mathcal{F}_i$ may not be known exactly. Furthermore, the inaccuracy in knowledge of $\mathcal{F}_i\left[x(T),T,T'\right]$ will depend on the time difference $T' - T$ -- i.e., on how far ahead you are predicting the elements.

Suppose, for example, that the prediction functions used in the stage-wise procedure are g^μ, $\mathcal{F}_i$, but that the correct functions should be h^μ, $\mathcal{F}_i^*$. Let $\mathcal{E}^\mu$ still represent the observation errors--i.e., $\mathcal{E}^\mu = F^\mu - h^\mu$-- and let, for $T \geqq t_\mu$,

$$\delta^\mu(T) = h^\mu\left\{\mathcal{F}_1^*\left[x(T),T,t_\mu\right], \cdots, \mathcal{F}_n^*\left[x(T),T,t_\mu\right], t_\mu\right\} \tag{71}$$

$$- g^\mu\left\{\mathcal{F}_1\left[x(T),T,t_\mu\right], \cdots, \mathcal{F}_n\left[x(T),T,t_\mu\right], t_\mu\right\} .$$

Suppose that, in the case $g^\mu = h^\mu$, $\mathcal{F}_i = \mathcal{F}_i^*$, the first-order error in $P(T)$, $T_s \leqq T < T_{s+1}$, could be expressed

$$P_i(T) - x_i(T) = \sum_{\mu=1}^{M_s} \Gamma_i^\mu(T)\, \mathcal{E}^\mu . \tag{72}$$

Then, in the case $g^\mu \neq h^\mu$, $\mathcal{F}_i \neq \mathcal{F}_i^*$, one would have

$$P_i(T) - x_i(T) = \sum_{\mu=1}^{M_s} \Gamma_i^\mu(T)\left[\mathcal{E}^\mu + \delta^\mu(T)\right] . \tag{73}$$

provided $\delta^\mu(T)$ and $\mathcal{E}^\mu$ are sufficiently small.

The errors $\delta^{\mu}(T)$ would, in general, grow as T increases. In such cases, it would be desirable to modify the smoothing procedure so as to give more recent observations greater weight, and to continuously diminish the weight given to past observations. There are a number of possibilities for accomplishing this; one way, for instance, would be to define the matrices $B^{(s)}$ in (60) - (62) by

$$B_{ij}^{(s)} = \left[\psi^{(s)}\right]_{ij}^{-1} \lambda_i^{(s)} \lambda_j^{(s)} \qquad (60')$$

where $\lambda_i^{(s)}$ and $\lambda_j^{(s)}$ are ≤ 1.

This would mean that the effective smoothing matrix for the observations would not be the matrix (η) appearing in (52) - (62); in fact, there would really be no fixed smoothing matrix. Also, the first-order error equations (63) - (70) would have to be modified appropriately.

There is an analogy between this class of smoothing procedures and the process of filtering of signals. One might regard the smoothing procedures as filters with input $x(T)$ and output $P(T)$. The case where the f^{μ} and $\mathcal{F}_i$ are perfectly known is equivalent to a constant signal input to the filter, in which case the filter should have infinite memory to provide maximum smoothing of observation errors. If say $\mathcal{F}_i$ are not exactly known, this corresponds to the existence of an unpredictable time-varying component of the filter input; in this case the filter memory must be reduced (its 'bandwidth' increased); the best 'time constant' is related to the rate of variation of the unpredictable part of $x(T)$. One could also, in effect, make the time constant different for the different elements.