Enlarging the Region of Convergence of Kalman Filters Employing Range Measurements

William S. Widnall*
Intermetrics Inc., Cambridge, Mass.

Advances in nonlinear filter theory have produced the Gaussian second-order filter. This extension of the Kalman filter compensates for measurement nonlinearity by appropriate bias compensation and a priori measurement variance increase. This paper develops the Gaussian second-order measurement equations appropriate for the nonlinear elongation of measured range. The required additions to a linearized Kalman filter are not complex. The performance improvement obtainable is demonstrated in a spacecraft rendezvous navigation simulation. Divergence of the standard linearized filter occurs if the initial error exceeds 1 km. Divergence in the second-order filter does not occur unless the initial error exceeds 10 km.

Nomenclature

b	= parameter defined by Eq. (A8)
c	= parameter defined by Eq. (A3)
d	= difference between measured range and estimated range
$\mathbf{e}^-$	= error in estimated state vector before a measurement
$\mathbf{e}^+$	= error in estimated state vector after a measurement
e_1, e_2, e_3	= components of position error
$E[\ ]$	= expectation of (mean value of) enclosed quantity
g	= measurement nonlinear function of state
$\bar{g}$	= mean value of g before a measurement
g_0	= function g evaluated at the estimated state before a measurement
$\mathbf{h}$	= measurement gradient vector at the estimated state before a measurement
J	= measurement matrix of second partials at the estimated state before a measurement
$\mathbf{k}$	= filter gain vector for measurement incorporation
P^-	= estimation error covariance matrix before a measurement
P^+	= covariance matrix after a measurement
P_{xx}	= 3×3 position partition of the covariance matrix
P_{22}, P_{23}, P_{33}	= elements of the position covariance matrix
Q	= added covariance between measurements
r	= variance of measurement additive random error
r_A	= actual range
r_C	= range calculated using the estimated position
s	= parameter defined by Eq. (18)
$tr[\ ]$	= trace of the enclosed square matrix
$[\ ]^T$	= transpose
$\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$	= unit vectors
v	= measurement additive random error
v_r	= range measurement additive random error
Var[ ]	= variance of the enclosed quantity
$\mathbf{x}$	= actual state vector
$\mathbf{x}^-$	= estimate of state vector before a measurement
$\mathbf{x}^+$	= estimate of state vector after a measurement
z	= scalar measurement
$\bar{z}$	= mean value of z before a measurement
Φ	= state transition matrix

Introduction

ONE of the important assumptions underlying Kalman filter theory is that the measurements are linear functions of the system state. However, in many practical applications, the measured quantities are nonlinear functions of the system state. The usual approach in practical filter design is to linearize the measurement function about the current estimated value of the state. The linearized measurement description is then utilized by the Kalman measurement incorporation algorithm.

A Kalman filter designed in this manner works satisfactorily provided the linear approximation is accurate for the actual level of state estimation errors. If, however, the nonlinearity is comparable to the measurement error, then the standard filter yields inferior performance.[1] Divergence can occur such that the state estimation errors become orders of magnitude larger than the filter's own computation of the RMS estimation errors.[2]

Denham and Pines[1] noted that a significant effect of measurement nonlinearity is the introduction of measurement bias. Improved filter performance can be obtained by subtracting the expected value of the bias from the measurement. They also noted that proper calculation of the covariance after a measurement would require third- and fourth-order statistics of the estimation error. An alternate suggested approach was the introduction of artificially increased a priori variance for the measurements. Numerical examples[1,2] have demonstrated the success of this approach. Simulation (rather than some direct analytical relationship) is suggested as the method of determining the proper increased variance.

Advances in nonlinear filter theory have provided analytical formulas for computing an appropriate variance increase as a function of the measurement nonlinearity and the estimation error covariance. The Gaussian second-order filter[3,4] incorporates both the measurement bias compensation and the measurement variance increase. The apparent complexity of the Gaussian second-order filter may be preventing its wider application.

This paper develops the Gaussian second-order measurement equations appropriate for the nonlinear elongation of measured range. These equations are not complex and are a straightforward addition to any Kalman filter utilizing range measurements.

The performance improvement obtainable (with compensation for the nonlinear elongation of measured range) is demonstrated in a spacecraft rendezvous navigation simulation.

Nonlinear Elongation of Measured Range

Range measurement nonlinearity is illustrated in Fig. 1. The actual range r_A may be expressed in terms of the estimated range r_C and the position estimate error components e_1, e_2, e_3 as

$$r_A = [(r_C + e_1)^2 + e_2{}^2 + e_3{}^2]^{1/2} \tag{1}$$

Received July 14, 1972; presented as Paper 72-879 at the AIAA Guidance and Control Conference, Stanford, Calif., August 14–16, 1972; revision received October 19, 1972.

Index category: Navigation, Control, and Guidance Theory.

* Senior Engineer.

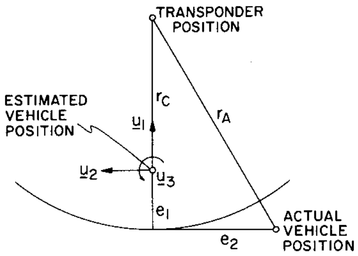

Fig. 1 Nonlinear elongation of measured range.

This may be expanded in a Taylor series. Retaining only the linear and quadratic terms yields

$$r_A = r_C + e_1 + (e_2^2 + e_3^2)/2r_C \tag{2}$$

The difference between the measured range and the estimated range is

$$d = v_r + e_1 + (e_2^2 + e_3^2)/2r_C \tag{3}$$

where v_r is the additive random error in the range measurement.

In many applications with very precise ranging equipment (v_r of the order of 1 m), the quadratic term can easily be the largest contributor to the measurement difference. Consider a position error e_2 of 4 km and a range r_C of 200 km. The quadratic term equals

$$e_2^2/2r_C = 40 \text{ m} \tag{4}$$

It is clear that if such a 40-m measurement difference were assumed to be evidence of a 40-m error e_1, then the subsequent filter performance would be unpredictable, to say the least. Such an assumption underlies the linear Kalman filter implementation.

Gaussian Second-Order Filter Equations

A scalar measurement z is assumed to be a nonlinear function of the state vector $\mathbf{x}$ plus an additive random noise v

$$z = g(\mathbf{x}) + v \tag{5}$$

The nonlinear function can be expanded in a Taylor series about the estimate $\mathbf{x}^-$ of the state before the measurement

$$g = g_0 - \mathbf{h}^T\mathbf{e}^- + \mathbf{e}^{-T}J\mathbf{e}^-/2 + \ldots \tag{6}$$

$$g_0 = g(\mathbf{x}^-) \tag{7}$$

$$\mathbf{h} = \partial g/\partial \mathbf{x} \,|\, \mathbf{x} = \mathbf{x}^- \tag{8}$$

$$J = \partial^2 g/\partial \mathbf{x}\,\partial \mathbf{x} \,|\, \mathbf{x} = \mathbf{x}^- \tag{9}$$

$$\mathbf{e}^- = \mathbf{x}^- - \mathbf{x} \tag{10}$$

The commonly used extended (linearized) Kalman filter incorporates the measurement z with the following equations

$$\mathbf{k} = P^-\mathbf{h}/(\mathbf{h}^T P^- \mathbf{h} + r) \tag{11}$$

$$\mathbf{x}^+ = \mathbf{x}^- + \mathbf{k}(z - g_0) \tag{12}$$

$$P^+ = (I - \mathbf{k}\mathbf{h}^T)P^-(I - \mathbf{k}\mathbf{h}^T)^T + \mathbf{k}r\mathbf{k}^T \tag{13}$$

The Gaussian second-order filter[3,4] takes into account a quadratic measurement nonlinearity. The filter is derived assuming the estimation error before the measurement has zero mean and has Gaussian distribution. A simplified derivation of the equations for a scalar measurement is presented in the Appendix. The resulting filter measurement incorporation equations are

$$\mathbf{k} = P^-\mathbf{h}/(\mathbf{h}^T P \mathbf{h} + r + tr[JP^-JP^-]/2) \tag{14}$$

$$\mathbf{x}^+ = \mathbf{x}^- + \mathbf{k}(z - g_0 - tr[JP^-]/2) \tag{15}$$

$$P^+ = (I - \mathbf{k}\mathbf{h}^T)P^-(I - \mathbf{k}\mathbf{h}^T)^T + \mathbf{k}(r + tr[JP^-JP^-]/2)\mathbf{k}^T \tag{16}$$

These equations differ from the more commonly used extended (linearized) Kalman filter equations in two ways: 1) the computed measurement g_0 (using the estimated state) is replaced by the a priori expected measurement, which takes into account the biasing effect of the nonlinearity

$$\bar{z} = g_0 + tr[JP^-]/2 \tag{17}$$

2) the variance r of the additive random measurement error is increased to account for the added variance due to the measurement quadratic nonlinearity J

$$s = r + tr[JP^-JP^-]/2 \tag{18}$$

In the case of the range measurement, these added terms are quite simple. For the coordinate system defined in Fig. 1 (unit-1 direction from the estimated vehicle position *toward* the transponder), the general coefficients of the Taylor series expansion are

$$g_0 = r_c \tag{19}$$

$$\mathbf{h}^T = [-1, 0, 0] \tag{20}$$

$$J = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1/r_c & 0 \\ 0 & 0 & 1/r_c \end{bmatrix} \tag{21}$$

Equations (17) and (18) reduce to

$$\bar{z} = r_c + (P_{22} + P_{33})/2r_c \tag{22}$$

$$s = r + (P_{22}^2 + 2P_{23}^2 + P_{33}^2)/2r_c^2 \tag{23}$$

The required elements of the position covariance matrix must also be expressed in the coordinate system shown in Fig. 1. The position partition P_{xx} of the filter covariance matrix in general is expressed in some other coordinate system. Let $\mathbf{u}_2$ and $\mathbf{u}_3$ be two orthogonal unit vectors constructed orthogonally to the estimated direction of the range measurement. Then the required elements are

$$P_{22} = \mathbf{u}_2^T P_{xx} \mathbf{u}_2 \tag{24}$$

$$P_{23} = \mathbf{u}_2^T P_{xx} \mathbf{u}_3 \tag{25}$$

$$P_{33} = \mathbf{u}_3^T P_{xx} \mathbf{u}_3 \tag{26}$$

Rendezvous Navigation Simulation

The performance improvement obtainable using the Gaussian second-order filter has been demonstrated in a spacecraft rendezvous navigation simulation. The rendezvous target is in a 500-km-alt circular orbit. The pursuer begins rendezvous navigation at $t = 0$ while in a 315-km-alt circular orbit 2300 km behind the target. The state vector estimated is the six elements of pursuer position and velocity. Target ephemeris is assumed known. Range to the target is the sole source of measurement until within 100 km of the target, at which time optical direction measurements are added. The rendezvous timeline is presented in Table 1.

The pursuer initially is actually in a coplanar circular orbit. The estimated pursuer state is in error in all six components. The initial covariance matrix assumed is in good agreement with the actual level of errors.. The choice of initial covariance matrix is such that if it is integrated forward, using no measurements, it exhibits a steady radial error variance, a steady cross-range error variance, and a down-range error variance that grows with time.

The actual and estimated state vectors are computed by integrating their differential equations assuming an inverse-square-law gravity model. In addition, the 6 × 6 transition

Table 1 Rendezvous timeline

Time, hr:min	Alt. diff., km	Range, km	Event
0:00	315	2326	Covariance matrix initialization
0:24	315	1565	First of 4 range measurements
0:27	315	1470	Last of the 4 range measurements
0:44	315	940	Coelliptic-sequence-initiation burn
1:09	137	220	First of 4 range measurements
1:12	109	177	Last of the 4 range measurements
1:29	20	124	Constant-differential-height burn
1:56	18	72	First of 4 range and 4 direction measurements
1:59	17	67	Last of 4 range and 4 direction measurements
2:16	18	41	Terminal-phase-initiation burn

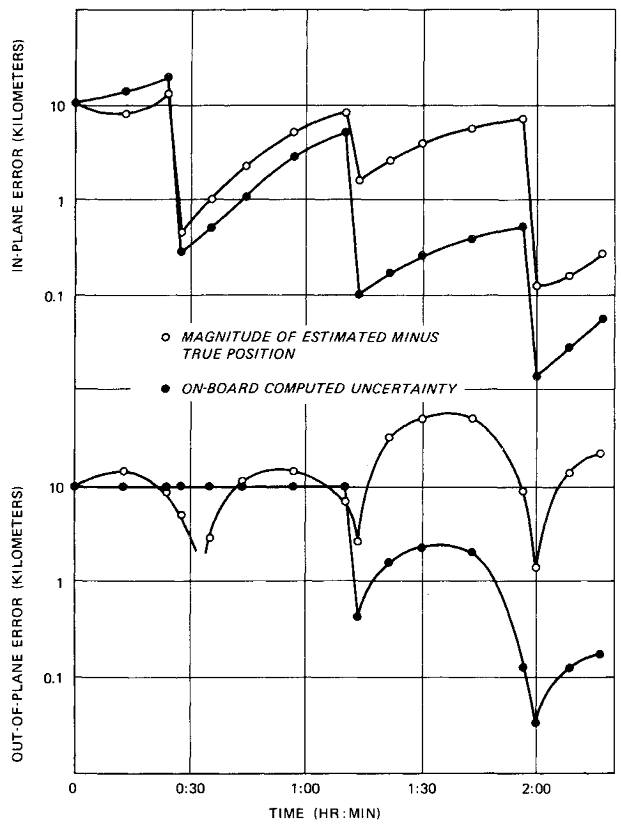

Fig. 2 Divergence of the standard Kalman filter.

matrix Φ, which governs the pursuer estimated state errors, is calculated by integrating the appropriate matrix differential equation. The 6×6 covariance matrix P is propagated from after a measurement at t_0 to before the next measurement at t_1 by

$$P(t_1) = \Phi(t_1, t_0)P(t_0)\Phi^T(t_1, t_0) + Q \qquad (27)$$

where Q is a positive definite diagonal matrix representing the growth in estimation error variance from t_0 to t_1 due to computational and other errors.

The periods of powered flight are simulated as impulsive changes in the pursuer actual and estimated velocity vector. It is assumed that inertial navigation is used during powered flight. The simulation includes added velocity navigation error during powered flight due to inertial system misalignment and accelerometer errors. The covariance of the velocity navigation error is increased an appropriate amount after each period of powered flight.

When the pursuer is within 1600 km of the target, direct measurements of the range between the pursuer and target become available. The range measurements are assumed to be highly precise. There is a bias of 1 m, a random error of 1 m 1σ, plus a scale factor error of 1 ppm. When processing a range measurement, the navigation filter assumes a measurement variance equal to the sum of the squares of the 1σ bias, the 1σ random error, plus the 1σ scale factor error times range.

When the pursuer is within 100 km of the target, optical measurements of the direction from the pursuer to the target become available. The error model used in the simulation includes a random error of 2 arc min 1σ in each axis plus a bias error of 1 arc min in each axis.

Three periods of navigation measurement processing are simulated. In the first period four range measurements are processed, one measurement every 60 sec. (Four measurements are the minimum necessary to observe the four inplane state vector errors.) In the second period, four more range measurements are processed. In the last period, four range measurements plus four optical direction measurements are processed, one set of range plus direction data every 60 sec.

The performance of the standard (extended or linearized) Kalman filter is shown in Fig. 2. The magnitude of the estimated minus true position is shown for the inplane and out-of-plane components of the state vector. (Note the naturally sinusoidal out-of-plane position error appears as a series of positive half cycles, because it is the magnitude that is plotted.) Also shown is the onboard computed navigation uncertainty (square root of the appropriate combination of elements of the covariance matrix). The first set of range measurements (from 24 to 27 min) is processed satisfactorily. The onboard computed uncertainty is in good agreement with the actual navigation error.

At the second set of range measurements (from 1 hr 09 min to 1 hr 12 min), serious problems are evident. While the actual pursuer trajectory is coplanar with the target trajectory, the estimated pursuer trajectory is not coplanar. With the estimated out-of-plane distance of the order of 10 km and a range 200 km to the target, the navigation filter believes it has adequate geometry to begin reducing the uncertainty in out-of-plane position. The true coplanar geometry presents no such opportunity. The subsequent incorrect processing of the range data produces a large discrepancy between the actual navigation errors and the computed uncertainty.

The third and final period of navigation measurements includes both range and direction measurements. But the filter is unable to recover from the large discrepancy. There is a factor of 100 disagreement between the actual estimation error and the computed uncertainty at the end of the simulation (at the time of the terminal-phase-initialization burn $t = 2$ hr 16 min). With an out-of-plane navigation error of more than 10 km at this point, the rendezvous would be unsuccessful.

The vastly improved performance using the compensation for the nonlinear elongation of the measured range is shown in Fig. 3. The onboard computed uncertainty remains in good agreement with the actual level of the navigation errors. During

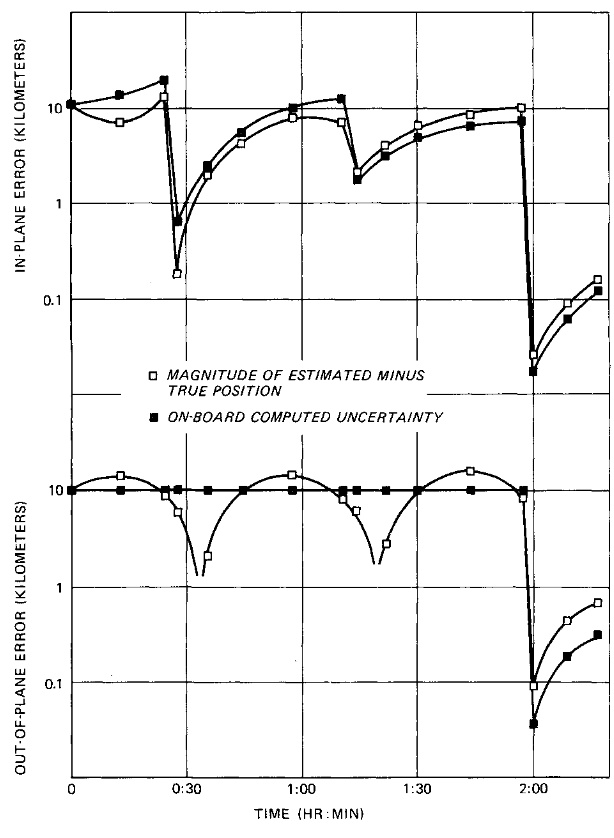

Fig. 3 Improved performance with compensation for nonlinear elongation of measured range.

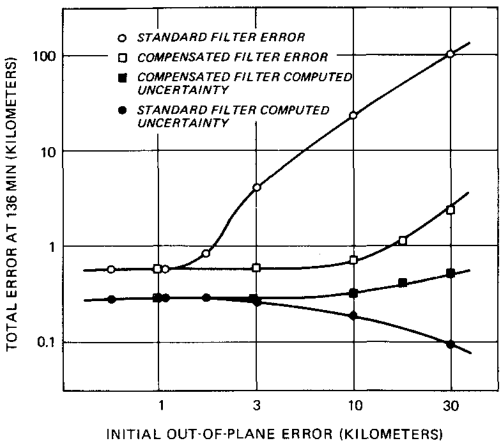

Fig. 4 Comparison of compensated and standard filter performance as a function of the initial out-of-plane error.

the two periods of range-only measurements, the computed uncertainty in out-of-plane position correctly remains at the initial level. The good covariance matrix, preceding the third and final navigation measurement period, permits optimal utilization of the precise optical direction data. The actual navigation error of less than 1 km at the end of the simulation is adequate for successful rendezvous.

The simulation has been repeated for various initial navigation errors. In each simulation the covariance matrix has been initialized in agreement with the actual level of errors. The resulting experimentally-determined domains of convergence of both the compensated and the standard Kalman filters are shown in Fig. 4. Beyond a 1 km initial error, the standard filter produces divergent results. Divergence in the compensated filter is delayed beyond 10 km initial error.

Conclusions

The Gaussian second-order measurement equations appropriate for the nonlinear elongation of measured range have been derived. These equations are a simple addition to any Kalman filter utilizing range measurements. The added terms are estimates of the bias and variance increase (of the difference between the measured and calculated range) caused by measurement nonlinearity and state estimate covariance.

The enlarged domain of convergence of one navigation filter incorporating these additions has been demonstrated in this paper in a spacecraft rendezvous navigation simulation. The author has also utilized these equations successfully in the design of precision-distance-measuring-equipment aided inertial navigation systems for aircraft.

One word of caution: the added variance is proportional to the square of the filter-computed covariance. It is necessary to initialize the covariance in agreement with the actual order of magnitude of the estimation errors for the Gaussian second-order filter to function properly. If too large an initial covariance is assumed, the added measurement variance will be excessively large and as a result measurement data will be rejected.

Appendix: Minimum Variance Estimator for Quadratic Measurement Nonlinearity and Gaussian Errors

Before incorporating a measurement, the state vector $\mathbf{x}$ is known to have mean $\mathbf{x}^-$ and covariance P^-

$$E[\mathbf{x}] = \mathbf{x}^- \tag{A1}$$

$$E[(\mathbf{x}-\mathbf{x}^-)(\mathbf{x}-\mathbf{x}^-)^T] = P^- \tag{A2}$$

It can be shown that the minimum-variance a priori (before the measurement) estimate of the state is $\mathbf{x}^-$.

A scalar measurement z is available, which is a nonlinear function $z(\mathbf{x}, v)$ of the state $\mathbf{x}$ and a noise v. Assume the following-form filter is used to combine the a priori state estimate with the measurement:

$$\mathbf{x}^+ = \mathbf{x}^- + \mathbf{k}(z-c) \tag{A3}$$

What should be the scalar c and the gain vector $\mathbf{k}$ to provide the minimum variance estimate $\mathbf{x}^+$? The estimation error is

$$\mathbf{e}^+ = \mathbf{e}^- + \mathbf{k}(z-c) \tag{A4}$$

where

$$\mathbf{e}^+ = \mathbf{x}^+ - \mathbf{x} \tag{A5}$$

$$\mathbf{e}^- = \mathbf{x}^- - \mathbf{x} \tag{A6}$$

The covariance of the estimation error after incorporating the measurement is

$$P^+ = P^- + \mathbf{k}E[(z-c)\mathbf{e}^{-T}] + E[\mathbf{e}^-(z-c)]\mathbf{k}^T + \mathbf{k}E[(z-c)^2]\mathbf{k}^T \tag{A7}$$

Let c be expressed in terms of the a priori mean value of z as

$$c = \bar{z} + b \tag{A8}$$

The resulting covariance formula is

$$P^+ = P^- + \mathbf{k}E[(z-\bar{z})\mathbf{e}^{-T}] + E[\mathbf{e}^-(z-\bar{z})]\mathbf{k}^T + \mathbf{k}\,\mathrm{Var}[z]\mathbf{k}^T + \mathbf{k}b^2\mathbf{k}^T \tag{A9}$$

To minimize the covariance for any $\mathbf{k}$, choose $b = 0$. Thus, the optimum value for c is the a priori mean value of the measurement z. The assumed filter Eq. (A3) then becomes

$$\mathbf{x}^+ = \mathbf{x}^- + \mathbf{k}(z-\bar{z}) \tag{A10}$$

A necessary condition that $\mathbf{k}$ be the optimum (minimum variance) gain vector is that

$$\partial\, tr[P^+]/\partial\mathbf{k} = \mathbf{0} \tag{A11}$$

From Eq. (A9) and with $b = 0$, it can be shown

$$\partial\, tr[P^+]/\partial\mathbf{k} = 2E[\mathbf{e}^-(z-\bar{z})] + 2\mathbf{k}\,\mathrm{Var}[z] \tag{A12}$$

Thus, the optimum gain vector for Eq. (A10) is

$$\mathbf{k} = -E[\mathbf{e}^-(z-\bar{z})]/\mathrm{Var}[z] \tag{A13}$$

In many applications the noise can be modeled additive. That is, the measurement is of the form

$$z(\mathbf{x}, v) = g(\mathbf{x}) + v \tag{A14}$$

where

$$E[v] = 0 \tag{A15}$$

$$E[v^2] = r \tag{A16}$$

$$E[v\mathbf{x}] = \mathbf{0} \tag{A17}$$

In this case the measurement a priori mean is

$$\bar{z} = \bar{g} \tag{A18}$$

The measurement a priori variance is

$$\mathrm{Var}[z] = \mathrm{Var}[g] + r \tag{A19}$$

The numerator in Eq. (A13) is

$$E[\mathbf{e}^-(z-\bar{z})] = E[\mathbf{e}^-(g-\bar{g})] \tag{A20}$$

Thus, the optimum gain vector is

$$\mathbf{k} = -E[\mathbf{e}^-(g-\bar{g})]/(\mathrm{Var}[g]+r) \tag{A21}$$

Assume the measurement nonlinearity is adequately represented in the vicinity of the a priori state estimate as

$$g(\mathbf{x}) = g_0 - \mathbf{h}^T\mathbf{e}^- + \mathbf{e}^{-T}J\mathbf{e}^-/2 \tag{A22}$$

where $\mathbf{e}^-$ was defined in Eq. (A6) and

$$g_0 = g(\mathbf{x}^-) \tag{A23}$$

$$\mathbf{h} = \partial g/\partial\mathbf{x} \,|\, \mathbf{x} = \mathbf{x}^- \tag{A24}$$

$$J = \partial^2 g/\partial\mathbf{x}\,\partial\mathbf{x} \,|\, \mathbf{x} = \mathbf{x}^- \tag{A25}$$

The mean value of g is

$$\bar{g} = g_0 + tr[JP^-]/2 \tag{A26}$$

Note $\bar{g}$ is *not* equal to g_0 unless the nonlinearity J is zero. The mean square value of g is

$$E[g^2] = g_0^{\,2} + \mathbf{h}^T P^- \mathbf{h} + E[(\mathbf{e}^{-T}J\mathbf{e}^-)^2]/4 + g_0\, tr[JP^-] - \mathbf{h}^T E[\mathbf{e}^-\mathbf{e}^{-T}J\mathbf{e}^-] \tag{A27}$$

If the estimation error has zero mean and has Gaussian distribution it can be shown that Eq. (A27) reduces to

$$E[g^2] = g_0{}^2 + \mathbf{h}^T P^- \mathbf{h} + tr[JP^-JP^-]/2 + (tr[JP^-])^2/4 + g_0\, tr[JP^-] \qquad \text{(A28)}$$

The variance of g is the mean square minus the squared mean

$$\mathrm{Var}[g] = \mathbf{h}^T P^- \mathbf{h} + tr[JP^-JP^-]/2 \qquad \text{(A29)}$$

The numerator in Eq. (A21) becomes

$$-E[\mathbf{e}^-(g-\bar{g})] = -E[\mathbf{e}^-(-\mathbf{h}^T\mathbf{e}^- + \mathbf{e}^{-T}J\mathbf{e}^-/2 - tr[JP^-]/2)] = P^-\mathbf{h} \qquad \text{(A30)}$$

Thus, the optimum gain vector is

$$\mathbf{k} = P^-\mathbf{h}/(\mathbf{h}^T P^- \mathbf{h} + r + tr[JP^-JP^-]/2) \qquad \text{(A31)}$$

The covariance of the estimation error after incorporating the measurement was given in Eq. (A9). Recall $b = 0$. Using Eqs. (A19, 20, 29, 30) the covariance is

$$P^+ = (I - \mathbf{k}\mathbf{h}^T)P^-(I - \mathbf{k}\mathbf{h}^T)^T + \mathbf{k}(r + tr[JP^-JP^-]/2)\mathbf{k}^T \qquad \text{(A32)}$$

Note if the nonlinearity J is zero, then Eq. (A31) and Eq. (A32) are equivalent to the extended (linearized) Kalman measurement incorporation equations.

To summarize, the measurement incorporation equations are

$$\mathbf{k} = P^-\mathbf{h}/(\mathbf{h}^T P^- \mathbf{h} + r + tr[JP^-JP^-]/2) \qquad \text{(A33)}$$

$$\mathbf{x}^+ = \mathbf{x}^- + \mathbf{k}(z - g_0 - tr[JP^-]/2) \qquad \text{(A34)}$$

$$P^+ = (I - \mathbf{k}\mathbf{h}^T)P^-(I - \mathbf{k}\mathbf{h}^T)^T + \mathbf{k}(r + tr[JP^-JP^-]/2)\mathbf{k}^T \qquad \text{(A35)}$$

where g_0, $\mathbf{h}$, and J were defined in Eqs. (A23–25). These equations are equivalent to those proposed previously by Athans-Wishner-Bertolini[3] and by Jazwinski.[4]

References

[1] Denham, W. F. and Pines, S., "Sequential Estimation When Measurement Function Nonlinearity Is Comparable to Measurement Error," *AIAA Journal*, Vol. 4, No. 6, June 1966, pp. 1071–1076.

[2] Toda, N. F., Schlee, F. H., and Obsharsky, P., "Region of Kalman Filter Convergence for Several Autonomous Navigation Modes," *AIAA Journal*, Vol. 7, No. 4, April 1969, pp. 622–627.

[3] Athans, M., Wishner, R. P., and Bertolini, A., "Suboptimal State Estimation for Continuous-Time Nonlinear Systems from Discrete Noisy Measurements," *IEEE Transactions on Automatic Control* Vol. AC-13, No. 5, Oct. 1968, pp. 504–514.

[4] Jazwinski, A. H., *Stochastic Processes and Flitering Theory*, Academic Press, New York, 1970, pp. 340–346.

MARCH 1973 AIAA JOURNAL VOL. 11, NO. 3

Shape Functions and the Accuracy of Arch Finite Elements

ISAAC FRIED*
Boston University, Boston, Mass.

The rate of convergence of the finite element method is determined by the ability of the approximate strains to assume arbitrary polynomial states. A systematic way to generate shape functions assuring a certain accuracy is to start, then, with assumed polynomial strains and integrate for the displacements. In curved structures, this leads to a coupled system of shape functions for the normal and tangential displacements, which are not necessarily polynomials. Some further simplifications reduce these to coupled polynomial shape functions which might prove more efficient than uncoupled shape functions assumed independently for the different displacements. The spectral condition number of the global stiffness matrix for the curved beam increases with $(r/t)^2$—the ratio of radius of curvature to thickness squared. The extensional energy portion of the elastic energy is proportional to $(t/r)^2$. For a coarse mesh this might be smaller than the discretization error, not warranting the introduction of the exact small t/r into the extensional energy. By relating r/t to the number of elements so as to balance the extensional energy and the energy discretization error, r/t is eliminated from the extensional energy and consequently from the condition number. The resulting element converges with its full rate to the inextensional solution but without undue ill-conditioning in the stiffness matrix.

Introduction

THE completeness requirement for the shape functions in the finite element analysis of flat membranes and plates is well understood now. If the shape functions for the displacements in the element include a complete polynomial of degree p then it follows from Taylor's theorem that the energy error in boundary value problems[1,2] and the error in the eigenvalues in eigenproblems[3,4] is $O[h^{2(p+1-m)}]$ where h denotes the diameter of the element, and where $m = 1$ for the membrane and $m = 2$ for the plate (the first proof for the convergence of the finite element method was given by Synge[5,6] and it is remarkable that his proof is more realistic that some recent proofs[7] appearing in the mathematical literature). Polynomial interpolation schemes do not only provide an automatic means for increasing the rate of convergence of the finite element method but also permit, in a systematic way, the satisfaction of the interelement continuity requirements posed by the variational principle.

Essentially, the rate of convergence of the finite element method is determined by the degree of the highest complete polynomial in the approximate strains. In flat structures, polynomial shape functions of degree p for the displacements produce polynomial shape functions of degree p–m for the strains and vice-versa. Thus, in flat structures, starting with a polynomial expression for the displacements and differentiating for the strains is equivalent to

Received February 11, 1972; revision received October 13, 1972.
Index categories: Structural Static Analysis; Structural Dynamic Analysis.

* Assistant Professor, Department of Mathematics. Associate Member AIAA.