

AN EMPIRICAL STATE ERROR COVARIANCE MATRIX FOR THE WEIGHTED LEAST SQUARES ESTIMATION METHOD

Joseph H. Frisbee, Jr.*

State estimation techniques effectively provide mean state estimates. However, the theoretical state error covariance matrices provided as part of these techniques often suffer from a lack of confidence in their ability to describe the uncertainty in the estimated states. By a reinterpretation of the equations involved in the weighted least squares algorithm, it is possible to directly arrive at an empirical state error covariance matrix. This proposed empirical state error covariance matrix will contain the effect of all error sources, known or not. Results based on the proposed technique will be presented for a simple, two observer, measurement error only problem.

INTRODUCTION

Over the last decade or so interest has grown significantly with regards to the question of just how well state error covariance matrices represent the errors in estimated states. A principle motivating factor in this growth in interest in “covariance accuracy” is the issue of estimating the collision risk due to the close approach of any two Earth orbiting objects. Typically this involves an active satellite and either a debris object or an inactive/nonmaneuvering satellite. The International Space Station (ISS) is an example of a satellite subject to such collision risks. Due to an intended long operational lifetime and as it is part of the NASA Human Space Flight Program, it was determined before the launch of the first component of the ISS that a probability based method was necessary in order to have an effective and efficient collision risk reduction process.

Error covariance matrices arising from estimation processes have been viewed in either a qualitative or quantitative way. When viewed as a qualitative judgment of a state vector’s accuracy the only real requirement was that the error covariance matrix be an acceptable guide as to whether or not the estimate was useful for operations. Viewed as a quantitative measure, the error covariance matrix has usually come with doubt as to whether it really characterized the uncertainty in a state estimate. The theoretical covariance matrix coming from the usual weighted least squares (batch) estimation process is understood to represent only the state uncertainty associated with the assumed measurement errors. The effects of errors arising from mismodeling of the physics of the problem, algorithmic/computational simplifications and incorrect knowledge of the true measurement errors are generally excluded from the state error covariance matrix[†]. The usual approach to a better batch filter covariance matrix is to increase its size. That could simply mean “just multiply it by a big number”. As long as the resulting state uncertainty is operational-

* ISS Trajectory Operations, NASA/JSC (United Space Alliance, LLC),

[†] Sequential methods, Kalman for example, typically rely on process noise to account for such errors. More will be said on sequential filters and process noise later.

ly within reason, this method might be deemed acceptable. Other covariance scaling techniques have been and are being used. The most sophisticated scaling techniques involve a systematic and statistically formal study of the performance of the estimation process over a large population of appropriate estimates. The result of such a study is a set of error covariance scale factors (perhaps many sets for different scenarios) that, on an axis by axis basis, account for the average misrepresentation of the errors in the covariance matrices of the studied process. These statistical correction methods, regardless as to how effective they may be, all suffer from the problem that they are based on historical performance and rely on the invariability of the problem being investigated for their reliable application.

Dissatisfaction with the performance of the typical batch estimation algorithm to always provide an appropriate state error covariance matrix has lead to that algorithm being reexamined. Reexamination of the weighted least squares estimation method, invoking a literal and simple interpretation of the elements of the algorithm, results in the realization that the algorithm has, all along, provided a direct and simple path for the calculation of a formally correct empirical state error covariance matrix. This path also permits the direct calculation of approximate confidence intervals for the elements of the state error covariance matrix in its theoretical form. The empirical error covariance matrix includes all errors in the process without regard to the degree to which they are understood or whether they are even known to exist at all.

First will be a review of a typical presentation of a batch algorithm. Next, the derivation of the batch empirical error covariance matrix and the confidence intervals associated with the theoretical matrix elements will be presented. A simple example of the results of the use of the empirical method follows. Finally, ideas for further study either directly or indirectly connected to the empirical state error covariance matrix will be presented.

THE TYPICAL BATCH (WEIGHTED LEAST SQUARES) DERIVATION

This presentation of the batch weighted least squares algorithm reflects the contents of section 4.3.3 of Reference 1. An attempt will be made to be consistent with the notation and variable definitions from that text throughout this document. For that reason, explicit definitions for many variables will not be supplied here. While there is some variance among texts discussing this subject, they are all similar enough that any one of them should be sufficient to help in understanding the equations presented here. Additionally noted is that equations taken directly from Reference 1 will be noted as “TSB n.n.n” where n.n.n will be the equation number from that reference.

Typically, batch algorithms are based on the concept of minimizing the sum of the normalized squared deviations of the observations from their anticipated values. The normalization process is invoked so that measurements of different forms and qualities might all be brought to a common statistical basis involving unitless, scale similar quantities. Thus, statistically similar measurement errors will contribute similarly to the solution without regard to the type or absolute error of the measurement source. With this in mind, the weighted least squares cost function (TSB 4.3.19) is now presented.

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{y} - \mathbf{H}\mathbf{x})^T \mathbf{W}(\mathbf{y} - \mathbf{H}\mathbf{x}) \quad (1)$$

This particular function represents the total, normalized variance of all of the measurements to be included. That is, it is the sum of the squared, normalized deviations of all of the measurements from their estimated values. The next step is to differentiate this function with respect to the unknown state variable, \mathbf{x} . The result (TSB 4.3.20) of such a differentiation is:

$$\partial J / \partial \mathbf{x} = -\mathbf{H}^T \mathbf{W}(\mathbf{y} - \mathbf{H}\mathbf{x}) \quad (2)$$

When this derivative function is set equal to zero the result will be indicative of the function being at its minimum value. Solving directly for the unknown variable, with the addition of the usual notation change to indicate it is the value of the unknown variable which minimizes the function, the result (TSB 4.3.22) is obtained:

$$\hat{\mathbf{x}} = (\mathbf{H}^T \mathbf{W} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{W} \mathbf{y} \quad (3)$$

This solution is recognized as the least square estimate of the state for linear problems or the differential correction to the state for a problem having been linearized with respect to an initial estimate of the unknown state. For the linear differential correction problem, it is common to iterate on the solution until the estimated correction is essentially zero. At that point, the reference state is the value which results in a minimum of the linearized total squared variance of the measurement residuals. The final step in this typical formulation is to identify the initial coefficient in equation 3 with the covariance matrix of the error in the final state estimate (TSB 4.3.23):

$$\mathbf{P} = (\mathbf{H}^T \mathbf{W} \mathbf{H})^{-1} \quad (4)$$

Any text on this subject will almost always present more specific details of the justification of the result in equation 4. Those details are omitted here. However, in the context of this result, it is curious that no mention is made of the natural connection of this result with the experimental nature of the solution and the generally followed path in such problems. That is, to compute an empirical state error covariance matrix from the existing data.

A last point to emphasize here is that the estimate above is that of a statistical mean value. While this is generally obvious, seldom is it stated directly. The result is that most users may lose the connection between the solutions, their accompanying state error covariance matrices, and the statistical theory on which they are based. It will be this connection that is formally taken advantage of in the following derivation of the empirical, batch estimate, state error covariance matrix.

DERIVATION OF THE EMPIRICAL STATE ERROR COVARIANCE MATRIX

Derivation of the empirical state error covariance matrix for the batch estimate begins with the typical minimization of the weighted least squares cost function. This presentation will vary slightly from those in most current references on this subject. The cost function is written in a more traditional form using explicit summation notation and as a mean squared variance rather than as a total variance. These notational changes can be omitted from the final forms of the solutions but are useful in the derivation in terms of how they tie the batch estimate to usual statistical methods. The mean form of the cost function is thus represented by:

$$\tilde{J}(\mathbf{x}) = \frac{1}{2N} \sum_{i=1}^N (\mathbf{y}_i - \mathbf{H}_i \mathbf{x})^T \mathbf{W}_i (\mathbf{y}_i - \mathbf{H}_i \mathbf{x}) \quad (5)$$

This function, just as with the previous version, is differentiated with respect to the unknown variable, \mathbf{x} . Reference 1 has a coefficient of 1/2 in the original cost function, equation 1. This

coefficient is not necessary but serves to permit cancellation of the factor of two that results from differentiating a squared term. The divisor*, N, which might be assumed to represent the number of measurements could be canceled out as well. However, this “sample size” divisor will be carried along as it is significant to the process of deriving the empirical state error covariance matrix.

Differentiation of the mean squared variance cost function is performed just as is done in the traditional least squares derivation. The result is:

$$\partial \tilde{J} / \partial \mathbf{x} = -\frac{1}{N} \sum_{i=1}^N [\mathbf{H}_i^T \mathbf{W}_i (\mathbf{y}_i - \mathbf{H}_i \mathbf{x})] \quad (6)$$

Again, this derivative is set equal to zero. The result is then solved for the associated value of the unknown state variable. As noted above, the sample size is not discarded but instead is carried along with each term. Solving for the unknown mean value of the state variable yields:

$$\hat{\mathbf{x}} = \left(\frac{1}{N} \sum_{j=1}^N \mathbf{H}_j^T \mathbf{W}_j \mathbf{H}_j \right)^{-1} \left(\frac{1}{N} \sum_{i=1}^N \mathbf{H}_i^T \mathbf{W}_i \mathbf{y}_i \right) \quad (7)$$

Note that the coefficient matrix in equation 7 is just the standard theoretical estimate of the state error covariance matrix multiplied by the factor N, as is illustrated in equation 8. The derivation will also show that the coefficient matrix in equation 7 is also an estimate of a hypothetical population covariance matrix.

$$\text{NP} = N \left(\sum_{j=1}^N \mathbf{H}_j^T \mathbf{W}_j \mathbf{H}_j \right)^{-1} = \left(\frac{1}{N} \sum_{i=1}^N \mathbf{H}_i^T \mathbf{W}_i \mathbf{H}_i \right)^{-1} \quad (8)$$

At this point, a short review of basic statistics is useful. For experimental data, assumed here to be of vector form, the average of the sample is just the sum of all of the samples divided by the number of samples. This average is the estimate of the mean of the sampled population:

$$\bar{\mathbf{z}} = \frac{1}{N} \sum_{i=1}^N \mathbf{z}_i \quad (9)$$

Similarly, an unbiased estimate of the covariance matrix of the population from which such a sample is drawn is given by:

$$\tilde{\mathbf{P}}_p = \frac{1}{N-1} \sum_{i=1}^N [(\mathbf{z}_i - \bar{\mathbf{z}})(\mathbf{z}_i - \bar{\mathbf{z}})^T] \quad (10)$$

The relationship between this estimate of the variance of the population and the estimate of the variance of the mean is just the estimate of the population variance divided by the number of samples:

* While the “N” used here reasonably might be interpreted to be the number of observations, the correct divisor might be another integer. For example, this other integer could represent the total number of statistical degrees of freedom of the problem. The actual value of the divisor is irrelevant as the divisor will eventually cancel out in the calculation of the empirical error covariance matrix just as it does in the usual theoretical approach.

$$\tilde{\mathbf{P}} = \frac{1}{N(N-1)} \sum_{i=1}^N [(z_i - \bar{z})(z_i - \bar{z})^T] \quad (11)$$

Recall from statistics that the distributional character of the uncertainty in the sample mean of a population tends to normality as the sample size gets large. This is without regard to the character of the distribution of the population itself. This directly implies that the uncertainty in an estimated state vector should follow an approximately multivariate normal distribution, at least with large sample sizes.

Considering equation 7, all of the coefficient terms (excepting the second sample size divisor) may be combined in one term within the second summation. This result is rewritten as a single, measurement related coefficient matrix multiplying the corresponding measurement residual within the sum:

$$\hat{\mathbf{x}} = \left(\frac{1}{N} \sum_{i=1}^N \mathbf{A}_i \mathbf{y}_i \right) \quad (12)$$

Note the similarity of this expression to equation 9. If each sample vector within the sum of equation 9 is just recognized as the corresponding product term within the sum of equation 12, shown in equation 13, then the usual batch estimate solution can be interpreted as a direct application of the standard statistical approach to finding the mean of a sample of a population.

$$\mathbf{x}_i = \mathbf{A}_i \mathbf{y}_i = \left[\left(\frac{1}{N} \sum_{j=1}^N \mathbf{H}_j^T \mathbf{W}_j \mathbf{H}_j \right)^{-1} \mathbf{H}_i^T \mathbf{W}_i \right] \mathbf{y}_i \quad (13)$$

By way of analogy with the sample vectors in equation 9, and for convenience in relating the standard batch estimate to the usual statistical methods, the product vectors in equation 12 and individually in equation 13, may be interpreted as vectors in a hypothetical population. The statistical characteristics of this hypothetical population are such that the mean of a sample drawn from this population has exactly the same distributional character as the estimate from the batch method solution.

Now recall the iterative nature of the typical batch estimate. On convergence, the estimate itself is approximately zero. This allows equations 10 and 11 (with a large sample size assumption) to be rewritten^{*†} as:

$$\tilde{\mathbf{P}}_p = \frac{1}{N} \sum_{i=1}^N (z_i z_i^T) \quad (14)$$

* It is important to note here that the z_i appearing in equations 9, 10, 11, 13, 14 and 15 DO NOT represent the solutions (using generalized inverses) of the individual residual expressions appearing in equations 5 and 6. These z_i are presented in 9-11 and 13-15 to illustrate standard statistical forms and to relate the batch estimate to those forms. The specific definition of each of the z_i is given by equation 13.

† Using the converged solution is not just a matter of convenience. It is the most useful. Without thinking about how, it might be possible to employ the empirical error covariance matrix in the iteration process. However, with the nature of empirical estimates it is anticipated that more iterations would be required for convergence, provided the process just wouldn't diverge.

$$\tilde{\mathbf{P}} = \frac{1}{N^2} \sum_{i=1}^N (\mathbf{z}_i \mathbf{z}_i^T) \quad (15)$$

Based on the logic in the paragraph following equation 12 and using the expanded version of the notation shown in equation 13, the population covariance matrix in equation 14 and the mean covariance matrix in 15 may both be rewritten using more standard notation typical of the batch estimate process. This gives for these two matrices:

$$\tilde{\mathbf{P}}_p = \frac{1}{N} \sum_{i=1}^N [(\mathbf{NPH}_i \mathbf{W}_i \mathbf{y}_i)(\mathbf{NPH}_i \mathbf{W}_i \mathbf{y}_i)^T] \quad (16)$$

$$\tilde{\mathbf{P}} = \frac{1}{N^2} \sum_{i=1}^n [(\mathbf{NPH}_i \mathbf{W}_i \mathbf{y}_i)(\mathbf{NPH}_i \mathbf{W}_i \mathbf{y}_i)^T] \quad (17)$$

Each of these may be further simplified by rearranging and factoring out the covariance matrix terms. Note that the sample size will remain in the population covariance expression but is absent from the expression for the covariance of the mean. Also, the symmetry of \mathbf{P} is used to eliminate some matrix transpose symbols.

$$\tilde{\mathbf{P}}_p = \mathbf{NP} \left[\sum_{i=1}^N (\mathbf{H}_i \mathbf{W}_i \mathbf{y}_i \mathbf{y}_i^T \mathbf{W}_i \mathbf{H}_i^T) \right] \mathbf{P} \quad (18)$$

$$\tilde{\mathbf{P}} = \mathbf{P} \left[\sum_{i=1}^N (\mathbf{H}_i \mathbf{W}_i \mathbf{y}_i \mathbf{y}_i^T \mathbf{W}_i \mathbf{H}_i^T) \right] \mathbf{P} \quad (19)$$

The first result, equation 18, is the empirical covariance matrix of the hypothetical parent population associated with the batch estimate of the mean vector. Equation 19 is the empirical error covariance matrix of the batch estimate itself. Because each residual contains the contributions of all error sources, known or not, this empirical error covariance matrix will fully describe the uncertainty in a state estimate to within sampling accuracy. To show both the equivalence of the expression in 18 to that in equation 4, as well as the relationship between the population and mean covariance matrices it is only necessary to consider the expected values of the above two expressions. Since only the measurement residual terms are random values, the above may be rewritten using expectation notation as:

$$\mathbf{E}[\tilde{\mathbf{P}}_p] = \mathbf{NP} \left[\sum_{i=1}^N (\mathbf{H}_i \mathbf{W}_i \mathbf{E}[\mathbf{y}_i \mathbf{y}_i^T] \mathbf{W}_i \mathbf{H}_i^T) \right] \mathbf{P} \quad (20)$$

$$\mathbf{E}[\tilde{\mathbf{P}}] = \mathbf{P} \left[\sum_{i=1}^N (\mathbf{H}_i \mathbf{W}_i \mathbf{E}[\mathbf{y}_i \mathbf{y}_i^T] \mathbf{W}_i \mathbf{H}_i^T) \right] \mathbf{P} \quad (21)$$

With the assumption that each measurement weighting matrix is just the inverse of the covariance matrix of the error associated with that measurement, equations 20 and 21 become respectively:

$$\mathbf{E}[\tilde{\mathbf{P}}_p] = \mathbf{N}\mathbf{P} \left[\sum_{i=1}^N (\mathbf{H}_i \mathbf{W}_i \mathbf{H}_i^T) \right] \mathbf{P} = \mathbf{N}\mathbf{P} \quad (22)$$

$$\mathbf{E}[\tilde{\mathbf{P}}] = \mathbf{P} \left[\sum_{i=1}^N (\mathbf{H}_i \mathbf{W}_i \mathbf{H}_i^T) \right] \mathbf{P} = \mathbf{P} \quad (23)$$

The second equality on the right in each equation follows directly from the fact that the bracketed summation in each of these equations is just the inverse of the original description, in summation notation, of the standard, theoretical estimate of the state error covariance matrix of the batch estimation method (equation 8). Equation 23 shows that the empirical error covariance matrix of equation 19 is a proper representation of the state error, under measurement error only. The relationship between equations 22 and 23 similarly shows that equation 22 is a proper estimate of the population covariance.

Just to place these results in a frequently observed format, each of the above may be rewritten in the original matrix notation of Reference 1. As the population covariance is just the sample size (degrees of freedom?) multiplied times the covariance of the estimate of the mean, only the covariance of the mean is presented.

$$\tilde{\mathbf{P}} = \mathbf{P}\mathbf{H}^T \mathbf{W}\mathbf{Y}\mathbf{W}\mathbf{H}\mathbf{P} \quad (24)$$

The central matrix, \mathbf{Y} , is not simply the outer product of the appended residual vector with itself but is a block diagonal matrix with each block being the outer product of a specific measurement residual with itself.

$$\mathbf{Y} = \begin{bmatrix} y_1 y_1^T & 0 & \cdots & 0 \\ 0 & y_2 y_2^T & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & y_N y_N^T \end{bmatrix} \quad (25)$$

This block diagonal form implies independence of the measurements. There may be the possibility of this not being the case*. With independence, the expected value of the outer product of the appended residual vector with itself will be the same as the expected value of the matrix in equation 25. However, the simple outer product of the entire appended residual vector with itself, instead of the general block diagonal form in equation 25, will result in the empirical matrix in equation 24 being singular. In general, without strong evidence of a dependent condition, the empirical state error covariance matrix should be computed using equation 25, if equation 24 is

* An example of when dependence could occur might be when two or more observations belong to one track (a single sensor pass). In computing the empirical state error covariance matrix, it might be appropriate to define an appended super observation composed of the dependent observation residuals. This would result in a super block within equation 23 which would be the outer product of the appended super observation with itself. If a large number of such tracks for a specific sensor were included in an estimate, that should empirically account for the correlation between the observations in the tracks having such correlated measurements.

chosen over equation 19. The form of equation 19* implies independence of measurements so there is no ambiguity with regard to its meaning.

This concludes the derivation of the empirical state error covariance matrix for the batch estimate. Either one of equations 19 or 24 may be viewed as the description of how to actually compute the empirical state error covariance matrix. The empirical error covariance matrix may be computed for each of the iterations but only if a form of the empirical matrix consistent with equation 10 is used. The result will be iterative values of the empirical matrix. However, the final forms in equations 19 and 24 imply that the batch iteration is allowed to converge and only the results of the last iteration give rise to the final, working value of the empirical state error covariance matrix.

BATCH ESTIMATE COVARIANCE MATRIX ELEMENT CONFIDENCE INTERVALS

One aspect of any sampled data problem is that statistics derived from such data are inherently uncertain. For this reason, it is very common for confidence intervals to be determined for such statistics. In batch state estimation, the expected value of the state error covariance matrix has always represented, under measurement error only, a theoretical statement of the uncertainty in the estimated state. Given the sample data estimate of the uncertainty in the estimated state, it is appropriate to consider the uncertainty in the elements of the state error covariance matrix. In order to determine the error covariance matrix uncertainty both the variance and the third central moment of each element of the error covariance matrix need to be computed. These element by element statistics are determined in their theoretical form. The degree of difficulty in doing so depends on the assumed form of the distributions associated with the measurement residuals. Here it will be assumed that the measurement uncertainties have zero mean values and are normally distributed.

Derivation of the theoretical variance in each element of the batch state error covariance matrix follows directly from the standard statistical approach. The variance (second central moment) of any variable is computed as the expected value of the squared difference between the expected value of that variable and the possible random values of the same variable. To approach this for each element of the error covariance matrix, it is convenient to first repose the description of the error covariance matrix of equation 19 in terms of the contribution of each observation. This is motivated by the fact that the variance of the sum of independent random variables is just the sum of the variances of those same independent random variables. So under the assumption that the measurements are independent, the contribution of each to the variance of the empirical state error covariance matrix may be individually determined. The results are combined to determine the variance of any element of that error covariance matrix. Thus, the i -th measurement's contribution to the empirical state error covariance matrix is given by:

$$\tilde{P}_i = PH_i W_i y_i y_i^T W_i H_i^T P \quad (26)$$

Following common practice and as noted previously in section III, the measurement weighting matrix may be interpreted as the inverse of the covariance of the corresponding measurement error. This allows each measurement to be written in terms of the product of the measurement covariance matrix square root matrix (obtained through Cholesky decomposition of the inverse of the weighting matrix) and a normalized random column vector, u_i .

* Recall that this form specifically applies to the converged solution where the estimated state correction is assumed to be equal to zero.

$$y_i = S_i u = W_i^{-1/2} u_i \quad (27)$$

With substitution for y_i , and after simplification, equation 26 becomes:

$$\tilde{P}_i = P H_i S_i^{-1} u_i u_i^T S_i^{-T} H_i^T P \quad (28)$$

At this point it is useful to redefine the first three matrices on the right hand side of equation 28 as a single matrix:

$$B_i = P H_i S_i^{-1} \quad (29)$$

This new matrix, B_i , is dimensioned $p \times k_i$ where p is the number of elements of the state vector to be estimated and k_i is the number of elements of the measurement residual vector y_i . With this matrix substitution, equation 28 becomes:

$$\tilde{P}_i = B_i u_i u_i^T B_i^T \quad (30)$$

The last part of this preliminary is to give the element by element representation of the contribution of the i -th measurement to the state error covariance matrix. This is accomplished by noting that the (m,n) element of the contribution of the i -th measurement involves only rows m , $b_{i:m}$, and n , $b_{i:n}$, of B_i .

$$\tilde{P}_{i:mn} = b_{i:m} u_i u_i^T b_{i:n}^T \quad (31)$$

The expected value of the outer product of the normalized measurement residual u_i with itself is just the identity matrix. Therefore, the expected value of equation 31 is given by:

$$\mathbf{E}[\tilde{P}_{i:mn}] = P_{i:mn} = b_{i:m} b_{i:n}^T \quad (32)$$

This result, equation 32, is the element by element representation of the contribution of the i -th measurement to the theoretical state error covariance matrix in equation 23. Note that it is just the inner product of row m of B_i with row n of B_i .

To determine the variance of the element expressed in equation 31, the usual definition of the variance being the expected value of the square of the difference between that element and its expected value is employed. This definition directly gives the following relationship:

$$\mathbf{V}[\tilde{P}_{i:mn}] = \mathbf{E}[(b_{i:m} u_i u_i^T b_{i:n}^T - P_{i:mn})^2] \quad (33)$$

By way of a term by term expansion of the squared expression in equation 33, followed by appropriate application of the expectation operator to the random variable terms, it can be shown that equation 33 reduces to:

$$\mathbf{V}[\tilde{P}_{i:mn}] = b_{i:m} b_{i:m}^T b_{i:n} b_{i:n}^T + (b_{i:m} b_{i:n}^T)^2 \quad (34)$$

At this point, equations 32 and 34 give explicit representations of both the mean expected value of $\tilde{P}_{i:mn}$ and of its variance. Both mean and variance terms will be necessary in defining the characteristics of the distributions of the theoretical uncertainties of the covariance matrix ele-

ments. For some of the terms of the matrix the third central moment may also be necessary and so it is presented next. By definition, the third central moment is given by:

$$\mathbf{M}_3[\tilde{\mathbf{P}}_{i:mn}] = \mathbf{E}[(\mathbf{b}_{i:m} \mathbf{u}_i \mathbf{u}_i^T \mathbf{b}_{i:n}^T - \mathbf{P}_{i:mn})^3] \quad (35)$$

Through substitution and simplification, the right hand side of equation 35 reduces to:

$$\mathbf{M}_3[\tilde{\mathbf{P}}_{i:mn}] = 2\mathbf{b}_{i:m} \mathbf{b}_{i:n}^T \left(3\mathbf{b}_{i:m} \mathbf{b}_{i:m}^T \mathbf{b}_{i:n} \mathbf{b}_{i:n}^T + (\mathbf{b}_{i:m} \mathbf{b}_{i:n}^T)^2 \right) \quad (36)$$

The expressions in 32, 34 and 36 are used to determine the total element by element values through simple summation over the observation index i . These results are given by:

$$\mathbf{E}[\tilde{\mathbf{P}}_{mn}] = \mathbf{P}_{mn} = \sum_{i=1}^N \mathbf{b}_{i:m} \mathbf{b}_{i:n}^T \quad (37)$$

$$\mathbf{V}[\tilde{\mathbf{P}}_{mn}] = \mathbf{V}_{mn} = \sum_{i=1}^N \left[\mathbf{b}_{i:m} \mathbf{b}_{i:m}^T \mathbf{b}_{i:n} \mathbf{b}_{i:n}^T + (\mathbf{b}_{i:m} \mathbf{b}_{i:n}^T)^2 \right] \quad (38)$$

$$\mathbf{M}_3[\tilde{\mathbf{P}}_{mn}] = \mathbf{M}_{mn} = \sum_{i=1}^N \left[2\mathbf{b}_{i:m} \mathbf{b}_{i:n}^T \left(3\mathbf{b}_{i:m} \mathbf{b}_{i:m}^T \mathbf{b}_{i:n} \mathbf{b}_{i:n}^T + (\mathbf{b}_{i:m} \mathbf{b}_{i:n}^T)^2 \right) \right] \quad (39)$$

Before continuing, it is convenient to note that all of the outer products appearing in 37-39 are, by way of equation 32, nothing more than the contribution of each of the individual observations to the empirical error covariance matrix. Using this observation, it is possible to rewrite equations 37-39 as the following:

$$\mathbf{E}[\tilde{\mathbf{P}}_{mn}] = \mathbf{P}_{mn} = \sum_{i=1}^N \mathbf{P}_{i:mn} \quad (40)$$

$$\mathbf{V}[\tilde{\mathbf{P}}_{mn}] = \mathbf{V}_{mn} = \sum_{i=1}^N \left[\mathbf{P}_{i:mn} \mathbf{P}_{i:mn} + (\mathbf{P}_{i:mn})^2 \right] \quad (41)$$

$$\mathbf{M}_3[\tilde{\mathbf{P}}_{mn}] = \mathbf{M}_{mn} = \sum_{i=1}^N \left[2\mathbf{P}_{i:mn} \left(3\mathbf{P}_{i:mn} \mathbf{P}_{i:mn} + (\mathbf{P}_{i:mn})^2 \right) \right] \quad (42)$$

The convenience of these forms are that they show, as the empirical error covariance matrix is being calculated through observation by observation accumulation, the variance and third central moment of each of the covariance matrix elements may be computed from only the terms used to compute the matrix itself.

The results in equations 40, 41 and 42 (which use equation 32) are specifically based on the a priori information. As such, the confidence intervals to be determined using these results will describe how empirical error covariance matrix terms vary with respect to their corresponding theoretical covariance matrix terms under perfect system and measurement modeling. Such precise knowledge is seldom the case.

Given the results in equations 40, 41 and 42 the important question now is ‘‘How are the mean, variance and third central moment used to determine the characteristics of the uncertainties

in the error covariance matrix?” The characteristics of the uncertainty for any random variable are defined by the form of the appropriate probability density function (pdf). Generally speaking, any pdf is described by a set of parameters. For example, a univariate normal distribution is described using both its mean and its variance or, equivalently, its standard deviation. For the uncertainties in the elements of the error covariance matrix of the batch estimate in theoretical form, there are three reasonably good candidate distributions. These candidate distributions are the normal distribution, the chi-square distribution and the gamma distribution. As error covariance matrices are composed of two distinct classes of “random” variables, diagonal and off-diagonal elements, the discussion of appropriate distributions describing each will be done separately. Addressed first will be the diagonal elements of the error covariance matrix.

The expectation for the diagonal terms of an error covariance matrix is that they will be positive-definite since each represents the variance of the corresponding state element. Those with sufficient familiarity of probability and statistics might anticipate that the diagonal elements might have chi-square distributions. It turns out that independent samplings of an i.i.d. (independent, identically distributed) random variable do result in the variances of multiple random samples having a distribution that is chi-square with a degree of freedom associated with the size of the sample drawn. In the more general estimation problem being discussed here, the random variables being drawn are, in general, not i.i.d. While a chi-square fit might be reasonable to try, it is unlikely that empirical data will lead to integer value estimates for the degrees of freedom associated with the distributions of the diagonal covariance matrix elements of an given estimation problem. A better distributional form would be that associated with the gamma distribution. The gamma distribution allows for noninteger shape values whereas the chi-square distribution is intended to reflect integer degrees of freedom, which correspond to the “shape” of the gamma distribution. When the shape parameter of a gamma distribution is sufficiently large (analogous to when the degrees of freedom for a chi-square distribution is large) the gamma distribution may be approximated as a normal distribution. At this point it is appropriate to present the forms of the pdfs for both the gamma and normal distributions to be used as candidates for representation of the uncertainty of the diagonal elements of the error covariance matrix. Presented first is the gamma distribution pdf followed by the normal approximation pdf.

$$f_{\text{gamma}}(C_{mm}) = \frac{C_{mm}^{\alpha_{mm}-1} e^{-C_{mm}/\beta_{mm}}}{\Gamma(\alpha_{mm})\beta_{mm}^{\alpha_{mm}}} \quad (43)$$

$$f_{\text{normal}}(C_{mm}) = \frac{e^{-\frac{(C_{mm}-\mu_{mm})^2}{2\sigma_{mm}^2}}}{\sigma_{mm}\sqrt{2\pi}} \quad (44)$$

The constant parameters α_{mm} and β_{mm} in equation 43 represent, respectively, the shape and scale parameters of the gamma distribution function. The terms μ_{mm} and σ_{mm} in equation 44 represent the mean and standard deviation of the normal approximation. For the gamma distribution, the expected value (mean value) and the variance of the distribution are given by:

$$\mathbf{E}(C_{mm}) = \alpha_{mm}\beta_{mm} \quad (45)$$

$$\mathbf{V}(C_{mm}) = \alpha_{mm} \beta_{mm}^2 \quad (46)$$

For the theoretical state error covariance matrix it is necessary to equate (with m equal to n) equation 37 or 40 to equation 45 and equation 38 or 41 to equation 46. Solving for the unknown parameters α_{mm} and β_{mm} will yield the following two results:

$$\alpha_{mm} = \frac{\mathbf{P}_{mm}^2}{\mathbf{V}_{mm}} \quad (47)$$

$$\beta_{mm} = \frac{\mathbf{V}_{mm}}{\mathbf{P}_{mm}} \quad (48)$$

For the situation in which the normal distribution approximation might be appropriate^{*}, the results are even more simple. These normal distribution approximation results are for the theoretical estimate:

$$\mu_{mm} = \mathbf{P}_{mm} \quad (49)$$

$$\sigma_{mm} = \sqrt{\mathbf{V}_{mm}} \quad (50)$$

Having presented the parameters describing the distributions of the diagonal elements of the state error covariance matrix, the off-diagonal elements are next. Generally speaking, the procedure is the same as that for the diagonal elements of the matrix. The complicating factor is that the off-diagonal terms may be positive or negative. If it is clear that the magnitude of the mean value of any (m,n) off-diagonal element is significantly larger than the square root of the variance of that same (m,n) element then a normal approximation might serve as a reasonable pdf function for that element. Also, if the mean value of an off-diagonal element is essentially zero and its magnitude is much less than the square root of its variance then the normal approximation might again work. When the normal distribution does appear to be a good candidate for the distributions of off-diagonal elements of the error covariance matrix, equations 49 and 50 may be used to determine the means and standard deviations of appropriate off-diagonal elements of the error covariance matrix[†].

For cases that are much less clear, a more generalized version of the gamma distribution is suggested. This more generalized gamma distribution is of the same form as that shown in equation 43 except that an origin shift, μ_{mno} , is included. This more generalized gamma pdf is:

^{*} The shape, akin to the degrees of freedom, of the general estimation problem probably will not correspond to the number of observations. In fact the shape parameter may be substantially less than the number of observations. Therefore, in order to know if the normal approximation might be appropriate for the uncertainty in the diagonal matrix elements, it is necessary to actually compute the shape parameter from equation 47 first.

[†] As the form of the equations used to solve for the off-diagonal covariance element normal distribution fit parameters μ and σ are exactly the same as those in 49 and 50 the equations will not be explicitly given. It is just noted that, to create such a set of equations, one only needs to replace the “mm” subscripts by “mn” subscripts so that the expressions will represent any off-diagonal term of the error covariance matrix.

$$f_{\text{gamma}}(C_{mn}) = \frac{(C_{mn} - \mu_{mno})^{\alpha_{mn}-1} e^{-(C_{mn} - \mu_{mno})/\beta_{mn}}}{\Gamma(\alpha_{mn})\beta_{mn}^{\alpha_{mn}}} \quad (51)$$

Based upon this generalized pdf form the three parameters μ_{mno} , α_{mn} and β_{mn} may be determined, specifically for the theoretical case, from the following three relationships:

$$\mu_{mno} = \frac{(\mathbf{M}_{mn} \mathbf{P}_{mn} - 2\mathbf{V}_{mn}^2)}{\mathbf{M}_{mn}} \quad (52)$$

$$\alpha_{mn} = \frac{4\mathbf{V}_{mn}^3}{\mathbf{M}_{mn}^2} \quad (53)$$

$$\beta_{mn} = \frac{\mathbf{M}_{mn}}{(2\mathbf{V}_{mn})} \quad (54)$$

Having determined the parameters (theoretical or empirical) for any selected pdf form, the distributions may be used in the usual way to determine the confidence intervals associated with any selected error covariance matrix element. Typically, this would be something of the form, ‘‘Given the covariance element \mathbf{P}_{mn} , what is the interval for \mathbf{P}_{mn} ($\mathbf{P}_{high} > \mathbf{P}_{mn} > \mathbf{P}_{low}$) such that the probability that \mathbf{P}_{mn} lies within the interval is at least of the chosen content?’’ At this point, it might seem that the problem of the uncertainty in the estimated state error covariance matrix has been substantially solved (at least as by the methods presented in this paper). That in fact is not the case.

One reason is that the proposed confidence intervals for the error covariance matrix elements may not simply just be applied independently. Specifically, considering both diagonal and off-diagonal elements, the covariance matrix must remain positive definite regardless as to the apparent individual confidence intervals of its elements. So, perhaps, if one attempted to randomly generate error covariance matrices from the proposed confidence intervals it would be necessary to check for positive definiteness after each random matrix draw. Nonpositive definite matrices would have to be rejected and a new one drawn. But even this conditional Monte Carlo technique is not actually possible at this point.

Connected to the above, as to why the covariance matrix uncertainty described in this document is not complete is that the covariance between any two matrix terms has not been presented. All that have been given are the variances of each matrix element. To make this clear, suppose one constructs a vector where each element of the vector is a unique element of the empirical state error covariance matrix. Since each element is a statistic, a new random vector has been created. This random covariance vector then clearly implies, in general, the existence of an associated covariance matrix describing its uncertainty. What are presented in this paper are just the diagonal elements of the covariance matrix describing the uncertainty of the random covariance vector. The off-diagonal terms not presented are each as simple to determine as the presented diagonal variances. Any formal attempt to randomly generate state error covariance matrices would require the determination of all of the off diagonal elements of the covariance matrix describing the uncertainty in the random covariance vector.

A last point to be noted (though not necessarily the only remaining issue) is that of the interpretation of a confidence interval of an error covariance matrix. How that might come into play in analyzing the effect of the uncertainty of an error covariance matrix on some function of

the original state vector and/or its estimated error covariance matrix is difficult to anticipate. This issue also includes the characterization of the confidence interval given that the different elements of the state error covariance matrix may have different distributional types (normal, chi-square, or gamma). This complication may actually imply that the best approach is a Monte Carlo generation of error covariance matrices, and perhaps state vectors as well, followed by calculation of the functional value of interest. Such an involved Monte Carlo scheme would eventually build up the character of the uncertainty of any function of interest.

A SIMPLE EXAMPLE

In order to illustrate the potential usefulness of the empirical error covariance approach, a simple example is presented. This example is a basic geometric triangulation problem involving a stationary target and multiple, range only, observations from two independent stationary observers. The observer locations are fixed and known. For this simple problem only measurement noise will be introduced to corrupt the true, exact measurements. Two variations of this problem will be presented. The first variation will illustrate the effectiveness of the empirical method when the problem is ideal, which is, modeled correctly. So for the batch solutions the measurement errors associated with the observations from each of the two observers are correctly described by their reported, but different, sensor measurement uncertainties. The second variation will involve mismodeling of the anticipated measurement errors for each of the two observers. This will be accomplished for the batch solutions by simply switching the assumed measurement uncertainties for the two observers. Thus the first observer will have, as its reported measurement uncertainty, the true uncertainty of the second observer. The second observer will have as its reported measurement uncertainty the true uncertainty of the first observer. In both cases, the first observer will report 10 observations for each estimate and the second observer will report 20 observations to go into the same estimate. In fact, the same set of observations will be used for both variations with the only difference being that the assumed measurement errors will be switched in the mismodeled solution. The solutions will be repeated, with new and independent normal random error draws, for a total of 500 trials. Because the empirical method is only intended as an adjunct to the existing batch estimate, it is based on the final converged theoretical solution. For this reason, there is only one actual state estimate for each trial in each case. That is, the empirical covariance matrix is referenced to the same solution as the theoretical covariance matrix. The Table below documents the data used in both the ideal and mismodeled solutions. Note that for each solution, the assumed location of the target at the start of each estimate's iteration cycle was chosen to be the target's true location.

Table 1. Data for the Ideal and Mismodeled Problem Variations.

Data type	Observer 1 Location		Observer 2 Location		Target Location		Observer 1 Range Sigma	Observer 2 Range Sigma
	X1	Y1	X2	Y2	Xt	Yt		
True	0	0	14000	0	9000	12000	30	10
Ideal	0	0	14000	0	9000	12000	30	10
Mismodeled	0	0	14000	0	9000	12000	10	30

What follow now are the results for the ideal variation of the triangulation problem. Figure 1 displays the dispersion of the deviations of the 500 trial solutions of the target location from its

true position. Superimposed over this field of deviations are ellipses^{*†}, representing the theoretical error covariance matrix, the average (over the 500 trials) empirical error covariance matrix and the covariance matrix computed directly from the 500 trials. The average empirical error covariance matrix is used here as it best displays the goal of having a method which, on average, represents the actual uncertainty. Individual empirical error covariance matrices typically will be close to the mean value but there is some small likelihood of an empirical error covariance matrix being markedly different from the mean expectation. This is just a fundamental aspect of any sampled data process. In any case, it is clear that both the theoretical error covariance matrix and the average empirical error covariance matrix are consistent with the observed dispersion of the solutions as well as the dispersion covariance ellipse computed from those dispersed solutions.

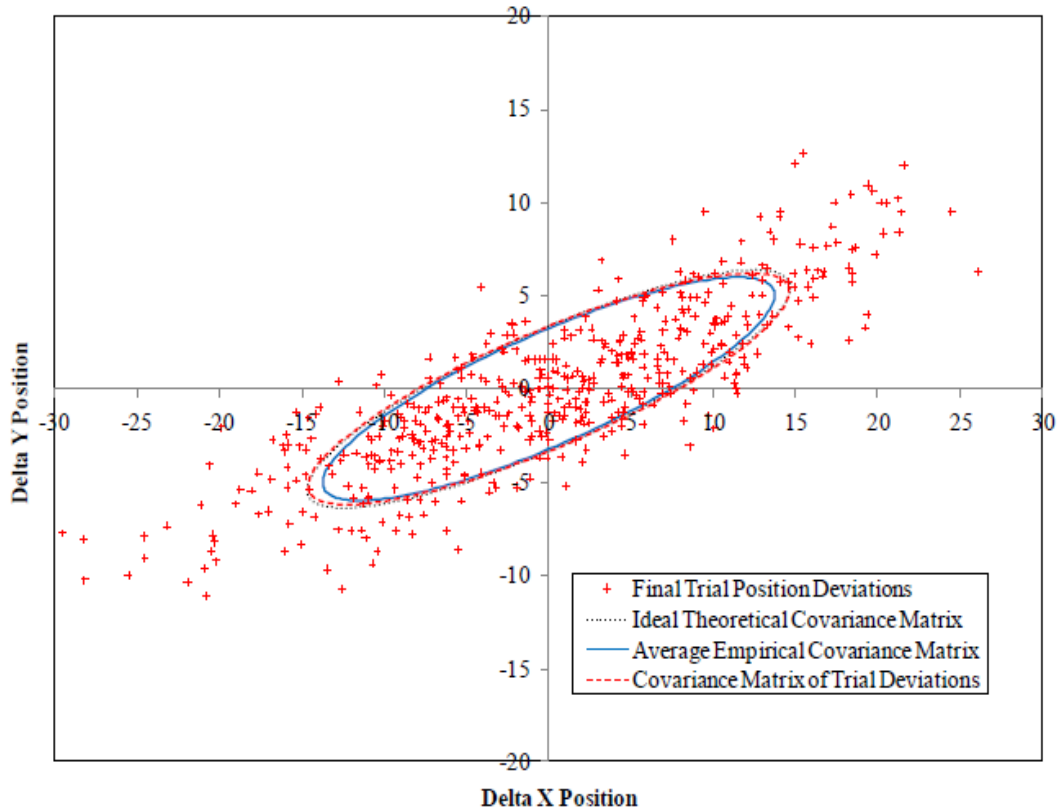


Figure 1. Ideal Case Solutions with Position Error Covariance Matrix Ellipses.

With some confidence being implied from Figure 1 that the empirical error covariance matrix does represent the uncertainty of an estimate, it is reasonable to consider what happens if the method is applied to the case where some mismodeling exists. As described by Table 1 and asso-

* The ellipses correspond to a chi-square value of 2, equivalent to a sigma level (Mahalanobis number) of the square root of 2. The choice of ellipse size is somewhat arbitrary and usually a value of 1 is used but the value of 2 appears most compatible with the data presented.

† All ellipses are referenced to the true target position.

ciated prior text, the case is considered where the presumed measurement errors are switched. This switching of measurement errors is not intended to imply that such a mistake is apt to happen. The intent is to directly show how the empirical error covariance matrix compares to the theoretical result under any measurement error misrepresentation. Indirectly, it implies that the empirical method should also accommodate errors induced by other sources.

Before presenting the results for the mismodeled case, it is of interest to note a special characteristic (presented without proof) of the specific, two observer, problem that is being examined here. It turns out that both the state updates for any iteration as well as the converged empirical error covariance matrices are independent of the assumed measurement errors. The result of this is that the set of 500 estimated solutions and their corresponding empirical error covariance matrices are unchanged from the results obtained in the ideal case. As a result, the covariance matrix of the 500 estimates and the average empirical error covariance matrix to be displayed in Figure 2 are exactly the same as those displayed in Figure 1. For other problems this condition should not be expected to exist. In any case, Figure 2 is now presented to display the effect of the measurement error assumption. Note that, due to its specific dependence on the assumed measurement error characteristics, the theoretical state error covariance matrix is drastically different from the mean empirical error covariance, the covariance matrix of the set of trial solutions as well and the theoretical error covariance matrix of the ideal case.

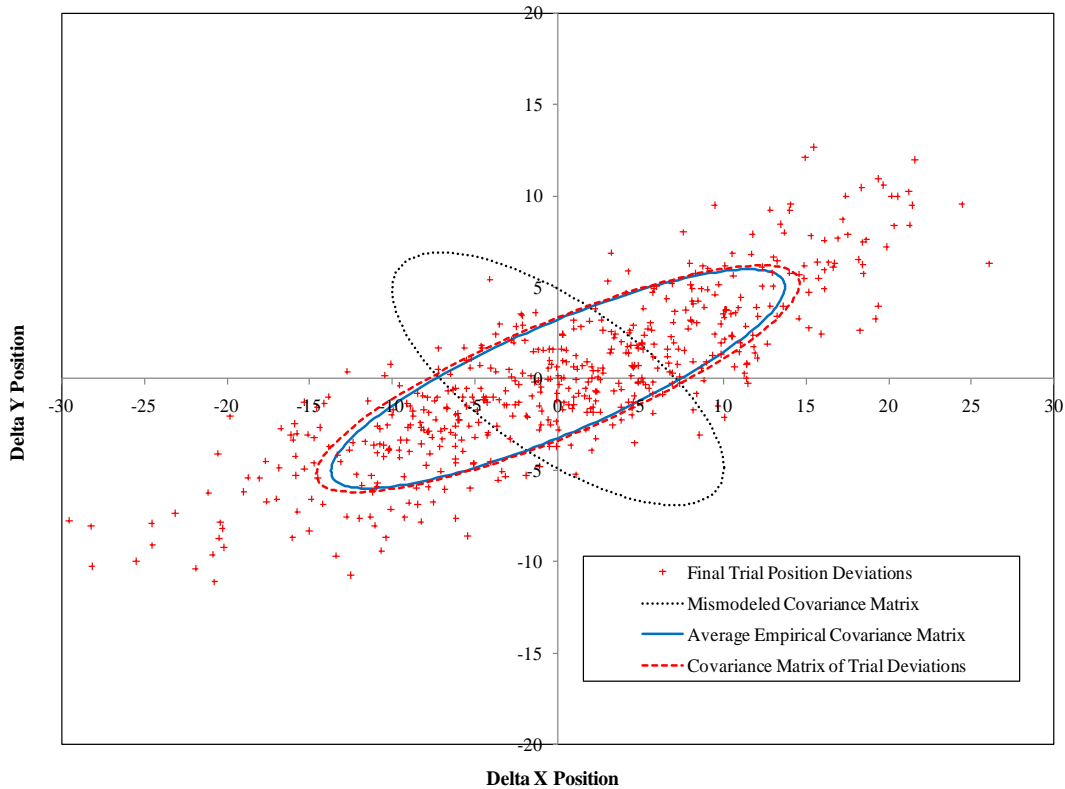


Figure 2. Mismodeled Case Solutions with Position Error Covariance Matrix Ellipses.

Figures 1 and 2 indicate clearly that the empirical error covariance matrix has an ability to reflect the actual uncertainty characteristics of an estimate even when the theoretical error cova-

riance matrix is incorrect. The summary statistics of both the ideal and the mismodeled cases are given in Table 2. The data in Table 2 represents average values over the 500 trials for theoretical (ideal and mismodeled) and empirical cases. The “Collective Solution Statistics” category represents the covariance of the 500 individual trial solutions.

Table 2. Summary of Position Error Covariance Matrix Values.

Matrix Element:	P_{xx}	P_{xy}	P_{yy}
Theory (ideal case average):	107.630	39.814	20.361
Theory (mismodeled case average):	50.280	-24.318	23.819
Empirical (both cases, average):	94.086	34.497	17.899
Collective Solution Statistics (both cases):	106.527	37.988	19.196

The question is how would a user know if any observed difference between the two error covariance matrices (or more specifically, their elements) might be due to mismodeling or is just a reflection of sampling variability? In other words, how is any difference between corresponding elements of the two error covariance matrices determined to be statistically significant? Using the technique described earlier, the constants, α and β , of the gamma distribution are determined for the x element covariance, P_{xx} , of the theoretical error covariance matrix in both the ideal and mismodeled cases. Using the gamma function constants, 95% confidence intervals are determined for each case. The empirical and collective values of P_{xx} are compared to the theoretical confidence intervals for each case. If the empirical or collective values fail to lie within the 95% confidence interval for either of the two theoretical cases then the failing case is said to disagree at the 5% significance level. In other words, at the 5% significance level, the difference is not explained purely by random variation. Comparing the P_{xx} values for the empirical and collective results in Table 2 to the confidence intervals in Table 3 it can be seen that the Pass/Fail notations in Table 3 are consistent with the desired confidence interval requirement. The conclusion is that the results shown in Tables 2 and 3 indicate that modeling for the theoretical position error covariance is consistent with the observation residuals in the ideal case but, for the mismodeled case, the modeling is inconsistent with the observation residuals.

Table 3. Summary of Theoretical Solution Gamma Distribution Results for P_{xx} .

Theoretical Case	α_{xx}	β_{xx}	$P_{xx_upper\ 95\%}$	$P_{xx_lower\ 95\%}$	Empirical	Collective
Ideal	5.421	19.853	215.455	36.970	Pass	Pass
Mismodeled	14.291	3.518	79.500	27.682	Fail	Fail

CONCLUSIONS

The principal and most significant conclusion within this work is that for the batch least squares estimate an empirical error covariance matrix exists and that this matrix will reflect all errors present in the modeling and measurement processes. This empirical error covariance matrix is a specific result for any single estimate and does not rely on historical data. Furthermore

the empirical error covariance matrix development also gives rise to a method by which uncertainty in error covariance matrices may be investigated. Substantial difference between the theoretical error covariance matrix elements and their corresponding empirical forms may indicate that modeling or measurement process error exists. Through confidence interval analysis it can be concluded whether or not the differences between the two matrix forms are significant. The result of this will be when to accept the empirical error covariance matrix as the appropriate representation of the state uncertainty rather than the usual theoretical error covariance matrix.

FUTURE WORK AND RELATED CONCEPTS

The principal goal of intended and in-progress work is to fully resolve the issue surrounding the proper sample size divisor. Resolution of this factor will open the door to complete utilization of both the empirical error covariance matrix and of the equivalent parent population matrix. A specific goal is to develop empirically based confidence intervals so that, when the empirical matrix is an appropriate alternative, the confidence intervals will empirically reflect the true physics and measurement processes at work in the problem. While minimally mentioned, the parent population covariance matrix may offer a great deal of utility. For example, it is suggested that the parent population matrix is the appropriate matrix for testing for observation outliers and more generally is desired for any search algorithm based on a least squares state estimate. An example of this would be searches for lost satellites, asteroids or aircraft.

Lastly, a very specific goal is to extend this theory to sequential estimates. Difficulties with this are more related to associated aspects of sequential estimates rather than their actual sequential character. Specifically, most sequential estimates involve the use of a process noise term. Process noise is commonly understood to account for unmodeled effects in the physics and measurement processes. While it may do this, process noise is more specifically understood to control over convergence of the state error covariance matrix which becomes vanishingly small as observations continue to be accumulated. This statistical contraction of the error covariance matrix is to be expected in typical sequential algorithms. For the empirical error covariance matrix, process noise is inherently included over the fit interval. In fact, it is likely possible to estimate any unknown process noise by comparing the empirical error covariance to the theoretical covariance and so have a real time process noise estimate to apply to propagated matrices. Regardless, applying the empirical matrix concept to sequential estimates will require various detailed considerations.

REFERENCES

1) Statistical Orbit Determination, Byron D. Tapley, Bob E. Schutz and George H. Born; Boston, Elsevier Academic Press, 2003.