Pursuing Appropriate Specs on Covariance Fidelity Measures by Analysis and Simulation

Thomas H. Kerr III, Ph.D.[1], TeK Associates

Abstract

This summarizes and explicitly documents all the topics TeK associates recently discussed with Boeing in 2003 and we offer our further assistance with the Tasks discussed in Sec. 9. We need inputs from Boeing to know how to properly bound the effort so that we can offer corresponding costs and schedules. The Boeing inputs that we solicit pertain to how many different cases should be run to have sufficient coverage (otherwise the evaluation task could be somewhat open-ended and we would need to bill for time and materials as cost plus fixed fee).

Table of Contents

1. **Application Constraints: estimating target states using Kalman Filter–like algorithms**
2. **Seeking Uniformity in Test Conditions Imposed in Performing Covariance Fidelity Tests**
3. **Impact of Covariance Fidelity on System Performance**
4. **An Appealing Candidate Approach to Covariance Fidelity Testing**
5. **History of Mahalanobis Distance use and Its Interpretation**
6. **Specifying reasonable alternative confidence regions and revealing the interrelationships between the two 3-D (for position and velocity separately) and 6-D (both handled jointly)**
7. **A More Robust Test Statistic Recommended for use to Avoid Ill-Conditioning**
8. **A Recommendation to use Proximity to Cramer-Rao Lower Bounds as a Gauge of on-line Covariance Fidelity**
9. **An Offer to Perform Specific Tasks that Support Boeing's Covariance Fidelity Analysis and Simulation Effort to Arrive at Appropriate Specs**

Appendix A: Simple Review of Some Equations Describing Ellipsoids
Appendix B: Proofs of Lemmas associated with TeK's new Results
References

1. Application Constraints: estimating target states using Kalman Filter–like algorithms

For a purely linear system (with Gaussian white process and measurement noises and Gaussian independent initial condition) using an optimal linear Kalman filter (with an identical underlying linear system model) for tracking the state as depicted in Fig. 1, the exact conditional distribution of the true (nx1) state vector **x**(t), given the measurements up to time t, is described by the following multidimensional Gaussian probability density function:

$$p_{x(t)}\left(\boldsymbol{\alpha} \mid \mathbf{Z}(t)\right) = \frac{1}{(2\pi)^{n/2}\left|\mathbf{P}(t)\right|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}\left(\boldsymbol{\alpha}-\hat{\mathbf{x}}(t)\right)^{\mathrm{T}}\left[\mathbf{P}(t)\right]^{-1}\left(\boldsymbol{\alpha}-\hat{\mathbf{x}}(t)\right)\right\}, \tag{1}$$

where $\hat{x}(t)$ is the outputted Kalman filter state estimate at time t (i.e., the optimal being the conditional expectation) and **P**(t) is the covariance Matrix computed on-line within the Kalman filter, generated using the recursive Matrix Riccati Equation.

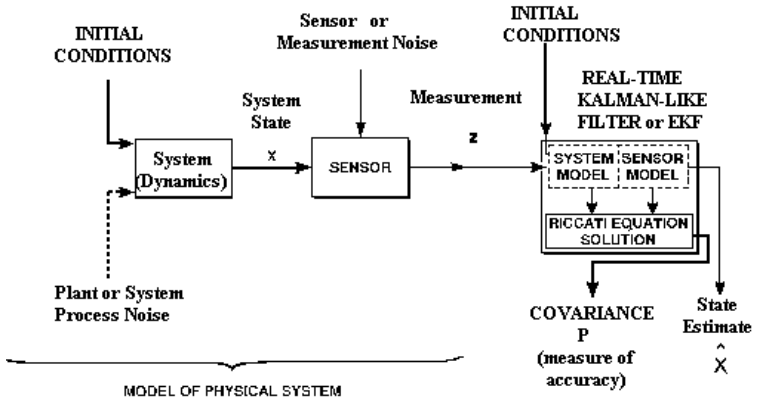

Figure 1: *Overview Functional Block Diagram of the Internal Structure of a Kalman Filter, or of an Extended Kalman Filter, or of every likely candidate NMD target tracking algorithm*

[1] Principal Investigator, **TeK Associates,** P.O. Box 459, 9 Meriam St., Suite 7-R, Lexington, MA 02420-5312. e-mail: thomas_h_kerr@msn.com.

Equation 1 is the solution to the Fokker-Planck equation (also called the forward Kolmogorov equation):

$$\frac{\partial p(x,t;y,\tau)}{\partial t} = -\frac{\partial\left[p(x,t;y,\tau)f(x,t)\right]}{\partial x} + \frac{1}{2}\frac{\partial^2\left[p(x,t;y,\tau)g^2(x,t)\right]}{\partial x^2} \tag{2}$$

that the ideal pdf of an optimal estimator, such as the Kalman filter, satisfies (as discussed in [1, pp. 129-135]). In Eq. 1, the covariance matrix **P**(t) is not diagonal in general so cross-correlations exist between the various variables that constitute the underlying system state **x**(t), as conveyed in the associated pdf underlying estimation algorithm mechanization. The NMD application departs somewhat from the above idealization, as discussed next.

NMD has nonlinear target and observation models and currently provides a target tracking function by expedient approximate nonlinear filtering via use of an EKF that is, of necessity, suboptimal in order to be tractably mechanized in real-time. Arguments supporting a Gaussian assumption for tractably assessing Covariance fidelity are offered here; and, moreover, can be numerically confirmed by specified statistical hypothesis *tests of Normality* before proceeding any further. In Figs. 1 and 2, NMD target system dynamics are described by an ordinary differenttial equation of the following form: $d\mathbf{x}(t)/dt = \mathbf{f}(\mathbf{x}(t))$, with $\mathbf{x}(t_o)$ being merely unknown and not random (but frequently do possess some randomness in the actual UEWR because of how the initial condition is provided, as deduced from pulse pairs on the same target [23]). Discrete-time radar sensor measurements are described by a vector algebraic equation: $\mathbf{z}(t_i) = \mathbf{h}(\mathbf{x}(t_i)) + \mathbf{v}(t_i)$, where $\mathbf{v}(t_i)$ is white Gaussian noise. The tracking estimate $\hat{\mathbf{x}}(t_i)$ is obtained from the *approximate* nonlinear filtering tracking algorithm (where it is *approximate* for reasons of practicality so it will yield real-time results).

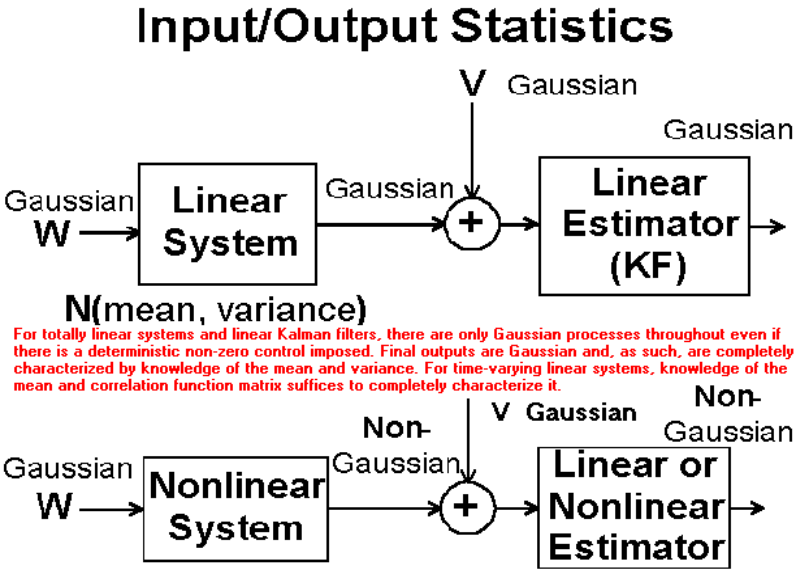

Figure 2: *Consequences of either a Linear or Nonlinear System Structure on Output Statistics*

Referring to Fig. 2, if the system were totally linear, then a transition matrix could be used to propagate all the states throughout. However, since an exoatmospheric target trajectory is nonlinear, its solution is calculated directly using Runge-Kutta integration in simulations and within the EKF target tracker model. The exoatmospheric target trajectory can be viewed as purely deterministic. The state variables describing the motion are not redundant because, by the

definition of what constitutes an adequate state variable model, they consist of the minimum number of variables (not necessarily unique) needed to completely describe the state of the system. The ballistic target described by this system model is accelerating (decelerating) because inverse squared gravity is present along with its harmonic of gravity J_2, which accounts for the oblateness of the earth (i.e., its departure from a perfectly spherical earth due to flattening at the poles) and its presence induces two more characteristic motions known as the "regression of nodes'' and the "rotation of Apsides''.

As mentioned above, initial condition specification and the gravity profile experienced cause the output of the exatmospheric trajectory modeled within UEWR to be purely deterministic and then Gaussian observation noise is added to represent what the UEWR radar receiver sees as a response to the presence of a target. UEWR tends to rely on position (range) being very good (due to design trade-offs having been historically made that favor high position accuracy over reaping good Doppler information). A factor is introduced in the tracking filter representation of the observation vector to better account for and compensate for the presence of significant range-Doppler ambiguity errors and the current system still utilizes this because it improves tracking filter performance [3]-[5]. This compensation is accomplished by modeling the radar range measurements as nonlinear observations of the target motions, as viewed from Earth-fixed Cartesian coordinates erected in the face plane of the UEWR. Measurements received after passing through the radar receiver electronics, matched filters, Automatic Gain Control (AGC) and associated signal processing have noisy spiky output. It's a random process with a significant deterministic component being the ballistic target of interest.

The tracking filter is then applied which better reconstructs the target of interest by smoothing out the effect of the noise spikes but does still correspond to a level of noise being present in the system. Moreover, to effectively increase the tracking filter's bandwidth, there is a nonzero process noise covariance intensity level parameter Q included in the tracking filter as an artifice that corresponds to a fictitious process noise level even though there is no process noise present in the actual ballistic target model. The outputs of the 6-state tracking filter are the 6 state estimates as a function of time and the accompanying covariances that serve as a rough on-line gauge of their quality, as both depicted in Fig. 1. The almost linear EKF used as a tracking filter operates on its input random process and yields this output random process. (The EKF possesses a linear filter structure, except that the *Jacobian* derivatives of the nonlinear system and measurement models are linearized about the state estimate from the last previously available time step thus making on-line Covariance and subsequent gain calculation no longer purely linear, an aspect that is also inherited by the outputted state estimate too as a consequence.) All 6 states are correlated in general and this feature is confirmable by the outputted covariance not being exclusively zero in the block off-diagonal terms.

The nonlinearities mentioned above constitute nonlinear operation on a Gaussian-like process (that yields a non-Gaussian output in general). However, it may be argued that the operations of an EKF are "almost" linear and so can be approximated as yielding a Gaussian process output but with a bias (which is unknown but can be estimated from realistic Monte-Carlo simulations at each time point t_i, of interest). This non-stationary aspect or time varying trend can be handled entirely within the framework offered here by comparison to non-central Chi-square statistics. (There are two different tests of Normality within The Math Work's Statistics Toolbox as well as programs to handle non-central Chi-Square, all to be identified explicitly in Sec. 7.)

2. Seeking Uniformity in Test Conditions Imposed in Performing Covariance Fidelity Tests

A somewhat obvious observation is that covariances provided on-line by the tracking filter exhibit greater accuracy when sensor measurements (radar target returns) are made available more frequently (by the Radar Scheduler) and used within the tracking filter more frequently as measurement updates. It is reminded that the associated track filter outputted estimates have correspondingly improved accuracy as well. What is to be tested is how well the on-line covariances compare to the accuracy of the outputted estimates (under the same conditions) as a measure of consistency. Since the outputted covariances are looked to as an indicator of quality of associated tracking estimates or estimate accuracy (since the true state is not available for comparison in the application scenario), expressing how reliable such interpretations can be based on viewing indications only from the covariance is called Covariance fidelity. The common source of both entities (as shown in Fig. 1) is the tracking algorithm being used.

Both tracking errors and covariances must be obtained from the same tracking algorithm in common under controlled conditions. To obtain an apples-to-apples comparison, results should be based on a particular single target track (with no other targets present that would otherwise complicate the effect with interactions and conflicts of radar resources allotted), in a particular trajectory, as viewed from a particular designated radar, (no matter which are used) as long as it is the same throughout (with all other pertinent underlying parameters fixed, e.g., target detection threshold setting, antenna pattern, electronic and atmospheric noise levels, etc). The target pulse return history on target should be identically the same (as much as practicable) in order to cross-compare the performance of alternative tracking algorithms as yet another use of this same statistical test defined below.

3. Impact of Covariance Fidelity on System Performance

As depicted in Fig. 3, having reasonably truthful covariances as the output of the tracking algorithm (Covariance Fidelity) is important in setting adequate gates for comparing new returns with existing target tracks previously started by the multi-target tracking algorithm since these available, less computationally intense covariances are used within the Multi-Hypothesis Test (MHT) for this purpose. (Once a track has persisted long enough to be declared *mature* by the MHT, its entire [saved] time history of detected returns above the detection threshold is run through the iterative Batch algorithm for an improved track estimate and better [possessing greater veracity] associated covariance [21]. It is only this Batch covariance that is used for interceptor handover.) However, before the target track is designated as *mature* by the MHT, "handover" on the same target as it comes into view from another face of the same radar uses the tracking algorithm's covariance. In this critical role of correctly categorizing the accuracy of the estimates that they accompany as an interpretation gauge, the goal is to have covariances that are neither larger than justified (i.e., being too pessimistic or conservatively large) or smaller than justified (i.e., being too optimistic in the quality of the estimates) but to better reflect the true situation (without compromising the accuracy already achieved by the tracking algorithm's estimates). Since tracker covariances are used in its gain calculation, better covariances should improve the accuracy of outputted estimates as well.

There are many system roles that would be improved by availing better tracker accuracy throughout or by meeting tight tracker accuracy goals earlier. The superior state vector estimation accuracy of an algorithm deemed worthy of replacing the existing UEWR filters, in

particular its more precise velocity estimate, is a direct benefit to the UEWR portion of the mission because it:

- Would enable earlier launch of the interceptor merely by satisfying accuracy guidelines sooner and therefore would provide more accurate in-flight-target-update (IFTU) for the kinetic kill vehicle;
- Would provide earlier support for phase-ambiguity-resolution and contextual feature discrimination;
- Would estimate better orbital elements for space object identification, and would thereby result in better satellite vs. missile discrimination;
- Would offer better launch and impact accuracy for legacy early warning functions;
- Would be potentially more robust in the ionospheric scintillation, range, and Doppler error environment;
- Should offer considerable dB savings of radar energy exerted to achieve a specified level of success in comparison with current EKF object tracking.

Regarding the first item above, the more accurate (non-optimistic) on-line prediction of 1-sigma bounds of the improved target tracking algorithm should help properly constrain the region that the interceptor needs to search for target acquisition. Use of an optimistic bound in this role would result in limiting search to too small a volume of space and therefore risk missing the target although supporting theoretical numerical calculations would falsely assure success (because they expect the available 1-sigma to be trustworthy, which it is not in general for earlier EKFs).

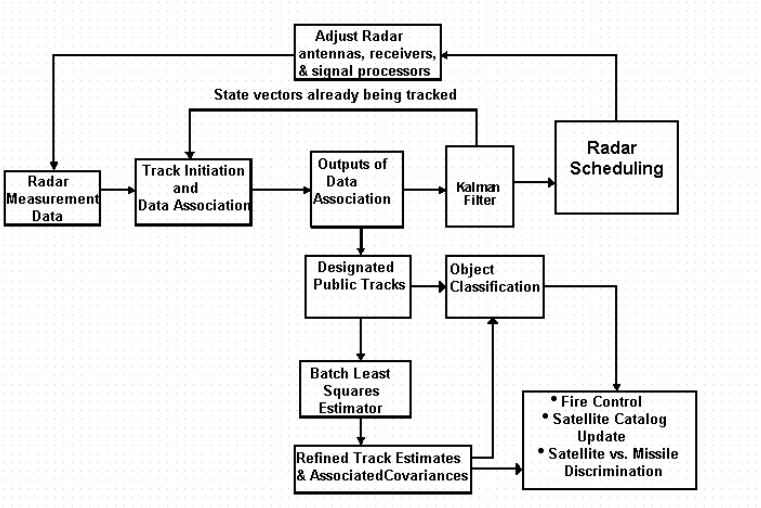

Figure 3: *Simplified High Level Overview of the UEWR Processing Architecture*

4. An Appealing Candidate Approach to Covariance Fidelity Testing

This test is a gauge of how consistent a particular tracking algorithm's covariance time history is with the algorithm's accompanying estimation errors accrued. Covariance fidelity is to be determined by performing a Monte-Carlo experiment to gather tracking algorithm state estimates, whose errors (or departure from the true target trajectory as the reference), are sample statistics that are then compared within this framework to the analytical covariance matrix also obtained from the on-line track algorithm under scrutiny.

The Mahalanobis normalized distance, familiar from Pattern Recognition, is defined here for a single target's track error at time t_i as:

$$r(t_i) \triangleq \sqrt{\mathbf{e}^{\mathbf{T}}(t_i)\left[\mathbf{P}(t_i)\right]^{-1}\mathbf{e}(t_i)} \tag{3}$$

where

$r(t_i)$ is the Mahalanobis distance (a scalar),

$\mathbf{e}(t_i)$ the Mahalanobis error vector consists of either all the (6x1) state errors at once or consists of 3 position error components in one evaluation along with 3 velocity error components in a separate evaluation for this NMD application,

$\mathbf{P}(t_i)$ is the corresponding submatrices of the (6x6) covariance matrix associated with position and velocity estimates to correspondingly match the grouping ultimately selected for the Mahalanobis error vector above.

At each test time, t_i, the intermediate quantity to be calculated for the test is the squared Mahalanobis distance:

$$r_j^2(t_i) \triangleq \mathbf{e}_\mathbf{j}^{\mathbf{T}}(t_i)\left[\mathbf{P}_\mathbf{j}(t_i)\right]^{-1}\mathbf{e}_\mathbf{j}(t_i) \tag{4}$$

where

$$\mathbf{e}_\mathbf{j}(t_i) \triangleq \hat{\mathbf{x}}_\mathbf{j}(t_i) - \mathbf{x}_{\mathbf{true}}(t_i) = \text{constituents of Mahalanobis error vector, defined above.} \tag{5}$$

For an on-line covariance $\mathbf{P}(t_i)$ generated within the tracking algorithm, there should be no additional diagonal loading or weighting used other than what comes from the tracking algorithm itself. A sampled estimate $\mathbf{S}^2(t_i)$ of the covariance $\mathbf{P}(t_i)$ could be used instead within Eq. 4, as now calculated exclusively from sample statistics. In using a sampled $\mathbf{S}^2(t_i)$, a sufficient number of Monte-Carlo trials, N, should be used for this estimated $\mathbf{P}(t_i)$ to ensure that this sample matrix $\mathbf{S}^2(t_i)$ is satisfactory to use by being invertible. The fact that $\mathbf{S}^2(t_i)$ needs to be inverted in Eq. 4 means that it is more sensitive to errors.

5. History of Mahalanobis Distance use and Its Interpretation

The Mahalanobis "distance" measure arises naturally within pattern recognition problems involving discriminant analysis, statistical communication and statistical detection theory, and more fundamentally in statistics. The simplest case is for an underlying scalar 1-D situation involving two different Gaussian populations, each of the same variance but with two different means, as depicted in Fig. 4. Since an obvious separating hyperplane can be erected at the origin that enables decisions with the smallest errors of the 1st and 2nd kinds, respectively, because of the symmetry exhibited in this problem. A random sample can be taken and its parent population or affiliation is decided based on whether the sample realization is to the left or right of the separating hyperplane (in this case the origin). There would be no decision error if the two Gaussian pdfs did not overlap but they do. This choice of a decision criterion gauges the effectiveness of the test by minimizing the decision error incurred and the critical spread is

$$|\text{mean1-mean2}|/\text{standard deviation} = \frac{|\mu_1 - \mu_2|}{\sigma}.$$

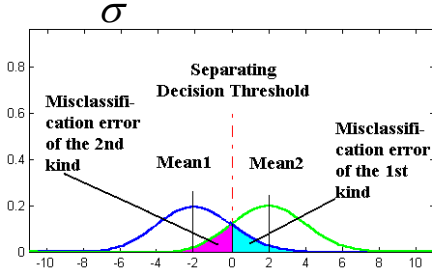

Figure 4: *Decision to be made for two underlying Gaussian Populations with same variance but different means*

For statistical decision situations involving two underlying multidimensional Gaussians, the above idea of what constitute the critical parameters underlying the making of correct decisions $\sqrt{\left[\frac{|\mu_1-\mu_2|}{\sigma}\right]^2} = \sqrt{\frac{|\mu_1-\mu_2|^2}{\sigma^2}}$ generalizes to $\sqrt{\frac{|\mu_1-\mu_2|^2}{\sigma^2}} \approx \sqrt{(\mu_1-\mu_2)^{\mathrm{T}}\,\mathrm{P}^{-1}(\mu_1-\mu_2)}$, which is recognized to be the Mahalanobis distance measure.

Now addressing the use of the name "distance" measure for this concept. It is well known from textbooks on Hilbert Space, like that of Paul Halmos or Berberian, that in a complete finite dimensional linear vector space, the following constitutes an inner product on the space when defined for any two points x and y as $(x|y) = x^{\mathrm{T}}\mathbf{W}y$, where **W** is symmetric and positive definite. Moreover, a norm can be defined in terms of the underlying inner product as $\|x-y\| \triangleq \sqrt{(x-y\,|\,x-y)} = \sqrt{(x-y)^{\mathrm{T}}\,\mathrm{W}(x-y)}$ and a distance measure can be defined in terms of the norm as $d(x,y) \triangleq \|x-y\| = \sqrt{(x-y\,|\,x-y)} = \sqrt{(x-y)^{\mathrm{T}}\,\mathrm{W}(x-y)}$ and these definitions are good for any x and y within the entire vector space.

Now for the vector space of random vectors, let W in the above be the inverse of the positive definite variance **P**, then a similar structure arises as $d(x,y) \triangleq \|x-y\| = \sqrt{(x-y\,|\,x-y)} = \sqrt{(x-y)^{\mathrm{T}}\,\mathrm{P}^{-1}(x-y)}$ but this expression is only appropriate for $d(\mu_1,\mu_2) \triangleq \|\mu_1-\mu_2\| = \sqrt{(\mu_1-\mu_2\,|\,\mu_1-\mu_2)} = \sqrt{(\mu_1-\mu_2)^{\mathrm{T}}\,\mathrm{P}^{-1}(\mu_1-\mu_2)}$ and is not a significant measure of distance for all other elements in the underlying vector space.

Brief theoretical motivation for using this statistic and its relationship to Chi-square is provided next. For an ideal mean

$$E[\,\hat{x}(t_i)\,] = x_{true}\,(\mathrm{t_i}) \tag{6}$$

and the *True Covariance*:

$$\mathbf{P}(\mathrm{t_i}) \triangleq E[(\,\hat{\mathbf{x}}\,(\mathrm{t_i}) - \mathbf{x}_{true}(\mathrm{t_i}))^{\mathrm{T}}(\,\hat{\mathbf{x}}\,(\mathrm{t_i}) - \mathbf{x}_{true}(\mathrm{t_i}))], \tag{7}$$

define the following coordinate transformation (that is time-varying in general but with the time index being suppressed here for clarity) by letting

$$\mathbf{y} = \mathbf{P}^{-1/2}(\hat{\mathbf{x}} - \mathbf{x}_{true}), \tag{8}$$

then

$$E[\mathbf{y}] = \mathbf{P}^{-1/2}E[(\hat{\mathbf{x}} - \mathbf{x}_{true})] = 0 \tag{9}$$

$$E[\mathbf{y}\mathbf{y}^{\mathbf{T}}] = \mathbf{P}^{-1/2}E[(\hat{\mathbf{x}} - \mathbf{x}_{true})(\hat{\mathbf{x}} - \mathbf{x}_{true})^{\mathbf{T}}]\mathbf{P}^{-1/2} = \mathbf{P}^{-1/2}\mathbf{P}\mathbf{P}^{-1/2} = \mathbf{I}_{n\ x\ n} \tag{10}$$

and if $(\,\hat{\mathbf{x}} - \mathbf{x}_{true})$ is Gaussianly distributed, then the linear transformation of Eq. 8 on a Gaussianly distributed random vector is again Gaussianly distributed, with mean zero, and unity variance. Moreover, the following scalar measure

$$\ell = y^T y = y_1^2 + y_2^2 + \cdots + y_n^2 \tag{11}$$

has statistics that are time-invariant and is central Chi-square with *n* degrees-of-freedom, hence

$$\text{Prob}\left[\left(\hat{\mathbf{x}}-\mathbf{x}_{true}\right)^{T}\mathbf{P}^{-1}\left(\hat{\mathbf{x}}-\mathbf{x}_{true}\right)\le K^{2}\right]=\text{Prob}\left[\mathbf{y}^{\mathrm{T}}\mathbf{y}\le K^{2}\right]=\text{Prob}\left[\ell\le K^{2}\right]=\text{Prob}\left[\chi_{n}^{2}\le K^{2}\right]. \tag{12}$$

Now consider the following specific scalar test statistic with time index no longer suppressed:

$$s\left(t_i\right)=\left[\hat{x}\left(t_i\right)-x_{true}\left(t_i\right)\right]^{\mathrm{T}}\ [\mathbf{P}\ (\mathrm{t_i})\ ]^{-1}\left[\hat{x}\left(t_i\right)-x_{true}\left(t_i\right)\right], \tag{13}$$

where each item above is a function of time t_i and so should correspondingly be plotted as a function of time. This test statistic **s** can be compared to 1, 2^2, 3^2, or whatever value desired, depending, respectively, on whether spec compliance dictates minimum acceptable performance within 1-, 2-, 3-, or whatever-sigma's away from the $\boldsymbol{x_{true}}$ for each particular candidate target tracking algorithm. Please see Fig. 8 in Sec. 7 for how we propose to view the time-history of **s** evaluations in comparison to K^2-sigma levels reached or passed through.

The same statistic can apparently also be applied to the results of an ensemble average when the <u>*covariance is treated as unknown*</u> by summarizing estimator performance as the following:

$$s\left(t_i\right)=\left(\bar{\hat{\mathbf{x}}}\left(t_i\right)-\mathbf{x}_{true}\left(t_i\right)\right)^{\mathrm{T}}\left[\mathbf{S^2}\left(t_i\right)\right]^{-1}\left(\bar{\hat{\mathbf{x}}}\left(t_i\right)-\mathbf{x}_{true}\left(t_i\right)\right) \tag{14}$$

where the unbiased sample mean and sample variance, respectively, are defined next

for the **state vector estimate** $\hat{\mathbf{x}}^{\mathbf{T}}\left(t_i\right)=\left[\hat{x}_1\left(t_i\right)\quad\hat{x}_2\left(t_i\right)\quad\hat{x}_3\left(t_i\right)\quad\hat{x}_4\left(t_i\right)\quad\hat{x}_5\left(t_i\right)\quad\hat{x}_6\left(t_i\right)\right]$**,** (15)

as the **Sample Mean vector**: $\bar{\hat{\mathbf{x}}}\left(t_i\right)=\dfrac{1}{\mathrm{N}}\sum_{\mathrm{j}=1}^{\mathrm{N}}\hat{\mathbf{x}}_{\mathrm{j}}\left(t_i\right)$**,** (16)

and the **Sample Covariance Matrix**: $\mathbf{S^2}\left(t_i\right)=\dfrac{1}{\mathrm{N}-1}\sum_{\mathrm{j}=1}^{\mathrm{N}}\left(\hat{\mathbf{x}}_{\mathbf{j}}\left(\mathbf{t_i}\right)-\bar{\mathbf{x}}\left(\mathbf{t_i}\right)\right)\left(\hat{\mathbf{x}}_{\mathbf{j}}\left(\mathbf{t_i}\right)-\bar{\mathbf{x}}\left(\mathbf{t_i}\right)\right)^{\mathbf{T}}$. (17)

Whether the test point $\hat{\mathbf{x}}$ is within the bounding ellipsoid of *K*-sigma's at each calculation time point t_i depends (as an if and only if statement) on whether the calculated statistic **s** of Eqs. 13 or 14 is less than or equal to K^2. If **s** exceeds K^2, then the test point is outside the *K*-sigma ellipsoid centered on x_{true}. One can plot the instantaneous **s** vs. time (see Fig. 8 below) and just look at the possible crossing trend relative to horizontal lines at ordinate 1, ordinate 2^2, or ordinate 3^2, representing, respectively, 1-, 2- or 3-sigma's away from the goal of $\boldsymbol{x_{true}}$ (where the ellipsoid is centered).

There is no need to perform this covariance fidelity test on the well-known standard expressions for the sampled mean and sampled variance (as, respectively, Eqs. 16 and 17) since they both should be obtained from the exact same Monte-Carlo data set in common and not be at odds with each other. However, since this Covariance Fidelity test is a likely candidate for abuse if Eq. 17 were used in Eq. 4 instead of the covariance obtained from within the particular tracking algorithm under scrutiny, an obligation exists to discuss the differences that underlie these two different situations.

When Eq. 4 is evaluated using the $\mathbf{P}(\mathrm{t_i})$ from the tracking algorithm, it is treated as an underlying covariance that is known beforehand and tested for consistency against the sampled errors accrued afterwards from Eq. 16. If Eq. 17 were used in Eq. 4 along with the errors from Eq. 16, both the mean and variance are now effectively being treated as unknown and are both inferred from the Monte-Carlo data. The statistics underlying the latter test are more challenging and more complex if both are treated as unknown and needing to be inferred than if just one is treated

as unknown to be inferred. This situation arises because of the distribution of the statistics themselves. Both Eqs. 16 and 17 are functions of the random vector $x(t_i)$ and as such are random themselves with particular distributions and exhibiting a particular resulting *realization*. Only if $x(t_i)$ is Gaussian distributed do Eqs. 16 and 17 have well-known distributions that are tractable and the distribution for Eq. 16 is simpler if the underlying population covariance $\mathbf{P}(t_i)$ is known beforehand [2, Table 8.5, pp. 232-233].

A corroborating reference for this topic of relating these statistics to the appropriate corresponding probabilities is [6]. In case the correct associated theoretical covariance matrix is not known in Eqs. 3 and 13, Ref. [6] shows on page 129 when to use the Hotelling T-distribution, as the multidimensional analog of Student's t-distribution (and also provides its underlying connection with the well-tabulated F-distribution of Snedecor for numerical evaluations).

We will return to the test statistic formulation of Eq. 14 above in Sec. 7 and explain why, although particularly appealing on the face of it, a more robust version should be used instead to better match the NMD application. The appropriate covariance that should be used (in the middle of Eq. 13) also varies with each Monte-Carlo trial as a result of on-line linearization (unlike the case for an exclusively linear Kalman filter, where it would be identically the same for each trial) and so must be included within the averaging statistics in order to be correct.

6. Specifying reasonable alternative confidence regions and revealing the interrelationships between the two 3-D (for position and velocity separately) and 6-D (both handled jointly)
For Gaussian random vectors, two components being uncorrelated means that they are indeed independent in the statistical sense. For more general random variables other than Gaussian, being uncorrelated is not enough to conclude that they are statistically independent. (There are many existing examples in statistics books that illustrate this aspect.) When random variables are independent and their joint and individual pdfs exist, their joint pdf factors into the product of the individual pdfs.

A multidimensional Gaussian probability density function (pdf) for a general n-dimensional random vector x is of the following form:

$$p_{\mathbf{x}}(\boldsymbol{\alpha}) = \frac{1}{(2\pi)^{n/2}|\mathbf{P}|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\boldsymbol{\alpha}-\bar{\mathbf{x}})^{\mathrm{T}}\,\mathbf{P}^{-1}(\boldsymbol{\alpha}-\bar{\mathbf{x}})\right\} \tag{18}$$

where $\mathbf{P}$ is the covariance and $\bar{x}$ is the mean. For a known specified (nxn) covariance matrix $\mathbf{P}$, the associated (nxn) normalized eigenvector matrix U is known to diagonalize $\mathbf{P}$ in the following way:

$$\mathbf{U}^{\mathbf{T}}\,\mathbf{P}\,\mathbf{U} = \mathrm{diag}(\lambda_1,\lambda_2,\ldots,\lambda_n) \tag{19}$$

and also

$$\mathbf{U}^{\mathbf{T}}\,\mathbf{P}^{-1}\,\mathbf{U} = \mathrm{diag}\left(\frac{1}{\lambda_1},\frac{1}{\lambda_2},\ldots,\frac{1}{\lambda_n}\right) \tag{20}$$

Now when the following linear transformation is applied to the original random vector x (using the above described normalized eigenvector as a rotation of the underlying coordinate axes) as:

$$\mathbf{w} = \mathrm{U}\mathbf{x} \tag{21}$$

the result is

$$p_{\mathbf{w}}(\boldsymbol{\beta}) = \frac{1}{(2\pi)^{n/2}\left|\mathbf{U}^{\mathrm{T}}\mathbf{P}\mathbf{U}\right|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(\mathbf{U}\boldsymbol{\alpha}-\mathbf{U}\bar{\mathbf{x}})^{\mathrm{T}}\left[\mathbf{U}^{\mathrm{T}}\mathbf{P}\,\mathbf{U}\right]^{-1}(\mathbf{U}\boldsymbol{\alpha}-\mathbf{U}\bar{\mathbf{x}})\right\}$$

$$= \frac{1}{(2\pi)^{n/2}\left|\mathbf{U}^{\mathrm{T}}\mathbf{P}\mathbf{U}\right|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(\boldsymbol{\beta}-\bar{\mathbf{w}})^{\mathrm{T}}\left[\mathbf{U}^{\mathrm{T}}\mathbf{P}\,\mathbf{U}\right]^{-1}(\boldsymbol{\beta}-\bar{\mathbf{w}})\right\}$$

$$= \frac{1}{(2\pi)^{n/2}\left|\begin{bmatrix}\lambda_1 & 0 & 0 & \cdots & 0\\ 0 & \lambda_2 & 0 & \cdots & 0\\ 0 & 0 & \ddots & \ddots & \vdots\\ \vdots & \vdots & \ddots & \ddots & 0\\ 0 & 0 & 0 & \cdots & \lambda_n\end{bmatrix}\right|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(\boldsymbol{\beta}-\bar{\mathbf{w}})^{\mathrm{T}}\begin{bmatrix}\lambda_1 & 0 & 0 & \cdots & 0\\ 0 & \lambda_2 & 0 & \cdots & 0\\ 0 & 0 & \ddots & \ddots & \vdots\\ \vdots & \vdots & \ddots & \ddots & 0\\ 0 & 0 & 0 & \cdots & \lambda_n\end{bmatrix}^{-1}(\boldsymbol{\beta}-\bar{\mathbf{w}})\right\} \tag{22}$$

$$= \frac{1}{(2\pi)^{n/2}(\lambda_1\lambda_2\cdots\lambda_n)^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(\boldsymbol{\beta}-\bar{\mathbf{w}})^{\mathrm{T}}\begin{bmatrix}\frac{1}{\lambda_1} & 0 & 0 & \cdots & 0\\ 0 & \frac{1}{\lambda_2} & 0 & \cdots & 0\\ 0 & 0 & \ddots & \ddots & \vdots\\ \vdots & \vdots & \ddots & \ddots & 0\\ 0 & 0 & 0 & \cdots & \frac{1}{\lambda_n}\end{bmatrix}(\boldsymbol{\beta}-\bar{\mathbf{w}})\right\}$$

$$= \frac{1}{(2\pi)^{n/2}(\lambda_1\lambda_2\cdots\lambda_n)^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\left[\frac{(\beta_i-\bar{w}_i)^2}{\lambda_i}\right]\right\} = \Pi_{i=1}^{n}\frac{1}{(2\pi)^{1/2}(\lambda_i)^{\frac{1}{2}}}\exp\left\{-\frac{(\beta_i-\bar{w}_i)^2}{2\lambda_i}\right\}$$

The final expression is the product of individual scalar Gaussian pdfs for each of the coordinate directions defined by the eigenvector matrix and just represents a rotation of the original coordinate axes. The resulting axes are still orthogonal in n-dimensional space. Since the pdfs are all Gaussian, being uncorrelated, they are now independent in the resulting coordinate system since their joint pdf can be represented as the product of individual pdfs.

Within NMD, there is a precedent in the use of two different collections of 3 states each (as 3 position states and 3 velocity states) when examining and evaluating mission performance measures but there is nothing theoretically wrong with handling all 6 states together as a joint consideration. **Indeed, when only the two 3 state collections are utilized (as, say, 3 position states and 3 velocity states), cross-correlations between these groupings is being ignored in the evaluations that is present in the covariance from the tracking algorithm.**

The equation of a K-sigma tilted (non-diagonal covariance **P**) ellipsoid in 6-dimensions about the point $\left[\bar{x}_1, \bar{x}_2, \bar{x}_3, \ldots, \bar{x}_6\right]^{\mathrm{T}}$ is:

$$K^2 = \left[x_1-\bar{x}_1, x_2-\bar{x}_2, x_3-\bar{x}_3, \ldots, x_6-\bar{x}_6\right]\mathrm{P}^{-1}\begin{bmatrix}x_1-\bar{x}_1\\ x_2-\bar{x}_2\\ x_3-\bar{x}_3\\ \vdots\\ x_6-\bar{x}_6\end{bmatrix} \tag{23}$$

and the corresponding n-dimensional Gaussian pdf is of the following form:

$$p_x(\mathbf{x}) = \frac{1}{(2\pi)^3 |\mathbf{P}|^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2} \left[x_1 - \overline{x}_1, x_2 - \overline{x}_2, x_3 - \overline{x}_3, \ldots, x_6 - \overline{x}_6 \right] \begin{bmatrix} \mathbf{P}_1 & \mathbf{P}_{12} \\ \mathbf{P}_{12}^{\mathrm{T}} & \mathbf{P}_2 \end{bmatrix}^{-1} \begin{bmatrix} x_1 - \overline{x}_1 \\ x_2 - \overline{x}_2 \\ x_3 - \overline{x}_3 \\ \vdots \\ x_6 - \overline{x}_6 \end{bmatrix} \right\} \tag{24}$$

with its corresponding countour as depictured in Fig. 5.

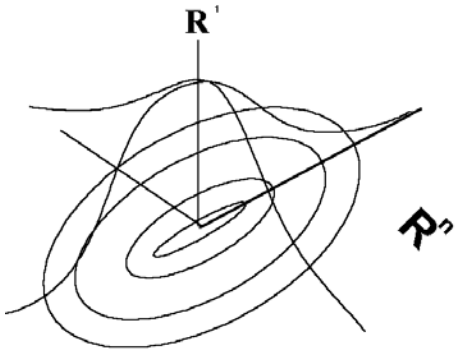

Figure 5: *The pdf of an 6-dimensional Gaussian with non-diagonal covariance **P** (n=6)*

Notice that the pdf of the 6-dimensional Gaussian is a scalar function of 6-variables. The appropriate confidence regions of constant pdf align perfectly with the underlying ellipsoidal contours in $\Re^6$ defined by the structure of the covariance matrix **P**.

For the following three different confidence region interior and boundaries in $\Re^6$:

$$K'^2 \geq \left[x_1 - \overline{x}_1, x_2 - \overline{x}_2, x_3 - \overline{x}_3, \ldots, x_6 - \overline{x}_6 \right] \begin{bmatrix} \mathbf{P_1} & \mathbf{P_{12}} \\ \mathbf{P_{12}^T} & \mathbf{P_2} \end{bmatrix}^{-1} \begin{bmatrix} x_1 - \overline{x}_1 \\ x_2 - \overline{x}_2 \\ x_3 - \overline{x}_3 \\ \vdots \\ x_6 - \overline{x}_6 \end{bmatrix}, \tag{25}$$

$$K'^2 \geq \left[x_1 - \overline{x}_1, x_2 - \overline{x}_2, x_3 - \overline{x}_3 \right] \left[\mathbf{P}_1 \right]^{-1} \begin{bmatrix} x_1 - \overline{x}_1 \\ x_2 - \overline{x}_2 \\ x_3 - \overline{x}_3 \end{bmatrix}, \tag{26}$$

$$K'^2 \geq \left[x_4 - \overline{x}_4, x_5 - \overline{x}_5, x_6 - \overline{x}_6 \right] \left[\mathbf{P}_2 \right]^{-1} \begin{bmatrix} x_4 - \overline{x}_4 \\ x_5 - \overline{x}_5 \\ x_6 - \overline{x}_6 \end{bmatrix}, \tag{27}$$

where, in the above, the score value K'^2, the vector entries $\left[\overline{x}_1, \overline{x}_2, \overline{x}_3, \ldots, \overline{x}_6 \right]^{\mathrm{T}}$, and $\begin{bmatrix} \mathbf{P}_1 & \mathbf{P}_{12} \\ \mathbf{P}_{12}^{\mathrm{T}} & \mathbf{P}_2 \end{bmatrix}$ are all <u>particular known fixed constants</u> but the values of x_1, x_2, …, x_6 are variables that define the specific unique confidence regions for that particular score value (K'^2) and center located at $\left[\overline{x}_1, \overline{x}_2, \overline{x}_3, \ldots, \overline{x}_6 \right]^{\mathrm{T}}$.

Recall that in order to prove that one set, say, **A** is a subset of another set **B** (as $\mathbf{A} \subset \mathbf{B}$) it suffices to show that for any arbitrary element of **A** it is also an element of **B** or, equivalently, that any element of $\mathbf{B}^{\mathbf{c}}$ implies that it is also a member of $\mathbf{A}^{\mathbf{c}}$ (since $\mathbf{B}^{\mathbf{c}} \subset \mathbf{A}^{\mathbf{c}} \Leftrightarrow \mathbf{A} \subset \mathbf{B}$), where superscript **c** denotes the set's complement. In seeking to establish a rigorous connection or relationship, please consider the following proof that uses the following four intermediate Lemmas as stepping-stones. We relegate the proofs of these four Lemmas to Appendix B to avoid obscuring our primary thrust here in the main body.

We first need a standard result from the partitioning of a matrix and its inverse as

Lemma 1:
$$\begin{bmatrix} \mathbf{P}_1 & \mathbf{P}_{12} \\ \mathbf{P}_{12}^{\mathrm{T}} & \mathbf{P}_2 \end{bmatrix} \begin{bmatrix} \left(\mathbf{P}_1 - \mathbf{P}_{12}\mathbf{P}_2^{-1}\mathbf{P}_{12}^{\mathrm{T}}\right)^{-1} & -\mathbf{P}_1^{-1}\mathbf{P}_{12}\left(\mathbf{P}_2 - \mathbf{P}_{12}^{\mathrm{T}}\mathbf{P}_1^{-1}\mathbf{P}_{12}\right)^{-1} \\ -\mathbf{P}_2^{-1}\mathbf{P}_{12}^{\mathrm{T}}\left(\mathbf{P}_1 - \mathbf{P}_{12}\mathbf{P}_2^{-1}\mathbf{P}_{12}^{\mathrm{T}}\right)^{-1} & \left(\mathbf{P}_2 - \mathbf{P}_{12}^{\mathrm{T}}\mathbf{P}_1^{-1}\mathbf{P}_{12}\right)^{-1} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_3 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_3 \end{bmatrix}, \quad (28)$$

where the matrix inverse is unique and the block diagonal inverses depicted above also exist by virtue of intermediate steps encountered in the proofs of Lemmas 2 to 4 below. (A proof is also in [7, p. 28]. However, for the reader's benefit, we still include a simple abbreviated proof here in Appendix B to make this document more self-contained.)

Lemma 2: The partitioned square matrix $\begin{bmatrix} \mathbf{P}_1 & \mathbf{P}_{12} \\ \mathbf{P}_{12}^{\mathrm{T}} & \mathbf{P}_2 \end{bmatrix}$ is positive definite (and therefore invertible) if and only if all four of the following are true:

$$\left(\mathbf{P}_1 - \mathbf{P}_{12}\mathbf{P}_2^{-1}\mathbf{P}_{12}^{\mathrm{T}}\right)^{-1} > \mathbf{0}, \left(\mathbf{P}_2 - \mathbf{P}_{12}^{\mathrm{T}}\mathbf{P}_1^{-1}\mathbf{P}_{12}\right)^{-1} > \mathbf{0}, \mathbf{P}_1 > \mathbf{0}, \mathbf{P}_2 > \mathbf{0}, \quad (29)$$

where the "greater than" symbol appearing here is in the sense of matrix positive definiteness.

Lemma 3: $\left(\mathbf{P}_1 - \mathbf{P}_{12}\mathbf{P}_2^{-1}\mathbf{P}_{12}^{\mathrm{T}}\right)^{-1} \geq \mathbf{P}_1^{-1}$. (30)

Lemma 4: $\left(\mathbf{P}_2 - \mathbf{P}_{12}^{\mathrm{T}}\mathbf{P}_1^{-1}\mathbf{P}_{12}\right)^{-1} \geq \mathbf{P}_2^{-1}$. (31)

Again, the simple proofs of these four Lemmas have been relegated to Appendix B.

Using these Lemmas, we will now establish how the above three confidence regions of Eqs. 25-27 are related as subsets and offset subspaces. First recall that for confidence regions, the interior and boundary is where the right hand side (RHS) quadratic form is less than or equal to ($K'^2 \geq$) as in Eqs. 25-27 and that the *exterior* of the confidence region is correspondingly where the RHS quadratic form is switched to be greater than the specific score value ($K'^2 <$) as given by:

$$K'^2 < \begin{bmatrix} x_1 - \overline{x}_1, x_2 - \overline{x}_2, x_3 - \overline{x}_3, \ldots, x_6 - \overline{x}_6 \end{bmatrix} \begin{bmatrix} \mathbf{P}_1 & \mathbf{P}_{12} \\ \mathbf{P}_{12}^{\mathrm{T}} & \mathbf{P}_2 \end{bmatrix}^{-1} \begin{bmatrix} x_1 - \overline{x}_1 \\ x_2 - \overline{x}_2 \\ x_3 - \overline{x}_3 \\ \vdots \\ x_6 - \overline{x}_6 \end{bmatrix}, \quad (32)$$

$$K'^2 < \begin{bmatrix} x_1 - \overline{x}_1, x_2 - \overline{x}_2, x_3 - \overline{x}_3 \end{bmatrix} \begin{bmatrix} \mathbf{P}_1 \end{bmatrix}^{-1} \begin{bmatrix} x_1 - \overline{x}_1 \\ x_2 - \overline{x}_2 \\ x_3 - \overline{x}_3 \end{bmatrix}, \quad (33)$$

$$K'^2 < \left[x_4 - \overline{x}_4, x_5 - \overline{x}_5, x_6 - \overline{x}_6\right] \left[\mathbf{P}_2\right]^{-1} \begin{bmatrix} x_4 - \overline{x}_4 \\ x_5 - \overline{x}_5 \\ x_6 - \overline{x}_6 \end{bmatrix}, \tag{34}$$

Case 1: Notice that for arbitrary $\left[x_1', x_2', x_3'\right]^{\mathrm{T}} \in \Re^3 - \left\{\left[\overline{x}_1, \overline{x}_2, \overline{x}_3\right]^{\mathrm{T}}\right\}$ that satisfies Eq. 33 by virtue of being in the 3-dimensional offset subspace (i.e., manifold) with coordinates $x_4' = \overline{x}_4, x_5' = \overline{x}_5$, and $x_6' = \overline{x}_6$ as fixed, specified constants: $\forall \left[x_1', x_2', x_3'\right]^{\mathrm{T}} \in \mathbf{F}_1$, where

$$\mathbf{F}_1 \triangleq \left\{ \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \in \Re^3 - \left\{ \begin{bmatrix} \overline{x}_1 \\ \overline{x}_2 \\ \overline{x}_3 \end{bmatrix} \right\} \middle| K'^2 < \begin{bmatrix} x_1 - \overline{x}_1 \\ x_2 - \overline{x}_2 \\ x_3 - \overline{x}_3 \end{bmatrix}^{\mathrm{T}} \left[\mathbf{P}_1\right]^{-1} \begin{bmatrix} x_1 - \overline{x}_1 \\ x_2 - \overline{x}_2 \\ x_3 - \overline{x}_3 \end{bmatrix} \right\} \Rightarrow K'^2 < \begin{bmatrix} x_1' - \overline{x}_1 \\ x_2' - \overline{x}_2 \\ x_3' - \overline{x}_3 \end{bmatrix}^{\mathrm{T}} \left[\mathbf{P}_1\right]^{-1} \begin{bmatrix} x_1' - \overline{x}_1 \\ x_2' - \overline{x}_2 \\ x_3' - \overline{x}_3 \end{bmatrix}, \tag{35}$$

also satisfies Eq. 32 with $\left[x_1', x_2', x_3', \overline{x}_4, \overline{x}_5, \overline{x}_6\right]^{\mathrm{T}} \in \Re^6 - \left\{\left[\overline{x}_1, \overline{x}_2, \overline{x}_3, \ldots, \overline{x}_6\right]^{\mathrm{T}}\right\}$ since

$$K'^2 < \left[x_1' - \overline{x}_1, \quad x_2' - \overline{x}_2, \quad x_3' - \overline{x}_3\right] \mathbf{P_1^{-1}} \begin{bmatrix} x_1' - \overline{x}_1 \\ x_2' - \overline{x}_2 \\ x_3' - \overline{x}_3 \end{bmatrix} \leq \left[x_1' - \overline{x}_1, \quad x_2' - \overline{x}_2, \quad x_3' - \overline{x}_3\right] \left(\mathbf{P_1 - P_{12} P_2^{-1} P_{12}^{\mathrm{T}}}\right)^{-1} \begin{bmatrix} x_1' - \overline{x}_1 \\ x_2' - \overline{x}_2 \\ x_3' - \overline{x}_3 \end{bmatrix}$$

$$= \begin{bmatrix} x_1' - \overline{x}_1 \\ x_2' - \overline{x}_2 \\ x_3' - \overline{x}_3 \\ 0 \\ 0 \\ 0 \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} \left(\mathbf{P_1 - P_{12} P_2^{-1} P_{12}^{\mathrm{T}}}\right)^{-1} & -\mathbf{P_1^{-1} P_{12}} \left(\mathbf{P_2 - P_{12}^{\mathrm{T}} P_1^{-1} P_{12}}\right)^{-1} \\ -\mathbf{P_2^{-1} P_{12}^{\mathrm{T}}} \left(\mathbf{P_1 - P_{12} P_2^{-1} P_{12}^{\mathrm{T}}}\right)^{-1} & \left(\mathbf{P_2 - P_{12}^{\mathrm{T}} P_1^{-1} P_{12}}\right)^{-1} \end{bmatrix} \begin{bmatrix} x_1' - \overline{x}_1 \\ x_2' - \overline{x}_2 \\ x_3' - \overline{x}_3 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} x_1' - \overline{x}_1 \\ x_2' - \overline{x}_2 \\ x_3' - \overline{x}_3 \\ 0 \\ 0 \\ 0 \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} \mathbf{P}_1 & \mathbf{P}_{12} \\ \mathbf{P}_{12}^{\mathrm{T}} & \mathbf{P}_2 \end{bmatrix}^{-1} \begin{bmatrix} x_1' - \overline{x}_1 \\ x_2' - \overline{x}_2 \\ x_3' - \overline{x}_3 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{36}$$

Please notice in the above that the result of Lemma 3 was used in establishing Eq. 36 and that

$$\mathbf{A}_1^{\mathrm{C}} \triangleq \left\{ \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \overline{x}_4 \\ \overline{x}_5 \\ \overline{x}_6 \end{bmatrix} \in \Re^6 - \left\{ \begin{bmatrix} \overline{x}_1 \\ \overline{x}_2 \\ \overline{x}_3 \\ \overline{x}_4 \\ \overline{x}_5 \\ \overline{x}_6 \end{bmatrix} \right\} \middle| K'^2 < \begin{bmatrix} x_1 - \overline{x}_1 \\ x_2 - \overline{x}_2 \\ x_3 - \overline{x}_3 \\ 0 \\ 0 \\ 0 \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} \mathbf{P}_1 & \mathbf{P}_{12} \\ \mathbf{P}_{12}^{\mathrm{T}} & \mathbf{P}_2 \end{bmatrix}^{-1} \begin{bmatrix} x_1 - \overline{x}_1 \\ x_2 - \overline{x}_2 \\ x_3 - \overline{x}_3 \\ 0 \\ 0 \\ 0 \end{bmatrix} \right\}, \Rightarrow \mathbf{F}_1^{\mathrm{C}} \subset \mathbf{A}_1^{\mathrm{C}} \Leftrightarrow \mathbf{A}_1 \subset \mathbf{F}_1, \tag{37}$$

and $\mathbf{D^c} \triangleq \left\{ \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} \in \Re^6 - \left\{ \begin{bmatrix} \overline{x}_1 \\ \overline{x}_2 \\ \overline{x}_3 \\ \overline{x}_4 \\ \overline{x}_5 \\ \overline{x}_6 \end{bmatrix} \right\} \middle| K'^2 < \begin{bmatrix} x_1 - \overline{x}_1 \\ x_2 - \overline{x}_2 \\ x_3 - \overline{x}_3 \\ x_4 - \overline{x}_4 \\ x_5 - \overline{x}_5 \\ x_6 - \overline{x}_6 \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} \mathbf{P}_1 & \mathbf{P}_{12} \\ \mathbf{P}_{12}^{\mathrm{T}} & \mathbf{P}_2 \end{bmatrix}^{-1} \begin{bmatrix} x_1 - \overline{x}_1 \\ x_2 - \overline{x}_2 \\ x_3 - \overline{x}_3 \\ x_4 - \overline{x}_4 \\ x_5 - \overline{x}_5 \\ x_6 - \overline{x}_6 \end{bmatrix} \right\} \Rightarrow \mathbf{A_1^C} \subset \mathbf{D^C} \Leftrightarrow \mathbf{D} \subset \mathbf{A_1}$,(38)

since $\mathbf{D^C}$ is less restricted in its domain than is $\mathbf{A_1}^{\mathbf{C}}$ by allowing it to take on any values in $\Re^6 - \left\{ \left[\overline{x}_1, \overline{x}_2, \overline{x}_3, \overline{x}_4, \overline{x}_5, \overline{x}_6 \right]^{\mathrm{T}} \right\}$.

Case 2: Notice that for arbitrary $\left[x_4', x_5', x_6' \right]^{\mathrm{T}} \in \Re^3 - \left\{ \left[\overline{x}_4, \overline{x}_5, \overline{x}_6 \right]^{\mathrm{T}} \right\}$ that satisfies Eq. 34 by virtue of being in the 3-dimensional offset subspace (i.e., manifold) with coordinates $x_1' = \overline{x}_1, x_2' = \overline{x}_2$, and $x_3' = \overline{x}_3$ as fixed, specified constants: $\forall \left[x_4', x_5', x_6' \right]^{\mathrm{T}} \in \mathbf{F_2^C}$

$$\mathbf{F_2^C} \triangleq \left\{ \begin{bmatrix} x_4 \\ x_5 \\ x_6 \end{bmatrix} \in \Re^3 - \left\{ \begin{bmatrix} \overline{x}_4 \\ \overline{x}_5 \\ \overline{x}_6 \end{bmatrix} \right\} \middle| K'^2 < \begin{bmatrix} x_4 - \overline{x}_4 \\ x_5 - \overline{x}_5 \\ x_6 - \overline{x}_6 \end{bmatrix}^{\mathrm{T}} \left[\mathbf{P_2} \right]^{-1} \begin{bmatrix} x_4 - \overline{x}_4 \\ x_5 - \overline{x}_5 \\ x_6 - \overline{x}_6 \end{bmatrix} \right\} \Rightarrow K'^2 < \begin{bmatrix} x_4' - \overline{x}_4 \\ x_5' - \overline{x}_5 \\ x_6' - \overline{x}_6 \end{bmatrix}^{\mathrm{T}} \left[\mathbf{P_2} \right]^{-1} \begin{bmatrix} x_4' - \overline{x}_4 \\ x_5' - \overline{x}_5 \\ x_6' - \overline{x}_6 \end{bmatrix}, \quad (39)$$

also satisfies Eq. 32 with $\left[\overline{x}_1, \overline{x}_2, \overline{x}_3, x_4', x_5', x_6' \right]^{\mathrm{T}} \in \Re^6 - \left\{ \left[\overline{x}_1, \overline{x}_2, \overline{x}_3, \overline{x}_4, \overline{x}_5, \overline{x}_6 \right]^{\mathrm{T}} \right\}$ since

$$K'^2 < \left[x_4' - \overline{x}_4, \quad x_5' - \overline{x}_5, \quad x_6' - \overline{x}_6 \right] \mathbf{P_2^{-1}} \begin{bmatrix} x_4' - \overline{x}_4 \\ x_5' - \overline{x}_5 \\ x_6' - \overline{x}_6 \end{bmatrix} \leq \left[x_4' - \overline{x}_4, \quad x_5' - \overline{x}_5, \quad x_6' - \overline{x}_6 \right] \left(\mathbf{P_2 - P_{12}^T P_1^{-1} P_{12}} \right)^{-1} \begin{bmatrix} x_4' - \overline{x}_4 \\ x_5' - \overline{x}_5 \\ x_6' - \overline{x}_6 \end{bmatrix}$$

$$= \begin{bmatrix} 0 \\ 0 \\ 0 \\ x_4' - \overline{x}_4 \\ x_5' - \overline{x}_5 \\ x_6' - \overline{x}_6 \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} \left(\mathbf{P_1 - P_{12} P_2^{-1} P_{12}^T} \right)^{-1} & -\mathbf{P_1^{-1} P_{12}} \left(\mathbf{P_2 - P_{12}^T P_1^{-1} P_{12}} \right)^{-1} \\ -\mathbf{P_2^{-1} P_{12}^T} \left(\mathbf{P_1 - P_{12} P_2^{-1} P_{12}^T} \right)^{-1} & \left(\mathbf{P_2 - P_{12}^T P_1^{-1} P_{12}} \right)^{-1} \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \\ x_4' - \overline{x}_4 \\ x_5' - \overline{x}_5 \\ x_6' - \overline{x}_6 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ x_4' - \overline{x}_4 \\ x_5' - \overline{x}_5 \\ x_6' - \overline{x}_6 \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} \mathbf{P}_1 & \mathbf{P}_{12} \\ \mathbf{P}_{12}^{\mathrm{T}} & \mathbf{P}_2 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 0 \\ 0 \\ x_4' - \overline{x}_4 \\ x_5' - \overline{x}_5 \\ x_6' - \overline{x}_6 \end{bmatrix}. \quad (40)$$

Please notice in the above that the result of Lemma 4 was used in establishing Eq. 40 and that

$$\mathbf{A_2^C} \triangleq \left\{ \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} \in \Re^6 - \left\{ \begin{bmatrix} \overline{x}_1 \\ \overline{x}_2 \\ \overline{x}_3 \\ \overline{x}_4 \\ \overline{x}_5 \\ \overline{x}_6 \end{bmatrix} \right\} \middle| K'^2 < \begin{bmatrix} 0 \\ 0 \\ 0 \\ x_4 - \overline{x}_4 \\ x_5 - \overline{x}_5 \\ x_6 - \overline{x}_6 \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} \mathbf{P}_1 & \mathbf{P}_{12} \\ \mathbf{P}_{12}^{\mathrm{T}} & \mathbf{P}_2 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 0 \\ 0 \\ x_4 - \overline{x}_4 \\ x_5 - \overline{x}_5 \\ x_6 - \overline{x}_6 \end{bmatrix} \right\} \Rightarrow \mathbf{F_2^C} \subset \mathbf{A_2^C} \Leftrightarrow \mathbf{A_2} \subset \mathbf{F_2}, \quad (41)$$

and

$$\mathbf{D^c} \triangleq \left\{ \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} \in \Re^6 - \left\{ \begin{bmatrix} \overline{x}_1 \\ \overline{x}_2 \\ \overline{x}_3 \\ \overline{x}_4 \\ \overline{x}_5 \\ \overline{x}_6 \end{bmatrix} \right\} \middle| K'^2 < \begin{bmatrix} x_1 - \overline{x}_1 \\ x_2 - \overline{x}_2 \\ x_3 - \overline{x}_3 \\ x_4 - \overline{x}_4 \\ x_5 - \overline{x}_5 \\ x_6 - \overline{x}_6 \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} \mathbf{P}_1 & \mathbf{P}_{12} \\ \mathbf{P}_{12}^{\mathrm{T}} & \mathbf{P}_2 \end{bmatrix}^{-1} \begin{bmatrix} x_1 - \overline{x}_1 \\ x_2 - \overline{x}_2 \\ x_3 - \overline{x}_3 \\ x_4 - \overline{x}_4 \\ x_5 - \overline{x}_5 \\ x_6 - \overline{x}_6 \end{bmatrix} \right\} \Rightarrow \mathbf{A_2^C} \subset \mathbf{D^C} \Leftrightarrow \mathbf{D} \subset \mathbf{A_2}, \qquad (42)$$

since $\mathbf{D^C}$ is less restricted in its domain than is $\mathbf{A_2}^{\mathbf{C}}$ by allowing it to take on any values in $\Re^6 - \left\{ \left[\overline{x}_1, \overline{x}_2, \overline{x}_3, \overline{x}_4, \overline{x}_5, \overline{x}_6 \right]^{\mathrm{T}} \right\}$.

Now from both Cases 1 and 2 above, $\mathbf{F_2^C} \subset \mathbf{A_2^C} \subset \mathbf{D^C}$ and $\mathbf{F_1^C} \subset \mathbf{A_1^C} \subset \mathbf{D^C}$ (where $\mathbf{D^C}$ is the same identical set in Cases 1 and 2) so that

$$\mathbf{F_2^C} \cup \mathbf{F_1^C} \subset \mathbf{A_2^C} \cup \mathbf{A_1^C} \subset \left(\mathbf{D^C} \cup \mathbf{D^C} \right) = \mathbf{D^C}, \qquad (43)$$

and by de Morgan's laws yields

$$\mathbf{F_2} \cap \mathbf{F_1} \supset \mathbf{A_2} \cap \mathbf{A_1} = \left(\mathbf{A_2^C} \cup \mathbf{A_1^C} \right)^{\mathbf{c}} \supset \left(\mathbf{D^C} \right)^{\mathbf{c}} = \mathbf{D}, \qquad (44)$$

and all the more

$$\mathbf{F_2} \supset \left(\mathbf{F_2} \cap \mathbf{F_1} \right) \supset \left(\mathbf{A_2} \cap \mathbf{A_1} \right) \supset \mathbf{D}, \qquad (45)$$

and

$$\mathbf{F_1} \supset \left(\mathbf{F_2} \cap \mathbf{F_1} \right) \supset \left(\mathbf{A_2} \cap \mathbf{A_1} \right) \supset \mathbf{D}. \qquad (46)$$

Q.E.D.

Now please consider Table 1 below.

Probability of being within an uncertainty ellipsoid (for a Gaussian Distribution) both centered on the mean for:

Dimension: N	Probability of 1-sigma containment	Probability of 3-sigma containment	Probability of 4-sigma containment
1	0.68269	0.9973	0.9999
3	0.19876	0.9707	0.9989
6	0.01439	0.8264	0.9863

Table 1: Multivariate Gaussian containment probabilities within K-sigma

Notice from Table 1 that the value of K-sigma to use when considering the 3 position states and the 3 velocity states together within a single joint 6-dimensional confidence region in order to have 0.99 containment in (6-D) is 4-sigma. The same containment probabilities are provided for 3-D collections with slightly greater than 3-sigma ellipsoids. Use of 4-sigma for a 3-dimensional confidence region provides 0.9989 probability of containment. (Table 1 was prepared using Chi-square statistics for 1-, 3-, and 6-degrees-of-freedom from [8, Table 26.7] and crosschecked using the computer program **chi2cdf.m** from The Math Work's *Statistics Toolbox* for MatLab, Ver. 6.5, Release 13. The row entries for N=1 allows a useful cross-check with what is usually already familiar to most technologist for the Gaussian scalar case.)

For $K'^2 = (4)^2$ or 4-sigma in Table 1, and for the following set nestings established above:

$$\mathbf{F}_1 \supset (\mathbf{F}_2 \cap \mathbf{F}_1) \supset (\mathbf{A}_2 \cap \mathbf{A}_1) \supset \mathbf{D} \Rightarrow \underbrace{\text{Prob}(\mathbf{F}_1)}_{0.9989} \geq \text{Prob}(\mathbf{F}_2 \cap \mathbf{F}_1) \geq \text{Prob}(\mathbf{A}_1 \cap \mathbf{A}_2) \geq \underbrace{\text{Prob}(\mathbf{D})}_{0.9863} \quad (47)$$

as expected; and similarly by parallelism for

$$\mathbf{F}_2 \supset (\mathbf{F}_2 \cap \mathbf{F}_1) \supset (\mathbf{A}_2 \cap \mathbf{A}_1) \supset \mathbf{D} \Rightarrow \underbrace{\text{Prob}(\mathbf{F}_2)}_{0.9989} \geq \text{Prob}(\mathbf{F}_2 \cap \mathbf{F}_1) \geq \text{Prob}(\mathbf{A}_1 \cap \mathbf{A}_2) \geq \underbrace{\text{Prob}(\mathbf{D})}_{0.9863} \quad (48)$$

Notice that the probability that both 3-dimensional confidence regions are satisfied simultaneously is upper and lower bounded as a bracketed quantity (both from above and below) in Eqs. 47 and 48 and that this probability is less than any one of the 3-dimensional confidence regions being satisfied when considered all by itself.

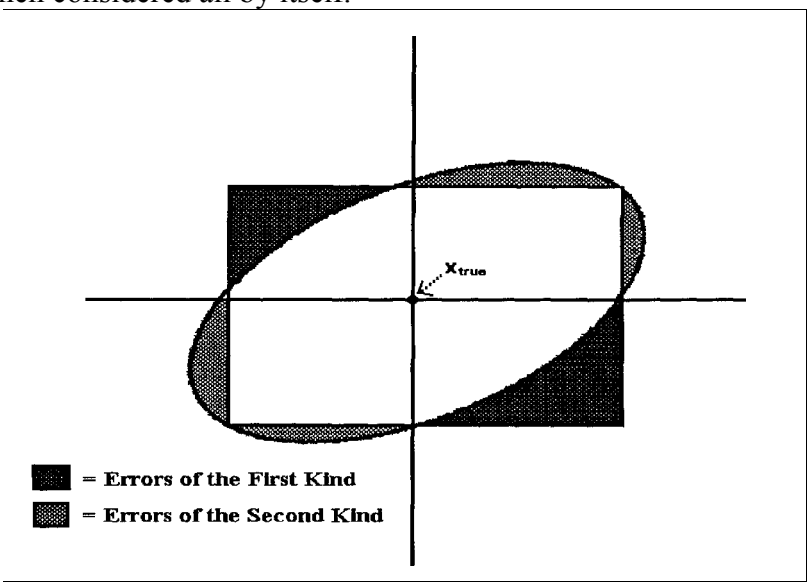

Figure 6: *Utility of decision provided by one 2-D Ellipsoid Inclusion Test versus conclusion of decisions provided by Two independent scalar inclusion tests administered component-by-component*

While the gist of the above discussion has been to show that both 3-dimesional confidence regions are nested subsets of what would otherwise be handled in a full consideration of the 6-dimensional confidence region, and moreover, each 3-dimensional confidence region just views an offset subspace of the full confidence region, the main idea that we are warning about here is that this myopic view of concentrating only on the two larger confidence regions misses valuable information there (that is, in effect, ignored). While it is impossible to drawn pictures of the 6-dimensional space and pictures of the 3 dimensional subspaces and represent that in a 2-dimensional flat picture here, it is still rather easy to conceptualize it using the analogy of Figs. 5 and 6, moreover, it doesn't elude the rigorous but tedious simple proof offered here.

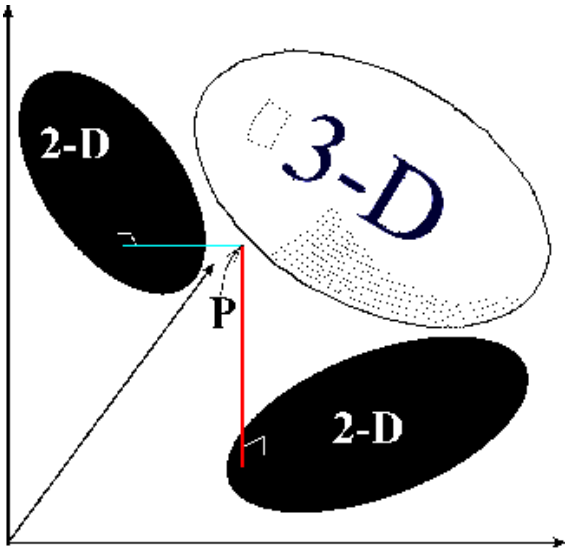

Figure 7: *Point **P** is contained in both lower dimensional 2-D projections yet is not actually contained in higher dimensional 3-D ellipsoid*

The actual underlying situation depicted in Fig. 6 is an ellipse in 2-space and conclusions drawn on two simplified tests performed in the 2 constituent scalar subspaces incur increased errors of the first and second kind of "incorrectly thinking there is membership when there isn't" and "incorrectly thinking there is no membership when there is". Fig. 7 depicts the wrong conclusion about containment being arrived at for point P (outside the 3-D ellipsoid) although both 2-D elliptical projections appear to contain it. This figure uses perspective in an attempt to adequately portray an ellipsoid in 3-space and its elliptical projections in a plane not yielding satisfactory decisions for the same reason of not seeing the whole story even when the 3-space ellipsoid is projected onto two separate planes (for a total of 4-dimensions) that are still not availed of the true full 3-dimensional view. A similar example can be concocted with 3 misleading 2-D projections, all orthogonal. (Recall in the standard engineering drawing class that three orthogonal planar perspectives were usually needed to remove the ambiguity in most 3-D objects [for a total of 6-D] and that still there are famous objects for which hidden lines still caused ambiguity to arise almost as if it were from an M. C. Escher painting).

7. A More Robust Test Statistic Recommended for use to Avoid Ill-Conditioning

The appropriate generalization of the statistic introduced in Sec. 4 as Eq. 13, applicable to the results of an ensemble average, representing NMD tracking algorithm estimator performance is:

$$s(t_i) = \frac{1}{N}\sum_{j=1}^{N}\left\{\left[\hat{\mathbf{x}}_j(\mathbf{t}_i) - \mathbf{x}_{true}(\mathbf{t}_i)\right]^T \left[\mathbf{P}_j(\mathbf{t}_i)\right]^{-1} \left[\hat{\mathbf{x}}_j(\mathbf{t}_i) - \mathbf{x}_{true}(\mathbf{t}_i)\right]\right\}, \tag{49}$$

where all the entries in the ensemble average are scalars and all the realizations of the covariance matrix obtained from the target tracking algorithm should be invertible. The subscript j in the above represents the different sample realizations or trials (and not the vector components). Each covariance instantiation, $\mathrm{P}_j(t_i)$ obtained from the particular tracking algorithm under scrutiny should be positive definite (and not merely non-negative definite) because:

1. it is the covariance for the entire state estimate, whose underlying filter model has independent constituents as a valid state variable model (only linearly dependent random vectors have an accompanying covariance that is not of full rank);
2. as the covariance outputted from the tracking algorithm that is ostensibly a stable tracker, as a necessary condition for estimator stability, the associated Lyapunov function used to argue stability for the tracker is of the standard form:

$$V(\hat{\mathbf{x}}) = \hat{\mathbf{x}}^{\mathrm{T}}(\mathbf{t_i})\left[\mathbf{P_j}(\mathbf{t_i})\right]^{-1}\hat{\mathbf{x}}(\mathbf{t_i}) \tag{50}$$

and, as such, should possess an invertible covariance matrix. The above is necessary and sufficient for Kalman filter stability when there is an underlying linear system as detailed in [9, Sec. 4.2 and App. A], [10, App. C]

Another possible variation of the above that comes to mind (a variant of Eq. 14 that is even less appropriate to use for NMD) is of the form:

$$\overline{\mathrm{s}}(t_i) \triangleq \left(\overline{\hat{\mathbf{x}}}(\mathbf{t_i}) - \mathbf{x_{true}}(\mathbf{t_i})\right)^{\mathrm{T}}\left[\frac{1}{\mathrm{N}}\sum_{\mathrm{j=1}}^{\mathrm{N}}\mathbf{P}_{\mathrm{j}}(t_i)\right]^{-1}\left(\overline{\hat{\mathbf{x}}}(\mathbf{t_i}) - \mathbf{x_{true}}(\mathbf{t_i})\right), \tag{51}$$

where $\mathbf{P}(t_i)$ is obtained from the Target tracking algorithm and not from sampled statistics and, as such, each should be positive definite and invertible (so the sum should be invertible as well and scaling by 1/N does not alter the outcome of being invertible). However, this version in Eq. 51 isn't theoretically justified and shouldn't be used even though an appealing aspect is that the averaging sum of the Monte-Carlo estimates (Eq. 16) would likely go to Gaussian in distribution (by an invocation of the Central Limit Theorem) for a large enough N even if the underlying estimates from approximate nonlinear filtering algorithm were not Gaussian.

Whether the test point $\overline{\hat{\mathbf{x}}}$ is within the bounding ellipsoid of *K*-sigma's at each calculation time point t_i depends (as an if and only if statement) on whether the calculated statistic **s** is less than or equal to *K*. If **s** exceeds *K*, then the test point is outside the *K*-sigma ellipsoid centered on x_{true}. One can plot the instantaneous **s** vs. time (see Fig. 8 below) and just look at the possible crossing trend relative to horizontal lines at ordinate 1, ordinate 2^2, or ordinate 3^2, representing, respectively, 1-, 2- or 3-sigma's away from the goal of x_{true} (where the ellipsoid is centered).

Multiplier of uncertainty ellipsoid (for a Gaussian Distribution) both centered on the mean for:

Dimension : N	Probability of Containment = 0.75	Probability of Containment = 0.98	Probability of Containment = 0.99
1	$K^2 = (1.150217)^2 = 1.323$	$K^2 = (2.326156)^2 = 5.4119$	$K^2 = (2.575849)^2 = 6.635$
3	$K^2 = (2.02682)^2 = 4.1083$	$K^2 = (3.136463)^2 = 9.8374$	$K^2 = (3.368234)^2 = 11.345$
6	$K^2 = (2.80018)^2 = 7.8408$	$K^2 = (3.877267)^2 = 15.0332$	$K^2 = (4.100244)^2 = 16.812$

Table 2: K-sigma's for specific Multivariate Gaussian containment probabilities

(Table 2 was prepared using Chi-square statistics from [11, Table 2] and cross-checked using the computer program **chi2inv.m** from The Math Work's *Statistics Toolbox* for MatLab, Ver. 6.5, Release 13. An alternate title for Table 2 would be "Score Values for Chi-Squared.")

From statistical considerations (assuming the errors are Gaussian, as can be explicitly tested for beforehand as discussed below), it follows from the supporting theory that:

$$\frac{1}{N}\sum_{j=1}^{N} r_j^2(t_i) \sim \chi^2 \text{ with 6+N-1 degrees-of-freedom.} \tag{52}$$

Although the tracking filter approximation to ideal non-linear filtering looks to be linear if it is a variant of Extended Kalman Filtering (EKF), in reality it is nonlinear since internal parameters are linearized about the EKF's previous estimate of the state. So this is yet another nonlinear operation on a Gaussian-like process (that yields a non-Gaussian output in general). However, it may be argued that the operations of an EKF are "almost" linear and so can be approximated as yielding a Gaussian process output but with a bias (which is unknown but can be estimated from realistic Monte-Carlo simulations at each time point t_i, of interest). Later, this non-stationary aspect or time varying trend can be handled entirely within this framework by comparison to non-central Chi-square statistics. (There are two different tests of Normality within The Math Work's Statistics Toolbox [**jbtest.m**, the Bera-Jarque parametric hypothesis test of composite normality, **lillietest.m**, single sample Lilliefors' hypothesis test of composite normality] and programs there as well to handle non-central Chi-Square [**ncx2cdf.m** and **ncx2inv.m**] and both assume that the population's true mean and variance are unknown and thus utilize Eqs. 16 and 17 in deducing whether to accept or reject the null hypothesis H_0 with significance level α that the sample is Gaussian [35].)

If sufficient Monte-Carlo evaluation trials substantiate (from applying the averaging of Eq. 16) that:

$$E[\hat{\mathbf{x}}(t_i)] \neq \mathbf{x}_{\mathbf{true}}(t_i), \tag{53}$$

by revealing that **the particular estimation algorithm under test** has a pronounced bias at the designated time point t_i after the start of track **for a specific target trajectory and designated radar sensor location**, then this nonzero mean can be used with non-central Chi-square statistics replacing the indicated Chi-square statistics/table usage to be used throughout all of the above Eqs. 8 to12, 49, and 52. The appropriate scalar non-centrality parameter to be used in the look-up tables is:

$$\delta^2(t_i) = \mathbf{b}^{\mathrm{T}}(t_i)\mathbf{b}(t_i), \tag{54}$$

where the algorithm's bias at the specific time point for the designated algorithm, target trajectory, and radar sensor location is:

$$\mathbf{b}(t_i) \triangleq E[\hat{\mathbf{x}}(t_i)] - \mathbf{x}_{\mathbf{true}}(t_i) \simeq \frac{1}{N}\sum_{j=1}^{N}\left[\hat{\mathbf{x}}_j(t_i) - \mathbf{x}_{\mathbf{true}}(t_i)\right]. \tag{55}$$

Please see discussion just before Eq. 53 above on recommended computer programs to perform calculations for non-central chi-square statistics.

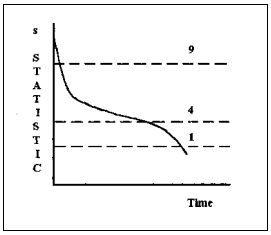

Figure 8: *Comparisons of test statistic* ***s*** *to constant levels reveals when estimates are acceptably close to goal of* x_{true}

Notice from Eq. 49 that when the sample test statistic is below a designated threshold, as $s(t_i) \le K^2$, this situation corresponds to:

$$\frac{1}{\mathrm{N}}\sum_{\mathrm{j}=1}^{\mathrm{N}}\left[\hat{\mathbf{x}}_{\mathrm{j}}(t_i)-\mathbf{x}_{true}(t_i)\right]^{\mathrm{T}}\left[\mathbf{P}_{\mathrm{j}}(t_i)\right]^{-1}\left[\hat{\mathbf{x}}_{\mathrm{j}}(t_i)-\mathbf{x}_{true}(t_i)\right]\le K^2 \text{ or, equivalently,}$$

$$\sum_{\mathrm{j}=1}^{\mathrm{N}}\left[\hat{\mathbf{x}}_{\mathrm{j}}(t_i)-\mathbf{x}_{true}(t_i)\right]^{\mathrm{T}}\left[\mathbf{P}_{\mathrm{j}}(t_i)\right]^{-1}\left[\hat{\mathbf{x}}_{\mathrm{j}}(t_i)-\mathbf{x}_{true}(t_i)\right]\le \mathrm{N}\cdot K^2, \tag{56}$$

which, for each index j within the range $1 \le \mathrm{j} \le \mathrm{N}$, the following expression

$$\left[\hat{\mathbf{x}}_{\mathrm{j}}(t_i)-\mathbf{x}_{true}(t_i)\right]^{\mathrm{T}}\left[\mathbf{P}_{\mathrm{j}}(t_i)\right]^{-1}\left[\hat{\mathbf{x}}_{\mathrm{j}}(t_i)-\mathbf{x}_{true}(t_i)\right]\le K^2 \tag{57}$$

corresponds to each realization in the sum being within its corresponding K^2-bounding ellipsoidal confidence region and by summing both sides yields:

$$\sum_{\mathrm{j}=1}^{\mathrm{N}}\left[\hat{\mathbf{x}}_{\mathrm{j}}(t_i)-\mathbf{x}_{true}(t_i)\right]^{\mathrm{T}}\left[\mathbf{P}_{\mathrm{j}}(t_i)\right]^{-1}\left[\hat{\mathbf{x}}_{\mathrm{j}}(t_i)-\mathbf{x}_{true}(t_i)\right]\le \sum_{\mathrm{j}=1}^{\mathrm{N}}K^2 \le \mathrm{N}\cdot K^2, \tag{58}$$

which, when rearranged, is just Eq. 56 above. For perspective, recall that the following (with an equals sign) represents the bounding surface of the n-dimensional ellipsoidal confidence region centered about $\mathrm{x}_{\mathrm{true}}$:

$$\left[\hat{\mathbf{x}}(t_i)-\mathbf{x}_{true}(t_i)\right]^{\mathrm{T}}\left[\mathbf{P}(t_i)\right]^{-1}\left[\hat{\mathbf{x}}(t_i)-\mathbf{x}_{true}(t_i)\right]=K^2, \tag{59}$$

while for values of $\hat{\mathbf{x}}$ such that the quadratic form on the left hand side of the above being **less than** K^2 corresponds to the **inside or being contained within the interior** of the ellipsoidal confidence region; while for values of $\hat{\mathbf{x}}$ such that the quadratic form on the left hand side of the above being **greater than** K^2 corresponds to **being external or outside of** the ellipsoidal confidence region. Some simple transparent low dimensional examples are provided in Appendix A as a refresher to illustrate correctly choosing K^2, manipulating it, and properly interpreting the confidence regions that it represents when there is a Gaussian probability measure associated with it.

8. A Recommendation to use Proximity to Cramer-Rao Lower Bounds as a Gauge of on-line Covariance Fidelity

8.1 Review of analytical basis and procedure for evaluating CRLBs for NMD Target Tracking

Under the standard assumption that the *estimator is unbiased*[2], then the familiar form of Cramer-Rao inequality encountered or invoked most frequently is:

$$E\left[(x-\hat{x})(x-\hat{x})^T|x\right]\ge\left[-E\left\{\left(\frac{\partial}{\partial x}\right)^T\left(\frac{\partial}{\partial x}\right)\ln\{p(z|x)\}\right\}\right]^{-1}\equiv\left[\mathrm{I}+\frac{\partial \mathrm{B}}{\partial \mathrm{x}}\right]^{\mathrm{T}}\overbrace{\left[\text{Fisher Information Matrix}\right]^{-1}}^{\boldsymbol{\mathcal{I}}^{-1}}\left[\mathrm{I}+\frac{\partial \mathrm{B}}{\partial \mathrm{x}}\right] \tag{60}$$

where the inequality here for these matrices is interpreted in the matrix positive semi-definite sense (i.e., $A \ge B \Rightarrow A\text{-}B \ge 0$). Please see references cited in [15] for details. It is this form[3] (under the widely invoked assumption that the estimator bias is non-existent or negligible) that has a RHS that is independent of the particular estimator being used and that may be

[2] The bias referred to here is inherent to a particular estimator and is generally **not** directly related to any underlying fundamental biases arising for reasons other than the structure of the estimator being employed within a particular application.

compared to a wide variety of distinctly different estimators as a single relative gauge throughout. The CRLB methodology is used here to gauge the quality of filter performance in the tracking task.

The above Cramer-Rao inequality arises in seeking to estimate an unknown parameter *x* using **any** estimator $\hat{x}$ and the measurement z(t), referenced above and available from the sensor as a time record, must be a non-trivial function of the unknown parameter *x* as

$$z(t) = h(x, t, v(t)). \tag{61}$$

In the above, *p(x|z)* is the conditional probability density function (pdf) of *x* given all the measurements *z*, and *v(t)* is the measurement noise. In exoatmospheric target tracking, *x* is deterministic and satisfies a known nonlinear ordinary differential equation and *v* is additive Gaussian white noise of known covariance intensity, hence *p(z|x)* is known.

Although other bounds exist like that of Barankin, the CRLB was selected for use as the familiar bound most appropriate for NMD application because it matches the situation and is tractable. A high-level overview of the CRLB methodology and its benefits and limitations may be found in [15] while it is specialized specifically for NMD exoatmospheric tracking in [16]-[18].

The CRLB being achieved means that the error of estimation term on the LHS of Eq. 60 touches the CRLB term on the RHS by satisfying the indicated inequality as an exact equality. For NMD/GMD target tracking, the lower bound should not generally be achievable (hence this CRLB is NOT expected to **exactly** match the average sampled tracking error variance compiled from *N* Monte-Carlo trials).

This CRLB was derived by adapting a time-varying radar SNR to realistically correspond to fluctuating PRF and other underlying signal processing as an enhancement of the fundamental methodology that evolved as tailored to this NMD radar application using the conventions laid out in [16]-[18]. The procedure of [16]-[19] already considered Pd < 1 since it included explicit consideration of the detection threshold settings and, moreover, used measurement reception times that corresponded to the time-tags for when measurements were actually received (so these CRLBs are a posteriori bounds).While [33] initially tries to tackle a more general case of a priori bounds, it found that approach to be intractable and so [33] then merely resorts to using the structure of CRLBs for handling process noise that is not zero (as occurs in indoatmospheric reentry tracking and not in exoatmospheric tracking, where Q=0). This is the big distinction between the CRLB approach of [33] and that of [16]-[19].

8.2 Insights into When and Why CRLBs sometimes Appear to have Weird Behavior

There is frequently a small initial time segment in the beginning of an estimation error plot when an estimator's covariance lies below the CRLB (as it should, considering the approximations that are usually invoked up to that point, as will be explained) before switching to the usual situation of the CRLB lying below. Sometimes the initial values are so large and far off that, by the automatically adjusted vertical scale inherent in many plot packages, this initial switch appears to be such a proportionately small segment of the figure that it doesn't raise suspicion or concern enough to be explained to an audience of readers. The same type of thing occurred for each of

[3] The summarizing notation $\mathcal{I}$ appearing on the Right Hand Side (RHS) in Eq. 60 is known as the *Information matrix* prior to matrix inversion, after which the entire expression (after $\partial B/\partial x \to 0$) is the so-called or so-designated Cramer-Rao Lower Bound (CRLB), which can be numerically evaluated.

Farina's four estimators in [34]. To the present author, this situation is a mark of "honesty in the preparation of the results" (but impedes a presentation somewhat when the speaker has to stop and explain why an apparent bewildering situation occurs of the direction of the expected inequality flipping and CRLB being above the sampled σ). An explanation was not given in [34] nor has it been given anywhere else to this author's knowledge so we will do so here now.

The explanation is because of that numerator factor $\left(\mathrm{I}+\partial \mathrm{B}/\partial x\right)^2$ in Eq. 60 that is needed for CRLB to be below ALL the time, where B is the *bias* in the estimator. Since we don't usually have B explicitly available and even when we do, its sensitivity to the parameter (in this case the state vector) being estimated needs to be evaluated as the indicated derivative (which is usually not conveniently tractable) so it is usually ignored entirely since it can't be evaluated anyway and we focus instead on the denominator term (which is the inverse of the Fisher Information matrix), which we can evaluate numerically. The numerator term can be a magnifier or a minimizer, depending on whether it is greater that 1 or less than 1 and it changes with time. Since estimators frequently proceed to have a steady-state bias (where $\partial \mathrm{B}/\partial x$ becomes zero), the exact CRLB expression (numerator and denominator) eventually converges to the approximate CRLB expression (involving denominator alone). Since we frequently only have explicit access to the denominator alone, we usually use just that and wait past the initial transient until it is appropriate to compare against because only then do we have that the right hand side CRLB $\approx \mathcal{I}^{-1}$, where it is reminded that $\mathcal{I}$ here is the *Fisher Information Matrix*.

8.3 Analytical Derivation of the CRLB for Target Tracking situations devoid of Process Noise (e.g., as occur in ballistic trajectories that are exclusively exoatmospheric)

The CRLB that is treated here goes beyond just using the historically familiar per pulse CRLB angle measurement error[4]: $\sigma_\theta = \dfrac{\theta_3}{1.6\sqrt{2 \cdot SNR(t)}}$ (in [16]-[19]) since our CRLB goes further to additionally utilize **(1)** information provided by the target dynamics model over time in an inverse square gravity field, **(2)** the initial (starting) covariance *P(0)* of the tracking filter as handed-over[5], and **(3)** the structure of the radar as a measurement sensor/device having additive Gaussian measurement noise with parameters including **(3a)** explicit use of the radar range uncertainty due to resolution size of the range gates and **(3b)** the monopulse SNR time-record with its adaptive step size (as a consequence of a realistically varying PRF) as it affects the corresponding angle uncertainty. [However, one CRLB version used the SNR records simulated by TD/SAT and each sample function realization interpolated to common times throughout (and interpolated) and then averaged (by Dan Pulido, previously at General Dynamics using MatLab) to obtain SNR values at designated periodic times, thus providing smooth CRLBs as an envelope for comparison to estimator performance.]

For UEWR application, which has additive Gaussian white measurement noise *v(t)*, Eq. 61 has this further more benign and accommodating structure to be exploited:

$$z(t) = h(x,t) + v(t), \tag{62}$$

and since the equation for the system evolution is essentially deterministic (with $Q_c=0$), then the pdf's of interest here (to be used in numerically evaluating the CR lower bound of Eq. 60) are of the form:

[4] Within this notation, Θ_3= 3 dB receive sum-pattern beamwidth.
[5] We used a standard hand-over covariance of $(100\ \mathrm{km})^2$ for all three components of the position block and $(100\ \mathrm{m/sec})^2$ for all three components of the velocity block. Physically, this should come from SBIRS for NMD.

$$p(z|x) = \frac{1}{(2\pi)^{n/2}|R|^{1/2}} \cdot e^{-\frac{1}{2}(z-h(x))^T R^{-1}(z-h(x))} \tag{63}$$

Now taking natural logarithms on both sides of the above pdf yields:

$$\ln\{p(z|x)\} = -\frac{1}{2}(z-h(x))^{\mathrm{T}} R^{-1}(z-h(x)) - \ln(2\pi)^{n/2}|R|^{\frac{1}{2}} \tag{64}$$

which upon taking the gradient is:

$$\left(\frac{\partial}{\partial x}\right)^{\mathrm{T}} \ln\{p(z|x)\} = \frac{\partial^{\mathrm{T}} h(x)}{\partial x} R^{-1}(z-h(x)). \tag{65}$$

When the above expression is post-multiplied by its transpose and expectation taken throughout, the result is:

$$\begin{aligned} \mathrm{E}\left[\left(\frac{\partial}{\partial x}\right)^{\mathrm{T}} \ln\{p(z|x)\} \frac{\partial}{\partial x} \ln\{p(z|x)\}\right] &= \frac{\partial^{\mathrm{T}} h(x)}{\partial x} R^{-1} \overbrace{E\left[(z-h(x))(z-h(x))^{\mathrm{T}}\right]}^{R} R^{-1} \frac{\partial h(x)}{\partial x} \\ &= \frac{\partial^{\mathrm{T}} h(x)}{\partial x} R^{-1} \frac{\partial h(x)}{\partial x} \end{aligned} \tag{66}$$

Finally, over corresponding discrete-time steps (not necessarily uniform), the total pdf of independent (white) measurements is the product of each individual pdf of the form of Eq. 63 as $p(z_1|x(0))p(z_2|x(0))p(z_3|x(0))$ ... $p(z_k|x(0))$, where each pdf for each constituent measurement here focuses on or is conditioned on the initial condition for the deterministic system equation. Once the initial condition $x(0)$ is known with confidence, then the time evolution of the deterministic system is completely determined (as a consequence of initial condition *observability*). The corresponding information matrix for each of these measurement time points is of the form of Eq. 66 so the aggregate is of the form[6]:

$$\boldsymbol{\mathcal{I}}(k,0) = \sum_{j=1}^{k} \Phi^{-T}(k,j) \frac{\partial^T h(x)}{\partial x}\Big|_j R^{-1}(j,j) \frac{\partial h(x)}{\partial x}\Big|_j \Phi^{-1}(k,j), \tag{67}$$

for $k \geq j$, where the transition matrix $\Phi^{-1}(k,j) \overset{\Delta}{=} \left[\Phi(k,j)\right]^{-1} = \Phi(j,k)$ and, likewise, corresponds to an evaluation of the system matrix linearized about the true state. Now, when there is a finite initial covariance being utilized by the estimator as tracking commences, then there is an additional term that should appear in the above Information matrix to properly reflect this situation, as depicted as the first term on the RHS here:

$$\boldsymbol{\mathcal{I}}(k,0) = \Phi^{-\mathrm{T}}(k,0) P^{-1}(0) \Phi^{-1}(k,0) + \sum_{j=1}^{k} \Phi^{-\mathrm{T}}(k,j) \frac{\partial h(x)}{\partial x}\Big|_j R^{-1}(j,j) \frac{\partial h(x)}{\partial x}\Big|_j \Phi^{-1}(k,0) \tag{68}$$

In either the case of Eq. 67 or Eq. 68 holding, the Information matrix can be interpreted or formulated as evolving recursively with each received measurement arrival time as:

$$\boldsymbol{\mathcal{I}}(k,0) = \Phi^{-\mathrm{T}}(k,j) \boldsymbol{\mathcal{I}}(j,0) \Phi^{-1}(k,j) + \frac{\partial^{\mathrm{T}} h(x)}{\partial x}\Big|_k R^{-1}(k,k) \frac{\partial h(x)}{\partial x}\Big|_k \tag{69}$$

[6] After taking the natural logarithm of the aggregate pdf, the exponents in the Gaussian distribution correspond to the indicated sum, after performing a gradient and taking expectations, as illustrated in detail above in Eqs. 63 to 66 for just a single measurement for clarity.

and, as such, may be implemented within software as merely a loop (but by observing all the constraints and coordinate conventions, where $\frac{\partial h}{\partial x}$ is evaluated within the ECI frame and $\frac{\partial h}{\partial x}$ is evaluated in the (E,N,U) frame[7] with corresponding translation offset to the location of the tracking radar)[8]. We have particular interest in the total position error and the corresponding total velocity error to determine how well we are actually doing in tracking a target complex. To this end, we must rigorously contort the inequality of Eq. 60 to a form that we can use. This is accomplished by properly applying matrix operations that yield the expressions that we seek[9] as:

$$\begin{bmatrix} \sigma_{11}^2 & \sigma_{12}^2 & \sigma_{13}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 & \sigma_{23}^2 \\ \sigma_{31}^2 & \sigma_{32}^2 & \sigma_{33}^2 \end{bmatrix} = \begin{bmatrix} 100000 \\ 010000 \\ 001000 \end{bmatrix} E\left[\left(x_{true}(t) - \hat{x}(t) \right) \left(x_{true}(t) - \hat{x}(t) \right)^{\mathrm{T}} \left| \boldsymbol{\mathcal{Z}}(t) \right. \right] \begin{bmatrix} 100000 \\ 010000 \\ 001000 \end{bmatrix}^{\mathrm{T}} \tag{70}$$
$$\geq \begin{bmatrix} 100000 \\ 010000 \\ 001000 \end{bmatrix} \boldsymbol{\mathcal{I}}^{-1} \begin{bmatrix} 100000 \\ 010000 \\ 001000 \end{bmatrix}^{\mathrm{T}} = \begin{bmatrix} crlb_{11} & crlb_{12} & crlb_{13} \\ crlb_{21} & crlb_{22} & crlb_{23} \\ crlb_{31} & crlb_{32} & crlb_{33} \end{bmatrix}$$

and

$$\begin{bmatrix} \sigma_{44}^2 & \sigma_{45}^2 & \sigma_{46}^2 \\ \sigma_{54}^2 & \sigma_{55}^2 & \sigma_{56}^2 \\ \sigma_{64}^2 & \sigma_{65}^2 & \sigma_{66}^2 \end{bmatrix} = \begin{bmatrix} 000100 \\ 000010 \\ 000001 \end{bmatrix} E\left[\left(x_{true}(t) - \hat{x}(t) \right) \left(x_{true}(t) - \hat{x}(t) \right)^{\mathrm{T}} \left| \boldsymbol{\mathcal{Z}}(t) \right. \right] \begin{bmatrix} 000100 \\ 000010 \\ 000001 \end{bmatrix}^{\mathrm{T}} \tag{71}$$
$$\geq \begin{bmatrix} 000100 \\ 000010 \\ 000001 \end{bmatrix} \boldsymbol{\mathcal{I}}^{-1} \begin{bmatrix} 000100 \\ 000010 \\ 000001 \end{bmatrix}^{\mathrm{T}} = \begin{bmatrix} crlb_{44} & crlb_{45} & crlb_{46} \\ crlb_{54} & crlb_{55} & crlb_{56} \\ crlb_{64} & crlb_{65} & crlb_{66} \end{bmatrix}$$

and then by taking the trace of a matrix throughout[10], respectively, yields **radial position error variance**:

$$\sigma_{position}^2 = \sigma_{11}^2 + \sigma_{22}^2 + \sigma_{33}^2 = \mathrm{tr} \begin{bmatrix} \sigma_{11}^2 & \sigma_{12}^2 & \sigma_{13}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 & \sigma_{23}^2 \\ \sigma_{31}^2 & \sigma_{32}^2 & \sigma_{33}^2 \end{bmatrix} \geq \mathrm{tr} \begin{bmatrix} crlb_{11} & crlb_{12} & crlb_{13} \\ crlb_{21} & crlb_{22} & crlb_{23} \\ crlb_{31} & crlb_{32} & crlb_{33} \end{bmatrix} = crlb_{11} + crlb_{22} + crlb_{33} = CRLB_{position} \tag{72}$$

and **total velocity error variance:**

[7] A representation in sine space centered within the antenna array is recommended for consistency with UEWR.
[8] Notice that nothing was presumed of the estimator in deriving and evaluating Eq. 60 beyond the underlying measurement structure of Eqs. 61, 62 and the availability of all measurements **up to the current time** k. Alternative estimators that "smooth" by estimating the state x_k using measurements beyond k may violate this assumption and this CRLB but they aren't real-time. The appropriate CRLB to correspond to an estimator that uses measurements beyond the current time of interest (such as in "sliding window" smoothing or in "fixed point" smoothing, or BLS) should just have the additional corresponding terms beyond the current time also included in Eqs. 68 and 69.
[9] Pre- and post-multiplying $A \geq B$ by the same matrix L yields $LAL^T \geq LBL^T$.
[10] The matrix inequality $A \geq B$ implies that **trace**[A] $\geq$ **trace**[B].

$$\sigma^2_{velocity} = \sigma^2_{44} + \sigma^2_{55} + \sigma^2_{66} = \text{tr}\begin{bmatrix} \sigma^2_{44} & \sigma^2_{45} & \sigma^2_{46} \\ \sigma^2_{54} & \sigma^2_{55} & \sigma^2_{56} \\ \sigma^2_{64} & \sigma^2_{65} & \sigma^2_{66} \end{bmatrix} \geq \text{tr}\begin{bmatrix} crlb_{44} & crlb_{45} & crlb_{46} \\ crlb_{54} & crlb_{55} & crlb_{56} \\ crlb_{64} & crlb_{65} & crlb_{66} \end{bmatrix} = crlb_{44} + crlb_{55} + crlb_{66} = CRLB_{velocity} \tag{73}$$

and, finally, by taking squareroots throughout[11], respectively, yields:

$$\sigma_{position} = \sqrt{\sigma^2_{11} + \sigma^2_{22} + \sigma^2_{33}} \geq \sqrt{crlb_{11} + crlb_{22} + crlb_{33}} \stackrel{\Delta}{=} \sqrt{CRLB_{position}} \tag{74}$$

and

$$\sigma_{velocity} = \sqrt{\sigma^2_{44} + \sigma^2_{55} + \sigma^2_{66}} \geq \sqrt{crlb_{44} + crlb_{55} + crlb_{66}} \stackrel{\Delta}{=} \sqrt{CRLB_{velocity}} \tag{75}$$

Please notice in the above that we do not decouple position and velocity states but merely project both of the 6 x 6 matrices of Eq. 60, respectively, into the position subspace (as Eqs. 70, 72, 74) and into the velocity subspace (as Eqs. 71, 73, 75) for viewing in a plotter display. These instantaneous inequalities are now the theoretically justified comparisons that we invoke in monitoring performance of any target tracking algorithm under scrutiny as a function of time.

8.4 Assessing EKF Tracking Performance BASELINE: Comparing Existing Standard EKF to the Computed CRLB

We illustrated the CRLB calculations relative to ensemble sampled Monte-Carlo results for the BMEWS radar: Thule[12] tracking an RV on a ballistic trajectory (post cut-off) having the following position and velocity states at cut-off time normalized to $t_o=0$ seconds:

$$x^{\text{T}}(t_0) = \left[\text{-3217302.678, } 3527834.349, \text{ } 4535013.695, \text{ -767.670, -2520.638, } 5065.414\right]^{\text{T}}, \tag{76}$$

where in the above, the units are in meters for position and meters/sec for velocity, respectively. The simulations of the radar case [using known BMEWS published Cobra Dane measurement covariance's for range and angle being[13]

$$\sigma_{range} = 30 \text{ meters (per pulse)}; \ \sigma_{angle} = \frac{2.2}{1.6\sqrt{2 \cdot SNR(t)}} \text{ degrees (per pulse)}, \tag{77}$$

respectively[14]] appear to be performing properly, as depicted in Fig. 3-1 of [18] for the case of a nonlinear target (corresponding to use of the system truth model used for simulating the trajectory, but linearized about the estimate within the EKF) while both situations utilized the same nonlinear measurement model. Both parameters in Eq. A.26 of [18] (with SNR varying with time) are used in Eq. A.69 of [18] with $\sigma_E \equiv \sigma_{angle}$. For position error at time t (and

[11] Scalar $a \geq b \geq 0$ implies that $\sqrt{a} \geq \sqrt{b}$.

[12] This 10 MHz bandwidth Thule radar (AN/FPS-123V5), with a beamwidth of 1.8° is located in Greenland at Latitude = 76.56° N, Longitude = 297.70° E. The actual range resolution is determined by beam forming to reduce sidelobes and assumptions on range accuracy of from as little as 15 meters (for the 10 MHz signal) up to more than 30 meters (for the 5 MHz signal) shouldn't significantly alter the subsequently computed results since sensitivity to the range uncertainty parameter is low as compared to the effect of the more dominant angle uncertainty.

[13] Expressed within our software in MKS units with angles in radians, respectively.

[14] The radar's intrinsic range gate size dictates the effective range resolution, which is a constraint that is less restrictive than the angle acuity.

similarly for corresponding velocity with obvious direct replacement substitutions in the LHS of Eqs. 74 and 75), calculated as

$$\sqrt{\left(x_{true}(t)-\hat{x}_i(t)\right)^2+\left(y_{true}(t)-\hat{y}(t)\right)^2+\left(z_{true}(t)-\hat{z}(t)\right)^2}\,, \tag{78}$$

the corresponding sampled variance over N trials ($N=250$) being[15]:

$$\sum\nolimits_N = \left[\frac{1}{N}\sum_{i=1}^{N}\left(x_{true}(t)-\hat{x}_i(t)\right)^2+\left(y_{true}(t)-\hat{y}_i(t)\right)^2+\left(z_{true}(t)-\hat{z}_i(t)\right)^2\right] - \left[\frac{1}{N}\sum_{i=1}^{N}\sqrt{\left(x_{true}(t)-\hat{x}_i(t)\right)^2+\left(y_{true}(t)-\hat{y}_i(t)\right)^2+\left(z_{true}(t)-\hat{z}_i(t)\right)^2}\right]^2 \tag{79}$$

were depicted for UEWR as diagrammatic plots in [22]-[24], [53] (*N=1,000 in 1997 results*).

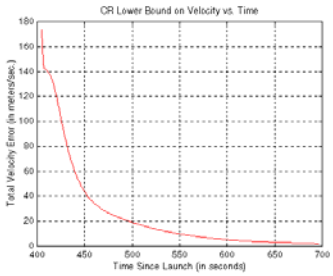

Figure 9: *CRLB on radial velocity accuracy (Threat 1)[18]*

9. An Offer to Perform Specific Tasks that Support Boeing's Covariance Fidelity Analysis and Simulation Effort to Arrive at Appropriate Specs

Precedents for using a test statistic **s** of the form of Eq. 49 as a consistency test between estimates and variances within a tracking filter context for EKF's may be found in [12, p. 42, Eqs. 1.6.3-1,-2] and in [13, Sec. 5.4]. While we at TeK Associates have prior experience using this particular test statistic ourselves (since it was originally used for the application of [30]), we eventually modified it to use the methodology of [26]–[30] as an application outcome of original directed research. This was needed for a real-time monitoring test of a newly introduced navigation system to avail the test with more leeway before it alarmed (to alert the technician to a failure of the new system component, where the old was still being conservatively retained as a warm stand-by system for back-up and could then be switched in if needed).

Under a contract from Boeing, TeK Associates would be happy to review plots of the **s** statistic of Eq. 49 versus time in trying to access appropriate behavior as compared to the probability of containment of its associated underlying Gaussian Confidence Region(s) interpretation. Before deciding whether to use two 3-D confidence region assessments or one 6-D confidence region assessment, both options should be used for testing before a final decision is made to use the best (by figuring out which is better based on performance and not just on analytics such as what we have provided here). Notice from the form of Eq. 49 that too large (conservative or pessimistic) a value appearing as the inner product matrix $\mathbf{P}^{-1}(t)$ would yield a value for the test statistic **s**(t) that is smaller than it should be. We seek veracity in the tracker algorithm outputted covariance **P**(t) not just smallness in **s**(t) resulting from largeness in **P**(t).

[15] Notice that this is of the form $E[(W-E[W])^2]=E[W^2]-(E[W])^2$.

Beyond just having a measure to correctly appraise and gauge the outputted **P**(t), would it not be better to get to the source of the problem and improve the outputted values of **P**(t) itself by further tuning the two EKF's used as the UEWR target tracking algorithms? Other alternatives would be to augment them as either an iterated EKF (to improve the linearization) or by using an additional Hessian (2^{nd} derivative) term as well as the current Jacobian (1^{st} derivative) term to better capture the effect of the explicitly known nonlinearities present in the system and measurement models. There are other approaches to estimator amelioration as well and TeK Associates has recently submitted an unsolicited proposal on 20 May 2003 (still under review) to the Missile Defense Agency (MDA) to perform these tasks.

In seeking to use the CRLB methodology as a way of gauging the Covariance Fidelity for on-line tracker specs, we now realize (from dealing with the three confidence region alternatives considered herein) that instead of just looking at 3 x 3 position error and 3 x 3 velocity error as summarized, respectively, into just total position error and total velocity error (as we did before), in some cases we should look at each off-diagonal component as well since these define how the confidence regions tilt in the state space. We realize that this means looking at $n(n+1)/2 = (6)(7)/2=21$ plots (but at least not having to look at 6 x 6 =36 plots since the covariance as well as CRLB's are symmetric)! TeK Associates already has the above-described CRLB methodology coded up in MatLab. Of course, it would now need to be modified to also accommodate plotting all 21 covariance components but that is a minor modification. TeK Associates has already coded up a simple ballistic target simulator in Simulink with inverse squared gravity (but with no J_2 present) that is appropriate to use for this application. As before, TeK Associates expects to be given the SNR time record (from a more detailed simulator like TD/SAT was and in terms of seconds in time since launch) for all missions that it handles. TeK already has corresponding target initial conditions in terms of position and velocity for several trajectories (~30+) in their unclassified post-launch phase. The lure or motivation for using CRLB as a comparison to target tracker on-line covariances is that these CRLB's are: (1) linearized about the ideal true trajectory and are error free while target tracker EKF covariances are obtained from non-ideal linearizations about their last estimate; (2) these CRLB's are for Q=0 while target tracker EKF's have tuning corresponding to non-zero Q and so should be larger. Hence proximity of target tracker on-line covariances to CRLB should be a good gauge of Covariance fidelity (especially if we look at all 21 components for closeness).

A detail is offered here now regarding how to conveniently handle $n(n+1)/2=6(6+1)/2=21$ components constituting the comparison test of tracker-algorithm-generated covariance for better (closer) proximity to Cramer-Rao Lower Bound. Each component of interest in the comparison to a CRLB template is a function of time over a specified mission time interval. Since we seek the algorithm with "best proximity to ideal CRLB templates" as possessing greater covariance fidelity, we should state how we would automate such comparisons or else risk being overwhelmed with a high dimensional comparison of 21 items where some components could be closer to their CRLB templates than other components at various times during a particular mission (of specified finite duration from t_0 until t_f). The comparison problem is compounded when other candidate tracker algorithms are considered in the mix. While this initially appears to be a hard problem, we state here how it may be simply handled.

Our guide here will be three observations: two about distance measures and one about the structure of the practical problem facing us involving CRLBs with initial transients (Sec. 8.2):

1. Several candidate distance measures are available for us to use as provided by the field of mathematics known as *Functional Analysis*, where these distances are routinely used to measure the closeness of one function to another function within the "space of functions" such as $d(f,g) \triangleq \sqrt{\int_{t_0}^{t_f} \left(f(t)-g(t)\right)^2 dt}$, for example. The situation is simplified (or less stressing) when the domain of the functions of concern are limited to a finite interval of the real line, as is the case for out application on $[t_0, t_f]$ rather than over an infinite or semi-infinite interval (i.e., $(-\infty,\infty)$ or $[0, \infty)$).
2. The linear combination of distance measures is again a distance measure.
3. It is acknowledged that initial transient that is present in a standard CRLB evaluation clouds the issue. It is safer to wait until after the initial transient has died out (as at, say, by t_{aithdo}) before making comparisons of outputted tracker algorithm covariances to CRLB templates: $\mathrm{d}(f,g) \triangleq \sqrt{\int_{t_{aithdo}}^{t_f} \left(f(t)-g(t)\right)^2 dt}$. Of course, there is a discrete-time analog to all of the above that uses sums rather than integrals so it is even simpler.

The above observation 2 enters in defining the final scalar criterion to be used as a gauge to order the results of the comparisons. Since principal diagonal terms are more important than off –diagonal terms that only represent the tilts, we suggest using something like

$$
\begin{aligned}
\mathrm{d}^*(\mathbf{P(t)},\mathbf{CRLB(t)}) = (1-\mu_1)&\left\{\begin{matrix}\left[\mathrm{d}(\mathbf{P}_{11}\mathbf{(t)},\mathbf{CRLB}_{11}\mathbf{(t)})\right]+\left[\mathrm{d}(\mathbf{P}_{22}\mathbf{(t)},\mathbf{CRLB}_{22}\mathbf{(t)})\right]+\left[\mathrm{d}(\mathbf{P}_{33}\mathbf{(t)},\mathbf{CRLB}_{33}\mathbf{(t)})\right] \\ +\left[\mathrm{d}(\mathbf{P}_{44}\mathbf{(t)},\mathbf{CRLB}_{44}\mathbf{(t)})\right]+\left[\mathrm{d}(\mathbf{P}_{55}\mathbf{(t)},\mathbf{CRLB}_{55}\mathbf{(t)})\right]+\left[\mathrm{d}(\mathbf{P}_{66}\mathbf{(t)},\mathbf{CRLB}_{66}\mathbf{(t)})\right]\end{matrix}\right\} \\
+\mu_1&\left\{\begin{matrix}\left[\mathrm{d}(\mathbf{P}_{12}\mathbf{(t)},\mathbf{CRLB}_{12}\mathbf{(t)})\right]+\left[\mathrm{d}(\mathbf{P}_{13}\mathbf{(t)},\mathbf{CRLB}_{13}\mathbf{(t)})\right]+\left[\mathrm{d}(\mathbf{P}_{14}\mathbf{(t)},\mathbf{CRLB}_{14}\mathbf{(t)})\right]+\left[\mathrm{d}(\mathbf{P}_{15}\mathbf{(t)},\mathbf{CRLB}_{15}\mathbf{(t)})\right]+ \\ \left[\mathrm{d}(\mathbf{P}_{16}\mathbf{(t)},\mathbf{CRLB}_{16}\mathbf{(t)})\right]\end{matrix}\right\} \\
+\mu_1&\left\{\left[\mathrm{d}(\mathbf{P}_{23}\mathbf{(t)},\mathbf{CRLB}_{23}\mathbf{(t)})\right]+\left[\mathrm{d}(\mathbf{P}_{24}\mathbf{(t)},\mathbf{CRLB}_{24}\mathbf{(t)})\right]+\left[\mathrm{d}(\mathbf{P}_{25}\mathbf{(t)},\mathbf{CRLB}_{25}\mathbf{(t)})\right]+\left[\mathrm{d}(\mathbf{P}_{26}\mathbf{(t)},\mathbf{CRLB}_{26}\mathbf{(t)})\right]\right\} \\
+\mu_1&\left\{\left[\mathrm{d}(\mathbf{P}_{34}\mathbf{(t)},\mathbf{CRLB}_{34}\mathbf{(t)})\right]+\left[\mathrm{d}(\mathbf{P}_{35}\mathbf{(t)},\mathbf{CRLB}_{35}\mathbf{(t)})\right]+\left[\mathrm{d}(\mathbf{P}_{36}\mathbf{(t)},\mathbf{CRLB}_{36}\mathbf{(t)})\right]\right\} \\
+\mu_1&\left\{\left[\mathrm{d}(\mathbf{P}_{45}\mathbf{(t)},\mathbf{CRLB}_{45}\mathbf{(t)})\right]+\left[\mathrm{d}(\mathbf{P}_{46}\mathbf{(t)},\mathbf{CRLB}_{46}\mathbf{(t)})\right]\right\} \\
+\mu_1&\left\{\left[\mathrm{d}(\mathbf{P}_{56}\mathbf{(t)},\mathbf{CRLB}_{56}\mathbf{(t)})\right]\right\},
\end{aligned}
\tag{80}
$$

where $\mathrm{d}(\cdot,\cdot)$ has already been defined above in Items 1 and 3, and the setting of the weight as $0 < \mu_1 < 1$ should be fixed beforehand. In order to emphasize the proximity of main diagonal components more than the off-diagonal components, it should be the case that $0 < \mu_1 < 0.5$. There are other candidate function space distances to possibly be used for such a comparison. The one that is actually used and the form of weightings in the linear combinations could ostensibly be specifically tailored to be different for each separate component. The results should be compared to what common-sense human-in-the-loop eye-ball comparisons would also conclude from these plots (for calibration) before they are relied upon and calculated automatically as numerical summaries of proximity.

TeK Associates should also be able to find some multidimensional hypothesis tests in [35] that are useful for interpreting or augmenting the **s** test statistic of Sec. 7. TeK Associates seeks to go further with the associated interpretation of **s**(t) as confidence regions [two 3-D vs. one 6-D] and look into applying results of [36] in this effort to further delineate and understand.

How much time all this takes depends on how much of this job Boeing wants us to do and how many diverse mission scenarios of target aim point/launch point/radar location are to be considered and how many alternative target trackers need to be evaluated. Based on our prior experience as well as literature searches, TeK Associates has a certain proprietary perspective and preferred prioritization on what EKF modifications should be tried (as discussed in our MDA proposal, mentioned above).

Appendix A: Simple Review of Some Equations Describing Ellipsoids

The equation of a 1-sigma ellipsoid in 2-D (i.e., an ellipse) about the origin is:

$$1=\frac{x^2}{4^2}+\frac{y^2}{3^2}=\begin{bmatrix}x,y\end{bmatrix}\begin{bmatrix}\frac{1}{4^2} & 0\\ 0 & \frac{1}{3^2}\end{bmatrix}\begin{bmatrix}x\\ y\end{bmatrix}=\begin{bmatrix}x,y\end{bmatrix}\begin{bmatrix}4^2 & 0\\ 0 & 3^2\end{bmatrix}^{-1}\begin{bmatrix}x\\ y\end{bmatrix} \tag{A-1}$$

The equation of a 1-sigma ellipsoid in 3-D about the origin is:

$$1=\frac{x^2}{4^2}+\frac{y^2}{3^2}+\frac{z^2}{2^2}=\begin{bmatrix}x,y,z\end{bmatrix}\begin{bmatrix}\frac{1}{4^2} & 0 & 0\\ 0 & \frac{1}{3^2} & 0\\ 0 & 0 & \frac{1}{2^2}\end{bmatrix}\begin{bmatrix}x\\ y\\ z\end{bmatrix}=\begin{bmatrix}x,y,z\end{bmatrix}\begin{bmatrix}4^2 & 0 & 0\\ 0 & 3^2 & 0\\ 0 & 0 & 2^2\end{bmatrix}^{-1}\begin{bmatrix}x\\ y\\ z\end{bmatrix} \tag{A-2}$$

The above two numerical examples have their major and minor axes of the ellipse aligned with the underlying coordinate axes. The equation of a 1-sigma tilted (non-diagonal Σ) ellipsoid in 3-D about the point (x_o, y_o, z_o) is:

$$1=\begin{bmatrix}x-x_0, y-y_0, z-z_0\end{bmatrix}\sum{}^{-1}\begin{bmatrix}x-x_0\\ y-y_0\\ z-z_0\end{bmatrix} \tag{A-3}$$

The equation of a $\sqrt{c}$-sigma tilted (non-diagonal Σ) ellipsoid in 3-D about the point (x_o, y_o, z_o) is:

$$c=\begin{bmatrix}x-x_0, y-y_0, z-z_0\end{bmatrix}\sum{}^{-1}\begin{bmatrix}x-x_0\\ y-y_0\\ z-z_0\end{bmatrix} \tag{A-4}$$

Notice that by taking $c=K^2$ in Tables 1 or 2, the above ellipsoids correspond to the confidence regions with prescribed probability of containment for Gaussianly distributed random variables. It is K^2 times the covariance but only K times the standard deviation (as the squareroot of the covariance).

Two planar numerical examples to aid intuition and interpretation:

The equation of a **1-sigma ellipse** is:

$$1 = \frac{x^2}{4^2} + \frac{y^2}{3^2} = [x, y] \begin{bmatrix} \frac{1}{4^2} & 0 \\ 0 & \frac{1}{3^2} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = [x, y] \begin{bmatrix} 4^2 & 0 \\ 0 & 3^2 \end{bmatrix}^{-1} \begin{bmatrix} x \\ y \end{bmatrix} \quad \text{(A-5)}$$

In the above, observe that, $\sum = \begin{bmatrix} 4^2 & 0 \\ 0 & 3^2 \end{bmatrix}$ and $\lambda_{min}=3^2$ and $\lambda_{max}=4^2$.

Circumscribed Circle: $x^2 + y^2 = \lambda_{max} \Rightarrow$ circle radius = $\sqrt{\lambda_{max}} = 4$

Inscribed Circle: $x^2 + y^2 = \lambda_{min} \Rightarrow$ circle radius = $\sqrt{\lambda_{min}} = 3$

The equation of a $\sqrt{C}$ **-sigma ellipse** is:

$$C = \frac{x^2}{4^2} + \frac{y^2}{3^2} = [x, y] \begin{bmatrix} \frac{1}{4^2} & 0 \\ 0 & \frac{1}{3^2} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = [x, y] \begin{bmatrix} 4^2 & 0 \\ 0 & 3^2 \end{bmatrix}^{-1} \begin{bmatrix} x \\ y \end{bmatrix} \quad \text{(A-6)}$$

or, equivalently, by dividing throughout by *C* yields:

$$1 = \frac{1}{C}\left(\frac{x^2}{4^2} + \frac{y^2}{3^2}\right) = [x, y] \frac{1}{C} \begin{bmatrix} \frac{1}{4^2} & 0 \\ 0 & \frac{1}{3^2} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = [x, y] \begin{bmatrix} C4^2 & 0 \\ 0 & C3^2 \end{bmatrix}^{-1} \begin{bmatrix} x \\ y \end{bmatrix} \quad \text{(A-7)}$$

where, $C\sum = \begin{bmatrix} C4^2 & 0 \\ 0 & C3^2 \end{bmatrix}$ and $\lambda_{min}\left[C\sum\right] = C\,3^2$ and $\lambda_{max}\left[C\sum\right] = C\,4^2$.

Circumscribed Circle: $x^2 + y^2 = \lambda_{max} \Rightarrow$ circle radius = $\sqrt{\lambda_{max}} = \sqrt{C}\ 4$

Inscribed Circle: $x^2 + y^2 = \lambda_{min} \Rightarrow$ circle radius = $\sqrt{\lambda_{min}} = \sqrt{C}\ 3$,

Where for 0.97 SEP, $\sqrt{C} = \sqrt{8.85} = 2.97 \approx 3$

Same spherical radii result for rotated ellipsoids

For an un-rotated original:

$$1 = \mathbf{x}^{\mathbf{T}} \sum^{-1} \mathbf{x} \quad \text{(A-8)}$$

A subsequent rotation is merely represented as a unitary transformation

$$\mathbf{x} = \mathbf{U}\mathbf{w}\ , \text{ with } \mathbf{U}^{\mathbf{T}} = \mathbf{U}^{-1}, \quad \text{(A-9)}$$

Which, upon applying, results in:

$$1 = \mathbf{w}^{\mathbf{T}}\mathbf{U}^{\mathbf{T}} \sum^{-1} \mathbf{U}\mathbf{w}\ , \quad \text{(A-10)}$$

However the eigenvalues (hence the radius of the circumscribed and inscribed spheres) remain the same since

$$
\begin{aligned}
\mathbf{0} &= \det\left[\lambda \mathbf{I}_{3x3} - \mathbf{U}^{\mathrm{T}} \sum \mathbf{U} \underline{\mathbf{w}}\right] \\
&= \det\left[\lambda \mathbf{U}^{\mathrm{T}}\mathbf{U} - \mathbf{U}^{\mathrm{T}} \sum \mathbf{U}\right] \\
&= \det\left[\mathbf{U}^{\mathrm{T}}\right] \det\left[\lambda \mathbf{I}_{3x3} - \sum\right] \det[\mathbf{U}] \\
&= \det\left[\mathbf{U}^{-1}\right] \det\left[\lambda \mathbf{I}_{3x3} - \sum\right] \det[\mathbf{U}] \\
&= \det\left[\lambda \mathbf{I}_{3x3} - \sum\right]
\end{aligned}
\tag{A-11}
$$

exactly as before the rotation. The above steps carry over exactly for the n-dimensional case as well. When the **U** in the above is the normalized eigenvector matrix corresponding to **Σ,** then the new coordinate axes are again orthogonal and the ellipsoids' major and minor axes are aligned with the underlying coordinate axes. This corresponds to the original matrix **Σ** now being diagonalized in the resulting coordinate system. Notice that the result of evaluating the quadratic form on the right hand side of Eq. A-8 remains the same no matter what coordinate system it is expressed in (as is also the case for both determinates and traces of a matrix being invariant to similarity transformations).

Appendix B: Proofs of Lemmas associated with TeK's new Results

Proof of Lemma 1:
Block multiply out the following:

$$
\begin{bmatrix} \mathbf{P}_1 & \mathbf{P}_{12} \\ \mathbf{P}_{12}^{\mathrm{T}} & \mathbf{P}_2 \end{bmatrix} \begin{bmatrix} \mathbf{L} & \mathbf{M} \\ \mathbf{N} & \mathbf{Q} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_3 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_3 \end{bmatrix} \tag{B-1}
$$

to yield the following four equations to be solved as indicated next.

$$
\begin{aligned}
&\mathbf{P}_1\mathbf{L} + \mathbf{P}_{12}\mathbf{N} = \mathbf{I}_3, \\
&\mathbf{P}_1\mathbf{M} + \mathbf{P}_{12}\mathbf{Q} = \mathbf{0} \Rightarrow \overbrace{\left(\mathbf{P}_1^{-1}\mathbf{P}_1\right)}^{\mathbf{I}_3}\mathbf{M} = -\mathbf{P}_1^{-1}\mathbf{P}_{12}\mathbf{Q}, \\
&\mathbf{P}_{12}^{\mathrm{T}}\mathbf{L} + \mathbf{P}_2\mathbf{N} = \mathbf{0} \Rightarrow \overbrace{\left(\mathbf{P}_2^{-1}\mathbf{P}_2\right)}^{\mathbf{I}_3}\mathbf{N} = -\mathbf{P}_2^{-1}\mathbf{P}_{12}^{\mathrm{T}}\mathbf{L}, \\
&\mathbf{P}_{12}^{\mathrm{T}}\mathbf{M} + \mathbf{P}_2\mathbf{Q} = \mathbf{I}_3.
\end{aligned}
\tag{B-2}
$$

Combine the first with the result of the third above to yield an equation that may be solved for **L** as follows:

$$
\mathbf{P}_1\mathbf{L} + \mathbf{P}_{12}\left(-\mathbf{P}_2^{-1}\mathbf{P}_{12}^{\mathrm{T}}\mathbf{L}\right) = \mathbf{I}_3 \Rightarrow \mathbf{P}_1\mathbf{L} - \mathbf{P}_{12}\mathbf{P}_2^{-1}\mathbf{P}_{12}^{\mathrm{T}}\mathbf{L} = \left[\mathbf{P}_1 - \mathbf{P}_{12}\mathbf{P}_2^{-1}\mathbf{P}_{12}^{\mathrm{T}}\right]\mathbf{L} = \mathbf{I}_3 \Rightarrow \mathbf{L} = \left[\mathbf{P}_1 - \mathbf{P}_{12}\mathbf{P}_2^{-1}\mathbf{P}_{12}^{\mathrm{T}}\right]^{-1}. \tag{B-3}
$$

Similarly, combine the second with the result of the fourth above to yield an equation that may likewise be solved for **Q** as follows:

$$
\mathbf{P}_{12}^{\mathrm{T}}\left(-\mathbf{P}_1^{-1}\mathbf{P}_{12}\mathbf{Q}\right) + \mathbf{P}_2\mathbf{Q} = \mathbf{I}_3 \Rightarrow -\mathbf{P}_{12}^{\mathrm{T}}\mathbf{P}_1^{-1}\mathbf{P}_{12}\mathbf{Q} + \mathbf{P}_2\mathbf{Q} = \mathbf{I}_3 \Rightarrow \left[\mathbf{P}_2 - \mathbf{P}_{12}^{\mathrm{T}}\mathbf{P}_1^{-1}\mathbf{P}_{12}\right]\mathbf{Q} = \mathbf{I}_3 \Rightarrow \mathbf{Q} = \left[\mathbf{P}_2 - \mathbf{P}_{12}^{\mathrm{T}}\mathbf{P}_1^{-1}\mathbf{P}_{12}\right]^{-1}, \tag{B-4}
$$

and, finally, from the above original second and third results, respectively, we have that

$$\mathbf{M} = -\mathbf{P}_1^{-1}\mathbf{P}_{12}\mathbf{Q} \Rightarrow \mathbf{M} = -\mathbf{P}_1^{-1}\mathbf{P}_{12}\left[\mathbf{P}_2 - \mathbf{P}_{12}^{\mathrm{T}}\mathbf{P}_1^{-1}\mathbf{P}_{12}\right]^{-1},$$
$$\mathbf{N} = -\mathbf{P}_2^{-1}\mathbf{P}_{12}^{\mathrm{T}}\mathbf{L} \Rightarrow \mathbf{N} = -\mathbf{P}_2^{-1}\mathbf{P}_{12}^{\mathrm{T}}\left[\mathbf{P}_1 - \mathbf{P}_{12}\mathbf{P}_2^{-1}\mathbf{P}_{12}^{\mathrm{T}}\right]^{-1}. \quad \text{(B-5)}$$

That the two indicated inverses encountered above do in fact exist (by the necessary intermediate quantities being nonsingular) is established as intermediate steps in the proofs of Lemmas 3 and 4 coming next. Since the matrix inverse is unique, that which was to be proved has now been demonstrated.

A more general result for matrix partitioning and their corresponding inverses even when the original matrix is not symmetric is available in [22, p. 99].

Theorem 8.5.11. Let **T** represent an m x m matrix, U an m x n matrix, **V** an n x m matrix, and **W** an n x n matrix. Suppose that **T** is nonsingular. Then, $\begin{bmatrix}\mathbf{T} & \mathbf{U}\\ \mathbf{V} & \mathbf{W}\end{bmatrix}$, or, equivalently, $\begin{bmatrix}\mathbf{W} & \mathbf{V}\\ \mathbf{U} & \mathbf{T}\end{bmatrix}$, is nonsingular if and only if the n x n matrix

$$\mathbf{Q} = \mathbf{W} - \mathbf{V}\mathbf{T}^{-1}\mathbf{U} \quad \text{(B-6)}$$

is nonsingular, in which case

$$\begin{aligned}\begin{bmatrix}\mathbf{T} & \mathbf{U}\\ \mathbf{V} & \mathbf{W}\end{bmatrix}^{-1} &= \begin{bmatrix}\mathbf{T}^{-1} + \mathbf{T}^{-1}\mathbf{U}\mathbf{Q}^{-1}\mathbf{V}\mathbf{T}^{-1} & -\mathbf{T}^{-1}\mathbf{U}\mathbf{Q}^{-1}\\ -\mathbf{Q}^{-1}\mathbf{V}\mathbf{T}^{-1} & \mathbf{Q}^{-1}\end{bmatrix}\\ &= \begin{bmatrix}\mathbf{T}^{-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix} + \begin{bmatrix}-\mathbf{T}^{-1}\mathbf{U}\\ \mathbf{I}_n\end{bmatrix} Q^{-1}\left[-\mathbf{V}\mathbf{T}^{-1}, \quad \mathbf{I}_n\right]\end{aligned} \quad \text{(B-7)}$$

$$\begin{aligned}\begin{bmatrix}\mathbf{W} & \mathbf{V}\\ \mathbf{U} & \mathbf{T}\end{bmatrix}^{-1} &= \begin{bmatrix}\mathbf{Q}^{-1} & -\mathbf{Q}^{-1}\mathbf{V}\mathbf{T}^{-1}\\ -\mathbf{T}^{-1}\mathbf{U}\mathbf{Q}^{-1} & \mathbf{T}^{-1} + \mathbf{T}^{-1}\mathbf{U}\mathbf{Q}^{-1}\mathbf{V}\mathbf{T}^{-1}\end{bmatrix}\\ &= \begin{bmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{T}^{-1}\end{bmatrix} + \begin{bmatrix}\mathbf{I}_n\\ -\mathbf{T}^{-1}\mathbf{U}\end{bmatrix} Q^{-1}\left[\mathbf{I}_n, \quad -\mathbf{V}\mathbf{T}^{-1}\right]\end{aligned} \quad \text{(B-8)}$$

Proof appears in the reference cited above and is straightforward. **Q.E.D.**

Proof of Lemma 2: Please see result in [24]. **Q.E.D.**

Proof of Lemma 3 (all inequalities below are in the matrix positive definite or positive semi-definite sense):

$\mathbf{P}_2 > \mathbf{0}$, (all block diagonal elements of a positive definite P must also be positive definite),

$$\mathbf{P}_2^{-1} > \mathbf{0},$$

$\mathbf{P}_{12}\mathbf{P}_2^{-1}\mathbf{P}_{12}^{\mathrm{T}} \geq \mathbf{0}$, ($\mathbf{P}_{12}$ need not be of full rank since only seek positive semidefiniteness in this step),

$$-\mathbf{P}_{12}\mathbf{P}_2^{-1}\mathbf{P}_{12}^{\mathrm{T}} \leq \mathbf{0},$$
$$\mathbf{0} < \left[\mathbf{P}_1 - \mathbf{P}_{12}\mathbf{P}_2^{-1}\mathbf{P}_{12}^{\mathrm{T}}\right] < \mathbf{P}_1,$$
$$\left[\mathbf{P}_1 - \mathbf{P}_{12}\mathbf{P}_2^{-1}\mathbf{P}_{12}^{\mathrm{T}}\right]^{-1} > \mathbf{P}_1^{-1} > \mathbf{0}.$$

Q.E.D.

Proof of Lemma 4 (all inequalities below are in the matrix definite or semi-definite sense):

$\mathbf{P}_1 > \mathbf{0}$, (all block diagonal elements of a positive definite P must also be positive definite),

$$\mathbf{P}_1^{-1} > \mathbf{0},$$

$\mathbf{P}_{12}^{T}\mathbf{P}_{1}^{-1}\mathbf{P}_{12} \geq \mathbf{0}$, ($\mathbf{P}_{12}$ need not be of full rank since only seek positive semidefiniteness in this step),

$$-\mathbf{P}_{12}^{T}\mathbf{P}_{1}^{-1}\mathbf{P}_{12} \leq \mathbf{0},$$

$$\mathbf{0} < \left[\mathbf{P}_{2} - \mathbf{P}_{12}^{T}\mathbf{P}_{1}^{-1}\mathbf{P}_{12}\right] < \mathbf{P}_{2},$$

$$\left[\mathbf{P}_{2} - \mathbf{P}_{12}^{T}\mathbf{P}_{1}^{-1}\mathbf{P}_{12}\right]^{-1} > \mathbf{P}_{2}^{-1} > \mathbf{0}.$$

Q.E.D.

In the proof of both Lemmas 3 and 4, we invoked properties of Lemma 2 by requiring the matrix difference within brackets above to also be strictly positive definite as well by establishing that all the indicated inverses exist. This aspect is mentioned here to let the cognizant reader know that we are aware of the subtle distinction regarding a slight lapse of rigor in [7] along these lines in assuming that the intermediate inverses existed for submatrices $\mathbf{P}_1$ and $\mathbf{P}_2$. However, in the interest of expediency, we have slightly simplified just this step here that we have previously rigorously proved before in [24, Eq. 7] (being strictly positive definite when the original matrix is so) and merely quote in Lemma 2. If the matrix manipulations immediately above as steps in the proof are unfamiliar, they are proved rigorously in the Appendix of [27] that reports on one aspect of work done earlier [26]-[32]. Several historical numerical tests of positive definiteness/semi-definiteness contained in 1980's software code (in navigation systems and sonobuoy target trackers) is revealed in [25] to be fallacious.

References

1. Jazwinski, A., *Stochastic Processes and Filtering Theory*, Academic Press, NY, 1970.
2. Johnson, N. L., and Leone, F. C., *Statistics and Experimental Design: In engineering and the physical sciences*, Wiley Publications in Statistics, John Wiley & Sons, Inc., NY, 1964.
3. Fitzgerald, R. J., "Effects of Range-Doppler Coupling on Chirp Radar Tracking,'' *IEEE Trans. on Aerospace & Electronic Systems*, Vol. AES-10, No. 4, pp. 528-532, July 1974.
4. Fitzgerald, R. J., "On Reentry Vehicle Tracking in Various Coordinate Systems,'' *IEEE Trans. on Automatic Control*, Vol. AC-19, No. 5, pp. 581-582, July 1974.
5. Daum, F. E. and Fitzgerald, R. J., "Decoupled Kalman Filters for Phased Array Radar Tracking,'' *IEEE Trans. on Autom. Control*, Vol. AC-28, No. 3, pp. 269-283, Mar. 1983.
6. Paradowski, L. R., "Uncertainty Ellipses and Their Application to Interval Estimation of Emitter Position," *IEEE Trans. on Aerospace and Electronic Systems*, Vol. 33, No. 1, pp. 126-133, January 1997.
7. Paul B. Liebelt, *An Introduction to Optimal Estimation*, Addison-Wesley Publishing Company, Reading, MA, 1967.
8. Abramowitz, M., Stegun, I. A., *Handbook of Mathematical Functions,* Applied Mathematics Series 5, National Bureau of Standards, Washington, D.C., 1964.
9. Kerr, T. H., Chin, L., "The Theory and Techniques of Discrete-Time Decentralized Filters," in *Advances in the Techniques and Technology in the Application of Nonlinear Filters and Kalman Filters*, C.T. Leondes (ed.), NATO Advisory Group for Aerospace Research and Development, AGARDograph No. 256, Noordhoff International Publishing, Lieden, 1981.
10. McGarty, T. P., *Stochastic Systems and State Estimation*, Wiley-Interscience, NY, 1974.
11. Bowker, A. H., Lieberman, G. J., *Engineering Statistics*, 2nd Ed., Prentice-Hall, Englewood Cliffs, NJ, 1972.
12. Bar-Shalom, Y., Li, Xiao-Rong, *Multitarget-Multisensor Tracking: Principles and Techniques*, YBS Publishing, 1995.
13. Bar-Shalom, Y., Li, Xiao-Rong, *Estimation and Tracking: Principles, Techniques, and Software*, Artech House, Boston, MA, 1993.

14. Kerr, T. H., “A New Multivariate Cramer-Rao Inequality for Parameter Estimation (Application: Input Probing Function Specification),” *Proceedings of IEEE Conference on Decision and Control*, Phoenix, AZ, pp. 97-103, December 1974.
15. Kerr, T. H., “Status of CR-Like Lower bounds for Nonlinear Filtering,” *IEEE Transactions on Aerospace and Electronic Systems*, Vol. AES-25, No. 5, pp. 590-601, September 1989 (Author's reply in Vol. AES-26, No. 5, pp. 896-898, September 1990).
16. Kerr, T. H., “Cramer-Rao Lower Bound Implementation and Analysis: CRLB Target Tracking Evaluation Methodology for NMD Radars,” MITRE Technical Report, Contract No. F19628-94-C-0001, Project No. 03984000-N0, Bedford, MA, February 1998.
17. Kerr, T. H., “Developing Cramer-Rao Lower Bounds to Gauge the Effectiveness of UEWR Target Tracking Filters,” Proceedings of AIAA/BMDO Technology Readiness Conference and Exhibit, Colorado Springs, CO, 3-7 August 1998.
18. Kerr, T. H., UEWR *Design Notebook-Section 2.3: Track Analysis*, TeK Associates, Lexington, MA, (for XonTech, Hartwell Rd, Lexington, MA), XonTech Report No. D744-10300, 29 March 1999.
19. Kerr, T. H., “Perspectives on New Tracking Methods as Compared to Current NMD Use,” prepared for *Missile Defense Conference* to occur in 4-7 March 2003 and submitted to *IEEE Proceedings: Special Issue on Recursive Estimation*, Ed. By Simon Haykin, (in review and expected to appear in 2004).
20. Kerr, T. H., “Apparent Drawbacks that Exist in the Underlying Structure of Covariance Intersection.” submitted to *IEEE Transactions on Automatic Control* in 2003.
21. Satz, H. S., Kerr, T. H., “Comparison of Batch and Kalman Filtering for Radar Tracking,” *Proceedings of 10th Annual AIAA/BMDO Conference*, Williamsburg, VA, 25 July 2001 (Unclassified paper but Conference Proceedings are SECRET).
22. David A. Harville, *Matrix Algebra from a Statis-tician’s Perspective*, Springer, NY, 1997.
23. Kerr, T. H., “Considerations in whether to use Marquardt Nonlinear Least Squares vs. Lambert Algorithm for NMD Cue Track Initiation (TI) calculations,” TeK Associates Technical Report No. 2000-101, Lexington, MA, (for Raytheon, Sudbury, MA), 27 September 2000.
24. Kerr, T. H., “On Misstatements of the Test for Positive Semi-definite Matrices,” *AIAA Journal of Guidance, Control, and Dynamics*, Vol. 13, No. 3, pp. 571-572, May-June 1990.
25. Kerr, T. H., “Fallacies in Computational Testing of Matrix Positive Definiteness/Semi-definiteness,” *IEEE Transactions on Aerospace and Electronic Systems*, Vol. AES-26, No. 2, pp. 415-421, March 1990.
26. Kerr, T. H., ``A Two Ellipsoid Overlap Test for Real-Time Failure Detection and Isolation by Confidence Regions," *Proceedings of IEEE Conference on Decision and Control*, Phoenix, AZ, December 1974.
27. Kerr, T. H., “Real-Time Failure Detection: A Static Nonlinear Optimization Problem that Yields a Two Ellipsoid Overlap Test,” *Journal of Optimization Theory and Applications*, Vol. 22, No. 4, August 1977.
28. Kerr, T. H., “Statistical Analysis of a Two Ellipsoid Overlap Test for Real-Time Failure Detection,” *IEEE Transactions on Automatic Control*, Vol. AC-25, No. 4, August 1980.
29. Kerr, T. H., “False Alarm and Correct Detection Probabilities Over a Time Interval for Restricted Classes of Failure Detection Algorithms,” *IEEE Transactions on Information Theory*, Vol. IT-28, No. 4, pp. 619-631, July 1982.
30. Kerr, T. H., “Poseidon Improvement Studies: Real-Time Failure Detection in the SINS/ESGM,” TASC Report TR-418-20, Reading, MA, June 1974 (Confidential) for Navy, SP-24.
31. Kerr, T. H., “Failure Detection in the SINS/ESGM System,” TASC Report TR-528-3-1, Reading, MA, July 1975 (Confidential) for Navy, SP-24.
32. Kerr, T. H., “Improving ESGM Failure Detection in the SINS/ESGM System (U),” TASC Report TR-678-3-1, Reading, MA, October 1976 (Confidential) for Navy, SP-24.

33. Farina, A., Ristic, B., Timmoneri, L., "Cramer-Rao Bound for Nonlinear Filtering with Pd < 1 and its Application to Target Tracking," *IEEE Trans. on Signal Processing*, Vol. 50, No. 8, pp. 1916-1924, Aug. 2002.
34. Farina, A., Ristic, B., Benvenuti, D., "Tracking a Ballistic Target: Comparison of Several Nonlinear Filters," *IEEE Trans. on AES*, Vol. 38, No. 3, pp.854-867, July 2002.
35. Alvin C. Rencher, *Multivariate Statistical Inference and Applications*, John Wiley, Series in Probability and Statistics, NY, 1998.
36. Alfano, S., Greer, M. L., "Determining if Two Solid Ellipsoids Intersect," *AIAA Journal of Guidance, Control, and Dynamics*, Vol. 26, No. 1, pp. 106-110, Jan.-Feb. 2003.