Overbounding False-Alarm Probability for a Chi-Square Monitor with Natural Biases

Jason H. Rife, Tufts University

ABSTRACT

This paper presents methods for overbounding the false-alarm probability of chi-square monitors with natural biases. A significant example in navigation is the signal deformation monitor (SDM), which is a critical component of ground-based augmentation systems (GBAS) and satellite-based augmentation systems (SBAS) for GPS. The key contribution of the paper is a proof that a simple analytical bound for false-alarm risk can be guaranteed conservative, even in realistic circumstances when the mean and covariance of the monitor’s inputs are poorly characterized.

INTRODUCTION

Aviation sensing systems must meet strict requirements for continuity risk, which is the likelihood that the sensing system becomes unavailable during a safety-critical operation. Continuity loss occurs if the radio signals or data channels needed for navigation are lost. Continuity breaks may occur, for example, when a satellite sets or when signal tracking fails abruptly due to blockage by physical structures, unintentional radio-frequency interference (RFI), or intentional jamming. In a safety of life system, where signal quality is continually verified by algorithms called integrity monitors, continuity loss may also occur when an integrity monitor deems that the signal is anomalous and that its continued use would pose a threat to life or property.

Because integrity monitors are subject to noise, sometimes they may detect an anomaly when no anomaly actually exists. Such events are called false alarms. Though false alarms are, in a sense, avoidable, there is always a tradeoff between the sensitivity of the monitor and its rate of false alarms. This tradeoff is exacerbated by the short response time required for most safety-of-life applications. Cases in which a monitor fails to detect an anomaly within this short period after the anomaly occurs are called missed detections. From a practical point of view, it is necessary to tune monitor sensitivity to obtain a balance between false alarms and missed detections.
In order to avoid missed detections, safety-critical systems are engineered to be as sensitive as possible without violating continuity requirements. Continuity breaks interrupt operations and may pose a safety hazard. For instance, loss of continuity during a precision aircraft landing triggers an emergency response, typically a balked landing procedure in which the aircraft attempts to regain altitude and perform a go-around maneuver. As such events are disconcerting and somewhat dangerous, system specifications impose a maximum probability of continuity loss. (e.g., For a Category I landing, continuity risk is limited to about $10^{-5}$ per 30 sec.)

In order to demonstrate that continuity requirements can be met, rigorous bounds are needed to compute conservative estimates of risk under various environmental conditions. Within the navigation field, the process of developing rigorous risk bounds is called overbounding. Overbounding position-error distributions has been the subject of intense prior research [1]-[9]. In recent years, a particular emphasis has been placed on overbounding for integrity monitors, too, with the goal of obtaining rigorous bounds for false-alarm and missed-detection risks [10]-[13]. As experience has been gained in operating integrity augmentations for GPS, such as ground-based augmentation systems (GBAS) and satellite-based augmentation systems (SBAS) [14]-[17], it has become increasingly clear that error distributions for integrity monitoring algorithms are often poorly characterized and that monitor performance analyses must take this fact into account. Mitigating environmental uncertainty is particularly important for monitors that operate on a vector of input signals, when analysis is highly dependent on quantities such as the relative variances of each signal and the correlation between signal pairs [11].

Another important issue is that some integrity monitors must process biased signals. Ideally, biases could be mitigated via calibration in advance or via online estimation. Unfortunately, mitigation is not always possible, particularly when biases are small (but not insignificant) and vary strongly over time. Such biases are present, for example, in signal deformation monitoring (SDM). SDM is an algorithm used by GBAS and SBAS to detect deformations of the correlator peak used for code tracking [18]-[20]. Deformations of the correlator peak may be indicative of severe multipath or, worse, of faulty electronics on board a satellite vehicle.

Though severe signal deformation faults are rare [18], small nominal deformations have now been observed on all GPS satellites [21]-[26]. Nominal deformations were initially found using a high-gain dish antenna [21] but can now also be detected by applying appropriate signal processing techniques to large data sets acquired using conventional GNSS antennas and receivers [25]. These biases are sometimes called natural biases, because they are present at all times, even under nominal conditions. Specific effects are different for different receivers, even for different instances of the same receiver model, due to manufacturing variations in the receiver front-end electronics. Because of the relatively small magnitude and unit-to-unit
variability of this bias, the bias is extremely difficult to remove; nonetheless, nominal signal deformation biases are large enough to have a significant impact on the SDM statistic, noticeably increasing false-alarm risk above predicted levels.

The main purpose of this paper is to derive a model for false-alarm risk for integrity monitors subject both to natural biases and to random noise modeled by parameters that are themselves uncertain. In particular, the paper focuses on chi-square integrity monitors, like SDM, which quadratically combine a vector of noisy signals into a single monitor statistic. A bound for false-alarm risk in chi-square monitors has previously been derived for the case when natural biases are zero [13]; however, no prior work has identified a bound for false-alarm risk when natural biases are present. This paper shows that deriving such a bound is possible; however, the derivation presented here to bound natural biases is necessarily more complicated than the bias-free bound presented in [13].

The paper is organized as follows. A first section discusses models of the monitor noise, including the effects of uncertain model parameters and biases. A subsequent section introduces a rigorous bound for false-alarm risk that accounts for this noise model. A third section discusses the implications of the bound for integrity-monitor applications. Following a brief conclusion, the Appendix offers a detailed proof that the proposed method for bounding false-alarm risk is, indeed, conservative.

NOISE MODEL

This section provides a model of the random signal processed by a chi-square monitor.

Chi-square monitors operate by computing a monitor statistic $m \in \mathbb{R}^1$ from a regularized signal vector $y \in \mathbb{R}^N$.

$$m = y'y$$

(1)

In the presence of an anomaly, the mean value of $y$ grows large, and so the anomaly can be detected by continually checking if $m$ ever exceeds a threshold $T$. In the absence of an anomaly, random noise may still occasionally cause $m$ to exceed $T$. Such an event is called a false alarm.

Analysis of false-alarm risk requires that the distribution of the random signal vector $y$ be modeled. This distribution of $y$ is the result of a regularization process, which conditions raw data signals [11]. The regularization process decorrelates and normalizes raw signals so that the elements of $y$ are nominally unbiased, independent and of approximately equal variance. This is to say that, under ideal conditions, the regularization process would result in $y$ having a distribution with a mean equal to zero and a covariance matrix equal to the identity matrix $I$. 

3
As discussed in [11], the regularization process is always somewhat inexact and may sometimes be very imprecise. In the presence of natural biases, for instance, the mean \( \mu = E[y] \) can never be made to be zero. Likewise, when the covariance of the raw data varies strongly over time, the covariance \( Q = E[y'y] \) cannot be regularized to be the identity matrix. In practice, in developing a monitor like SDM, these deviations of \( \mu \) and \( Q \) from their ideal values are significant. Deviations from nominal must therefore be modeled in estimating the probability of false alarms, where \( m \) exceeds \( T \) in the absence of an anomaly.

In order to account for these nonideal effects, consider that the false-alarm probability \( P_{fa} \) is

\[
P_{fa} = \int_{T}^{\infty} p(m)dm ,
\]

where \( p(m) \) is the probability density function of the monitor statistic \( m \). Equivalently,

\[
P_{fa} = 1 - \int_{0}^{T} p(m)dm
\]

Assuming that \( y \) is Gaussian distributed, the integral term in the above expression is simply the probability that the Gaussian random vector lies inside a spherical threshold. In other words, the false-alarm probability (3) can be written

\[
P_{fa} = 1 - P_{gr2} (T; Q, \mu)
\]

where the Gaussian integral is

\[
P_{gr2} (T; Q, \mu) = \int_{\{y'y < T\}} 2\pi |Q|^{-\frac{1}{2}} e^{-\frac{1}{2}(y'Q^{-1}y'} 1 dy'.
\]

The label \( P_{gr2} \) refers to generalized chi-square distribution, which is a common name for this integral. The integral is a function of the threshold \( T \) and the covariance and mean parameters (\( Q \) and \( \mu \)) that describe the probability distribution over \( y \).

The primary application of (4) is to solve for the threshold \( T \) that keeps the false-alarm risk at an acceptable level. Unfortunately, inverting (4) to compute the threshold \( T \) is complicated by the fact that the actual covariance \( Q \) and mean \( \mu \) are unknown (despite regularization trying to make these values approximately \( I \) and \( 0 \), respectively). If a reasonable range of possible \( Q \) and \( \mu \) values can be identified, however, then a conservative estimate of \( T \) can be obtained that meets the desired \( P_{fa} \) specification for all these values of \( Q \) and \( \mu \).
The goal of the next section will be to identify a conservative estimate of $T$ assuming that the range of allowed values of $Q$ and $\mu$ are defined as follows. The natural bias $\mu$ is allowed to take any value so long as its magnitude is less than $b_{\text{max}}$.

$$\|\mu\| \leq b_{\text{max}} \quad (6)$$

The covariance $Q$ is allowed to take any value such that its largest eigenvalue does not exceed one. This constraint is equivalent to a comparison of positive definite matrices:

$$Q \prec I \quad (7)$$

Here the curly comparison means that $x^T Q x \leq x^T I x$ for all vectors $x \in \mathbb{R}^N$.

**PROPOSED BOUND**

This section presents a conservative method to estimate threshold for a specified false-alarm risk given natural biases and uncertain covariance, with unknown parameters bounded on the ranges given by (6) and (7).

The final result of this analysis is a simple equation for the threshold $T$. The equation computes a conservative estimate of $T$ from the maximum-allowable false-alarm risk $r_{fa}$. The equation is written in terms of the function $P_{ncx}^{-1}$, which is the inverse of the noncentral chi-square cumulative distribution with $N$ degrees of freedom and noncentrality parameter $b_{\text{max}}^2$:

$$T = P_{ncx}^{-1} \left( 1 - r_{fa} \cdot N, b_{\text{max}}^2 \right) \quad (8)$$

This equation can easily be evaluated using standard analysis software, for example, using the $\text{ncx2inv}$ command in Matlab.

To demonstrate that the threshold $T$ is conservative, consider that the maximum-allowable false-alarm risk $r_{fa}$ is a system specification, where the true false-alarm risk $P_{fa}$ must satisfy

$$r_{fa} \geq P_{fa} \quad (9)$$

Because it is difficult to evaluate the true false-alarm risk $P_{fa}$ over all possible values of the unknown parameters $Q$ and $\mu$, it is useful to introduce a conservative approximation $\overline{P}_{fa}$, where

$$\overline{P}_{fa} \geq P_{fa} \quad (10)$$
As long as the approximation $P_{fa}$ can be guaranteed to be an upper bound for $P_{fa}$, then requirement (9) can be satisfied by setting

$$P_{fa} = r_{fa}. \quad (11)$$

The next step is to define an approximation that ensures $P_{fa} \geq P_{fa}$. One way to do this is to examine (4) and introduce an analogous definition for $P_{fa}$.

$$P_{fa} = 1 - P_{gr2}(T). \quad (12)$$

Here the approximation $P_{gr2}(T)$ is a lower bound for the generalized chi-square integral (5) over all possible values of $Q$ and $\mu$. As long as

$$P_{gr2}(T) \leq P_{gr2}(T; Q, \mu), \quad (13)$$

then $P_{fa} \geq P_{fa}$, as can be seen by substituting (3) and (12) into (10). The important detail is that the negative sign in front of

$P_{gr2}$ means that a lower bound for the integral $P_{gr2}$ results in an upper bound for $P_{fa}$.

A relatively tight lower-bound for $P_{gr2}$ is

$$P_{gr2}(T) = P_{ncx} \left( \frac{\sqrt{T} - b}{\sqrt{\lambda_{\max}} + b} \right)^2; N, b^2 \right) \quad (14)$$

where $\lambda_{\max}$ is the largest eigenvalue of $Q$, where the degrees of freedom $N$ match the dimension of $Q \in \mathbb{R}^{N \times N}$, and where the magnitude $b$ of the natural bias is required to be smaller than $\sqrt{T}$. A proof for this result is provided in the Appendix (Theorem 3).

As long as the largest eigenvalue of $Q$ is no larger than one, as required by (7), then setting $\lambda_{\max} = 1$ simplifies the above equation to be

$$P_{gr2}(T) = P_{ncx} (T; N, b^2). \quad (15)$$
It is well known that the magnitude of the noncentral chi-square distribution grows opposite the noncentrality parameter $b^2$. That is

$$\frac{\partial P_{ncx}(T;N,b^2)}{\partial(b^2)} \leq 0.$$  \hfill (16)

Hence a lower-bound $P_{na2}$ can be obtained if $b$ is set to its highest allowed value, $b = b_{\text{max}}$. Inserting this worst-case bias into (15) and (11) into (12) gives

$$r_{fa} = 1 - P_{ncx}(T;N,b_{\text{max}}^2).$$  \hfill (17)

Rearranging terms, it is possible to solve for $T$ as described by equation (8), above.

**DISCUSSION**

The formulation presented in the prior section for obtaining a threshold $T$ is reasonably intuitive. In essence, the idea is that the false-alarm probability can be conservatively estimated from an overbound distribution, where the actual distribution of the input signal (characterized by $Q$ and $\mu$) is replaced by an inflated distribution, where all the principal axes of the covariance matrix $Q$ are expanded to their largest allowed value (unity) and where the bias magnitude is assigned to its largest possible value $b_{\text{max}}$. Though this concept applies to a vector signal, it is thematically consistent with overbounding methods developed for one-dimensional Gaussian distributions in earlier work, such as [1].

This “inflation” analogy should not be taken for granted. Missed-detection risk is quite different from false-alarm risk, in that an overbound for missed-detection risk $P_{md}$ depends not only on an inflated upper bound for covariance $Q$, but also on a deflated lower bound. Consider the overbound for $P_{md}$ found in reference [13]:

$$P_{md} = P_{ncx}\left(\frac{\lambda_{\text{max}} - \lambda_{\text{min}}}{\lambda_{\text{min}}} T;\text{DoF.}\left(\frac{b \sqrt{T}}{\lambda_{\text{max}} T} + \frac{\sqrt{\lambda_{\text{max}} T}}{\lambda_{\text{min}}}\right)^2\right).$$  \hfill (18)

This result depends both on a largest eigenvalue $\lambda_{\text{max}}$ and on a smallest eigenvalue $\lambda_{\text{min}}$ for the covariance $Q$. In this light, it is a surprising (and useful) result that (14), by comparison, requires only and inflated upper bound and not a deflated lower bound for $Q$.  


The false-alarm bounding formulation presented in this paper has practical implications for designing monitors in the presence of natural biases, such as in the case of SDM. The essential result is that larger natural biases require larger thresholds. The dependence of threshold on natural bias size can best be understood with a plot. For clarity in plotting, the threshold $T$ computed by (8) is compared to a nominal value $T_{nom}$, which is the value when bias size is zero. In the plot, bias size is similarly normalized, with the bias magnitude $b_{max}$ taken as a ratio with $\sqrt{T_{nom}}$. Using these normalizations, threshold is plotted as a function of bias magnitude in Fig. 1 for the case where $r_{fa}$ is set to a characteristic value ($r_{fa} = 10^{-7}$) and where the number of elements in the random vector $\mathbf{y}$ is seven ($N = 7$).

![Fig. 1. Conservative threshold (normalized) as a function of natural-bias magnitude.](image)

The figure shows that the threshold inflation is modest when the bias is small. When the bias is 10% of the nominal threshold or smaller ($b_{max} / \sqrt{T_{nom}} < 0.1$) the threshold inflation is less than 6% ($T/T_{nom} < 1.06$). However, when the bias grows large, the threshold-inflation penalty rapidly grows large. For example, as $b_{max}$ approaches $T_{nom}$, the inflation is more than 200% ($T/T_{nom} > 3$).

It is important to keep in mind that the proposed formulation for $T$ is tight but still somewhat conservative, in the sense that a worst-case bias $\mathbf{u}$ is assumed. As discussed in the Appendix (Theorem 3), the worst case occurs when the bias direction aligns with the direction of the largest principal axis of the covariance matrix $\mathbf{Q}$. In a real monitor, it is likely that this worst-case condition will not, in fact, be observed. As such, it is possible that a tighter estimate of $T$ might be obtained by tuning the threshold to real data rather than by applying the conservative estimate (8). The utility of (8) is nonetheless justified by its...
simplicity. It is much easier to obtain the threshold $T$ by evaluating equation (8) than to obtain $T$ by analyzing vast data sets, which grow to be quite large when $r_{fa}$ is small.

Ultimately, the bottom line is that the inflation introduced using (8) is small when the natural bias is modest, as illustrated in Fig. 1. Given that, typically, natural bias magnitude is a relatively small fraction of the threshold, the required inflation for using (8) is much smaller than inflation introduced for other effects, such as inflation for heavy-distribution tails in SDM. In such cases, equation (8) has great utility as a practical yet rigorous method for setting the threshold of a chi-square integrity monitor.

**SUMMARY**

This paper considers chi-square integrity monitors that process random signal vectors characterized by uncertain parameters (i.e. mean and covariance). Of particular interest are non-zero mean values that may occur during nominal, fault-free operation. This class of so-called natural biases appears in Signal Deformation Monitoring (SDM) used to ensure the integrity of GNSS signals. To date, no rigorous method has previously been developed to account for natural biases in developing conservative estimates of monitor false-alarm risk.

This paper introduces a method to evaluate false-alarm risk accounting for unknown but bounded natural bias and covariance parameters, with the goal of identifying a conservative threshold. The end result is a relatively simple and intuitive formulation for threshold, where threshold can be evaluated by inverting a noncentral chi-square distribution.

**ACKNOWLEDGEMENTS**

The authors gratefully acknowledge the Federal Aviation Administration GBAS Program (Grant FAA-10-G-006) for supporting this research. The opinions discussed here are those of the authors and do not necessarily represent those of the FAA or other affiliated agencies.
APPENDIX

The results presented in the body of the paper depend strongly on (14), which was presented without proof. A rigorous proof is presented here, in the Appendix. The final result, Theorem 3, depends on two intermediate results, which are presented as Theorems 1 and 2.

Theorem 1: Family of Common-Extremum Ellipsoids. An ellipsoid can be described as the set of points $z \in \mathbb{R}^N$ that comprise the surface $\partial \Theta$ where:

$$\partial \Theta = \left\{ z \mid (z - \mu_c)^T \Lambda^{-1} (z - \mu_c) = \beta \right\}. \quad (19)$$

Consider the family of ellipsoids that all share the same shape and orientation (defined by a diagonal covariance matrix $\Lambda \in \mathbb{R}^{N \times N}$) and that all pass through a common point $z_* \in \partial \Theta$. Assume that members of the family have center locations $\mu_c \in \mathbb{R}^N$ and sizes $\beta \in [0, \infty)$ defined by

$$\mu_c = \left( I - \frac{\Lambda}{\gamma_*} \right) z_* \quad (20)$$

and

$$\beta = \frac{z_*^T \Lambda z_*}{\gamma_*^2}. \quad (21)$$

Define $\lambda_{\text{max}}$ to be the largest element of $\Lambda$. So long as the scalar parameter $\gamma_*$ falls on the domain $\gamma_* \in [\lambda_{\text{max}}, \infty)$, then no point on any ellipsoid $\partial \Theta$ is farther from the origin than $z_*$. 

Context: In order to define a lower bound $P_{\text{edr}}$, it will be necessary to identify related ellipsoids contained within a sphere. Theorem 1 is useful in that it defines such a family of ellipsoids, all of which are contained within a sphere of radius $\|z_*\|$ and, moreover, which are nested within each other. A visualization of a representative family of such ellipsoids is illustrated in Fig. 2.
Fig. 2. A family of ellipsoids (light green) all sharing the same outermost point $\mathbf{z}_*$. This particular family is illustrated for a point $\mathbf{z}_*$ (small red circle) rotated 0.4 radians from the x-axis around the unit circle (dark blue).

Proof: Consider a point $\mathbf{z}_* \in \mathbb{R}^N$ that lies on an ellipsoid $\partial \Theta$ with shape defined by the covariance matrix $\mathbf{\Lambda}$. For this point to be at least as far from the origin as any other point on $\partial \Theta$, the point $\mathbf{z}_*$ must maximize $||\mathbf{z}||^2$ subject to the ellipsoid constraint (19). This maximization problem can be solved by defining an objective function $J$ that combines the square-distance $||\mathbf{z}||^2$ with the ellipsoid constraint, via the help of a Lagrange multiplier $\gamma_c \in \mathbb{R}^1$.

$$J = \mathbf{z}_*^T \mathbf{z} - \gamma_c \left( (\mathbf{z} - \mathbf{\mu}_*)^T \mathbf{\Lambda}^{-1} (\mathbf{z} - \mathbf{\mu}_*) - \beta \right)$$

(22)

From this objective function, it is possible to derive expressions for the ellipsoid size $\beta$ and center $\mathbf{\mu}_*$. Before deriving these expressions, it is first important to acknowledge that there may be more than one global maximum for objective function (22), as two or more global maxima may be equally distant from the origin.

$$\mathbf{z}_* \in \{ \text{argmax}(J) \}$$

(23)

The domain of $\mathbf{z}$ is unbounded, so all global maxima $\mathbf{z}_*$ (and more generally, all extrema $\mathbf{z}_{ex}$) for the objective function must reside where the slope of $J$ is zero:

$$\nabla J|_{\mathbf{z}_*} = \mathbf{0}.$$  

(24)

Setting the gradient of (22) equal to the zero vector gives
\[ 2z_{ex} - 2\gamma_c \Lambda^{-1}(z_{ex} - \mu_c) = 0. \]  

(25)

Rearranging terms of the above equation gives

\[ \mu_c = \left( I - \frac{\Lambda}{\gamma_c} \right) z_{ex}. \]  

(26)

Given that \( z_{ex} \in \partial \Theta \), equation (26) can be combined with (19), to show that

\[ \beta = \frac{z_{ex}^T \Lambda z_{ex}}{\gamma_c^2}. \]  

(27)

Thus we have expressions for the ellipsoid center \( \mu_c \) and magnitude \( \beta \). These expressions are written in terms of the location of an extremum (e.g. \( z_{ex} = z_* \)), the covariance matrix \( \Lambda \), and the Lagrange multiplier \( \gamma_c \).

In other words, for a given ellipsoid shape \( \Lambda \), there may be multiple ellipsoids with different center locations and sizes that all pass through the same global maximum point \( z_* \). Each of the ellipsoids in this family is associated with a different scalar value \( \gamma_c \).

Not all values of \( \gamma_c \) necessarily result in ellipsoids where \( z_* \) is a global maximum, however. For \( z_* \) to be a global maximum, \( z_* \) must fall in the same orthant as \( \mu_c \) (meaning that the \( i \)-th elements of each vector either have the same sign, or at least one of the pair is zero).

To prove that this is true, define a set of indices \( i \in \Gamma_{\text{opp}} \) that includes all indices for vector element pairs that are non-zero and have opposite signs.

\[ \Gamma_{\text{opp}} = \{ i | \mu_{i,j}z_{i,j} < 0 \} \]  

(28)

The vectors \( z_* \) and \( \mu_c \) fall in the same orthant if \( \Gamma_{\text{opp}} \) is empty. Now construct a vector \( z_r \) that reflects only the \( \Gamma_{\text{opp}} \) elements of \( z_* \) about the ellipse center \( \mu_c \). The \( i \)-th element of \( z_r \) is

\[ z_{r,i} = \begin{cases} 
  z_{i,j} & i \not\in \Gamma_{\text{opp}} \\
  -z_{i,j} + 2\mu_{i,j} & i \in \Gamma_{\text{opp}}
\end{cases}. \]  

(29)
By construction, $z_r$ and $\mu$ are in the same orthant. The magnitude of any reflected element is increased. This is true because $\mu_{c,i}z_{c,i} < 0$ for elements in $\Gamma_{opp}$, and so $|z_{c,i} + 2\mu_{c,i} | > |z_{c,i}|$ for those elements. By extension, the magnitude of $z_r$ is

$$
\|z_r\| = \sqrt{\sum_{i \in \Gamma_{opp}} \left( |z_{c,i}| + 2|\mu_{c,i}| \right)^2 + \sum_{i \in \Gamma_{opp}} |z_{c,i}|^2}.
$$

(30)

From this equation, it is evident that $\|z_r\| = \|z_r\|$ if $\Gamma_{opp}$ is empty and $\|z_r\| > \|z_r\|$ otherwise. Because the magnitude of $z_r$ must be larger than that of all other points on the ellipsoid, $\Gamma_{opp}$ must be empty for $z_r$ to be a global maximum. Hence $z_r$ must lie in the same orthant as $\mu$. As shorthand, this result will be called the same-orthant condition.

The same-orthant condition limits the allowable range of $\gamma_c$ for which $z_r$ is a global maximum. To see this, consider (26). If the extremum $z_{es}$ in (26) is to be a global maximum $z_r$, the same-orthant condition requires corresponding elements of $z_{es}$ and $\mu$ not to have opposite signs. Since $\Lambda$ is a diagonal covariance matrix, all its diagonal elements $\lambda_i$ are positive and its off-diagonal elements are zero, and so the same-orthant constraint implies that $1 - \lambda_i / \gamma_c \geq 0$ for each $i$. Considering all elements $i$, the two vectors can only share the same orthant if $\gamma_c \geq \lambda_S$ and

$$
\lambda_S = \max \left\{ \lambda_j \mid z_{c,j} \neq 0 \right\}.
$$

(31)

Generally speaking, $\lambda_S$ is simply the largest eigenvalue $\lambda_{max}$, except in special cases when one or more elements of $z_r$ are zero. In those cases, $\lambda_S$ is the largest eigenvalue with a nonzero associated element in $z_r$.

Thus $\gamma_c \geq \lambda_S$ is a necessary condition for $z_r$ to be a global maximum. Although it is necessary to select the Lagrange multiplier parameter on this specified range $\gamma_c \in [\lambda_S, \infty)$, this alone is not a sufficient condition to guarantee $z_r$ is a global maximum. The reason is that we have only guaranteed that $z_r$ is an extremum located in the correct orthant (same orthant as $\mu$). We do not yet know how many such extrema might exist.

To develop a sufficient condition, it is necessary to invert the problem and determine how many extrema might share the same orthant with $\mu$. To look for other extrema in the same orthant as $\mu$, consider a particular ellipsoid (19) characterized by the
parameters $\mu, \beta$, and $\Lambda$. Conceptually these values are set by specifying a baseline value of $z$, and a baseline Lagrange multiplier $\gamma_e = \gamma_e$. Now invert (26) and (27) to obtain other extrema $z_{ex}$ and Lagrange multipliers $\gamma_e$ that are solutions to these nonlinear equations. When inverting (26) and (27), the number of unknown scalars $(N+1$ including $\gamma_e$ and elements of $z_{ex})$ is matched to the number of scalar equations $(N+1$ including the $\beta$ equation and each row of the $\mu$ equation). Because the equations are nonlinear, multiple solutions may exist. By construction, at least one solution is known to exist (i.e., $z$, and $\gamma_e$).

To determine if multiple solutions to (26) and (27) exist, substitute the former into the latter to eliminate $z_{ex}$. This substitution gives one scalar equation for the unknown $\gamma_e$:

$$\beta \gamma_e^2 = \mu^T \left( I - \frac{\Lambda}{\gamma_e^2} \right) \Lambda \left( I - \frac{\Lambda}{\gamma_e^2} \right) \mu_e.$$ (32)

Recognizing that $I$ and $\Lambda$ are diagonal, and invoking the same-orthant condition, equation (32) can be rewritten:

$$\beta = \sum_{i \in \Gamma_{\gamma}} \frac{\lambda_i \mu_i^2}{(\gamma_e - \lambda_i)^2} + c_s \text{ for } \gamma_e \geq \lambda_s.$$ (33)

This equation is constructed to avoid singular terms in the summation by moving them into

$$c_s = \frac{1}{\lambda_s} \sum_{j \in \Gamma_{z}} z_{ex,j}^2.$$ (34)

The indices in the set

$$\Gamma_s = \left\{ j \mid \lambda_j = \gamma_e \right\},$$ (35)

are those that would be singular if included in the first summation of (33). The singularity is avoided by applying (27) directly for each of these elements and moving the result to the second summation labeled $c_s$. Note that $\Gamma_s$ is empty except on the domain boundary, where $\gamma_e = \lambda_{max}$. On this boundary, the rows of (26) corresponding to the indices in $\Gamma_s$ are all zero, so the corresponding $z_{ex,j}$ values are unconstrained. By contrast, on the domain interior, where $\gamma_e > \lambda_s$, $\Gamma_s$ is empty, and so the system of equations is nonsingular and (26) can be used to obtain all elements of the vector $z_{ex}$. 

14
Analyze the domain interior, where \( \gamma > \lambda_s \). On the interior, \( c_s \) is zero and the solution to (26) and (27) is nonsingular.

Importantly, equation (33) is monotonic in \( \beta \), with \( \beta \) decreasing to zero as \( \gamma \) is increased from \( \lambda_s \) to infinity. Because the function is monotonic, there can exist at most one consistent solution for \( \gamma \) on \( \gamma > \lambda_s \). Hence there is at most one extremum \( z_{\text{ex}} \) associated with values of \( \gamma \) on the domain interior.

Now analyze the domain boundary, where \( \gamma = \lambda_s \). On this boundary, the right side of (26) is zero for all indices \( \Gamma_s \). If any \( \mu_{s,j} \neq 0 \) for \( j \in \Gamma_s \), then a contradiction exists, because the left side of the equation would be nonzero. If all \( \mu_{s,j} = 0 \) for \( j \in \Gamma_s \), then no contradiction exists, and the value of \( c_s \) is unconstrained on the range of non-negative real numbers. In the contradictory case (any \( \mu_{s,j} \neq 0 \)), there is no solution at \( \gamma = \lambda_s \). In the consistent case (all \( \mu_{s,j} = 0 \)), there are an infinite number of solutions at \( \gamma = \lambda_s \), spanning the range of \( \beta \in [\beta_{\text{min,bnd}}, \infty) \), where

\[
\beta_{\text{min,bnd}} = \sum_{i \in \Omega_1} \frac{\lambda_i \mu_{i,j}^2}{(\gamma_i - \lambda_i)^2} \quad \text{for} \quad \gamma = \lambda_s. \tag{36}
\]

In other words, the plot of \( \beta \) as a function of \( \gamma \) is a spike at \( \gamma = \lambda_s \), where the spike extends upward from the minimum \( \beta \). This minimum \( \beta \) is a continuous extension of the monotonic function (33), noting that \( \mu_{s,j} = 0 \) for \( j \in \Gamma_s \) in the consistent case. In this sense, the spike is part of a monotonic function \( \beta \) on the domain (\( \gamma \geq \lambda_s \)).

Since there are two cases of solution to (33) – with consistent and inconsistent solutions at the domain boundary – it is useful to summarize the two cases graphically, as is done in Fig. 3. The case in which no consistent solution occurs at the domain boundary is shown in Fig. 3(a). In this case, the plot of \( \beta \) as a function of \( \gamma \) is asymptotic at \( \gamma = \lambda_s \). The case in which a consistent solution occurs at the domain boundary is shown in Fig. 3(b). In this case \( \beta \) is smooth over the interior where \( \gamma > \lambda_s \) but spikes sharply at the boundary where \( \gamma = \lambda_s = 1 \). Despite the rapid change in slope at the boundary, it is evident that \( \beta \) is continuous and monotonically decreasing from infinity to zero over the closed domain \( \gamma \geq \lambda_s \).

A final degenerate case exists in which \( z \) is identically zero and \( \lambda_s \) is undefined, according to (31). In this case the ellipsoid is degenerate and consists only of one point (the origin). Though it is possible to plot, \( \beta \) as a function of \( \gamma \), as shown in Fig.
3(c), the result is trivial. In other words, since the ellipsoid is confined to the origin, the global maximum point \( z_s \) is unique and lies at the origin.

![Graphs showing \( \beta \) as a function of \( \gamma_c \) for different bias vectors \( \mu_c \).

Fig. 3. Counting Maxima. This plot shows \( \beta \) as a function of \( \gamma_c \), for \( \lambda = \{0.5, 0.8, 1.0\} \) and for bias vectors \( \mu_c \) labeled in each plot. In all cases shown, there is exactly one value of \( \gamma_c \) in the domain \( \gamma_c \geq \lambda_{\text{max}} \) that corresponds to any strictly positive value of \( \beta \). In other words, for a fixed \( \beta \) there is only one possible solution for \( \gamma_c \) on the domain \( \gamma_c \geq \lambda_{\text{max}} \) and hence only one possible \( \gamma_c \) corresponding to a global maximum.

A key result is that there is only one value of \( \gamma_c \) where an extremum can appear in the same orthant as \( \mu_c \). This result extends from the observation that \( \beta \) is monotonic between infinity and zero on the domain \( \gamma_c \geq \lambda_S \). Hence for any specified ellipse, a strictly positive value of \( \beta \) maps to a unique value of \( \gamma_c \) in that domain, such that equations (26) and (27) are solved simultaneously by this unique value of \( \gamma_c \). Importantly, it still remains to be determined whether this unique value of \( \gamma_c \) corresponds to multiple values of \( z_{ex} \).

To determine whether multiple extremum points \( z_{ex} \) may correspond to the same value of \( \gamma_c \), consider the domain interior \( \gamma_c > \lambda_S \) separately from the domain boundary \( \gamma_c = \lambda_S \).
In the interior region where $\gamma_c > \lambda_S$, equation (26) is linear and nonsingular, so the matrix on the right-hand side can be inverted to solve for $z_e$. When $\mu_c$ is nontrivial, the linear equations have a unique solution. (When $\mu_c$ is trivial, the ellipsoid is a point and there is still a unique though trivial solution, as discussed above.) In short, when $\gamma_c > \lambda_S$, there is only one extremum $z$ in the same orthant as $\mu_c$.

By contrast, multiple extrema are possible on the boundary where $\gamma_c = \lambda_S$, but all these extrema are equal in magnitude. To see this, consider that, by the definition of the two-norm,

$$\|z_e\| = \sum_{i \in I_2} z_{e,i}^2 + \sum_{j \in I_2} z_{e,j}^2.$$ (37)

For the boundary case, substituting (34) into (37) gives

$$\|z_e\| = \sum_{i \in I_2} z_{e,i}^2 + \lambda_S c_s.$$ (38)

Eliminate the $z_{e,i}$ in the first summation with (26) and eliminate $c_s$ in the second summation with (33) to obtain.

$$\|z_e\| = \sum_{i \in I_2} \left(1 - \frac{\lambda_i}{\lambda_S}\right)^2 \mu_{i,j}^2 + \lambda_S \left(\beta - \sum_{i \in I_2} \frac{\lambda_i \mu_{i,j}^2}{(\lambda_S - \lambda_i)^2}\right).$$ (39)

This equation depends only on fixed parameters for the ellipse: $\beta$, $\Lambda$, and $\mu_c$. Thus, when the value of $\gamma_c = \lambda_S$, the equation has a constant value at all extrema $z_e$. Since all extrema have the magnitude given by (39), no other extremum $z_e$ has a magnitude greater than that of $z$.

This final result establishes sufficiency.

In summary, we initially proposed the global maximum to be the point $z$ and the corresponding Lagrange multiplier to be $\gamma_c = \gamma$. The Lagrange multiplier was restricted to the range $\gamma_c \geq \lambda_S$ in order to impose the same-orthant condition, which was necessary for $z$ to be a global maximum. Then we considered the ellipsoid with a particular shape defined by $\Lambda$ and with magnitude $\beta$ and center $\mu_c$ computed from (20) and (21), which are equivalent to (26) and (27) with $z_e = z$, and $\gamma_c = \gamma$. We
then looked for any other extremum points \( \mathbf{z}_{\text{es}} \) (and corresponding Lagrange multipliers \( \gamma_c \)) that might appear in the same orthant as \( \mathbf{u}_c \), given the set of ellipsoid parameters \( \mathbf{A}, \beta, \text{ and } \mathbf{u}_c \) associated with the proposed global maximum. In looking for additional extrema, we found that \( \beta \) was a monotonic function of \( \gamma_c \), such that there could be no other extremum on the same orthant as \( \mathbf{u}_c \) unless \( \gamma_c = \gamma_z \). Thus, the only way for multiple extrema to exist on the same orthant as \( \mathbf{u}_c \) would be in special cases when elements of \( \mathbf{u}_c \) are zero, which allows for multiple \( \mathbf{z}_{\text{es}} \) that are symmetric reflections of each other. In this special case with multiple extrema in the same orthant as \( \mathbf{u}_c \), it was shown that all the symmetric extrema \( \mathbf{z}_{\text{es}} \) have the same magnitude as the proposed global maximum \( \mathbf{z}_z \). This justifies Theorem 1, which states for the assumptions used above, that no point on the ellipsoid defined by \( \mathbf{z}_z, \gamma_z, \mathbf{A}, (20), \text{ and } (21) \) is farther from the origin than \( \mathbf{z}_z \).

For simplicity, the theorem relaxes the same-orthant condition in a conservative manner, noting that the requirement is that \( \gamma_z \in [\lambda_z, \infty) \), but that since \( \lambda_{\max} \geq \lambda_z \), it is sufficient to require that \( \gamma_z \in [\lambda_{\max}, \infty) \).

Commentary: As a side note related to Theorem 1, it is interesting to visualize some of the cases in which multiple solutions appear. These cases are all radially symmetric in some sense, with the multiple extrema \( \mathbf{z}_{\text{es}} \) located, according to (34), at a radius of \( \sqrt{c_{\max} \lambda_{\max}} \) from some center point.

One example is when global maxima appear at two points, e.g., in an ellipse centered at zero, where the maximum occurs on either end of the major axis. Another example is when the extrema are arrayed in a circle, e.g., in the case of an oblate spheroid, shaped like a soccer ball squished along one axis, where the maxima appear around the girth of the ball. A third example is the case when \( \mathbf{u}_c \) is the zero vector and all of the \( \lambda_i \) are the same, implying the ellipsoidal surface is a sphere centered at the origin.

In this case, all points on the entire surface \( \partial \Theta \) are global maxima. These three examples are illustrated in Fig. 4.
Fig. 4. Multiple global maxima may occur. Three examples are shown including an ellipse with two global maximum points (small red circles), an oblate ellipsoid with a circular ring of maxima (red band), and a sphere on which all points are maxima.
**Theorem 2**: Weighted squares of a positive diagonal matrix. Consider a unit vector $e \in \mathbb{R}^N$ and a positive diagonal matrix $\Lambda \in \mathbb{R}^{N \times N}$. Then

$$e^T \Lambda^2 e \geq (e^T \Lambda e)^2.$$ \hspace{1cm} \text{(40)}

**Context**: This theorem is an intermediate result, critical to the proof of Theorem 3. Theorem 2 also happens to be an alternative form of a result commonly used in statistical analysis (e.g., in the estimation of covariance from a large population of random samples).

**Proof**: Start by considering the following matrix

$$\Gamma = (\Lambda - (e^T \Lambda e) I)^2.$$ \hspace{1cm} \text{(41)}

Since $\Lambda$ is diagonal, $\Gamma$ is also diagonal. Furthermore, since $\Lambda$ is real and since all of the diagonal terms of $\Gamma$ are therefore squares of real numbers, the diagonal elements of $\Gamma$ are positive or zero, making $\Gamma$ positive semidefinite:

$$e^T (\Lambda - (e^T \Lambda e) I)^2 e \geq 0.$$ \hspace{1cm} \text{(42)}

Expanding this quadratic gives

$$e^T \left( \Lambda^2 - 2(e^T \Lambda e) \Lambda + (e^T \Lambda e)^2 \right) e \geq 0.$$ \hspace{1cm} \text{(43)}

Distributing the leading and trailing unit vectors gives

$$e^T \Lambda^2 e - 2(e^T \Lambda e)^2 + (e^T \Lambda e)^2 \geq 0.$$ \hspace{1cm} \text{(44)}

Collapsing the two similar terms and moving them to the right side of the inequality gives (40).
**Theorem 3: Generalized Chi-Square Lower Bound.** Consider a random vector $\mathbf{y} \in \mathbb{R}^N$ that is Gaussian distributed with covariance $\mathbf{Q}$ and mean $\mathbf{\mu}$. The integral of the Gaussian density function inside a spherical threshold with radius $\sqrt{T}$ is the generalized chi-square cumulative distribution function $P_{\chi^2}(T; \mathbf{Q}, \mathbf{\mu})$, as described by (5). Assume the values of the parameters $\mathbf{Q}$ and mean $\mathbf{\mu}$ are uncertain but bounded, such that $\|\mathbf{\mu}\| < b < \sqrt{T}$ and $\mathbf{Q} < \lambda_{\max} \mathbf{I}$, where $\lambda_{\max} \geq 0$. Define

$$P_{\chi^2}(T) = P_{\text{ncx}} \left( \sqrt{\frac{T - b}{\lambda_{\max}}} + b ; N, b^2 \right),$$  

where $P_{\text{ncx}}$ is a noncentral chi-square distribution with $N$ degrees-of-freedom and noncentrality parameter $b^2$. This noncentral chi-square distribution is a lower bound for the original distribution, with $P_{\chi^2}(T) \leq P_{\chi^2}(T; \mathbf{Q}, \mathbf{\mu})$.

**Context:** This theorem is the key result of the paper. Application of this theorem makes it possible to bound false alarms even for a monitor that has unknown (but bounded) mean and covariance in fault-free conditions.

**Proof:** As a first step in proving the lower bound, it is useful to rotate the signal vector $\mathbf{y}$ such that its coordinate axes align with the principal axes of the covariance matrix $\mathbf{Q}$. This can be accomplished using the eigenvalue decomposition $\mathbf{Q} = \mathbf{X} \Lambda \mathbf{X}^T$, where $\mathbf{X}$ is a rotation (or Eigenvector) matrix and where $\Lambda$ is the diagonal eigenvalue matrix. A suitable eigen-decomposition always exists for a covariance matrix. The rotated signal variable is $\mathbf{z} = \mathbf{X}^T \mathbf{y}$.

After changing variables, the generalized chi-square integral (5) becomes

$$P_{\chi^2}(T; \mathbf{Q}, \mathbf{\mu}) = \int_{\Omega} [2\pi\Lambda]^{-\frac{N}{2}} e^{-\frac{1}{2}(\mathbf{y}^T \Lambda^{-1} \mathbf{y} + \mathbf{\mu}^T \Lambda^{-1} \mathbf{\mu})} \prod_i dz_i,$$

where the domain of integration is now

$$\Omega = \{ \mathbf{z} | \mathbf{z}^T \mathbf{z} \leq T \},$$

and where the bias has also been rotated, with $\mathbf{\mu}_z = \mathbf{X}^T \mathbf{\mu}$. Because $\mathbf{X}$ is a unitary matrix, the rotation operation does not change the magnitude of the bias vector, and so

$$\|\mathbf{\mu}_z\| = \|\mathbf{\mu}\| = b.$$
As a next step, approximate the true integral (46) by shrinking its domain of integration. Without changing the integrand, label the new domain \( \Theta \), and label the approximation \( P_{gr2} \), such that

\[
P_{gr2}(T; Q, \mu) = \int_{\Theta} [2\pi \Lambda]^\frac{1}{2} e^{-\frac{1}{2}(x - \mu)^\top \Lambda^{-1} (x - \mu)} \prod_i dz_i .
\]

(49)

As long as the new domain \( \Theta \) is contained inside of the original domain \( \Omega \) (i.e., as long as \( \Theta \subset \Omega \)), then the approximate integral is smaller than the original integral.

\[
P_{gr2}(T; Q, \mu) = P_{gr2}(T; Q, \mu) + \int_{\Omega \setminus \Theta} [2\pi \Lambda]^\frac{1}{2} e^{-\frac{1}{2}(x - \mu)^\top \Lambda^{-1} (x - \mu)} \prod_i dz_i .
\]

(50)

Hence if \( \Theta \subset \Omega \), then the approximation is a lower bound: \( P_{gr2}(T; Q, \mu) \leq P_{gr2}(T; Q, \mu) \).

The next step is to specify \( \Theta \) and ensure that it is contained within the original domain \( \Omega \). In specifying the new domain, it is helpful to introduce another change of variables, introducing \( w \), which is a warped version of \( z \) shifted by an offset \( \mu_c \):

\[
w = \Lambda^{-1/2} (z - \mu_c) .
\]

(51)

The vector \( \mu_c \) can be transformed into \( \mu_w \) similarly:

\[
\mu_w = \Lambda^{-1/2} (\mu_z - \mu_c) .
\]

(52)

Substituting these definitions, the Gaussian distribution in integral (49) becomes spherically symmetric.

\[
P_{gr2}(T; Q, \mu) = \int_{\Theta} [2\pi \Lambda]^\frac{1}{2} e^{-\frac{1}{2}(w - \mu_c)^\top \Lambda^{-1} (w - \mu_c)} \prod_i dw_i
\]

(53)

The shape of the new domain of integration \( \Theta \) has not yet been specified; however, analysis can be simplified if the new domain of integration is spherical. For this reason, define \( \Theta \) to be spherical in \( w \) coordinates.

\[
\Theta = \{ w | w^\top w \leq \beta \}
\]

(54)
Because the new threshold surface, described by the parameter $\beta$, and the constant-probability contours for the Gaussian are both spherical, the integral (53) is equivalent to a noncentral chi-square distribution. This distribution $P_{ncx}$ has $N$ degrees of freedom (where $w \in \mathbb{R}^N$) and noncentrality parameter $\|\mu_w\|^2$.

\[
\left( T; Q, \mu \right) = P_{ncx} \left( \beta; N, \|\mu_w\|^2 \right) \tag{55}
\]

For the new domain of integration to be contained in the original, a careful choice of parameters is required. Specifically, the shift parameter $\mu_c$, from (51), and the modified threshold $\beta$, from (54), must be selected to ensure $\Theta \subset \Omega$.

To select appropriate parameter values, start by considering an arbitrary point on the boundary $\partial \Omega$ of the original domain of integration, and label this point $z_*$. Given that $z_* \in \partial \Omega$ and that the radius of the original spherical domain of integration is $\sqrt{T}$,

\[
\|z_*\| = T. \tag{56}
\]

Now define this point also to be the outermost point (farthest from the origin) for the approximated domain of integration $\Theta$. Since all points in $\Theta$ have values with $\|z\| \leq \|z_*\|$, they all lie within the closed domain $\Omega$, and so $\Theta \subset \Omega$.

Theorem 1 can be used to obtain a family of ellipsoids, all with their outermost point at $z_*$. Ellipsoids in this family are defined for a range of possible values of the parameter $\gamma_e \in [\lambda_{\text{max}}, \infty)$. For each value of $\gamma_e$, the ellipsoid parameters $\mu_*$ and $\beta$ are computed from (20) and (21). Note that the new domain of integration $\Theta$ is described here as an ellipsoid in $z$ coordinates, though it is also a sphere in $w$ coordinates, as mapped through transformation (51).

Rather than allow $z_*$ to be arbitrary, define $z_*$ such that the mean of the density function $\mu_*$ is one of the possible ellipsoid centers $\mu_*$. This particular value of $z_*$ may be obtained by inverting (20) to give

\[
z_* = \left( I - \frac{\Lambda}{\gamma_e} \right)^{-1} \mu_. \tag{57}
\]

Here the value $\gamma_e$ is selected to enforce constraint (56). If the constraint cannot be satisfied by evaluating (57) on the interior domain $\gamma_e < \lambda_{\text{max}}$, then a special case must be considered where $\gamma_e = \lambda_{\text{max}}$. For this case, (57) should be inverted only for the
nonsingular indices; the singular indices of \( z \) are arbitrary, and can be set as needed to ensure \( \| z \| = T \). In special cases, when more than one of the \( \lambda_i \) is \( \lambda_{\text{max}} \), there are multiple \( z \) that map to \( \mu_c \). For the purposes of subsequent analysis, it is sufficient to pick any of these equivalent \( z \).

With the point \( z \) determined from \( \mu_c \), it is now possible to consider the family of associated ellipsoids, all with an outermost point at \( z \). As a starting point, note that if \( \gamma_c = \gamma \), the domain of integration is centered on \( \mu_c \), the mean of the Gaussian density function. As \( \gamma_c \) decreases below \( \gamma \), the ellipsoid grows larger, meaning that \( \beta \) increases according to (21) and that the ellipse center \( \mu_c \) shifts away from the mean \( \mu_c \). These two effects impact the probability integral (49) in opposite ways. Increasing \( \beta \) means that the ellipse grows to contain more probability, tending to increase \( \mu_{P_x} \). By contrast, motion of \( \mu_c \) away from \( \mu_c \) increases the noncentrality parameter, which is the magnitude squared of (52). This increase in noncentrality parameter decreases \( P_{\mu_{P_x}} \), and therefore competes with the increase in \( \beta \).

In concept, an optimization problem might be defined to balance these competing effects by choosing the threshold \( T \) and the parameter \( \gamma_c \) to maximize the lower bound \( P_{\mu_{P_x}} \) over all possible values of the unknown Gaussian parameters \( Q \) and \( \mu_c \).

\[
P_N' = \min_{\mu_c} \left( \max_{\gamma_c} \left( P_{\mu_{P_x}} \right) \right).
\]

Although this optimization problem might be solved numerically, the result requires an involved computation for each different bias direction \( \mu_c \).

As an alternative, a suboptimal but useful analytic result can be obtained that provides a simple form of \( P_{\mu_{P_x}} \) for arbitrary bias direction. The remainder of this proof will focus on this analytic result. To obtain the desired simplification, set \( \gamma_c \) using the following equation.

\[
\gamma_c = \left(1 - \frac{b}{\sqrt{\beta}} \right) \gamma_z
\]

The utility of this constraint is clear when (20), (52), and (57) are combined to give the following equation for the noncentrality parameter \( \| \mu_c \| \).
\[\|\mu_x\|^2 = \beta \left(1 - \frac{\gamma_x}{\gamma_z}\right)^2. \quad (60)\]

Applying constraint (59) to the above equation means that

\[\|\mu_x\|^2 = b^2. \quad (61)\]

This result greatly simplifies the problem by making the noncentrality parameter constant (and therefore independent of the effective threshold \(\beta\)). It should be noted, however, that (59) only makes sense if \(b / \sqrt{\beta} < 1\); otherwise a contradiction would arise since one of the pair \(\gamma_x\) and \(\gamma_z\) would be zero or negative, which is not compatible with the requirement that both variables be greater than or equal to \(\lambda_{max}\).

Introducing constraint (59) simplifies the expression for the lower bound \(P_{\gamma^2}\).

\[P_{\gamma^2}(T; Q, \mu) = P_{\text{ncz}} \left(\beta; N, b^2\right). \quad (62)\]

The expression can be further simplified by noting that the only remaining user-defined parameter is the threshold \(T\). The parameter \(\beta\) in the above equation is implicitly related to \(T\). To make this relationship explicit, start by noting that \(\|\mu_x\| = b\), as defined in (48), and combine this fact with (57) to give

\[b^2 = \mathbf{z}_u^\top \left(I - \frac{\Lambda}{\gamma_z}\right)^2 \mathbf{z}_u. \quad (63)\]

Now expand the right-hand side.

\[b^2 = \mathbf{z}_u^\top \mathbf{z}_u - 2 \mathbf{z}_u^\top \frac{\Lambda}{\gamma_z} \mathbf{z}_u + \mathbf{z}_u^\top \frac{\Lambda^2}{\gamma_z} \mathbf{z}_u. \quad (64)\]

Note that \(\|\mathbf{z}_u\|^2 = T\) and introduce the variable \(\mathbf{e}\) to represent the unit vector aligned with and parallel to \(\mathbf{z}_u\).

\[\mathbf{z}_u = \sqrt{T} \mathbf{e}. \quad (65)\]

Substituting this definition into (64) gives
\[ b^2 = T - 2T \mathbf{e}^\top \frac{\Lambda}{\gamma_z} \mathbf{e} + T \mathbf{e}^\top \frac{\Lambda^2}{\gamma_z} \mathbf{e}. \]  

(66)

To simplify the last term, invoke Theorem 2.

\[ b^2 \geq T - 2T \mathbf{e}^\top \frac{\Lambda}{\gamma_z} \mathbf{e} + T \left( \mathbf{e}^\top \frac{\Lambda}{\gamma_z} \mathbf{e} \right)^2. \]  

(67)

The result is an inequality, but one that can still be used to relate \( \beta \) to \( T \). To develop this relationship further, factor the above to give

\[ b^2 \geq T \left( 1 - \mathbf{e}^\top \frac{\Lambda}{\gamma_z} \mathbf{e} \right)^2. \]  

(68)

Rearranging terms:

\[ \mathbf{e}^\top \Lambda \mathbf{e} \geq \gamma_z \left( 1 - \frac{b}{\sqrt{T}} \right). \]  

(69)

Introducing constraint (59), the free parameter \( \gamma_z \) can be eliminated from the above equation.

\[ \mathbf{e}^\top \frac{\Lambda}{T} \mathbf{e} \geq \gamma_z \left( 1 - \frac{b}{\sqrt{T}} \right) \left( 1 - \frac{b}{\sqrt{T}} \right). \]  

(70)

The parameter \( \gamma_z \) can in turn be eliminated and replaced with \( \beta \) by rearranging the terms of (21) to give

\[ \gamma_z = \sqrt{\frac{\mathbf{z}^\top \Lambda \mathbf{z}}{\beta}}. \]  

(71)

Substituting (65) and the above equation and into (70) gives

\[ \sqrt{\mathbf{e}^\top \Lambda \mathbf{e}} \geq \left( \sqrt{T} - b \right)^{-1} \left( \sqrt{T} - b \right). \]  

(72)

Solving for \( \beta \), we have:
Now find the smallest possible value for the right-hand side of the above equation, which occurs when \( e \) aligns with the largest principal axis of \( \Lambda \) (associated with the eigenvalue \( \lambda_{\text{max}} \)). The result is that

\[
\sqrt{\beta} \geq \sqrt{\beta_{\text{min}}}
\]

(74)

where

\[
\sqrt{\beta_{\text{min}}} = \frac{\sqrt{T} - b}{\sqrt{\lambda_{\text{max}}}} + b.
\]

(75)

For the limiting case with \( e = e_{\text{max}} \), equation (67) is in fact an equality, since \( e_{\text{max}}^T \Lambda e_{\text{max}} = e_{\text{max}}^T \Lambda^2 e_{\text{max}} \). Hence, no excess margin was introduced by invoking Theorem 2, in this case where \( \sqrt{\beta} = \sqrt{\beta_{\text{min}}} \).

Returning to (62), we can now replace \( \beta \) with \( \beta_{\text{min}} \) as defined by (75). The result is the desired, simple form of the noncentral chi-square distribution, stated above as (45). By setting \( \sqrt{\beta} = \sqrt{\beta_{\text{min}}} \), the lower bound no longer depends on the \( Q \) and \( \mu \) that describe the actual distribution, hence equation (45) is not parameterized by these variables.

This completes the theorem, noting that for the analysis to be valid, constraint (59) requires \( b / \sqrt{\beta} < 1 \), which is true according to (75) so long as \( \sqrt{T} > b \).

\[ \rightarrow \]

REFERENCES


