Asymptotic Global Confidence Regions in Parametric Shape Estimation Problems

Jong Chul Ye, *Member, IEEE*, Yoram Bresler, *Fellow, IEEE*, and Pierre Moulin, *Senior Member, IEEE*

***Abstract*—We introduce confidence region techniques for analyzing and visualizing the performance of two-dimensional parametric shape estimators. Assuming an asymptotically normal and efficient estimator for a finite parameterization of the object boundary, Cramér–Rao bounds are used to define an asymptotic confidence region, centered around the true boundary. Computation of the probability that an entire boundary estimate lies within the confidence region is a challenging problem, because the estimate is a two-dimensional nonstationary random process. We derive lower bounds on this probability using level crossing statistics. The same bounds also apply to asymptotic confidence regions formed around the estimated boundaries, lower-bounding the probability that the entire true boundary lies within the confidence region. The results make it possible to generate asymptotic confidence regions for arbitrary prescribed probabilities. These asymptotic global confidence regions conveniently display the uncertainty in various geometric parameters such as shape, size, orientation, and position of the estimated object, and facilitate geometric inferences. Numerical simulations suggest that the new bounds are quite tight.**

***Index Terms*—Confidence regions, Cramér–Rao bounds, geometric inference, imaging, level crossing probability, maximum-likelihood estimation, parametric shape estimation, random process.**

I. Introduction

SEVERAL important imaging problems, including tomographic shape reconstruction, nonlinear inverse scattering, and computer vision, involve estimation of two-dimensional object shapes. We consider the following generic model for estimation of a shape $\boldsymbol{s}$ from a collection of N independent noisy data vectors $\boldsymbol{Y}_i$:

$$\boldsymbol{Y}_i \sim p(\boldsymbol{Y}_i|\boldsymbol{s}), \qquad \boldsymbol{Y}_i \in \mathbb{C}^M,\ i = 1, \cdots, N \tag{1}$$

where M denotes the number of sensor elements, $\boldsymbol{s}$ is a two-dimensional boundary, which describes the unknown *true* shape, and $p(\boldsymbol{Y}_i|\boldsymbol{s})$ is the observation model. This paper deals with parametric shape representations of the form

$$\boldsymbol{s}(t) \triangleq \boldsymbol{s}(t;\boldsymbol{\theta}) = \begin{bmatrix} s_x(t;\boldsymbol{\theta}) \\ s_y(t;\boldsymbol{\theta}) \end{bmatrix}, \qquad t \in I \tag{2}$$

where t is an index to points along the boundary, taking values in an interval $I = [0, T]$ for some fixed $0 < T < \infty$, $s_x(t;\boldsymbol{\theta})$, and $s_y(t;\boldsymbol{\theta})$ are the x and y coordinates of the boundary point indexed by t, and $\boldsymbol{\theta} \in \mathbb{R}^K$ is a K-dimensional parameter vector. The functions s_x and s_y are known functions,[1] assumed to be bounded and continuously twice differentiable with respect to t and continuously differentiable with respect to $\boldsymbol{\theta} \in \mathbb{R}^K$.

Parameterizations such as Fourier descriptors (FD) [1], [2], B-splines [3], and wavelet descriptors [4], [5] are special cases of this model and have been widely used for shape representation. In these examples, $\boldsymbol{s}(t;\boldsymbol{\theta})$ is linear in $\boldsymbol{\theta}$, but the general, nonlinear relationship arising for example in a spline model with free knots, is also of interest, and is covered by the analysis in this paper. In the shape estimation problem under the model (2), the true vector parameter $\boldsymbol{\theta}$ is unknown, and the problem can be reduced to finding an estimate $\hat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}$ from the statistics

$$\boldsymbol{Y}_i \sim p(\boldsymbol{Y}_i|\boldsymbol{\theta}), \qquad \boldsymbol{\theta} \in \mathbb{R}^K,\ \boldsymbol{Y}_i \in \mathbb{C}^M,\ i = 1, \cdots, N. \tag{3}$$

The boundary estimate is then

$$\hat{\boldsymbol{s}}(t) \triangleq \boldsymbol{s}(t;\hat{\boldsymbol{\theta}}). \tag{4}$$

The parametric formulation (3) of the shape estimation problem offers several advantages. First, for inverse problems, such as tomography and nonlinear inverse scattering [6], [7], a parametric formulation can alleviate the ill-posedness of the estimation problem (1). Second, fundamental bounds on the performance of shape estimation can, in principle, be derived [8]. This can be done via the Cramér–Rao lower bound (CRB) on estimation of the vector parameter $\boldsymbol{\theta}$, which provides a benchmark for assessing the performance of any specific unbiased estimator of $\boldsymbol{\theta}$. Such predictions can be useful *a priori* for purposes of design and analysis, or *a posteriori*, after measurements have been taken and the shape reconstructed, for purposes of assessing the reliability of the reconstruction.

In practice, because $\boldsymbol{s}(t)$ describes the geometry of an object, one is more interested in assessing the quality of estimates of $\boldsymbol{s}(t;\boldsymbol{\theta})$ in easily interpreted geometric terms. Rather than the quality of estimates of $\boldsymbol{\theta}$ itself, what is needed is a global quality measure for the entire boundary $\{\hat{\boldsymbol{s}}(t), \forall t \in I\}$. For example, in a clinical radiotherapy application, an uncertainty estimate of the shape of a tumor would allow surgeons to focus radiation effectively on a tumor with prescribed probability, while

Manuscript received September 13, 1999; revised February 4, 2000. This work was supported by a grant from DARPA under Contract F49620-98-1-0498, administered by AFOSR.

The authors are with the Coordinated Science Laboratory, Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801 (e-mail: jong@ifp.uiuc.edu; ybresler@uiuc.edu; moulin@ifp.uiuc.edu).

Communicated by A. O. Hero, Guest Editor.
Publisher Item Identifier S 0018-9448(00)06064-8.

[1]When $\{s(t),\ t \in I\}$ is a closed boundary, s_x and s_y are also periodic in t with period T. Inclusion of the nonperiodic case makes our analysis applicable to parts of an object boundary, allowing focus on parts of the boundary and production of tighter confidence regions for specific features of the object. Likewise, the results for nonperiodic $\boldsymbol{s}(t)$ may be applicable to the derivation of performance bounds for segmentation.

minimizing unnecessary exposure to other healthy tissue. Likewise, in brain surgery, it is critical to avoid blood vessels. An uncertainty estimate of blood vessel boundaries may help neurosurgeons plan an optimum path for the insertion of a surgical instrument. Similar uncertainty quantification for the shape of an object would be useful in automated assembly lines, where robots solder and screw, and autonomous vehicles convey assembled parts by avoiding obstacles, all under guidance of an imperfect imaging system. In general, such quantification of the uncertainty in shape estimation problems may answer many useful questions: what accuracy we can expect at any given point, which features of a shape are difficult to estimate, what is the effect of different parameterizations and data collecting geometries, etc.

There are significant difficulties, however, in obtaining a geometrically meaningful global performance measure. While the estimate $\hat{\boldsymbol{\theta}}$ for unknown deterministic parameter $\boldsymbol{\theta}$ is a random vector, the estimate $\hat{\boldsymbol{s}}(t) = \boldsymbol{s}(t; \hat{\boldsymbol{\theta}})$ for unknown deterministic shape $\boldsymbol{s}(t; \boldsymbol{\theta})$ is a stochastic process continuously indexed by t. Performance bounds for $\hat{\boldsymbol{s}}(t)$ can be derived by extension of the Cramér–Rao inequality [9]. Defining the shape estimation error for an unbiased estimator $\hat{\boldsymbol{s}}(t)$ by

$$\tilde{\boldsymbol{s}}(t) = \hat{\boldsymbol{s}}(t) - \boldsymbol{s}(t), \qquad t \in I = [0, T] \tag{5}$$

its integrated mean-square error $\boldsymbol{e_s}$ is bounded below by

$$\boldsymbol{e_s} = E\left[\frac{1}{T}\int_0^T ||\tilde{\boldsymbol{s}}(t)||^2\,dt\right] \geq \frac{1}{T}\int_0^T \boldsymbol{J}^{-1}(t, t)\,dt \tag{6}$$

where $\boldsymbol{J}(t, s)$ denotes the information kernel [9, pp. 437-455]. While the bound (6) for the shape estimate $\boldsymbol{s}(t; \hat{\boldsymbol{\theta}})$ is a useful criterion, it reduces information about an infinite-dimensional random process into a single performance index $\boldsymbol{e_s}$. Furthermore, the bound (6) cannot answer some important questions related to shape estimation: for example, it cannot reveal what accuracy we can expect at any given point t, which features of a shape are difficult to estimate, what is the effect of different parameterizations and data collecting geometries, etc. Pointwise bounds are also easy to derive

$$E||\tilde{\boldsymbol{s}}(t)||^2 \geq \boldsymbol{J}^{-1}(t, t). \tag{7}$$

However, the bound (7) does not provide quantitative information about the entire boundary estimate $\{\boldsymbol{s}(t; \hat{\boldsymbol{\theta}}), \forall t \in I\}$ since it is derived for each individual point t, without considering the entire correlation structure of the random process $\hat{\boldsymbol{s}}(t)$.

The main contribution of this paper is a technique for constructing small-size *global confidence regions* in the asymptotic regime where the estimate is unbiased, efficient, and Gaussian. We provide bounds of the probability that the entire boundary estimate lies in the global confidence region. These confidence regions can be conveniently visualized, see Fig. 1 for an example. They incorporate limits on the estimation performance for interesting geometric parameters such as shape, size, orientation, and the position of the object into an uncertainty band $\mathcal{U}_\beta$, which is created by moving local confidence ellipses $\{\mathcal{U}_\beta(t)\}$

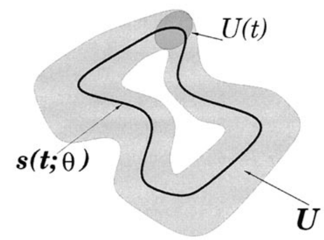


Fig. 1. An example of local and global confidence regions $\mathcal{U}_\beta(t)$ and $\mathcal{U}_\beta$, respectively, for a target boundary.

around the *true shape* $\boldsymbol{s}(t; \boldsymbol{\theta})$. Then, one can investigate the fundamental performance of shape estimation from the geometric properties of the confidence regions. We present two related applications of our confidence regions: one to pattern recognition; the other to an image reconstruction problem.

In the formulation of the confidence regions, a key parameter is the probability that the entire estimated boundary $\{\boldsymbol{s}(t; \hat{\boldsymbol{\theta}}), \forall t \in I\}$ lies in the asymptotic confidence region. We call this probability the *global confidence level*. Its evaluation is a *level-crossing* problem [10]–[24]. In the 1940s, Rice [12]–[14] developed fundamental techniques for solving such problems. In level-crossing terminology, our problem is interpreted as finding statistics that the stochastic process $\boldsymbol{s}(t; \hat{\boldsymbol{\theta}})$ stay between the barriers for all $t \in I$, where the barriers correspond to the inner and outer boundaries of the confidence region. The difficulty of the analysis is due to our confidence region problem having two-dimensional barriers, while all other conventional results for level-crossing by a scalar-indexed process are for one-dimensional barriers. Furthermore, as shown later, even for periodic $\boldsymbol{s}(t)$, the stochastic process $\boldsymbol{s}(t; \hat{\boldsymbol{\theta}})$ is usually nonstationary, so classical results for level-crossing of stationary random processes cannot be applied. It is not possible, in general, to find the exact global confidence level. Instead, we derive lower bounds which appear to be quite tight. Such lower bounds are useful in practice, since one can generate somewhat larger confidence regions, and still guarantee the required global confidence level.

In this paper, we also investigate bounds for the classical definition of confidence regions for interval estimation [25]. Here, the confidence region is generated *a posteriori* around the estimate $\boldsymbol{s}(t; \hat{\boldsymbol{\theta}})$ (rather than around the true shape $\boldsymbol{s}(t; \boldsymbol{\theta})$ as discussed above), and one is interested in finding the probability that the true shape $\boldsymbol{s}(t; \boldsymbol{\theta})$ lies in the confidence region. We show that our lower bounds are also valid for this type of problem.

In Section II, we present an asymptotic statistical model for the shape estimate $\boldsymbol{s}(t; \hat{\boldsymbol{\theta}})$, and define the global confidence regions and probabilities of interest in this asymptotic regime. In Section III, a conservative but still useful bound, the *incomplete gamma bound*, is first computed based on basic projection inequalities, then a tighter *level-crossing bound* is derived by decoupling the projection into a two-dimensional column space using Gram–Schmidt orthonormalization. Comparative studies with Monte Carlo simulations in Section IV suggest that the level-crossing bound is quite tight and useful for parametric shape estimation. Section V discusses further extensions of the analysis, and conclusions are presented in Section VI.

II. Problem Formulation

Under appropriate regularity conditions [25], [26], the maximum-likelihood estimator (MLE) $\hat{\boldsymbol{\theta}}$ of (3) is asymptotically unbiased, efficient and Gaussian-distributed (or best asymptotically normal (BAN))

$$\hat{\boldsymbol{\theta}} \sim \mathcal{N}(\boldsymbol{\theta}, \boldsymbol{C}_{\boldsymbol{\theta}}) \tag{8}$$

where the $K \times K$ covariance matrix $\boldsymbol{C}_{\boldsymbol{\theta}} = E[\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}][\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}]^T$ is the inverse of the Fisher information matrix. Examples of asymptotics in the observation model (3) include collecting an infinite number of snapshots ($N \rightarrow \infty$), or an increasing number of sensor elements ($M \rightarrow \infty$), with independent sensor noise realizations. Moreover, the identically distributed observations assumption can be relaxed if the observation model $P(\boldsymbol{Y}_i|\boldsymbol{\theta})$ satisfies the Lindeberg condition [27]. The independence assumption too can be relaxed to one of several weaker conditions [27].

Similar properties apply to smooth functions of $\boldsymbol{\theta}$ such as $\boldsymbol{s}(t; \boldsymbol{\theta})$. In particular, by the invariance property of the MLE, the asymptotic distribution of $\boldsymbol{s}(t; \hat{\boldsymbol{\theta}})$ is likewise BAN

$$\boldsymbol{s}(t; \hat{\boldsymbol{\theta}}) \sim \mathcal{N}(\boldsymbol{s}(t; \boldsymbol{\theta}), \boldsymbol{C}_{\boldsymbol{s}}(t)) \tag{9}$$

where the 2×2 covariance matrix

$$\boldsymbol{C}_{\boldsymbol{s}}(t) = E[\boldsymbol{s}(t; \hat{\boldsymbol{\theta}}) - \boldsymbol{s}(t; \boldsymbol{\theta})][\boldsymbol{s}(t; \hat{\boldsymbol{\theta}}) - \boldsymbol{s}(t; \boldsymbol{\theta})]^T$$

is given by

$$\boldsymbol{C}_{\boldsymbol{s}}(t) = [\nabla_{\boldsymbol{\theta}} \boldsymbol{s}(t; \boldsymbol{\theta})]^T \boldsymbol{C}_{\boldsymbol{\theta}} [\nabla_{\boldsymbol{\theta}} \boldsymbol{s}(t; \boldsymbol{\theta})] \tag{10}$$

where $\nabla_{\boldsymbol{\theta}} \boldsymbol{s}(t; \boldsymbol{\theta}) \in \mathbb{R}^{K \times 2}$ is the gradient of $\boldsymbol{s}(t; \boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$.[2] In the remainder of the paper, we assume the appropriate regularity conditions are satisfied, and the estimator operates in the asymptotic regime, so that the distribution (8) can be used.

For *a fixed t*, according to classical estimation theory [25], a confidence region $\hat{\mathcal{U}}_\beta(t)$ for $\boldsymbol{s}(t)$ at confidence level $\alpha \in [0, 1]$ is any subset of $\mathbb{R}^2$ such that

$$\Pr\left\{\boldsymbol{s}(t) \in \hat{\mathcal{U}}_\beta(t)\right\} = \alpha. \tag{11}$$

While there are infinitely many choices of $\hat{\mathcal{U}}_\beta(t)$ for each specified α [25], Wilks and Daly [28] showed that the smallest size confidence region *on the average* (because $\hat{\mathcal{U}}_\beta(t)$ is random) is

$$\hat{\mathcal{U}}_\beta(t) = \left\{\boldsymbol{x} \in \mathbb{R}^2 \colon (\boldsymbol{x} - \hat{\boldsymbol{s}}(t))^T \boldsymbol{C}_{\boldsymbol{s}}(t)^{-1} (\boldsymbol{x} - \hat{\boldsymbol{s}}(t)) \leq \beta^2\right\} \tag{12}$$

for an appropriate $\beta > 0$ such that $\Pr\{\chi_2^2 \leq \beta^2\} = \alpha$, where $\boldsymbol{C}_{\boldsymbol{s}}(t)$ is the covariance matrix for $\boldsymbol{s}(t; \hat{\boldsymbol{\theta}})$ given by (10), and χ_2^2 denotes the chi-square distribution with two degrees of freedom. For each t, $\hat{\mathcal{U}}_\beta(t)$ is an ellipse centered on the boundary point $\hat{\boldsymbol{s}}(t)$. Note that the signal-to-noise ratio (SNR) of the measurement model (3) does not enter the definition of the CRB ellipse (12) and the confidence level α. However, the spectral radius of $\boldsymbol{C}_{\boldsymbol{\theta}}$ is inversely proportional to the SNR, hence the size of the overall CRB ellipse radius is also inversely proportional to the SNR for a given value of β.

It is now desired to construct a global confidence region $\hat{\mathcal{U}}_\beta$ for the entire function $\{\boldsymbol{s}(t), \forall t \in I\}$. A possible design is obtained by moving $\hat{\mathcal{U}}_\beta(t)$ along the boundary

$$\hat{\mathcal{U}}_\beta \triangleq \bigcup_{t \in I} \hat{\mathcal{U}}_\beta(t). \tag{13}$$

The region $\hat{\mathcal{U}}_\beta$ forms a "tube" around the estimated boundary $\hat{\boldsymbol{s}}(t)$. Because $\hat{\boldsymbol{s}}(t)$ is a random process, $\hat{\mathcal{U}}_\beta$ is a random set. The goal of this paper is to find a method to calculate the global confidence level for the so-formed confidence region $\hat{\mathcal{U}}_\beta$

$$\gamma \triangleq \Pr\left\{\boldsymbol{s}(t) \in \hat{\mathcal{U}}_\beta, \forall t \in I\right\}. \tag{14}$$

Clearly, γ is smaller than the probability in (11).

Another important probability of theoretical interest is

$$\Pr\{\hat{\boldsymbol{s}}(t) \in \mathcal{U}_\beta, \forall t \in I\} \tag{15}$$

where $\hat{\boldsymbol{s}}(t)$ is the estimate of $\boldsymbol{s}(t)$, and the deterministic confidence region $\mathcal{U}_\beta$ is defined as

$$\mathcal{U}_\beta \triangleq \bigcup_{t \in I} \mathcal{U}_\beta(t) \tag{16}$$

where

$$\mathcal{U}_\beta(t) = \left\{\boldsymbol{x} \in \mathbb{R}^2 \colon (\boldsymbol{x} - \boldsymbol{s}(t))^T \boldsymbol{C}_{\boldsymbol{s}}(t)^{-1} (\boldsymbol{x} - \boldsymbol{s}(t)) \leq \beta^2\right\}, \tag{17}$$

Again, the region $\mathcal{U}_\beta$ is a tube, but this time centered around the true boundary. While the probability in (14) is the *a posteriori* probability that the true shape lies in a confidence region generated around the MLE $\hat{\boldsymbol{s}}(t)$, the *a priori* probability in (15) focuses on predicting the fundamental uncertainty region for *any* asymptotically normal and efficient estimator.[3] Therefore, in general they are different, and both are useful in practice.

The probabilities (14) and (15) are difficult to compute owing to the overlaps of individual ellipses $\hat{\mathcal{U}}_\beta(t)$ (resp., $\mathcal{U}_\beta(t)$) for all $t \in I$. We therefore wish to determine lower bounds on (14) and (15) that are reasonably tight. Such lower bounds are useful since one can generate somewhat larger confidence regions, and still guarantee the required probability that the shape lies within that region.

III. Main Results

The probabilities (14) and (15) satisfy the following bounds:

$$\Pr\left\{\boldsymbol{s}(t) \in \hat{\mathcal{U}}_\beta, \forall t \in I\right\} \geq \Pr\left\{\boldsymbol{s}(t) \in \hat{\mathcal{U}}_\beta(t), \forall t \in I\right\} \tag{18}$$

$$\Pr\{\hat{\boldsymbol{s}}(t) \in \mathcal{U}_\beta, \forall t \in I\} \geq \Pr\{\hat{\boldsymbol{s}}(t) \in \mathcal{U}_\beta(t), \forall t \in I\} \tag{19}$$

because there may exist some $t \neq t'$ such that $\boldsymbol{s}(t) \in \hat{\mathcal{U}}_\beta(t')$ while $\boldsymbol{s}(t) \notin \hat{\mathcal{U}}_\beta(t)$ (or $\hat{\boldsymbol{s}}(t) \in \mathcal{U}_\beta(t')$ while $\hat{\boldsymbol{s}}(t) \notin \mathcal{U}_\beta(t)$). Even though the probabilities in (14) and (15) are in general different,

[2]The Cramér–Rao inequality [25] tells us that the covariance matrix for any unbiased estimate of $\boldsymbol{s}(t; \boldsymbol{\theta})$ at a given t satisfies

$$\mathrm{Cov}\,(\boldsymbol{s}(t; \hat{\boldsymbol{\theta}})) \geq \boldsymbol{C}_{\boldsymbol{s}}(t)$$

where the inequality $A \geq B$ for matrices means that $A - B$ is nonnegative semi-definite.

[3]If an estimator is asymptotically normal and efficient, then it is asymptotically equivalent to the MLE.

the right-hand sides of (18) and (19) are identical. Indeed, the definitions (12) and (17) imply that the following events are the same:

$$\begin{aligned}&\left\{\boldsymbol{s}(t)\in\hat{\mathcal{U}}_\beta(t),\,\forall\,t\in I\right\}\\&\quad=\left\{(\hat{\boldsymbol{s}}(t)-\boldsymbol{s}(t))^T\boldsymbol{C}_{\boldsymbol{s}}(t)^{-1}(\hat{\boldsymbol{s}}(t)-\boldsymbol{s}(t))\le\beta^2,\,\forall\,t\in I\right\}\\&\quad=\{\hat{\boldsymbol{s}}(t)\in\mathcal{U}_\beta(t),\,\forall\,t\in I\}.\end{aligned}\tag{20}$$

Due to the equivalence in (20), in the sequel we derive the bounds for

$$\Pr\{\hat{\boldsymbol{s}}(t)\in\mathcal{U}_\beta(t),\,\forall\,t\in I\}\tag{21}$$

as a means to obtain lower bounds on the probabilities in (14) and (15).

A. Incomplete Gamma Bound

For convenience, we introduce the following notation:

$$\boldsymbol{Z}\triangleq\boldsymbol{C}_{\boldsymbol{\theta}}^{-1/2}\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)\in\mathbb{R}^K\tag{22}$$

and

$$\boldsymbol{B}(t)=[\boldsymbol{B}_x(t)\quad\boldsymbol{B}_y(t)],\tag{23}$$
$$\boldsymbol{B}_x(t)\triangleq\nabla_{\boldsymbol{\theta}}s_x(t;\boldsymbol{\theta})\in\mathbb{R}^{K\times1}\tag{24}$$
$$\boldsymbol{B}_y(t)\triangleq\nabla_{\boldsymbol{\theta}}s_y(t;\boldsymbol{\theta})\in\mathbb{R}^{K\times1}\tag{25}$$

where the dependencies on $\boldsymbol{\theta}$ of $\boldsymbol{B}$, $\boldsymbol{B}_x$, and $\boldsymbol{B}_y$ are implicit.[4] It follows from (8) that the transformed variable $\boldsymbol{Z}$ in (22) is normally distributed: $\boldsymbol{Z}\sim\mathcal{N}(\boldsymbol{0},\boldsymbol{I})$. Then, for all $t\in I$ and for all nonnegative β, we have

$$\begin{aligned}&\Pr\left\{\boldsymbol{s}(t;\hat{\boldsymbol{\theta}})\in\mathcal{U}_\beta(t)\right\}\\&=\Pr\left\{(\boldsymbol{s}(t;\hat{\boldsymbol{\theta}})-\boldsymbol{s}(t;\boldsymbol{\theta}))^T\boldsymbol{C}_{\boldsymbol{s}}(t)^{-1}(\boldsymbol{s}(t;\hat{\boldsymbol{\theta}})-\boldsymbol{s}(t;\boldsymbol{\theta}))\le\beta^2\right\}\\&=\Pr\left\{(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})^T\boldsymbol{B}(t)(\boldsymbol{B}(t)^T\boldsymbol{C}_{\boldsymbol{\theta}}\boldsymbol{B}(t))^{-1}\boldsymbol{B}^T(t)(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})\le\beta^2\right\}\\&=\Pr\left\{\boldsymbol{Z}^T\boldsymbol{C}_{\boldsymbol{\theta}}^{1/2}\boldsymbol{B}(t)(\boldsymbol{B}^T(t)\boldsymbol{C}_{\boldsymbol{\theta}}\boldsymbol{B}(t))^{-1}\boldsymbol{B}^T(t)\boldsymbol{C}_{\boldsymbol{\theta}}^{1/2}\boldsymbol{Z}\le\beta^2\right\}\end{aligned}\tag{26}$$

where the first equality follows from the definition of $\mathcal{U}_\beta(t)$ in (17), the second equality uses (10) and (23) and holds asymptotically by the consistency of $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{s}(t;\hat{\boldsymbol{\theta}})$ and differentiability of $\boldsymbol{s}(t;\hat{\boldsymbol{\theta}})$, and the third follows from the definition (22). We assume that $\boldsymbol{C}_{\boldsymbol{\theta}}^{1/2}\boldsymbol{B}(t)$ has full column rank for every t (this is satisfied by $\boldsymbol{B}(t)$ with full column rank). The $K\times K$ matrix appearing in (26) is

$$\boldsymbol{P}(t)\triangleq\boldsymbol{C}_{\boldsymbol{\theta}}^{1/2}\boldsymbol{B}(t)(\boldsymbol{B}(t)^T\boldsymbol{C}_{\boldsymbol{\theta}}\boldsymbol{B}(t))^{-1}\boldsymbol{B}(t)^T\boldsymbol{C}_{\boldsymbol{\theta}}^{1/2}\tag{27}$$

which is a rank-two projection on the range space of $\boldsymbol{C}_{\boldsymbol{\theta}}^{1/2}\boldsymbol{B}(t)$. With this notation, we rewrite (26) as

$$\Pr\left\{\boldsymbol{s}(t;\hat{\boldsymbol{\theta}})\in\mathcal{U}_\beta(t),\forall t\in I\right\}=\Pr\left\{\boldsymbol{Z}^T\boldsymbol{P}(t)\boldsymbol{Z}\le\beta^2,\forall t\in I\right\}\tag{28}$$

[4]Fourier descriptors, B-splines, and wavelet descriptors are special cases of (2) that are linearly parameterized by $\boldsymbol{\theta}$. For such representations, $\boldsymbol{B}(t)$ is independent of $\boldsymbol{\theta}$.

where $\boldsymbol{Z}\sim\mathcal{N}(\boldsymbol{0},\boldsymbol{I})$. Since $\boldsymbol{Z}^T\boldsymbol{P}(t)\boldsymbol{Z}\le\boldsymbol{Z}^T\boldsymbol{Z}$ for any projection $\boldsymbol{P}(t)$, (28) provides a simple lower bound

$$\begin{aligned}\Pr\left\{\boldsymbol{s}(t;\hat{\boldsymbol{\theta}})\in\mathcal{U}_\beta(t),\,\forall\,t\in I\right\}&\ge\Pr\{\boldsymbol{Z}^T\boldsymbol{Z}\le\beta^2\}\\&=\Pr\{\chi_K^2\le\beta^2\}.\end{aligned}\tag{29}$$

The equality in (29) follows because $\boldsymbol{Z}^T\boldsymbol{Z}$ is distributed as χ^2 with K degrees of freedom. Hence, $\Pr\{\chi_K^2\le\beta^2\}$ can be determined using tables of the cumulative distribution function of χ_K^2, series expansions, or using the relation between χ_K^2 and the incomplete gamma function (for the details of these techniques, see [29] and the references therein). We obtain

$$\begin{aligned}\Pr\{\chi_K^2\le\beta^2\}&=\frac{1}{\Gamma\left(\frac{K}{2}\right)}\int_0^{\frac{\beta^2}{2}}e^{-y}y^{\frac{K}{2}-1}\,dy\\&=\Gamma\left(\frac{K}{2},\,\frac{\beta^2}{2}\right).\end{aligned}\tag{30}$$

While (30) provides a valid bound on the probability (21), the bound can be very conservative for large K. This is because the inequality in the first line of (29) removes the dependence on the characteristics of the projection $\boldsymbol{P}(t)$ (i.e., the basis matrix $\boldsymbol{B}(t)$ or the covariance matrix $\boldsymbol{C}_{\boldsymbol{\theta}}(t)$). Indeed, the Fisher–Cochran theorem [30] tells that $\boldsymbol{Z}^T\boldsymbol{P}(t)\boldsymbol{Z}$ is a χ^2 process with two degrees of freedom, and

$$\Pr\{\boldsymbol{Z}^T\boldsymbol{P}(t)\boldsymbol{Z}\le\beta^2,\forall t\in I\}=\Pr\{\chi_2^2(t)\le\beta^2,\forall t\in I\}.\tag{31}$$

Therefore, the incomplete Gamma bound (30), which bounds the random process $\chi_2^2(t)$ by a random variable χ_K^2, provides a very conservative bound for large K.

In order to derive a tighter bound, note that calculation of (31) becomes an extreme probability problem [10], [11] of a χ^2 process with two degrees of freedom. As will be explained later, the problem is quite involved since the random process $\chi_2^2(t)$ is usually nonstationary. Hence, the stationary case results by Aronowich and Adler [31] are not applicable. While Piterbarg [32] investigated extreme probability problems for nonstationary χ^2 processes, the results therein are asymptotic large-deviation results which hold for sufficiently large β. We derive a new nonasymptotic result which is valid for all positive β for this nonstationary χ^2 process. The next section describes the new result.

B. Level-Crossing Bound

A tighter bound for the probability in (28) can be obtained by taking into account of the characteristics of $\boldsymbol{C}_{\boldsymbol{\theta}}^{1/2}\boldsymbol{B}(t)$. We assume that the $K\times2$ matrix $\boldsymbol{B}(t)$ is continuously twice-differentiable with respect to t and has full column rank for all $t\in I$. Apply the Gram–Schmidt orthonormalization procedure to $\boldsymbol{C}_{\boldsymbol{\theta}}^{1/2}\boldsymbol{B}(t)$ as follows:

$$\boldsymbol{u}_1(t)=\boldsymbol{C}_{\boldsymbol{\theta}}^{1/2}\boldsymbol{B}_x(t)\qquad\boldsymbol{v}_1(t)=\frac{\boldsymbol{u}_1(t)}{\|\boldsymbol{u}_1(t)\|}\tag{32}$$

and

$$\begin{aligned}\boldsymbol{u}_2(t)&=\boldsymbol{C}_{\boldsymbol{\theta}}^{1/2}\boldsymbol{B}_y(t)-\left(\boldsymbol{C}_{\boldsymbol{\theta}}^{1/2}\boldsymbol{B}_y(t)\right)^T\boldsymbol{v}_1(t)\boldsymbol{v}_1(t)\\\boldsymbol{v}_2(t)&=\frac{\boldsymbol{u}_2(t)}{\|\boldsymbol{u}_2(t)\|}.\end{aligned}\tag{33}$$

Then, the corresponding QR factorization [33] of $\boldsymbol{C}_{\boldsymbol{\theta}}^{1/2}\boldsymbol{B}(t)$ becomes

$$\boldsymbol{C}_{\boldsymbol{\theta}}^{1/2}\boldsymbol{B}(t) = \boldsymbol{V}(t)\boldsymbol{D}(t) \tag{34}$$

where

$$\begin{aligned}\boldsymbol{V}(t) &= [\boldsymbol{v}_1 \quad \boldsymbol{v}_2]\\ \boldsymbol{D}(t) &= \begin{bmatrix} \|\boldsymbol{u}_1(t)\| & -\left(\boldsymbol{C}_{\boldsymbol{\theta}}^{1/2}\boldsymbol{B}_y(t)\right)^T \boldsymbol{v}_1(t) \\ 0 & \|\boldsymbol{u}_2(t)\| \end{bmatrix}.\end{aligned} \tag{35}$$

Note that the Gram–Schmidt procedure (32), (33) guarantees the differentiability of the orthonormal vectors $\boldsymbol{v}_1(t)$ and $\boldsymbol{v}_2(t)$, because the elements of matrix $\boldsymbol{C}_{\boldsymbol{\theta}}^{1/2}\boldsymbol{B}(t)$ are differentiable with respect to t, and neither of the norms $\|\boldsymbol{u}_1(t)\|$ or $\|\boldsymbol{u}_2(t)\|$ vanishes, because $\boldsymbol{C}_{\boldsymbol{\theta}}^{1/2}\boldsymbol{B}(t)$ has full rank for all t. Furthermore, a general form of *differentiable* orthonormal basis vectors that span *Range* $(\boldsymbol{C}_{\boldsymbol{\theta}}^{1/2}\boldsymbol{B}(t))$, can be represented using $\boldsymbol{v}_1(t)$ and $\boldsymbol{v}_2(t)$

$$\boldsymbol{q}_1(t,\,\omega) = \boldsymbol{v}_1(t)\cos\omega(t) + \boldsymbol{v}_2(t)\sin\omega(t) \tag{36}$$

$$\boldsymbol{q}_2(t,\,\omega) = -\boldsymbol{v}_1(t)\sin\omega(t) + \boldsymbol{v}_2(t)\cos\omega(t) \tag{37}$$

where $\omega(t)$ is a differentiable function with respect to t. The projection matrix $\boldsymbol{P}(t)$ is then given by

$$\begin{aligned}\boldsymbol{P}(t) &= \boldsymbol{Q}(t)\boldsymbol{Q}(t)^T\\ &= \boldsymbol{q}_1(t,\,\omega)\boldsymbol{q}_1(t,\,\omega)^T + \boldsymbol{q}_2(t,\,\omega)\boldsymbol{q}_2(t,\,\omega)^T, \qquad t \in I.\end{aligned} \tag{38}$$

Note that although $\boldsymbol{q}_1$ and $\boldsymbol{q}_2$ depend on ω, $\boldsymbol{P}(t)$ does not, because it is determined by the *subspace* spanned by these vectors, and is invariant to the specific choice of basis for this subspace. Hence, the dependence on ω will be suppressed in the sequel, and only displayed when considering the choice of ω to simplify the computation of the bound.

Using (38), we have

$$\{\boldsymbol{Z}^T\boldsymbol{P}(t)\boldsymbol{Z} \le \beta^2\} = \{|\boldsymbol{Z}^T\boldsymbol{q}_1(t)|^2 + |\boldsymbol{Z}^T\boldsymbol{q}_2(t)|^2 \le \beta^2\}. \tag{39}$$

Define two scalar random processes $I_c(t)$ and $I_s(t)$

$$I_c(t) \triangleq \boldsymbol{Z}^T\boldsymbol{q}_1(t) \qquad I_s(t) \triangleq \boldsymbol{Z}^T\boldsymbol{q}_2(t), \qquad t \in I \tag{40}$$

and their envelope process

$$R(t) \triangleq \sqrt{I_c(t)^2 + I_s(t)^2}, \qquad t \in I. \tag{41}$$

Now (28) can be written as

$$\begin{aligned}\Pr\left\{\boldsymbol{s}(t;\,\hat{\boldsymbol{\theta}}) \in \mathcal{U}_\beta(t), \forall t \in I\right\} &= \Pr\{\boldsymbol{Z}^T\boldsymbol{P}(t)\boldsymbol{Z} \le \beta^2, \forall t \in I\}\\ &= \Pr\{\chi_2^2(t) \le \beta^2, \forall t \in I\}\\ &= \Pr\{0 \le R(t) \le \beta, \forall t \in I\}\end{aligned} \tag{42}$$

where $\boldsymbol{Z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$. Compared to the original Problem (15), Problem (42) is relatively simple, because for each $t \in I$, asymptotically $I_c(t)$ and $I_s(t)$ in (40) are independent, zero-mean, unit-variance Gaussian random variable; $R(t)$ in (41) is a Rayleigh process; and the barrier of the level-crossing is reduced to a one-dimensional barrier. However, the analysis is still complicated by the fact that $\chi_2^2(t)$ is usually nonstationary except for some special cases, see Appendix A. This implies that the classical results on level-crossing based on stationary assumptions [14], [18]–[22], [24] cannot be employed in our problem.

There is an analysis due to Shinozuka [16] for computing level-crossing probabilities, which we have found to be useful for our problem. Shinozuka's analysis does not require stationarity of the random process. In the following, we introduce a simplified version of his approach, which is adapted to our problem in which the barrier is one-sided instead of two-sided as in Shinozuka's problem. Following the notation in [16], define the probability

$$\begin{aligned}P[T;\,\beta] &\triangleq \Pr\left\{\max_{t\in[0,\,T]} R(t) \ge \beta\right\}\\ &= 1 - \Pr\{0 \le R(t) \le \beta, \forall\, t \in [0,\,T]\}.\end{aligned} \tag{43}$$

Then, for any $0 \le \Delta T \le T$, we have

$$\begin{aligned}P[T;\,\beta] &= P[T-\Delta T;\,\beta] + \Pr\left[\left\{\max_{0\le t\le T-\Delta T} R(t) < \beta\right\} \cap \left\{\max_{T-\Delta T\le t\le T} R(t) \ge \beta\right\}\right]\\ &\le P[T-\Delta T;\,\beta] + \Pr\left[\{R(T-\Delta T) < \beta\} \cap \left\{\max_{T-\Delta T\le t\le T} R(t) \ge \beta\right\}\right]\end{aligned} \tag{44}$$

since

$$\left\{\max_{T-\Delta T\le t\le T} R(t) < \beta\right\} \subseteq \{R(T-\Delta T) < \beta\}.$$

For infinitesimal ΔT, (44) can be evaluated using Rice's formula for level-crossing rate with positive slope [12]

$$\lim_{\Delta T\to 0}\frac{1}{\Delta T}\Pr\left[\{R(T-\Delta T) < \beta\} \cap \left\{\max_{T-\Delta T\le t\le T} R(t) \ge \beta\right\}\right] = \int_0^\infty d\dot{R}\,\dot{R}\,p(\beta,\,\dot{R};\,T) \tag{45}$$

where $\dot{R}$ denotes the derivative of the envelope process $R(t)$ with respect to t, and $p(R,\,\dot{R};\,T)$ is the joint probability density function of R and $\dot{R}$ at instant T. Note that in order to use the Rice formula (45), the sample path $R(t)$ should be continuous and differentiable for all $t \in I$ with probability one. For general random processes, Cramér and Leadbetter [10] summarized sufficient conditions for the sample path continuity and differentiability. However, the structure of our random process

$$R(t) = \sqrt{\boldsymbol{Z}^T\boldsymbol{P}(t)\boldsymbol{Z}}$$

is special so that Lemma 1 below provides simpler sufficient conditions. The proof may be found in Appendix B.

Lemma 1: Suppose the basis matrix $\boldsymbol{B}(t)$ is continuously differentiable, and has full column rank for all $t \in I$. Then, the sample path $R(t)$ of (41) is continuous and has a sample derivative for all $t \in I$ with probability one.

Hence, as $\Delta T \to 0$, (44) and (45) yield

$$\begin{aligned}\frac{\partial P[T;\,\beta]}{\partial T} &= \lim_{\Delta T\to 0}\frac{P[T;\,\beta] - P[T-\Delta T;\,\beta]}{\Delta T}\\ &\le \int_0^\infty d\dot{R}\,\dot{R}\,p(\beta,\,\dot{R};\,T)\\ &\triangleq h(T,\,\beta).\end{aligned} \tag{46}$$

The function $h(T, \beta)$ is termed the *level-crossing rate*. Because $h(T, \beta)$ is nonnegative, integrating (46) from 0 to T yields

$$P[T; \beta] \leq \int_0^T h(t, \beta)\, dt + P[0; \beta] \tag{47}$$

where $P[0; \beta]$ is the initial probability value at $t = 0$. Note that stationarity is not needed for (47) to hold. (The stationarity assumption would only simplify the integral of $h(t, \beta)$ to $Th(T, \beta)$). The initial value $P[0; \beta]$ can be readily computed from the definition

$$\begin{aligned} P[0; \beta] &= \Pr\{R(0) > \beta\} \\ &= \exp\left(-\frac{\beta^2}{2}\right) \end{aligned} \tag{48}$$

since $R(0)$ is Rayleigh-distributed. Therefore, the problem of finding a lower bound for the probability in (21) is reduced to finding $h(t, \beta)$ for the nonstationary envelope process $R(t)$ of (41). At this point, our analysis departs from that of Shinozuka because, as shown in Appendix C, a closed-form solution for $h(t, \beta)$ is difficult to find. Instead of finding $h(t, \beta)$, we replace it by a simpler function $h^*(t, \beta)$ such that

$$h(t, \beta) \leq h^*(t, \beta) \qquad \forall\, t \in [0, T]. \tag{49}$$

Hence, (47) can be replaced with the bound

$$P[T; \beta] \leq \int_0^T h^*(t, \beta)\, dt + P[0; \beta]. \tag{50}$$

One candidate for $h^*(t, \beta)$ is given by the following lemma which is proven in Appendix D.

Lemma 2: Suppose the basis matrix $\boldsymbol{B}(t)$ is continuous and twice-differentiable and has full column-rank for all $t \in I$. Let $h(t, \beta)$ denote the level-crossing rate of the nonstationary envelope process $R(t)$ of (41), as defined in (46). Then, for every $t \in [0, T]$

$$h(t, \beta) \leq h^*(t, \beta) = \beta \exp\left(-\frac{\beta^2}{2}\right)\left(\frac{\zeta(t, \omega^*)}{2\pi}\right)^{1/2} \tag{51}$$

where the function $\zeta(t, \omega^*)$ is given by

$$\begin{aligned} \zeta(t, \omega^*) \triangleq \max\{||\dot{\boldsymbol{q}}_1(t, \omega^*)||^2, ||\dot{\boldsymbol{q}}_2(t, \omega^*)||^2\} \\ -(\dot{\boldsymbol{q}}_1(t, \omega^*)^T \boldsymbol{q}_2(t, \omega^*))^2 \end{aligned} \tag{52}$$

$\dot{\boldsymbol{q}}_1$ and $\dot{\boldsymbol{q}}_2$ denote derivatives with respect to t, and $\omega^*(t)$ is given by

$$\omega^*(t) = \frac{1}{2}\tan^{-1}\left(\frac{2\dot{\boldsymbol{v}}_1^T \dot{\boldsymbol{v}}_2}{||\dot{\boldsymbol{v}}_1||^2 - ||\dot{\boldsymbol{v}}_2||^2}\right), \qquad -\frac{\pi}{4} < \omega^*(t) \leq \frac{\pi}{4} \tag{53}$$

where $\boldsymbol{v}_1(t)$ and $\boldsymbol{v}_2(t)$ denote the orthonormal basis vectors computed by the Gram–Schmidt procedure (32) and (33).

Note that $h^*(t, \beta)$ depends on t only through $\zeta(t, \omega^*)$. Therefore, using (42), (43), (47), (48), and (50), we readily arrive at a new bound.

Proposition 1:

$$\Pr\{0 \leq R(t) \leq \beta, \forall t \in I\} \geq 1 - (1 + \kappa\beta) \exp\left(-\frac{\beta^2}{2}\right) \tag{54}$$

where the constant κ is given by

$$\kappa \triangleq \int_0^T \left(\frac{\zeta(t, \omega^*)}{2\pi}\right)^{1/2} dt \tag{55}$$

and $\zeta(t, \omega^*)$ given by (52).

Note that the bound in (54) is a useful lower bound for (42), but its value depends on ω^* of (53), hence the computation of $\dot{\boldsymbol{q}}_1$ and $\dot{\boldsymbol{q}}_2$ is very complicated. Therefore, we find a lower bound for (54), which does not depend on ω^*. Proposition 2 gives such a lower bound. The proof may be found in Appendix E.

Proposition 2:

$$\Pr\left\{\boldsymbol{s}(t; \hat{\boldsymbol{\theta}}) \in \mathcal{U}_\beta, \forall\, t \in I\right\} \geq 1 - (1 + \kappa_o \beta) \exp\left(-\frac{\beta^2}{2}\right) \tag{56}$$

where the constant $\kappa_o(t)$ is given by

$$\kappa_o \triangleq \int_0^T \left(\frac{\zeta(t)}{2\pi}\right)^{1/2} dt \tag{57}$$

$\zeta(t)$ is given by

$$\zeta(t) = ||\dot{\boldsymbol{v}}_1(t)||^2 + ||\dot{\boldsymbol{v}}_2(t)||^2 - 2(\dot{\boldsymbol{v}}_1(t)^T \boldsymbol{v}_2(t))^2 \tag{58}$$

and $\boldsymbol{v}_1(t)$ and $\boldsymbol{v}_2(t)$ denote the orthonormal basis vectors computed by the Gram–Schmidt procedure (32) and (33).

The computation of the derivatives of the orthonormal vectors $\boldsymbol{v}_1(t)$ and $\boldsymbol{v}_2(t)$ can be done using the explicit formulation of (32) and (33). However, Proposition 3 presents a simpler technique to compute the function $\zeta(t)$ of (58). The proof may be found in Appendix F.

Proposition 3: Suppose $\boldsymbol{C}_{\boldsymbol{\theta}}^{1/2}\boldsymbol{B}(t)$ has full column rank, the corresponding QR factorization is given by (34), and

$$\boldsymbol{A}(t) = \boldsymbol{C}_{\boldsymbol{\theta}}^{1/2}\boldsymbol{B}(t)\boldsymbol{B}(t)^T\boldsymbol{C}_{\boldsymbol{\theta}}^{1/2} = \boldsymbol{V}(t)\boldsymbol{G}(t)^T\boldsymbol{V}(t)^T \tag{59}$$

where $\boldsymbol{G}(t) = \boldsymbol{D}(t)\boldsymbol{D}(t)^T$. Then, $\zeta(t)$ in (58) is given by

$$\begin{aligned} \zeta(t) &= \sum_{i=1}^{2}\{\boldsymbol{e}_i^T \boldsymbol{G}^{-1}\boldsymbol{V}^T \dot{\boldsymbol{A}}\boldsymbol{P}_{\boldsymbol{A}}^{\perp}\dot{\boldsymbol{A}}\boldsymbol{V}\boldsymbol{G}^{-1}\boldsymbol{e}_i\} && (60) \\ &= \mathrm{trace}\{\boldsymbol{P}_{\boldsymbol{A}}^{\perp}\dot{\boldsymbol{A}}\boldsymbol{A}^{\dagger}\boldsymbol{A}^{\dagger}\dot{\boldsymbol{A}}\boldsymbol{P}_{\boldsymbol{A}}^{\perp}\} && (61) \end{aligned}$$

where $\boldsymbol{e}_1 = [1, 0]^T$ and $\boldsymbol{e}_2 = [0, 1]^T$; $\boldsymbol{A}^{\dagger} = \boldsymbol{V}\boldsymbol{G}^{-1}\boldsymbol{V}^T$ denotes the pseudo-inverse of $\boldsymbol{A}$; and $\boldsymbol{P}_{\boldsymbol{A}}^{\perp}$ denotes the projection on the orthogonal complement of the range space of $\boldsymbol{A}$.

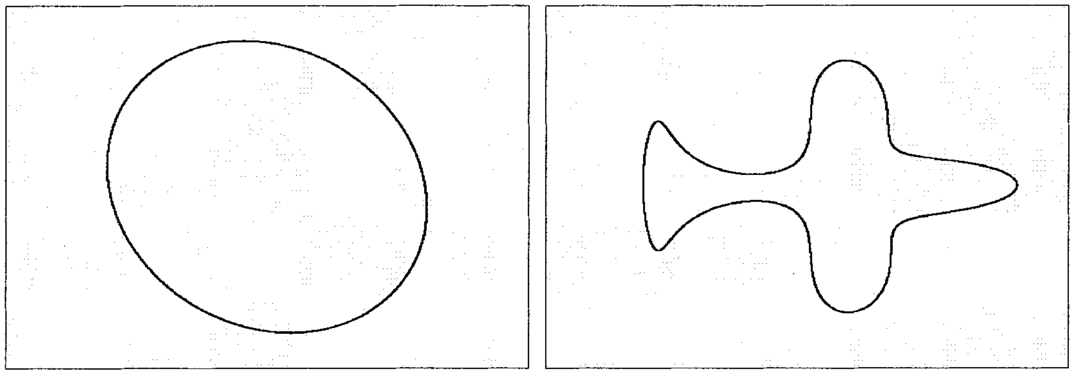
As observed in (61), the value of $\zeta(t)$ is independent of the choice of the orthonormal basis for $\boldsymbol{C}_{\boldsymbol{\theta}}^{1/2}\boldsymbol{B}(t)$, because all choices of orthonormal basis vectors yield the same subspace $\boldsymbol{P}_{\boldsymbol{A}}^{\perp}$. Furthermore, Proposition 3 suggests a numerically stable way to compute $\zeta(t)$ since derivatives are required only for $\boldsymbol{A}$, and the rest of the computation can be done numerically by QR factorization of $\boldsymbol{C}_{\boldsymbol{\theta}}^{1/2}\boldsymbol{B}(t)$. The derivative of $\boldsymbol{A}$ is readily given by

$$\dot{\boldsymbol{A}} = \boldsymbol{C}_{\boldsymbol{\theta}}^{1/2}\left(\dot{\boldsymbol{B}}\boldsymbol{B}^T + \boldsymbol{B}\dot{\boldsymbol{B}}^T\right)\boldsymbol{C}_{\boldsymbol{\theta}}^{1/2} \tag{62}$$

where, in many parameterizations such as Fourier descriptors, the derivative $\dot{\boldsymbol{B}}$ of the basis matrix can be computed analytically. Therefore, we can obtain accurate numerical values for $\zeta(t)$ and the level-crossing bounds by using (61).

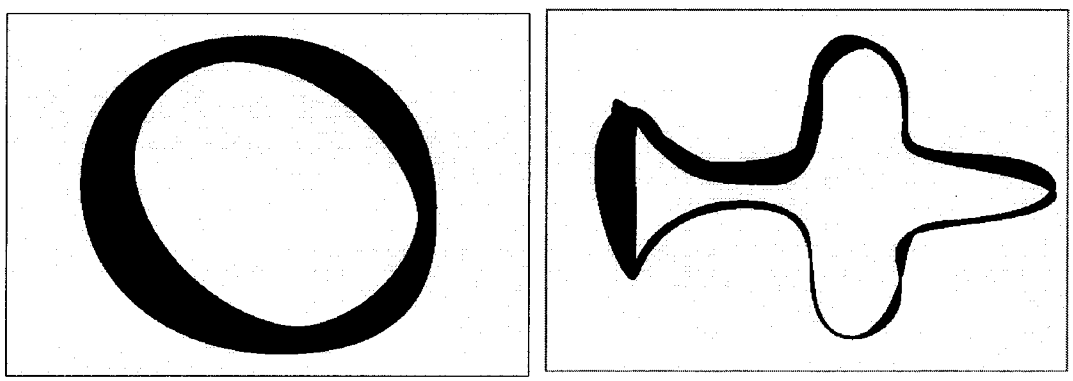
IV. Numerical Results

We now compute bounds for shape estimation in a practical scenario that arises in a passive radar imaging problem and compare the results with Monte Carlo simulations (for more details about passive radar imaging problems, see [34]). In this



(a) (b)

Fig. 2. Examples of shape $\boldsymbol{s}(t;\,\theta)$ parametrized by the Fourier descriptors of (64). (a) $L = 1,\ K = 5$. (b) $L = 7,\ K = 29$.



(a) (b)

Fig. 3. Examples of confidence regions with $\beta = 4$ for the shapes of Fig. 2. (a) $L = 1,\ K = 5$. (b) $L = 7,\ K = 29$.

problem, the target for shape estimation is parameterized by Fourier descriptors

$$\boldsymbol{s}(t;\,\boldsymbol{\theta}) = \boldsymbol{B}(t)^T\boldsymbol{\theta}, \qquad t \in [0,\,T] \tag{63}$$

where the matrix of basis functions $\boldsymbol{B}(t)$ is given by

$$\boldsymbol{B}(t)^T = \begin{bmatrix} \boldsymbol{a}(t) & \boldsymbol{0} \\ \boldsymbol{r}(t) & \boldsymbol{c}(t) \end{bmatrix} \in \mathbb{R}^{2\times K} \tag{64}$$

where $K = 4L + 1$ and

$$\boldsymbol{a}(t) = [1\ \cos(t)\ \cdots\ \cos(Lt)\ \sin(t)\ \cdots\ \sin(Lt)] \in \mathbb{R}^{2L+1} \tag{65}$$

$$\boldsymbol{r}(t) = \Big[\overbrace{0\cdots0}^{L+1}\ \ \cos(t)\ \ \overbrace{0\cdots0}^{L-1}\Big] \in \mathbb{R}^{2L+1} \tag{66}$$

$$\boldsymbol{c}(t) = [1\ \cos(2t)\ \cdots\ \cos(Lt)\ \sin(t)\ \cdots\ \sin(Lt)] \in \mathbb{R}^{2L}. \tag{67}$$

The measurement model (3) and the derivation of the Cramér–Rao covariance matrix $\boldsymbol{C_\theta}$ in the passive radar problem are very involved due to the nonlinear nature of inverse scattering phenomena, and are explained in detail in a separate publication [34]. Therefore, in this simulation, we just use the covariance matrix computed in [34]. Two examples of shape $\boldsymbol{s}(t;\,\boldsymbol{\theta})$ parameterized by (64) are given in Fig. 2(a) and (b) for $L = 1 (K = 5)$ and $L = 7 (K = 29)$, respectively. The true parameter $\boldsymbol{\theta} \in \mathbb{R}^5$ and an example of the CRB matrix $\boldsymbol{C_\theta} \in \mathbb{R}^{5\times5}$ for Fig. 2(a) are given by $\boldsymbol{\theta} = [0.0,\ 8.6,\ 0.7,\ 0.0,\ 11.4]^T$ and

$$\boldsymbol{C_\theta} = \begin{bmatrix} 0.0309 & -0.0287 & 0.0078 & -0.0130 & 0.0069 \\ -0.0287 & 0.0424 & -0.0160 & 0.0157 & -0.0073 \\ 0.0078 & -0.0160 & 0.0758 & -0.0026 & -0.0143 \\ -0.0130 & 0.0157 & -0.0026 & 0.0312 & 0.0024 \\ 0.0069 & -0.0073 & -0.0143 & 0.0024 & 0.0367 \end{bmatrix} \tag{68}$$

while the true parameter $\boldsymbol{\theta} \in \mathbb{R}^{29}$ and an example of $\boldsymbol{C_\theta} \in \mathbb{R}^{29\times29}$ for Fig. 2(b) can be obtained from [35]. Examples of the deterministic confidence regions $\mathcal{U}_\beta$ for Fig. 2(a) and (b) with $\beta = 4$ are shown in Fig. 3(a) and (b), using the $\boldsymbol{C_\theta}$ matrices of (68) and [35], respectively. It is interesting to note that the overall confidence region $\mathcal{U}_\beta$ is produced by a morphological dilation of $\boldsymbol{s}(t)$ with varying structural element [36]. In these examples, thicker confidence regions are generated where the information of the scattered waves is insufficient, implying difficulties of shape estimation in that portion of the target (for more discussion of the relation of the geometry of the confidence regions to receiver locations in passive radar problems, see [34] or [35]).

The actual level-crossing probability

$$p_e = 1 - \Pr\left\{\boldsymbol{s}(t;\,\hat{\boldsymbol{\theta}}) \in \mathcal{U}_\beta,\ \forall\, t \in I\right\} \tag{69}$$

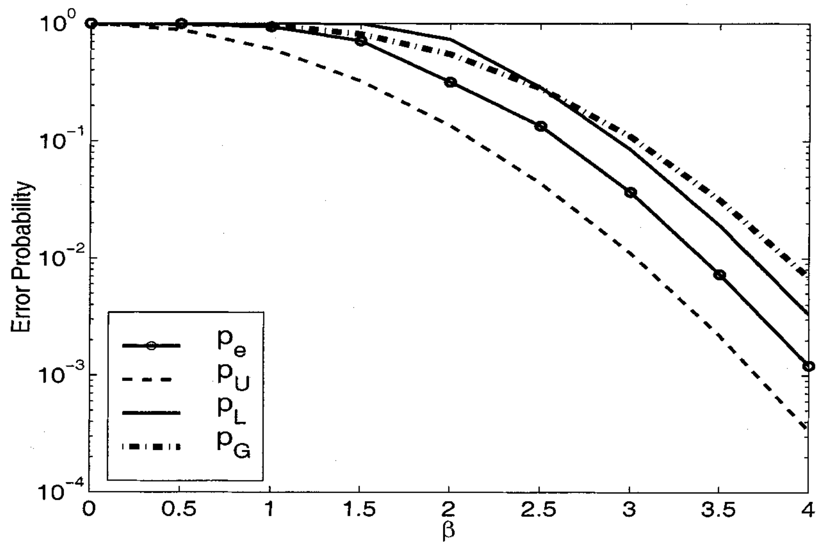


Fig. 4. Comparison of Monte Carlo simulation (p_e) with level-crossing bound (p_L), incomplete gamma bound (p_G), and an approximation (p_U) for the shape in Fig. 2 (a) $(K = 5)$.

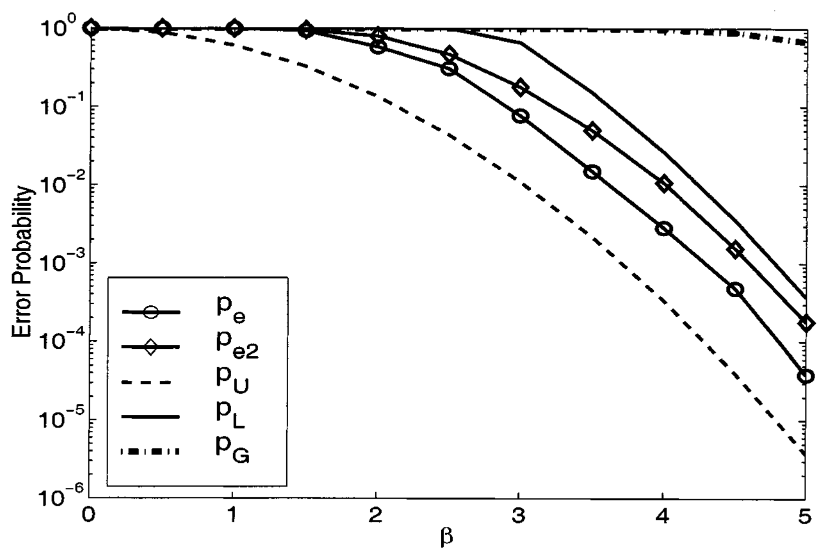


Fig. 5. Comparison of Monte Carlo simulation $(p_e,\ p_{e2})$ with level-crossing bound (p_L), incomplete gamma bound (p_G), and an approximation (p_U) for the shape in Fig. 2 (b) $(K = 29)$.

is computed by Monte Carlo simulation, and is compared with the incomplete-gamma bound and the level-crossing bound. The results are shown in Figs. 4 and 5. A display of p_e in a log-scale plot is chosen because it is more informative for large β than displaying $1 - p_e$. Monte Carlo simulations were performed by drawing a random vector $\hat{\boldsymbol{\theta}}$ from a Gaussian distribution with mean $\boldsymbol{\theta}$ and covariance matrix $\boldsymbol{C}_{\boldsymbol{\theta}}$, generating the shape $\boldsymbol{s}(t;\, \hat{\boldsymbol{\theta}})$ and counting the events $\{\boldsymbol{s}(t;\, \hat{\boldsymbol{\theta}}) \in \mathcal{U}_\beta, \forall t \in I\}$. The number of Monte Carlo runs N is determined by a confidence measure known as the *coefficient of variation*, $\sqrt{(1-p_e)/Np_e}$ [37]. When p_e is very small

$$N \geq \frac{1}{\eta^2 p_e} \tag{70}$$

where η is a confidence level for the Monte Carlo estimate (in this paper, we selected $\eta = 0.1$).

The incomplete-gamma bound and the level-crossing bound for the error probability p_e are rewritten as follows:

$$p_G = 1 - \Gamma\left(\frac{K}{2}, \frac{\beta^2}{2}\right) \tag{71}$$

$$p_L = (1 + \kappa\beta)\exp\left(-\frac{\beta^2}{2}\right) \tag{72}$$

so that $p_e \leq p_G$ and $p_e \leq p_L$. We found that the following probability is also useful for comparison with p_L and p_G:

$$p_U = \exp\left(-\frac{\beta^2}{2}\right). \tag{73}$$

This quantity would be an approximation to p_e if the pointwise probability that $\boldsymbol{s}(t;\, \hat{\boldsymbol{\theta}})$ lies outside of $\mathcal{U}_\beta(t)$ were approximately equal to p_e.

Figs. 4 and 5 illustrate the simulation results for Fig. 2(a) and (b), respectively, for different values of β, using the CRB matrices $\boldsymbol{C_\theta}$ of (68) and [35], respectively. The incomplete gamma bound p_G is useful for small K and β, but is very conservative for large K. The approximation p_U degrades with increasing number of parameters K. However, the level-crossing bound p_L is reasonably tight for a wide range of β and K. Fig. 5 also plots Monte Carlo estimates of

$$p_{e2} = 1 - \Pr\left\{\boldsymbol{s}(t; \hat{\boldsymbol{\theta}}) \in \mathcal{U}_\beta(t), \forall\, t \in I\right\} \tag{74}$$

which, as discussed below (19), is an upper bound on p_e. Recall that our level-crossing bound is derived as an upper bound on (74). Fig. 5 clearly confirms that our level-crossing bound is a tight bound on p_{e2} even for large K.

V. Further Extensions and Applications

In this section, we consider several extensions and applications of our basic setup.

A. (Nearly) Nondifferentiable Boundary Points

It would be interesting to consider the case when there are (nearly) nondifferentiable points along the boundary. The situation would be of interest because, for example, man-made objects have sharp boundaries. For nearly nondifferentiable boundaries, in general, the level-crossing rate $h^*(t, \beta)$ becomes large since $\dot{\boldsymbol{q}}_1$ and $\dot{\boldsymbol{q}}_2$ (and, therefore, $\zeta(t, \omega^*)$) are large. Furthermore, in nondifferentiable cases, the level-crossing rate becomes infinite and hence Rice's formula provides a trivial bound. However, there exists a large deviation result for stationary Gaussian random processes [11, Lemma 12.2.1] which provides nontrivial bounds even for nondifferentiable random processes. We conjecture that a result similar to [11, Lemma 12.2.1] could be obtained for our nonstationary χ^2 random processes providing nontrivial bounds even in the nondifferentiable case (but rigorous analysis is beyond the scope of this paper). Therefore, we expect the level-crossing bounds presented in this paper to become weaker in the presence of (nearly) nondifferentiable points along the boundary.

B. Alternative Designs of Global Confidence Region

Recall that we fix the value of β of (12) and (17) for all $t \in I$ to construct global confidence regions. (A similar construction is used in the robust control literature for estimation of the Nyquist plot for certain systems, although no global confidence level have been studied. See [38] and references therein). However, a different global confidence region can be constructed by making β dependent upon t. In this case, it is not difficult to show that the level crossing bound becomes

$$\begin{aligned}\Pr&\left\{\boldsymbol{s}(t; \hat{\boldsymbol{\theta}}) \in \mathcal{U}, \forall\, t \in I\right\}\\ &\geq 1 - \exp\left(-\frac{\beta(0)^2}{2}\right) - \int_0^T \left(\frac{\zeta(t)}{2\pi}\right)^{1/2}\\ &\quad\cdot \beta(t) \exp\left(-\frac{\beta(t)^2}{2}\right)\, dt\end{aligned} \tag{75}$$

where the constant $\zeta(t)$ is again given by (61). How should $\beta(t)$ be chosen? One possible choice may be the function $\beta(t)$ that minimizes the area of the global confidence region for a given confidence level. This would be analogous to the minimum area local confidence region by Wilks and Daly [28]. We are currently investigating this issue.

C. Global Confidence Regions for Subset of Boundary

Proposition 2 can be extended to consider a subset of the boundary, say for $[t_1, t_2] \subset [0, T]$. In this case, the only modification to Proposition 2 is to change the limits of the integral of (57) to $[t_1, t_2]$, making the bounds tighter.

D. Bounds on Hausdorff Distances

The level-crossing bound can be used to derive a lower bound for measuring the quality of boundary estimation in terms of the Hausdorff distance $d_H(\boldsymbol{s}, \hat{\boldsymbol{s}})$ between the true boundary $\boldsymbol{s}$ and its estimate $\hat{\boldsymbol{s}}$. The Hausdorff distance is defined as [39, p. 279]

$$d_H(\boldsymbol{s}, \hat{\boldsymbol{s}}) = \text{glb}\{\gamma | \boldsymbol{s} \subset U(\hat{\boldsymbol{s}}, \gamma), \hat{\boldsymbol{s}} \subset U(\boldsymbol{s}, \gamma)\} \tag{76}$$

where glb$\{\cdot\}$ denotes the greatest lower bound, and

$$U(\boldsymbol{s}, \gamma) = \bigcup_{t \in I} B_d(\boldsymbol{s}(t), \gamma) \tag{77}$$

where $B_d(\boldsymbol{s}(t), \gamma)$ is the open ball with radius γ, centered at $\boldsymbol{s}(t)$. Here, the regions $U(\boldsymbol{s}, \gamma)$ and $U(\hat{\boldsymbol{s}}, \gamma)$ form "tubes" around $\boldsymbol{s}(t)$ and $\hat{\boldsymbol{s}}(t)$, respectively. However, unlike the global confidence region in (13) and (16), $U(\boldsymbol{s}, \gamma)$ and $U(\hat{\boldsymbol{s}}, \gamma)$ are tubes with uniform widths.

The Hausdorff distance has been widely used in pattern recognition applications such as pattern matching algorithms, because this distance is a measure of differences between two objects. Hence, the probability

$$\Pr\{d_H(\boldsymbol{s}, \hat{\boldsymbol{s}}) \leq \epsilon\} \tag{78}$$

is useful to quantify shape estimator performance, as well as classifier performance [40]. For example, in a template-based pattern matching problem, when the Hausdorff distance between the template $\boldsymbol{s}$ and the estimated object boundary $\hat{\boldsymbol{s}}$ is compared to a threshold ϵ, $\Pr\{d_H(\boldsymbol{s}, \hat{\boldsymbol{s}}) \leq \epsilon\}$ is the probability that $\hat{\boldsymbol{s}}$ is recognized as a noisy instance of $\boldsymbol{s}$. To relate the Hausdorff distance to our global confidence regions $\mathcal{U}_\beta$ and $\hat{\mathcal{U}}_\beta$, we pick β so that

$$\mathcal{U}_\beta \subset U(\boldsymbol{s}, \epsilon) \qquad \hat{\mathcal{U}}_\beta \subset U(\hat{\boldsymbol{s}}, \epsilon). \tag{79}$$

To this end, we choose

$$\begin{aligned}\beta &= \frac{\epsilon}{\sqrt{\lambda_{\max}}}\\ \lambda_{\max} &= \max_{t \in I} \lambda_M(t)\end{aligned} \tag{80}$$

where $\lambda_M(t)$ denotes the largest eigenvalue of $\boldsymbol{C_s}(t)$. This ensures the largest width of $\mathcal{U}_\beta$ and $\hat{\mathcal{U}}_\beta$ is less than or equal to $\beta\sqrt{\lambda_{\max}}$. Then, we have

$$\begin{aligned}\Pr\{d_H(\boldsymbol{s}, \hat{\boldsymbol{s}}) \leq \epsilon\} &\geq \Pr\{\hat{\boldsymbol{s}}(t) \in \mathcal{U}_\beta, \forall\, t \in I\}\\ &\geq 1 - \left(1 + \kappa_o \frac{\epsilon}{\sqrt{\lambda_{\max}}}\right) \exp\left(-\frac{\epsilon^2}{2\lambda_{\max}}\right)\end{aligned} \tag{81}$$

where κ_o is given by (57). The lower bound (81) can be easily improved if we allow β to depend upon t

$$\beta(t) = \frac{\epsilon}{\sqrt{\lambda_M(t)}} \qquad \forall t \in I. \tag{82}$$

Using (75), we have

$$\Pr\{d_H(\boldsymbol{s}, \hat{\boldsymbol{s}}) \leq \epsilon\} \geq 1 - \exp\left(-\frac{\epsilon^2}{2\lambda_M(0)}\right) - \int_0^T \left(\frac{\zeta(t)}{2\pi}\right)^{1/2} \frac{\epsilon}{\sqrt{\lambda_M(t)}} \exp\left(-\frac{\epsilon^2}{2\lambda_M(t)}\right) dt \tag{83}$$

where $\zeta(t)$ is given by (61).

E. Application to Medical Imaging

Finally, we discuss a possible application of global confidence regions to a medical imaging problem: emission computed tomography (ECT) with high-resolution anatomical side information obtained from magnetic resonance imaging (MRI) [8]. ECT is a functional imaging modality that reconstructs the tracer uptake from gamma rays emitted from a patient after injection of a radioactive tracer. MRI is an anatomical imaging modality which images tissue structures based on proton spin transverse magnetization differences. While MRI has very high spatial resolution and produces detailed images of anatomical structures, ECT has much lower spatial resolution than MRI but can track a large number of different biochemical compounds as they are metabolized by organs within the body. Recently, it has been recognized that when functional and anatomical organ boundaries are spatially correlated, MRI and ECT data can be combined to improve the accuracy of the ECT image. Hero *et al.* [8] presented a minimax methodology for combining information from these two imaging modalities and showed that the estimator is asymptotically equivalent to a penalized maximum-likelihood (PML) estimator with resolution–selective Gibbs penalty. More specifically, if the MRI boundary is again parameterized by $\boldsymbol{s}(t; \boldsymbol{\theta}) = \boldsymbol{B}^T(t)\boldsymbol{\theta}$, $t \in I$ and the measured emission data is a vector of counts $\boldsymbol{Y}_E$, then the asymptotic minimax estimator for the discretized spatial density $\boldsymbol{\lambda} = [\lambda_1, \cdots, \lambda_P]^T$ is given by

$$\hat{\boldsymbol{\lambda}} = \arg\max_{\boldsymbol{\lambda}} \left\{ \ln f(\boldsymbol{Y}_E|\boldsymbol{\lambda}) - \alpha \sum_{j=1}^{P} \sum_{k \in \mathcal{N}_j} \tilde{w}_{jk}(\hat{\boldsymbol{\theta}})(\lambda_j - \lambda_k)^2 \right\} \tag{84}$$

where α is the regularization parameter, $\mathcal{N}_j$ denote the neighborhood of λ_j, and

$$\tilde{w}_{jk}(\hat{\boldsymbol{\theta}}) = \int_{\mathbb{R}^K} \omega_{jk}(\boldsymbol{\theta}) \frac{|\boldsymbol{F}_{\hat{\boldsymbol{\theta}}}|^{\frac{K}{2}}}{(2\pi)^{\frac{K}{2}}} \cdot \exp\left(-\frac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T \boldsymbol{F}_{\hat{\boldsymbol{\theta}}}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\right) d\boldsymbol{\theta} \tag{85}$$

where $\boldsymbol{F}_{\hat{\boldsymbol{\theta}}}$ denotes the empirical Fisher-information matrix. In (85), $\omega_{jk}(\boldsymbol{\theta}) = 1$ if pixels j and k are both within the boundary specified by $\boldsymbol{s}(t; \boldsymbol{\theta}) = \boldsymbol{B}^T(t)\boldsymbol{\theta}$ or if they are completely outside of the boundary. Otherwise, $\omega_{jk}(\boldsymbol{\theta})$ is set to zero when j and k are on opposite sides of the boundary. If $\tilde{\omega}_{jk}$ in (84) were set equal to ω_{jk}, this assignment would completely decouple pixels on either side of the boundary yet still encourage smoothness within and exterior to the boundary. Instead, (85) accounts for the uncertainty in the estimated boundary by smoothing the ω_{jk}. While (85) is optimal in minimax sense, the exact computation of (85) is quite expensive since it requires Monte Carlo integration or a k-dimensional Fourier transform (for example, K may be a few tens). We are aware of fast Fourier transform results in [41], but their application is limited to star-shape objects. Instead, our global confidence region may be used as a computationally efficient alternative. For a given confidence level determined by (56), a posterior confidence region $\hat{\mathcal{U}}_\beta$ is constructed by (13), and we may set the weights $\tilde{w}_{jk}(\hat{\boldsymbol{\theta}}) = 0$ if pixels j or k are within the global confidence region $\hat{\mathcal{U}}_\beta$ of (13), and $\tilde{w}_{jk}(\hat{\boldsymbol{\theta}}) = 1$ otherwise. This assignment would completely decouple pixels in the uncertainty region, while encouraging smoothness outside of the uncertainty region.

VI. Conclusions

In this paper, we proposed a new concept for analyzing the performance of two-dimensional parametric shape estimators using a confidence region visualization technique. Cramér–Rao bounds for estimation of unknown geometric parameters such as shape, size, orientation, and the position of objects are used to construct a confidence region within which the shape estimate resides with a prescribed probability. The geometric properties of the confidence region convey information about which parts of the shape are difficult to estimates and can be used to optimize imaging system parameters. Computing the actual probability that the estimate lies in the confidence region is, however, quite a challenging problem because a shape estimate is usually a two-dimensional nonstationary random process. We derived two bounds on this probability: the incomplete-gamma and level-crossing bounds. The incomplete-gamma bound was derived using the basic inequalities of a projection without considering the characteristics of the CRB matrix and the basis functions for parameterization. As expected, this bound turned out to be very conservative. However, using the level-crossing statistics, we derived a tighter level-crossing bound which can be obtained in a numerically stable way using QR decomposition. We also showed that our bounds are valid lower bounds for the probabilities that the true shape lies in *a posteriori* confidence regions which are created around the MLE, which is often encountered in practical estimation algorithms.

Numerical simulations for a nonlinear inverse scattering problem showed that our level-crossing bound is tight enough for a broad range of confidence levels and number of unknown parameters. Based on the theoretical results and several test applications, our global confidence region may prove useful in a variety of other applications of parametric shape estimation. Furthermore, owing to the generality of our model and approach, the result is applicable to linear and nonlinear shape representations, complicated observation models, etc.

APPENDIX A
NONSTATIONARITY OF $R^2(t)$

Let $\boldsymbol{u} \triangleq \boldsymbol{P}(t)\boldsymbol{Z}$ and $\boldsymbol{v} \triangleq \boldsymbol{P}(s)\boldsymbol{Z}$, where the projection matrix $\boldsymbol{P}(t)$ is given in (38). Then, $E[\boldsymbol{u}^T\boldsymbol{v}] = \mathrm{tr}\{\boldsymbol{P}(s)\boldsymbol{P}(t)\}$, $E[\boldsymbol{Z}^T\boldsymbol{v}] = \mathrm{tr}\{\boldsymbol{P}(s)\} = 2$, and $E[\boldsymbol{Z}^T\boldsymbol{u}] = \mathrm{tr}\{\boldsymbol{P}(t)\} = 2$. The mean of $R^2(t)$ is $E[R^2(t)] = E[\boldsymbol{Z}^T\boldsymbol{v}] = 2$, and the autocorrelation function $\phi(t, s) \triangleq E[R^2(t)R^2(s)]$ is computed using the Gaussian moment theorem

$$\begin{aligned}\phi(t, s) &= E\left[\boldsymbol{Z}^T\boldsymbol{P}(t)\boldsymbol{Z}\boldsymbol{Z}^T\boldsymbol{P}(s)\boldsymbol{Z}\right]\\ &= E\left[\boldsymbol{Z}^T\boldsymbol{u}\boldsymbol{Z}^T\boldsymbol{v}\right]\\ &= 2E\left[\boldsymbol{Z}^T\boldsymbol{u}\right]E\left[\boldsymbol{Z}^T\boldsymbol{v}\right] + E\left[\boldsymbol{Z}^T\boldsymbol{Z}\right]E[\boldsymbol{v}^T\boldsymbol{u}]\\ &= 8 + K\mathrm{tr}\{\boldsymbol{P}(s)\boldsymbol{P}(t)\}.\end{aligned} \tag{A.1}$$

The form of $\phi(t, s)$ in (A.1) suggests that for most choices of $\boldsymbol{P}(s)$, $\phi(t, s) \neq \phi(t - s)$, thus $R^2(t)$ is nonstationary. However, there are some nontrivial examples of $\boldsymbol{P}(t)$ that allow the process $R^2(t)$ to be wide-sense stationary. For example, the projection

$$\boldsymbol{P}(t) = \frac{1}{2}\begin{bmatrix} 1+\cos(2t) & \sin(2t) & 0 & 0\\ \sin(2t) & 1-\cos(2t) & 0 & 0\\ 0 & 0 & 1+\cos(2t) & \sin(2t)\\ 0 & 0 & \sin(2t) & 1-\cos(2t)\end{bmatrix} \tag{A.2}$$

is a rank-two projection that depends on t, but (A.1) shows that $R^2(t)$ is wide-sense stationary since

$$\phi(t, s) = 8 + K\mathrm{tr}\{\boldsymbol{P}(s)\boldsymbol{P}(t)\} = 12 + 4\cos 2(t - s). \tag{A.3}$$

Because $\boldsymbol{P}(t)$ is the projection onto the range space of $\boldsymbol{C}_{\boldsymbol{\theta}}^{1/2}\boldsymbol{B}(t)$, the condition for stationarity

$$\mathrm{tr}\{\boldsymbol{P}(s)\boldsymbol{P}(t)\} = \psi(t - s)$$

for some function $\psi(\cdot)$, depends on both $\boldsymbol{C}_{\boldsymbol{\theta}}$ and $\boldsymbol{B}(t)$. An explicit necessary and sufficient condition in terms of the measurement plus object model is an open problem.

APPENDIX B
PROOF OF LEMMA 1

In order to prove Lemma 1, the following lemmas are useful. For notational convenience, we define

$$\mu(t) \triangleq R^2(t) = \boldsymbol{Z}^T\boldsymbol{P}(t)\boldsymbol{Z}.$$

Lemma B.1 (Cramer and Leadbetter [10, pp. 76–77]): Let u be fixed. If the one-dimensional probability density function of the process $\xi(t)$ is bounded at every $0 \le t \le 1$, and if $\xi(t)$ has, with probability one, a continuous sample derivative, $\dot{\xi}(t)$, then the probability is zero that $\dot{\xi}(t) = 0$, $\xi(t) = u$ simultaneously, for any point t in $0 \le t \le 1$.

Lemma B.2: Suppose the basis matrix $\boldsymbol{B}(t)$ is continuous, continuously differentiable, and has full column rank for all $t \in I = [0, T]$. Then, the sample path $\mu(t) = \boldsymbol{Z}^T\boldsymbol{P}(t)\boldsymbol{Z}$ is continuous and has a continuous sample derivative for all $t \in I$ with probability one.

Proof: $\boldsymbol{P}(t)$ is the projection on the column space of $\boldsymbol{C}_{\boldsymbol{\theta}}^{1/2}\boldsymbol{B}(t)$, and $\boldsymbol{B}(t)$ is continuous and full column rank for all $t \in I$. Hence, the matrix-valued function $\boldsymbol{P}(t)$ is continuous. Therefore, $\boldsymbol{Z}^T\boldsymbol{P}(t)\boldsymbol{Z}$ is continuous for any fixed vector $\boldsymbol{Z}$. The derivative of the projection matrix $\boldsymbol{P}(t)$ is given by [42]

$$\dot{\boldsymbol{P}} = \boldsymbol{P}^{\perp}\dot{\boldsymbol{H}}\boldsymbol{H}^{\dagger} + \left(\boldsymbol{P}^{\perp}\dot{\boldsymbol{H}}\boldsymbol{H}^{\dagger}\right)^T \tag{B.1}$$

where $\boldsymbol{H} = \boldsymbol{C}_{\boldsymbol{\theta}}^{1/2}\boldsymbol{B}(t)$, $\boldsymbol{P}^{\perp}$ is the orthogonal complement of the range space of $\boldsymbol{C}_{\boldsymbol{\theta}}^{1/2}\boldsymbol{B}(t)$, $\dot{\boldsymbol{P}}$ and $\dot{\boldsymbol{H}}$ denote the derivatives of $\boldsymbol{P}(t)$ and $\boldsymbol{H}(t)$ with respect to t, and $\boldsymbol{H}^{\dagger}(t)$ denotes the pseudoinverse of $\boldsymbol{H}(t)$. In (B.1), $\boldsymbol{P}^{\perp}$ and $\boldsymbol{H}^{\dagger}$ are well defined and continuous if $\boldsymbol{B}(t)$ is continuous. Moreover, if $\boldsymbol{B}(t)$ is continuously differentiable with respect to t, then $\dot{\boldsymbol{H}}$ is also well defined, thus $\boldsymbol{P}(t)$ is continuously differentiable. Therefore, $\boldsymbol{Z}^T\boldsymbol{P}(t)\boldsymbol{Z}$ is continuously differentiable for any fixed vector $\boldsymbol{Z}$. Hence, all sample paths of $\mu(t)$ are continuous and continuously differentiable. $\square$

Lemma B.3: If $\mu(t) = 0$, then $\dot{\mu}(t) = 0$.

Proof: Using (B.1), we have

$$\begin{aligned}\dot{\mu}(t) &= \boldsymbol{Z}^T\dot{\boldsymbol{P}}(t)\boldsymbol{Z}\\ &= 2\boldsymbol{Z}^T\boldsymbol{P}^{\perp}(t)\dot{\boldsymbol{H}}\boldsymbol{H}^{\dagger}\boldsymbol{Z}\\ &= 2\boldsymbol{Z}^T\boldsymbol{P}^{\perp}(t)\dot{\boldsymbol{H}}\boldsymbol{H}^{\dagger}\boldsymbol{P}(t)\boldsymbol{Z}.\end{aligned} \tag{B.2}$$

But $\mu(t) = 0 = \boldsymbol{Z}^T\boldsymbol{P}(t)\boldsymbol{Z}$ implies $\boldsymbol{P}(t)\boldsymbol{Z} = 0$. Therefore,

$$\dot{\mu}(t) = 2\boldsymbol{Z}^T\boldsymbol{P}^{\perp}(t)\dot{\boldsymbol{H}}\boldsymbol{H}^{\dagger}\boldsymbol{P}(t)\boldsymbol{Z} = 0. \qquad \square$$

Proof of Lemma 1: First, note that Lemma B.1 trivially generalizes to the domain $0 \le t \le T$ with $T < \infty$. Furthermore, for each t, the random process $\mu(t) = \boldsymbol{Z}^T\boldsymbol{P}(t)\boldsymbol{Z}$ is a χ^2 random variable, so its density function is bounded. In addition, Lemma B.2 guarantees that the random process $\mu(t)$ has, with probability one, a continuous sample derivative. Hence, all the assumptions for Lemma B.1 are satisfied for the random process $\mu(t)$. Now let $u = 0$ for Lemma B.1. Since

$$R(t) = \sqrt{\boldsymbol{Z}^T\boldsymbol{P}(t)\boldsymbol{Z}}$$

we have

$$\dot{R}(t) = \frac{\boldsymbol{Z}^T\dot{\boldsymbol{P}}(t)\boldsymbol{Z}}{2\sqrt{\boldsymbol{Z}^T\boldsymbol{P}(t)\boldsymbol{Z}}} = \frac{\dot{\mu}(t)}{2\sqrt{\mu(t)}} \tag{B.3}$$

which may not be defined for $t \in [0, T]$ such that $\mu(t) = 0$. According to Lemma B.3, at these nondifferentiable points, $\dot{\mu}(t) = 0$. Hence, $R(t)$ is not differentiable at $t \in [0, T]$ if and only if $\mu(t) = 0$ and $\dot{\mu}(t) = 0$ simultaneously. Now, Lemma B.1 tells us that the probability is zero that $\dot{\mu}(t) = 0$ and $\mu(t) = 0$ simultaneously, for any point $t \in [0, T]$. In other words, the probability is zero that $R(t)$ is not differentiable for any point $t \in [0, T]$. Therefore, $R(t)$ is differentiable for all $t \in [0, T]$, with probability one. This concludes the proof. $\square$

APPENDIX C
CALCULATION OF $p(R, \dot{R}; t)$

In order to compute the zero crossing rate $h(t, \beta)$ for the envelope process $R(t)$, we need the joint probability density function $p(R, \dot{R}; t)$. According to Rice's original work [12], $p(R, \dot{R}; t)$ is derived from the joint density function for

$$\boldsymbol{x} = [x_1, x_2, x_3, x_4]^T = \left[I_c(t), \dot{I}_s(t), I_s(t), \dot{I}_c(t)\right]^T \quad \text{(C.1)}$$

by coordinate transformation to $(R, \dot{R}, \phi, \dot{\phi})$ such that

$$\begin{aligned} x_1 &= R\cos\phi \qquad & x_2 &= \dot{x}_3 = \dot{R}\sin\phi + R\cos\phi\dot{\phi} \\ x_3 &= R\sin\phi \qquad & x_4 &= \dot{x}_1 = \dot{R}\cos\phi - R\sin\phi\dot{\phi} \end{aligned} \quad \text{(C.2)}$$

for $0 \le \phi \le 2\pi$ and $-\infty \le \dot{\phi} \le \infty$. From definition (40), the four-dimensional vector $\boldsymbol{x}$ of (C.1) can be represented by

$$\boldsymbol{x}(t) \triangleq \boldsymbol{U}(t)^T \boldsymbol{Z} \quad \text{(C.3)}$$

where the deterministic matrix $\boldsymbol{U}(t)$ is given by

$$\boldsymbol{U}(t) = [\boldsymbol{q}_1(t) \quad \dot{\boldsymbol{q}}_2(t) \quad \boldsymbol{q}_2(t) \quad \dot{\boldsymbol{q}}_1(t)]^T \quad \text{(C.4)}$$

where $\dot{\boldsymbol{q}}_1(t)$ and $\dot{\boldsymbol{q}}_2(t)$ denote derivatives with respect to t. Note that $\boldsymbol{x}(t)$ is zero-mean Gaussian with covariance matrix $\boldsymbol{M} \in \mathbb{R}^{2\times 2}$

$$\boldsymbol{M} = E[\boldsymbol{x}\boldsymbol{x}^T] = \boldsymbol{U}(t)\boldsymbol{U}(t)^T \quad \text{(C.5)}$$

since $\boldsymbol{Z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$. Because $\boldsymbol{q}_1(t)$ and $\boldsymbol{q}_2(t)$ are unit vectors, we have

$$\boldsymbol{q}_1(t)^T\dot{\boldsymbol{q}}_1(t) = \boldsymbol{q}_2(t)^T\dot{\boldsymbol{q}}_2(t) = 0.$$

Furthermore, the orthonormality of $\boldsymbol{q}_1(t)$ and $\boldsymbol{q}_2(t)$ ($\boldsymbol{q}_1(t)^T\boldsymbol{q}_2(t) = 0$) implies

$$\boldsymbol{q}_1(t)^T\dot{\boldsymbol{q}}_2(t) + \boldsymbol{q}_2(t)^T\dot{\boldsymbol{q}}_1(t) = 0 \quad \text{(C.6)}$$

Therefore, the covariance matrix M becomes

$$\boldsymbol{M} = \begin{bmatrix} 1 & b_1(t) & 0 & 0 \\ b_1(t) & b_2(t) & 0 & \xi(t) \\ 0 & 0 & 1 & -b_1(t) \\ 0 & \xi(t) & -b_1(t) & b_4(t) \end{bmatrix} \quad \text{(C.7)}$$

where

$$\begin{aligned} b_1(t) &\triangleq \boldsymbol{q}_1(t)^T\dot{\boldsymbol{q}}_2(t) \qquad & b_2(t) &\triangleq \|\dot{\boldsymbol{q}}_2(t)\|^2 \\ b_4(t) &\triangleq \|\dot{\boldsymbol{q}}_1(t)\|^2 \qquad & \xi(t) &\triangleq \dot{\boldsymbol{q}}_1(t)^T\dot{\boldsymbol{q}}_2(t). \end{aligned} \quad \text{(C.8)}$$

The up-crossing rate of $R(t)$ is given by [12]

$$h(t, \beta) = \int_0^\infty d\dot{R}\,\dot{R}\,p(\beta, \dot{R}; t). \quad \text{(C.9)}$$

The probability density function $p(\beta, \dot{R}; t)$ is computed from the joint probability of $\boldsymbol{x}$ using the coordinate transformation (C.2)

$$\begin{aligned} p(R, \dot{R}, \phi, \dot{\phi}; t) &\triangleq \frac{p(x_1, x_2, x_3, x_4)}{|J(x_1, x_2, x_3, x_4)|} \\ &= \frac{R^2}{(2\pi)^2|\boldsymbol{M}|^{1/2}} \exp\left(-\frac{\boldsymbol{x}^T\boldsymbol{M}^{-1}\boldsymbol{x}}{2}\right) \end{aligned} \quad \text{(C.10)}$$

since the Jacobian is $|J(x_1, x_2, x_3, x_4)| = R^{-2}$. Note that in our formulation $b_4(t) \ne b_2(t)$ and $\xi(t) \ne 0$ in general, unlike Rice's original formulation for the stationary envelope process [14], [15]. This makes the computation of M^{-1} and hence $p(R, \dot{R}, \phi, \dot{\phi}; t)$ very complicated. Fortunately, $\xi(t)$ can be made zero by an appropriate choice of $\omega(t)$ in (36) and (37).

Lemma C.1: Let $\boldsymbol{v}_1(t)$ and $\boldsymbol{v}_2(t)$ denote the orthonormal basis vectors computed by the Gram–Schmidt procedure (32) and (33), and $\boldsymbol{v}_1(t)$ and $\boldsymbol{v}_2(t)$ be twice-differentiable. Then, for the general form of the differentiable orthonormal basis functions $\boldsymbol{q}_1(t, \omega)$ and $\boldsymbol{q}_2(t, \omega)$ given by (36) and (37), we have $\dot{\boldsymbol{q}}_1(t, \omega)^T\dot{\boldsymbol{q}}_2(t, \omega) = 0$ if and only if $\omega(t)$ is given by $\omega(t) = \omega^*(t)$ such that

$$\omega^*(t) = \frac{1}{2}\tan^{-1}\left(\frac{2\dot{\boldsymbol{v}}_1^T\dot{\boldsymbol{v}}_2}{\|\dot{\boldsymbol{v}}_1\|^2 - \|\dot{\boldsymbol{v}}_2\|^2}\right), \qquad -\frac{\pi}{4} < \omega^*(t) \le \frac{\pi}{4}. \quad \text{(C.11)}$$

Proof: Suppose $\omega(t)$ is differentiable with respect to t. Then

$$\dot{\boldsymbol{q}}_1(t, \omega) = \dot{\boldsymbol{v}}_1\cos\omega + \dot{\boldsymbol{v}}_2\sin\omega - \boldsymbol{v}_1\dot{\omega}\sin\omega + \boldsymbol{v}_2\dot{\omega}\cos\omega \quad \text{(C.12)}$$

$$\dot{\boldsymbol{q}}_2(t, \omega) = -\dot{\boldsymbol{v}}_1\sin\omega + \dot{\boldsymbol{v}}_2\cos\omega - \boldsymbol{v}_1\dot{\omega}\cos\omega - \boldsymbol{v}_2\dot{\omega}\sin\omega \quad \text{(C.13)}$$

where $\dot{\boldsymbol{v}}_1$, $\dot{\boldsymbol{v}}_2$ and $\dot{\omega}$ denote the derivative with respect to t, which are well defined owing to the Gram–Schmidt orthonormalization procedure (32), (33). Using $\boldsymbol{v}_1^T\boldsymbol{v}_1 = \boldsymbol{v}_2^T\boldsymbol{v}_2 = 1$, $\dot{\boldsymbol{v}}_1^T\boldsymbol{v}_1 = \dot{\boldsymbol{v}}_2^T\boldsymbol{v}_2 = 0$, and $\dot{\boldsymbol{v}}_1^T\boldsymbol{v}_2 = -\dot{\boldsymbol{v}}_2^T\boldsymbol{v}_1$, we have

$$\dot{\boldsymbol{q}}_1^T\dot{\boldsymbol{q}}_2 = -\frac{1}{2}\sin(2\omega)(\|\dot{\boldsymbol{v}}_1\|^2 - \|\dot{\boldsymbol{v}}_2\|^2) + \cos(2\omega)(\dot{\boldsymbol{v}}_1^T\dot{\boldsymbol{v}}_2). \quad \text{(C.14)}$$

Therefore, $\dot{\boldsymbol{q}}_1(t, \omega)^T\dot{\boldsymbol{q}}_2(t, \omega) = 0$ if and only if

$$\tan(2\omega) = \frac{2\dot{\boldsymbol{v}}_1^T\dot{\boldsymbol{v}}_2}{\|\dot{\boldsymbol{v}}_1\|^2 - \|\dot{\boldsymbol{v}}_2\|^2}$$

$$\Leftrightarrow \omega(t) = \omega^*(t) \triangleq \frac{1}{2}\tan^{-1}\left(\frac{2\dot{\boldsymbol{v}}_1^T\dot{\boldsymbol{v}}_2}{\|\dot{\boldsymbol{v}}_1\|^2 - \|\dot{\boldsymbol{v}}_2\|^2}\right). \quad \text{(C.15)}$$

To complete the proof, we must show that $\omega(t)$ of (C.15) is differentiable. This can be done by restricting the range of $\omega(t)$ such that $-\frac{\pi}{4} < \omega(t) \le \frac{\pi}{4}$, since $\boldsymbol{v}_1(t)$ and $\boldsymbol{v}_2(t)$ are twice-differentiable and $\tan^{-1}(\cdot)$ is differentiable. This concludes the proof. □

From the structure of M in (C.7) with $\xi(t) = 0$, we have

$$\boldsymbol{M}^{-1} = \begin{bmatrix} g_0 & g_1 & 0 & 0 \\ g_1 & g_2 & 0 & 0 \\ 0 & 0 & f_0 & -f_1 \\ 0 & 0 & -f_1 & f_2 \end{bmatrix} \quad \text{(C.16)}$$

where g_0, g_2, f_0, and f_2 are positive. Therefore,

$$\begin{aligned} \boldsymbol{x}^T\boldsymbol{M}^{-1}\boldsymbol{x} = &\, g_o x_1^2 + 2g_1x_1x_2 + g_2x_2^2 \\ &+ f_0x_3^2 - 2f_1x_3x_4 + f_2x_4^2. \end{aligned} \quad \text{(C.17)}$$

Using (C.2), $\boldsymbol{x}^TM^{-1}\boldsymbol{x}$ of (C.10) can be represented as

$$\begin{aligned} \boldsymbol{x}^T\boldsymbol{M}^{-1}\boldsymbol{x} = &\, g_0R^2\cos^2\phi + 2g_1\left(R\dot{R}\cos\phi\sin\phi + R^2\cos^2\phi\dot{\phi}\right) \\ &+ g_2\left(\left(\dot{R}\right)^2\sin^2\phi + R^2\cos^2\phi(\dot{\phi})^2\right. \\ &\left. + 2R\dot{R}\sin\phi\cos\phi\dot{\phi}\right) \\ &+ f_0R^2\sin^2\phi - 2f_1\left(R\dot{R}\cos\phi\sin\phi - R^2\sin^2\phi\dot{\phi}\right) \\ &+ f_2\left(\left(\dot{R}\right)^2\cos^2\phi + R^2\sin^2\phi(\dot{\phi})^2\right. \\ &\left. - 2R\dot{R}\sin\phi\cos\phi\dot{\phi}\right). \end{aligned} \quad \text{(C.18)}$$

After substituting (C.18) into (C.10), the probability density function for the level-crossing rate is obtained by integration

$$p\left(\beta, \dot{R}; t\right) = \int_0^{2\pi} d\phi \int_{-\infty}^{\infty} d\dot{\phi}\, p\left(\beta, \dot{R}, \phi, \dot{\phi}; t\right). \quad (C.19)$$

If $b_4(t) = b_2(t)$ in (C.7), then $g_i = f_i$, $i = 0, 1, 2$ and (C.18) can be greatly simplified as

$$\boldsymbol{x}^T \boldsymbol{M}^{-1} \boldsymbol{x} = g_0 R^2 + 2g_1 R^2 \dot{\phi} + g_2 \left(\left(\dot{R}\right)^2 + R^2 (\dot{\phi})^2 \right) \quad (C.20)$$

and the integration (C.19) is fairly easy, which was done in Rice's original work [14], [15]. However, the simplification of (C.20) cannot be applied to our problem since the magnitudes of derivatives $\|\dot{\boldsymbol{q}}_1\|$ and $\|\dot{\boldsymbol{q}}_2\|$ are in general different. As expected from (C.18), a closed-form expression for the integral in (C.19) is, in general, impossible and its evaluation may requires table lookup or series expansion.

APPENDIX D
PROOF OF LEMMA 2

Even though closed-form expressions for the integral in (C.19) are in general unavailable, the following property of $p(R, \dot{R}, \phi, \dot{\phi}; t)$ is useful for the proof of Lemma 2:

Lemma D.1: $p(R, \dot{R}, \phi, \dot{\phi}; t)$ is differentiable with respect to $\dot{R}$ and $\dot{\phi}$, and has exponential decay as $\dot{\phi} \to \pm\infty$ or $\dot{R} \to \pm\infty$.

Proof: Using (C.18), $\boldsymbol{x}^T M^{-1} \boldsymbol{x}$ can be represented as polynomial with respect to $\dot{R}$ or $\dot{\phi}$

$$\boldsymbol{x}^T M^{-1} \boldsymbol{x} = \left(\dot{R}\right)^2 (g_2 \sin^2\phi + f_2 \cos^2\phi) + k_1 \dot{R} + k_2 \quad (D.1)$$

$$\boldsymbol{x}^T M^{-1} \boldsymbol{x} = (\dot{\phi})^2 R^2 (g_2 \sin^2\phi + f_2 \cos^2\phi) + k_3 \dot{\phi} + k_4 \quad (D.2)$$

for appropriate constants k_i, $i = 1, \cdots, 4$. Note that

$$g_2 \sin^2\phi + f_2 \cos^2\phi > 0 \quad (D.3)$$

since g_2 and f_2 are positive. Therefore, (C.10) implies that the conditions of (D.1), (D.2), and (D.3) are enough to guarantee the differentiability and exponential decay of $p(R, \dot{R}, \phi, \dot{\phi}; t)$ with respect to $\dot{R}$ and $\dot{\phi}$. This concludes the proof. $\square$

Lemma D.2: Let M_{ij} denote the (i, j)th element of M matrix of (C.7) with $\xi(t) = 0$. Then

$$\frac{\partial h(t, \beta)}{\partial M_{44}} \geq 0 \quad (D.4)$$

where $h(t, \beta)$ is the level-crossing rate of (C.9).

Proof: The main idea of the proof is similar to the proof of Slepian's Lemma 1 [18]. Using (C.9) and (C.19), we have

$$h(t, \beta) = \int_0^{\infty} d\dot{R}\dot{R} \int_0^{2\pi} d\phi \int_{-\infty}^{\infty} d\dot{\phi} p\left(\beta, \dot{R}, \phi, \dot{\phi}; t\right). \quad (D.5)$$

By the characteristic function for the Gaussian

$$p(x_1, x_2, x_3, x_4) = \int_{-\infty}^{\infty} d\omega_1 \cdots d\omega_4 \exp\left(i \sum x_j \omega_j - \frac{1}{2} \sum M_{jk} \omega_j \omega_k \right) \quad (D.6)$$

we have

$$\frac{\partial p(x_1, x_2, x_3, x_4)}{\partial M_{44}} = \frac{1}{2} \frac{\partial^2 p(x_1, x_2, x_3, x_4)}{\partial x_4^2}. \quad (D.7)$$

From Rice's transformation (C.2), the inverse transformation is

$$R = \sqrt{x_1^2 + x_3^2} \qquad \phi = \tan^{-1}\left(\frac{x_3}{x_1}\right) \quad (D.8)$$

$$\dot{R} = x_2 \sin\phi + x_4 \cos\phi \qquad \dot{\phi} = x_2 \frac{\cos\phi}{R} - x_4 \frac{\sin\phi}{R}. \quad (D.9)$$

Equations (D.5) and (D.7) yield

$$\frac{\partial h(t, \beta)}{\partial M_{44}} = \frac{1}{2} \int_0^{\infty} d\dot{R}\dot{R} \int_0^{2\pi} d\phi \cdot \int_{-\infty}^{\infty} d\dot{\phi} \frac{\partial^2 p\left(\beta, \dot{R}, \phi, \dot{\phi}; t\right)}{\partial x_4^2}. \quad (D.10)$$

Using the chain rule as well as (D.8) and (D.9), we get

$$\frac{\partial^2 p\left(\beta, \dot{R}, \phi, \dot{\phi}; t\right)}{\partial x_4^2} = \left(\cos^2\phi \frac{\partial^2}{\partial \left(\dot{R}\right)^2} - \frac{\sin 2\phi}{\beta} \frac{\partial^2}{\partial \dot{R} \partial \dot{\phi}} + \frac{\sin^2\phi}{\beta^2} \frac{\partial^2}{\partial (\dot{\phi})^2} \right) \cdot p\left(\beta, \dot{R}, \phi, \dot{\phi}; t\right). \quad (D.11)$$

Therefore, the integral (D.10) has three terms owing to (D.11). The first term is computed by

$$\int_0^{\infty} d\dot{R}\dot{R} \int_0^{2\pi} d\phi \int_{-\infty}^{\infty} d\dot{\phi} \cos^2\phi \frac{\partial^2 p}{\partial \left(\dot{R}\right)^2} = \int_0^{2\pi} d\phi \cos^2\phi \int_{-\infty}^{\infty} d\dot{\phi} \left[\int_0^{\infty} d\dot{R}\dot{R} \frac{\partial^2 p}{\partial \left(\dot{R}\right)^2} \right] \quad (D.12)$$

where the term in the brackets is computed by integration by parts

$$\int_0^{\infty} d\dot{R}\dot{R} \frac{\partial^2 p}{\partial \left(\dot{R}\right)^2} = \left[\dot{R} \frac{\partial p}{\partial \dot{R}} \right]_{\dot{R}=0}^{\infty} - \int_0^{\infty} \frac{\partial p}{\partial \dot{R}} d\dot{R} = p(\beta, 0, \phi, \dot{\phi}). \quad (D.13)$$

In (D.13), we use the decay condition of Lemma D.1 to simplify. The last two terms of (D.10) are simpler to compute

$$\int_0^{\infty} d\dot{R}\dot{R} \int_0^{2\pi} d\phi \int_{-\infty}^{\infty} d\dot{\phi} \frac{\sin 2\phi}{\beta} \frac{\partial^2 p}{\partial \dot{R} \partial \dot{\phi}} = \frac{1}{\beta} \int_0^{\infty} d\dot{R}\dot{R} \int_0^{2\pi} d\phi \sin 2\phi \left[\frac{\partial p}{\partial \dot{R}} \right]_{\dot{\phi}=-\infty}^{\infty} = 0$$

$$\int_0^{\infty} d\dot{R}\dot{R} \int_0^{2\pi} d\phi \int_{-\infty}^{\infty} d\dot{\phi} \frac{\partial^2 p}{\partial (\dot{\phi})^2} \frac{\sin^2\phi}{\beta^2} = \frac{1}{\beta^2} \int_0^{\infty} d\dot{R}\dot{R} \int_0^{2\pi} d\phi \sin^2\phi \left[\frac{\partial p}{\partial \dot{\phi}} \right]_{\dot{\phi}=-\infty}^{\infty} = 0$$

where we again use Lemma D.1. Therefore, we have

$$\frac{\partial h(t, \beta)}{\partial M_{44}} = \frac{1}{2} \int_0^{2\pi} d\phi \cos^2\phi \int_{-\infty}^{\infty} d\dot{\phi} p(\beta, 0, \phi, \dot{\phi}) \geq 0 \quad (D.14)$$

since $p(\beta, 0, \phi, \dot{\phi}) \cos^2 \phi \geq 0$ for all ϕ and $\dot{\phi}$. This concludes the proof. □

Proof of Lemma 2: Choosing $\omega(t) = \omega^*(t)$ as in (53), Lemma C.1 guarantees that $\dot{\boldsymbol{q}}_1 = \dot{\boldsymbol{q}}_1(t, \omega^*)$ and $\dot{\boldsymbol{q}}_2 = \dot{\boldsymbol{q}}_2(t, \omega^*)$ are orthogonal and $\xi(t) = 0$ for the matrix $\boldsymbol{M}$ in (C.7). Therefore, we can use Lemma D.2. First, assume that $||\dot{\boldsymbol{q}}_1||^2 \leq ||\dot{\boldsymbol{q}}_2||^2$. Let $\boldsymbol{M}^*$ be a new covariance matrix formed by replacing $b_4(t)$ of (C.7) with $||\dot{\boldsymbol{q}}_2||^2$, and $h^*(t, \beta)$ be the corresponding level-crossing rate. Then, $\tilde{\boldsymbol{M}} = \rho \boldsymbol{M}^* + (1-\rho)\boldsymbol{M}$ is positive-definite for each $\rho \in [0, 1]$. Let $h(t, \beta; \tilde{\boldsymbol{M}})$ denote the level-crossing rate for the covariance matrix $\tilde{\boldsymbol{M}}$. Then, it is a function of ρ with derivative

$$\begin{aligned} \frac{dh\left(t, \beta; \tilde{\boldsymbol{M}}\right)}{d\rho} &= \frac{dh\left(t, \beta; \tilde{\boldsymbol{M}}\right)}{d\tilde{M}_{44}} \frac{d\tilde{M}_{44}}{d\rho} \\ &= \frac{dh\left(t, \beta; \tilde{\boldsymbol{M}}\right)}{d\tilde{M}_{44}} (||\dot{\boldsymbol{q}}_2||^2 - ||\dot{\boldsymbol{q}}_1||^2). \end{aligned} \tag{D.15}$$

Since $||\dot{\boldsymbol{q}}_2||^2 - ||\dot{\boldsymbol{q}}_1||^2 \geq 0$, Lemma D.2 gives

$$\frac{dh\left(t, \beta; \tilde{\boldsymbol{M}}\right)}{d\rho} \geq 0. \tag{D.16}$$

Integration with respect to ρ from 0 to 1 yields

$$h^*(t, \beta) = h(t, \beta; \boldsymbol{M}^*) \geq h(t, \beta; \boldsymbol{M}) = h(t, \beta). \tag{D.17}$$

Now $h^*(t, \beta)$ can be computed by Rice's formula for Rayleigh process with $b_4(t) = b_2(t)$ [14], [15]

$$h^*(t, \beta) = \left(\frac{||\dot{\boldsymbol{q}}_2||^2 - (\dot{\boldsymbol{q}}_1^T \boldsymbol{q}_2)^2}{2\pi}\right)^{1/2} \beta \exp\left(-\frac{\beta^2}{2}\right). \tag{D.18}$$

Now, assume that $||\dot{\boldsymbol{q}}_2||^2 \leq ||\dot{\boldsymbol{q}}_1||^2$. Redefine $\boldsymbol{x}$ such that

$$[x_1, x_2, x_3, x_4]^T = [I_s(t), \dot{I}_c(t), I_c(t), \dot{I}_s(t)]^T$$

to make $b_4(t) = ||\dot{\boldsymbol{q}}_2||^2$. Using a similar technique, we arrive at

$$h^*(t, \beta) = \left(\frac{||\dot{\boldsymbol{q}}_1||^2 - (\dot{\boldsymbol{q}}_2^T \boldsymbol{q}_1)^2}{2\pi}\right)^{1/2} \beta \exp\left(-\frac{\beta^2}{2}\right). \tag{D.19}$$

Since $(\dot{\boldsymbol{q}}_1^T \boldsymbol{q}_2)^2 = (\dot{\boldsymbol{q}}_2^T \boldsymbol{q}_1)^2$, this concludes the proof. □

APPENDIX E
PROOF OF PROPOSITION 2

Using (C.12) and (C.13), and $\boldsymbol{v}_1^T \boldsymbol{v}_1 = \boldsymbol{v}_2^T \boldsymbol{v}_2 = 1$, $\dot{\boldsymbol{v}}_1^T \boldsymbol{v}_1 = \dot{\boldsymbol{v}}_2^T \boldsymbol{v}_2 = 0$, and $\dot{\boldsymbol{v}}_1^T \boldsymbol{v}_2 = -\dot{\boldsymbol{v}}_2^T \boldsymbol{v}_1$, we have

$$\begin{aligned} ||\dot{\boldsymbol{q}}_1||^2 &= (\dot{\omega})^2 + 2\dot{\omega}(\boldsymbol{v}_2^T \dot{\boldsymbol{v}}_1) + ||\dot{\boldsymbol{v}}_1||^2 \cos^2 \omega \\ &\quad + ||\dot{\boldsymbol{v}}_2||^2 \sin^2 \omega + 2(\dot{\boldsymbol{v}}_1^T \dot{\boldsymbol{v}}_2) \sin \omega \cos \omega \qquad \text{(E.1)} \\ ||\dot{\boldsymbol{q}}_2||^2 &= (\dot{\omega})^2 + 2\dot{\omega}(\boldsymbol{v}_2^T \dot{\boldsymbol{v}}_1) + ||\dot{\boldsymbol{v}}_1||^2 \sin^2 \omega \\ &\quad + ||\dot{\boldsymbol{v}}_2||^2 \cos^2 \omega - 2(\dot{\boldsymbol{v}}_1^T \dot{\boldsymbol{v}}_2) \sin \omega \cos \omega \qquad \text{(E.2)} \\ \dot{\boldsymbol{q}}_1^T \boldsymbol{q}_2 &= (\dot{\boldsymbol{v}}_1^T \boldsymbol{v}_2) + \dot{\omega} = -\dot{\boldsymbol{q}}_2^T \boldsymbol{q}_1. \qquad \text{(E.3)} \end{aligned}$$

Hence, we have

$$\begin{aligned} ||\dot{\boldsymbol{q}}_1||^2 - (\dot{\boldsymbol{q}}_1^T \boldsymbol{q}_2)^2 &= ||\dot{\boldsymbol{v}}_1||^2 \cos^2 \omega + ||\dot{\boldsymbol{v}}_2||^2 \sin^2 \omega \\ &\quad + \sin(2\omega)(\dot{\boldsymbol{v}}_1^T \dot{\boldsymbol{v}}_2) - (\dot{\boldsymbol{v}}_1^T \boldsymbol{v}_2)^2 \qquad \text{(E.4)} \\ ||\dot{\boldsymbol{q}}_2||^2 - (\dot{\boldsymbol{q}}_2^T \boldsymbol{q}_1)^2 &= ||\dot{\boldsymbol{v}}_1||^2 \sin^2 \omega + ||\dot{\boldsymbol{v}}_2||^2 \cos^2 \omega \\ &\quad - \sin(2\omega)(\dot{\boldsymbol{v}}_1^T \dot{\boldsymbol{v}}_2) - (\dot{\boldsymbol{v}}_1^T \boldsymbol{v}_2)^2. \qquad \text{(E.5)} \end{aligned}$$

Therefore,

$$\begin{aligned} &\max\{||\dot{\boldsymbol{q}}_1||^2, ||\dot{\boldsymbol{q}}_2||^2\} - (\dot{\boldsymbol{q}}_1^T \boldsymbol{q}_2)^2 \\ &\quad \leq \left(||\dot{\boldsymbol{q}}_1||^2 - (\dot{\boldsymbol{q}}_1^T \boldsymbol{q}_2)^2\right) + \left(||\dot{\boldsymbol{q}}_2||^2 - (\dot{\boldsymbol{q}}_2^T \boldsymbol{q}_1)^2\right) \\ &\quad = ||\dot{\boldsymbol{v}}_1||^2 + ||\dot{\boldsymbol{v}}_2||^2 - 2(\dot{\boldsymbol{v}}_1^T \boldsymbol{v}_2)^2. \end{aligned} \tag{E.6}$$

This implies

$$\zeta(t, \omega) \leq \zeta(t) \triangleq ||\dot{\boldsymbol{v}}_1||^2 + ||\dot{\boldsymbol{v}}_2||^2 - 2(\dot{\boldsymbol{v}}_1^T \boldsymbol{v}_2)^2 \tag{E.7}$$

and

$$\kappa = \int_0^T \left(\frac{\zeta(t, \omega)}{2\pi}\right)^{1/2} dt \leq \kappa_o \triangleq \int_0^T \left(\frac{\zeta(t)}{2\pi}\right)^{1/2} dt. \tag{E.8}$$

This concludes the proof.

APPENDIX F
PROOF OF PROPOSITION 3

First, we show that (60) is equivalent to (58). Using (59), we have

$$\dot{\boldsymbol{A}} = \dot{\boldsymbol{V}}\boldsymbol{G}\boldsymbol{V}^T + \boldsymbol{V}\dot{\boldsymbol{G}}\boldsymbol{V}^T + \boldsymbol{V}\boldsymbol{G}\dot{\boldsymbol{V}}^T. \tag{F.1}$$

Hence

$$\dot{\boldsymbol{A}}\boldsymbol{P}_{\boldsymbol{A}}^{\perp}\dot{\boldsymbol{A}} = \boldsymbol{V}\boldsymbol{G}\dot{\boldsymbol{V}}^T \boldsymbol{P}_{\boldsymbol{A}}^{\perp} \dot{\boldsymbol{V}}\boldsymbol{G}\boldsymbol{V}^T \tag{F.2}$$

$$\boldsymbol{G}^{-1}\boldsymbol{V}^T \dot{\boldsymbol{A}}\boldsymbol{P}_{\boldsymbol{A}}^{\perp}\dot{\boldsymbol{A}}\boldsymbol{V}\boldsymbol{G}^{-1} = \dot{\boldsymbol{V}}^T \boldsymbol{P}_{\boldsymbol{A}}^{\perp} \dot{\boldsymbol{V}} = \dot{\boldsymbol{V}}^T \dot{\boldsymbol{V}} - \dot{\boldsymbol{V}}^T \boldsymbol{V}\boldsymbol{V}^T \dot{\boldsymbol{V}} \tag{F.3}$$

since $\boldsymbol{V}^T \boldsymbol{V} = \boldsymbol{I}$ and $\boldsymbol{P}_{\boldsymbol{A}}^{\perp} = \boldsymbol{I} - \boldsymbol{V}\boldsymbol{V}^T$. Therefore, we have

$$\boldsymbol{e}_1^T \boldsymbol{G}^{-1}\boldsymbol{V}^T \dot{\boldsymbol{A}}\boldsymbol{P}_{\boldsymbol{A}}^{\perp}\dot{\boldsymbol{A}}\boldsymbol{V}\boldsymbol{G}^{-1}\boldsymbol{e}_1 = \dot{\boldsymbol{v}}_1^T \dot{\boldsymbol{v}}_1 - \dot{\boldsymbol{v}}_1^T \boldsymbol{V}\boldsymbol{V}^T \dot{\boldsymbol{v}}_1. \tag{F.4}$$

Furthermore, from the orthonormality $\boldsymbol{v}_1^T \boldsymbol{v}_1 = 1$, we obtain $\dot{\boldsymbol{v}}_1^T \boldsymbol{v}_1 = 0$, $\dot{\boldsymbol{v}}_1^T \boldsymbol{V} = [0 \quad \dot{\boldsymbol{v}}_1^T \boldsymbol{v}_2]$, and

$$\boldsymbol{e}_1^T \boldsymbol{G}^{-1}\boldsymbol{V}^T \dot{\boldsymbol{A}}\boldsymbol{P}_{\boldsymbol{A}}^{\perp}\dot{\boldsymbol{A}}\boldsymbol{V}\boldsymbol{G}^{-1}\boldsymbol{e}_1 = \dot{\boldsymbol{v}}_1^T \dot{\boldsymbol{v}}_1 - (\dot{\boldsymbol{v}}_1^T \boldsymbol{v}_2)^2. \tag{F.5}$$

Similarly, we have

$$\boldsymbol{e}_2^T \boldsymbol{G}^{-1}\boldsymbol{V}^T \dot{\boldsymbol{A}}\boldsymbol{P}_{\boldsymbol{A}}^{\perp}\dot{\boldsymbol{A}}\boldsymbol{V}\boldsymbol{G}^{-1}\boldsymbol{e}_2 = \dot{\boldsymbol{v}}_2^T \dot{\boldsymbol{v}}_2 - (\dot{\boldsymbol{v}}_2^T \boldsymbol{v}_1)^2. \tag{F.6}$$

Since $(\dot{\boldsymbol{v}}_1^T \boldsymbol{v}_2)^2 = (\dot{\boldsymbol{v}}_2^T \boldsymbol{v}_1)^2$, after substituting (F.5) and (F.6) into (58), we obtain (60).

Second, we show that (60) and (61) coincide. This can be shown using

$$\begin{aligned} &\boldsymbol{e}_i^T \boldsymbol{G}^{-1}\boldsymbol{V}^T \dot{\boldsymbol{A}}\boldsymbol{P}_{\boldsymbol{A}}^{\perp}\dot{\boldsymbol{A}}\boldsymbol{V}\boldsymbol{G}^{-1}\boldsymbol{e}_i \\ &\quad = \operatorname{trace}\left\{\boldsymbol{P}_{\boldsymbol{A}}^{\perp}\dot{\boldsymbol{A}}\boldsymbol{V}\boldsymbol{G}^{-1}\boldsymbol{e}_i \boldsymbol{e}_i^T \boldsymbol{G}^{-1}\boldsymbol{V}^T \dot{\boldsymbol{A}}\boldsymbol{P}_{\boldsymbol{A}}^{\perp}\right\}, \quad i = 1, 2. \end{aligned} \tag{F.7}$$

Therefore,

$$\sum_{i=1}^{2}\left\{\boldsymbol{e}_i^T\boldsymbol{G}^{-1}\boldsymbol{V}^T\dot{\boldsymbol{A}}\boldsymbol{P}_{\boldsymbol{A}}^{\perp}\dot{\boldsymbol{A}}\boldsymbol{V}\boldsymbol{G}^{-1}\boldsymbol{e}_i\right\} = \mathrm{trace}\left\{\boldsymbol{P}_{\boldsymbol{A}}^{\perp}\dot{\boldsymbol{A}}\boldsymbol{V}\boldsymbol{G}^{-1}\left(\sum_{i=1}^{2}\boldsymbol{e}_i\boldsymbol{e}_i^T\right)\boldsymbol{G}^{-1}\boldsymbol{V}^T\dot{\boldsymbol{A}}\boldsymbol{P}_{\boldsymbol{A}}^{\perp}\right\}. \quad \text{(F.8)}$$

Using

$$\left(\sum_{i=1}^{2}\boldsymbol{e}_i\boldsymbol{e}_i^T\right) = \boldsymbol{I} = \boldsymbol{V}^T\boldsymbol{V}$$

and the pseudoinverse $\boldsymbol{A}^{\dagger} = \boldsymbol{V}\boldsymbol{G}^{-1}\boldsymbol{V}^T$, we have

$$\sum_{i=1}^{2}\left\{\boldsymbol{e}_i^T\boldsymbol{G}^{-1}\boldsymbol{V}^T\dot{\boldsymbol{A}}\boldsymbol{P}_{\boldsymbol{A}}^{\perp}\dot{\boldsymbol{A}}\boldsymbol{V}\boldsymbol{G}^{-1}\boldsymbol{e}_i\right\} = \mathrm{trace}\left\{\boldsymbol{P}_{\boldsymbol{A}}^{\perp}\dot{\boldsymbol{A}}\boldsymbol{A}^{\dagger}\boldsymbol{A}^{\dagger}\dot{\boldsymbol{A}}\boldsymbol{P}_{\boldsymbol{A}}^{\perp}\right\}. \quad \text{(F.9)}$$

This concludes the proof.

REFERENCES

[1] C. T. Zahn and R. Z. Roskies, "Fourier descriptors for plane closed curves," *IEEE Trans. Comput.*, vol. C-21, pp. 269–281, Mar. 1972.
[2] E. Persoon and K. S. Fu, "Shape discrimination using Fourier descriptors," *IEEE Trans. Syst.. Man, Cybern.*, vol. SMC-7, pp. 170–179, Mar. 1977.
[3] C. K. Chui, *Multivariate Splines*. Philadelphia, PA: SIAM, 1988.
[4] G. C. H. Chuang and C. C. J. Kuo, "Wavelet descriptor of planar curves: Theory and applications," *IEEE Trans. Image Processing*, vol. 5, pp. 56–70, Jan. 1996.
[5] P. Wunsch and A. F. Laine, "Wavelet descriptors for multiresolution recognition of handprinted characters," *Pattern Recogn.*, no. 8, pp. 1237–1249, 1995.
[6] D. Colton and A. Kirsch, *Inverse Acoustic and Electromagnetic Scattering*. Berlin, Germany: Springer-Verlag, 1992.
[7] A. Kirsch, "The domain derivative and two applications in inverse scattering," *Inverse Probl.*, no. 9, pp. 81–96, 1993.
[8] A. O. Hero, R. Piramuthu, J. A. Fessler, and S. R. Titus, "Minimax estimation computed tomography using high resolution anatomical side information and B-spline models," *IEEE Trans. Inform. Theory*, vol. 45, pp. 920–938, Apr. 1999.
[9] H. V. Trees, *Detection, Estimation and Modulation Theory*. New York: Wiley, 1968.
[10] H. Cramér and M. R. Leadbetter, *Stationary and Related Stochastic Processes*. New York: Wiley, 1967.
[11] M. R. Leadbetter, G. Lindgren, and H. Rootzén, *Extremes and Related Properties of Random Sequences and Processes*. New York: Springer-Verlag, 1983.
[12] S. O. Rice, "Mathematical analysis of random noise," *Bell Syst. Tech. J.*, pp. 282–332, July 1944.
[13] ——, "Mathematical analysis of random noise—Part II," *Bell Syst. Tech. J.*, pp. 46–156, Jan. 1945.
[14] ——, "Statistical properties of a sine wave plus random noise," *Bell Syst. Tech. J.*, pp. 107–157, Jan. 1948.
[15] ——, "Distribution of the duration of fades in radio transmission: Gaussian noise model," *Bell Syst. Tech. J.*, pp. 581–635, May 1958.
[16] M. Shinozuka, "On the two sided barrier problem," *J. Appl. Probab.*, pp. 79–879, 1965.
[17] D. Slepian, "First passage time for a particular Gaussian process," *Ann. Math. Statist.*, pp. 610–6129, 1961.
[18] ——, "The one-sided barrier problem for Gaussian noise," *Bell Syst. Tech. J.*, pp. 1383–1387, Nov. 1962.
[19] A. J. Rainal, "First and second passage times of Rayleigh processes," *IEEE Trans. Inform. Theory*, vol. IT-33, pp. 419–425, May 1987.
[20] A. H. Gray, "On Gaussian noise envelopes," *IEEE Trans. Inform. Theory*, vol. IT-16, pp. 522–5289, Sept. 1970.
[21] H. Cramér, "On the maximum of a normal stationary stochastic process," *Bull. Amer Math. Soc.*, pp. 513–516, 1962.
[22] G. F. Newell and M. Rosenblatt, "Zero crossing probabilities for Gaussian stationary processes," *Ann. Math. Statist.*, pp. 1306–1313, 1962.
[23] D. A. Darling and A. J. F. Siegert, "The first passage problem for a continuous Markov process," *Ann. Math. Statist.*, pp. 624–639, July–Aug. 1953.
[24] M. G. Shur, "On the maximum of a Gaussian stationary process," *Theory Prob. and Appl.*, pp. 1306–1313, 1965.
[25] M. Kendall and A. Stuart, *The Advanced Theory of Statistics*, 4th ed. New York: MacMillan, 1977.
[26] H. V. Poor, *An Introduction of Signal Detection and Estimation*, 2nd ed. New York: Springer-Verlag, 1994.
[27] I. A. Ibragimov and R. Z. Has'minskii, *Statistical Estimation: Asymptotic Theory*. New York: Springer-Verlag, 1979.
[28] S. S. Wilks and J. F. Daly, "An optimum property of confidence regions associated with the likelihood function," *Ann. Math. Statist.*, no. 10, pp. 225–235, 1939.
[29] H. Ruben, "Probability content of regions under spherical normal distributions I," *Ann. Math. Statist.*, pp. 598–618, 1960.
[30] T. W. Anderson, *An Introduction to Multivariate Anaysis*. New York: Wiley, 1958.
[31] M. Aronowich and R. J. Adler, "Extrema and level crossings of χ^2 processes," *Adv. Appl. Probab.*, no. 18, pp. 901–920, 1986.
[32] V. I. Piterbarg, "High excursions for nonstationary generalized chi-square processes," *Stoch. Processes their Applic.*, no. 53, pp. 307–337, 1994.
[33] G. H. Golub and C. F. V. Loan, *Matrix Computations*, 2 ed. Baltimore, MD: Johns Hopkins Univ. Press, 1989.
[34] J. C. Ye, Y. Bresler, and R. Moulin, "Cramér–Rao bounds for parametric estimation of target boundaries in nonlinear inverse scattering problems," *IEEE Trans. Antennas Propagat.*, submitted for publication.
[35] ——, "Cramér–Rao bounds for parametric estimation of target boundaries in nonlinear inverse scattering problems," Coordinated Sci. Lab. (CSL), Dep. Elec. Comput. Eng., Univ. Illinois at Urbana-Champaign, Tech. Rep. UILU-ENG-00-2202, Jan. 2000.
[36] P. Maragos, "Differential morphology and image processing," *IEEE Trans. Image Processing*, vol. 5, pp. 922–937, June 1996.
[37] D. H. Johnson and D. E. Dudgeon, *Array Signal Processing*. Englewood Cliffs, NJ: Prentice Hall, 1993.
[38] R. G. Hakvoort and P. M. J. V. den Hof, "Identification of probabilistic system uncertainty regions by explicit evaluation of bias and variance errors," *IEEE Trans. Automat. Contr.*, vol. 42, pp. 1516–1528, Nov. 1997.
[39] J. R. Munkres, *Topology: A First Course*. Englewood Cliffs, NJ: Prentice-Hall, 1975.
[40] M. Lindenbaum, "Bounds on shape recognition performance," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 17, pp. 666–680, July 1995.
[41] R. Piramuthu and A. O. Hero, "Theory and implementation of minmax ECT image reconstruction with MRI side information," Comm. Signal Processing Lab. (CSPL), Dep. Elec. Eng. Comput. Sci., Univ. Michigan, Ann Arbor, MI, Tech. Rep. 317, Aug. 1998.
[42] M. Viberg and B. Ottersten, "Sensor array processing based on subspace fitting," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 39, pp. 1110–1121, May 1991.