Ambiguity in Radar and Sonar

M. João D. Rendas, *Member, IEEE*, and José M. F. Moura, *Fellow, IEEE*

***Abstract*— We introduce a new ambiguity function for general parameter estimation problems in curved exponential families. We focus the presentation on passive and active radar and sonar location mechanisms. The new definition is based on the Kullback directed divergence and reflects intrinsic properties of the model. It is independent of any specific algorithms used in the processing of the signals. For the active single target problem, we show that our definition is equivalent to Woodward's radar narrowband ambiguity function. However, the new ambiguity is much broader, handling radar/sonar problems when there are unknown parameters (e.g., unknown power level in active systems), when the signals are random (e.g., passive systems), when the signals are wideband, or when there are model mismatches. We illustrate the new ambiguity in localization problems in multipath channels.**

***Index Terms*— Ambiguity function, exponential family, Kullback divergence, radar, Sanof, sonar, Woodward.**

I. Introduction

In PARAMETER estimation, it is important to determine how the observed behavior is sensitive to the parameters of the system under study. If *small* parameter changes affect *significantly* the behavior of the system, then we can determine accurately the actual values of the parameters from measurements of the system behavior. On the contrary, if the system behavior is *insensitive* to *large* variations of the parameters, then we cannot expect to be able to estimate accurately the parameters. These are structural issues, independent of specific algorithms. They address the fundamental question of what can and cannot be done.

In principle, such basic issues are resolved by computing the conditional probability density (pdf) of the estimates of the parameters given the parameters actual values. If this function is multimodal, with several important peaks, then the argument values corresponding to any of these peaks correspond to highly probable estimates. In other words, and regardless of the processing algorithm used, we cannot expect to determine from the data the true values of the parameters. This pdf is usually expressed in terms of iterated integrals of densities that depend on specific values that are hypothesized for the parameters. Lacking closed-form expressions, this pdf is computed through expensive numerical procedures, which limits its usefulness as a means of gaining insight into the system's design procedure.

Manuscript received May 15, 1991; revised July 28, 1997. The work of J. M. F. Moura was supported in part by DARPA through AFOSR Grant F49620-96-0436 and by ONR through Grant N00014-97-1-0040. The associate editor coordinating the review of this paper and approving it for publication was Prof. Don H. Johnson.

M. J. D. Rendas is with the Université de Nice Sophia Antipolis, CNRS Laboratoire, 06560 Valbonne, France.

J. M. F. Moura is with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213 USA.

Publisher Item Identifier S 1053-587X(98)01354-3.

Rather than computing this pdf, it is common to resort to bounds on the mean square parameter error, e.g., Cramér–Rao bounds (CRB's) or, as an alternative, to a sensitivity index like the ambiguity function. These tools complement each other. CRB's are usually optimistic, being local bounds, while the ambiguity function is used to assess the global resolution and large error properties. The ambiguity function establishes global conditions under which the local bounds are accurate predictions of the expected error performance and identifies the regions of the parameter space where large errors may occur.

Ambiguity functions were first introduced by Woodward [13] in active radar. Woodward's definition requires a deterministic context with complete specification of the transmitted signal, assumes known all intervening parameters (e.g., signal power), and applies to the narrowband problem. Extensions to stochastic narrowband signals (fading channel) exist; see [12]. For example, [8] uses these extensions to study the performance of passive narrowband systems in localization of moving targets. Other extensions include [11].

The available definitions of ambiguity function are restrictive. They require complete knowledge of the statistical characterization of the signal by the receiving mechanism, or they assume additional constraints that may not be verified in practice. They do not apply when the signals are wideband, stochastic with unknown signal specifications, or to multipath problems where the received power depends on the parameters being estimated.

The paper introduces a new definition of ambiguity function; see also [9]. The definition is based on a geometric interpretation of maximum likelihood (ML) parameter estimation with exponential families and on concepts of information theory. Our definition applies to wideband signals, stochastic processes, and multipath problems. It has been used in [10] to determine the test points in the computation of the Barankin bounds in DOA estimation problems. When applied to the classical context, it recovers Woodward's definition. In this sense, we consider ours to be an extension of the classical narrowband ambiguity function.

We focus the paper on radar/sonar localization. This should not distract the reader from the much broader applicability of our definition of ambiguity to general problems of parameter estimation in curved exponential families.

The paper consists of the following. In Section II, we formulate the problem and introduce notation. A number of facts from statistical theory that relate ML in exponential families to the Kullback divergence are reviewed in the Appendix. In Section III, we present our definition of ambiguity in the general context of estimation of parameters embedded in waveforms. We consider three cases where there are

1) no unwanted parameters, as in active systems;
2) unknown unwanted parameters like in passive systems;
3) model mismatches.

In Section IV, we recover Woodward's definition as a special case of our new ambiguity function. When there is uncertainty about the signal parameters, our ambiguity function differs from Woodward's definition. We consider also the effect of wrong modeling assumptions (modeling mismatches). We illustrate the applicability of our new definition by applying it to source localization. Section V concludes the paper.

II. Problem Formulation: Model

We introduce our concept of ambiguity in the context of localization systems. The transmitted signal f travels through a channel with transfer function $\mathcal{H}$ and is corrupted by an additive noise w. More formally, the observations are modeled by

$$r(t) = \mathcal{H}_\theta[f(t{:}\, \gamma)] + w(t), \quad t \in T. \tag{1}$$

For example, in a multipath channel, $\mathcal{H}_\theta[f(t{:}\ \gamma)] = \sum a_k f(t - \tau_k)$, where a_k and τ_k are the attenuations and delays. Our setup is the following.

- Observation interval T: The observation interval is discrete. This is not a fundamental limitation; it simplifies the presentation, avoiding unnecessary technicalities.
- Observed signal r: $r(t)$ is a K-dimensional vector, the outputs of an array of K omnidirectional sensors. We denote by $r_T = \{r(t), t \in T\}$ the set of available observations.
- Source signal f: $f(t, \gamma)$ is either a deterministic signal of known structure (active radar or sonar) or a sample function of a random process (passive context). In either case, the source signal is parametrized by γ. This vector may include the power level, the bandwidth, or even signal samples. Some of these parameters may be unknown. When S multiple sources are being located, f is a vector of dimension S.
- Channel operator $\mathcal{H}$: $\mathcal{H}_\theta[\cdot]$ represents the action of the channel, e.g., multipath or the effects of boundaries, or also the action of the receiving array, e.g., its spatial diversity. These combined actions depend on the vector of location parameters θ. This vector parametrizes the operator $\mathcal{H}_\theta[\cdot]$. For the multisource problem of S sources detected by K sensors, $\mathcal{H}_\theta[\cdot]$ is a $K \times S$ matrix operator.
- Noise w: $w(t)$ is the observation noise. It includes interferences, other directional point or distributed noise sources, modeling errors not accounted for, and other background disturbances. Here, we treat interferences and directional sources as signals and reduce $w(t)$ to background Gaussian white noise.

A. Estimation Problem

The goal is to estimate from the observations r_T the vector of location parameters θ. When γ is unknown, these parameters are termed nuisance parameters. It is well known that ML estimation of θ also requires the ML estimation of γ.

The vector of unknown parameters is in general represented by α. When the source parameters γ are not known, $\alpha^T = [\theta^T \ \ \gamma^T]$. If γ is known, $\alpha = \theta$. The dimension of α is represented by q. The parameter α takes values in the parameter space $\mathcal{A}$, where $\mathcal{A} \subset \mathbf{R}^q$. The location parameter vector θ takes values in Θ and the vector of source parameters γ takes values in Γ such that $\mathcal{A} = \Theta \times \Gamma$.

The estimation problem is completely defined by the family of pdf's[1]

$$\mathcal{G}_\alpha = \{p(r_T|\alpha), \, \alpha \in \mathcal{A}\}$$

where $p(r_T|\alpha)$ is the pdf of the observed data[2] indexed by the vector α. For the sake of illustration, we can take $p(r_T|\alpha)$ to be a multivariate Gaussian, where the mean and/or the covariance are parametric functions of α. We also refer to $\mathcal{G}_\alpha$ as the probabilistic model, the model manifold, or the probabilistic manifold.

The family $\mathcal{G}_\alpha$ summarizes both the structural and the statistical prior knowledge because that is all that is needed to derive the optimal ML estimator of the unknown parameter α. The geometry associated with this parametric family $\mathcal{G}_\alpha$ of densities determines the fundamental ambiguity structure of the estimation procedure and is basic to our definition of ambiguity. We overview briefly our approach.

The family $\mathcal{G}_\alpha$ is embedded in a more general family $\mathcal{G}$. For example, when $\mathcal{G}_\alpha$ is the multivariate Gaussian with moments parametrized by $\alpha \in \mathcal{A}$, $\mathcal{G}$ is the family of multivariate Gaussian densities of the same order. We restrict $\mathcal{G}$ to be an exponential family; see the Appendix. The family $\mathcal{G}_\alpha$, as a subset of $\mathcal{G}$, is a curved exponential family. This setup allows for the use of well-known statistical tools for studying its geometric properties; see the Appendix.

When $\alpha^T = [\theta^T \ \ \gamma^T]$, two submanifolds of the model manifold $\mathcal{G}_\alpha$ are relevant in our study.

$$\mathcal{G}_\gamma^\theta = \{p(r_T|\theta, \gamma), \gamma \in \Gamma\} \ \text{ and } \ \mathcal{G}_\theta^\gamma = \{p(r_T|\theta, \gamma), \, \theta \in \Theta\}. \tag{2}$$

The first submanifold $\mathcal{G}_\gamma^\theta$ corresponds to the pdf's where θ is fixed. It describes data from sources at a particular location θ. The second submanifold $\mathcal{G}_\theta^\gamma$ has γ fixed. It is the family of densities that describes data from sources with a fixed parameter γ. It parametrizes the model by the source location parameters $\theta \in \Theta$. For example, with active location systems, the source signal is usually taken to be completely known so that γ is fixed and known. $\mathcal{G}_\theta^\gamma$ is the relevant probabilistic manifold $\mathcal{G}_\alpha$ that describes the observed data. When γ is not known, as, for example, when fading is present or in passive problems, $\mathcal{G}_\alpha = \bigcup_{\gamma \in \Gamma} \mathcal{G}_\theta^\gamma$. This family is clearly larger than any of its components $\mathcal{G}_\theta^\gamma$. Geometrically, uncertainty in the source signal results in an increased dimensionality of the probabilistic model.

[1] The application Φ: $\mathcal{A} \to \mathcal{G}_\alpha$ satisfies regularity and invertibility conditions such that Φ constitutes a global coordinate system for the manifold $\mathcal{G}_\alpha$.

[2] Often, we will drop the explicit dependence on T, i.e., we will refer to $p(r_T|\alpha)$ simply as $p(r|\alpha)$.

III. Ambiguity

In the Appendix, which should be read now, we introduce a number of facts from statistical geometry that relate ML in exponential families to the Kullback directed divergence. In this section, we use Facts 5 and 6 in the Appendix, together with a limiting argument, to present our definition of ambiguity. We consider first the simple case of signals with no unwanted parameters. Then, we study the general case of signals dependent on unknown unwanted parameters. Finally, we show how to modify the definition of ambiguity to address the case when the receiver uses a wrong model, i.e., the values used by the receiver for some parameters although thought to be correct are actually in error (model mismatches).

A. Ambiguity: No Nuisance Parameters

We introduce ambiguity in the simplest context of active systems with completely known signal. In this case, the only unknowns are the location parameters $\alpha = \theta$. The manifold of interest is $\mathcal{G}_\alpha = \mathcal{G}_\theta = \{p(r_T|\theta) : \theta \in \Theta\}$. Our definition of ambiguity depends solely on the geometric properties of the model manifold $\mathcal{G}_\theta$ and provides an index on the ability to discriminate between different values of θ in the model $\mathcal{G}_\theta$.

In assessing the difficulty of estimating the actual value θ_0 of an unknown parameter θ, we can interpret the estimation problem as the binary decision test

$$H_0 : r_T \to p_{\theta_0} = p(r_T|\theta_0) \quad \text{versus} \quad H_1 : r_T \to p_\theta = p(r_T|\theta)$$

i.e., as the problem of distinguishing between the two points $p(r_T|\theta_0)$ and $p(r_T|\theta)$ in the family $\mathcal{G}_\theta$. From the discussion in the Appendix (Fact 5), we know that the functional that we should use to compare different points of the model is the directed divergence.

Denote by r_T^* the observed sample. Let it correspond to the occurrence of θ_0, i.e., the actual value of the desired parameter is θ_0. We make the ideal sampling assumption given by (39) in the Appendix that r_T^* perfectly determines p_{θ_0}. This assumption is essentially an asymptotic type hypothesis, implying that our ambiguity function captures the large sample size limit of the problem.

With this assumption, the first term on the right-hand side of (41) following Fact 6 in the Appendix is zero, and the log-likelihood ratio for deciding between θ_0 and θ is

$$\Lambda(\theta_0 : \theta) = -I(\theta_0 : \theta).$$

$I(\theta_0 : \theta)$ is shorthand for the directed divergence $I(p_{\theta_0} : p_\theta)$ between the pdf's p_{θ_0} and p_θ.

Based on the above considerations, we propose the following definition of ambiguity.

Definition 1—Ambiguity: Consider the problem of estimation of the parameter θ from observations described by the model $\mathcal{G}_\theta$. We define *ambiguity* in the estimation of θ, conditioned on the occurrence of θ_0, by

$$\mathcal{A}(\theta_0, \theta) \triangleq 1 - \frac{I(\theta_0 : \theta)}{I_{\text{ub}}(\theta_0)} \tag{3}$$

where $I_{\text{ub}}(\theta_0)$ is an upper bound of $I(\theta_0 : \theta)$

$$I(\theta_0 : \theta) \le I_{\text{ub}}(\theta_0), \qquad \forall\, \theta \in \Theta. \tag{4}$$

The ambiguity function $\mathcal{A}(\theta_0, \theta)$ is equal to its maximal value of one at the true parameter value θ_0. It takes values close to one at the points that are "difficult to distinguish" from θ_0. The upper bound $I_{\text{ub}}(\theta_0) = \max_\theta I(\theta_0 : \theta)$ corresponds to the distribution in the parametric class furthest from the nominal true distribution. From the point of view of the user, $\mathcal{A}(\theta_0, \theta)$ summarizes the impact of the geometry of the statistical model on the ability to differentiate between different parameter values.

Our ambiguity function provides, as discussed before, a prediction, for large sample sizes, of the normalized shape of the score function in parameter estimation problems. We can also relate it to the probability of error in the binary test introduced above $H0 : p_{\theta_0}$ versus $H1 : p_\theta$. We use Sanov's theorem and Stein's lemma, [2], [3], [7] applied to optimal Neyman–Pearson decision tests. Let α_n and β_n be the probabilities of false alarm and missed detection for sample size n

$$\alpha_n = \Pr(H1|\theta_0), \qquad \beta_n = \Pr(H0|\theta).$$

Fix $\beta_n = \beta$, $0 < \beta < 1$. Then, by Stein's lemma,[3] [2], [7]

$$\lim_{n\to\infty} \frac{1}{n} \log \alpha_n = -I(p_{\theta_0} : p_\theta)$$

i.e., the rate of false alarms (errors in estimation of θ_0) tends to zero exponentially fast, with exponent $I(p_{\theta_0} : p_\theta)$. The larger $I(p_{\theta_0} : p_\theta)$ is, the smaller the amount of data required to achieve the same level of false alarms.

Remark: For Definition 1 to be well defined, we need to guarantee that $I_{\text{ub}}(\theta_0)$ is neither zero nor ∞. We discuss these two points separately.

1) It may happen that $I_{\text{ub}}(\theta_0) = 0$. Since the directed divergence is zero only if the two distributions being compared are equal [see (32)], we may conclude that except over sets of measure zero

$$I_{\text{ub}}(\theta_0) = 0 \implies \forall\, \theta \in \Theta: \quad p(r_T|\theta_0) = p(r_T|\theta)$$

i.e., all different points of the parameter space Θ define the same measure in the observation space. This means that the family of distributions is a manifold of dimension 0 or, in more intuitive terms, that the model is not sensitive to parameter variations. In this case, we define

$$\forall\, \theta \in \Theta: \quad \mathcal{A}(\theta_0, \theta) = 1$$

meaning that the model is not informative with respect to the parameter. When this happens, we conclude that the parameter θ is not observable from the data.

2) We state conditions for $I_{\text{ub}}(\theta_0) < \infty$.

A necessary condition is clearly to require the absolute continuity of the members of the family of distributions, i.e., that whenever either $p(r_T|\theta_0)$ or $p(r_T|\theta)$ is zero, the other is also zero. This is however not sufficient, i.e.,

[3] According to [7], Chernoff (1956) derived this lemma and related results. Chernoff attributes the results to unpublished work of Stein.

the absolute continuity assumption does not guarantee the finiteness of the Kullback divergence.

From the definition of Kullback divergence [see (31) in the Appendix] and for members in the curved exponential family [see (37)], we get

$$\begin{aligned} I(\theta_0 : \theta) &= \mathrm{E}_{\theta_0}\left\{\log \frac{p(r_T|\theta_0)}{p(r_T|\theta)}\right\} \\ &= \mathrm{E}_{\theta_0}\left\{[a(\theta_0) - a(\theta)]^T b(r_T) - [c(\theta_0) - c(\theta)]\right\} \quad (5) \\ &= [a(\theta_0) - a(\theta)]^T \beta(\theta_0) - [c(\theta_0) - c(\theta)] \quad (6) \end{aligned}$$

where $\beta(\theta_0) = \mathrm{E}_{\theta_0}\{b(r_T)\}$ is referred to as the *expectation* parameter [4]. We obtain (5) by simply using the definition of curved exponential families [see (37)] and noting that the term $d(r_T)$ is canceled in the Kullback divergence.

Since the Kullback divergence is nonnegative, it is finite and upperbounded if

$$[a(\theta_0) - a(\theta)]^T \beta(\theta_0) - [c(\theta_0) - c(\theta)] < M \quad (7)$$

where M is positive. A set of sufficient conditions for (7) to hold is

$$\forall\, \theta \in \Theta: \; |a(\theta)| < \infty, \; |\beta(\theta)| < \infty \; \text{ and } \; |c(\theta)| < \infty. \quad (8)$$

Condition (8) rules out pathological cases such as singularity of covariance matrices. In general, under reasonable assumptions, e.g., compactness of the parameter space Θ, these conditions are satisfied. We hypothesize these conditions in the sequel.

We make two additional comments:

1) In contrast with the classical ambiguity function of Woodward [13], $\mathcal{A}(\theta_0, \theta)$ is not a symmetric function of its arguments. The symmetry of the classical ambiguity is a consequence of the particular geometry of the model for which it was developed. Symmetry, however, is not a natural requirement for an ambiguity function since in a binary test, the probability of error depends generally on which hypothesis is true.
2) The definition depends, somewhat arbitrarily, on the upper bound $I_{\mathrm{ub}}(\theta_0)$. This was introduced to enforce a positive normalized measure of ambiguity in between zero and one, which is a natural requirement. The bound should be calculated as tight as possible so that the ambiguity values span the entire range from 0 to 1. If a loose bound is used, this range is squeezed.

B. Ambiguity: Unwanted Parameters

We consider here the practically interesting situation where, besides the parameters θ of interest, the data is dependent on a distinct set of nuisance parameters γ, which are unknown to the receiver. This occurs, for example, in passive location problems, where θ describes the target location and γ the source signal.

Again, consider the problem of deciding between two distinct values θ and θ_0 to which subfamilies $\mathcal{G}_\gamma^\theta = \{p(r_T|\theta, \gamma), \gamma \in \Gamma\}$ and $\mathcal{G}_\gamma^{\theta_0} = \{p(r_T|\theta_0, \gamma), \gamma \in \Gamma\}$ in $\mathcal{G}_\theta$ correspond

$$H_0 : p(r_T) \in \mathcal{G}_\gamma^{\theta_0} \quad \text{versus} \quad H_1 : p(r_T) \in \mathcal{G}_\gamma^{\theta}.$$

Consider that the observed sample r_T^* corresponds to $\alpha_0 = [\theta_0 \gamma_0]$, i.e., that the actual true value of the desired parameter is θ_0, and the value of the nuisance parameter is γ_0. Under ideal sampling, i.e., when the true pdf $p_0 = p(r_T|\theta_0, \gamma_0)$ is identified from the data, the generalized likelihood ratio for deciding between θ_0 and θ is

$$\Lambda(\theta_0 : \theta) = -\min_{p_1 \in \mathcal{G}_1} I(p_0 : p_1) = -\tilde{I}(\theta_0 : \theta)_{\gamma_0}$$

where $\Lambda(\theta_0 : \theta)$ is shorthand for $\Lambda(H_0 : H_1)$, and we have defined $\tilde{I}(\theta_0 : \theta)_{\gamma_0}$ as the minimum value of the Kullback divergence

$$\forall\, f \in \mathcal{G}_\gamma^\theta: \tilde{I}(\theta_0 : \theta)_{\gamma_0} \triangleq I\{p_0 : p[r_T|\theta, \tilde{\gamma}(\theta)_{p_0}]\} \leq I(p_0 : f) \quad (9)$$

where $\tilde{\gamma}(\theta)_{p_0}$ is the value of γ to which the element of $\mathcal{G}_\gamma^\theta$ closest to p_0 corresponds

$$\tilde{\gamma}(\theta)_{p_0} \triangleq \arg \min_{\gamma \in \Gamma} I[p_0 : p(r_T|\theta, \gamma)].$$

Based on the above considerations, we propose the following definition of ambiguity.

Definition 2—Ambiguity: Consider the problem of estimation of the parameter θ from observations described by the model $\mathcal{G}_\alpha$, where γ is an unknown nonrandom vector of parameters. We define *ambiguity in the estimation of* θ *conditioned on the occurrence of* $\alpha_0 = (\theta_0, \gamma_0)$, by

$$\mathcal{A}(\theta_0, \theta)_{\gamma_0} \triangleq 1 - \frac{\tilde{I}(\theta_0 : \theta)_{\gamma_0}}{\tilde{I}_{\mathrm{ub}}(\theta_0)_{\gamma_0}} \quad (10)$$

where $\tilde{I}_{\mathrm{ub}}(\theta_0)_{\gamma_0}$ is an upper bound of $\tilde{I}(\theta_0 : \theta)_{\gamma_0}$ [defined by (9)]

$$\forall\, \theta \in \Theta: \quad \tilde{I}(\theta_0 : \theta)_{\gamma_0} \leq \tilde{I}_{\mathrm{ub}}(\theta_0)_{\gamma_0}. \quad (11)$$

Comments similar to the ones following Definition 1 apply here regarding the upper bound.

When comparing the global performance of different models of the same physical problem, for instance, when studying the effect of lack of knowledge about source parameters, it is desirable that I_{ub} be defined in a way that allows comparison of the ambiguity values. Since

$$\tilde{I}(\theta_0 : \theta)_{\gamma_0} \leq I[p(r_T|\theta_0, \gamma_0) : p(r_T|\theta, \gamma_0)]$$

we use the upper bound for the case of known γ as an upper bound for $\tilde{I}(\cdot : \cdot)_{\gamma_0}$; then

$$\mathcal{A}(\theta_0, \theta)_{\gamma_0} \geq \mathcal{A}[(\theta_0, \gamma_0), (\theta, \gamma_0)]$$

implying that ambiguity always increases when some parameters are not known precisely.

Definition 2 is conditioned on the true parameter values: both the wanted parameter θ_0 and the nuisance parameter

γ_0. It would be desirable that a global measure of ambiguity depends only on the two values θ_0 and θ. Mathematically, this corresponds to applying an operator A_γ to the conditioned ambiguity defined by (10)

$$\mathcal{A}(\theta_0, \theta) = A_\gamma[\mathcal{A}(\theta_0, \theta)_\gamma].$$

Obvious candidates for the operator $A_\gamma[\cdot]$ are the mean value, the maximum, or the minimum taken over all possible values of γ. We will see, for a specific problem, the consequences of using one of these alternative operators. We think, however, that it is preferable to keep the conditioned ambiguity given in Definition 2. This definition gives insight into the geometry of the location problem and clarifies the impact of both the source signal and of the medium characteristics on the resolution of the localization mechanism.

Our definition captures the impact of uncertainty on the observability of the wanted parameter. This is the key difference between passive and active systems and one of the fundamental reasons why the same tools cannot be used to study both. A problem with a nice ambiguity structure when each $\mathcal{G}_\gamma^\theta$ is considered individually can be unobservable (extreme case of ambiguity) when the complete manifold $\mathcal{G}_\alpha$ is considered. It suffices that for $\theta \neq \theta_0$, $\mathcal{G}_\gamma^\theta \bigcap \mathcal{G}_\gamma^{\theta_0} \neq \emptyset$. Our tool allows the study of all intermediate situations.

C. Ambiguity: Modeling Inaccuracies

The ambiguity function defined previously can be extended to handle the impact of wrong modeling assumptions. We continue the presentation in the framework of location systems, assuming observations generally described by (1), and, for simplicity, we consider that there are no unknown signal parameters γ. We now make explicit the dependence of the observation operator on both physical medium parameters and on the source position vector θ. Rewrite the observation equation as

$$r(t) = \mathcal{H}_{\theta,\xi}[f(t)] + w(t), \qquad t \in T. \tag{12}$$

The vector ξ denotes parameters related to the propagation channel that may include sound speed profile, water column depth, or boundaries reflectivity. In practice, these parameters are only approximately known. The uncertainty regarding the propagation channel characteristics ξ will in general result in a degradation of the system performance in the form of nonnegligible biases. We address now the issue of how the ambiguity function we have introduced can be extended to predict this form of performance degradation.

Denote by ξ_0 the true value of the propagation parameters. At the receiver, a wrong value $\xi \neq \xi_0$ is used to model the observed data. We are not concerned with the receiver using inaccurate estimates of ξ but with the receiver's willful use of wrong parameter values. Our goal is to study the impact of this mismatch on the ability of correctly predicting the source location θ.

We introduce additional notation. We consider only the simpler case of active location systems. The discussion is generalized with no great additional effort to passive systems. Let $\mathcal{G}_\theta^0$ be the probabilistic manifold that describes the received data for the true value of the physical parameters ξ_0 and $\mathcal{G}_\theta^1$ the probabilistic model used at the receiver. Denote members of $\mathcal{G}_\theta^i$ by p_θ^i, $i = 0, 1$. Under the ideal sampling assumption (see D3 of the Appendix), $p_{\theta_0}^0$ is perfectly identified [see (39)], and Fact 6 states that the generalized log-likelihood ratio for deciding between source locations θ_0 and θ is

$$\Lambda(\theta_0, \theta) = I(p_{\theta_0}^0 : p_{\theta_0}^1) - I(p_{\theta_0}^0 : p_\theta^1).$$

This expression is illustrative of the performance degradation due to model mismatches. The first term is a constant bias accounting for the distance between the densities corresponding to the true value in the correct and incorrect model. It may happen that the maximum of the likelihood ratio does not occur, even under ideal sampling, at the true source location, indicating the presence of a nonzero estimation bias, since the likelihood estimator searches for the value θ that yields the maximum value of $I(p^0(r_T|\theta_0) : p^1(r_T|\theta))$.

The definition of ambiguity is modified according to the following.

Definition 3—Ambiguity: Model Mismatch: Consider the parameter estimation problem described by the curved exponential family $\mathcal{G}_\theta^0$ using the probabilistic model $\mathcal{G}_\theta^1$ at the receiver. We define *ambiguity in the estimation of* θ, *given that* θ_0 *is the true value of the parameter*, as

$$\mathcal{A}(\theta_0, \theta) \triangleq 1 - \frac{I(p_{\theta_0}^0 : p_\theta^1)}{I_{\mathrm{ub}}(\theta_0)} \tag{13}$$

where $I_{\mathrm{ub}}(\theta_0)$ is an upper bound on the Kullback directed divergence between $p_{\theta_0}^0$ and any member of $\mathcal{G}_\theta^1$

$$I(p_{\theta_0}^0 : p_\theta^1) \leq I_{\mathrm{ub}}(\theta_0), \; \theta \in \Theta.$$

IV. The Classical Ambiguity Function: Generalizations

In this section, we recover the classical definition of ambiguity when Definition 2 is applied to the particular observation model of Woodward. We show also that introduction of uncertainty in this model leads to different measures of ambiguity.

A. Woodward's RADAR Ambiguity Function

Consider the problem of simultaneous estimation of delay and Doppler shift in a narrowband signal of *known* complex envelope, transmitted through a Rayleigh channel. Let θ denote the vector of wanted parameters, i.e., the delay τ and the Doppler shift ω. In complex notation, the observation is an N-dimensional complex Gaussian vector

$$p(r|\theta) \sim \mathcal{N}[0, \gamma f(\theta) f(\theta)^H + \sigma^2 I] \tag{14}$$

where we assumed that the observation noise is white and the power parameters γ and σ^2 are known. The N-dimensional vector $f(\theta)$ is the complex envelope of the received waveform known at the receiver $\forall\, \theta \in \Theta$.

This is the model underlying the classical definition of ambiguity. To completely fall under the framework of the classical definition, we need to make an additional assumption, namely, that the received signal energy does not depend on the particular value of the parameter being estimated, i.e., $\forall\,\theta \in \Theta$, $||f(\theta)||^2 = K$, where K is a positive constant.

Woodward introduced his ambiguity function for the problem formulated above. Our proposed definition of ambiguity is Definition 1.

The Kullback-directed divergence between two N-dimensional normal densities with zero mean and covariance matrices R_{θ_0} and R_θ is (see [7] or the Appendix)

$$I(\theta_0 : \theta) = \tfrac{1}{2}\left[\operatorname{tr}[R_\theta^{-1} R_{\theta_0}] - N - \ln\left|R_\theta^{-1} R_{\theta_0}\right|\right]. \tag{15}$$

Using elementary facts from linear algebra

$$R_\theta^{-1} = \frac{1}{\sigma^2}\left[I - \frac{\gamma}{\sigma^2 + K\gamma} f(\theta) f(\theta)^H\right]$$
$$|R_\theta| = \sigma^{2N}\left(1 + \frac{K\gamma}{\sigma^2}\right) = \sigma^{2N}(1 + \text{ SNR})$$
$$\text{SNR } \triangleq \frac{\gamma||f(\theta)||^2}{\sigma^2} = \frac{K\gamma}{\sigma^2}.$$

In these equations, I denotes the identity matrix of order N, and SNR is the received signal-to-noise ratio. Using (15)

$$I(\theta_0 : \theta) = \frac{1}{2}\frac{\text{SNR}^2}{1 + \text{SNR}}\left[1 - \frac{1}{K^2}|f(\theta)^H f(\theta_0)|^2\right]. \tag{16}$$

Since the second term is always positive, we can compute the upper bound

$$I_{\text{ub}} = K^2\gamma^2/2\sigma^2(\sigma^2 + K\gamma) = \frac{\text{SNR}^2}{2[1 + \text{SNR }]}.$$

From this and (3), we get the following fact.

Fact 1: Consider the problem of estimation of θ from observations following model (14), where γ and the application $f(\theta)$ are known. Then, Definition 1 coincides with the classical ambiguity function

$$\mathcal{A}(\theta_0, \theta) = \frac{1}{K^2}|f(\theta)^H f(\theta_0)|^2. \tag{17}$$

Equation (17), which we refer to as Woodward's ambiguity function, is sometimes called the ambiguity surface, the ambiguity function being then reserved to the inner product $f(\theta)^H f(\theta_0)$.

If the set of signal envelopes $\{f(\theta), \theta \in \Theta\}$ is designed such that

$$\forall\,\theta_0\,\exists\,\theta : \; f(\theta_0) \perp f(\theta) \tag{18}$$

then ambiguity is maximally stretched between 0 and 1. Otherwise, its range is limited to a smaller interval $[\epsilon, 1]$, with $\epsilon > 0$.

B. Uncertainty in Signal Parameters

As stressed before, the model that leads to Woodward's classical definition of ambiguity admits no uncertainty regarding the signal parameters, namely, the Rayleigh coefficient γ and, for each θ, the vector $f(\theta)$.

To see that even for this simple model, dropping any of these assumptions changes the ambiguity structure, we study now the effects of i) the unknown received signal power and ii) the unknown shape of the signal correlation, i.e., unknown $f(\theta)$. In either case, associated with each value of θ, we now have a manifold of dimension greater than 0.

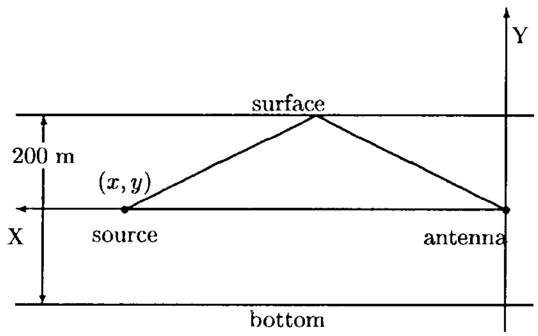


Fig. 1. Acoustic source localization in a bounded channel.

1) Unknown Received Signal Power: Consider first that γ *is not known* and that we drop the assumption of constant power. This corresponds to unknown emitted signal power or to propagation in multipath channels, where the power of the received signal depends on the specific source location that dictates the pattern of energy recombination of the individual paths.

Denote the covariance matrix of the observed field by $R_{\theta,\gamma}$. The directed divergence between $p(r|\theta_0, \gamma_0)$ and any member of $\mathcal{G}_\gamma^\theta$ is

$$I[p(r|\theta_0, \gamma_0) : p(r|\theta, \gamma)] = \tfrac{1}{2}[\operatorname{tr}[R_{\theta,\gamma}^{-1} R_{\theta_0,\gamma_0}] - N - \ln|R_{\theta,\gamma}^{-1} R_{\theta_0,\gamma_0}|]. \tag{19}$$

Minimization over γ leads to

$$\tilde{\gamma}(\theta)_{p_0} = \gamma_0 \frac{|f(\theta)^H f(\theta_0)|^2}{||f(\theta)||^4}$$

and the divergence between $p(r|\theta_0, \gamma_0)$ and $\mathcal{G}_\gamma^\theta$ is

$$\tilde{I}(\theta_0 : \theta)_{\gamma_0} = \frac{1}{2}\left\{\text{SNR}_0\left[1 - \frac{|f(\theta)^H f(\theta_0)|^2}{||f(\theta)||^2||f(\theta_0)||^2}\right] + \ln\frac{1 + \text{SNR}_0\dfrac{|f(\theta)^H f(\theta_0)|^2}{||f(\theta)||^2||f(\theta_0)||^2}}{1 + \text{SNR}_0}\right\} \tag{20}$$

$$\text{SNR}_0 = \frac{\gamma_0||f(\theta_0)||^2}{\sigma^2}. \tag{21}$$

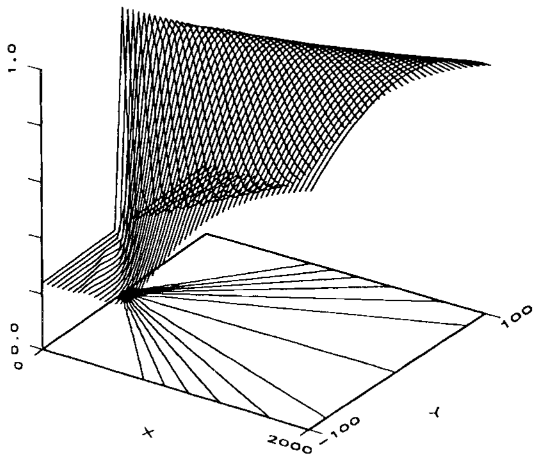
For each value of γ_0, from Schwartz's inequality, it follows that the argument of the logarithm in the previous expression is never larger than one. Thus

$$\tilde{I}(\theta_0 : \theta)_{\gamma_0} \leq \frac{1}{2}\text{SNR}_0\left[1 - \frac{|f(\theta)^H f(\theta_0)|^2}{||f(\theta)||^2||f(\theta_0)||^2}\right].$$

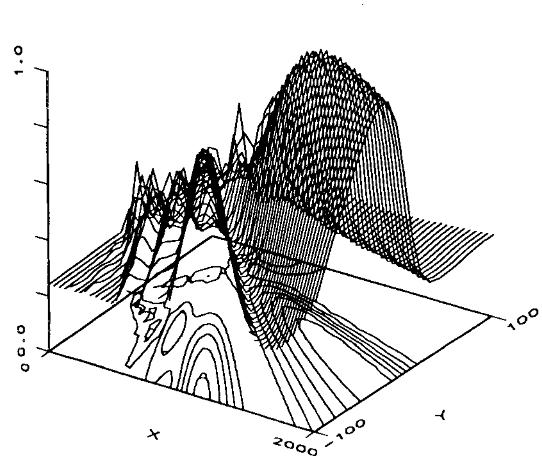
Using Schwartz's inequality, again

$$\tilde{I}(\theta_0 : \theta)_{\gamma_0} \leq \tfrac{1}{2}\text{SNR}_0 \triangleq \tilde{I}_{ub_{\gamma_0}}.$$

Using this bound and (20) in (10) of Definition 2, we obtain the following fact.



(a)



(b)

Fig. 2. Ambiguity surface for (a) single direct ray and (b) direct and surface reflected rays.

Fact 2: Consider the problem of estimation of θ from observations following model (14), where $f(\theta)$ is a known application and γ an unknown nonrandom parameter. Then, the ambiguity function conditioned on the true power level γ_0, which is denoted by $\mathcal{A}(\theta_0, \theta)_{\gamma_0}$, is related to the ambiguity function for known γ_0, $\mathcal{A}(\theta_0, \theta)$ in the following way.

$$\mathcal{A}(\theta_0, \theta)_{\gamma_0} = \mathcal{A}(\theta_0, \theta) - \frac{1}{\text{SNR}_0} \ln \frac{1 + \text{SNR}_0 \quad \mathcal{A}(\theta_0, \theta)}{1 + \text{SNR}_0}. \tag{22}$$

The argument of ln is smaller than 1; therefore, $\mathcal{A}(\theta_0, \theta)_{\gamma_0}$ is always larger than $\mathcal{A}(\theta_0, \theta)$, showing that ambiguity increases relative to the classical situation of Section IV-A.

Since $\mathcal{A}(\theta_0, \theta)_{\gamma_0}$ is a monotonic function of $\mathcal{A}(\theta_0, \theta)$, its maxima in the parameter space Θ occur in the same points as the maxima of the classical ambiguity function, which are dictated by the local maxima of the angle $|f(\theta)^H f(\theta_0)|^2$. However, the relative heights of the maxima are not the same, as we can see from the previous equation.

While Woodward's ambiguity is independent of the signal-to-noise ratio, this is no longer true in the present situation. In fact, although for very high signal-to-noise ratios the influence of the second term in (22) is negligible, so that (17) and (22) predict essentially the same behavior, in the low signal-to-noise ratio limit, the second term of (22) prevents ambiguity from being zero, even when $f(\theta_0) \perp f(\theta)$. We get a lower bound on ambiguity. From (22)

$$\mathcal{A}(\theta_0, \theta)_{\gamma_0} = \mathcal{A}(\theta_0, \theta) - \frac{1}{\text{SNR}_0} \ln[1 + \text{SNR}_0 \mathcal{A}(\theta_0, \theta)] + \frac{1}{\text{SNR}_0} \ln(1 + \text{SNR}_0). \tag{23}$$

Using the inequality $x \geq \ln(1+x)$ in this equation yields

$$\mathcal{A}(\theta_0, \theta)_{\gamma_0} \geq \frac{1}{\text{SNR}_0} \ln(1 + \text{SNR}_0). \tag{24}$$

We illustrate this discussion with an example where the classical ambiguity is not applicable, and we instead use our definition. We study the localization of a narrowband Gaussian acoustic source (100 Hz) in a shallow water channel of 200 m depth. The boundaries are perfectly flat; see Fig. 1. The medium is homogeneous, with a constant sound speed velocity of 1500 m/s. The receiving array is vertical, with ten sensors, located at the center of the water column. The acoustic source is located at an horizontal distance of 1600 m of the array, which is also equidistant from surface and bottom. The SNR is 0 dB.

We consider two distinct reflection behaviors of the boundaries: i) when both boundaries are perfectly absorbing and only the direct path reaches the receiver and ii) when there is a nonzero reflection at the surface, and thus, two distinct rays arrive at each sensor. The 3-D plots, as well as the contour plots (lines on X–Y plane) of the resulting ambiguity surfaces are shown in Fig. 2(a) and (b), respectively.

Both figures clearly exhibit a threshold due to the low SNR, as predicted by (24). The shapes of the ambiguity surfaces in Fig. 2(a) and (b) are distinctly different. The main lobe of the ambiguity in Fig. 2(b) is sharper than the shallower main lobe of Fig. 2(a). This is due to the focusing effect provided by the reflected ray. Its presence leads to an ambiguity [see Fig. 2(b)] with a marked lobe structure with an almost periodic nature, resulting from the variation of the differential delay between the two paths. The absence of this ray leads to an almost perfectly radial ambiguity structure; see Fig. 2(a).

This example illustrates the role that the ambiguity function as introduced here plays as a diagnostic tool in analyzing the degree of global observability of the estimation problem.

2) Unknown Signal Shape: A more radical change is obtained when, instead of a perfectly known application $f: \Theta \to \mathbb{C}^N$, as in the previous section, we know only that $f(\theta) \in \mathcal{M}_\theta$, where $\mathcal{M}_\theta$ is a known proper P_θ-dimensional subspace of $\mathbb{C}^N$. This is the relevant model for location of multiple perfectly correlated narrowband sources or, in a multipath channel, when only the spatial structure of the observations is modeled, i.e., only the intersensor delays, or directions of arrival, for the plane wave approximation are taken into account, any temporal structure being ignored.

Assume that the signal component of the observations is known to belong to $\mathcal{M}_\theta = \text{Sp}\{a_i(\theta)\}_{i=1}^{P_\theta}$, i.e., to the linear span of the P_θ vectors $a_i(\theta)$, such that the covariance matrix of the observations can be written as

$$R(\theta,\, s) = \sigma^2 I + A(\theta) s s^H A(\theta)^H \tag{25}$$

where $s \in \mathbb{C}^{P_\theta}$ is an unknown vector, and $A(\theta)$ is the matrix that collects the basis vectors $a_i(\theta)$. Using, in (19), the expression for the covariance matrix yields

$$\begin{aligned} I(\theta_0,\, s_0 : \theta,\, s) = \frac{1}{2}\Bigg\{ &\ln \frac{\sigma^2 + s^H A^H(\theta) A(\theta) s}{\sigma^2 + s_0^H A^H(\theta_0) A(\theta_0) s_0} \\ &+ \frac{s_0^H A^H(\theta_0) A(\theta_0) s_0}{\sigma^2} \\ &- \frac{s^H A^H(\theta) R(\theta_0, s_0) A(\theta) s}{\sigma^2[\sigma^2 + s^H A^H(\theta) A(\theta) s]}\Bigg\}. \end{aligned} \tag{26}$$

The vector s that minimizes this expression $\tilde{s}(\theta)_{p_0}$ satisfies

$$A(\theta)\tilde{s}(\theta)_{p_0} = \Pi_{\mathcal{M}_\theta}[A(\theta_0) s_0]$$

where $\Pi_{\mathcal{M}_\theta}$ denotes the orthogonal projection operator onto $\mathcal{M}_\theta$. Define

$$\gamma_{s_0}^2(\theta_0,\, \theta) \triangleq \|\Pi_{\mathcal{M}_\theta}[A(\theta_0) s_0]\|^2. \tag{27}$$

Then, the minimum of (26) is

$$\begin{aligned} \tilde{I}(\theta_0 : \theta)_{s_0} = \frac{1}{2}\Bigg[&\ln \frac{\sigma^2 + \gamma_{s_0}^2(\theta_0, \theta)}{\sigma^2 + \|A(\theta_0) s_0\|^2} \\ &+ \frac{\|A(\theta_0) s_0\|^2}{\sigma^2} - \frac{\gamma_{s_0}^2(\theta_0, \theta)}{\sigma^2}\Bigg]. \end{aligned}$$

Since $0 \leq \gamma_{s_0}^2(\theta_0,\, \theta) \leq \|A(\theta_0) s_0\|^2$, we can set

$$I_{\mathrm{ub}}(\theta_0)_{s_0} = \frac{1}{2} \frac{\|A(\theta_0) s_0\|^2}{\sigma^2}.$$

Analogous to the previous subsection, define the SNR as

$$\mathrm{SNR} \triangleq \frac{\|A(\theta_0) s_0\|^2}{\sigma^2}.$$

Then, the ambiguity function for the present scenario can be put into the form

$$\begin{aligned} \mathcal{A}(\theta_0,\, \theta)_{s_0} = &\frac{\gamma_{s_0}^2(\theta_0, \theta)}{\|A(\theta_0) s_0\|^2} - \frac{1}{\mathrm{SNR}} \\ &\cdot \ln \frac{1 + \mathrm{SNR}\gamma_{s_0}^2(\theta_0, \theta)/\|A(\theta_0) s_0\|^2}{1 + \mathrm{SNR}}. \end{aligned}$$

Denote the first term in the previous expression by $\mathcal{A}(\theta_0,\, \theta)_{s_0}^d$

$$\mathcal{A}(\theta_0,\, \theta)_{s_0}^d \triangleq \frac{\gamma_{s_0}^2(\theta_0, \theta)}{\|A(\theta_0) s_0\|^2}. \tag{28}$$

We can now state the following fact.

Fact 3: Consider the problem of estimation of θ from zero mean Gaussian observations with covariance matrix given by (25), where $A(\theta)$ is a known matrix, and s is an unknown vector. Then, the ambiguity function conditioned on the true value $f(\theta_0) = A(\theta_0) s_0$ is

$$\mathcal{A}(\theta_0,\, \theta)_{s_0} = \mathcal{A}(\theta_0,\, \theta)_{s_0}^d - \frac{1}{\mathrm{SNR}} \ln \frac{1 + \mathrm{SNR} \quad \mathcal{A}(\theta_0,\, \theta)_{s_0}^d}{1 + \mathrm{SNR}} \tag{29}$$

where $\mathcal{A}(\theta_0,\, \theta)_{s_0}^d$ is defined in (28).

Comparing expressions (29) and (22), we conclude that the role of the classical ambiguity function in (22) is here played by $\mathcal{A}(\theta_0,\, \theta)_{s_0}^d$, which is determined by the norm of the projection of $A(\theta_0) s_0$ onto $\mathcal{M}_\theta$.

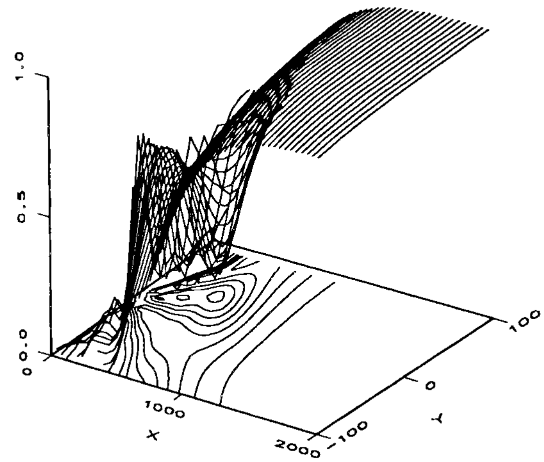


Fig. 3. Ambiguity surface for spatial model.

While in Section IV-B1 (γ unknown) the location of the peaks of the ambiguity function was independent of the unwanted parameter (see the discussion following Fact 2) now, this is not so. Different directions of the complex vector s_0 result in different shapes of the conditional ambiguity function. Comparing with the expression obtained for unknown γ [see (22)], we see that uncertainty in the direction of $f(\theta)$ results in an increased ambiguity. Even for large SNR's, when the second term in (29) becomes negligible, the value of the ambiguity is larger in the present framework since

$$\mathcal{A}(\theta_0,\, \theta)_{s_0}^d \geq \frac{\left|s_0^H A(\theta)^H A(\theta_0) s_0\right|^2}{\|A(\theta_0) s_0\|^2}.$$

Fig. 3 illustrates the ambiguity surface for the same acoustic source localization problem, but when only a spatial model is considered, i.e., when the differential delay between the direct and the reflected rays is ignored. The ambiguity function is now flat around the source position (position $x = 1600$, $y = 0$). Comparing Fig. 3 with the ambiguity structure in Fig. 2(b) of the complete model, we see that the deep valleys in Fig. 2(b) are now absent, causing the widening of the lobe structure. The ambiguity surface in Fig. 3 is much smoother, with fewer secondary lobes. The sharper main lobe in Fig. 2(b) leads to better local performance (marked peaks), as CRB studies show [9], but may be prone to ambiguity problems associated with the presence of important secondary lobes of the likelihood function. Another interesting comparison between Figs. 2(b) and 3 is the improved range observability provided by accounting in the model for multipath. This is illustrated by the fact that beyond a certain value of the horizontal distance, the surface in Fig. 3 exhibits a flat structure. This predicts the well-known problems associated with range estimation of distant sources. The detailed analysis of the shape of the two surfaces is closely linked to the features of the underwater acoustic channel and will be pursued elsewhere.

C. Upper and Lower Ambiguity Bounds

We consider now (28) as a function of the true source signal parameter vector γ_0 and verify that in this case, we are able to find upper and lower bounds to the surface $\mathcal{A}(\theta_0,\, \theta)_{\gamma_0}$

$$\mathcal{A}(\theta_0,\, \theta)_{\mathrm{MIN}} \leq \mathcal{A}(\theta_0,\, \theta)_{\gamma_0} \leq \mathcal{A}(\theta_0,\, \theta)_{\mathrm{MAX}}.$$

Assume, for simplicity, that $R_w = \sigma^2 I$ so that we are dealing with the usual Euclidean metric. The material presented below holds in the general case, with a convenient reinterpretation of the inner products.

Let $\{v_i^{\theta}\}_{i=1}^{P_\theta}$ and $\{v_i^{\theta_0}\}_{i=1}^{P_{\theta_0}}$ denote an orthonormal basis for $\mathcal{M}_\theta$ and $\mathcal{M}_{\theta_0}$, respectively, and define the $P_\theta \times P_{\theta_0}$ matrix Q^{θ,θ_0} of generic element

$$[Q^{\theta,\theta_0}]_{ij} \triangleq \langle v_i^{\theta}, v_j^{\theta_0}\rangle.$$

Let $s(\theta_0, \gamma_0)$ denote the vector of coordinates of $f_0 = A(\theta_0)s_0$ in the basis $\{v_i^{\theta_0}\}$

$$f_0 = A(\theta_0)s_0 = \sum_{i=1}^{P_{\theta_0}} s_i(\theta_0, \gamma_0)v_i^{\theta_0}.$$

Then, an equivalent expression for the ambiguity is

$$\mathcal{A}(\theta_0, \theta)_{\gamma_0} = \frac{\|Q^{\theta,\theta_0}s(\theta_0,\gamma_0)\|^2}{\|s(\theta_0,\gamma_0)\|2}.$$

The function $\mathcal{A}(\theta_0, \theta)_{\gamma_0}$ is a quadratic form in the Hermitean matrix $(Q^{\theta,\theta_0})^H Q^{\theta,\theta_0}$. As is well known, the largest and smallest eigenvalues determine upper and lower bounds on the value of a quadratic form.

Let X be a $P \times P$ matrix, and denote by $\lambda_1(X)$ and $\lambda_P(X)$ the largest and smallest eigenvalues of X, respectively. Then,

$$\lambda_{P_{\theta_0}}[(Q^{\theta,\theta_0})^H Q^{\theta,\theta_0}] \le \mathcal{A}(\theta_0, \theta)_{\gamma_0} \le \lambda_1[(Q^{\theta,\theta_0})^H Q^{\theta,\theta_0}].$$

Since the eigenvalues of $X^H X$ are the squares of the singular values of the matrix X and recognizing the singular values of the matrix Q^{θ,θ_0} as the cosines of the principal angles between the subspaces $\mathcal{M}_\theta$ and $\mathcal{M}_{\theta_0}$ [6], we can state the following fact.

Fact 4: The ambiguity (28) verifies

$$\cos^2(\sigma_{\min}) \le \mathcal{A}(\theta_0, \theta)_{\gamma_0} \le \cos^2(\sigma_{\max}) \qquad (30)$$

where $\sigma_{\min}$ and $\sigma_{\max}$ are the largest and smallest principal angles between the subspaces between $\mathcal{M}_\theta$ and $\mathcal{M}_{\theta_0}$.

An alternative lower bound can be found, noting that

$$\mathcal{A}(\theta_0, \theta)_{s_0} = 1 - \frac{\|\Pi_{(\mathcal{M}_\theta)^\perp}[A(\theta_0)s_0]\|^2}{\|A(\theta_0)s_0\|^2} \ge 1 - \cos^2(\rho_{\max})$$

where $\rho_{\max}$ is the largest principal angle between $(\mathcal{M}_\theta)^\perp$ and $\mathcal{M}_{\theta_0}$. We note that $\sin^2(\rho_{\max}) = 1 - \cos^2(\rho_{\max})$ is the gap metric between the two subspaces [5].

If $\mathcal{M}_\theta$ and $\mathcal{M}_{\theta_0}$ have an intersection of dimension greater than or equal to 1, then $\cos^2(\sigma_{\max}) = 1$. In this way, the upper bound in (30) indicates the structural ambiguities of the problem, i.e., that there may be pairs of $\theta \ne \theta_0$ for which there are γ, γ_0 such that $p(r|\theta, \gamma) = p(r|\theta_0, \gamma_0)$.

V. Conclusions

We present a novel definition for the ambiguity function that generalizes the classical definition of ambiguity function introduced by Woodward. Woodward's definition considers narrowband, completely known deterministic signals. Ours is valid under much broader conditions that include wideband signals, stochastic signals, signals with unknown (nuisance) parameters in their specification, or problems where the model is incorrectly specified. Our definition of ambiguity is based on the Kullback directed divergence between two probability densities used to describe the observed data and follows from a geometric interpretation of ML estimation. By using Sanov's theorem and Stein's lemma, [2], [3], [7], we relate our ambiguity function to the probability of certain types of error in decision problems.

We consider three definitions of ambiguity. The first definition is a direct generalization of the classical ambiguity function of Woodward. We make no specific assumptions regarding the structure of the observed data beyond complete knowledge of its statistical description. This definition shows that in a wider context, the angular interpretation of Woodward's ambiguity generalizes to direct divergence. Our definition reveals interesting features of the ambiguity function, for example, in general, it is not symmetric and depends on the noise characteristics, in particular, the signal-to-noise ratio. The second definition allows for the presence of nuisance parameters like for example unknown power level or when the parameters of interest only restrict the signals to a particular finite-dimensional linear space. In the latter case, we show that the principal angles between subspaces determine upper and lower bounds on the ambiguity function. The third definition of ambiguity addresses the problems where there are mismatches between the assumed model parameters and the actual values these parameters take in the real world. This is particularly important in applications like matched field processing, where detailed physical models are coupled with the signal processing algorithms. Our definition provides a tool to identify systematic biases induced by the mismatches in the estimation of the desired parameters.

The paper illustrates our definitions in the context of source localization in a multipath environment. We compare the ambiguity surfaces associated with distinct channel models: single direct path, direct and surface reflected path, and an incomplete model that ignores the temporal structure provided by the incoming rays. This study shows that modeling the temporal structure can be a determining factor to resolve ambiguity in source localization.

Appendix

A. Geometric Interpretation of ML

We recall in this Appendix concepts and results from information theory and ML estimation. We omit proofs referring the reader to the relevant bibliography.

B. Kullback Directed Divergence

The Kullback directed divergence (KDD) [7] is a measure of similarity between probability densities and bears a fundamental relationship to ML theory; see [1]. The KDD is defined by

$$I(p:q) = \mathrm{E}_p\left\{\ln \frac{p}{q}\right\} \qquad (31)$$

where $\mathrm{E}_p\{\cdot\}$ is expectation with respect to the pdf p. The KDD does not satisfy all the properties of a distance. As seen from its definition (31), the KDD is not symmetric and does not satisfy the triangular inequality. However, it can be shown that [7]

$$I(p:q) \geq 0 \quad \text{and} \quad I(p:q) = 0 \Leftrightarrow p = q \tag{32}$$

where the second equality holds except, possibly, on sets of measure zero.

C. Kullback Directed Divergence of a pdf to a Submanifold

Let $\mathcal{G}$ be a submanifold of densities. We introduce the KDD of the pdf q to the submanifold $\mathcal{G}$

$$I(q:\mathcal{G}) = \min_{p \in \mathcal{G}} I(q:p). \tag{33}$$

As an illustration, we consider the KDD between two multivariate Gauss pdf's p and q parametrized by parameters α and α_0. If the pdf's have the same covariance R and different means μ_α and μ_{α_0}, the KDD is given by the square of the Mahalanobis distance

$$\begin{aligned} I(p:q) &= \tfrac{1}{2} \|\mu_\alpha - \mu_{\alpha_0}\|_R^2 \\ &= \tfrac{1}{2} (\mu_\alpha - \mu_{\alpha_0})^T R^{-1} (\mu_\alpha - \mu_{\alpha_0}). \end{aligned} \tag{34}$$

If the multivariate Gauss pdf's p and q have the same mean μ and distinct covariance matrices R_α and R_{α_0}, the KDD is

$$I(p:q) = \tfrac{1}{2} [\text{tr}[R_\alpha^{-1} R_{\alpha_0}] - N - \ln |R_\alpha^{-1} R_{\alpha_0}|]. \tag{35}$$

As a short-hand notation, when the pdf's are parametrized by α and α_0 as in (34) and (35), we will often write $I(\alpha, \alpha_0)$ for the KDD.

D. Exponential Family

In many parameter estimation applications, the pdf's of interest are a subfamily of the family of exponential densities. The family of exponential densities $\mathcal{G}$ is well known to statisticians. Exponential densities enjoy a number of properties that make them attractive when dealing with inference procedures; see [1]. The exponential family is a parametric family of densities of the form

$$p(r_T|a) = \exp\{\langle a, b(r_T)\rangle - c(a) + d(r_T)\} \tag{36}$$

where the parameter vector a is called the *natural* parameter of the family, $b(\cdot)$ and $d(\cdot)$ are known functions, $\exp[-c(a)]$ is the normalizing constant, and $\langle \cdot, \cdot \rangle$ is a suitably defined inner product. This family of densities includes several distributions such as the Gauss, gamma, binomial, Wishart, and Poisson. However, from a practical viewpoint, of these distributions, the one with most practical significance as a noise model is the Gauss multivariate distribution so that when we refer to the exponential family, we can in fact think of the Gauss pdf.

We illustrate this definition for multivariate Gauss probability density functions (pdf) $p(r_T)$ with mean μ and covariance R. If the covariance is known, the mean μ is the natural parameter of the exponential family $\mathcal{G}_\mu = \{p(r_T|\mu): \mu \in \mathbf{R}^p\}$. The quantities defining the family are

$$a = \mu \qquad c(a) = \frac{1}{2} \mu^T R^{-1} \mu$$

$$b(r_T) = R^{-1} \sum_{i=1}^{N} r_{t_i}$$

$$d(r_T) = -\tfrac{1}{2} \sum_{i=1}^{N} r_T^T R^{-1} r_{t_i} - \frac{pN}{2} \ln(2\pi) - \frac{p}{2} \ln |R|.$$

The inner product in the definition of the exponential distribution is the usual l_2 inner product.

If the mean is known and the covariance R unknown, the inverse R^{-1} is the natural parameter of the family $\mathcal{G}_R = \{p(r_T|R): R \in \mathcal{S}\}$, where $\mathcal{S}$ is the cone of symmetric positive definite matrices. With the following identifications, $\mathcal{G}_R$ is an exponential family.

$$a = R^{-1} \qquad c(a) = \tfrac{1}{2} \ln |R|$$

$$b(r_T) = -\frac{1}{2} \frac{1}{N} \sum_{i=1}^{N} (r_{t_i} - \mu)(r_{t_i} - \mu)^T$$

$$d(r_T) = -\frac{pN}{2} \ln(2\pi).$$

The inner product is defined in the cone of positive definite matrices as $\langle A, B \rangle = \text{tr}[AB]$.

1) Curved Exponential Family: In many applications, the natural parameter a of the exponential family is itself parametrized by a vector of parameters α. In parameter estimation problems, these may represent the unknown quantities in the model. We indicate this dependence of the natural parameter on the vector of parameters α by writing $a(\alpha)$.

The subfamily of exponential pdf's that results by parametrizing the natural parameter a by α is called the curved exponential family $\mathcal{G}_\alpha$. In other words, $\mathcal{G}_\alpha$ is a curved exponential family if its members are written as

$$\begin{aligned} \forall\, \alpha \in \mathcal{A}: \quad p(r_T|\alpha) &= p[r_T|a(\alpha)] \\ &= \exp\{\langle a(\alpha), b(r_T)\rangle \\ &\quad - c[a(\alpha)] + d(r_T)\} \end{aligned} \tag{37}$$

where $\mathcal{A}$ is the space of the parameter α. We denote by $A_\alpha \subset A$ the image of $\mathcal{A}$ under the mapping $\alpha \longrightarrow a(\alpha)$ and similarly for $B_\alpha \subset B$.

2) Exponential Data pdf $\hat{p}(r_T)$: In the geometric interpretation of parameter estimation and detection, we associate with each observation $r_T = r_T^*$ the *exponential data* pdf $\hat{p}(r_T)$. We show how to construct this exponential density from r_T^*. To compute the (unstructured) maximum-likelihood (ML) estimate a^* of the natural parameter a, evaluate the gradient of the log of $p(r_T|a)$ with respect to a, i.e., the score function $s(a)$ and equate to zero

$$b(r_T) = \left. \frac{\partial c(a)}{\partial a} \right|_{a^*}. \tag{38}$$

The *exponential data* pdf $\hat{p}(r_T)$ is now defined as the pdf $p(r_T|a)$ corresponding to a^*, i.e., $\hat{p}(r_T) = p(r_T|a^*)$. With Gaussian densities, it is easy to show that $\hat{p}(r_T)$ is just the Gaussian density with mean and covariance equal to the sample mean and sample covariance matrix.

3) Ideal Sampling Assumption: Our definition of ambiguity in Section IV makes the following assumption. We work with the curved exponential family $\mathcal{G}_\alpha$ of pdf's parametrized by α. Let the observation be $r_T^\star$. With these observations, we associate two pdf's. The first is the exponential data pdf $\hat{p}(r_T)$ determined from the observed data $r_T^\star$, as indicated above. The second is the true pdf $p_0 = p(r_T|\alpha_0)$, where α_0 is the true value of the parameter that gives rise to the observation $r_T^\star$. The *ideal sampling assumption* states that $r_T^\star$ perfectly determines the true pdf $p(r_T|\alpha_0)$, i.e., that

$$\hat{p}(r_T) = p(r_T|\alpha_0). \tag{39}$$

This is an asymptotic type assumption since, as it is well known, the ML estimate is consistent, i.e., it converges in the large sample size limit with probability one to the true value of the estimate.

E. Geometric Interpretation of ML Estimation and Detection

In parameter estimation problems, the family of pdf's of interest is the curved exponential family $\mathcal{G}_\alpha$; see Fig. 4. Let the observed data be $r_T^\star$. It determines the exponential data pdf $\hat{p}(r_T)$ as given above. This pdf is a point in the larger set of exponential family of pdf's $\mathcal{G}$ of which $\mathcal{G}_\alpha$ is a subset. The following fact establishes the role played by the Kullback directed divergence in ML estimation. Optimal ML estimators $\hat{\alpha}$ are obtained by finding the element $p(r_T|\hat{\alpha}) \in \mathcal{G}_\alpha$ that is closer to $\hat{p}$ in the sense of a conveniently defined distance.

Fact 5 [4]: Let $\mathcal{G}$ be an exponential family of probability density functions and α a vector of parameters. Let $\mathcal{G}_\alpha$, which is the probabilistic model that describes the observed data, be a curved exponential family in $\mathcal{G}$. Then, the ML estimate of α is the point $\hat{\alpha}$ such that

$$\hat{\alpha} = \arg\min_{\alpha \in \mathcal{A}} I\{\hat{p}(r_T) : p[r_T|a(\alpha)]\}$$

where $I(p : q)$ is the Kullback-directed divergence between densities p and q. Further, the pdf $\hat{p}(r_T)$ is a member of the exponential family $\mathcal{G}$ of which $\mathcal{G}_\alpha$ is a subset and is uniquely determined from the data r_T. It follows from this fact that the geometry of $\mathcal{G}_\alpha$ as a subset of $\mathcal{G}$ is the factor that determines the global performance of the estimation process. Two distinct values of α are ambiguous, or easily mistaken one with the other, if the distance between them is small. This is the key idea underlying our definition of ambiguity.

Fig. 4 illustrates the relationships between i) the unstructured ML-estimate a^* in the natural parameter space A and the exponential data pdf $\hat{p}(r_T)$ in the probabilistic model $\mathcal{G}$ and between ii) the structured ML estimate of the natural parameter $a(\hat{\alpha}) \in A_\alpha$ and the minimizing density for Fact 5 $p[r_T|a(\alpha)] \in \mathcal{G}_\alpha$.

Fact 5 provides us with a simple geometric picture of ML on which we will base our definition of ambiguity function. We quote a second important fact that establishes the relationship between binary decision and Kullback divergence. We consider *composite* decision problems, where to each hypothesis corresponds a set of possible density functions. The simpler case is when we have a parametric description of each of these sets, but we do not know the values of

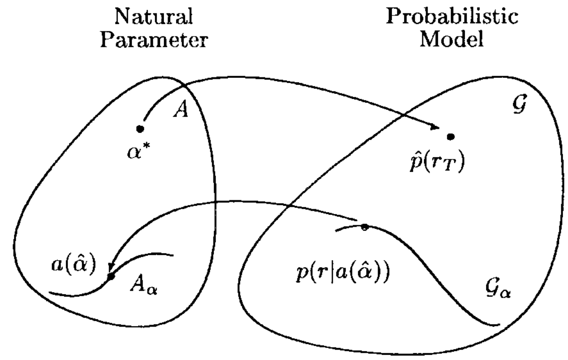


Fig. 4. ML estimation.

these parameters, which are called nuisance parameters. For simplicity, consider the binary composite hypotheses problem. Denote the two hypotheses by H_0 and H_1, and let $\mathcal{G}_0$ and $\mathcal{G}_1$ be the families of densities corresponding to each hypothesis. The test decides which of the hypotheses (H_0 or H_1) is true in the absence of knowledge of what values the unwanted, unknown parameters take.

Fact 6 [7]: The generalized log-likelihood ratio for deciding between H_0 ($\mathcal{G}_0$) and H_1 ($\mathcal{G}_1$) is

$$\begin{aligned}\Lambda(H_0, H_1) &= \ln \frac{\max_{p_1 \in \mathcal{G}_1} p_1(r_T)}{\max_{p_0 \in \mathcal{G}_0} p_0(r_T)} \\ &= \min_{p_0 \in \mathcal{G}_0} I[\hat{p}(r_T) : p_0] - \min_{p_1 \in \mathcal{G}_1} I[\hat{p}(r_T) : p_1]) \\ &= I[\hat{p}(r_T) : \mathcal{G}_0] - I[\hat{p}(r_T) : \mathcal{G}_1] \end{aligned} \tag{40}$$

where $\hat{p}(r_T)$ is the exponential density determined from the data (see D2 in this Appendix).

This fact asserts that the generalized likelihood ratio between two alternative hypotheses corresponds to comparing (in the sense of the Kullback directed divergence) the densities that under each hypothesis best explain the observed data.

For simple decision binary tests $\mathcal{G}_0$ and $\mathcal{G}_1$ are singletons with pdf's p_0 and p_1, respectively. In this case, the log-likelihood ratio is simply

$$\Lambda(H_0, H_1) = I[\hat{p}(r_T) : p_0] - I[\hat{p}(r_T) : p_1]. \tag{41}$$

REFERENCES

[1] S. Amari, "Differential geometrical theory of statistics," in *IMS Monograph vol. 10, Differential Geometry in Statistical Inference,* Hayward, CA, 1987, pp. 20–94.
[2] J. A. Bucklew, *Large Deviation Techniques in Decision, Simulation, and Estimation.* New York: Wiley, 1990.
[3] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* New York: Wiley, 1991.
[4] B. Effron, "The geometry of the exponential families," *Ann. Statist.,* vol. 6, pp. 362–376, 1978.
[5] I. Gohberg, P. Lancaster, and L. Rodman, *Invariant Subspaces of Matrices with Applications.* New York: Wiley, 1986.
[6] G. H. Golub and C. F. Van Loan, *Matrix Computations.* Baltimore, MD: John Hopkins Univ. Press, 1989.
[7] S. Kullback, *Information Theory and Statistics.* London, U.K.: Peter Smith, 1978.
[8] J. M. F. Moura, "Passive systems theory with narrow-band and linear constraints: Part III—Spatial/temporal diversity," *IEEE J. Oceanic Eng.,* vol. OE-9, pp. 113–119, July 1979.
[9] M. J. D. Rendas and J. M. F. Moura, "Ambiguity analysis in source localization with unknown signals," in *Int. Conf. Acoust., Speech Signal Process. (ICASSP),* Toronto, Ont., Canada, May 1991.

[10] I. Reuven and H. Messer, "The use of the Barankin bounds for determining the threshold SNR in estimating the bearing of a source in the presence of another," in *Int. Conf. Acoust., Speech Signal Processing, (ICASSP)*, Detroit, MI, May 1995, vol. 3, pp. 1645–1648.

[11] S. S. Soliman and R. A. Scholtz, "Spread ambiguity functions," *IEEE Trans. Inform. Theory*, vol. 34, Mar. 1988.

[12] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part III*. New York: Wiley, 1971.

[13] P. M. Woodward, *Probability and Information Theory, With Applications to RADAR*. New York: Pergamon, 1953.

M. João D. Rendas (S'83–M'91) received the engenheiro electrotécnico degree in 1980, the M.Sc. degree in electrical and computer engineering in 1985, and the Ph.D. degree in electrical and computer engineering in 1991, all from Instituto Superior Técnico (IST), Lisbon, Portugal.

She was a visiting researcher at Carnegie Mellon University (CMU), Pittsburgh, PA, from 1987 to 1989. In 1991, she joined the faculty of IST as an Assistant Professor and was promoted to Associate Professor in 1994. She held a Fellowship from the European Human Capital and Mobility Program from 1993 to 1995 at the CNRS Laboratoire I3S, Sophia–Antipolis, France. Since 1995, she has been a chargée de recherches at CNRS (Laboratoire I3S). Her research interests are in statistical signal processing, fuzzy and statistical characterization of uncertainty, and the geometry of statistical inference, with applications in perception for autonomous robots and inversion problems in underwater communications. Her research has been supported by several European programs, including Esprit, MAST, ACTS, and by Portuguese and French supported research projects.

José M. F. Moura (S'71–M'75–SM'90–F'94) received the engenheiro electrotécnico degree in 1969 from Instituto Superior Técnico (IST), Lisbon, Portugal, and the M.Sc., E.E., and the D.Sc. in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge, in 1973 and 1975, respectively.

He is presently a Professor of Electrical and Computer Engineering at Carnegie Mellon University (CMU), Pittsburgh, PA, which he joined in 1986. Prior to that, he was on the faculty of IST, where he was an Assistant Professor (1975), Professor Agregado (1978), and Professor Catedrático (1979). He has had visiting appointments at several Institutions, including MIT (Genrad Associate Professor of Electrical Engineering and Computer Science from 1984 to 1986) and the University of Southern California (Research Scholar, Department of Aerospace Engineering, Summers 1978 to 1981). His research interests include statistical signal processing (one- and two-dimensional), digital communications, image and video processing, radar and sonar, and multiresolution techniques. He has organized and codirected two international scientific meetings on signal processing theory and applications. He has published more than 170 technical contributions and is co-editor of two books.

Dr. Moura is currently the Editor-in-Chief for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and a member of the *Underwater Acoustics* and *Multimedia Signal Processing* Technical Committees of the IEEE Signal Processing Society. He was a member of the IEEE Press Board from 1991 ro 1995, a Technical Associate Editor for the IEEE SIGNAL PROCESSING LETTERS from 1993 to 1995, and an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 1988 to 1992. He was a program committee member for the IEEE International Conference on Image Processing (ICIP'95) and for the IEEE International Symposium on Information Theory (ISIT'93). He is a corresponding member of the *Academy of Sciences of Portugal* (Section of Sciences). He is affiliated with several IEEE societies, Sigma Xi, AMS, IMS, and SIAM.