

Randomized Data Selection in Detection with Applications to Distributed Signal Processing

CHARLES K. SESTOK, STUDENT MEMBER, IEEE, MAYA R. SAID, STUDENT MEMBER, IEEE, AND ALAN V. OPPENHEIM, FELLOW, IEEE

Invited Paper

Performing robust detection with resource limitations such as low-power requirements or limited communication bandwidth is becoming increasingly important in contexts involving distributed signal processing. One way to address these constraints consists of reducing the amount of data used by the detection algorithms. Intelligent data selection in detection can be highly dependent on a priori information about the signal and noise. In this paper, we explore detection strategies based on randomized data selection and analyze the resulting algorithms' performance. Randomized data selection is a viable approach in the absence of reliable and detailed a priori information, and it provides a reasonable lower bound on signal processing performance as more a priori information is incorporated. The randomized selection procedure has the added benefits of simple implementation in a distributed environment and limited communication overhead. As an example of detection algorithms based upon randomized selection, we analyze a binary hypothesis testing problem, and determine several useful properties of detectors derived from the likelihood ratio test. Additionally, we suggest an adaptive detector that accounts for fluctuations in the selected data subset. The advantages and disadvantages of this approach in distributed sensor networks applications are also discussed.

Keywords—Distributed signal processing, likelihood ratio test, nonparametric detection, randomized algorithms, randomized sampling, sensor networks.

Manuscript received October 20, 2002; revised March 17, 2003. This work was supported in part by participation in the Advanced Sensors Collaborative Technology Alliance (CTA) sponsored by the U.S. Army Research Laboratory under Cooperative Agreement DAAD19-01-2-008, in part by the Space and Naval Warfare Systems Center—San Diego and the Defense Advanced Research Projects Agency (DARPA) IXO under Contract N66001-00-C-8054, in part by DARPA under Contract MDA972-00-1-0030, and in part by the Texas Instruments Leadership University Consortium Program. The work of M.R. Said was supported by a Merck/MIT Fellowship.

The authors are with the Research Laboratory for Electronics, Massachusetts Institute of Technology, Cambridge, MA, 02139 USA (e-mail: cksestok@mit.edu; mayasaid@mit.edu; avo@mit.edu).

Digital Object Identifier 10.1109/JPROC.2003.814922

I. INTRODUCTION

With the recent advances in device and computing technologies, static as well as dynamic networks such as cellular systems, the Internet, and sensor networks, are becoming increasingly ubiquitous. Designing signal processing algorithms that satisfy constraints imposed by these networks is, therefore, becoming a necessity [1], [2]. Specifically, these algorithms need to be efficient and robust, and in the case of battery-powered networks, they need to operate under power constraints [3], [4]. In addition, there is usually the added requirement of restricted communication bandwidth for wireless networks and, therefore, careful management of the network's data transmission volume is important [5].

While it may be appropriate to design networks that densely populate a region with microsensors during sensor deployment, operation of the network may not require that all network nodes be operating and communicating at once. Indeed, for efficient operation, extended network lifetime, and efficient use of communication bandwidth, it may be desirable to select a subset of nodes to communicate at any fixed time [6]. The selected subset changes over time, varying usage among the nodes to extend their effective lifetime.

An appropriate algorithm for node subset selection in a densely populated network can be highly dependent on the *a priori* information about the characteristics of both the signal and noise for a specific task or environment, and, consequently, it would be unreasonable to attempt to formulate a generic optimal procedure. In this paper, we explore a randomized approach to node selection. Specifically, we abstract and simplify the problem by considering nodes as only communicating data for a signal processing task. We use the average rate at which an individual node is included in the selected subset as the basic design parameter. The selection reduces communication bandwidth requirements by limiting the amount of data transferred between nodes. Under the assumption that communication dominates the

energy usage of the nodes, the network lifetime increases as this average rate decreases. Using this model, we consider randomized data selection and analyze the network performance as a function of this average rate for several classes of detection algorithms. In addition, this approach can be used to model intermittent hardware or communication failures and, therefore, the algorithms presented in this paper offer ways to compensate for these failures.

In Section II, we consider randomized data selection and discuss its impact on signal statistics. Section III analyzes the performance of randomized data selection algorithms in detection problems, quantifying the performance versus the average sampling or selection rate. We consider a wide class of potential target signals, analyze a specific example, and develop a low-complexity, suboptimal detector for the entire signal class. Section IV reviews our results and discusses directions for further investigation.

II. RANDOMIZED DATA SELECTION

Distributed networks are composed of interacting hardware and software systems. The energy efficiency of these networks can be improved by modifying the hardware [7] or any of the algorithms, such as data routing [8], source coding [10], [11], or signal processing. In this paper, we focus on data selection as an algorithmic approach to improving the network's energy efficiency. Similar strategies have been used as a design for efficient systems in such diverse fields as filter approximation [12], statistical regression [13], and multiple-input multiple-output (MIMO) wireless communication [14], [15]. Selection, by reducing the amount of communication congestion throughout the network and avoiding the computational burden of processing all available data, relieves two major sources of energy dissipation. Since we are concerned primarily with the selection algorithm and its impact upon signal processing performance, we do not attempt to quantify the energy savings because it would be highly dependent upon specific algorithm or hardware properties.

While data selection algorithms accounting for many aspects of the network's state can be useful in practice, we choose a generic approach requiring limited *a priori* information and communication overhead. Specifically, we consider a randomized data selection strategy. This approach leads to useful algorithms in distinct fields such as estimation, hardware failure modeling, low-power design [17], and theoretical computer science [18].

A. Notation

Throughout this paper, we use the following notation to describe the data selection and signal processing algorithms used by the distributed system. Our analysis focuses on signal processing procedures using data collected in a single time slot. We assume that each measurement is assigned an identifier index i , arbitrarily chosen between 1 and N ,

where N is the total number of network nodes. Subject to this model, the total data available in time slot m is denoted

$$\mathbf{x}[m] = \begin{bmatrix} x_1[m] \\ x_2[m] \\ \vdots \\ x_N[m] \end{bmatrix}. \quad (1)$$

In our randomized selection rule, the decision to select measurement $x_a[m]$ depends on the outcome of an indicator random variable denoted $g_a[m]$. The random variable is independent of all other indicator random variables and from other physically measurable quantities available to the detector. In our model, each measurement in the current time slot is selected with probability γ_g , i.e. $g_a[m]$ ¹ has the probability mass function

$$p_{g_a[m]}(g) = \begin{cases} \gamma_g, & g = 1 \\ (1 - \gamma_g), & g = 0. \end{cases} \quad (2)$$

This selection rule reduces the expected complexity of the detector implementation by a factor of γ_g .

The randomly selected data vector, $\mathbf{x}_g[m]$, can be represented by the equation

$$\mathbf{x}_g[m] = \mathbf{G}[m]\mathbf{x}[m]. \quad (3)$$

Here, $\mathbf{x}[m]$ is the data vector defined in (1), and $\mathbf{G}[m]$ is an $N \times N$ diagonal matrix with the i th entry given by $\mathbf{G}_{ii}[m] = g_i[m]$. Consequently, the vector $\mathbf{x}_g[m]$ is N -dimensional with each entry being either zero or a data measurement. In each time slot m , the number of nonzero entries of $\mathbf{x}_g[m]$ is a random variable, which we denote by $K[m]$. Prior to discussing the specific detection problems, we examine the signal statistics for $\mathbf{x}_g[m]$.

Since the detector only receives a portion of the data, we must base our algorithms upon the conditional density for $\mathbf{x}_g[m]$ given $\mathbf{G}[m] = \mathbf{G}$. To establish the notation for the conditional density, let the set

$$\mathcal{S}[m] = \{j | \mathbf{G}_{jj}[m] = 1\} \quad (4)$$

denote the selected measurements in time slot m . If a particular realization of this set is $\mathcal{S} = \{j_1, j_2, \dots, j_K\}$, the conditional density for $\mathbf{x}_g[m]$ is the joint density for $x_{j_1}[m], x_{j_2}[m], \dots, x_{j_K}[m]$. The indicator random variables are independent of $\mathbf{x}[m]$, so no useful information about the signal is gained by observing $\mathbf{G}[m]$.

For notational convenience, we will denote the conditional density of $\mathbf{x}_g[m]$ given $\mathbf{G}[m] = \mathbf{G}$ by the expression

$$p_{\mathbf{x}_g | \mathbf{G}}(\mathbf{x} | \mathbf{G}) \quad (5)$$

with the dependence on m understood.

B. Application to Distributed Signal Processing

The properties of a sensor network strongly influence the choice of appropriate distributed signal processing algorithms. Important variables influencing this choice include the number and density of sensor nodes, the area

¹Throughout this paper, random variables will be denoted with a time slot index, such as $g_a[m]$. Scalars or vectors, such as g_a or g , are denoted without the indexes.

covered by the network, available information about external environment, and the communication capabilities of the network.

Randomized data selection may be an attractive approach to distributed signal processing in a variety of situations. For example, many simple sensor nodes may be densely distributed throughout the extent of the network. Additionally, no *a priori* information about where targets are likely to appear may be available to aid the sensor selection algorithm. Due to the high node density, data from neighboring nodes can be highly redundant. In this situation, random selection with a small value of γ_g can lead to acceptable detector performance, while limiting the energy dissipated by communication of sensor data through the network. Additionally, the randomized selection procedure avoids computational or communication overhead that may be incurred from more complicated iterative selection procedures [16], or from centralized coordination of sensor selection.

Randomized selection is compatible with common architectures for ad hoc wireless networking. Many networks use a combination local clustering (where a group of nodes communicate with a base-station node) and multihop routing in their networking protocols. Clustering can be combined with randomized data selection; in every time slot, each node in the cluster randomly determines whether to transmit its sensor measurement to the base station. The selection procedure limits the expected amount of data processed by each local base-station node. In multihop routing, each of the selected sensor measurements follows a path through several nodes in the network. Any node in the network sees a random number of packets from an individual time slot. Thus, if necessary, any node in the network can use the detection algorithms we describe.

Finally, note that the techniques we use to adapt the detector to fluctuations in the size of $\mathbf{x}_g[m]$ can be applied to situations where unreliable sensor or communication hardware lead to intermittent loss of data in the network.

III. RANDOMIZED SELECTION IN DETECTION

In this section, we analyze the interaction of randomized selection and signal detection in a background of additive white Gaussian noise. The canonical detector from a binary hypothesis testing model is the likelihood ratio test [19]. The test compares the likelihood ratio $L(\mathbf{x})$, defined as the ratio between the conditional densities for $\mathbf{x}[m]$, with a fixed threshold η . If the detector observes \mathbf{x} in a region of sample space where $L(\mathbf{x}) \geq \eta$, it makes the decision $\hat{H} = H_1$. Otherwise, it decides $\hat{H} = H_0$. We denote decision rules of this form with the notation

$$L(\mathbf{x}) \underset{\hat{H}=H_0}{\overset{\hat{H}=H_1}{\geq}} \eta. \quad (6)$$

Our analysis of the likelihood ratio test highlights two key issues inherent to randomized data selection. First, we discuss the binary hypothesis test, and account for random selection in its statistical model. Second, we suggest low-complexity detectors that adapt to fluctuation in the amount of selected

data. Additionally, we consider the robustness of the detector to inaccuracies or unknown parameters in the *a priori* model for the target signal's probability density. While this issue is not directly related to random sampling, it illustrates the challenging signal processing environment in which detectors often operate.

A. General Signal Model

In order to derive useful properties of the likelihood ratio test in the presence of random selection, we impose restrictions on the statistical model for the target signature. To balance the generality of the signal model with its special statistical structure, we assume that the probability density of the target signal is symmetric about the origin of the sample space. We shall refer to random vectors that satisfy this condition as even random vectors or even-symmetric signals. The precise definition of an even random vector is given in Definition 1.

Definition 1: An N -dimensional random vector $\mathbf{s}[m]$ is referred to as even if, for every $\mathbf{s}_o \in \mathfrak{R}^N$, its probability density function satisfies $p_{\mathbf{s}}(\mathbf{s}_o) = p_{\mathbf{s}}(-\mathbf{s}_o)$.

This signal model establishes a useful structure on the probability density of the signal, enabling us to determine key properties of the likelihood ratio test. Additionally, the signal model is broad enough to model many interesting target signatures. For example, a sinusoid with an unknown, uniformly distributed phase satisfies the condition in Definition 1, as does a zero-mean, Gaussian random vector with a known covariance matrix.

The general binary hypothesis test for signals in additive Gaussian noise obeys the following statistical model:

$$\begin{aligned} H_0 : \mathbf{x}[m] &= \mathbf{n}[m] \\ H_1 : \mathbf{x}[m] &= \mathbf{s}[m] + \mathbf{n}[m]. \end{aligned} \quad (7)$$

Here, we assume that $\mathbf{n}[m]$, is an N -dimensional, zero-mean, white Gaussian random vector with covariance $\Lambda_{\mathbf{n}} = \sigma^2 \mathbf{I}$. The signal vector $\mathbf{s}[m]$ has an even-symmetric probability density. Finally, we assume that $\mathbf{s}[m]$ and $\mathbf{n}[m]$ are independent random vectors. This model describes the statistics of the data without randomized selection.

In the presence of randomized data selection, the detector has access to the indicator random variables in $\mathbf{G}[m]$ and processes the subset of the available data contained in $\mathbf{x}_g[m]$. The likelihood ratio for detectors with randomized selection can be expressed as

$$\begin{aligned} L(\mathbf{x}_g, \mathbf{G}) &= \frac{p_{\mathbf{x}_g, \mathbf{G}|H}(\mathbf{x}_g, \mathbf{G}|H_1)}{p_{\mathbf{x}_g, \mathbf{G}|H}(\mathbf{x}_g, \mathbf{G}|H_0)} \\ &= \frac{p_{\mathbf{x}_g|\mathbf{G}, H}(\mathbf{x}_g|\mathbf{G}, H_1) p_{\mathbf{G}|H}(\mathbf{G}|H_1)}{p_{\mathbf{x}_g|\mathbf{G}, H}(\mathbf{x}_g|\mathbf{G}, H_0) p_{\mathbf{G}|H}(\mathbf{G}|H_0)} \\ &= L(\mathbf{x}_g|\mathbf{G}). \end{aligned} \quad (8)$$

The simplification in the likelihood occurs because the indicator random variables are independent of the hypotheses H_i .

Since conditioning upon $\mathbf{G}[m]$ does not affect the selected data in $\mathbf{x}_g[m]$, the detection problem based upon $\mathbf{x}_g[m]$ and $\mathbf{G}[m]$ reduces to an unconditional detection problem for the

data associated with the nonzero indicator random variables. For example, if three pieces of data are available, there are eight possible arrangements of the indicator random variables. If measurements 1 and 2 are selected in time slot m , the detector must make a decision \hat{H} based upon the joint densities

$$\begin{aligned} p_{\mathbf{x}_g|\mathbf{G},H}(\mathbf{x}|1,2,H_0) &= p_{x_1,x_2|H}(x_1,x_2|H_0) \\ p_{\mathbf{x}_g|\mathbf{G},H}(\mathbf{x}|1,2,H_1) &= p_{x_1,x_2|H}(x_1,x_2|H_1). \end{aligned} \quad (9)$$

Likewise, if measurements 2 and 3 are selected, the decision \hat{H} is determined from $p_{x_2,x_3|H}(x_2,x_3|H_0)$ and $p_{x_2,x_3|H}(x_2,x_3|H_1)$.

Based upon (8), the likelihood ratio test for $\mathbf{x}_g[m]$ and $\mathbf{G}[m]$ reduces to the comparison of $L(\mathbf{x}_g|\mathbf{G})$ to a fixed threshold. While the test is optimal under the Neyman–Pearson detection criteria, it poses some practical problems. First, determining the threshold can become computationally complex when there is a large amount of data available for selection. The threshold that achieves a desired false alarm rate \tilde{P}_F is determined by inverting the equation

$$\tilde{P}_F(\eta) = \sum_{\mathbf{G}} p_{\mathbf{G}|H}(\mathbf{G}|H_0) \Pr(L(\mathbf{x}_g) > \eta | \mathbf{G}, H_0). \quad (10)$$

If N samples of data are available, there are 2^N terms in the summation. Although it may be possible to approximate this function well by discarding terms with low $p_{\mathbf{G}|H}(\mathbf{G}|H_0)$, determining the functional form of such an approximation may be troublesome. The functions of the threshold η given by $\Pr(L(\mathbf{x}_g) > \eta | \mathbf{G}, H_0)$ may not be easily parameterized. Second, since the threshold η is constant while \mathbf{G} fluctuates, the conditional false alarm rate $P_F(\mathbf{G}, \eta) = \Pr(L(\mathbf{x}_g) > \eta | \mathbf{G}, H_0)$ fluctuates as well. In a situation where actions taken following a false alarm are costly, however, this fluctuation may not be desirable, since it is induced by the random data selection rather than an information-bearing signal.

Faced with the practical difficulties of solving (10) for η , a suboptimal yet tractable alternative seems desirable. A reasonable approach fixes the conditional false alarm rate

$$P_F(\mathbf{G}, \eta) = \tilde{P}_F \quad (11)$$

for each realization of \mathbf{G} . Similar procedures have been discussed in [22] in a two-sensor situation. This constraint eliminates the fluctuations in the conditional false alarm rate and may simplify the implementation of the resulting detector, since the constraint has a constant rather than exponential number of terms. It does, however, require the detector to adapt the test threshold to the arrangement of $\mathbf{G}[m]$. In the remainder of the paper, we will focus on detectors designed with randomized data selection and the constraint imposed by (11).

In the remainder of this section, we discuss detector adaptation from several perspectives. First, we analyze the example problem of detecting a sinusoidal signal, a familiar target signature that satisfies the condition in Definition 1. Second, we demonstrate some properties of the likelihood

ratio test for the general detection problem. Finally, we discuss nonparametric detection for any signal with an even-symmetric density.

B. Example—Detecting a Sinusoidal Signal

We consider detection of a sinusoidal signal in the presence of randomized data selection. Our analysis illustrates the difficulties associated with detection in the presence of uncertainty in the target signal and the challenges imposed by the fixed false alarm requirement from (11).

Consider a set of data generated by sampling a signal at several locations, denoted by v_i , $i = 1, \dots, N$. We shall assume that these locations can be modeled by a set of independent, identically distributed uniform random variables over an interval significantly larger than the sinusoid's wavelength.

Let H_0 denote the state in which the sinusoid is absent, and H_1 denote the state when it is present. The i th measurement under each hypothesis is given by

$$\begin{aligned} H_0 : x_i &= n_i \\ H_1 : x_i &= A \cos\left(2\pi \frac{v_i}{\lambda} + \phi\right) + n_i. \end{aligned} \quad (12)$$

The random variable n_i is a zero-mean Gaussian random variable with variance σ^2 . The probability density for $\mathbf{x}_g[m]$, conditioned upon $\mathbf{G}[m]$ and H_0 is white Gaussian. In order to determine the likelihood ratio and the resulting receiver operating characteristic, we also need the probability density for $\mathbf{x}_g[m]$ conditioned upon $\mathbf{G}[m]$ and H_1 . This conditional density depends, in turn, on the joint density of

$$w_i = 2\pi \frac{v_i}{\lambda} + \phi \quad (13)$$

for the selected data in $\mathcal{S}[m]$. The probability density for the signal is a function of the joint density of the phase random variables. Since the signal and noise are independent under H_1 , the overall conditional density for $\mathbf{x}_g[m]$ is the convolution of the signal density and the noise density. The determination of the joint density for the phase random variables is a key step in this calculation.

Since $\{v_i\}$ are independent and uniform over a large interval, we can approximate $\{w_i\}$ as independent, identically distributed uniform random variables over the region $[-\pi, \pi)$. Using this model, we can analyze the form of the likelihood ratio test for the model suggested in (12). Here, we assume that the base station knows A exactly. The signal is $\mathbf{c}[m]$, where $\mathbf{c}[m]$ is a K -dimensional random vector. Each entry takes the form $c_i[m] = A \cos(w_i)$. Based upon our approximation, the probability density for $\mathbf{c}[m]$ is

$$p_{\mathbf{c}|K}(\mathbf{c}|K) = \prod_{i=1}^K \frac{u(A - |c_i|)}{\pi \sqrt{A^2 - c_i^2}} \quad (14)$$

where $u(\cdot)$ denotes the unit step function. This density is nonzero over the K -dimensional hypercube of side A . For fixed K , we denote² the randomly selected data by $\mathbf{x}_K[m]$.

²When applied to a vector, the subscript K indicates its dimension. This does not contradict our earlier notation, where the subscript of a scalar random variable indicated the identity of the measurement. The dimension subscript is always attached to a vector, not a scalar.

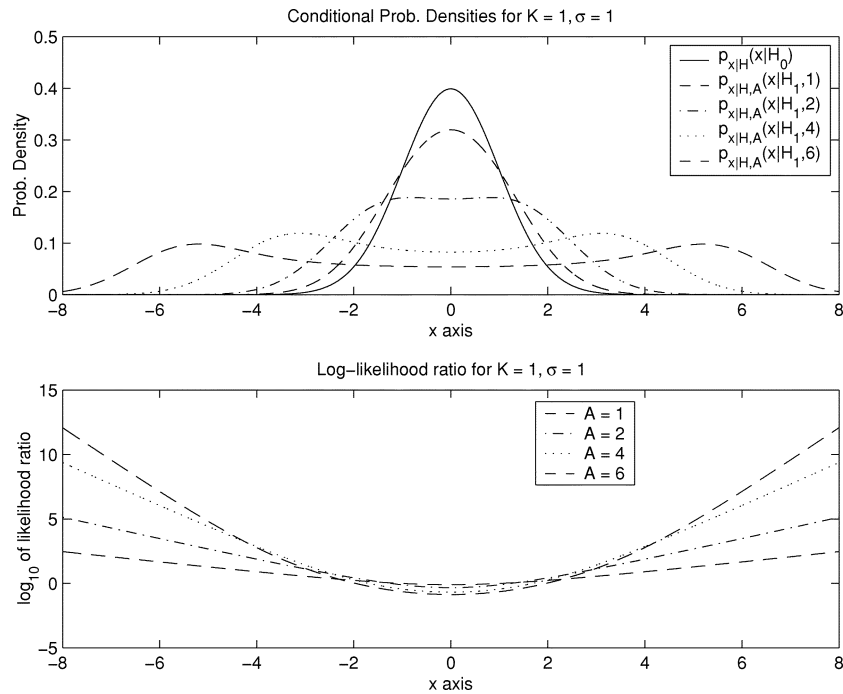


Fig. 1. One-dimensional conditional densities for the signal model defined in (15). The top plot shows the density for H_0 with a solid line. The densities for H_1 with $A = 1, 2, 4,$ and 6 become progressively wider. The second plot shows the corresponding log-likelihood ratios, $L(x|K = 1)$. In all plots, $\sigma^2 = 1$.

This random vector lists the selected data contiguously, rather than with zeros as in \mathbf{x}_g . For notational convenience, we assume that measurements 1 to K are selected,³ so $\mathbf{x}_K = [x_1 \ x_2 \ \dots \ x_K]^T$. The resulting signal model is

$$\begin{aligned} H_0 : \mathbf{x}_K[m] &= \mathbf{n}[m] \\ H_1 : \mathbf{x}_K[m] &= \mathbf{c}[m] + \mathbf{n}[m]. \end{aligned} \quad (15)$$

Based on these probability density functions, we can construct the likelihood ratio test for fixed values of K and A . The conditional density under H_0 is Gaussian with zero mean. Under H_1 , the conditional density is the convolution of the Gaussian with the density for $\mathbf{c}[m]$ given in (14). The conditional density for $\mathbf{x}_K[m]$ under H_1 can be written in terms of a one-dimensional (1-D) convolution, since both conditional densities are separable. The conditional density is

$$\begin{aligned} p_{\mathbf{x}_K|K,H_1}(\mathbf{x}|K, H_1) &= p_{\mathbf{c}|K}(\mathbf{x}_K|K) * p_{\mathbf{n}|K}(\mathbf{x}_K|K) \\ &= \prod_{i=1}^K \int_{-\infty}^{\infty} \frac{u(A - |a_i|)}{\pi \sqrt{A^2 - a_i^2}} \frac{1}{\sqrt{2\pi}\sigma} \\ &\quad \times e^{\left(-\frac{(x_i - a_i)^2}{2\sigma^2}\right)} da_i \\ &= \prod_{i=1}^K p_{x|H}(x_i|H_1). \end{aligned} \quad (16)$$

³This notation does not reduce the applicability of the analysis, since our modeling assumptions make the measurements statistically indistinguishable. Their joint statistics depend only on K and not on the measurement identifiers.

Since $\mathbf{x}_K[m]$ under H_0 is a white, Gaussian random vector with variance σ^2 , the likelihood ratio is

$$\begin{aligned} L(\mathbf{x}_K|K) &= \prod_{i=1}^K \int_{-\infty}^{\infty} \frac{u(A - |a_i|)}{\pi \sqrt{A^2 - a_i^2}} e^{\left(-\frac{(x_i - a_i)^2}{\sigma^2}\right)} e^{\left(-\frac{a_i^2}{2\sigma^2}\right)} da_i \\ &= \prod_{i=1}^K L(x_i). \end{aligned} \quad (17)$$

Analysis of the likelihood ratio test derived from (17) for $K = 1$ and $K = 2$ provides useful intuition about the general properties of the detector. The 1-D conditional densities, $p_{x|H}(x|H_1)$ and $p_{x|H}(x|H_0)$, and the associated log-likelihood ratios are shown in Fig. 1. When $K = 1$, $L(x)$ is symmetric and increasing, so the likelihood ratio test from (6) simplifies to a threshold test of the form

$$|x| \underset{\hat{H}=H_0}{\overset{\hat{H}=H_1}{\gtrless}} t. \quad (18)$$

Since the detector compares the magnitude of the received data with a threshold, the implementation is simple.

Typically, the performance of a detector is shown by an operating characteristic, which plots the detection probability P_D as a function of the false alarm probability P_F . Both P_D and P_F can be calculated by integrating, respectively, the conditional densities $p_{\mathbf{x}_K|K,H}(\mathbf{x}_K|K, H_1)$ and $p_{\mathbf{x}_K|K,H}(\mathbf{x}_K|K, H_0)$ over the $\hat{H} = H_1$ decision region. Thus, the operating characteristic is generated as the threshold in (18) ranges over $0 \leq t < \infty$. It can be shown that the operating characteristic calculated from the

likelihood ratio test gives the maximum achievable P_D for each false alarm rate $0 \leq P_F \leq 1$.

For $K = 1$, the detector described by (18) has an important universality property over the set of binary hypothesis tests for $A > 0$. The threshold that achieves a certain P_F can be determined in terms of the Q function [23] and σ^2 . Since the threshold can be determined without knowledge of the wave amplitude A , the test in (18) is a uniformly most powerful (UMP) test [19]. For such a test, the decision regions that maximize P_D subject to a constraint on P_F are invariant to the actual value of the parameter A . The actual value of P_D , however, does depend on A .

The contrast between the likelihood ratio test for $K = 1$ and $K = 2$ indicates some implementation challenges in the presence of uncertain signal models and random data selection. When $K > 1$, the likelihood ratio test for $\mathbf{x}_K[m]$ is not a function of the received data magnitude $\|\mathbf{x}_K\|$, as shown in Fig. 2. Since the likelihood ratio is increasing in all directions, the likelihood ratio test will declare $\hat{H} = H_0$ in a simply connected region containing the origin. Outside this region, it will declare $\hat{H} = H_1$. Thus, the two-dimensional (2-D) test determines a closed curve, expressed in polar coordinates as $r(\theta)$, that gives the boundary between the decision regions for $\hat{H} = H_0$ and $\hat{H} = H_1$. Since $r(\theta)$ is not constant, the implementation of the likelihood ratio test is more complicated in two dimensions than in one.

For situations where $K > 1$, there is not a UMP detector. In order to determine $r(\theta)$ properly, we require $p_{\mathbf{x}_K|H}(\mathbf{x}|H_0)$, $p_{\mathbf{x}_K|H}(\mathbf{x}|H_1)$, and the desired value of P_F . As shown in Fig. 2, the detector requires A to determine the decision regions in the likelihood ratio test.

Finally, the likelihood ratio test's decision regions depend on the value of K . The shape of the decision regions varies as K changes, as they did when K increased from 1 to 2. Evidently, larger values of K lead to more complicated decision regions. For example, the decision regions for $K = 2$ can be complicated sets in the (x_1, x_2) plane.

The difficulty in determining the decision regions under uncertainty in A and K makes the exact likelihood ratio test on \mathbf{x}_K challenging to implement. First, the fluctuation in K means that the detector must be able to quickly adapt the decision regions for each time slot. Second, potential uncertainties in the target signal density prevent the detector from determining the exact likelihood ratio test. These challenges in the example detection problem persist for the general even signal model.

C. General Properties of the Likelihood Ratio for Even Signals

The sinusoid detection example illustrates several properties of the likelihood ratio in white, Gaussian noise. This section generalizes these properties to signals with even-symmetric probability densities. The qualitative behavior of the resulting decision regions is illustrated, and the prospects for practical implementation are discussed.

Following the derivation of (16) and (17), we can calculate the conditional likelihood ratio for an arbitrary signal with an even-symmetric probability density. The binary hypothesis

test's signal model is given by (7). The resulting expression for the conditional likelihood ratio is

$$L(\mathbf{x}_g|\mathbf{G}) = \int_{\mathbf{a}_K} p_{\mathbf{s}_g|\mathbf{G}}(\mathbf{a}_K|\mathbf{G}) e^{(-\frac{1}{2\sigma^2} \mathbf{a}_K^T \mathbf{a}_K)} e^{(\frac{1}{\sigma^2} \mathbf{x}_g^T \mathbf{a}_K)} d\mathbf{a}_K. \quad (19)$$

In the integral, the variable \mathbf{a}_K is a K -dimensional vector, and the density $p_{\mathbf{s}_g|\mathbf{G}}(\mathbf{a}_K|\mathbf{G})$ denotes the joint density for the selected signal measurements.

The likelihood ratio is easily expressed in Cartesian coordinates. Its qualitative description, however, is easiest in generalized, K -dimensional spherical coordinates. When $K > 3$, the spherical coordinates can be determined via induction. In general, the transformation between spherical and Cartesian coordinates is expressed as

$$\begin{aligned} x_1 &= r \sin(\theta) \prod_{j=3}^K \sin(\phi_j) \\ x_2 &= r \cos(\theta) \prod_{j=3}^K \sin(\phi_j) \\ x_3 &= r \cos(\phi_3) \prod_{j=4}^K \sin(\phi_j) \\ &\vdots \\ x_{K-1} &= r \cos(\phi_{K-1}) \sin(\phi_K) \\ x_K &= r \cos(\phi_K). \end{aligned} \quad (20)$$

The domain of the radius is $r \geq 0$, and the domain of the angular variables is $\theta \in [0, 2\pi)$ and $\phi_i \in [0, \pi)$ for $i = 3, 4, \dots, K$.

Using spherical coordinates, the boundary between the decision regions of the likelihood ratio test can be described. In the 2-D example, the curve dividing the decision regions is denoted by $r(\theta)$. In higher dimensions, we indicate the boundary surface by $r(\Theta)$, where the argument $\Theta = [\theta \ \phi_3 \ \dots \ \phi_K]^T$ is a vector containing all the angular variables.

Without specific knowledge of the signal probability density $p_{\mathbf{s}_g|\mathbf{G}}(\mathbf{a}|\mathbf{G})$, the integral in (19) cannot be evaluated. The properties of even-symmetric signals, however, enable us to discover qualitative properties of the likelihood ratio. In spherical coordinates, we denote the likelihood ratio, conditioned upon $\mathbf{G}[m] = \mathbf{G}$, by $L(r, \Theta|\mathbf{G})$. In Appendix I, we prove the following theorem, showing that the likelihood ratio test for an arbitrary even signal produces decision regions similar to those for the example.

Theorem 1: Consider a detection problem of the class defined in (7). Let $r = \|\mathbf{x}_g\|$ and $\Theta = [\theta \ \phi_3 \ \dots \ \phi_K]^T$. Then, the likelihood ratio $L(r, \Theta|\mathbf{G})$ given by (19) increases monotonically without bound for any fixed Θ .

As a consequence of this theorem, we can describe the likelihood ratio test in terms of $r(\Theta)$, a closed surface containing the origin. The interior of the surface is the decision region $\hat{H} = H_0$, and the remainder of the sample space is the decision region $\hat{H} = H_1$. In order to determine $r(\Theta)$, consider a fixed threshold η for the likelihood ratio

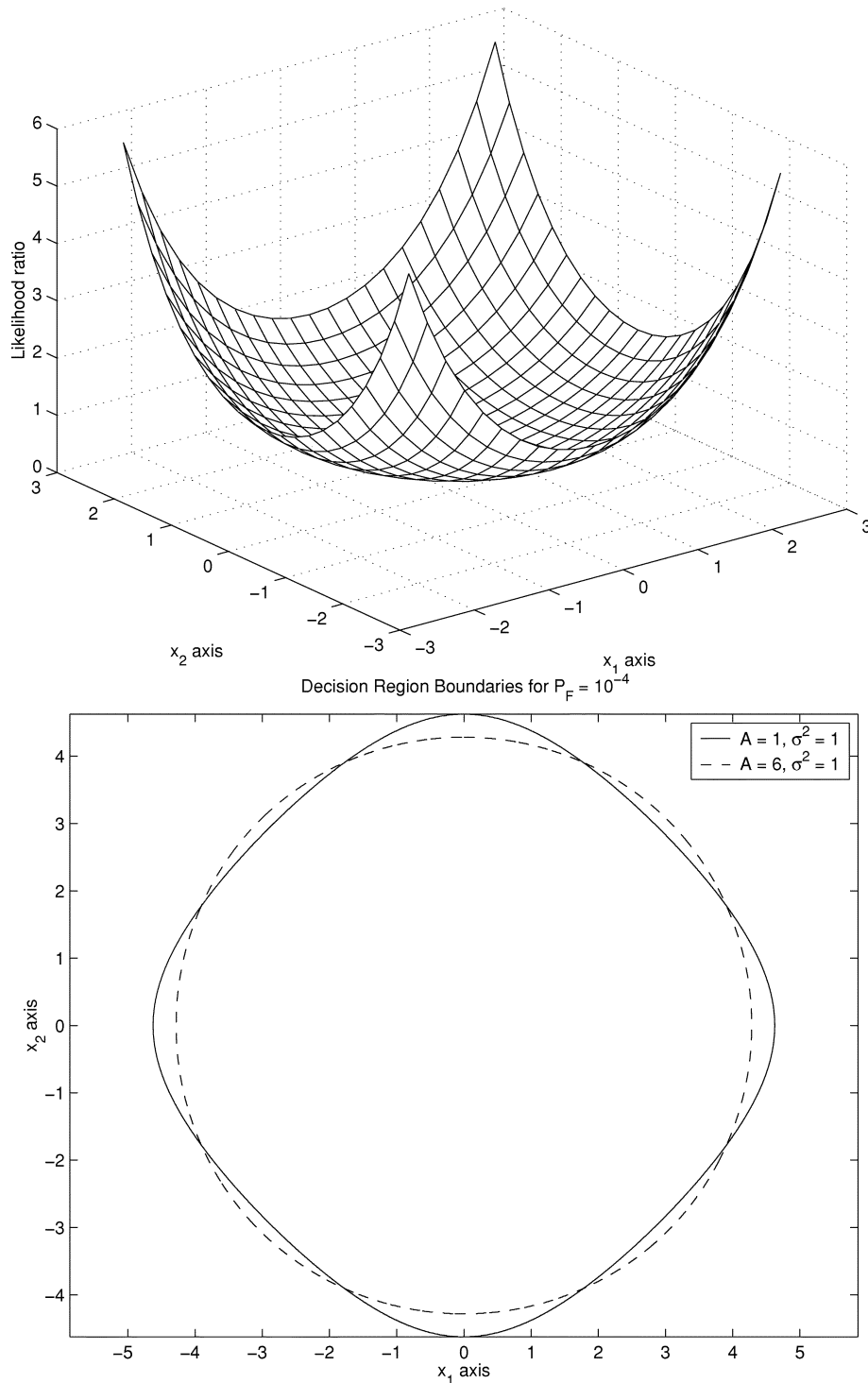


Fig. 2. The top plot shows the 2-D likelihood ratio when $A = 1$ and $\sigma^2 = 1$. The bottom plot shows the decision region boundary curves for $P_F = 10^{-4}$. In both cases, the noise variance is $\sigma^2 = 1$. The solid curve shows the boundary when $A = 1$, and the dashed curve shows the boundary when $A = 6$.

test. Assuming that the probability densities contain no point masses, the subset of sample space satisfying the condition $L(r, \Theta | \mathbf{G}) \leq \eta$ composes the decision region for H_0 . Likewise, the condition $L(r, \Theta | \mathbf{G}) > \eta$ determines the decision region for H_1 . For a fixed vector Θ_o , Theorem 1 implies that the function of r given by $L(r, \Theta_o | \mathbf{G})$ is strictly

monotonically increasing. Thus, there is a unique solution to the equation $L(r, \Theta_o | \mathbf{G}) = \eta$ if $\eta > L(0, \Theta_o | \mathbf{G})$. The set of solutions generated as Θ_o varies defines $r(\Theta)$, the boundary between the decision regions. Since the absolute minimum of the likelihood ratio occurs at $\mathbf{x}_g = \mathbf{0}$, the origin of the sample space is always included in the decision region for

H_0 , if it is nonempty. Hence, the qualitative description of the likelihood ratio test for the sinusoidal signal generalizes to any signal with an even-symmetric probability density.

Unfortunately, the implementation difficulties of the likelihood ratio test for the sinusoidal signal generalize as well. The detector must cope with fluctuations in the size of the selected subset and uncertainties in the *a priori* signal model. To implement the likelihood ratio test in all cases using randomized data selection, the detector should be able to determine the decision regions for each realization of $\mathbf{G}[m]$. Additionally, for a fixed realization of the indicator random variables, determining these decision regions depends on the exact density $p_{\mathbf{x}_g|\mathbf{G},H}(\mathbf{x}_g|\mathbf{G}, H_1)$ and the desired false alarm probability \tilde{P}_F . If, as in the sinusoid detection model, there are unknown parameters in the signal model, or the signal is difficult to model *a priori*, the decision regions that achieve \tilde{P}_F and maximize P_D are difficult to determine.

D. Nonparametric Detector

Techniques from the theory of nonparametric and robust statistics have been applied to detector design in situations without precise *a priori* models [9], [19], [20]. Application of two such techniques can combat the challenges identified in the previous section. One technique introduced to cope with uncertainty in a signal model is invariance [21]. If a signal belongs to a class that is closed under some transformation, it is useful to design the detector so that its performance is also invariant to the transformation. The second challenge for the detector is the fluctuation in \mathbf{G} and the associated task of rapidly updating the decision regions to satisfy the constraint in (11). This constraint enforces a constant false alarm rate (CFAR) condition on the detector: random fluctuations in the size of the selected data subset do not cause changes in the false alarm rate.

In this section, we propose a low-complexity, nonparametric detector addressing the implementation challenges of the exact likelihood ratio test. The detector is invariant to rotation of the target signal probability density, and maintains the CFAR property.

The intuition behind our nonparametric detector arises from the sinusoidal signal example when $K = 1$. In this case, the 1-D sample space simplifies the decision region and leads to a UMP detector. Even though we cannot find a UMP detector for $K > 1$, we can determine a test that has a weaker universality property over the class of even-symmetric random vectors. In this case, it is possible to determine a scalar-valued function of $\mathbf{x}_K[m]$ so that there is a UMP test for the resulting random variable.

In this section, we use a detector based upon the scalar test statistic $r = \|\mathbf{x}_g\|$. Note that this statistic is rotationally invariant. We analyze the properties of the likelihood ratio for r , and show that this statistic leads to a nonparametric detector resembling the detector derived for $K = 1$ earlier.

Based upon knowledge of the Gaussian conditional density $p_{r|\mathbf{G},H}(r|\mathbf{G}, H_0)$, we design a threshold test

$$r \underset{\hat{H}=H_0}{\overset{\hat{H}=H_1}{\gtrless}} t \quad (21)$$

where t is chosen to achieve a desired \tilde{P}_F . We show that this test has the maximum possible P_D for any set of decision regions based on r that achieve the false alarm rate \tilde{P}_F . Additionally, we show a method to calculate t that accounts for fluctuations in the selected subset and maintains the CFAR property.

1) *Properties of the Likelihood Ratio for $r[m]$* : In order for the intuition based on the scalar detector to work, two conditions must be satisfied. First, the likelihood ratio test for $r[m]$ given $\mathbf{G}[m]$ must reduce to a threshold test described in (21). This condition guarantees that the operating point of the test maximizes P_D over all decision regions that achieve the desired \tilde{P}_F , regardless of the target signal's density. Second, there ought to be an easy way to determine the proper threshold based upon a desired false alarm probability \tilde{P}_F . This allows the detector to be implemented with the required false alarm rate in real time. This subsection generates exact results for both these conditions, and the following subsection describes an approximate technique to calculate the test threshold.

In order to justify the simple threshold detector structure, we first verify that the likelihood ratio test simplifies to the form shown in (21). The likelihood ratio for $r[m]$ given $\mathbf{G}[m]$ is

$$L(r|\mathbf{G}) = \frac{p_{r|\mathbf{G},H}(r|\mathbf{G}, H_1)}{p_{r|\mathbf{G},H}(r|\mathbf{G}, H_0)}. \quad (22)$$

Rather than calculate the ratio by direct computation of the densities $p_{r|\mathbf{G},H}(r|\mathbf{G}, H_0)$ and $p_{r|\mathbf{G},H}(r|\mathbf{G}, H_1)$, we can show the necessary result using the properties of $L(\mathbf{x}_g|\mathbf{G})$ established in the proof of Theorem 1. The following theorem, proved in Appendix II, establishes the validity of the threshold test on r .

Theorem 2: For the statistical model established in (12), and a fixed value of K , let $r = \|\mathbf{x}_g\|$. If the conditional densities $p_{\mathbf{x}_g|\mathbf{G},H}(\mathbf{x}_g|\mathbf{G}, H_i)$ for $i = 0, 1$ are continuous and positive, the likelihood ratio for $r = \|\mathbf{x}_g\|$ increases monotonically without bound.

In order to determine an appropriate CFAR detector, we require a rule for selecting a threshold t as a function of $\mathbf{G}[m]$ and \tilde{P}_F . For a fixed t , the false alarm probability is

$$\begin{aligned} P_F &= Pr\{r > t|\mathbf{G}, H_0\} \\ &= \int_{r>t} p_{r|\mathbf{G},H}(r|\mathbf{G}, H_0) dr \\ &= \int_{\|\mathbf{x}_g\|>t} p_{\mathbf{x}_g|\mathbf{G},H}(\mathbf{x}_g|\mathbf{G}, H_0) d\mathbf{x}_g. \end{aligned} \quad (23)$$

The density $p_{\mathbf{x}_g|\mathbf{G},H}(\mathbf{x}_g|\mathbf{G}, H_0)$ is a multivariate Gaussian, so the integral (23) can be reduced the complementary dis-

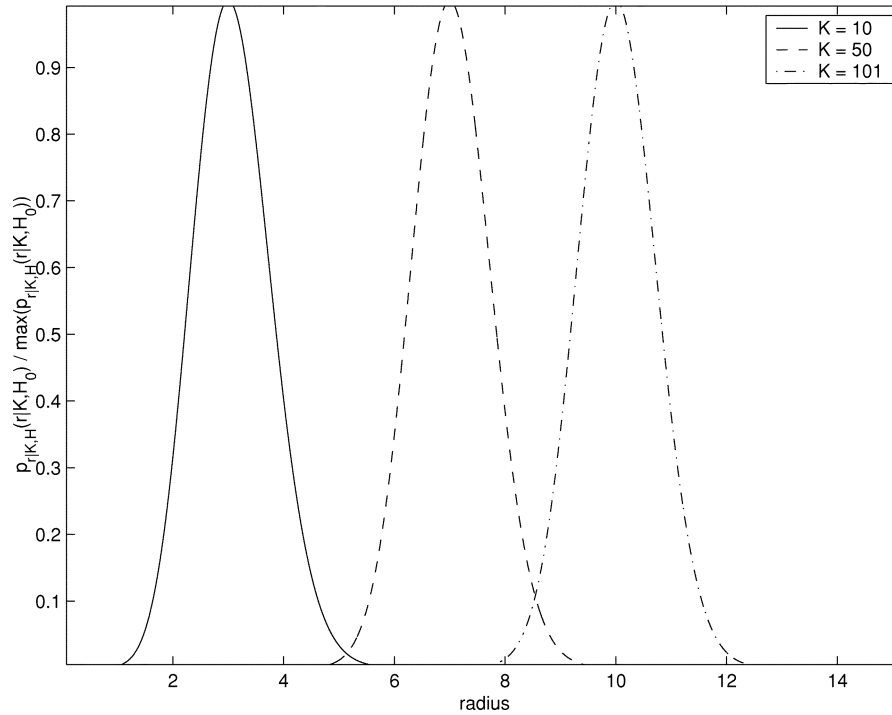


Fig. 3. Normalized radial densities for $K = 10, 50, \text{ and } 101$. Each radial density arises from a Gaussian density for $\mathbf{x}_g[m]$ with covariance matrix \mathbf{I} . The plotted curves are $((p_{r|K,H}(r|K, H_0))/(\max(p_{r|K,H}(r|K, H_0))))$. Each appears similar to a Gaussian with standard deviation near $(1/2)$.

tribution function (CDF) of a central χ^2 random variable of degree K [24]. In the case where $\sigma^2 = 1$, P_F is

$$P_F = \frac{1}{2^{\frac{K}{2}} - 1 \Gamma\left(\frac{K}{2}\right)} \int_t^\infty r^{K-1} e^{-\frac{r^2}{2}} dr. \quad (24)$$

This integral can be expressed using the incomplete gamma function denoted by $\Gamma(t^2/2, K/2)$. The false alarm rate in this situation is

$$P_F = 1 - \frac{\Gamma\left(\frac{t^2}{2}, \frac{K}{2}\right)}{\Gamma\left(0, \frac{K}{2}\right)}. \quad (25)$$

An exact determination of the test threshold that achieves a desired false alarm rate \tilde{P}_F requires inversion of (25). In situations where the false alarm rate is always fixed to a single value, a lookup table for t versus K may be appropriate. If the application scenario requires that \tilde{P}_f vary over time, the detector must be able to calculate the threshold numerically. Techniques to perform this calculation are described in [25].

2) *Low-Complexity Algorithm to Calculate the Detector Threshold:* In this subsection, we propose a threshold calculation based on an approximation of the conditional density for r under H_0 as a Gaussian with its parameters chosen as functions of K and \tilde{P}_F . The algorithm then uses the Gaussian approximation to calculate the threshold that would achieve \tilde{P}_F . This section argues for the plausibility of this approximation procedure, and evaluates its accuracy in determining t .

The algorithms for inverting (25) are iterative in nature. They may be computationally expensive to execute frequently. This section proposes an approximation appropriate for low false alarm rates that is based on the inverse of the Q function. This can be accomplished without iterative algorithms using a rational approximation [26], and is less expensive to compute.

As shown in (24), the conditional density for r under H_0 is proportional to the term $q_K(r) = r^{K-1} e^{-r^2/2}$. For large values of r , the exponential decay dominates the behavior of this function. For values of r near zero, however, the r^{K-1} term dominates. The overall behavior is that of a sharply peaked function cresting at $r_{\max} = \sqrt{K-1}$. Fig. 3 shows the normalized density for several values of K . Each peak resembles a Gaussian. Examining the logarithm of the function yields further insight on the resemblance. Taking the logarithm of $q_K(r)$ separates the terms in the function as

$$\log q_K(r) = (K-1) \log(r) - \frac{1}{2} r^2. \quad (26)$$

The function $\log q_K(r)$ is shown in Fig. 4. Since the $\log(r)$ term grows slowly for large r , the apparent drop of the function near r_{\max} resembles the quadratic $(1/2)r^2$. Specifically, the second derivative $q_K''(r_{\max}) = 2$ for all values of K . Consequently, the Gaussian with $\sigma = (1/2)$ approximates the behavior of $p_{r|H}(r|H_0)$ well near its peak at $\sqrt{K-1}$.

In order to estimate the density farther from the peak of $q_K(r)$, we extend the Gaussian approximation. Since the $\log(r)$ term in (26) varies slowly, it is possible to approximate $\log q_K(r)$ over several standard deviations by

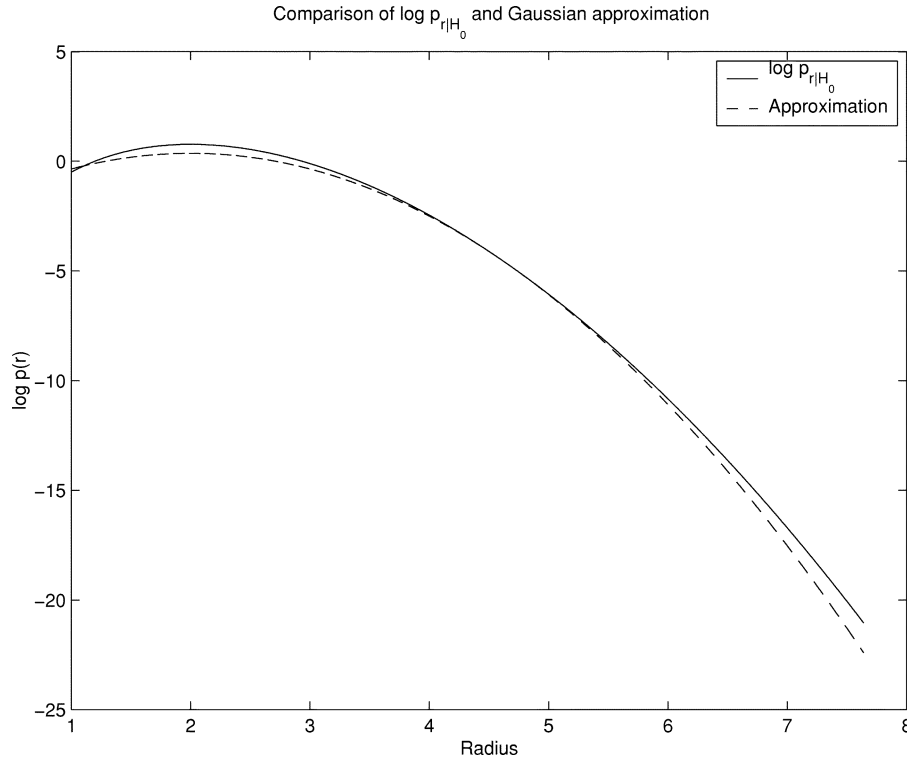


Fig. 4. Approximation of $\log a_K(r)$ by a parabola. The plot displays a case when $K = 5$. The parabolic approximation is designed to fit well enough to determine thresholds necessary to achieve $P_F = 10^{-4}$.

a quadratic with properly chosen peak and curvature. Essentially, the approximation of $\log q_K(r)$ resembles a Gaussian with its mean and variance adjusted to account for K and \tilde{P}_F . Fig. 4 shows an example of approximating $\log a_K(r)$ with a parabola given by $\log g_K(r) = D - ((r - \mu)^2)/(2\sigma_{\text{approx}}^2)$. For properly chosen values of m , D , and σ_{approx}^2 , the approximation can fit $\log a_K(r)$ closely over a particular interval. Since the conditional density decays quickly for large values of r , the interval where the approximation is accurate only needs to be a few standard deviations wide.

There are many possible ways to approximate $q_K(r)$ in order to determine an appropriate detector threshold. Our Gaussian approximation is designed to be accurate for small values of \tilde{P}_F , less than 10^{-1} . We suspect that this is a reasonable range of operation for many detectors, since false alarms will initiate subsequent processing, expending power and communication resources. Detectors designed to operate efficiently would typically avoid a high false alarm rate.

In order to have a low false alarm rate, the detector threshold should at least be greater than $\sqrt{K-1}$, the peak location of $\log q_k(r)$. Our approximation is based upon determining a parabola that accurately fits $\log q_K(r)$ over an interval of r sufficiently large to suggest that the exact P_F is near \tilde{P}_F .

The Gaussian approximation has three free parameters, the mean m , variance σ^2 , and the amplitude D . In our approximate fitting procedure, we assign $\mu = r_{\text{max}}$, which guarantees that the peak of the approximate density coincides with the peak of $p_{r|\mathbf{G},H}(r|\mathbf{G},H_0)$. Additionally, we will choose σ_{approx}^2 and D so that the approximation inter-

sects the true conditional density in two locations, denoted r_a and r_b . These fit points are chosen so that the interval $[r_a, r_b]$ is likely to contain the value of t that produces \tilde{P}_F .

The values r_a and r_b are determined by making an educated guess of the true threshold, and centering the fit points around it. The curvature of $\log q_K(r)$ at its peak is the same as a Gaussian with standard deviation $\sigma_{\text{init}} = 1/2$. Thus, we generate an initial guess for the threshold using a crude Gaussian approximation. The initial guess is

$$t_{\text{init}} = \mu + \sigma_{\text{init}} Q^{-1}(\tilde{P}_F). \quad (27)$$

Given this guess, the fit points are chosen with

$$\begin{aligned} r_a &= t_{\text{init}} - 0.01 \\ r_b &= t_{\text{init}} + 0.01. \end{aligned} \quad (28)$$

The choice of r_a and r_b is *ad hoc*. In general, however, it provides a close fit between $\log q_K(r)$ and the approximation on an interval extending several standard deviations past the true threshold. Since the true density is dominated by the $\exp(-r^2/2)$ term past its peak, the false alarm probability (24) is concentrated in the first several standard deviations past the threshold. The approximation does not fit accurately far from t_{init} , however, this will not have a significant impact on the approximation accuracy for thresholds near r_a or r_b .

The parameters of $\log g_K(r)$ depend on the fit points and K . The fitting error between $\log q_K(r)$ and $\log g_K(r)$ is

$$E(r) = D - \frac{(r - \mu)^2}{2\sigma_{\text{approx}}^2} - (K - 1) \log(r) + \frac{r^2}{2}. \quad (29)$$

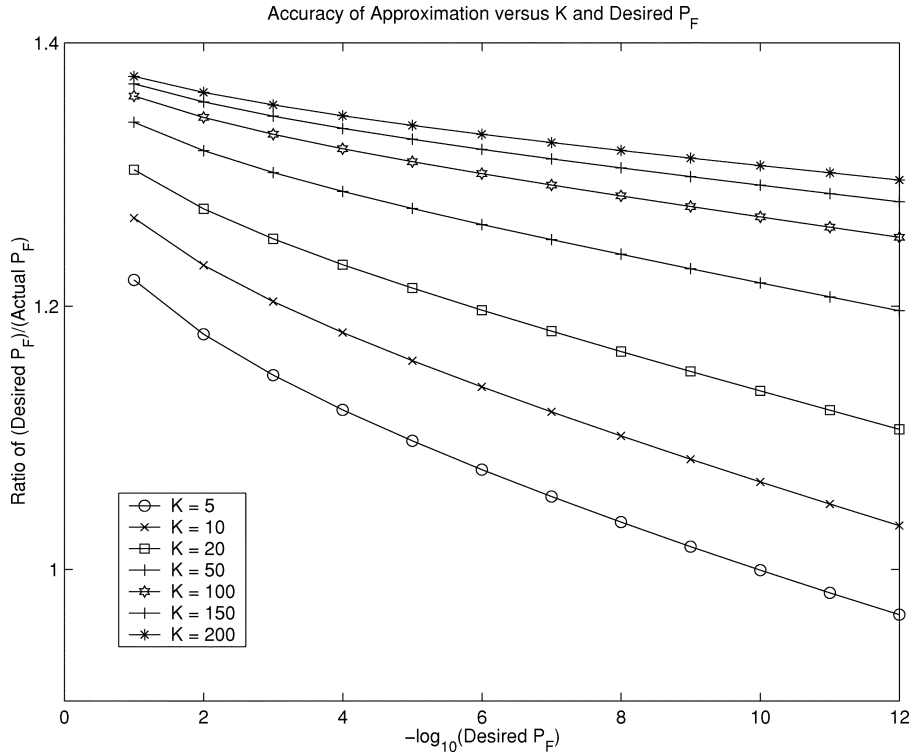


Fig. 5. Approximation accuracy. The ratio of the actual false alarm rate, P_F , to the desired value \tilde{P}_F is shown for $K = 5, 10, 20, 50, 100, 150,$ and 200 . The desired false alarm rate varies from $\tilde{P}_F = 10^{-2}$ to $\tilde{P}_F = 10^{-12}$.

The parameters are determined by requiring $m = \sqrt{K - 1}$, $E(r_a) = 0$, and $E(r_b) = 0$. Solving for A and σ^2 yields

$$\sigma_{\text{approx}}^2 = \frac{r_b + r_a - 2\mu}{r_b + r_a - 2\frac{\mu^2}{r_b - r_a} \log\left(\frac{r_b}{r_a}\right)}$$

$$D = \mu^2 \log(r_a) - \frac{r_a^2}{2} + \frac{(r_a - \mu)^2}{2\sigma_{\text{approx}}^2}. \quad (30)$$

The coefficients produce the approximation $\log g_K(r) = D - (((r - \mu)^2)/(2\sigma_{\text{approx}}^2))$. In practice, this function fits $\log a_K(r)$ well over a range of r near t_{init} . Fig. 4 shows an example of the curves for $K = 5$ and $\tilde{P}_F = 10^{-4}$.

The approximation gives a technique for determining the threshold to achieve a small \tilde{P}_F for a given K . Adjusting for the constant factors in the integrals (24), the threshold is

$$t = \mu + \sigma_{\text{approx}} Q^{-1} \left(\tilde{P}_F \frac{2^{\frac{K}{2}-1} \Gamma\left(\frac{K}{2}\right)}{e^D} \right). \quad (31)$$

In order to verify the approximation accuracy, we compare the actual P_F versus \tilde{P}_F over a range of exponentially spaced values from 10^{-2} to 10^{-12} . Each \tilde{P}_F is one-tenth the previous value. Fig. 5 shows the behavior of the approximation for several values of K . In general, the approximation appears to produce P_F values slightly smaller than \tilde{P}_F . The error increases with K . In general, the approximation is accurate to within a factor of 1.5 for desired false alarm rates between 0.01 and 10^{-12} and K between 5 and 200.

The actual performance of the threshold detector for the sinusoid detection problem described in Section III-B is shown

in Fig. 6. The figure shows the performance of the detector for several values of K . In all cases the noise variance is $\sigma^2 = 1$, and the wave's amplitude is $A = 1$. The simulations shown in Fig. 6 indicate that the performance of the detection algorithm improves as K or the ratio A/σ increase. These results are not particularly surprising, since increasing K gives the detector more raw data, and increasing A/σ improves the signal-to-noise ratio of each individual measurement. The results are useful primarily as a low-complexity baseline for the performance of detection schemes based upon more restrictive statistical models or data selection algorithms based on the details of such models.

IV. CONCLUSION

In this paper, we propose randomized data selection as a technique to cope with limited communication and computation resources in distributed networks. To illustrate the impact of randomized selection on signal processing performed by the network, we focus on a binary hypothesis testing problem: detection of a random vector with an even-symmetric probability density in white, Gaussian noise.

The detection problem shows several issues inherent in randomized selection. First, understanding the impact of randomized selection on the signal statistics is required to determine the new detector structure. Our assumptions about the network's communication protocols allow us to design our detectors based upon the conditional densities of the signals. Second, the fluctuation in the size of the selected sensor subset presents implementation challenges for the detector. It needs to perform the likelihood ratio test for a large number

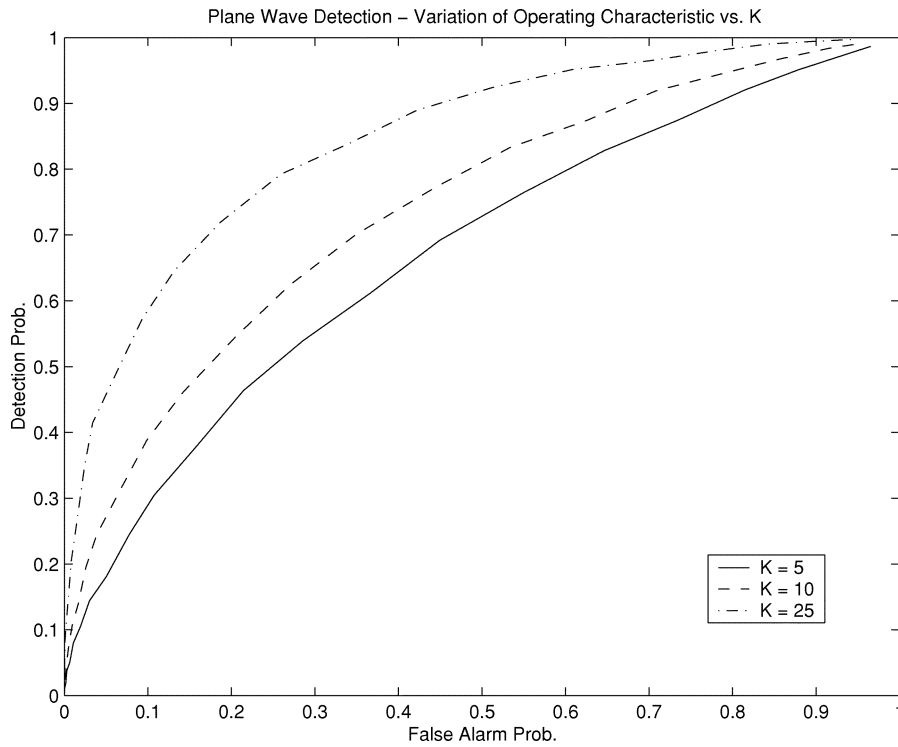


Fig. 6. The receiver operating characteristic of the example problem is shown for $K = 5, 10,$ and 25 .

of potential subsets. Finally, we consider the impact of uncertainties in the *a priori* model for the signal's probability density on the detector. This issue was independent of randomized selection, but it describes the potential applications of sensor networks. They may often be used in situations where the coefficients of the signal's density are not known in advance.

In order to deal with the randomized selection environment, we analyze a nonparametric detector. In lieu of a complex detector based upon a detailed statistical model for the target signal, we examine a threshold detector on the magnitude of $\mathbf{x}_g[m]$. For any even-symmetric target signature in white Gaussian noise, the detector's (P_F, P_D) operating point always achieves the maximum P_D for any set of rotationally invariant decision regions that produce a false alarm rate equal to P_F . In order to integrate the approach with randomized selection, we show an approximate technique to calculate the threshold from the desired false alarm rate \tilde{P}_F and K .

The proposed detection algorithm rests upon the statistical model for the noise. In practice, the noise may not obey a Gaussian density. The algorithm, however, serves as a template for designing detectors for use with random sensor selection. Detectors assuming different noise statistics will likely require similar approximations for the decision thresholds. Overall, the algorithm serves as a low-complexity baseline for evaluating the performance of sensor selection and detection algorithms in sensor networks.

Future work on distributed signal processing and sensor selection can take many directions. The analysis of random sensor selection can be expanded to improve the performance

baseline it provides. The selection algorithm can be modified to account for selection over multiple time slots or selection in the presence of colored noise. Both cases introduce an additional degree of freedom to the basic approach analyzed here. Additionally, exploring the interaction between selection and routing or distributed source coding algorithms may provide another interesting technique for balancing signal processing performance with communication cost.

Finally, the randomized selection approach may be useful in practical implementations of sensor networks. The issues of robustness, complexity, and energy efficiency are significant in this new environment, and randomized selection provides a way to balance these performance criteria. The randomized selection approach suggested in this paper has several implications for real systems. First, it gives a performance baseline to algorithms based upon more detailed state information and statistical models. Second, it may provide a desirable operating point in the tradeoff between robustness, complexity, and performance. Random selection requires no extra communication to collect the network state, and is robust to a wide class of signal models. While it will likely produce performance inferior to algorithms based upon detailed models, it may be more robust to modeling errors, and may strike a desirable balance between performance and complexity. Third, randomized selection is compatible with the cluster-based distributed signal processing techniques proposed for sensor networks. In networks where nodes cooperate in local clusters, random selection can be applied in each cluster. Finally, the selection algorithm depends only upon a single parameter. While we focused on single-shot processing

in this paper, these detection algorithms will be used over time. The selection procedure must balance the information quality in an individual time slot with the longevity of the sensor network. Since the randomized selection algorithms depend on a single parameter, it may be possible to illustrate the tradeoff between performance and network lifetime in an appealing way through further analysis of this class of sensor selection algorithms.

APPENDIX I PROOF OF THEOREM 1

Theorem 1: Consider a detection problem of the class defined in (7). Let $r = \|\mathbf{x}\|$ and $\Theta = [\theta \ \phi_3 \ \dots \ \phi_K]^T$. Then, the likelihood ratio $L(r, \Theta|\mathbf{G})$ given by (19) increases monotonically without bound.

Proof: Using the spherical coordinate system defined in (20), a vector $\mathbf{x} \in \mathcal{R}^N$ can be written

$$\mathbf{x} = r\mathbf{v}_\Theta \quad (32)$$

where \mathbf{v}_Θ is a unit vector defined by the angular variables. In terms of these coordinates, the conditional likelihood ratio is

$$L(r, \Theta|\mathbf{G}) = \int_{\mathbf{a}_K} p_{\mathbf{s}_g|\mathbf{G}}(\mathbf{a}_K|\mathbf{G}) e^{(-\frac{1}{2\sigma^2} \mathbf{a}_K^T \mathbf{a}_K)} e^{\left(r \frac{\mathbf{v}_\Theta^T \mathbf{a}_K}{\sigma^2}\right)} d\mathbf{a}_K. \quad (33)$$

Expressed in spherical coordinates, the derivative is

$$\frac{dL}{dr} = \int_{\mathbf{v}_\Theta^T \mathbf{a}_K \geq 0} p_{\mathbf{s}_g|\mathbf{G}}(\mathbf{a}_K|\mathbf{G}) e^{(-\frac{1}{2\sigma^2} \mathbf{a}_K^T \mathbf{a}_K)} * \left(\frac{\mathbf{v}_\Theta^T \mathbf{a}_K}{\sigma^2}\right) \left(2 \sinh\left(r \frac{\mathbf{v}_\Theta^T \mathbf{a}_K}{\sigma^2}\right)\right) d\mathbf{a}_K. \quad (34)$$

Each term in the integrand is positive over the region where $\mathbf{v}_\Theta^T \mathbf{a}_K \geq 0$; thus, the integrand is positive for every value of $r \geq 0$, and the likelihood ratio is always increasing. Similar analysis shows that the second derivative of the likelihood ratio is always positive.

Since $((dL)/(dr)) > 0$ and $((d^2L)/(dr^2)) > 0$ for all positive values of r , straightforward calculus shows that $L(r, \mathbf{v}_\Theta|\mathbf{G})$ satisfies the inequality

$$L(r, \Theta|\mathbf{G}) > L(r_o, \Theta|\mathbf{G}) + L'(r_o, \Theta|\mathbf{G})(r - r_o) \quad (35)$$

for every Θ and every pair $r \geq r_o \geq 0$. Thus, there is no upper bound on $L(r, \Theta|\mathbf{G})$; the likelihood ratio diverges along the direction identified by the unit vector \mathbf{v}_Θ . \square

APPENDIX II PROOF OF THEOREM 2

Theorem 2: For the statistical model established in (12), and a fixed value of K , let $r = \|\mathbf{x}_g\|$. If the conditional densities $p_{\mathbf{x}_g|\mathbf{G}, H}(\mathbf{x}_g|\mathbf{G}, H_i)$ for $i = 0, 1$ are continuous and positive, the likelihood ratio for $r = \|\mathbf{x}_g\|$ increases monotonically without bound.

Proof: For the random variable $r[m]$, the likelihood ratio is given by

$$L(r|\mathbf{G}) = \frac{p_{r|\mathbf{G}, H}(r|\mathbf{G}, H_1)}{p_{r|\mathbf{G}, H}(r|\mathbf{G}, H_0)}. \quad (36)$$

We shall derive a monotonically increasing lower bound for $p_{r|\mathbf{G}, H}(r|\mathbf{G}, H_1)$ in terms of the likelihood ratio for $\mathbf{x}_g[m]$.

In order to make the argument precise, let

$$f(r) = \min_{\|\mathbf{x}_g\|=r} L(\mathbf{x}_g|\mathbf{G}). \quad (37)$$

Under the assumption that this is a well-defined function, the conditional densities for \mathbf{x} satisfy

$$p_{\mathbf{x}_g|\mathbf{G}, H}(\mathbf{x}_g|\mathbf{G}, H_1) \geq p_{\mathbf{x}_g|\mathbf{G}, H}(\mathbf{x}_g|\mathbf{G}, H_0) f(r) \quad (38)$$

whenever $\|\mathbf{x}_g\| = r$. Thus, the density $p_{r|\mathbf{G}, H}(r|\mathbf{G}, H_1)$ satisfies

$$p_{r|\mathbf{G}, H}(r|\mathbf{G}, H_1) \geq B_K r^{K-1} \exp\left(\frac{-r^2}{2\sigma^2}\right) f(r). \quad (39)$$

This inequality shows that the likelihood ratio for the random variable r is bounded below by $f(r)$, i.e. $L(r|\mathbf{G}) \geq B_K f(r)$. In the remainder of this section, we will prove the existence and continuity of $f(r)$ and demonstrate that $f(r)$ increases monotonically without bound.

Consider the minimization from (37) in spherical coordinates. The domain for the angular variables is a compact set; each ϕ_i lies in the interval $0 \leq \phi_i < \pi$, and θ lines in the interval $0 \leq \theta < 2\pi$. With fixed r , $L(r, \Theta|\mathbf{G})$ is a positive, bounded function over a compact set, so the minimum exists [27].

The results of Theorem 1 show that $f(r)$ is an increasing function. For any $r \geq 0$ and $\epsilon > 0$, consider the value of

$$f(r + \epsilon) = \min_{\|\mathbf{x}_g\|=r+\epsilon} L(\mathbf{x}_g|\mathbf{G}). \quad (40)$$

Theorem 1 guarantees that $L((r + \epsilon), \Theta) > L(r, \Theta)$ for any fixed Θ . Additionally, the likelihood ratio satisfies $L(r, \Theta) \geq f(r)$ as a consequence of the definition of f . Combining these inequalities yields

$$L((r + \epsilon), \Theta|\mathbf{G}) > f(r) \text{ for all } \Theta, \text{ if } \epsilon > 0. \quad (41)$$

Since there is a strict inequality between $L((r + \epsilon), \Theta)$ and $f(r)$, $f(r + \epsilon)$ must satisfy

$$f(r + \epsilon) > f(r) \text{ for all } \epsilon > 0. \quad (42)$$

Thus, as a consequence of Theorem 1, the bound function $f(r)$ is strictly monotonically increasing.

To prove that $f(r)$ is continuous, we must verify that for every $\epsilon > 0$, we can find $\delta > 0$ such that

$$|f(r + \delta) - f(r)| < \epsilon. \quad (43)$$

The likelihood bound satisfies $f(r) = L(r, \Theta|\mathbf{G})$ for some unit vector Θ due to its definition as a minimization. Thus, using the fact that $L(r, \Theta|\mathbf{G})$ is an increasing function of r , we can determine the inequality

$$f(r + \delta) = \min_{\|\mathbf{x}_g\|=r+\delta} L(\mathbf{x}_g|\mathbf{G}) < L((r + \delta), \Theta_o|\mathbf{G}) \quad (44)$$

for any fixed $r > 0$ and $\delta > 0$. Using this inequality produces the following property of $f(r)$

$$|f(r + \delta) - f(r)| < |L((r + \delta), \Theta_o | \mathbf{G}) - L(r, \Theta_o | \mathbf{G})|. \quad (45)$$

Since $L(r, \Theta_o | \mathbf{G})$ is the ratio of two positive, continuous functions of r , it is also a continuous function. Thus, for any value of $\epsilon > 0$, we can find a value of δ such that $|L((r + \delta), \Theta_o | \mathbf{G}) - L(r, \Theta_o | \mathbf{G})| < \epsilon$. This property, along with the inequality (45) establish that $f(r)$ is a continuous function of r .

Finally, we can establish that $f(r)$ has a lower bound similar to (35). In the proof of Theorem 1, we showed that, with fixed Θ , $(dL)/(dr) > 0$ for any $r > 0$ and any Θ . Thus, we can guarantee that

$$d(r) = \min_{\|x_g\|=r} \frac{dL}{dr} \quad (46)$$

is a well-defined function because we are minimizing a function bounded from below over a compact set. Thus, for every Θ and every pair $r \geq r_o > 0$, the inequality

$$L(r, \Theta | \mathbf{G}) \geq L(r_o, \Theta | \mathbf{G}) + d(r_o)(r - r_o) \quad (47)$$

holds. Minimizing both sides of the expression guarantees that $f(r)$ satisfies

$$f(r) \geq f(r_o) + d(r_o)(r - r_o). \quad (48)$$

The function $f(r)$ increases at least linearly, so it is not bounded. Since $L(r | \mathbf{G}) \geq B_K f(r)$, the theorem is proven. \square

ACKNOWLEDGMENT

The authors would like to thank S. Blatt and C. Rohrs for a series of productive discussions, and B. Hajek, K. Ramchandran, and J. Reich for some stimulating suggestions.

REFERENCES

- [1] S. Kumar, F. Zhao, and D. Shepherd, "Collaborative signal and information processing in microsensor networks," *IEEE Signal Processing Mag.*, vol. 19, pp. 13–14, Mar. 2002.
- [2] J. Chen, K. Yao, and R. E. Hudson, "Source localization and beamforming," *IEEE Signal Processing Mag.*, vol. 19, pp. 30–39, Mar. 2002.
- [3] D. Li, K. Wong, Y. Hu, and A. Sayeed, "Detection, classification, and tracking of targets," *IEEE Signal Processing Mag.*, vol. 19, pp. 17–29, Mar. 2002.
- [4] S. Blatt, "Distributed sensor fusion for sensor networks," in *Proc. 4th Annu. Conf. Information Fusion*, 2001, pp. TuC3–TuC8.
- [5] V. Raghunathan, C. Schrugers, S. Park, and M. Srivastava, "Energy-aware wireless microsensor networks," *IEEE Signal Processing Mag.*, vol. 19, pp. 40–50, Mar. 2002.

- [6] F. Zhao, J. Shin, and J. Reich, "Information-driven dynamic sensor collaboration," *IEEE Signal Processing Mag.*, vol. 19, pp. 61–72, Mar. 2002.
- [7] A. P. Chandrakasan, R. Amirtharajah, S. Cho, J. Goodman, G. Konhuri, J. Kulik, W. Rabiner, and A. Wang, "Design considerations for distributed microsensor systems," in *Proc. CICC 1999*, pp. 279–286.
- [8] D. Estrin and R. Govindan, "Next century challenges: scalable coordination in sensor networks," in *Proc. Mobicom 1999*, pp. 263–270.
- [9] J. B. Thomas, "Nonparametric detection," *Proc. IEEE*, vol. 58, pp. 623–631, May 1970.
- [10] S. Pradhan, J. Kusuma, and K. Ramchandran, "Distributed compression in a dense microsensor network," *IEEE Signal Processing Mag.*, vol. 19, pp. 51–60, Mar. 2002.
- [11] S. C. Draper, "Successive structuring of source coding algorithms for data fusion, buffering, and distribution in networks," Ph.D. dissertation, M.I.T., Cambridge, MA, 2002.
- [12] D. Mittera, F. Palmieri, and S. Haykin, "Efficient sparse FIR filter design," in *Proc. ICASSP 2002*, vol. 2, pp. 1537–1540.
- [13] A. J. Miller, *Subset Selection in Regression*. New York: Chapman & Hall, 1990.
- [14] D. Gore, R. Nabar, and A. Paulraj, "Selecting an optimal set of transmit antennas for a low rank matrix channel," in *Proc. ICASSP 2001*, pp. 2785–2788.
- [15] A. Molisch, M. Win, and J. Winters, "Capacity of MIMO systems with antenna selection," in *Proc. ICC 2001*, pp. 570–574.
- [16] M. Nafie, A. Tewfik, and M. Ali, "Deterministic and iterative solutions to subset selection problems," *IEEE Trans. Signal Processing*, vol. 50, pp. 1591–1601, July 2002.
- [17] M. R. Said and A. V. Oppenheim, "Discrete-time randomized sampling," in *Proc. 8th Annu. IEEE Conf. Electronics, Circuits, and Systems*, vol. 3, 2001, pp. 1407–1411.
- [18] R. Motwani and P. Raghavan, *Randomized Algorithms*. Cambridge, U.K.: Cambridge Univ. Press, 1995.
- [19] H. V. Poor, *An Introduction to Signal Detection and Estimation*, 2nd ed. New York: Springer-Verlag, 1994.
- [20] H. V. Poor and S. A. Kassam, "Robust techniques for signal processing: a survey," *Proc. IEEE*, vol. 73, pp. 431–481, Mar. 1985.
- [21] L. L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*. Reading, MA: Addison-Wesley, 1991.
- [22] S. C. A. Thomopolous and N. A. Okello, "Distributed detection with consulting sensors and communication cost," *IEEE Trans. Automat. Contr.*, vol. 37, pp. 1398–1405, Sept. 1992.
- [23] J. Wozencraft and I. Jacobs, *Principles of Communication Engineering*. New York: Wiley, 1965.
- [24] J. G. Proakis, *Digital Communications*, 3rd ed. New York: McGraw-Hill, 1995.
- [25] A. R. DiDonato and A. H. Morris Jr., "Computation of the incomplete gamma function ratios and their inverse," *ACM Trans. Math. Software*, vol. 12, no. 4, pp. 377–393, Dec. 1986.
- [26] W. J. Cody, "Rational Chebyshev approximations for the error function," *Math. Comput.*, vol. 23, pp. 631–637, 1969.
- [27] D. Bertsekas, *Nonlinear Programming*, 2nd ed. Belmont, MA: Athena Scientific, 1999.



communications systems.

Mr. Sestok is a Member of Sigma Pi Sigma and Phi Beta Kappa.

Charles K. Sestok (Student Member, IEEE) received the S.B. degree in physics and the S.M. degree in electrical engineering from the Massachusetts Institute of Technology (MIT), Cambridge, in 1997 and 1999, respectively. He is currently working toward the Ph.D. degree in the Digital Signal Processing Group at MIT. He received a National Science Foundation Graduate Research Fellowship in 1997.

His research interests include signal processing, detection theory, and digital commu-



Maya R. Said (Student Member, IEEE) received the S.B. degree in biology and the S.B. degree in electrical engineering and computer science (EECS) from the Massachusetts Institute of Technology (MIT), Cambridge, in 1998, and the S.M. degree in toxicology and the M.Eng. degree in EECS from MIT in 2000 and 2001, respectively. She is currently working toward the Ph.D. degree in EECS in the Digital Signal Processing Group at MIT.

Her research interests are in the general area of signal processing and molecular biology with an emphasis on interdisciplinary research between the two fields.

Ms. Said is a member of Eta Kappa Nu, Sigma Xi, Tau Beta Pi, Phi Beta Kappa, and the American Association for the Advancement of Science. She received a number of awards at MIT, including the Carlton E. Tucker award for teaching excellence, and the Department Head Recognition Award for contributions to the EECS Department. She currently holds a Merck Fellowship in bioinformatics.



Alan V. Oppenheim (Life Fellow, IEEE) received the S.B. and S.M. degrees in electrical engineering from the Massachusetts Institute of Technology (MIT), Cambridge in 1961, the Sc.D. degree in electrical engineering from MIT in 1964, and the honorary Ph.D. degree from Tel Aviv University, Tel Aviv, Israel, in 1995.

In 1964, Dr. Oppenheim joined the faculty at MIT, where he is currently Ford Professor of Engineering and a MacVicar Faculty Fellow. Since 1967, he has been affiliated with MIT Lincoln Laboratory and since 1977 with the Woods Hole Oceanographic Institution, Woods Hole, MA. He is coauthor of several widely used textbooks. He is also editor of several advanced books on signal processing. His research interests are in the general area of signal processing and its applications.

Dr. Oppenheim is a Member of the National Academy of Engineering and a Member of Sigma Xi and Eta Kappa Nu. He has been a Guggenheim Fellow and a Sackler Fellow at Tel Aviv University. He has also received a number of awards for outstanding research and teaching, including the IEEE Education Medal, the IEEE Centennial Award, the Society Award, the Technical Achievement Award and the Senior Award of the IEEE Society on Acoustics, Speech, and Signal Processing. He has also received a number of awards at MIT for excellence in teaching, including the Bose Award and the Everett Moore Baker Award.