Robust Detection with the Gap Metric

Chuang He, *Student Member, IEEE,* and José M. F. Moura, *Fellow, IEEE*

***Abstract*—In a multipath communication channel, the optimal receiver is matched to the maximum likelihood (ML) estimate of the multipath signal. In general, this leads to a computationally intensive multidimensional nonlinear optimization problem that is not feasible in most applications. In this paper, we develop a detection algorithm that avoids finding the ML estimates of the channel parameters while still achieving good performance. Our approach is based on a geometric interpretation of the multipath detection problem. The ML estimate of the multipath signal is the orthogonal projection of the received signal on a suitable signal subspace $\mathcal{S}$. We design a second subspace $\mathcal{G}$, which is the representation subspace, that is close to $\mathcal{S}$ but whose orthogonal projection is easily computed. The closeness is measured by the gap metric. The subspace $\mathcal{G}$ is designed by using wavelet analysis tools coupled with a reshaping algorithm in the Zak transform domain. We show examples where our approach significantly outperforms the conventional correlator receiver (CR) and other alternative suboptimal detectors.**

I. Introduction

THE DETECTION of multipath signals is a problem of major concern in many applications, such as wireless communications, sonar, and radar. By multipath, we mean that the signal to be detected consists of multiple returns of the same transmitted signal. In a multipath channel, the received signal $r(t)$ is described mathematically as

$$r(t) = s_m(t) + n(t) \qquad 0 \le t \le T \tag{1}$$

$$= \sum_{k=1}^{K} \alpha_k s(\beta_k t - \tau_k) + n(t) \tag{2}$$

where

$s(t)$	transmitted signal,
$s_m(t)$	multipath noise-free signal,
K	number of paths,
α_k	attenuation factor for path k,
τ_k	time delay for path k,
β_k	Doppler shift factor for path k,
$n(t)$	additive noise.

In most practical situations of interest, the channel parameters $\{\alpha_k\}, \{\beta_k\}, \{\tau_k\}$, and K are unknown. For example, in wireless communications, surrounding obstacles create multiple unknown delayed replicas of the transmitted signal, and if the objects are also moving, which is the case in mobile wireless communications, unknown Doppler shifts occur, which further complicates the problem. In sonar and radar applications, time delays and Doppler shifts are due to a target's range and velocity, respectively. In sonar, the multipath signals result from surface and bottom reflections as well as refractive phenomena due to the nonhomogenous ocean media. In this paper, we will consider only the case where the Doppler shifts are assumed to be known; for simplicity, we take $\beta_k = 1, k = 1, \cdots, K$. In the future, we will extend the approach to the case where the Doppler is not known. Thus, in the following, we omit the Doppler shifts and focus on the K unknown delays $\tau_1, \tau_2, \cdots, \tau_K$.

When the transmitted signal $s(t)$ is deterministic known, the traditional detector is the correlator receiver (CR). It correlates the received signal $r(t)$ with the transmitted signal $s(t)$ and uses the peaks in the correlator output to estimate and detect the multipath signal. The advantage of this method is that it is simple and easy to implement. If different returns of the transmitted signal are separated in time by more than the effective duration of the signal autocorrelation function, the CR is practically optimal. Unfortunately, this condition is not satisfied in many practical situations. When the condition is not satisfied, the CR is not optimal in any sense.

On the other hand, the optimal generalized likelihood ratio test (GLRT) receiver requires the maximum likelihood (ML) estimates of the channel parameters. In general, this involves a computationally intensive multidimensional nonlinear optimization because the multipath signal is a nonlinear function of the channel delays. The dimension of the original parameter space is $2K+1$. If K is large or unknown, the optimal GLRT receiver is out of reach.

The goal of this paper is to develop a simple receiver that avoids finding the ML estimates of the channel parameters while still achieving good performance. Our approach is based on a geometric interpretation of the multipath signal detection problem. We see that the collection of all the possible multipath noise-free signals forms a linear subspace: the signal subspace $\mathcal{S}$. If we assume that the noise is white Gaussian and $s(t)$ is known, then the generalized log likelihood ratio function is the norm square of the orthogonal projection of the received signal $r(t)$ on $\mathcal{S}$. However, finding the orthogonal projection on the original signal subspace $\mathcal{S}$ involves the multidimensional nonlinear optimization over all the possible $\{\tau_k\}$ as well as the number of paths K. Our approach is to find an alternative subspace—the representation subspace $\mathcal{G}$ [1]–[4]—whose orthogonal projection is easily computed and approximates the original signal subspace $\mathcal{S}$ well. Once we have designed the representation subspace $\mathcal{G}$, the detector is practically as simple as the CR.

There are two major questions. The first is how to decide whether the subspace $\mathcal{G}$ is close to or a good approximation of $\mathcal{S}$. We propose to use the gap metric [5] to measure the

Manuscript received February 6, 1996; revised July 7, 1996. This work was supported in part by DARPA through Grants ONR N00014-91-J183, AFOSR F49620-96-1-0436, and by ONR N00014-97-0040. The associate editor coordinating the review of this paper and approving it for publication was Dr. Hagit Messer-Yaron.

The authors are with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213 USA.

Publisher Item Identifier S 1053-587X(97)04220-7.

closeness between $\mathcal{S}$ and $\mathcal{G}$. The gap metric is a measure of the distance between two closed subspaces. In other words, it tells us how two subspaces match each other. The gap metric is also related to a generalized ambiguity function; see [6]. With the gap metric, the problem is reduced to finding the representation subspace $\mathcal{G}$ that minimizes the gap between $\mathcal{S}$ and $\mathcal{G}$.

The second question is concerned with designing the subspace $\mathcal{G}$ so that the orthogonal projection on $\mathcal{G}$ is easily computed. Since we are studying the effects of translations, it is natural to turn to the wavelet transform that represents signals in terms of translates and dilations of a mother wavelet. Wavelet transforms have led to the theory of multiresolution analysis where the Hilbert space of square integrable functions is decomposed into a nested sequence of subspaces. These subspaces are used to approximate the original signal subspace $\mathcal{S}$. Using these subspaces as the representation subspace has several advantages. First, the subspaces in the multiresolution analysis are generated by the mother scaling functions and mother wavelets. This reduces the design of the representation subspace to the design of a single function, which is a much easier task. Second, nice parameterizations of the orthonormal compactly supported wavelets and scaling functions are available [7], [8] to facilitate the design of the subspace $\mathcal{G}$. Finally, the orthogonal projection on the multiresolution subspaces is easily computed by taking inner products.

The outline of the paper is as follows. In Section II, we discuss the geometric interpretation of the multipath signal detection problem; we introduce the definition and important properties of the gap metric. In Section III, we describe in detail the design of the representation subspace $\mathcal{G}$. In Section IV, we give the structure of the new minimum gap receiver (MGR) as well as the structure of some other alternative receivers for comparison. In Section V, we present experimental results to demonstrate the performance of our new receiver and compare it with the alternative receivers described in Section IV. Finally, in Section VI, we conclude the paper.

II. Geometric Interpretation

In Section II-A, we discuss a geometric interpretation of multipath signal detection. In Section II-B, we introduce the gap metric to be used throughout the paper. In Section II-C, we give a brief review of the compactly supported orthonormal scaling functions and wavelets.

A. Geometric Interpretation of Multipath Signal Detection

Many problems of interest to the signal processing community are cast in the general framework of detecting waveforms with unknown parameters, i.e., from an observation of the form

$$H_1: r(t) = (\mathcal{H}_\theta[s])(t) + n(t) \qquad 0 \le t \le T \tag{3}$$

$$H_0: r(t) = n(t) \tag{4}$$

to decide which hypothesis is true. We use here the notation in [6]. The received signal is $r(t)$, $s(t)$ is the transmitted signal, $\mathcal{H}_\theta[\cdot]$ is the channel operator parameterized by the vector $\theta \in \Theta$, and $n(t)$ is the additive noise. In other words, θ is the vector of channel parameters. With the multipath example of (2), θ is the vector $[\alpha_1, \alpha_2, \cdots, \alpha_K; \tau_1, \tau_2, \cdots, \tau_K; K]$, and

$$(\mathcal{H}_\theta[s])(t) = s_m(t) = \sum_{k=1}^{K} \alpha_k s(t - \tau_k). \tag{5}$$

We assume that the transmitted signal $s(t)$ is deterministic known, $n(t)$ is zero-mean white Gaussian, and $\theta \in \Theta$ is unknown deterministic.

Before proceeding, we clarify our notation. As in most engineering literature, we will use the notation $s(t), g(t), \cdots$, to represent a function of time or the value of the function at time t, depending on the context. In places where that may cause confusion, we will use instead $s(\cdot), g(\cdot), \cdots$, to represent explicitly the functions.

Denote by $\mathcal{S}$ the signal subspace

$$\mathcal{S} = \{(\mathcal{H}_\theta[s])(\cdot), \theta \in \Theta\}. \tag{6}$$

The GLRT is given by

$$\|P_{\mathcal{S}}[r]\|^2 \underset{H_0}{\overset{H_1}{\gtrless}} \eta_{\text{GLRT}}. \tag{7}$$

In (7), $\|\cdot\|$ is the L_2 norm, η_{GLRT} is a threshold, and

$$(P_{\mathcal{S}}[r])(t) = (\mathcal{H}_{\theta^*}[s])(t) \tag{8}$$

where

$$\theta^* = \arg \inf_{\theta \in \Theta} \|r(\cdot) - (\mathcal{H}_\theta[s])(\cdot)\| \tag{9}$$

is the vector that contains the ML estimates of the channel parameters, and $P_{\mathcal{S}}[r]$ is the orthogonal projection of $r(\cdot)$ on the signal subspace $\mathcal{S}$. Generally, computing the ML estimate θ^* is complicated because it involves a multidimensional nonlinear optimization.

In the multipath channel case, the signal subspace $\mathcal{S}$ is

$$\mathcal{S} = \left\{ s_m(\cdot) = \sum_{k=1}^{K} \alpha_k s(\cdot - \tau_k), \alpha_k, \tau_k \in \mathbb{R}, K \in \mathbb{Z}^+ \right\} \tag{10}$$

where $\mathbb{Z}^+$ is the set of natural numbers.

To compute the GLRT statistic, we need to solve (9), i.e., we need to optimize over all the channel parameters. Since the multipath noise-free signal $s_m(\cdot)$ is a nonlinear function of all the channel delays $\{\tau_k\}$, for a fixed K, solving (9) requires a nonlinear optimization over a K-dimensional parameter space [9]. For large K, this is not feasible from a practical point of view. Furthermore, if K is not known, we have to iterate on the value of K, which further increases the computational complexity.

Our approach is motivated by the fact that the GLRT statistic is only a function of the orthogonal projection of $r(\cdot)$ on $\mathcal{S}$. In other words, the goal is to find this orthogonal projection, or a good approximation of it. The receiver becomes simple if we can find a representation subspace $\mathcal{G}$ that approximates $\mathcal{S}$ well and whose orthogonal projection is easily computed. Then, the orthogonal projection on $\mathcal{G}$ is a good approximation to the orthogonal projection on $\mathcal{S}$. We propose to use the gap $\delta(\mathcal{S}, \mathcal{G})$ [5] as a measure of the closeness between $\mathcal{S}$ and $\mathcal{G}$. The

problem is then to find a linear subspace $\mathcal{G}$ that minimizes the gap metric $\delta(\mathcal{S},\mathcal{G})$. Once we have designed the representation subspace $\mathcal{G}$, we use $\|P_{\mathcal{G}}[r]\|^2$ as the test statistic instead of $\|P_{\mathcal{S}}[r]\|^2$. The choice of $\mathcal{G}$ will be discussed in Section III.

B. Gap Metric

The gap metric [5] is a measure of the distance between two closed subspaces. In other words, it tells us how two closed subspaces match each other. Two equivalent definitions are given in [5] for the gap between two closed subspaces in a Hilbert space. One is given in terms of the orthogonal projection operators on the subspaces, whereas the second definition, which is more mathematically involved, provides, in general, an analytic formula for the gap. We will start with the first definition because it is more intuitive and leads us to a natural interpretation of the problem that we are solving. Then, we will exploit the second definition in detail because we use this definition to calculate the gap.

1) First Definition: Given two closed subspaces $\mathcal{S}$ and $\mathcal{G}$ in a Hilbert space $\mathcal{H}$, we denote by $P_{\mathcal{S}}$ and $P_{\mathcal{G}}$ the orthogonal projection operators on $\mathcal{S}$ and $\mathcal{G}$, respectively. Then, the gap between $\mathcal{S}$ and $\mathcal{G}$ is

$$\delta(\mathcal{S},\mathcal{G}) = \|P_{\mathcal{S}} - P_{\mathcal{G}}\| \tag{11}$$

where $\|\cdot\|$ is the L_2-induced operator norm.

As we mentioned earlier in the last subsection, our goal is to find a representation subspace $\mathcal{G}$ such that the orthogonal projection of the received signal $r(\cdot)$ on $\mathcal{G}$ is close to the orthogonal projection of $r(\cdot)$ on the signal subspace $\mathcal{S}$. From (11), we can achieve this goal by designing $\mathcal{G}$ to be close to $\mathcal{S}$ in the gap sense. If the gap is small, then $P_{\mathcal{S}}$ is close to $P_{\mathcal{G}}$ in the L_2-induced operator norm sense, and we can approximate $P_{\mathcal{S}}[r]$ by $P_{\mathcal{G}}[r]$. This definition matches with our intuition. Unfortunately, calculating the gap using this definition is difficult because it requires computing the orthogonal projection operator on the signal subspace $\mathcal{S}$. That is exactly what we are trying to avoid. This difficulty leads to the second definition.

2) Second Definition: We denote by $S_{\mathcal{S}}$ the unit sphere of $\mathcal{S}$ (the set of all $u \in \mathcal{S}$ with $\|u\| = 1$), and let

$$\hat{\delta}(\mathcal{S},\mathcal{G}) = \sup_{u\in S_{\mathcal{S}}} \operatorname{dist}(u,\mathcal{G}) \tag{12}$$

where

$$\operatorname{dist}(u,\mathcal{G}) = \inf_{v\in\mathcal{G}} \|u-v\| = \sqrt{\|u\|^2 - \|P_{\mathcal{G}}(u)\|^2} \tag{13}$$

is the distance between the vector u and the subspace $\mathcal{G}$, which is equal to the distance between u and the orthogonal projection $P_{\mathcal{G}}(u)$ of u on $\mathcal{G}$. Likewise, we define $\hat{\delta}(\mathcal{G},\mathcal{S})$. The quantity

$$\delta(\mathcal{S},\mathcal{G}) = \max(\hat{\delta}(\mathcal{S},\mathcal{G}), \hat{\delta}(\mathcal{G},\mathcal{S})) \tag{14}$$

is called the gap between $\mathcal{S}$ and $\mathcal{G}$.

To ease the discussion in Section III, we call $\hat{\delta}(\mathcal{S},\mathcal{G})$ the left side of the gap and $\hat{\delta}(\mathcal{G},\mathcal{S})$ the right side of the gap, where $\mathcal{S}$ is the signal subspace, and $\mathcal{G}$ is the representation subspace. Fig. 1 illustrates the definition of the left side of the gap. From the definition and the figure, we see that to calculate $\hat{\delta}(\mathcal{S},\mathcal{G})$, we first pick a vector u on the unit sphere $S_{\mathcal{S}}$ of $\mathcal{S}$ and project u on $\mathcal{G}$ to get the distance $\operatorname{dist}(u,\mathcal{G})$ between u and $\mathcal{G}$. Once we have obtained this distance for all the vectors on the unit sphere of $\mathcal{S}$, we calculate $\hat{\delta}(\mathcal{S},\mathcal{G})$ by taking the supremum over all the vectors on the unit sphere of $\mathcal{S}$. The right side of the gap $\hat{\delta}(\mathcal{G},\mathcal{S})$ is defined in a similar manner with $\mathcal{S}$ and $\mathcal{G}$ switched. Thus, the gap metric is essentially a worst-case distance measure between two closed subspaces. The following relations follow directly from the second definition of the gap [5].

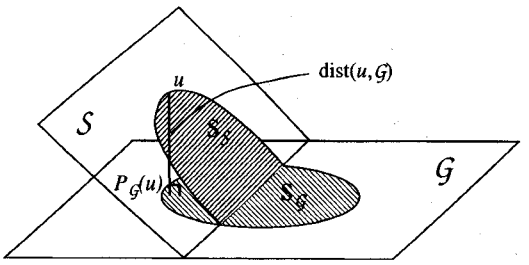


Fig. 1. One side of the gap.

i) $\hat{\delta}(\mathcal{S},\mathcal{G}) = 0$ if and only if $\mathcal{S} \subseteq \mathcal{G}$.
ii) $\delta(\mathcal{S},\mathcal{G}) = 0$ if and only if $\mathcal{S} = \mathcal{G}$.
iii) $\delta(\mathcal{S},\mathcal{G}) = \delta(\mathcal{G},\mathcal{S})$.
iv) $0 \le \delta(\mathcal{S},\mathcal{G}) \le 1$.
v) $\delta(\mathcal{S},\mathcal{G}) \le \delta(\mathcal{S},\mathcal{U}) + \delta(\mathcal{U},\mathcal{G})$.
vi) $\delta(\mathcal{S},\mathcal{G}) = 1$ or $\hat{\delta}(\mathcal{S},\mathcal{G}) = \hat{\delta}(\mathcal{G},\mathcal{S})$.

Property i) says that if the left side of the gap $\hat{\delta}(\mathcal{S},\mathcal{G})$ is zero, then $\mathcal{S}$ is included in $\mathcal{G}$; however, this does not imply that $\hat{\delta}(\mathcal{G},\mathcal{S})$ is also zero, i.e., that $\mathcal{G}$ is included in $\mathcal{S}$. Therefore, only one side of the gap metric is not enough to measure the similarity between two closed subspaces. Properties ii)–v) guarantee that the gap metric is indeed a distance measure for closed linear subspaces in a Hilbert space. Property iv) says that the gap is bounded by 1.

Property vi) is important because the gap, in general, requires that we compute both $\hat{\delta}(\mathcal{S},\mathcal{G})$ and $\hat{\delta}(\mathcal{G},\mathcal{S})$. However, in most cases, only one of them is easy to compute. What Property vi) tells us is that once we have computed $\hat{\delta}(\mathcal{S},\mathcal{G})$, we know immediately that $\hat{\delta}(\mathcal{G},\mathcal{S})$ is either 1 or equal to $\hat{\delta}(\mathcal{S},\mathcal{G})$, or vice versa. If we can guarantee in some way that both $\hat{\delta}(\mathcal{S},\mathcal{G})$ and $\hat{\delta}(\mathcal{G},\mathcal{S})$ are not 1, then we need only to compute one of them. This property avoids computing the orthogonal projection $P_{\mathcal{S}}[r]$ of $r(\cdot)$ on the signal subspace $\mathcal{S}$, which helps us solve the problem that we have with the first definition.

C. Compactly Supported Orthonormal Scaling Functions

In our approach, we use the subspaces spanned by integer shifts of the compactly supported orthonormal scaling functions to approximate the signal subspace. The advantage is that they are nicely parameterized. Thus, the subspace optimization is reduced to parameter optimization.

A scaling function satisfies the two-scale dilation equation [10]:

$$\phi(t) = \sum_k c_k \phi(2t-k). \tag{15}$$

The coefficients $\{c_k\}$ are called scaling coefficients. A scaling function is completely characterized by its scaling coefficients. Once the scaling coefficients $\{c_k\}$ are given, $\phi(t)$ can be constructed using an iterative procedure [10]. The wavelet corresponding to $\phi(t)$ is then defined [10] as

$$\psi(t) = \sum_k d_k \phi(2t - k) \tag{16}$$

where the coefficients $\{d_k\}$ are called wavelet coefficients.

Daubechies [7] provides a complete characterization of the power spectrum of the sequences of scaling coefficients that lead to compactly supported orthonormal scaling functions and wavelets. Zou and Tewfik [8] provide a further parameterization of these coefficients. They show that for a given support $2M-1$, there is an infinite number of scaling functions and mother wavelets. They also show that all scaling functions and wavelets of support $2M-1$ are parameterizable, or designable, by choosing just $M-1$ parameters $(\zeta_1, \zeta_2, \cdots, \zeta_{M-1})$ over $[0, 2\pi]^{M-1}$.

Specifically, denote by $H(z)$ and $G(z)$ the z transforms of the sequences $\{c_k\}$ and $\{d_k\}$; it can be shown as in [8] that

$$\begin{bmatrix} H(z) \\ z^{2(M-1)}G(z) \end{bmatrix} = \sqrt{2}E(z^2)\begin{bmatrix} 1 \\ z \end{bmatrix} \tag{17}$$

where the matrix $E(z)$ is called the polyphase matrix given by

$$\begin{aligned} E(z) &= V_{M-1}(z)V_{M-2}(z)\cdots V_1(z)V_0 \\ V_0 &= \begin{bmatrix} \cos\left(\frac{3\pi}{4}\right) & -\sin\left(\frac{3\pi}{4}\right) \\ \sin\left(\frac{3\pi}{4}\right) & \cos\left(\frac{3\pi}{4}\right) \end{bmatrix} \\ V_k(z) &= I + (z-1)v_k v_k^T, \qquad 1 \le k \le M-1. \end{aligned} \tag{18}$$

The vector v_k is given by

$$v_k = \begin{bmatrix} \cos(\zeta_k) \\ \sin(\zeta_k) \end{bmatrix} \qquad 1 \le k \le M-1. \tag{19}$$

III. Subspace Design

In this section, we discuss the computation and optimization of the gap metric in detail.

A. Problem Statement

As we mentioned early in Section I, in a multipath channel, the received signal $r(t)$ is described mathematically as

$$r(t) = \sum_{k=1}^{K} \alpha_k s(t - \tau_k) + n(t). \tag{20}$$

We assume that $\{\alpha_k\}, \{\tau_k\}$, and K are deterministic unknown. The signal subspace $\mathcal{S}$ is then given by (10). Our goal is to find a closed subspace $\mathcal{G}^*$ such that the gap $\delta(\mathcal{S}, \mathcal{G})$ between $\mathcal{S}$ and $\mathcal{G}$ is minimized.

Since each element in $\mathcal{S}$ is a linear combination of different delayed replicas of the transmitted signal $s(t)$, it is intuitive to design the representation subspace in a similar way. We use

$$\mathcal{G} = \left\{ g_{\text{int}}(\cdot) = \sum_{n=-\infty}^{+\infty} \beta_n g(\cdot - n), \beta_n \in \mathbb{R} \right\} \tag{21}$$

to approximate $\mathcal{S}$. Because of the structure of $\mathcal{G}$ given by (21), designing $\mathcal{G}$ is equivalent to designing its *generating* function $g(\cdot)$. We choose the function $g(\cdot)$, which is the generating function of the representation subspace, to be a compactly supported orthonormal scaling function. We will discuss in detail the reason for this choice in the following subsections.

B. Algorithm

Computing and minimizing the gap $\delta(\mathcal{S}, \mathcal{G})$ directly is not an easy task. For example, from the definition, we have the left side of the gap between $\mathcal{S}$ and $\mathcal{G}$

$$\hat{\delta}(\mathcal{S}, \mathcal{G}) = \sup_{\|s_m(\cdot)\|=1} \inf_{g_{\text{int}}(\cdot)\in\mathcal{G}} \|s_m(\cdot) - g_{\text{int}}(\cdot)\|. \tag{22}$$

Equation (22) says that in order to compute $\hat{\delta}(\mathcal{S}, \mathcal{G})$, we need to compute first the orthogonal projection of $s_m(\cdot)$ on $\mathcal{G}$ for a fixed $s_m(\cdot)$. This is accomplished by taking the infimum over all the $g_{\text{int}}(\cdot) \in \mathcal{G}$. Then, we need to maximize the distance between $s_m(\cdot)$ and its orthogonal projection on $\mathcal{G}$ subject to the constraint that $\|s_m(\cdot)\| = 1$. The first step is easy; the orthogonal projection of $s_m(\cdot)$ on $\mathcal{G}$ is obtained by calculating the inner products of $s_m(\cdot)$ with integer shifts of a biorthogonal function $\tilde{g}(\cdot)$ of $g(\cdot)$. Unfortunately, the second step is difficult. Since $s_m(\cdot)$ is a nonlinear function of $\{\tau_k\}$, maximizing this distance, i.e., taking the supremum, requires a multidimensional nonlinear optimization, which is precisely what we are trying to avoid. Computing the right side of the gap is even worse because getting the orthogonal projection on $\mathcal{S}$ involves, from the start, a multidimensional nonlinear optimization.

To circumvent these difficulties, we solve the problem in two steps:

1) Design the generating function $g(\cdot)$ to minimize the gap between $\mathcal{S}_{\text{int}}$ and $\mathcal{G}$.
2) Reshape the optimal $g^*(\cdot)$ obtained from step 1 to make it nearly shiftable.

In step 1, we restrict the delays $\{\tau_k\}$ to be integer valued rather than real valued. Instead of working with the subspace $\mathcal{S}$, we introduce the integer shift signal subspace $\mathcal{S}_{\text{int}}$ and solve the original problem with respect to $\mathcal{S}_{\text{int}}$ and not with respect to $\mathcal{S}$. We design the representation subspace $\mathcal{G}$ to be close to $\mathcal{S}_{\text{int}}$. In step 2, we modify the optimal $\mathcal{G}^*$ by reshaping its generating function $g^*(\cdot)$ to be as translation invariant, or shiftable, as possible so that the new representation subspace cannot only represent linear combinations of the integer shifts of $s(t)$ but linear combinations of its arbitrary real shifts as well. To do this, we use the reshaping algorithm of Benno and Moura; see [1]. In the following, we will explain these two steps in detail.

Step 1—Gap Minimization: The major reason for the complexity alluded to in the previous paragraphs is that in (22), we have to optimize over arbitrary real-valued $\{\tau_k\}$. We simplify the optimization by considering in this step the integer shifts only, i.e., rather than approximating the original subspace $\mathcal{S}$,

we approximate the subspace

$$\mathcal{S}_{\text{int}} = \left\{ s_{\text{int}}(\cdot) = \sum_{n=-\infty}^{+\infty} \alpha_n s(\cdot - n), \alpha_n \in \mathbb{R} \right\} \tag{23}$$

by the subspace $\mathcal{G}$ given in (21). Note that in $\mathcal{S}_{\text{int}}$, we consider only integer-shifted replicas of $s(t)$.

The problem has now been reduced to finding a closed subspace $\mathcal{G}^*$ such that the gap $\delta(\mathcal{S}_{\text{int}}, \mathcal{G})$ between $\mathcal{S}_{\text{int}}$ and $\mathcal{G}$ is minimized. In the following, we assume that $\{s(\cdot - n), n \in \mathbb{Z}\}$ is a Riesz basis [10] for $\mathcal{S}_{\text{int}}$, i.e., $\exists\ A, B$ such that $0 < A \leq B < \infty$ and

$$A \leq \sum_l |\mathcal{F}_s(f + l)|^2 \leq B \quad \text{a.e.} \tag{24}$$

where $\mathcal{F}_s(\cdot)$ is the Fourier transform of $s(\cdot)$.

We first need an explicit formula for the gap $\delta(\mathcal{S}_{\text{int}}, \mathcal{G})$ between $\mathcal{S}_{\text{int}}$ and $\mathcal{G}$. Theorem 1 provides this.

Theorem 1: Let

$$\mathcal{S}_{\text{int}} = \left\{ \sum_n \alpha_n s(\cdot - n), \alpha_n \in \mathbb{R} \right\}$$

$$\mathcal{G} = \left\{ \sum_n \beta_n g(\cdot - n), \beta_n \in \mathbb{R} \right\}$$

be two closed linear subspaces, where $\{s(\cdot - n), n \in \mathbb{Z}\}$ and $\{g(\cdot - n), n \in \mathbb{Z}\}$ satisfy the Riesz basis condition in (24). Then, the gap between $\mathcal{S}_{\text{int}}$ and $\mathcal{G}$ is given by

$$\begin{aligned} &\delta(\mathcal{S}_{\text{int}}, \mathcal{G}) \\ &\quad = \sqrt{1 - \inf_{f \in [0,1)} C_{sg}(f)} \\ &\quad = \sqrt{1 - \inf_{f \in [0,1)} \frac{\left| \sum_k \mathcal{F}_s(f + k) \overline{\mathcal{F}}_g(f + k) \right|^2}{\sum_l |\mathcal{F}_s(f + l)|^2 \sum_l |\mathcal{F}_g(f + l)|^2}} \end{aligned} \tag{25}$$

where $\mathcal{F}_s(\cdot)$ and $\mathcal{F}_g(\cdot)$ are the Fourier transforms of $s(\cdot)$ and $g(\cdot)$, respectively, and $\overline{\mathcal{F}}_g(f)$ is the complex conjugate of $\mathcal{F}_g(f)$. The infimum is taken over the regions where $C_{sg}(f)$ is continuous.

$C_{sg}(f)$ is essentially the periodization of the normalized magnitude squared cross-spectral density of the signals $s(t)$ and $g(t)$. The function $\Sigma_k\ \mathcal{F}_s(f + k)\overline{\mathcal{F}}_g(f + k)$ is the discrete time Fourier transform (DTFT) of the sampled cross-correlation function between $s(t)$ and $g(t)$, i.e., the DTFT of the discrete sequence

$$c_{sg}(n) = \langle s(\cdot), g(\cdot - n) \rangle \qquad n \in \mathbb{Z}. \tag{26}$$

The functions $\Sigma_l |\mathcal{F}_s(f + l)|^2$ and $\Sigma_l |\mathcal{F}_g(f + l)|^2$ are the DTFT of the sampled autocorrelation functions of $s(t)$ and $g(t)$, respectively.

The autocorrelation function of a deterministic energy signal $s(t)$ [11] is defined as

$$c_{ss}(t) = \int s(\tau) \overline{s}(t - \tau)\, d\tau. \tag{27}$$

The cross correlation between two deterministic energy signals $s(t)$ and $g(t)$ [11] is defined as

$$c_{sg}(t) = \int s(\tau) \overline{g}(t - \tau)\, d\tau. \tag{28}$$

By the Schwarz inequality, the second term in (25) is less than or equal to one, which makes $\delta(\mathcal{S}_{\text{int}}, \mathcal{G})$ satisfy Property iv) of the gap metric, i.e.,

$$0 \leq \delta(\mathcal{S}_{\text{int}}, \mathcal{G}) \leq 1. \tag{29}$$

As we know, the integral of the normalized magnitude squared cross-spectrum over frequency is an indication of the average similarity between $s(t)$ and $g(t)$. However, it is not enough to measure the similarity between the subspaces $\mathcal{S}_{\text{int}}$ and $\mathcal{G}$, which are spanned by integer shifts of $s(t)$ and $g(t)$, respectively. The infimum basically guarantees that it accounts for the worst case over frequency, thus providing a reasonable measure for the subspaces. This intuitively matches with our definition of the gap because the gap metric is a worst-case measure between the signals of two closed subspaces. Theorem 1 is proved in the Appendix. Theorem 1 can also be shown using the notation of angles between subspaces; see [12] for details.

We use Theorem 1 here to find the optimal subspace $\mathcal{G}^*$ that minimizes $\delta(\mathcal{S}_{\text{int}}, \mathcal{G})$, i.e.,

$$\mathcal{G}^* = \arg \min_{\mathcal{G}} \delta(\mathcal{S}_{\text{int}}, \mathcal{G}). \tag{30}$$

Equivalently, the generating function $g^*(\cdot)$ of $\mathcal{G}^*$ is

$$\begin{aligned} g^*(\cdot) &= \arg \max_{g(\cdot)} \inf_{f \in [0,1)} C_{sg}(f) \\ &= \arg \max_{g(\cdot)} \inf_{f \in [0,1)} \frac{\left| \sum_k \mathcal{F}_s(f + k) \overline{\mathcal{F}}_g(f + k) \right|^2}{\sum_l |\mathcal{F}_s(f + l)|^2 \sum_l |\mathcal{F}_g(f + l)|^2}. \end{aligned} \tag{31}$$

To perform the optimization over $g(\cdot)$ in (31), we restrict the search of the function $g(\cdot)$ to the set of compactly supported orthonormal scaling functions of a multiresolution analysis. There are three major reasons for doing this. First, it is clear that if $g(t) = s(t)$, then the gap in (25) is zero. However, $s(\cdot)$ is, in general, not shiftable; see (35) below. After reshaping $s(\cdot)$ using the algorithm in [1] to make it nearly shiftable, as explained in step 2 of our algorithm, we have observed that the subspace spanned by the integer shifts of the new reshaped signal does not approximate the subspace $\mathcal{S}_{\text{int}}$ well. On the other hand, from our simulations, the subspace spanned by the integer shifts of a reshaped compactly supported orthonormal scaling function does still approximate $\mathcal{S}_{\text{int}}$ well. Second, if we look at $\delta(\mathcal{S}_{\text{int}}, \mathcal{G})$ carefully, we notice that for a fixed value of $\delta(\mathcal{S}_{\text{int}}, \mathcal{G})$, there are many choices of $g(\cdot)$ because $\delta(\mathcal{S}_{\text{int}}, \mathcal{G})$ is only related to the minimum of $C_{sg}(f)$ over frequency f. We use this additional freedom to require $g(\cdot)$ to be a compactly supported orthonormal scaling function. Finally, compactly supported orthonormal scaling functions are nicely parameterized in terms of a vector ζ of parameters

[7], [8]. Using these parameterizations of the function $g(\cdot\colon \zeta)$, the optimization of $\delta(\mathcal{S}_{\text{int}}, \mathcal{G})$ is done by a search over the parameter space of ζ, i.e.,

$$\zeta^* = \arg\max_{\zeta} \inf_{f\in[0,1)} C_{sg}(f\colon \zeta) = \arg\max_{\zeta} \inf_{f\in[0,1)} \frac{\left|\sum_k \mathcal{F}_s(f+k)\overline{\mathcal{F}}_g(f+k\colon \zeta)\right|^2}{\sum_l |\mathcal{F}_s(f+l)|^2 \sum_l |\mathcal{F}_g(f+l\colon \zeta)|^2} \tag{32}$$

where $\zeta = [\zeta_1, \cdots, \zeta^{M-1}]^T$ is the vector of parameters defining the compactly supported orthonormal scaling functions, and $M-1$ is the number of unconstrained parameters [8]. The optimal scaling function is given by

$$g^*(t) = g(t\colon \zeta^*). \tag{33}$$

The corresponding representation subspace $\mathcal{G}^*$ approximates the integer shift signal subspace $\mathcal{S}_{\text{int}}$ well, i.e., it can represent linear combinations of integer shifts of the transmitted signal $s(t)$ well. In the next step, we consider representing linear combinations of arbitrary real shifts of $s(t)$.

Step 2—Reshaping: Once we have designed the closed subspace $\mathcal{G}^* = \{\Sigma_n\ \beta_n g^*(\cdot - n), \beta_n \in \mathbb{R}\}$ that minimizes $\delta(\mathcal{S}_{\text{int}}, \mathcal{G})$, we reshape $g^*(\cdot)$ to get a nearly shiftable function $g_r^*(\cdot)$. The corresponding representation subspace is

$$\mathcal{G}_r^* = \left\{ \sum_n \beta_n g_r^*(\cdot - n), \beta_n \in \mathbb{R} \right\}. \tag{34}$$

Hopefully, the new nearly shiftable $g_r^*(t)$ and its integer shifts cannot only represent well the integer shifts of $s(t)$ but also its arbitrary shifts.

A function $g(t)$ is shiftable if there exists a set of real coefficients $\{\beta_n(\tau), n \in \mathbb{Z}\}$ such that

$$g(t-\tau) = \sum_n \beta_n(\tau) g(t-n) \quad \forall \tau \in [0,1). \tag{35}$$

Equation (35) says that if a function $g(t)$ is shiftable, then any *arbitrary delayed* replica of the function $g(t-\tau)$ is well represented by a linear combination of *integer shifts* of the same function. A simple example of a shiftable function is the *sinc* function

$$\text{sinc}(t) = \frac{\sin(\pi t)}{\pi t}. \tag{36}$$

The *sinc* function is a bandlimitted signal with bandwith $\Delta f = 1$. It can be easily shown that this *sinc* function satisfies the shiftability condition. If the transmitted signal $s(t)$ is the *sinc* function, then we can just choose the generating function $g(t)$ as the transmitted signal itself. Thus, the gap between $\mathcal{S}_{\text{int}}$ and $\mathcal{G}$ is zero. In addition, because the *sinc* function is shiftable, the gap between $\mathcal{S}$ and $\mathcal{G}$ is zero. This means that linear combinations of the integer shifts of the *sinc* function represent exactly linear combinations of arbitrary real shifts of the *sinc* function itself.

However, in general, this is not the case, i.e., a function will not be shiftable. For example, it is not possible for a function to be shiftable and to have compact support [1], [13]. In [1], the authors relax the hard constraint of shiftability and consider instead the design of signals, for which in (35), they have only approximate equality. They design signals for which the mean square error in representing of these signals arbitrary delayed replicas by their integer shifted replicas

$$\epsilon(\tau) = \|g(\cdot - \tau) - \sum_n \beta_n(\tau) g(\cdot - n)\|^2 \tag{37}$$

is minimized. This translation error is a measure of the representation's *robustness* to continuous translations. A function is nearly shiftable if the maximum of this translation error is close to zero.

In [1], the authors show that the error given by (37) can be expressed in terms of the Zak transform [14]–[16] of the energy spectral density (ESD) of the function $g(t)$ as

$$\epsilon(\tau) = 1 - \int_0^1 \frac{|Z_{|\mathcal{F}_g|^2}(f,\tau)|^2}{Z_{|\mathcal{F}_g|^2}(f,0)}\, df \tag{38}$$

where $|\mathcal{F}_g(f)|^2$ is the energy density function of $g(t)$, and $Z_{|\mathcal{F}_g|^2}(f,\tau)$ its Zak transform.

If the function $g(t)$ is not shiftable, then the error $\epsilon(\tau)$ is nonzero, and it is a function of the delay τ. The error $\epsilon(\tau)$ can be reduced by modifying the term $Z_{|\mathcal{F}_g|^2}(f,\tau)$. Specifically, let

$$C(\tau) = \left[\int_0^1 \frac{|Z_{|\mathcal{F}_g|^2}(f,\tau)|^2}{Z_{|\mathcal{F}_g|^2}(f,0)}\, df \right]^{-1/2} \tag{39}$$

and reshape the Zak transform $Z_{|\mathcal{F}_g|^2}(f,\tau)$ as

$$Z_{\mathcal{F}_{\hat{g}}}(f,\tau) = C(\tau) Z_{|\mathcal{F}_g|^2}(f,\tau). \tag{40}$$

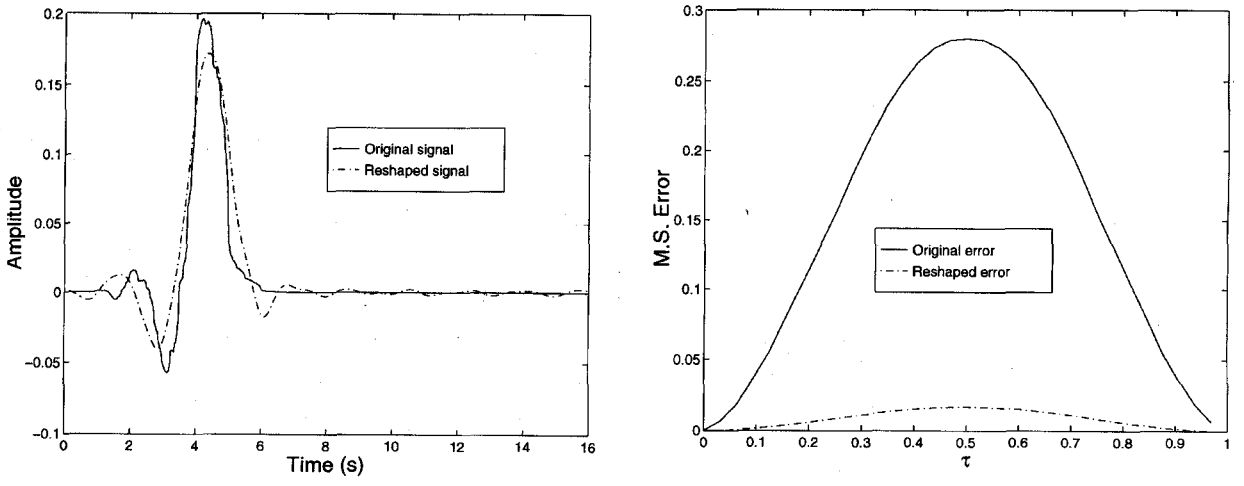
By choosing $C(\tau)$ as in (39), the representation error is zero if $\mathcal{F}_{\hat{g}}(f)$ is a valid ESD function. In general, that is not the case. However, it can be easily shown that $|\mathcal{F}_{\hat{g}}(f)|$ is indeed a valid ESD function. Therefore, in [1], the authors define in the frequency domain the reshaped signal $g_1(t)$ as

$$\mathcal{F}_{g_1}(f) = \sqrt{|\mathcal{F}\hat{g}(f)|} e^{j\theta_{\mathcal{F}_g}(f)} \tag{41}$$

where $\theta_{\mathcal{F}_g}(f)$ is the phase of $\mathcal{F}_g(f)$. The reshaped time domain signal $g_1(t)$ is obtained by taking the inverse Fourier transform of the expression in (41). The choice of the phase in (41) makes the reshaped signal closely resemble the original signal. Now, the representation error for $g_1(t)$, although reduced, is no longer zero. The error is further reduced by iterating the reshaping algorithm.

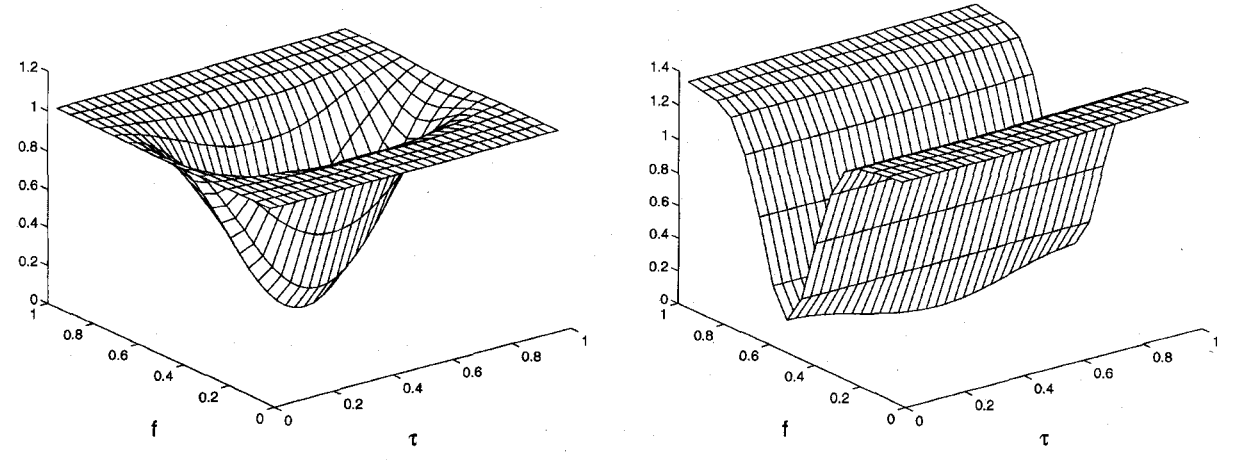
In the following, we illustrate the algorithm by reshaping a scaling function obtained using the parameterization described in Section II-C. Fig. 2(a) is a plot of the original signal and of the reshaped signal after 12 iterations of the reshaping algorithm. Fig. 2(b) shows the error given by (38) for the original and the reshaped signal. The maximum representation error is reduced from 0.275 to 0.014. Fig. 3(a) and (b) show the normalized Zak transforms of the ESD for the original signal and for the iterated reshaped signal, respectively. The normalized Zak transform of the ESD of the signal $g(t)$ is

$$\frac{|Z_{|\mathcal{F}_g|^2}(f,\tau)|^2}{Z_{|\mathcal{F}_g|^2}(f,0)}. \tag{42}$$



(a) (b)

Fig. 2. (a) Original signal and reshaped signal after 12 iterations. (b) Representation error for the original signal and for the 12-times iterated reshaped signal.



(a) (b)

Fig. 3. (a) Normalized Zak transform of the ESD of the original signal. (b) Normalized Zak transform of the ESD of the 12-times iterated reshaped signal.

The result of the reshaping algorithm is to make the normalized Zak transform more flat in the τ direction.

IV. Detection

In this section, we describe the structure of the new minimum gap receiver (MGR) and analyze its performance. We also discuss briefly the structure of other alternative receivers with which we compare our new receiver and their performance analysis. The alternative receivers that we consider are the unrealistic subspace receiver (USR), the energy detector (ED), the CR, and a suboptimal receiver called "matched filter with integer shifts" (MFIS).

A. Minimum Gap Receiver (MGR)

We are interested in the detection problem formulated in Section II-A. The GLRT is

$$L_{\text{GLRT}} = \|P_{\mathcal{S}}[r]\|^2 \underset{H_0}{\overset{H_1}{\gtrless}} \eta_{\text{GLRT}} \tag{43}$$

where $\mathcal{S}$ is the original signal subspace. We approximate the subspace $\mathcal{S}$ by the subspace $\mathcal{G}_r^*$ given by

$$\mathcal{G}_r^* = \left\{ \sum_n \beta_n g_r^*(\cdot - n), \beta_n \in \mathbb{R} \right\}. \tag{44}$$

The test statistic of the new MGR receiver is

$$L_{\text{MGR}} = \|P_{\mathcal{G}_r^*}[r]\|^2. \tag{45}$$

Since $g_r^*(t)$ is the reshaped scaling function, $\{g_r^*(\cdot - n), n \in \mathbb{Z}\}$ are not necessarily orthogonal to each other. We orthonormalize $\{g_r^*(\cdot - n), n \in \mathbb{Z}\}$ without changing the span of $\{g_r^*(\cdot - n), n \in \mathbb{Z}\}$ by

$$\mathcal{F}_{g^*_{r,\text{orth}}}(f) = \frac{\mathcal{F}_{g_r^*}(f)}{\sqrt{\sum_l |\mathcal{F}_{g_r^*}(f+l)|^2}} \tag{46}$$

where $\mathcal{F}_{g^*_{r,\text{orth}}}(\cdot)$, and $\mathcal{F}_{g_r^*}(\cdot)$ are the Fourier transforms of $g^*_{r,\text{orth}}(\cdot)$ and $g_r^*(\cdot)$, respectively, [10]. With the orthonormality

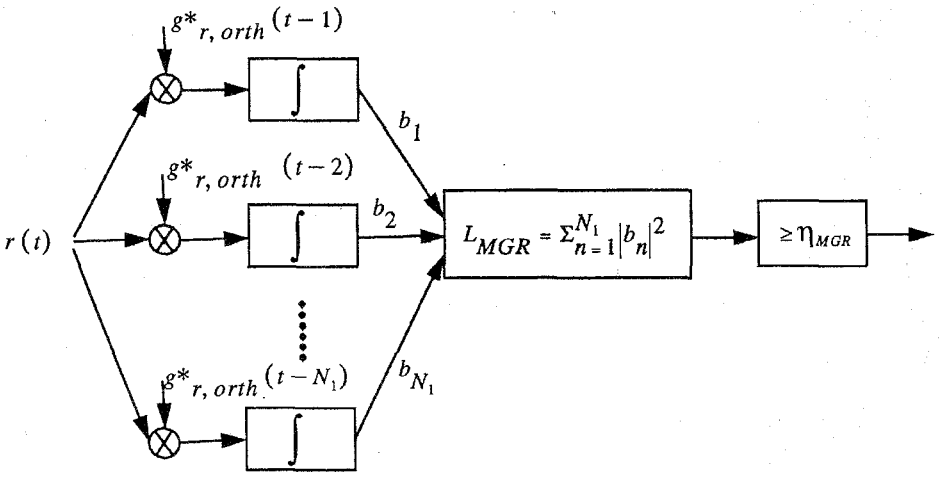


Fig. 4. Minimum gap receiver (MGR).

of $\{g^*_{r,\text{orth}} \cdot -n), n \in \mathbb{Z}\}$, $P_{\mathcal{G}^*_r}[r]$ is easily computed by

$$(P_{\mathcal{G}^*_r}[r])(t) = \sum_n b_n g^*_{r,\text{orth}}(t-n) = \sum_n \langle r(\cdot), g^*_{r,\text{orth}}(\cdot - n)\rangle \cdot g^*_{r,\text{orth}}(t-n) \quad (47)$$

where $\{b_n = \langle r(\cdot), g^*_{r,\text{orth}}(\cdot - n)\rangle, n \in \mathbb{Z}\}$ are the orthogonal projection coefficients. In general, the sum has to be taken from $-\infty$ to $+\infty$. Since, for all practical purposes, $r(t)$ and $g^*_{r,\text{orth}}(t)$ are time limited, we pick only those coefficients with significant value, and thus, we need only a finite sum. With proper alignment, we can take the sum from 1 to N_1.

Collecting the coefficients $\{b_n, n = 1, \cdots, N_1\}$ in the vector $B = [b_1, \cdots, b_{N_1}]^T$, we have the test

$$L_{\text{MGR}} = B^T B = \sum_{n=1}^{N_1} |b_n|^2 \underset{H_0}{\overset{H_1}{\gtrless}} \eta_{\text{MGR}}. \quad (48)$$

Under H_0, the test statistic L_{MGR} has a central chi-squared distribution with N_1 degrees of freedom. Under H_1, L_{MGR} has noncentral chi-squared distribution with N_1 degrees of freedom and noncentrality parameter

$$v = U^T U = \sum_{n=1}^{N_1} |u_n|^2$$

where

$$U = [u_1, \cdots, u_{N_1}]^T \quad \text{and}$$
$$u_n = \langle s_m(\cdot), g^*_{r,\text{orth}}(\cdot - n)\rangle, \qquad n = 1, \cdots, N_1.$$

Fig. 4 shows the structure of the MGR. The new receiver is essentially a bank of N_1 linear filters followed by an energy detector (ED). This structure is easy to implement, and no nonlinear optimization is involved. It avoids the ML estimation of the channel parameters.

B. Unrealistic Subspace Receiver (USR)

The USR assumes that the channel delays $\{\tau_k\}$ and the number of paths K are known. The attenuations $\{\alpha_k\}$ are not known. It is a K-dimensional subspace detector. The test statistic is the energy of the orthogonal projection of the received signal on the K-dimensional signal subspace. It is chi-squared distributed under H_0 and noncentral chi-squared distributed under H_1. In practice, the channel delays and the number of paths are not known; therefore, the ideal USR provides only an optimistic upper bound for the performance. The optimal GLRT receiver will degrade the performance of the USR.

C. Energy Detector (ED)

Suppose that the received signal is sampled at the Nyquist rate (T_N), i.e.,

$$r(n) = r(nT_N) \qquad n = 1, \cdots, N_s \quad (49)$$

where N_s is the number of samples, then the ED test statistic is given by

$$L_{ED} = \sum_{n=1}^{N_s} r(n)^2. \quad (50)$$

It is chi-squared distributed under H_0 and noncentral chi-squared distributed under H_1 with noncentrality

$$\sum_{n=1}^{N_s} s_m(nT_N)^2 \quad (51)$$

where $s_m(nT_N)$ is the sampled multipath noise free signal. This receiver is computationally simpler than the MGR receiver.

D. Correlator Receiver (CR)

The CR correlates the received signal $r(t)$ with the transmitted signal $s(t)$ and uses the peaks in the correlator output to form the test statistic. Mathematically, let

$$c_\tau = \langle r(\cdot), s(\cdot - \tau)\rangle \quad (52)$$

denote by $\tau_1^*, \cdots, \tau_K^*$ the K delays that maximize $|c_\tau|$, and let $C_\tau = [c_{\tau_1^*}, \cdots, c_{\tau_K^*}]^T$. Then, the test statistic is

$$L_{CR} = C_\tau^T \Lambda_\tau^{-1} C_\tau \quad (53)$$

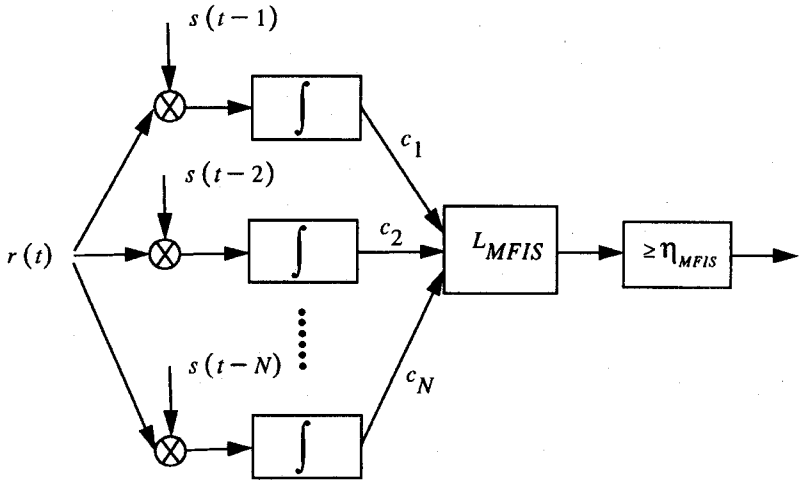


Fig. 5. Matched filter with integer shifts (MFIS) receiver.

where Λ_τ is the covariance matrix of C_τ. L_{CR} is chi-squared distributed under H_0 and noncentral chi-squared distributed under H_1. This receiver is computationally much more complex than the MGR receiver.

E. Matched Filter with Integer Shifts (MFIS)

Finally, we compare our approach with a simplification of the optimal GLRT receiver. The simplified GLRT receiver that we use [we call it matched filter with integer shifts (MFIS)], works in the following way: First, we take the inner product of $r(t)$ with $\{s(\cdot - n), n \in \mathbb{Z}\}$. Again, since in our examples they are time limited, we have a finite number of nonzero coefficients

$$c_n = \langle r(\cdot), s(\cdot - n)\rangle \qquad n = 1, \cdots, N. \tag{54}$$

In practice, we pick only those coefficients with significant energy. Denote as N_1 the number of "significant" coefficients and $N_2 = N - N_1$. Without loss of generality, we assume that we pick the first N_1 of them. Let

$$C = [c_1, \cdots, c_N]^T, \quad C_1 = [c_1, \cdots, c_{N_1}]^T \quad \text{and}$$
$$C_2 = [c_{N_1+1}, \cdots, c_N]^T.$$

The test statistic of the MFIS receiver is

$$L_{\text{MFIS}} = C^T \Lambda^{-1} C - C_2^T \Lambda_{22}^{-1} C_2 \tag{55}$$

where

$$\Lambda = \begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{bmatrix} \tag{56}$$

is the covariance matrix of C, and Λ_{22} is the covariance matrix of C_2.

Fig. 5 depicts the structure of the MFIS receiver. It is very similar to the MGR receiver. The difference is that the MFIS receiver matches with $\{s(\cdot - n)\}$, whereas the MGR receiver matches with $\{g^*_{r,\text{orth}}(\cdot - n)\}$. Under H_0, L_{MFIS} is central chi-squared distributed with N_1 degrees of freedom. Under H_1, L_{MFIS} is noncentral chi-squared distributed with N_1 degrees of freedom and noncentrality parameter [17],

$$U_1^T(\Lambda_{11} - \Lambda_{12}\Lambda_{22}^{-1}\Lambda_{21})U_1. \tag{57}$$

The vector U_1 is given by

$$U_1 = [u_1, \cdots, u_{N_1}]^T \tag{58}$$

where

$$u_n = \langle s_m(\cdot), s(\cdot - n)\rangle, \qquad n = 1, \cdots, N_1.$$

This receiver has a computational cost similar to the MGR receiver.

V. Experimental Results

In this section, we present numerical results to demonstrate the performance of the MGR and compare it with the alternative receivers that we describe in Section IV.

A. Parameter Setup

Again, the general formula for the multipath noise free signal is given by

$$s_m(t) = \sum_{k=1}^{K} \alpha_k s(t - \tau_k). \tag{59}$$

In our examples, the number of paths K is set to 15. To test the performance, we normalize the energy of $s_m(t)$ so that the absolute values of α_k do not affect the result; only their relative values matter. For simplicity, we set them all to be equal to 1. The delays $\{\tau_k, k = 1, \cdots, K\}$ are generated by a random number generator as uniformly distributed.

The transmitted signal $s(t)$ is a modulated chirp pulse

$$s(t) = \text{rect}\left(\frac{t}{T_c}\right)\cos(2\pi f_c t + 2\pi\gamma t^2) \tag{60}$$

where $\text{rect}(t)$ is the rectangular function with support $[0, 1]$, $T_c = 25\mu$s is the duration of the chirp pulse, $f_c = 16.8$ MHz is the carrier frequency, and $\gamma = 0.64 \times 10^4$ MHz/s is the chirp rate. These numbers are scaled typical values used in radar; see [18]. We sample equation (60) so that only the in phase component is nonzero. The signal $s(t)$ is shown in Fig. 6(a). Fig. 6(b) shows a multipath noise-free signal $s_m(t)$ with the parameters as we described in the previous paragraph.

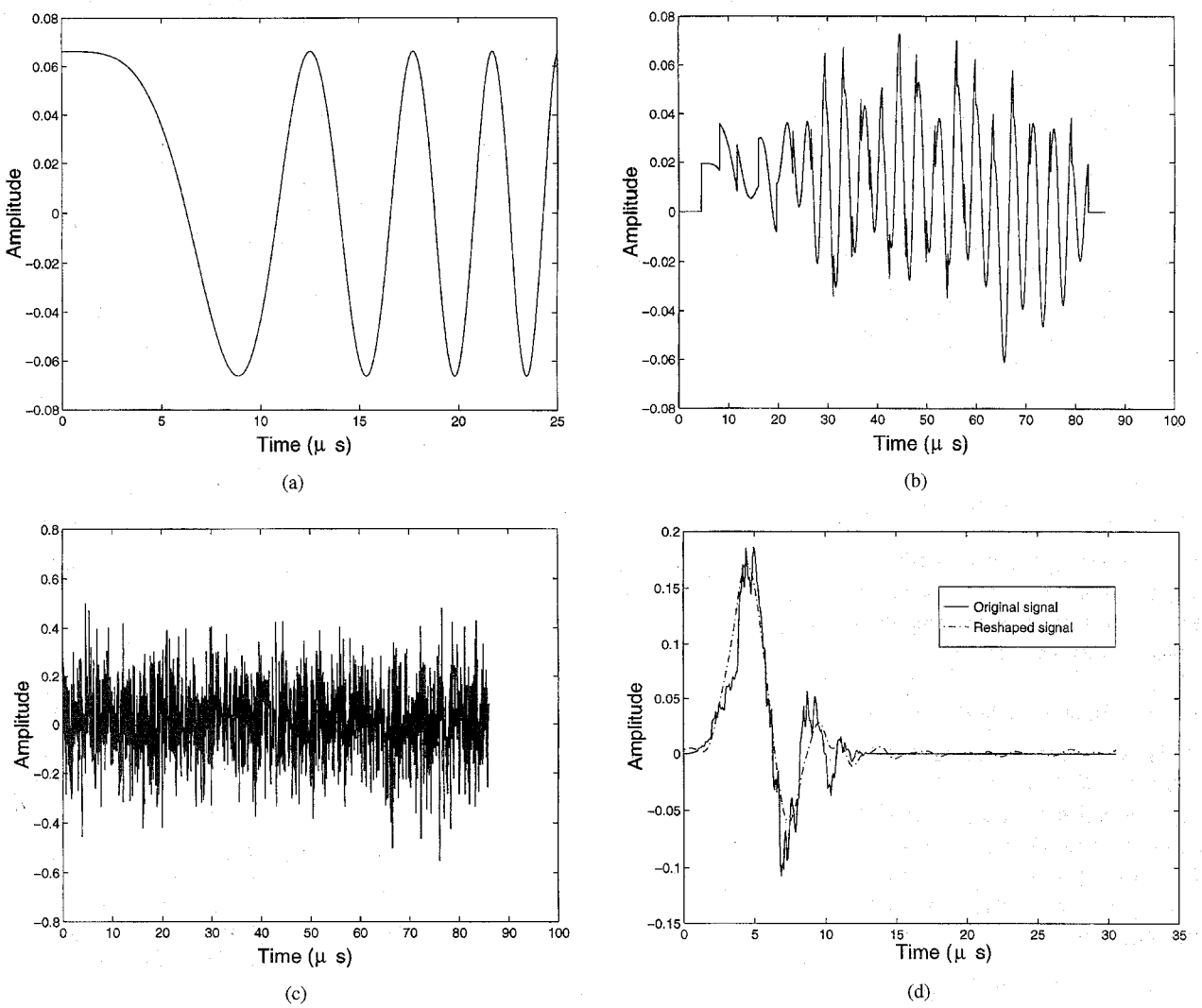


Fig. 6. (a) Transmitted signal $s(t)$. (b) Multipath noise-free signal $s_m(t)$ (15 paths). (c) Multipath noisy signal $r(t)$ (15 paths). (d) Minimum gap scaling function $g^*(t)$ and reshaped minimum gap scaling function $g_r^*(t)$ (baseband).

Fig. 6(c) shows the same multipath signal with additive white Gaussian noise. Since the signals in the figures are normalized to unit energy, and $r(t)$ is longer than $s(t)$, Fig. 6(c) is not at the same scale as Fig. 6(a).

B. Gap Minimization

In this case, the transmitted signal is a modulated chirp signal. We handle the modulation in two steps. We first design the generating function in baseband, i.e., we first let $\frac{s_b(t)=\text{rect}(t}{T_c)\cos(2\pi\gamma t^2)}$ be the transmitted signal and design an optimal generating function $g_r^*(t)$ to minimize the gap between $\mathcal{S}_b$ and $\mathcal{G}_r^*$, where

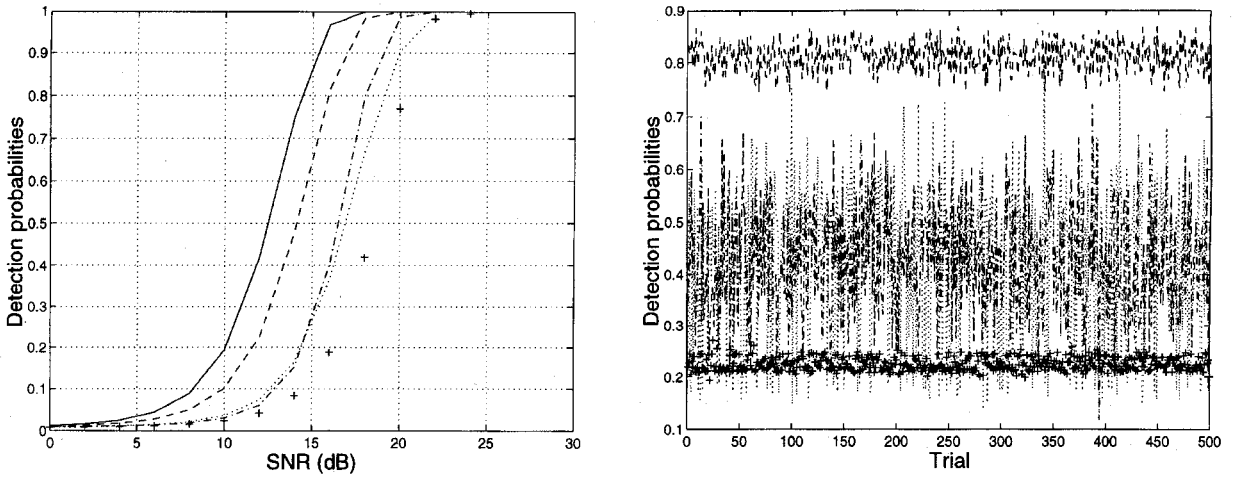
$$\mathcal{S}_b = \left\{ s_{m,b}(\cdot) = \sum_{k=1}^{K} \alpha_k s_b(\cdot - \tau_k), \alpha_k, \tau_k \in \mathbb{R}, K \in \mathbb{Z}^+ \right\} \tag{61}$$

and $\mathcal{G}_r^*$ is given in (44). Then, we modulate this baseband waveform to the carrier frequency.

To design the signal in baseband, we use the parameterization given by [8] to parameterize the scaling function $g(t;\zeta)$ in (31). In the parameterization described in Section II-C, we need to choose the number $M-1$ of unconstrained parameters. The choice of $M-1$ is a compromise between different requirements. The larger the number, the smoother the scaling function and the larger the dimension of the parameter space over which to carry out the optimization. We choose $M-1=3$, which leads to good acceptable performance, as we will see in our results below, while keeping the computational effort manageable. Denote by $[\zeta_1,\zeta_2,\zeta_3] \in [0,2\pi]$ the three unconstrained parameters. The optimization is done by searching the parameter space. We compute the gap at the vertices of a 3-D cubic grid given by

$$\zeta_1(l) = \zeta_2(l) = \zeta_3(l) = \frac{2\pi l}{50}, \qquad l = 0, \cdots, 49 \tag{62}$$

and find the $\zeta^* = [\zeta_1^*, \zeta_2^*, \zeta_3^*]$ that leads to the minimum value of the gap. The optimal scaling function $g^*(t)$ is reconstructed from ζ^* using the algorithm described in [10]. Fig. 6(d) shows the baseband scaling function that minimizes (31) and the reshaped $g_r^*(t)$ of $g^*(t)$ using the algorithm in [1]. Notice that the reshaped function does not look drastically different from the original function, but it is much more shiftable. The



(a) (b)

Fig. 7. (a) Average detection probabilities $K = 15, P_F = 0.01$. The solid line is the USR receiver, "- -" is the minimum gap receiver, "-." is the correlator receiver, "· · ·" is the MFIS receiver, and "+" is the energy detector. (b) Detection probabilities at 16.5 dB of the 500 trials. "- -" is the minimum gap receiver, "-." is the correlator receiver, "· · ·" is the MFIS receiver, and "+" is the energy detector.

maximum representation error given by (37) has been reduced from 0.3 to about 0.025.

C. Performance Comparison

We generate 500 delay patterns using a random number generator, and compute the performance for each of these delay patterns. The delay patterns of interest are where the closest replicas have considerable overlap since when they are well separated, the CR performs reasonably well. For the parameters described in Section V-A, below (60), we compute roughly the bandwidth of the frequency-modulated chirp signal at passband to be close to 480 KHz, corresponding to an envelope correlation time of approximately 4 μs. In our experiment, for 250 delay patterns, the 15 delays tested are

$$\tau_k = (4 \times (k-1) + \delta_k)\mu \text{ s} \qquad k = 1, \cdots, 15 \tag{63}$$

while the 15 delays for the remaining 250 delay patterns are

$$\tau_k = (4 \times (k-1) + \delta_k + 0.5)\mu\text{s} \qquad k = 1, \cdots, 15 \tag{64}$$

where $\{\delta_k\}$ are independent delays uniformly distributed in $[0, 1]\mu$s. The time spacing between consecutive filters in the bank of filters (MGR and MFIS receivers) is 2 μs. Fig. 7(a) shows the average detection probabilities for these 500 trials as a function of $10\log(1/N_0)$ with the false alarm probability $P_F = 0.01$. The signal energy is normalized to 1, and N_0 is the variance of the noise. Thus, $1/N_0$ is the SNR. The number of paths K and the attenuation factors $\{\alpha_k\}$, as described earlier in this section, are $K = 15$ and $\alpha_k = 1, k = 1, \cdots, K$. The performance of the optimal GLRT receiver is not tested because of its extremely high computational cost. Instead, we use the USR. As we mentioned early, the USR gives us only an optimistic performance upper bound. The detection probability of the GLRT receiver is between that of USR and MGR.

There are five curves in Fig. 7(a). The solid line represents the USR. The dashed line is for the MGR we have designed. The dash-dotted line is for the CR, and the dotted line is for the MFIS receiver. The "+" curve represents the ED. Table I presents the prior information used by each of the receivers and the computational complexity compared with the MGR receiver. The difference between the USR and the ED is about 6.3 dB. This is an optimistic estimate of how much SNR gain over the ED we have available.

TABLE I
PRIOR INFORMATION USED BY EACH OF THE RECEIVERS

Receivers	Transmitted Signal	Delays	Attenuations	Computational complexity compared to MGR
USR	known	known	unknown	unrealistically complex
MGR	known	unknown	unknown	
CR	known	unknown	unknown	much more complex
MFIS	known	unknown	unknown	similar
Energy Detector	unknown	unknown	unknown	simpler

Analysis of Fig. 7(a) shows that the MGR provides an average gain of about 2.5 dB over the CR, an average gain of about 3.1 dB over the MFIS and about a 5-dB gain over the ED.

The reason for the gain over the CR is that in our simulation, different delayed replicas of the transmitted signal overlap; therefore, the CR is not optimal in any sense. The reason for the gain over MFIS is that $\Sigma_n \beta_n g_r^*(\cdot - n)$ not only matches well with linear combinations of integer shifts of $s(t)$ but also with linear combinations of arbitrary shifts of $s(t)$, whereas the MFIS only matches with linear combinations of integer shifts of the transmitted signal $s(t)$.

Fig. 7(b) shows the detection probabilities at 16.5 dB for these 500 trials. We see that the fluctuation of the MGR receiver [top curve in Fig. 7(b)] is smaller than those of the MFIS receiver and of the CR receiver. In other words, the MGR is more robust to the multipath distortion. The ED

[bottom curve in Fig. 7(b)] is also robust, but its average performance is poor. This is because it does not utilize any multipath structure.

Simulation results presented elsewhere carried out under different conditions lead to gains similar to the ones in Fig. 7.

VI. Conclusion and Future Work

This paper develops a minimum gap receiver that is fine tuned to multipath detection. We design a representation subspace $\mathcal{G}$ that is matched to the multipath signal subspace $\mathcal{S}$ in the gap sense. The MGR is simple to implement. It avoids the multidimensional nonlinear optimization required by the optimal receiver while providing an average gain of about 3.1 dB over the MFIS and about a 2.5 dB gain over the CR. Further, the minimum gap receiver is robust to the delay patterns tested, whereas for these alternative receivers, the performance varies significantly with the actual delay patterns tested.

In this paper, we consider only the case where the transmitted signal $s(t)$ is completely known. For the cases where $s(t)$ is not known, for example, $s(t)$ is random, the original signal set $\mathcal{S}$ is no longer a subspace, and we need other measures to quantify the closeness. The details of this extension will be reported in future work.

Appendix

In this Appendix, we give the proof of Theorem 1.

From the second definition of the gap, we know that $\delta(\mathcal{S}_{\text{int}}, \mathcal{G})$ is the maximum of $\hat{\delta}(\mathcal{S}_{\text{int}}, \mathcal{G})$ and $\hat{\delta}(\mathcal{G}, \mathcal{S}_{\text{int}})$. We start with the left side $\hat{\delta}(\mathcal{S}_{\text{int}}, \mathcal{G})$. From (12) and (13), in order to compute $\hat{\delta}(\mathcal{S}_{\text{int}}, \mathcal{G})$, we need first to compute

$$\begin{aligned}\text{dist}^2(s_{\text{int}}(\cdot), \mathcal{G}) &= \inf_{g_{\text{int}}(\cdot)\in\mathcal{G}} \|s_{\text{int}}(\cdot) - g_{\text{int}}(\cdot)\|^2 \\ &= \|s_{\text{int}}(\cdot)\|^2 - \|P_{\mathcal{G}}[s_{\text{int}}]\|^2 \end{aligned} \tag{65}$$

where

$$s_{\text{int}}(\cdot) = \sum_n \alpha_n s(\cdot - n) \in \mathcal{S}_{\text{int}}$$

$$g_{\text{int}}(\cdot) = \sum_n \beta_n g(\cdot - n) \in \mathcal{G}$$

and $P_{\mathcal{G}}[s_{\text{int}}]$ is the orthogonal projection of $s_{\text{int}}(\cdot)$ on $\mathcal{G}$.

Since $\{g(\cdot - n)\}$ is a Riesz basis [10] of $\mathcal{G}$, i.e., $\exists A_g, B_g$ such that $0 < A_g \le B_g < \infty$ and

$$A_g \le \sum_l |\mathcal{F}_g(f + l)|^2 \le B_g \quad \text{a.e.} \tag{66}$$

the orthogonal projection $P_{\mathcal{G}}[s_{\text{int}}]$ is obtained by taking the inner product of $s_{\text{int}}(\cdot)$ with $\{\tilde{g}(\cdot - n), n \in \mathbb{Z}\}$, where $\tilde{g}(\cdot)$ is a biorthogonal function [10] of $g(\cdot)$, and its Fourier transform is

$$\mathcal{F}_{\tilde{g}}(f) = \frac{\mathcal{F}_g(f)}{\sum_l |\mathcal{F}_g(f + l)|^2}. \tag{67}$$

With $\tilde{g}(\cdot)$, we have

$$(P_{\mathcal{G}}[s_{\text{int}}])(t) = \sum_k \langle s_{\text{int}}(\cdot), \tilde{g}(\cdot - k)\rangle \cdot g(t - k). \tag{68}$$

Taking the Fourier transform of (68) and substituting (67), we get (69)–(72), shown at the bottom of the page. Taking the norm square, we have (73)–(77), shown at the top of the next page. Once we have $\text{dist}^2(s_{\text{int}}(\cdot), \mathcal{G})$, the square of the left side of the gap $\hat{\delta}^2(\mathcal{S}_{\text{int}}, \mathcal{G})$ is obtained by taking the supremum, i.e.,

$$\hat{\delta}^2(\mathcal{S}_{\text{int}}, \mathcal{G}) = \sup_{s_{\text{int}}(\cdot)\in\mathcal{S}_{\text{int}}} \text{dist}^2(s_{\text{int}}(\cdot), \mathcal{G}) \tag{78}$$

$$= \sup_{s_{\text{int}}(\cdot)\in\mathcal{S}_{\text{int}}} (\|s_{\text{int}}(\cdot)\|^2 - \|P_{\mathcal{G}}[s_{\text{int}}]\|^2) \tag{79}$$

$$= 1 - \inf_{s_{\text{int}}(\cdot)\in\mathcal{S}_{\text{int}}} \|P_{\mathcal{G}}[s_{\text{int}}]\|^2 \tag{80}$$

subject to

$$\|s_{\text{int}}(\cdot)\|^2 = \int_{-\infty}^{+\infty} |s_{\text{int}}(t)|^2 \, dt \tag{81}$$

$$\mathcal{F}[P_{\mathcal{G}}[s_{\text{int}}]](f) = \mathcal{F}_g(f) \cdot \sum_k \int_{-\infty}^{+\infty} \mathcal{F}_{s_{\text{int}}}(f_1)\overline{\mathcal{F}}_{\tilde{g}}(f_1) \exp\{-j2\pi k(f - f_1)\} \, df_1 \tag{69}$$

$$= \mathcal{F}_g(f) \cdot \sum_n \alpha_n \sum_k \int_{-\infty}^{+\infty} \frac{\mathcal{F}_s(f_1)\overline{\mathcal{F}}_g(f_1) \exp\{-j2\pi k(f - f_1)\} \exp\{-j2\pi n f_1\}}{\sum_l |\mathcal{F}_g(f_1 + l)|^2} \, df_1 \tag{70}$$

$$= \mathcal{F}_g(f) \cdot \sum_n \alpha_n \int_{-\infty}^{+\infty} \frac{\mathcal{F}_s(f_1)\overline{\mathcal{F}}_g(f_1) \sum_k \exp\{-j2\pi k(f - f_1)\} \exp\{-j2\pi n f_1\}}{\sum_l |\mathcal{F}_g(f_1 + l)|^2} \, df_1 \tag{71}$$

$$= \mathcal{F}_g(f) \cdot \left\{ \sum_n \alpha_n \exp\{-j2\pi n f\} \right\} \cdot \frac{\sum_k \mathcal{F}_s(f + k)\overline{\mathcal{F}}_g(f + k)}{\sum_l |\mathcal{F}_g(f + l)|^2} \tag{72}$$

$$\|P_{\mathcal{G}}[s_{\text{int}}]\|^2 = \|\mathcal{F}[P_{\mathcal{G}}[s_{\text{int}}]]\|^2 \tag{73}$$

$$= \int_{-\infty}^{+\infty} \left| \left\{ \sum_n \alpha_n \exp\{-j2\pi nf\} \right\} \cdot \mathcal{F}_g(f) \cdot \frac{\sum_k \mathcal{F}_s(f+k)\overline{\mathcal{F}}_g(f+k)}{\sum_l |\mathcal{F}_g(f+l)|^2} \right|^2 df \tag{74}$$

$$= \sum_i \int_i^{i+1} \left| \left\{ \sum_n \alpha_n \exp\{-j2\pi nf\} \right\} \cdot \mathcal{F}_g(f) \cdot \frac{\sum_k \mathcal{F}_s(f+k)\overline{\mathcal{F}}_g(f+k)}{\sum_l |\mathcal{F}_g(f+l)|^2} \right|^2 df \tag{75}$$

$$= \int_0^1 \left| \sum_n \alpha_n \exp\{-j2\pi nf\} \right|^2 \cdot \frac{\left| \sum_k \mathcal{F}_s(f+k)\overline{\mathcal{F}}_g(f+k) \right|^2}{\sum_l |\mathcal{F}_g(f+l)|^2} df \tag{76}$$

$$= \int_0^1 \left| \sum_n \alpha_n \exp\{-j2\pi nf\} \right|^2 \cdot \sum_l |\mathcal{F}_s(f+l)|^2 \cdot \frac{\left| \sum_k \mathcal{F}_s(f+k)\overline{\mathcal{F}}_g(f+k) \right|^2}{\sum_l |\mathcal{F}_s(f+l)|^2 \sum_l |\mathcal{F}_g(f+l)|^2} df \tag{77}$$

$$= \int_0^1 \left| \sum_n \alpha_n \exp\{-j2\pi nf\} \right|^2 \cdot \sum_l |\mathcal{F}_s(f+l)|^2 \, df \tag{82}$$

$$= 1. \tag{83}$$

Since $\{s(\cdot - n), n \in \mathbb{Z}\}$ is also a Riesz basis, i.e., $\exists\, A, B$ such that $0 < A \leq B < \infty$ and

$$A \leq \sum_l |\mathcal{F}_s(f+l)|^2 \leq B \quad \text{a.e.} \tag{84}$$

the periodic function, integrand of (82)

$$\left| \sum_n \alpha_n \exp\{-j2\pi nf\} \right|^2 \cdot \sum_l |\mathcal{F}_s(f+l)|^2 \tag{85}$$

can generate any positive function in $L_1[0,1)$. Using a limiting argument, we can choose it to be the Dirac delta function located at the frequency where

$$\frac{\left| \sum_k \mathcal{F}_s(f+k)\overline{\mathcal{F}}_g(f+k) \right|^2}{\sum_l |\mathcal{F}_s(f+l)|^2 \sum_l |\mathcal{F}_g(f+l)|^2} \tag{86}$$

has its infimum and continuous. Then, the quantity

$$\int_0^1 \left| \sum_n \alpha_n \exp\{-j2\pi nf\} \right|^2 \cdot \sum_l |\mathcal{F}_s(f+l)|^2 \cdot \frac{\left| \sum_k \mathcal{F}_s(f+k)\overline{\mathcal{F}}_g(f+k) \right|^2}{\sum_l |\mathcal{F}_s(f+l)|^2 \sum_l |\mathcal{F}_g(f+l)|^2} df \tag{87}$$

is minimized. The infimum is

$$\inf_{f \in [0,1)} \frac{\left| \sum_k \mathcal{F}_s(f+k)\overline{\mathcal{F}}_g(f+k) \right|^2}{\sum_l |\mathcal{F}_s(f+l)|^2 \sum_l |\mathcal{F}_g(f+l)|^2}. \tag{88}$$

This completes the calculation of $\hat{\delta}^2(\mathcal{S}_{\text{int}}, \mathcal{G})$. The square of the right side of the gap $\hat{\delta}^2(\mathcal{G}, \mathcal{S}_{\text{int}})$ can be obtained by interchanging $s(\cdot)$ and $g(\cdot)$. However, observing that (88) is symmetric about $s(\cdot)$ and $g(\cdot)$, interchanging $s(\cdot)$ and $g(\cdot)$ does not affect the result, which means that

$$\hat{\delta}^2(\mathcal{G}, \mathcal{S}_{\text{int}}) = \hat{\delta}^2(\mathcal{S}_{\text{int}}, \mathcal{G}) \tag{89}$$

Therefore, this proves Theorem 1.

REFERENCES

[1] S. A. Benno and J. M. F. Moura, "On translation invariant subspaces and critically sampled wavelet transforms," *Multidimensional Syst. Signal Processing*, vol. 8, no. 1–2, pp. 89–110, Jan. 1997.

[2] ———, "Scaling functions optimally robust to translations," submitted for publication.

[3] ———, "Nearly shiftable scaling functions," in *Proc. ICASSP*, May 1995, pp. II–1097–1100.

[4] C. He, J. M. F. Moura, and S. A. Benno, "Gap detector for multipath," in *Proc. ICASSP*, May 1996, pp. V–2650–2653.

[5] T. Kato, *Perturbation Theory for Linear Operators*, 2nd. ed. New York: Springer-Verlag, 1976.

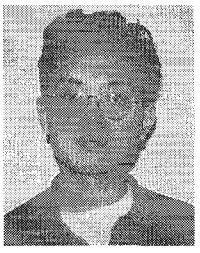
[6] M. J. D. Rendas and J. M. F. Moura, "Ambiguity analysis in source localization," Tech. Rep., ECE Dept., Carnegie Mellon Univ., Pittsburgh, PA, May 1991.

[7] I. Daubechies, "Orthonormal bases of compactly supported wavelets," *Commun. Pure Appl. Math*, vol. 41, pp. 909–996, Nov. 1988.

[8] H. Zou and A. H. Tewfik, "Parameterization of compactly supported orthonormal wavelets," *IEEE Trans. Signal Processing*, vol. 41, pp. 1428–1431, Mar. 1993.

[9] J. E. Ehrenberg, T. E. Ewart, and R. D. Morris, "Signal processing techniques for resolving individual pulses in a multipath signal," *J. Acoust. Soc. Amer.*, vol. 63, pp. 1861–1865, June 1978.

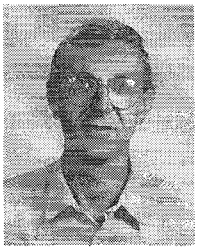
[10] C. K. Chui, *An Introduction to Wavelets*. New York: Academic, 1992.
[11] S. Haykin, *Communication Systems*, 2nd ed. New York: Wiley, 1983.
[12] C. He and J. M. F. Moura, "Robust detection via subspace design," in *Proc. SPIE, Wavelet Applications Signal Image Processing IV*, Aug. 1996.
[13] J. E. Odegard, R. A. Gopinath, and C. S. Burrus, "Optimal wavelets for signal decomposition and the existence of scale limited signals," in *Proc. ICASSP*, Oct. 1992, pp. IV–597–600.
[14] A. J. E. M. Janssen, "The Zak transform: A signal transform for sampled time-continuous signals," *Philips J. Res.*, vol. 43, no. 1, pp. 23–69, 1988.
[15] L. Auslander and R. Tolimieri, "Radar ambiguity functions and group theory," *SIAM J. Math. Anal.*, vol. 16, no. 3, pp. 577–601, May 1985.
[16] L. Auslander, I. Gertner, and R. Tolimieri, "The discrete Zak transform application to time-frequency analysis and synthesis of nonstationary signals," *IEEE Trans. Signal Processing*, vol. 39, pp. 825–835, Apr. 1992.
[17] B. Friedlander and B. Porat, "Detection of transient signals by the Gabor representation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 169–179, Feb. 1989.
[18] D. K. Barton, *RADARS*. Norwood, MA: Artech House, 1975, vol. 3.



Chuang He (S'93) received the B.S. degree from Tsinghua University, Beijing, China, in 1991 and the M.S. degree from Tulane University, New Orleans, LA, in 1993.

Since 1993, he has been a research assistant with the Electrical and Computer Engineering Department, Carnegie Mellon University, Pittsburgh, PA, where he is pursuing his Ph.D. degree in electrical and computer engineering. His research interests are digital signal processing, wavelet and time-frequency transforms, image processing, and communications.

Mr. He is a member of Eta Kappa Nu.



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. and E.E. degrees in 1973 and the D.Sc. degree in 1975 in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge.

He is presently a Professor of Electrical and Computer Engineering at Carnegie Mellon University (CMU), Pittsburgh, PA, which he joined in 1986. Prior to this, 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 from 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 co-directed two international scientific meetings on signal processing theory and applications. He has more than 160 published technical contributions and is co-editor of two books.

Dr. Moura is currently the Editor-in-Chief of the IEEE TRANSACTIONS ON SIGNAL PROCESSING and a member of the Underwater Accoustics Technical Committee of the Signal Processing Society. He was a member of the IEEE Press Board from 1991 to 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 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.