Improving the Generalized Likelihood Ratio Test for Unknown Linear Gaussian Channels

Elona Erez, *Student Member, IEEE*, and Meir Feder, *Fellow, IEEE*

***Abstract*—In this work, we consider the decoding problem for unknown Gaussian linear channels. Important examples of linear channels are the intersymbol interference (ISI) channel and the diversity channel with multiple transmit and receive antennas employing space–time codes (STC). An important class of decoders is based on the generalized likelihood ratio test (GLRT). Our work deals primarily with a decoding algorithm that uniformly improves the error probability of the GLRT decoder for these unknown linear channels. The improvement is attained by increasing the minimal distance associated with the decoder. This improvement is uniform, i.e., for all the possible channel parameters, the error probability is either smaller by a factor (that is exponential in the improved distance), or for some, may remain the same. We also present an algorithm that improves the average (over the channel parameters) error probability of the GLRT decoder. We provide simulation results for both decoders.**

***Index Terms*—Diversity channels, generalized likelihood ratio test (GLRT), intersymbol interference (ISI), maximum likelihood (ML).**

I. Introduction

WHEN a communication channel is band limited, signal transmission at a symbol rate that equals or exceeds the bandwidth of the channel results in intersymbol interference (ISI). One way to deal with ISI channels is to use an equalizer in order to remove the effects of the channel. From the probability of error viewpoint, the maximum-likelihood (ML) decoder, sometimes implemented via the ML sequence estimation (MLSE) algorithm [1], is optimal for *known* ISI channels. However, the best way to decode is not clear when the ISI coefficients are unknown.

Another class of linear channels is the class of diversity channels, with several transmit and receive antennas. The channel parameters are the fading coefficients between the transmitters and receivers. Space–time codes (STC), e.g., the codes introduced in [2], have been shown to significantly improve the communication performance over such multiple-antenna fading channels. In [2], as in many other STC schemes, the channel coefficients are assumed to be known to the decoder. But the question remains as to how to decode when the channel parameters are unknown.

A common approach in this situation, applied by many standard equalization methods, is to use a training sequence or a pilot sequence, to enable the receiver to identify the channel in use. Since the sequence is known at the receiver, the receiver can estimate the channel law by studying the received symbols corresponding to the known input sequence. The usage of training for diversity channels is discussed, e.g., in [3].

The training sequence approach, however, has many drawbacks. First, there is a mismatch penalty, since the channel estimate formed at the receiver is imprecise, which results in an increased error rate. Secondly, there is penalty in throughput, since the training sequence carries no information. This penalty increases as the training sequence is sent more frequently or as its length, compared with the length of the data sequence, is larger. When the channel changes rapidly over time, using training sequences might be completely inadequate. An example of such a rapidly changing environment is the underwater communication channel [4]. In mobile wireless communications, the varying locations of the mobile transmitter and receiver with respect to the scatterers lead to a rapidly changing channel as well. Another example where training fails is in broadcast multipoint communication networks. In this case, the training sequence must be sent (and received by all receivers) whenever any of the terminals goes down, even if it is desired to retain only that receiver. Furthermore, the reverse channel maybe loaded with requests for training retransmission. For all these reasons, the training approach can be problematic and so it is desirable to find methods that can decode without training sequences.

A possible way to deal with the problem of communication over unknown channels is to avoid signaling that requires the knowledge of the unknown parameters. One example is to use differential phase shift keying (DPSK), since the differential phase does not depend on the possibly unknown fading coefficients as long as they are time invariant. Clearly, in this case a training sequence is not necessary. An efficient differential detection scheme which does not require training sequences and has a linear complexity was developed in [5] for diversity channels. The detection scheme was developed for a simple transmit encoding design, known as the Alamouti block coding, first introduced in [6]. A different approach which, again, requires no pilot sequences is the unitary space–time modulation introduced in [7], where each matrix in the signal constellation is unitary (this decoder assumes a Rayleigh stochastic model on the channel coefficients with a known covariance matrix). If,

Manuscript received February 26, 2002; revised November 19, 2002. This work was supported in part by a grant from the Israeli Science Foundation. The material in this paper was presented in part at the 38th Allerton Conference on Communication, Control, and Computing, Monticello, IL, October 2000, and at the IEEE International Symposium on Information Theory and Its Applications, Honolulu, HI, November 2000.

The authors are with the Department of Electrical Engineering–Systems, Tel-Aviv University, Ramat-Aviv 69978, Israel (e-mail: elona@eng.tau.ac.il; meir@eng.tau.ac.il).

Communicated by V. V. Veeravalli, Associate Editor for Detection and Estimation.

Digital Object Identifier 10.1109/TIT.2003.809598

however, we do not have or do not want to impose a specific structure on the codewords or the signal set, the differential approach may not be applicable.

As noted above, the ML decoder is optimal, i.e., it leads to minimal error probability for known channels. In the situation considered in this work, the channel coefficients are unknown and, furthermore, do not have a known stochastic model. A possible decision rule for unknown channels is the generalized likelihood ratio test (GLRT), which essentially jointly maximizes the likelihood with respect to both channel parameters and the data. Some properties of the GLRT have been shown, e.g., the GLRT is asymptotically optimal in the Neyman–Pearson setting if the class is dense enough, see [8]. In our problem of unknown channel, if the family of possible channels consists of all discrete memoryless channels (DMCs) with finite input and output alphabets, the GLRT coincides with the maximum empirical mutual information (MMI) decoder. In this case, as shown in [9], if all the codewords have the same type, then the GLRT achieves the same error exponent as the ML decoder. However, the GLRT may no longer be optimal in this sense if the class of channels is a strict subset of the set of all DMCs, [10]. Furthermore, in general, there is no claim for the optimality of the GLRT under the error probability criterion. Indeed, our work deals primarily with a novel decoder that uniformly improves the error probability of the GLRT decoder for linear Gaussian channels. As we do not assume a stochastic model on the parameter space, in order to be superior to the GLRT our new decoder improves the performance for some channel parameters (in the parameter space) and does not worsen the error performance for any other possible channel parameter.

The outline of the paper is as follows. In Section II, we introduce the channel models. In Section III, we discuss the GLRT decoder for these channel models. We then briefly present, in Section IV, a decoding technique for a simple fading channel, described in [11] and in [12, the Appendix], that serves as the motivation for our novel decoder. The main new result appears in Sections V and VI, where we develop a new robust decoder for a special (hyperplane) case and the general case, respectively. This decoder is called the Uniformly improved GLRT (ULRT). In Section VII, we suggest an additional decoder, the energy weighted decoder (EWD), that improves the GLRT but only on the average over the channel parameters. A summary and discussion of further research concludes the paper.

II. The Channel Models

The problem of decoding one out of M codewords (hypotheses) observed after passing through a Gaussian ISI channel is modeled as

$$\mathcal{H}_m\colon y_n = \sum_{i=0}^{K-1} h_i x_{n-i}^{(m)} + z_n, \qquad n = 0, 1, \ldots, N-1, \; m = 1, \ldots, M \tag{1}$$

where $\{y_n\}_{n=0}^{N-1}$ are the observed data samples, $\{x_n^{(m)}\}_{n=0}^{P-1}$ are the transmitted symbols for the mth codeword, and $N = P + K - 1$, $\{h_i\}_{i=0}^{K-1}$ are the unknown ISI coefficients and $\{z_n\}_{n=0}^{N-1}$ are independent and identically distributed (i.i.d.) samples of white Gaussian noise with variance σ^2. Note that the length of the observation is N, which is longer than the length of the codewords which is P. We can write (1) as

$$H_m\colon \boldsymbol{y} = X_m \boldsymbol{h} + \boldsymbol{z} \tag{2}$$

where

$$\boldsymbol{y} = \begin{bmatrix} y_0 \\ \vdots \\ y_{N-1} \end{bmatrix}, \quad \boldsymbol{z} = \begin{bmatrix} z_0 \\ \vdots \\ z_{N-1} \end{bmatrix} \tag{3}$$

$$X_m = \begin{bmatrix} x_0^{(m)} & 0 & \cdots & 0 \\ x_1^{(m)} & x_0^{(m)} & \cdots & 0 \\ \vdots & \vdots & \ddots & x_0^{(m)} \\ x_{P-1}^{(m)} & \vdots & \ddots & \vdots \\ 0 & x_{P-1}^{(m)} & \ddots & \vdots \\ 0 & 0 & \cdots & x_{P-1}^{(m)} \end{bmatrix} \tag{4}$$

and

$$\boldsymbol{h} = \begin{bmatrix} h_0 \\ \vdots \\ h_{K-1} \end{bmatrix} \tag{5}$$

and where the matrices X_m, $m = 1 \cdots M$ are assumed to be full rank. It can be easily seen from the structure of the matrix that X_m is full rank unless $x_0^{(m)} = x_1^{(m)} = \cdots = x_{P-1}^{(m)} = 0$ since the diagonal shape of the columns ensures that they are linearly independent.

For convenience, we define the transmitted signal vectors given by

$$\boldsymbol{x}^{(m)} = \left[x_0^{(m)}, \ldots, x_{P-1}^{(m)}\right]^T. \tag{6}$$

Another linear Gaussian case is the diversity channel with L transmitting elements and J receiving antenna elements where

$$\mathcal{H}_m\colon y_{n,j} = \sum_{i=1}^{L} \alpha_{i,j} x_{n,i}^{(m)} + z_{n,j}, \qquad n = 0, 1, \ldots, N-1, \; m = 1, \ldots, M, \; j = 1, \ldots, J \tag{7}$$

and where $\{y_{n,j}\}_{n=0}^{N-1}$ are the observed data samples at receive antenna j, $\{x_{n,i}^{(m)}\}_{n=0}^{N-1}$ are the symbols transmitted by the ith antenna for the mth codeword, $\alpha_{i,j}$ is the unknown fading coefficient from transmit antenna i to receive antenna j, and $\{z_{n,j}\}$ are i.i.d. samples of white Gaussian noise with variance σ^2. We can write (7) as

$$\mathcal{H}_m\colon Y = X_m \alpha + Z \tag{8}$$

where

$$X_m = \begin{bmatrix} x_{0,1}^{(m)} & \cdots & x_{0,L}^{(m)} \\ \vdots & \ddots & \vdots \\ x_{N-1,1}^{(m)} & \ddots & x_{N-1,L}^{(m)} \end{bmatrix} \tag{9}$$

$$\alpha = \begin{bmatrix} \alpha_{1,1} & \cdots & \alpha_{1,J} \\ \vdots & \ddots & \vdots \\ \alpha_{L,1} & \ddots & \alpha_{L,J} \end{bmatrix} \tag{10}$$

$$Y = \begin{bmatrix} y_{0,1} & \cdots & y_{0,J} \\ \vdots & \ddots & \vdots \\ y_{N-1,1} & \cdots & y_{N-1,J} \end{bmatrix} \tag{11}$$

and,

$$Z = \begin{bmatrix} z_{0,1} & \cdots & z_{0,J} \\ \vdots & \ddots & \vdots \\ z_{N-1,1} & \cdots & z_{N-1,J} \end{bmatrix} \tag{12}$$

and where the matrices $X_m,\ m = 1, \ldots, M$ are assumed to be full rank. In many coding methods encountered in the literature, the matrices X_m turned out to be full rank. For example, in [13], each X_m has an orthogonal structure and in [7], the columns of X_m are designed to be (scaled) orthonormal.

Clearly, the ISI channel is a special case of the diversity channel with a single antenna at the receiver $(J = 1)$. In this paper, we discuss explicitly the ISI case, but the decoders we introduce can be directly extended to handle diversity channels [14].

III. The Generalized Likelihood Ratio Test (GLRT)

Decoding with unknown channel parameters leads to a composite hypothesis testing problem [12],[15]. In composite hypotheses testing, there is an uncertainty in the parameters that define the probability distribution associated with each hypotheses $H_i,\ i = 1, \ldots, M$. Specifically, for each hypothesis there is a family of possible probability assignments $\{p_\theta(\boldsymbol{y}|H_i),\ \theta \in \Lambda\}$, where $\boldsymbol{y} = (y_0, \ldots, y_n)$ is a sequence of observations, θ is the unknown parameter, and Λ is the set of unknown parameters. Note that in our case of unknown channel, the set of unknown parameters does not depend on the hypothesis. There is a family of channels

$$\mathcal{F} = \{p_\theta(\boldsymbol{y}|\boldsymbol{x}),\ \theta \in \Theta\} \tag{13}$$

and the hypotheses are the M possible codewords which are transmitted as an input to the channel.

If the channel is known, the decoding problem reduces to simple hypothesis testing, whose optimal solution in the sense of minimizing the error probability (assuming the codewords are equiprobable) is given by the ML decision rule

$$\phi_{\mathrm{ML}}(\boldsymbol{y}) = i \Leftrightarrow p_\theta\left(\boldsymbol{y}\,\middle|\,\boldsymbol{x}^{(i)}\right) = \max_{1\le j\le M} p_\theta\left(\boldsymbol{y}\,\middle|\,\boldsymbol{x}^{(j)}\right) \tag{14}$$

where $\boldsymbol{x}^{(i)}$ is the ith codeword. Since ML decoding in general leads to different rules for different channels it cannot be employed when the channel is unknown.

There are two major approaches to composite hypothesis testing [16]. The first is Bayesian, where the unknown parameters are considered as random variables with a specified prior probability. By taking the expectation of $p_\theta(\boldsymbol{y}|H_i)$ with respect to (w.r.t.) the unknown parameter θ, one obtains *a posteriori* probability distributions that are independent of θ and can be used for ML decision. The Bayesian approach can be computationally complex due to the expectation. Furthermore, it requires a subjective prior assumption. The second approach is the GLRT which has a lower computational complexity, and moreover, it does not make any assumption regarding a prior probability. The GLRT decoder can be defined as follows:

$$\phi_{GLRT}(\boldsymbol{y}) = i \Leftrightarrow \sup_{\theta\in\Theta} p_\theta\left(\boldsymbol{y}\,\middle|\,\boldsymbol{x}^{(i)}\right) = \max_{1\le j\le M} \sup_{\theta\in\Theta} p_\theta\left(\boldsymbol{y}\,\middle|\,\boldsymbol{x}^{(j)}\right). \tag{15}$$

While the GLRT is intuitively appealing as a joint channel and data estimation scheme, it does not have a solid theoretical justification in general. For ISI channels, as shown in this paper, the GLRT can be strictly suboptimal.

In the remainder of this section we present the GLRT decoding rule for ISI channels. Under the ISI linear Gaussian model previously described, the joint codeword and channel parameter estimation reduces to a joint minimization of the following Euclidean distance, and so the GLRT decoding rule becomes

$$\hat{m} = \arg\min_m \left\{ \min_{\boldsymbol{h}} \|\boldsymbol{y} - X_m\boldsymbol{h}\|^2 \right\}. \tag{16}$$

Since we assumed that X_m are full rank, the least squares (LS) solution for $\boldsymbol{h}$ is

$$\hat{\boldsymbol{h}}_m = (X_m^T X_m)^{-1} X_m^T \boldsymbol{y}. \tag{17}$$

Substituting into (16) yields the following closed-form solution:

$$\hat{m} = \arg\max_m \left\{ \boldsymbol{y}^T X_m (X_m^T X_m)^{-1} X_m^T \boldsymbol{y} \right\}. \tag{18}$$

For two codewords $(M = 2)$, define the two subspaces each of the codewords spans

$$C_m = \{\boldsymbol{r}_m : \boldsymbol{r}_m = X_m\boldsymbol{h},\ \boldsymbol{h} \in \Re^K\}. \tag{19}$$

The decoding regions D_1 and D_2 of $m = 1, 2$, respectively, are given by

$$\begin{aligned} D_1 = \{\boldsymbol{v} : \boldsymbol{v}^T (X_1(X_1^T X_1)^{-1} X_1^T \\ - X_2(X_2^T X_2)^{-1} X_2^T)\boldsymbol{v} > 0\} \end{aligned} \tag{20}$$

$$\begin{aligned} D_2 = \{\boldsymbol{v} : \boldsymbol{v}^T (X_1(X_1^T X_1)^{-1} X_1^T \\ - X_2(X_2^T X_2)^{-1} X_2^T)\boldsymbol{v} < 0\}. \end{aligned} \tag{21}$$

The surface that separates the decoding regions D_1 and D_2 (the separating surface of the decoder) is given by

$$S^G = \{\boldsymbol{v} : \boldsymbol{v}^T (X_1(X_1^T X_1)^{-1} X_1^T - X_2(X_2^T X_2)^{-1} X_2^T)\boldsymbol{v} = 0\} \tag{22}$$

We will use these definitions in the following sections, where we show how the GLRT can be uniformly improved.

IV. Uniformly Improving the GLRT: Motivation

Consider the two-codewords case, and let us analyze the GLRT decoder performance given an ISI coefficients vector $\boldsymbol{h}$. Define

$$d_1^G(\boldsymbol{h}) = \min_{\boldsymbol{v}\in S^G} \{\|X_1\boldsymbol{h} - \boldsymbol{v}\|\} \tag{23}$$

$$d_2^G(\boldsymbol{h}) = \min_{\boldsymbol{v}\in S^G}\{\|X_2\boldsymbol{h} - \boldsymbol{v}\|\} \tag{24}$$

where $\boldsymbol{h}$, $X_m (m = 1, 2)$ are defined in (2) and S^G is the separating surface of the GLRT decoder defined in (22). Since the noise is white Gaussian, and as we assume that the two messages are equiprobable, the error probability given $\boldsymbol{h}$ for the GLRT decoder can be approximated by

$$P_e^G(\boldsymbol{h}) \approx \frac{1}{2} Q\left(\frac{d_1^G(\boldsymbol{h})}{\sigma}\right) + \frac{1}{2} Q\left(\frac{d_2^G(\boldsymbol{h})}{\sigma}\right). \tag{25}$$

Assume that $d_1^G(\boldsymbol{h}) < d_2^G(\boldsymbol{h})$. The exponential order of $P_e^G(\boldsymbol{h})$ is given by

$$P_e^G(\boldsymbol{h}) \sim e^{-\frac{d_1^G(\boldsymbol{h})^2}{2\sigma^2}}. \tag{26}$$

Now, suppose we can find another decoder defined by a separating surface S^N, with respective distances

$$d_1^N(\boldsymbol{h}) = \min_{\boldsymbol{v}\in S^N}\{\|X_1\boldsymbol{h} - \boldsymbol{v}\|\} \tag{27}$$

$$d_2^N(\boldsymbol{h}) = \min_{\boldsymbol{v}\in S^N}\{\|X_2\boldsymbol{h} - \boldsymbol{v}\|\} \tag{28}$$

such that

$$\min_{m=1,2}\{d_m^N(\boldsymbol{h})\} \geq \min_{m=1,2}\{d_m^G(\boldsymbol{h})\}, \qquad \forall\, \boldsymbol{h} \in \Re^K \tag{29}$$

$$\exists\, \boldsymbol{h}\colon \min_{m=1,2}\{d_m^N(\boldsymbol{h})\} > \min_{m=1,2}\{d_m^G(\boldsymbol{h})\}, \qquad \boldsymbol{h} \in \Re^K. \tag{30}$$

These conditions ensure that for some $\boldsymbol{h}$ the error probability of the new decoder is improved exponentially, while for the rest it remains at least the same; thus, this decoder improves the GLRT uniformly.

We show now an example, originally presented in [11] and in [12, the Appendix], for such a decoder in the simple fading channel case. The fading channel is actually a single-parameter ISI channel where the observed data is given by

$$y_n = hx_n + z_n, \qquad n = 0, \ldots, N-1 \tag{31}$$

and where h is an unknown fading coefficient, and $\{z_n\}_{n=0}^{N-1}$ are i.i.d. zero-mean, Gaussian random variables with variance σ^2. Suppose we have two codewords of length N given by $\boldsymbol{x}^{(1)} = (a, 0, 0, \ldots, 0)$ and $\boldsymbol{x}^{(2)} = (0, b, 0, \ldots, 0)$. Note that any orthogonal code of two codewords can be transformed to this form. Since all of the coordinates of both codewords are zero for $n > 1$, the problem is essentially two dimensional.

The decoding regions for the GLRT decoder appear in Fig. 1. The GLRT projects the received signal (y_0, y_1) onto the directions of the two-dimensional vectors formed by the first two coordinates of $\boldsymbol{x}^{(1)}$ and $\boldsymbol{x}^{(2)}$, and decides according to the smaller between the distances of (y_0, y_1) to the vertical axis and to the horizontal axis of the coordinate system. The decoding rule decides $\boldsymbol{x}^{(1)}$ if $|y_0| \geq |y_1|$ and decides $\boldsymbol{x}^{(2)}$ if $|y_0| < |y_1|$. Thus, the boundaries between the two decision regions are straight lines through the origin at slopes of $\pm 45°$. Note that the decoding rule

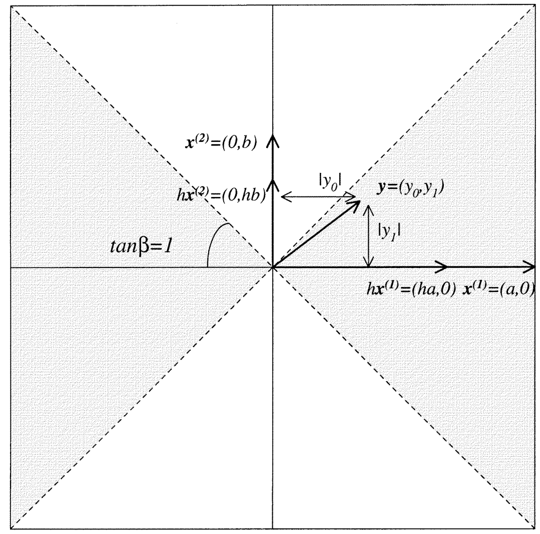

Fig. 1. Signal space diagram of the GLRT decoder.

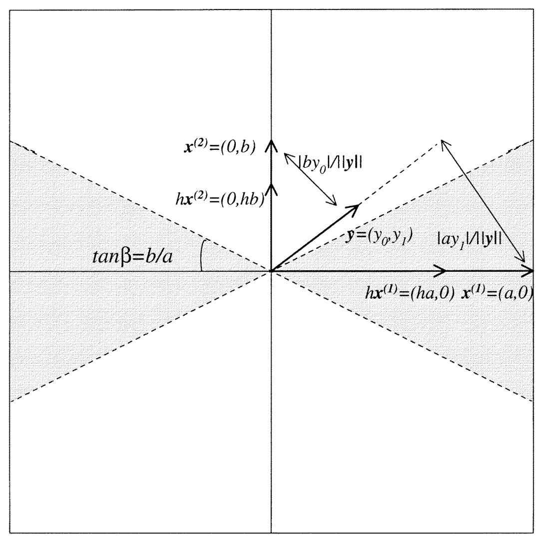

Fig. 2. Signal space diagram of the new decoder.

does not depend on the specific values of a and b. The distances of $h\cdot(a, 0)$ and $h\cdot(0, b)$ from the boundary lines dictate the error probability for the decoder. The distance d_1 of $h\cdot(a, 0)$ from the boundary lines at slope ± 1 is $ha/\sqrt{2}$ and the distance d_2 of $h\cdot(0, b)$ from the same lines is $hb/\sqrt{2}$. The leading term of the error probability behaves as $\exp\{-h^2 \min\{a^2, b^2\}/(4\sigma^2)\}$.

Following [12, the Appendix], the decoding regions of the new decoder appear in Fig. 2. This decoder projects the vector formed by the first two coordinates of each $\boldsymbol{x}^{(m)}$ in the direction of the first two coordinates of $\boldsymbol{y}$. The decoding rule decides $\boldsymbol{x}^{(1)}$ if $|y_0/a| \geq |y_1/b|$ and decides $\boldsymbol{x}^{(2)}$ if $|y_0/a| < |y_1/b|$. The boundary between the two decision regions is a pair of straight lines with slopes $\pm b/a$. For the new decoder, the distance of both $h\cdot(a, 0)$ and $h\cdot(0, b)$ from the boundary lines is $hab/\sqrt{a^2+b^2}$. Thus, the error probability has exponential order of $\exp\{-h^2a^2b^2/[2\sigma^2(a^2+b^2)]\}$, which is strictly better than that of the GLRT for any h, unless $a = b$.

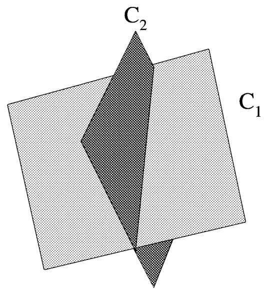

Fig. 3. Codewords hyperplanes for $N = 3$, $K = 2$.

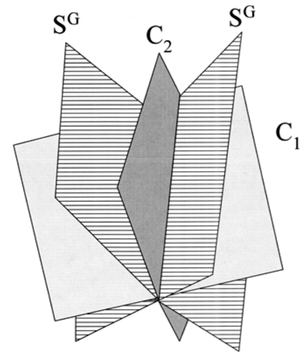

Fig. 4. GLRT separating surface for $N = 3$, $K = 2$.

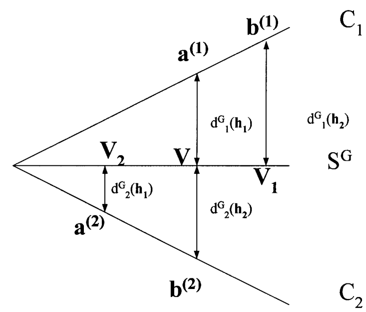

Fig. 5. Cross section for $N = 3$.

V. ULRT FOR A SPECIAL ISI CASE

A. Preliminaries

In this section, we analyze the special ISI case, with two codewords and where the ISI order is $K = N - 1$. A preliminary presentation of the ULRT for this case was given in [17], [18].

In this case, if $X_m(m = 1, 2)$ are full rank, $C_m(m = 1, 2)$ in (19) represent hyperplanes that pass through the origin. The intersection of the two hyperplanes is a subspace of dimension $N-2 = K-1$. As illustrated in Fig. 3, when $N = 3$, C_m, $m = 1, 2$ are planes and their intersection is a line. Thus, for $m = 1, 2$ we can find $\boldsymbol{p}^{(m)} \in \Re^N$ s.t.

$$C_m = \left\{ \boldsymbol{x} \in \Re^N : \boldsymbol{p}^{(m)T}\boldsymbol{x} = 0,\, \boldsymbol{p}^{(m)} \in \Re^N \right\}. \tag{32}$$

The distance between the hyperplane C_m and a vector $\boldsymbol{y} \in \Re^N$ is

$$d(\boldsymbol{y}, C_m) = \frac{\left|\boldsymbol{y}^T\boldsymbol{p}^{(m)}\right|}{\left\|\boldsymbol{p}^{(m)}\right\|}. \tag{33}$$

The GLRT metrics are given by $d(\boldsymbol{y}, C_m)$ according to definition. The GLRT separating surfaces are therefore two hyperplanes given by

$$S^G = \left\{ \boldsymbol{v}: \left(\frac{\boldsymbol{p}^{(1)}}{\left\|\boldsymbol{p}^{(1)}\right\|} \pm \frac{\boldsymbol{p}^{(2)}}{\left\|\boldsymbol{p}^{(2)}\right\|} \right)^T \boldsymbol{v} = 0 \right\} \tag{34}$$

or equivalently

$$S^G = \{\boldsymbol{v}: \boldsymbol{p}^T\boldsymbol{v} = 0\} \tag{35}$$

where

$$\boldsymbol{p} = \frac{\boldsymbol{p}^{(1)}}{\left\|\boldsymbol{p}^{(1)}\right\|} \pm \frac{\boldsymbol{p}^{(2)}}{\left\|\boldsymbol{p}^{(2)}\right\|}. \tag{36}$$

See illustration for $N = 3$, $K = 2$ in Fig. 4.

The normal to the surface at some $\boldsymbol{v} \in S^G$ intersects C_1 at $\boldsymbol{a}^{(1)}$ and C_2 at $\boldsymbol{b}^{(2)}$. Fig. 5 illustrates a cross section of the hyperplanes for $N = 3$.

For X_1, X_2 full rank, we can find (unique) ISI parameters $\boldsymbol{h_1}$ and $\boldsymbol{h_2}$ s.t.

$$\boldsymbol{a}^{(1)} = X_1\boldsymbol{h_1} \quad \text{and} \quad \boldsymbol{b}^{(2)} = X_2\boldsymbol{h_2}. \tag{37}$$

According to definitions (23) and (24)

$$d_1^G(\boldsymbol{h_1}) = \left\|\boldsymbol{a}^{(1)} - \boldsymbol{v}\right\| \tag{38}$$

$$d_2^G(\boldsymbol{h_2}) = \left\|\boldsymbol{b}^{(2)} - \boldsymbol{v}\right\|. \tag{39}$$

Define

$$\boldsymbol{a}^{(2)} = X_2\boldsymbol{h_1} \quad \text{and} \quad \boldsymbol{b}^{(1)} = X_1\boldsymbol{h_2}. \tag{40}$$

We can find $\boldsymbol{v_1}, \boldsymbol{v_2} \in S^G$ such that according to definitions (23) and (24)

$$d_2^G(\boldsymbol{h_1}) = \left\|\boldsymbol{a}^{(2)} - \boldsymbol{v_2}\right\| \tag{41}$$

$$d_1^G(\boldsymbol{h_2}) = \left\|\boldsymbol{b}^{(1)} - \boldsymbol{v_1}\right\|. \tag{42}$$

We make the following assumption on the code:

$$d_1^G(\boldsymbol{h_1}) > d_2^G(\boldsymbol{h_1}) \tag{43}$$

$$d_1^G(\boldsymbol{h_2}) > d_2^G(\boldsymbol{h_2}) \tag{44}$$

where the inequalities are strict.

Note that we could have chosen, without loss of generality, the same assumptions with both inequality signs reversed. If we cannot find any $\boldsymbol{v}$ such that these assumptions hold, we show in Section V-E that there is no decoder that uniformly improves the GLRT. Now, one can easily find examples where the assumptions hold for some region of S^G. For instance, define the sur-

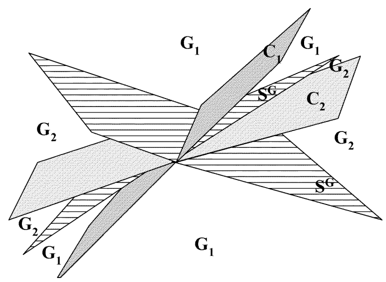

Fig. 6. The planes C_1, C_2, and S^G.

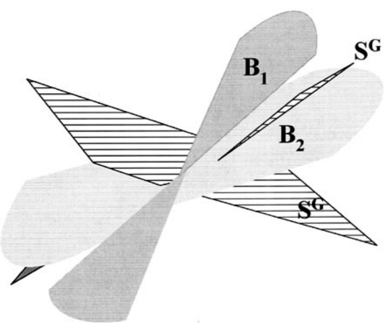

Fig. 7. The regions B_1 and B_2.

face $S_a(\boldsymbol{h_1}, \boldsymbol{h_2}) = (d_1^G(\boldsymbol{h_1}) - d_2^G(\boldsymbol{h_1}))\,(d_1^G(\boldsymbol{h_2}) - d_2^G(\boldsymbol{h_2}))$ where

$$X_1 = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix}, \qquad X_2 = \begin{bmatrix} 0.7 & 0 \\ 1.23 & 0.7 \\ 0 & 1.23 \end{bmatrix}. \tag{45}$$

One can verify that $S_a(\boldsymbol{h_1}, \boldsymbol{h_2})$ is positive for some values of $\boldsymbol{h_1}, \boldsymbol{h_2}$. Clearly, both assumptions (43) and (44) (or both with inequality signs reversed) hold for $\boldsymbol{h_1}, \boldsymbol{h_2}$ s.t. $S_a(\boldsymbol{h_1}, \boldsymbol{h_2})$ is positive.

Under these assumptions, we will find a decoder that uniformly improves (the exponential order of the error probability of) the GLRT. Interestingly, in the example in (45), the energies of the codewords are equal, yet the above assumptions hold, and therefore, according to the proof in Section IV-C, the GLRT may be uniformly improved. This is contrary to the fading example, where the GLRT is uniformly improved only when the codewords have unequal energy.

B. Example

Before getting into the formal proof we provide an example in order to demonstrate the general idea behind the construction of the ULRT. Fig. 6 shows the planes C_1, C_2, and the separating surfaces of the GLRT decoder S^G for $K = 2$ and $N = 3$. The decoding regions of $m = 1$ and $m = 2$ are denoted as G_1 and G_2, respectively. In Fig. 7, we have drawn around each point $X_1\boldsymbol{h} \in C_1$ a ball $\mathcal{B}_1(\boldsymbol{h})$ of radius $d_2^G(\boldsymbol{h})$ as defined in (24) and a ball $\mathcal{B}_2(\boldsymbol{h})$ of radius $d_1^G(\boldsymbol{h})$ as defined in (23) around each point $X_2\boldsymbol{h} \in C_2$. We define the regions

$$B_1 = \bigcup_{X_1\boldsymbol{h}\in C_1} \mathcal{B}_1(\boldsymbol{h}) \tag{46}$$

$$B_2 = \bigcup_{X_2\boldsymbol{h}\in C_2} \mathcal{B}_2(\boldsymbol{h}). \tag{47}$$

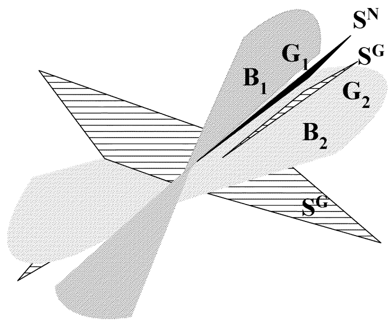

Fig. 8. The new separating surface.

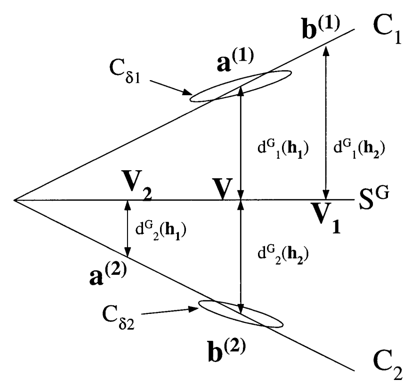

Fig. 9. Cross section for $N = 3$.

In Fig. 8, we observe that we can map the surface S^G to a new surface S^N (not necessarily a plane) such that S^N is outside B_1 and outside G_2, which together guarantee that the decoder maintains condition (29) and is within $B_2\backslash G_2$ which guarantees that (30) is maintained.

C. Formal Construction

We now return to a rigorous formulation. The assumptions (43) and (44) can be reformulated into

$$d_1^G(\boldsymbol{h_1}) = d_2^G(\boldsymbol{h_1}) + \alpha(\boldsymbol{h_1}) \tag{48}$$

$$d_1^G(\boldsymbol{h_2}) = d_2^G(\boldsymbol{h_2}) + \beta(\boldsymbol{h_2}) \tag{49}$$

where $\alpha(\boldsymbol{h_1}) > 0$ and $\beta(\boldsymbol{h_2}) > 0$ and finite.

Define the circle C_{δ_1} (see Fig. 9 for illustration) where $\delta_1 > 0$ and finite

$$C_{\delta_1} = \left\{X_1\boldsymbol{h}\colon X_1\boldsymbol{h} = \boldsymbol{a}^{(1)} + \boldsymbol{r},\ \|\boldsymbol{r}\| < \delta_1\right\} \subset C_1. \tag{50}$$

The distance function is continuous with respect to $X_1\boldsymbol{h}$ and $X_2\boldsymbol{h}$ and given by $d_m^G(\boldsymbol{h}) = |\boldsymbol{p^T}X_m\boldsymbol{h}|/\|\boldsymbol{p}\|$, $m = 1, 2$ (where $\boldsymbol{p}$ is defined in (36)). We can, therefore, find δ_1 finite and small enough and some $\alpha(\boldsymbol{h}) > 0$ s.t.

$$d_1^G(\boldsymbol{h}) = d_2^G(\boldsymbol{h}) + \alpha(\boldsymbol{h}), \qquad \alpha(\boldsymbol{h}) > 0 \tag{51}$$

for all $\boldsymbol{h}$ s.t. $X_1\boldsymbol{h} \in C_{\delta_1}$.

Any new separating surface has to pass through some point on the line between $\boldsymbol{a}^{(1)}$ and $\boldsymbol{b}^{(2)}$. The GLRT passes through $\boldsymbol{v}$. We look for a mapping of $\boldsymbol{v}$ to another point $\boldsymbol{u}$ that is between $\boldsymbol{v}$ and $\boldsymbol{a}^{(1)}$, for the ULRT. In other words

$$\boldsymbol{u} = \epsilon(\boldsymbol{v})\boldsymbol{a}^{(1)} + (1 - \epsilon(\boldsymbol{v}))\boldsymbol{v} \tag{52}$$

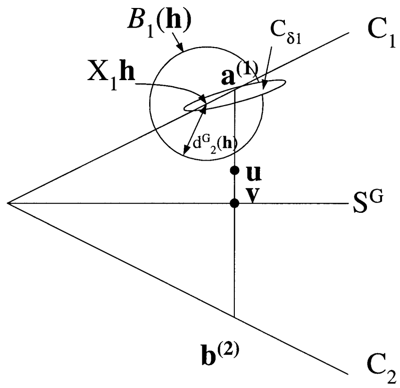

Fig. 10. Illustration for $N = 3$ (continued).

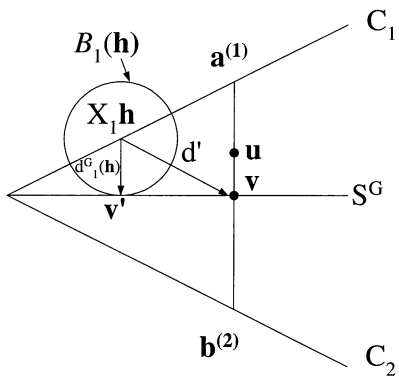

Fig. 11. Illustration for $N = 3$ (continued).

for some $0 < \epsilon(\boldsymbol{v}) < 1$. We first show that there exists $0 < \epsilon(\boldsymbol{v}) < 1$ small enough and finite such that the new surface will not worsen the exponential error for all possible $\boldsymbol{h}$. That is, $\forall \boldsymbol{h}$, the vector $\boldsymbol{u}$ is strictly outside the balls of radius $\min\{d_1^G(\boldsymbol{h}), d_2^G(\boldsymbol{h})\}$ around $X_1\boldsymbol{h}$ and around $X_2\boldsymbol{h}$. Clearly, the point $\boldsymbol{u}$, which is in the decision region (of the GLRT decoder) of the codeword $m = 1$, is outside a ball of radius $\min\{d_1^G(\boldsymbol{h}), d_2^G(\boldsymbol{h})\}$ around $X_2\boldsymbol{h}$. It remains to show that $\boldsymbol{u}$ is also outside a ball of radius $\min\{d_1^G(\boldsymbol{h}), d_2^G(\boldsymbol{h})\}$ around $X_1\boldsymbol{h}$.

We split C_1 into two sets. The first set contains all $X_1\boldsymbol{h} \in C_{\delta_1}$. For any $X_1\boldsymbol{h} \in C_{\delta_1}$ we define a ball of radius $d_2^G(\boldsymbol{h})$ around $X_1\boldsymbol{h}$

$$\mathcal{B}_1(\boldsymbol{h}) = \{\boldsymbol{r}\colon \|X_1\boldsymbol{h} - \boldsymbol{r}\| < d_2^G(\boldsymbol{h})\}. \tag{53}$$

See Fig. 10 for illustration. Since by (51), $d_2^G(\boldsymbol{h}) < d_1^G(\boldsymbol{h})$, the surface S^G is strictly separated from $\mathcal{B}_1(\boldsymbol{h})$. Therefore, $\boldsymbol{v}$ is strictly outside $\mathcal{B}_1(\boldsymbol{h})$. It follows that for any $\boldsymbol{h}$ s.t. $X_1\boldsymbol{h} \in C_{\delta_1}$, there exists a finite $\epsilon(\boldsymbol{v}, \boldsymbol{h}) > 0$ s.t. $\boldsymbol{u}$ defined in (52) is also strictly outside $\mathcal{B}_1(\boldsymbol{h})$.

The second set contains all $X_1\boldsymbol{h} \notin C_{\delta_1}$. We can, therefore, find finite $t > 0$ and unit vector $\boldsymbol{n}$ s.t.

$$X_1\boldsymbol{h} = \boldsymbol{a}^{(1)} + t\boldsymbol{n}. \tag{54}$$

Let $\boldsymbol{v}'$ be the projection of $X_1\boldsymbol{h}$ on S^G (see Fig. 11). The distance of $X_1\boldsymbol{h}$ from S^G is

$$d_1^G(\boldsymbol{h}) = \|X_1\boldsymbol{h} - \boldsymbol{v}'\|. \tag{55}$$

The vectors $\boldsymbol{v}$ and $\boldsymbol{v}'$ can be expressed as

$$\boldsymbol{v} = \boldsymbol{a}^{(1)} - \frac{\boldsymbol{a}^{(1)T}\boldsymbol{p}}{\boldsymbol{p}^T\boldsymbol{p}}\boldsymbol{p} \tag{56}$$

$$\boldsymbol{v}' = X_1\boldsymbol{h} - \frac{(X_1\boldsymbol{h})^T\boldsymbol{p}}{\boldsymbol{p}^T\boldsymbol{p}}\boldsymbol{p} \tag{57}$$

where $\boldsymbol{p}$ is defined in (36). It follows from (54), (56), and (57) that

$$\|\boldsymbol{v} - \boldsymbol{v}'\| = t\left(\boldsymbol{n} - \frac{\boldsymbol{n}^T\boldsymbol{p}}{\boldsymbol{p}^T\boldsymbol{p}}\boldsymbol{p}\right). \tag{58}$$

According to (54), the vector $\boldsymbol{n}$ is in the direction of $X_1\boldsymbol{h} - \boldsymbol{a}^{(1)}$ and according to (56), the vector $\boldsymbol{p}$ is in the direction of $\boldsymbol{a}^{(1)} - \boldsymbol{v}$. The three vectors $\boldsymbol{a}^{(1)} \in C_1$, $X_1\boldsymbol{h} \in C_1$, and $\boldsymbol{v} \notin C_1$ are not all on the same line. Therefore, $\boldsymbol{n} \neq \boldsymbol{p}/\|\boldsymbol{p}\|$ and $\boldsymbol{n} - (\boldsymbol{n}^T\boldsymbol{p})/(\boldsymbol{p}^T\boldsymbol{p})\boldsymbol{p} \neq 0$. It follows from (58) that since t is finite $\|\boldsymbol{v} - \boldsymbol{v}'\| > 0$ is also finite.

Denote by d' the distance of $X_1\boldsymbol{h}$ from $\boldsymbol{v}$. Since $\|\boldsymbol{v} - \boldsymbol{v}'\| > 0$ is finite it follows from the triangle inequality that there exists finite $\alpha(\boldsymbol{h}) > 0$ s.t.

$$d' = d_1^G(\boldsymbol{h}) + \alpha(\boldsymbol{h}). \tag{59}$$

For any $X_1\boldsymbol{h} \notin C_{\delta_1}$, we define a ball of radius $d_1^G(\boldsymbol{h})$ around $X_1\boldsymbol{h}$

$$\mathcal{B}_1(\boldsymbol{h}) = \{\boldsymbol{r}\colon \|X_1\boldsymbol{h} - \boldsymbol{r}\| < d_1^G(\boldsymbol{h})\}. \tag{60}$$

Since $d' > d_1^G(\boldsymbol{h})$, the vector $\boldsymbol{v}$ is strictly outside $\mathcal{B}_1(\boldsymbol{h})$. Thus, for any $\boldsymbol{h}$ s.t. $X_1\boldsymbol{h} \notin C_{\delta_1}$, there exists a finite $\epsilon(\boldsymbol{v}, \boldsymbol{h}) > 0$ s.t. $\boldsymbol{u}$ is also strictly outside $\mathcal{B}_1(\boldsymbol{h})$.

For the value of $\epsilon(\boldsymbol{v})$ in (52) a possible choice would be:

$$\epsilon(\boldsymbol{v}) = \min_{\boldsymbol{h}} \epsilon(\boldsymbol{v}, \boldsymbol{h}). \tag{61}$$

Since $\epsilon(\boldsymbol{v}, \boldsymbol{h})$ was shown to be strictly positive and finite $\forall h$, the choice for $\epsilon(\boldsymbol{v})$ in (61) is also positive and finite. We note that this choice for $\epsilon(\boldsymbol{v})$ is not necessarily optimal and is not unique, but it does guarantee that the error probability for any possible $\boldsymbol{h}$ will not be worsened as a result of the mapping.

So far $\boldsymbol{v}$ was mapped to $\boldsymbol{u}$ without worsening the error probability for any possible $\boldsymbol{h}$. We wish now to map an entire area around $\boldsymbol{v}$ to an area around $\boldsymbol{u}$. To that end we define now a circle of radius $\delta' < \delta_1$

$$C_{\delta'} = \left\{X_1\boldsymbol{h}'\colon X_1\boldsymbol{h}' = \boldsymbol{a}^{(1)} + \boldsymbol{r}, \|\boldsymbol{r}\| < \delta'\right\} \subset C_{\delta_1} \tag{62}$$

(see Fig. 12). The definition of $C_{\delta'}$ means that any $X_1\boldsymbol{h}' \in C_{\delta'}$ not only maintains $d_1^G(\boldsymbol{h}') > d_2^G(\boldsymbol{h}')$ but also has a set around it that also maintains the same condition. The existence of the set $C_{\delta'}$ follows again from the continuity of the distance function and the fact that $\alpha(\boldsymbol{h_1})$ is a finite positive number.

Any of the vectors $X_1\boldsymbol{h}' \in C_{\delta'}$ maintains the same conditions as $X_1\boldsymbol{h_1}$ does. That is, for all $X_1\boldsymbol{h}' \in C_{\delta'}$ there is $\alpha(\boldsymbol{h}') > 0$ s.t.

$$d_1^G(\boldsymbol{h}') = d_2^G(\boldsymbol{h}') + \alpha(\boldsymbol{h}') \tag{63}$$

and, in addition, there is a circle of a finite radius $\delta(\boldsymbol{h}') < \delta'$ around $X_1\boldsymbol{h}'$

$$C_{\delta(\boldsymbol{h}')} = \{X_1\boldsymbol{h}\colon X_1\boldsymbol{h} = X_1\boldsymbol{h}' + \boldsymbol{r}, \|\boldsymbol{r}\| < \delta(\boldsymbol{h}')\} \subset C_{\delta'} \tag{64}$$

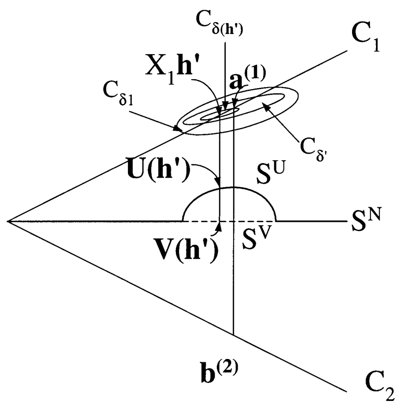

Fig. 12. Illustration for $N = 3$ (continued).

that also maintains for all $\boldsymbol{h}$ s.t. $X_1\boldsymbol{h} \in C_{\delta(\boldsymbol{h}')}$

$$d_1^G(\boldsymbol{h}) = d_2^G(\boldsymbol{h}) + \alpha(\boldsymbol{h}), \qquad \alpha(\boldsymbol{h}) > 0. \tag{65}$$

The projection vector of a certain $X_1\boldsymbol{h}' \in C_{\delta'}$ on S^G is $\boldsymbol{v}(\boldsymbol{h}')$. Using the same arguments as for $\boldsymbol{h_1}$, the vector $\boldsymbol{v}(\boldsymbol{h}')$ can be mapped to $\boldsymbol{u}(\boldsymbol{h}')$

$$\boldsymbol{u}(\boldsymbol{h}') = \epsilon(\boldsymbol{v}(\boldsymbol{h}'))X_1\boldsymbol{h}' + (1 - \epsilon(\boldsymbol{v}(\boldsymbol{h}')))\boldsymbol{v}(\boldsymbol{h}') \tag{66}$$

for some $0 < \epsilon(\boldsymbol{v}(\boldsymbol{h}')) < 1$. The parameter $0 < \epsilon(\boldsymbol{v}(\boldsymbol{h}')) < 1$ is finite and will be chosen such that for any possible $\boldsymbol{h} \in \mathcal{R}^K$, the vector $\boldsymbol{u}(\boldsymbol{h}')$ will be strictly outside a ball of radius $\min\{d_1^G(\boldsymbol{h}), d_2^G(\boldsymbol{h})\}$ around $X_1\boldsymbol{h}$. In summary, then, we have mapped an area of the separating surface without worsening the error probability exponential order for any possible channel vector.

Now it remains to show how the error probability for $\boldsymbol{h_2}$ has been modified. First, we have not worsened the error probability for $\boldsymbol{h_2}$ since the new surface is strictly separated, by construction, from balls $\mathcal{B}_1(\boldsymbol{h_2})$ and $\mathcal{B}_2(\boldsymbol{h_2})$ of radius $\min\{d_1^G(\boldsymbol{h_2}), d_2^G(\boldsymbol{h_2})\} = d_2^G(\boldsymbol{h_2})$ around both $\boldsymbol{b}^{(2)} = X_2\boldsymbol{h_2}$ and $\boldsymbol{b}^{(1)} = X_1\boldsymbol{h_2}$

$$\mathcal{B}_1(\boldsymbol{h_2}) = \{\boldsymbol{r}: \|X_1\boldsymbol{h_2} - \boldsymbol{r}\| < d_2^G(\boldsymbol{h_2})\} \tag{67}$$

$$\mathcal{B}_2(\boldsymbol{h_2}) = \{\boldsymbol{r}: \|X_2\boldsymbol{h_2} - \boldsymbol{r}\| < d_2^G(\boldsymbol{h_2})\}. \tag{68}$$

We show now that we have improved the error probability for $\boldsymbol{h_2}$. Denote by $S^V \subset S^G$ the area around $\boldsymbol{v}$ that was mapped to a different area (around $\boldsymbol{u}$). Denote by S^U the corresponding mapped area around $\boldsymbol{u}$ (see Fig. 13). The separating surface of the ULRT is defined by $(S^G \setminus S^V) \cup S^U$. For $S^G \setminus S^V$

$$d(X_2\boldsymbol{h_2}, S^G \setminus S^V) > d_2^G(\boldsymbol{h_2}) \tag{69}$$

according to the construction (where $d(\boldsymbol{v}, S) = \min_{\boldsymbol{v}' \in S} \|\boldsymbol{v}' - \boldsymbol{v}\|$). For S^U

$$d(X_2\boldsymbol{h_2}, S^U) > d_2^G(\boldsymbol{h_2}) \tag{70}$$

since every $\boldsymbol{u}' \in S^U$ is in the decision area (of the GLRT decoder) of $m = 1$. Since there is an area around $\boldsymbol{v}$ with finite radius that was mapped to a surface with larger distance from $\boldsymbol{b}^{(2)}$, there is a ball around $\boldsymbol{b}^{(2)}$ with radius $R_2 > d_2^G(\boldsymbol{h_2})$ that is strictly separated from the new separating surface. See Fig. 13 for illustration. As for $\boldsymbol{b}^{(1)} = X_1\boldsymbol{h_2}$

$$d(X_1\boldsymbol{h_2}, S^N) > d_2^G(\boldsymbol{h_2}) \tag{71}$$

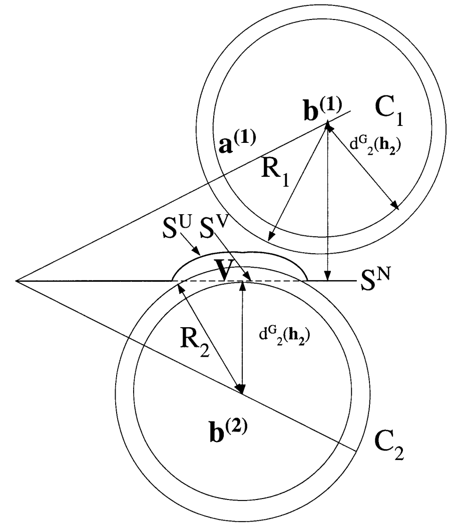

Fig. 13. Illustration for $N = 3$ (continued).

according to the construction, and the inequality is strict. Therefore, there is a ball of radius $d_2^G(\boldsymbol{h_2}) < R_1 \le d_1^G(\boldsymbol{h_2})$ around $\boldsymbol{b}^{(1)} = X_1\boldsymbol{h_2}$ that is strictly separated from the new separating surface. Thus, the distances of both $\boldsymbol{b}^{(1)}$ and $\boldsymbol{b}^{(2)}$ from the new surface is greater than $d_2^G(\boldsymbol{h_2})$ and, therefore, the error probability for $\boldsymbol{h_2}$ is improved.

Note that the procedure can be repeated for any $\boldsymbol{v} \in S^G$ which maintains assumptions (43) and (44) (or with both inequality signs reversed) and the separating surface can be modified accordingly. Thus, the error probability can be improved for additional channel coefficients as well without worsening the error probability for any possible channel coefficient.

The decoding is performed in the following way. Assume that the vector $\boldsymbol{y}$ is received. The projection vector of $\boldsymbol{y}$ on S^G intersects S^G, C_1, C_2 at $\boldsymbol{v}$, $\boldsymbol{a}^{(1)}$, $\boldsymbol{b}^{(2)}$ respectively. The channel coefficients corresponding to $\boldsymbol{a}^{(1)}$ and $\boldsymbol{b}^{(2)}$ are $\boldsymbol{h_1}$ and $\boldsymbol{h_2}$ respectively. The vectors corresponding to the same channel coefficients and the other codeword are $\boldsymbol{a}^{(2)}$ and $\boldsymbol{b}^{(1)}$, respectively. The distances $d_1^G(\boldsymbol{h_1})$, $d_1^G(\boldsymbol{h_2})$, $d_2^G(\boldsymbol{h_1})$, $d_2^G(\boldsymbol{h_2})$ can now be calculated and conditions (43) and (44) (or both with inequality reversed) verified. If (43) and (44) hold and assuming $\epsilon(\boldsymbol{v})$ is given, we can find $\boldsymbol{u}$ by (52). If $\boldsymbol{y}$ is on the line between $\boldsymbol{a}^{(1)}$ and $\boldsymbol{u}$ we decode $\hat{m} = 1$ otherwise $\hat{m} = 2$. Note that this decoding rule depends on $\epsilon(\boldsymbol{v})$. It turns out, however, that the optimal $\epsilon(\boldsymbol{v})$ is complicated to find.

The application of the ULRT to the simple fading example discussed in Section IV is illustrated in Fig. 14. The line S^G represents the GLRT and S^T represents the decoder described in IV. The line S^T corresponds to the choice

$$\epsilon^*(\boldsymbol{v}) = 1 - \sqrt{2}b \Big/ \sqrt{a^2 + b^2} \tag{72}$$

where $\epsilon^*(\boldsymbol{v})$ is the optimal choice for $\epsilon(\boldsymbol{v})$ in (52). Note that in the fading example $\epsilon^*(\boldsymbol{v})$ does not depend on $\boldsymbol{v}$ and the resulting separating line is a straight line. The line S^N corresponds to a different choice of $\epsilon(\boldsymbol{v})$ such that

$$0 < \epsilon(\boldsymbol{v}) < \epsilon_{\max} = \frac{\sqrt{2a^2 - b^2} - b}{\sqrt{2a^2 - b^2} + b}. \tag{73}$$

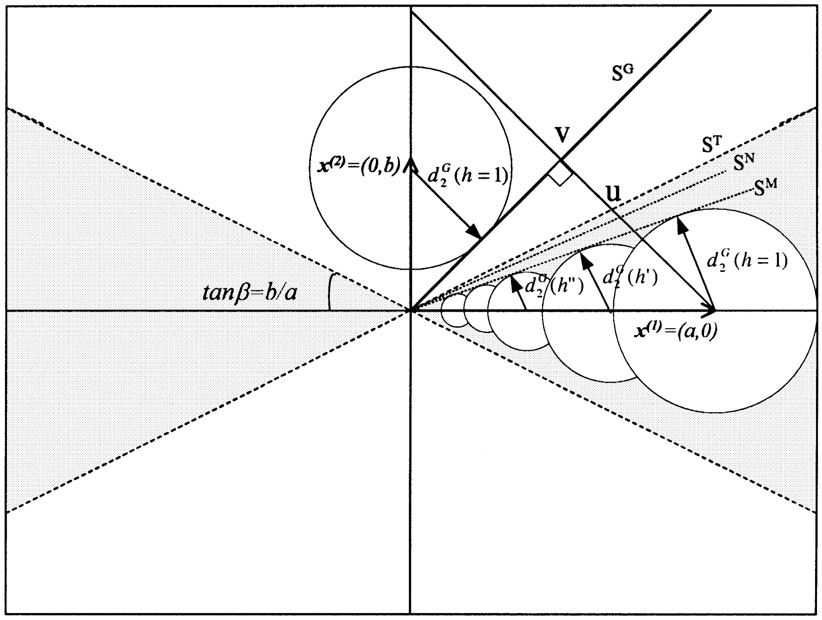

Fig. 14. Construction of the ULRT for the fading example.

Note that S^N is not optimal but it does improve the GLRT. The line S^M, corresponding to $\epsilon(\boldsymbol{v}) = \epsilon_{\max}$, is tangent to the circle of radius $d_2^G(h) = bh/\sqrt{2}$ around $a \cdot (h,\, 0), \forall\, h$ and, therefore, does not improve the GLRT. The optimal value $\epsilon(\boldsymbol{v}) = \epsilon^*(\boldsymbol{v})$ can be found via a search over the parameter $\epsilon(\boldsymbol{v})$.

In the Appendix, we explicitly present the structure of the ULRT for the hyperplane case. As previously mentioned, the value of $\epsilon(\boldsymbol{v})$ is not necessarily unique and determining its optimal value remains an open problem. Yet, we show in the Appendix that the optimal value of $\epsilon(\boldsymbol{v})$ is a function of only the direction of $\boldsymbol{v}$ and is independent of its magnitude. Thus, it turns out that the new surface consists of straight lines that emerge from the origin and together form a surface that is not a plane.

Another way to formulate our decision rule is as follows. Assume that $\boldsymbol{y}$ is in the decision region of codeword $m = 1$ for the GLRT decoder. Then if $d_1^G(\boldsymbol{h_1}) > d_2^G(\boldsymbol{h_1})$ and $d_1^G(\boldsymbol{h_2}) > d_2^G(\boldsymbol{h_2})$ the decision rule is

$$\frac{\left\|\boldsymbol{y} - \boldsymbol{b}^{(2)}\right\|}{\left\|\boldsymbol{y} - \boldsymbol{a}^{(1)}\right\| + 2\epsilon_1(\boldsymbol{v}) d_1^G(\boldsymbol{h_1})} \underset{\mathcal{H}_2}{\overset{\mathcal{H}_1}{\gtrless}} 1. \tag{74}$$

For the simulations we used a special case of this decoder, where $\epsilon_1(\boldsymbol{v}) = \xi_1(d_1^G(\boldsymbol{h_1}) - d_2^G(\boldsymbol{h_1}))/2d_1^G(\boldsymbol{h_1})$ and ξ_1 is a constant parameter of the decoder to be optimized so that the decoder would uniformly improve the error probability. Similarly, assume that $\boldsymbol{y}$ is in the decision region of codeword $m = 2$ for the GLRT decoder. Then if $d_1^G(\boldsymbol{h_1}) < d_2^G(\boldsymbol{h_1})$ and $d_1^G(\boldsymbol{h_2}) < d_2^G(\boldsymbol{h_2})$ the decision rule is

$$\frac{\left\|\boldsymbol{y} - \boldsymbol{b}^{(2)}\right\| + 2\epsilon_2(\boldsymbol{v}) d_2^G(\boldsymbol{h_2})}{\left\|\boldsymbol{y} - \boldsymbol{a}^{(1)}\right\|} \underset{\mathcal{H}_2}{\overset{\mathcal{H}_1}{\gtrless}} 1. \tag{75}$$

Again, for the simulations we used a special case of this decoder, where $\epsilon_2(\boldsymbol{v}) = \xi_2(d_2^G(\boldsymbol{h_2}) - d_1^G(\boldsymbol{h_2}))/2d_2^G(\boldsymbol{h_2})$ and ξ_2 is a constant parameter of the decoder to be optimized so that the decoder would uniformly improve the error probability.

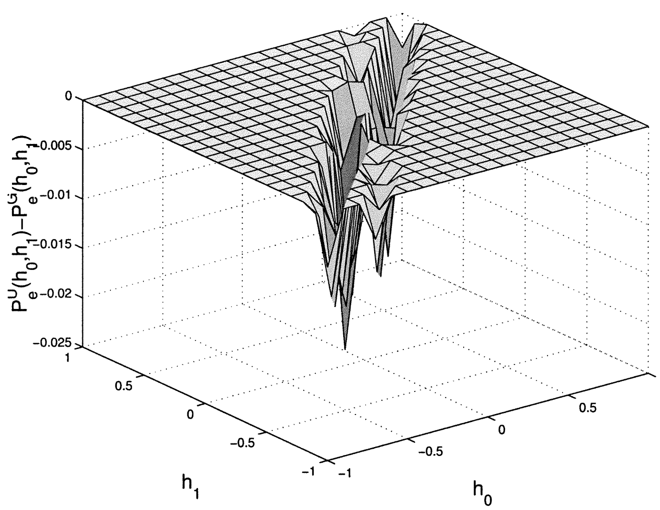

Fig. 15. Comparison between the GLRT and the ULRT for $N = 3, K = 2, M = 2$.

Fig. 15 compares the performance of the GLRT and the ULRT for a specific code with two codewords. The error probability for a certain choice of the parameter vector $(h_0,\, h_1)$ is given by $P_e^U(h_0,\, h_1)$ for the ULRT and by $P_e^G(h_0,\, h_1)$ for the GLRT. The graph shows the difference $P_e^U(h_0,\, h_1) - P_e^G(h_0,\, h_1)$. We see that for all $(h_0,\, h_1)$, $P_e^U(h_0,\, h_1) - P_e^G(h_0,\, h_1) \le 0$ with strict inequality for some $(h_0,\, h_1)$ and, therefore, the improvement is uniform. The values of the parameters $\xi_1,\, \xi_2$ were optimized by a search over a grid. The values chosen $\xi_1,\, \xi_2 = 0.2$ give optimal average performance (over the channel parameter space) while still uniformly improving the performance of the GLRT.

D. Hyperplane Case With M Codewords

Suppose we have M codewords, and each of the codewords represents a hyperplane. We assume that the codewords are

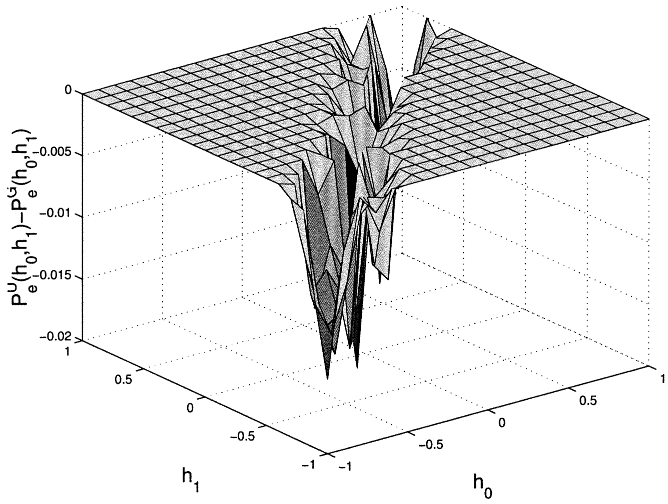

Fig. 16. Comparison between the GLRT and the ULRT for $N = 3, K = 2, M = 5$.

chosen such that all of the hyperplanes have the same intersection. The angle between hyperplane i and hyperplane j is $\beta_{i,j}$. We can construct a vector $\boldsymbol{b}$ with M components defined by

$$b_j = \arg\min_{i \neq j} \{\beta_{i,j}\}, \qquad j = 1 \cdots M. \tag{76}$$

The decoding is carried out in the following way. First, we employ the GLRT decoder. Suppose the selected word is $m = i$. We then look in $\boldsymbol{b}$ for j s.t $b_j = i$. If no such j is found, the decision remains that of the GLRT. Otherwise, for such j, $P_e(j \to i|\boldsymbol{h}) > P_e(j \to k|\boldsymbol{h}), \forall \boldsymbol{h},\ k \neq i$, where $P_e(j \to k|\boldsymbol{h})$ denotes the probability that the decoded codeword is $m = k$ while the transmitted one is $m = j$ and the channel coefficients vector is $\boldsymbol{h}$. Therefore, $P(\text{error}|j, \boldsymbol{h})$ (the probability of error when the transmitted word is j and the channel coefficients vector is $\boldsymbol{h}$) is of the exponential order of $P_e(j \to i|\boldsymbol{h})$. Now if we carry out the procedure in the previous subsection for codewords i and j we would uniformly improve the error probability, which follows from the same arguments.

For the simulations we have used a simplified version of this algorithm. The codewords were chosen so that the hyperplanes they represent have a common intersection. We have calculated the GLRT metrics for all the codewords. We then selected the two codewords with the two minimal metrics and performed the simplified version of the ULRT from Section V-C for these two codewords. Thus, existing GLRT decoders could be incorporated into the ULRT decoders. Fig. 16 compares the performance of the GLRT and the ULRT for a specific code with $M = 5$ codewords. We see that the improvement is uniform. The parameters ξ_1 and ξ_2 in (74) and (75), respectively, were optimized for each pair of codewords separately, by a search over a grid.

E. Converse Theorem

We prove now that the existence of $\boldsymbol{v} \in S^G$ such that both (43) and (44) hold (or both assumptions with inequality signs reversed) is also a necessary condition for the existence of a decoder that uniformly improves the GLRT decoder (for which (29) and (30) hold).

Assume there is no $\boldsymbol{v} \in S^G$ such that both (43) and (44) hold (or both assumptions with inequality signs reversed). Therefore, there are now only two possible cases for each $\boldsymbol{v} \in S^G$.

In case I

$$d_1^G(\boldsymbol{h_1}) < d_2^G(\boldsymbol{h_1}) \tag{77}$$

$$d_1^G(\boldsymbol{h_2}) > d_2^G(\boldsymbol{h_2}). \tag{78}$$

In case II

$$d_1^G(\boldsymbol{h_1}) > d_2^G(\boldsymbol{h_1}) \tag{79}$$

$$d_1^G(\boldsymbol{h_2}) < d_2^G(\boldsymbol{h_2}) \tag{80}$$

where $\boldsymbol{h_1}$ and $\boldsymbol{h_2}$ are defined in (37). Refer to case I. Any decoder is defined by a separating surface. Any separating surface has to pass at some point between $\boldsymbol{a}^{(1)}$ and $\boldsymbol{b}^{(2)}$. Clearly, in case I, the separating surface has to pass through $\boldsymbol{v}$ in order to maintain (29) for both $\boldsymbol{h_1}$ and $\boldsymbol{h_2}$. Therefore, we were not able to achieve a smaller error probability for $\boldsymbol{h_1}$ (and $\boldsymbol{h_2}$). Refer now to case II. Considering $\boldsymbol{a}^{(2)}$ (defined in (40)), we project it on S^G at point $\boldsymbol{v}'$, we define $\boldsymbol{d}^{(1)}$ to be the intersection of the difference vector $\boldsymbol{v}' - \boldsymbol{a}^{(2)}$ with C_1. We further define $\boldsymbol{h_3}$ to be the unique vector such that $\boldsymbol{d}^{(1)} = X_1\boldsymbol{h_3}$. Under our assumptions we have

$$d_1^G(\boldsymbol{h_1}) > d_2^G(\boldsymbol{h_1}) \tag{81}$$

$$d_1^G(\boldsymbol{h_3}) < d_2^G(\boldsymbol{h_3}). \tag{82}$$

Following the same arguments as in case I, we cannot map $\boldsymbol{v}'$ to a different point and, therefore, cannot improve the error probability of $\boldsymbol{h}_1$ (and $\boldsymbol{h}_2$) since any separating surface maintaining (29) has to pass through $\boldsymbol{v}'$. Since the above argument is valid for any $\boldsymbol{h}_1$ (and $\boldsymbol{h}_2$) the proof is complete.

VI. ULRT for the General ISI Case

As in the hyperplane case described in Section V, the construction of the ULRT for the general case is based on the GLRT decoder. Therefore, we begin this section by investigating the GLRT surface for the general case, in which the codewords span subspaces C_m, $m = 1, 2$ given in (19). Then, we present a decoding procedure, similar to that presented in Section V, with an additional assumption, made for simplicity, that the codewords span orthogonal subspaces.

The separating surface of the GLRT, S^G, is quadratic in the general case and given by (22). Define the matrices $A_m = 1, 2$

$$A_m = I - X_m(X_m^T X_m)^{-1} X_m^T, \qquad m = 1, 2. \tag{83}$$

Note that A_m is symmetric and idempotent, $A_m = A_m^T$, $A_m A_m = A_m$. Any vector $\boldsymbol{r_m} \in C_m$ for $m = 1$ or $m = 2$ satisfies

$$\begin{aligned} A_m\boldsymbol{r_m} &= (I - X_m(X_m^T X_m)^{-1} X_m^T)\boldsymbol{r_m} \\ &= (I - X_m(X_m^T X_m)^{-1} X_m^T) X_m\boldsymbol{h} \\ &= X_m\boldsymbol{h} - X_m\boldsymbol{h} = 0, \qquad m = 1, 2. \end{aligned} \tag{84}$$

Therefore, the subspaces C_m, $m = 1, 2$ can also be expressed as

$$C_m = \{\boldsymbol{r_m}: A_m\boldsymbol{r_m} = 0\}, \qquad m = 1, 2. \tag{85}$$

The separating surface is given by

$$S^G = \{\boldsymbol{v}: \boldsymbol{v}^T(A_2 - A_1)\boldsymbol{v} = 0\}. \tag{86}$$

Let $\boldsymbol{a}$ be a point on C_1. We analyze under what conditions $\|\boldsymbol{v} - \boldsymbol{a}\|$ represents the distance of $\boldsymbol{a}$ from S^G. A similar analysis can be performed for $\boldsymbol{b} \in C_2$. Consider the following (nonconvex) constrained optimization problem:

$$\min_{\boldsymbol{w}}\{\|\boldsymbol{a} - \boldsymbol{w}\|^2: \boldsymbol{w}^T(A_2 - A_1)\boldsymbol{w} = 0\}. \tag{87}$$

The constraint assures that the solution lies on S^G. The optimization problem can be relaxed to the following:

$$\min_{\boldsymbol{w}}\{\|\boldsymbol{a} - \boldsymbol{w}\|^2: \boldsymbol{w}^T(A_2 - A_1)\boldsymbol{w} \le 0\}. \tag{88}$$

The two problems are equivalent because the condition $\boldsymbol{w}^T(A_2 - A_1)\boldsymbol{w} < 0$ defines the decision region of $m = 2$. Therefore, the minimal distance of a point on C_1 to the region $\boldsymbol{w}^T(A_2 - A_1)\boldsymbol{w} \le 0$ is always achieved on the separating surface, where $\boldsymbol{w}^T(A_2 - A_1)\boldsymbol{w} = 0$. In what follows we state necessary conditions on the solution of (88).

Kuhn–Tucker conditions for a nonconvex constrained optimization problem

$$\min_{\boldsymbol{w}}\{f(\boldsymbol{w}): \boldsymbol{g}(\boldsymbol{w}) \le 0, \boldsymbol{w} \in \mathcal{R}^n\} \tag{89}$$

with $f: \mathcal{R}^n \to R$, $\boldsymbol{g}: \mathcal{R}^n \to \mathcal{R}^k$. Denote $I(\boldsymbol{w}) = \{i: g_i(\boldsymbol{w}) = 0\}$ i.e., the set of active constraints at $\boldsymbol{w}$ of the inequality constraints.

Let $\boldsymbol{w}^*$ be a local (global) minimum for (89). Assume that for $i \in I(\boldsymbol{w}^*)$, $\{\nabla g_i(\boldsymbol{w}^*)\}$ are linearly independent, where ∇ denotes the gradient operator (a point satisfying this condition is called regular). Then there exists a unique Lagrange vector λ^* satisfying

$$\begin{aligned} L(\boldsymbol{w}^*, \lambda^*) = \nabla f(\boldsymbol{w}^*) + \sum_{i=1}^{k} \lambda_i^* \nabla g_i(\boldsymbol{w}^*) = 0, \\ \lambda_i^* \ge 0,\ i = 1, \ldots, k, \quad \lambda_i^* = 0,\ \forall\, i \notin I(\boldsymbol{w}*). \end{aligned} \tag{90}$$

For the optimization problem in (88) $k = 1$. The Kuhn–Tucker conditions yield

$$\begin{aligned} L(\boldsymbol{v}, \lambda) &= -\boldsymbol{a} + \boldsymbol{v} + \lambda(A_2 - A_1)\boldsymbol{v} = 0 \\ \boldsymbol{a} &= [I + \lambda(A_2 - A_1)]\boldsymbol{v} \end{aligned} \tag{91}$$

where the Lagrange multiplier λ is nonnegative.

The gradient of the surface $\boldsymbol{w}^T(A_2 - A_1)\boldsymbol{w} = 0$ at $\boldsymbol{v}$ equals $(A_2 - A_1)\boldsymbol{v}$. In other words, the direction of the normal to the surface $\boldsymbol{w}^T(A_2 - A_1)\boldsymbol{w} = 0$ at $\boldsymbol{v}$ coincides with the direction of $(A_2 - A_1)\boldsymbol{v}$. Therefore, condition (91) is equivalent to requiring the vector $(\boldsymbol{a} - \boldsymbol{v})$ to be perpendicular to the separating surface S^G. In the general case, there could be several choices of perpendicular $\boldsymbol{v}$, each may be of different distance. The minimum of those projections is the global minimum.

We show now that each $\boldsymbol{v}$ is a regular solution of (88). For a single constraint, the requirement of linear independence of the Kuhn–Tucker condition reduces to the requirement that the gradient vector is not zero. In our case, for $\boldsymbol{v}$ to be regular we have to verify that $(A_2 - A_1)\boldsymbol{v} \neq 0$. Assume that $(A_2 - A_1)\boldsymbol{v} = 0$. Then from (91) it follows that $\boldsymbol{v} = \boldsymbol{a}$, which can occur only in the trivial case where $\boldsymbol{a}$ is in the intersection of C_1 and C_2. We thus conclude that $\boldsymbol{v}$ is regular.

Right-multiplying both sides of (91) by A_1 gives

$$\begin{aligned} A_1\boldsymbol{a} &= A_1[I + \lambda(A_2 - A_1)]\boldsymbol{v} \\ 0 &= A_1\boldsymbol{v} + \lambda A_1(A_2 - A_1)\boldsymbol{v} \end{aligned} \tag{92}$$

since $\boldsymbol{a} \in C_1$. In order for (92) to hold, $A_1\boldsymbol{v}$ and $A_1(A_2 - A_1)\boldsymbol{v}$ must be linearly dependent and λ must equal

$$\lambda = \frac{\|A_1\boldsymbol{v}\|}{\|A_1(A_2 - A_1)\boldsymbol{v}\|} \tag{93}$$

where λ was chosen to be nonnegative according to (90). For a certain $\boldsymbol{v}$, if $A_1\boldsymbol{v}$ and $A_1(A_2 - A_1)\boldsymbol{v}$ are not linearly **(?)** dependent, then the normal to S^G at $\boldsymbol{v}$ will not intersect C_1. Likewise, if $A_2\boldsymbol{v}$ and $A_2(A_1 - A_1)\boldsymbol{v}$ are not linearly **(?)** dependent, then the normal to S^G at $\boldsymbol{v}$ will not intersect C_2. Note that for the orthogonal case, i.e., $A_1A_2 = 0$ we have $A_1(A_2 - A_1) = -A_1$ and $A_2(A_1 - A_2) = -A_2$. Thus, for the orthogonal case the normal to S^G at any $\boldsymbol{v}$ intersects both C_1 and C_2. Returning to the general case, the relation between $\boldsymbol{v}$ and $\boldsymbol{a} \in C_1$ is given by

$$\boldsymbol{a} = \left[I + \frac{\|A_1\boldsymbol{v}\|}{\|A_1(A_2 - A_1)\boldsymbol{v}\|}(A_2 - A_1)\right]\boldsymbol{v} \tag{94}$$

and

$$\|\boldsymbol{a} - \boldsymbol{v}\| = \frac{\|A_1\boldsymbol{v}\|}{\|A_1(A_2 - A_1)\boldsymbol{v}\|}\|(A_2 - A_1)\boldsymbol{v}\|. \tag{95}$$

Analogously for $\boldsymbol{b} \in C_2$, if $A_2\boldsymbol{v}$ and $A_2(A_1 - A_2)\boldsymbol{v}$ are linearly **(?)** dependent

$$\boldsymbol{b} = \left[I + \frac{\|A_2\boldsymbol{v}\|}{\|A_2(A_1 - A_2)\boldsymbol{v}\|}(A_1 - A_2)\right]\boldsymbol{v}. \tag{96}$$

Note that given $\boldsymbol{a}$ (or $\boldsymbol{b}$), there can be more than one solution for $\boldsymbol{v}$. Returning to the original (equivalent) optimization problem in (87) we can find now a sufficient and necessary condition on $\boldsymbol{v}$ for global minimum. Consider $\boldsymbol{v}' = \boldsymbol{v} + \boldsymbol{d}$. Assume $\boldsymbol{v}'$ maintains the constraint of (87)

$$\boldsymbol{v}'^T(A_2 - A_1)\boldsymbol{v}' = (\boldsymbol{v} + \boldsymbol{d})^T(A_2 - A_1)(\boldsymbol{v} + \boldsymbol{d}) = 0 \tag{97}$$

or

$$\begin{aligned} \boldsymbol{v}^T(A_2 - A_1)\boldsymbol{v} + \boldsymbol{d}^T(A_2 - A_1)\boldsymbol{v} + \boldsymbol{v}^T(A_2 - A_1)\boldsymbol{d} \\ + \boldsymbol{d}^T(A_2 - A_1)\boldsymbol{d} = 0 \end{aligned} \tag{98}$$

and, since $\boldsymbol{v}^T(A_2 - A_1)\boldsymbol{v} = 0$, it follows that

$$\boldsymbol{d}^T(A_2 - A_1)\boldsymbol{v} + \boldsymbol{v}^T(A_2 - A_1)\boldsymbol{d} + \boldsymbol{d}^T(A_2 - A_1)\boldsymbol{d} = 0. \tag{99}$$

Defining $f(\boldsymbol{z}) = \|\boldsymbol{a} - \boldsymbol{z}\|^2$, we derive the following relation between $f(\boldsymbol{v})$ and $f(\boldsymbol{v}')$:

$$\begin{aligned} f(\boldsymbol{v}') &= (\boldsymbol{a} - \boldsymbol{v}')^T(\boldsymbol{a} - \boldsymbol{v}') \\ &= (\boldsymbol{a} - \boldsymbol{v} - \boldsymbol{d})^T(\boldsymbol{a} - \boldsymbol{v} - \boldsymbol{d}) \\ &= (\boldsymbol{a} - \boldsymbol{v})^T(\boldsymbol{a} - \boldsymbol{v}) - \boldsymbol{d}^T(\boldsymbol{a} - \boldsymbol{v}) - (\boldsymbol{a} - \boldsymbol{v})^T\boldsymbol{d} + \boldsymbol{d}^T\boldsymbol{d} \\ &= f(\boldsymbol{v}) - 2\lambda\boldsymbol{d}^T(A_2 - A_1)\boldsymbol{v} + \|\boldsymbol{d}\|^2 \\ &= f(\boldsymbol{v}) + \lambda\boldsymbol{d}^T(A_2 - A_1)\boldsymbol{d} + \|\boldsymbol{d}\|^2 \\ &= f(\boldsymbol{v}) + \boldsymbol{d}^T[I + \lambda(A_2 - A_1)]\boldsymbol{d} \end{aligned} \tag{100}$$

where the forth equality follows from (91) and the fifth follows from (99).

We have derived the following relation:

$$f(\boldsymbol{v}') = f(\boldsymbol{v}) + \boldsymbol{d}^T \left[I + \frac{\|A_1\boldsymbol{v}\|}{\|A_1(A_2 - A_1)\boldsymbol{v}\|} (A_2 - A_1) \right] \boldsymbol{d}. \tag{101}$$

Therefore, for $\boldsymbol{v}$ satisfying (91) (i.e., the vector $\boldsymbol{a} - \boldsymbol{v}$ is perpendicular to the separating surface), a necessary and sufficient condition for global minimum is

$$\begin{aligned} C(\boldsymbol{v}) &= I + \lambda(A_2 - A_1) \\ &= I + \frac{\|A_1\boldsymbol{v}\|}{\|A_1(A_2 - A_1)\boldsymbol{v}\|} (A_2 - A_1) \succeq 0 \end{aligned} \tag{102}$$

where $\succeq$ denotes semi-positive definiteness. If we further require

$$\begin{aligned} C(\boldsymbol{v}) &= I + \lambda(A_2 - A_1) \\ &= I + \frac{\|A_1\boldsymbol{v}\|}{\|A_1(A_2 - A_1)\boldsymbol{v}\|} (A_2 - A_1) \succ 0 \end{aligned} \tag{103}$$

where $\succ$ denotes positive definiteness, then $\boldsymbol{v}$ is also unique. Either condition (102) or condition (103) ensure that $\|\boldsymbol{v} - \boldsymbol{a}\|$ represents the distance of $\boldsymbol{a}$ from S^G.

We can define a matrix $C'(\boldsymbol{v})$, analogous to $C(\boldsymbol{v})$, for $\boldsymbol{b} \in C_2$ and require its positive definiteness in order to ensure $\|\boldsymbol{v} - \boldsymbol{b}\|$ represents the distance of $\boldsymbol{b}$ from S^G

$$\begin{aligned} C'(\boldsymbol{v}) &= I + \lambda(A_1 - A_2) \\ &= I + \frac{\|A_2\boldsymbol{v}\|}{\|A_2(A_1 - A_2)\boldsymbol{v}\|} (A_1 - A_2) \succ 0. \end{aligned} \tag{104}$$

Verifying the positive definiteness of the matrices $C(\boldsymbol{v})$ and $C'(\boldsymbol{v})$ may be complex, as it should be repeated for different choices of $\boldsymbol{v}$. To reduce the complexity, the eigenvalues of $C(\boldsymbol{v})$ can be related to those of the matrix $A = A_2 - A_1$, which are independent of $\boldsymbol{v}$ and so can be calculated off-line. Specifically, in order for (102) to hold, every eigenvalue μ of $C(\boldsymbol{v})$ must satisfy $\mu \geq 0$. Now, for every eigenvalue μ of $C(\boldsymbol{v})$

$$\det[C(\boldsymbol{v}) - \mu I] = 0. \tag{105}$$

Substituting $C(\boldsymbol{v})$, yields

$$\det[I + \lambda(A_2 - A_1) - \mu I] = 0. \tag{106}$$

Thus,

$$\det[(1-\mu)I + \lambda(A_2 - A_1)] = 0 \tag{107}$$

or

$$\det\left[-\frac{\mu - 1}{\lambda} I + (A_2 - A_1) \right] = 0. \tag{108}$$

It can be observed from (108) that $\delta = (\mu - 1)/\lambda$ is an eigenvalue of $A_2 - A_1$. Since we have required $\mu \geq 0$, δ has to satisfy $\delta \geq -1/\lambda$. In other words, $\boldsymbol{v}$ is a global minimum iff the minimal eigenvalue of $A_2 - A_1$, $\delta_{\min}$ satisfies

$$\delta_{\min} \geq -\frac{\|A_1(A_2 - A_1)\boldsymbol{v}\|}{\|A_1\boldsymbol{v}\|}. \tag{109}$$

We now determine the necessary conditions for a decoder that uniformly improves the GLRT, and present explicitly such a decoder. For simplicity, we assume in what follows an orthogonal case, i.e., $A_1 A_2 = 0$. It can be easily shown that in this orthogonal case $\delta_{\min} = -1$. Therefore, condition (109) is satisfied with equality and $C(\boldsymbol{v})$ is always a semipositive matrix. Thus, for any $\boldsymbol{v}$, we denote the intersection of the normal to S^G at $\boldsymbol{v}$ with C_1 by $\boldsymbol{a}^{(1)}$. Kuhn–Tucker conditions assure that global minimum for $\|\boldsymbol{a}^{(1)} - \boldsymbol{v}\|$ is achieved. Likewise, we denote the intersection of the normal to S^G at $\boldsymbol{v}$ with C_2 by $\boldsymbol{b}^{(2)}$.

Projecting $\boldsymbol{a}^{(1)} = X_1\boldsymbol{h}_1$ on S^G, there may be infinite solutions $\boldsymbol{v}$, such that $\|\boldsymbol{a}^{(1)} - \boldsymbol{v}\|$ is minimal. We denote by V the set of optimal solutions, given by

$$V = \left\{ \boldsymbol{v} \in S^G : \boldsymbol{v} - \boldsymbol{a}^{(1)} \perp S^G \right\}. \tag{110}$$

For each $\boldsymbol{v} \in V$, a unique $\boldsymbol{b}^{(2)} = X_2\boldsymbol{h}_2$ can be found using (96). We denote this set by $\mathcal{B}^{(2)}$

$$\mathcal{B}^{(2)} = \left\{ \boldsymbol{b}^{(2)} : \boldsymbol{v} \in V,\, \boldsymbol{v} - \boldsymbol{b}^{(2)} \perp S^G \right\}. \tag{111}$$

A sufficient and necessary condition for the existence of a decoder that improves the error probability for $\boldsymbol{h}_1$ and does not worsen the error probability for any other channel parameters vector, is that any $\boldsymbol{h}_2$, such that $X_2\boldsymbol{h}_2 \in \mathcal{B}^{(2)}$, satisfies

$$d_1^G(\boldsymbol{h}_1) < d_2^G(\boldsymbol{h}_1) \tag{112}$$

$$d_1^G(\boldsymbol{h}_2) < d_2^G(\boldsymbol{h}_2) \tag{113}$$

where $d_1^G(\boldsymbol{h})$ and $d_2^G(\boldsymbol{h})$ were defined in (23) and (24), respectively. Analogous conditions can be formed for the case $d_1^G(\boldsymbol{h}_1) > d_2^G(\boldsymbol{h}_1)$.

We describe now the decoding procedure. Assume that the observed vector $\boldsymbol{y}$ is in the region $m = 2$ of the GLRT decoder. The vector of $\boldsymbol{y}$ may have more than one projection on S^G. Denote this set by V_y

$$V_y = \{ \boldsymbol{v} \in S^G : \boldsymbol{v} - \boldsymbol{y} \perp S^G \}. \tag{114}$$

For a specific $\boldsymbol{v} \in V_y$, the normal to S^G intersects C_1, C_2 at $\boldsymbol{a}^{(1)}$, $\boldsymbol{b}^{(2)}$ according to (94) and (96), respectively. Since both X_1 and X_2 are full rank we can find unique ISI parameters $\boldsymbol{h}_1$ and $\boldsymbol{h}_2$ such that $\boldsymbol{a}^{(1)} = X_1\boldsymbol{h}_1$ and $\boldsymbol{b}^{(2)} = X_2\boldsymbol{h}_2$. Then, if conditions (112) and (113) hold, and $\epsilon(\boldsymbol{v})$ is given, a new mapping $\boldsymbol{u}$ can be found according to (52). If for some $\boldsymbol{v} \in V_y$ the observation $\boldsymbol{y}$ is on the line between $\boldsymbol{a}^{(1)}$ and $\boldsymbol{u}$ we decode $\hat{m} = 1$, otherwise $\hat{m} = 2$.

VII. Energy Weighted Decoder

For two hypotheses $\mathcal{H}_1, \mathcal{H}_2$, i.e., two codewords $\boldsymbol{x}^{(1)}$ and $\boldsymbol{x}^{(2)}$ defined in (6) the GLRT decoding rule in (15) can be reformulated as

$$\frac{\sup_{\theta\in\Theta} p_\theta\left(\boldsymbol{y}\,\middle|\,\boldsymbol{x}^{(1)};\,\theta\right)}{\sup_{\theta\in\Theta} p_\theta\left(\boldsymbol{y}\,\middle|\,\boldsymbol{x}^{(2)};\,\theta\right)} \underset{\mathcal{H}_2}{\overset{\mathcal{H}_1}{\gtrless}} 1. \tag{115}$$

A new decoder that improves the average error probability over all the possible unknown fading coefficients is given by

$$\frac{\sup_{\theta\in\Theta} p_\theta\left(\boldsymbol{y}\,\middle|\,\boldsymbol{x}^{(1)};\,\theta\right)}{\sup_{\theta\in\Theta} p_\theta\left(\boldsymbol{y}\,\middle|\,\boldsymbol{x}^{(2)};\,\theta\right)} \underset{\mathcal{H}_2}{\overset{\mathcal{H}_1}{\gtrless}} \gamma \tag{116}$$

where γ has yet to be optimized in order to minimize the average error probability. The motivation for the new decoding rule is the simple fading case. Using the notations in Section IV, the GLRT decoding rule was

$$\frac{|y_0|}{|y_1|} \underset{\mathcal{H}_2}{\overset{\mathcal{H}_1}{\gtrless}} 1. \tag{117}$$

The new decoding rule suggested, which reduces the exponential order of the error probability, is given by

$$\frac{|y_0|}{|y_1|} \underset{\mathcal{H}_2}{\overset{\mathcal{H}_1}{\gtrless}} \frac{|a|}{|b|}. \tag{118}$$

Therefore, a possible choice for the parameter γ can be a function of the ratio between the energies of the codewords. In the one-dimensional case, γ is given by the square root of this ratio. For ISI channels, we select γ as a certain power of the ratio of the energies. Denote the energies of the transmitted signals $\boldsymbol{x}^{(1)}$ and $\boldsymbol{x}^{(2)}$ as E_1 and E_2, respectively. Then

$$E_1 = \left\|\boldsymbol{x}^{(1)}\right\|^2 \tag{119}$$

$$E_2 = \left\|\boldsymbol{x}^{(2)}\right\|^2 \tag{120}$$

and the decoding rule is given by

$$\frac{\sup\limits_{\theta\in\Theta} p_\theta\left(\boldsymbol{y}\left|\boldsymbol{x}^{(1)}; \theta\right.\right)}{\sup\limits_{\theta\in\Theta} p_\theta\left(\boldsymbol{y}\left|\boldsymbol{x}^{(2)}; \theta\right.\right)} \underset{\mathcal{H}_2}{\overset{\mathcal{H}_1}{\gtrless}} \left(\frac{E_1}{E_2}\right)^\eta \tag{121}$$

for some $0 \le \eta \le 1$. For $\eta = 0$ it is the GLRT decoder.

According to (16), for Gaussian ISI channels the decoding rule is given by

$$\frac{\min\limits_{\boldsymbol{h}} \|\boldsymbol{y} - X_2\boldsymbol{h}\|^2}{\min\limits_{\boldsymbol{h}} \|\boldsymbol{y} - X_1\boldsymbol{h}\|^2} \underset{\mathcal{H}_2}{\overset{\mathcal{H}_1}{\gtrless}} \left(\frac{E_1}{E_2}\right)^\eta. \tag{122}$$

In Fig. 17, we compare the average performance (over messages and channel coefficients) of the GLRT decoder, the ULRT, and the energy weighted decoder (EWD) for a specific code with two codewords. The code we used for the simulation is

$$X_1 = \begin{bmatrix} 1 & 0 \\ 2 & 1 \\ 0 & 2 \end{bmatrix}, \qquad X_2 = \begin{bmatrix} 6 & 0 \\ 7 & 6 \\ 0 & 7 \end{bmatrix}. \tag{123}$$

The value of η was optimized by a search over a grid. The optimal value that minimizes the average error probability is about $\eta = 0.3$.

Fig. 18 compares the performance of the GLRT and the EWD. The error probability for a certain choice of parameter vector (h_0, h_1) is given by $P_e^N(h_0, h_1)$ for the EWD and by $P_e^G(h_0, h_1)$ for the GLRT. The graph shows the difference $P_e^N(h_0, h_1) - P_e^G(h_0, h_1)$ of the error probability for each choice of (h_0, h_1). We see that while for some (h_0, h_1)

$$P_e^N(h_0, h_1) - P_e^G(h_0, h_1) < 0$$

for others

$$P_e^N(h_0, h_1) - P_e^G(h_0, h_1) > 0$$

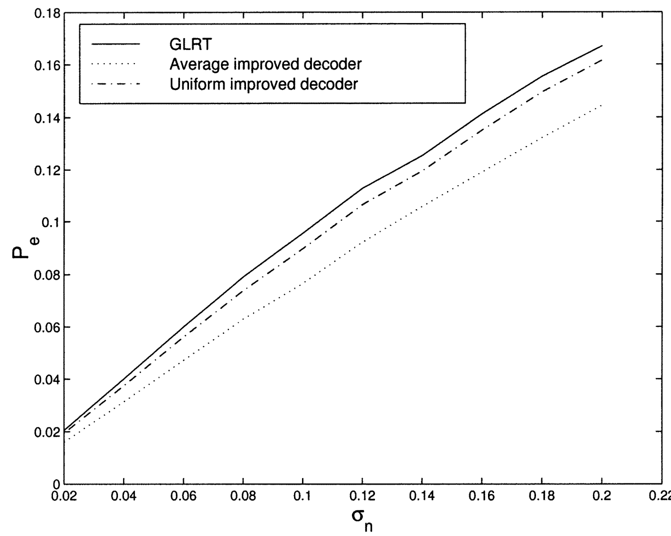

Fig. 17. Average performance $N = 3, K = 2, M = 2$.

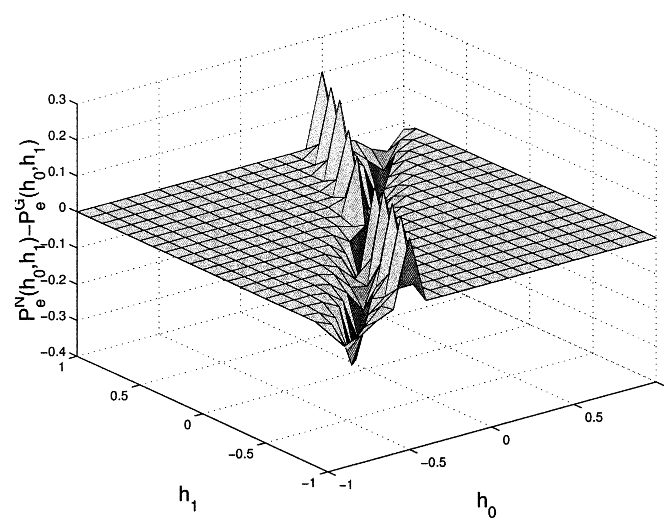

Fig. 18. Comparison between the GLRT and the EWD $N = 3$, $K = 2$, $M = 2$.

and, therefore, the improvement is not uniform. The comparison between the GLRT and the ULRT in the parameter space was already given above in Fig. 15.

VIII. Summary and Further Research

We have introduced in this work two classes of alternative new decoders for unknown linear channels that improve the GLRT under different criteria. Most of our work is dedicated to the ULRT that uniformly improves the error probability (actually the exponential order of the error probability) of the GLRT decoder. For this decoder we have distinguished between two cases: the hyperplane case and the general case, which are determined by K, the number of channel parameters and N, the block length. The hyperplane case turned out to be simpler and we found closed-form equations for implementing the algorithm. The general case turned out to be more complicated since it involved a nonconvex optimization problem. We have explicitly presented a decoder only for the case where the subspaces associated with the codewords are orthogonal. The fact that one can uniformly improve the GLRT is important since much research was directed to find theoretical justification to the GLRT decoder, and to develop implementation algorithms for it. Our result shows that the lack of theoretical justification is not coincidental. Yet, from the practical viewpoint, at least for the hyperplane case, the complexity of the ULRT is not significantly higher than that of the GLRT and can incorporate existing GLRT decoders.

Decoders of the second new class improve the average (over channel parameters) error probability. The resulting EWD rotates the separating surface of the GLRT in the direction of the less energetic codeword. Thus, while the separating surface maintains the same characteristics of the GLRT (e.g., hyperplane, quadratic surface) it improves the exponential order of the average error probability. In this respect, we note that while in many cases codewords have the same energy, there are cases where it is actually advantageous to use different energies. For example, in [19] it was shown that quadrature amplitude modulation (QAM), with codewords that are not necessarily equal in energy, has superior performance over the equal energy phase shift keying (PSK) modulation even for noncoherent reception employing the GLRT decoder. A simplified version of the EWD was introduced in [14], where it was implemented for a multiple-antenna system employing QAM.

For further research, one direction would be the implementation of the ULRT for practical systems. This may require an algorithm for determining the parameter ϵ that may involve iterative or recursive modifications of an initial value. For the general case, the implementation may require an algorithm for solving the resulting nonconvex optimization problem. Also, an explicit analysis in the general case, without the assumption that the codewords subspaces are orthogonal, should be completed. For practical systems, one should also find efficient implementation for the case of large M codewords.

Actually, the case of large M, especially the case where M grows exponentially with the block length N, i.e., $M = 2^{NR}$ for some rate R, is interesting and requires further theoretical analysis. Specifically, an interesting question is whether the ULRT can improve the error exponent attained by the GLRT. In this respect, it was shown [20] that GLRT decoders can achieve the rate attainable by an optimal ML decoder, yet the GLRT exponential error performance may be improved.

An additional direction for research is modifying the decoders to other channel models. Linear systems can, in general, be classified into four categories: time-invariant flat fading, time-invariant frequency selective fading, time-variant flat fading, and time-variant frequency fading. The first category is covered by the simple fading example, while our work here focused mainly on the second category. A natural generalization of the GLRT decoder to time-variant channels would modify the estimation of the channel coefficients involved in the algorithm. Instead of LS estimation it could involve weighted least squares (WLS) algorithm, where the weights are chosen to account for the changes in the channel. A new decoding algorithm that improves the performance of this decoder can be developed analogously to the improved decoder we have developed in this work for time invariant channels.

Another subject for further research involves performance bounds, and especially analysis of the error exponent achieved by the decoders. The decoders might be analyzed according to the *competitive min-max criterion* proposed in [12]. This criterion minimizes the worst ratio between the error probability of the proposed decoder and the error probability of the optimal ML rule, raised to a certain power $0 < \xi < 1$. It is interesting to see to what extent the new decoder proposed here satisfies this criterion.

A criterion for an optimal decision rule under channel uncertainty is not well defined. A certain decoder is superior to another decoder under any criterion only if it uniformly improves the error probability. In this work, we have shown that the GLRT is not an admissible decision rule, as it can be uniformly improved. This work might be a step toward a more general theory designed to determine whether a certain decision rule is admissible or not.

The problem of *encoder* design for unknown linear channels can be investigated more closely in order to achieve a complete view of robust communication systems for unknown channels. A general discussion of robust communication for various classes of unknown channels can be found in [10]. Clearly, the design of encoders for unknown channels could take into account the results here and other related results on universal decoding.

Appendix
ULRT Structure for the Hyperplane Case

In this appendix, we will look more closely at the structure of the decoder when $C_m,\ m = 1, 2$ in (19) represent hyperplanes as in (32). We provide a geometrical representation of the problem (i.e., the structure of the separating surface of the GLRT decoder). This will be the basis for the geometrical structure of the ULRT.

Assume that $\boldsymbol{p}$ in (35) is a unit vector. The distances $d_1^G(\boldsymbol{h})$ and $d_2^G(\boldsymbol{h})$ defined in (23) and (24), respectively, are given by

$$d_1^G(\boldsymbol{h}) = |\boldsymbol{p}^T X_1 \boldsymbol{h}|, \qquad d_2^G(\boldsymbol{h}) = |\boldsymbol{p}^T X_2 \boldsymbol{h}|. \tag{124}$$

Denote by L_1 the intersection of C_1 and C_2

$$L_1 = C_1 \bigcap C_2 \tag{125}$$

which is a subspace of dimension $N-2 = K-1$. From (32), (35), and (36) it follows that

$$L_1 = S^G \bigcap C_1 = S^G \bigcap C_2. \tag{126}$$

Consider the hyperplane C_1 (the following procedure is applicable to C_2 as well). The intersection of C_1 with S^G is L_1 and given by

$$L_1 = \{X_1\boldsymbol{h}\colon \boldsymbol{p}^T X_1 \boldsymbol{h} = 0,\ \boldsymbol{h} \in \Re^K\}. \tag{127}$$

See illustration for $N = 3,\ K = 2$ in Fig. 19.

Observe that in the parameter space (of dimension K) L_1 is a hyperplane (of dimension $K-1$). The hyperplane L_1 divides C_1 into two regions

$$L_1^+ = \{X_1\boldsymbol{h}\colon \boldsymbol{p}^T X_1 \boldsymbol{h} > 0,\ \boldsymbol{h} \in \Re^K\} \tag{128}$$

$$L_1^- = \{X_1\boldsymbol{h}\colon \boldsymbol{p}^T X_1 \boldsymbol{h} < 0,\ \boldsymbol{h} \in \Re^K\}. \tag{129}$$

We can also construct the $K-1$-dimensional hyperplane L_2 defined by

$$L_2 = \{X_1\boldsymbol{h}\colon \boldsymbol{p}^T X_2 \boldsymbol{h} = 0,\ \boldsymbol{h} \in \Re^K\} \tag{130}$$

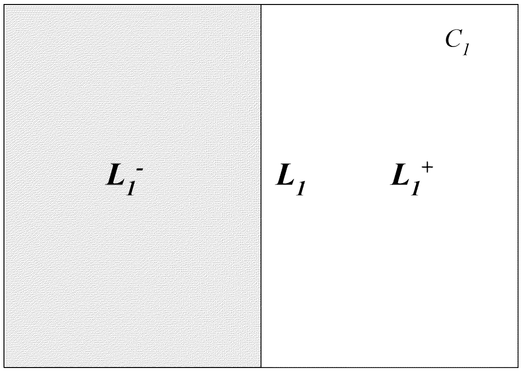

Fig. 19. Illustration for $N = 3, K = 2$.

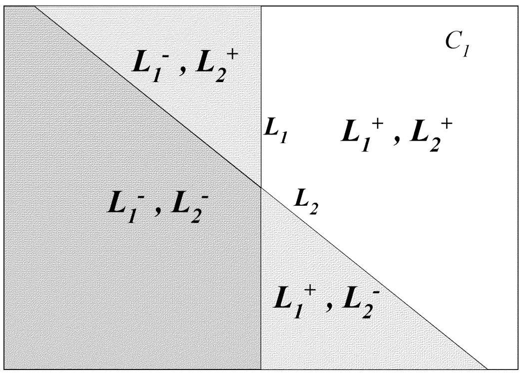

Fig. 20. Illustration for $N = 3, K = 2$ (continued).

which divides C_1 into L_2^+ and L_2^- (defined analogously to L_1^+ and L_1^-). See Fig. 20 for illustration. The hyperplane L_2 represents all the points $X_1\boldsymbol{h} \in C_1$ such that $X_2\boldsymbol{h} \in L_1$ ($d_2^G(\boldsymbol{h}) = 0$, see definition in (24)). The intersection of L_1 and L_2 is a subspace given by

$$L_1 \bigcap L_2 = \{X_1\boldsymbol{h}\colon \boldsymbol{p}^T X_1 \boldsymbol{h} = 0,\, \boldsymbol{p}^T X_2 \boldsymbol{h} = 0,\, \boldsymbol{h} \in \Re^K\} \tag{131}$$

which is of dimension $K - 2$.

Construct the $K-1$-dimensional hyperplanes L_3 and L_4 defined by

$$L_3 = \{X_1\boldsymbol{h}\colon \boldsymbol{p}^T X_1\boldsymbol{h} - \boldsymbol{p}^T X_2\boldsymbol{h} = 0,\, \boldsymbol{h} \in \Re^K\} \tag{132}$$

$$L_4 = \{X_1\boldsymbol{h}\colon \boldsymbol{p}^T X_1\boldsymbol{h} + \boldsymbol{p}^T X_2\boldsymbol{h} = 0,\, \boldsymbol{h} \in \Re^K\}. \tag{133}$$

It follows from (124) that for any $X_1\boldsymbol{h} \in L_3$ or $X_1\boldsymbol{h} \in L_4$, $d_1^G(\boldsymbol{h}) = d_2^G(\boldsymbol{h})$.

Define the set D

$$D = L_3 \bigcap L_4 \tag{134}$$

which is of dimension $K - 2$. For any $X_1\boldsymbol{h} \in D$, $\boldsymbol{p}^T X_1\boldsymbol{h} = 0$ and $\boldsymbol{p}^T X_2\boldsymbol{h} = 0$. Therefore,

$$D = L_3 \bigcap L_4 = L_1 \bigcap L_2 = L_1 \bigcap L_2 \bigcap L_3 \bigcap L_4 \tag{135}$$

or

$$\begin{aligned} D &= \{X_1\boldsymbol{h}\colon \boldsymbol{p}^T X_1\boldsymbol{h} + \boldsymbol{p}^T X_2\boldsymbol{h} = 0,\\ &\qquad \boldsymbol{p}^T X_1\boldsymbol{h} - \boldsymbol{p}^T X_2\boldsymbol{h} = 0,\, \boldsymbol{h} \in \Re^K\}\\ &= \{X_1\boldsymbol{h}\colon \boldsymbol{p}^T X_1\boldsymbol{h} = 0,\, \boldsymbol{p}^T X_2\boldsymbol{h} = 0,\, \boldsymbol{h} \in \Re^K\}. \end{aligned} \tag{136}$$

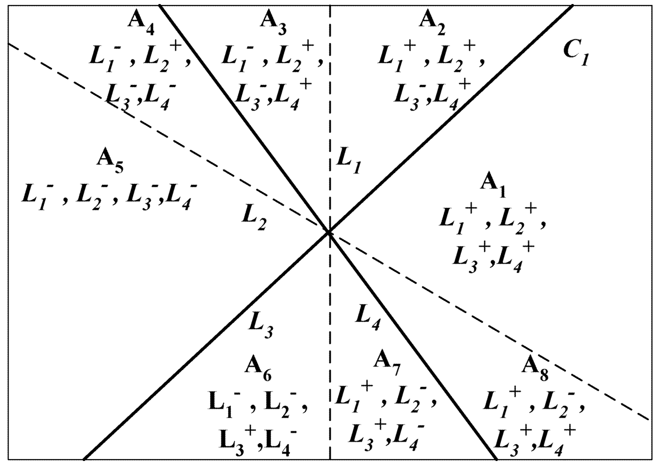

Fig. 21. Illustration for $N = 3, K = 2$ (continued).

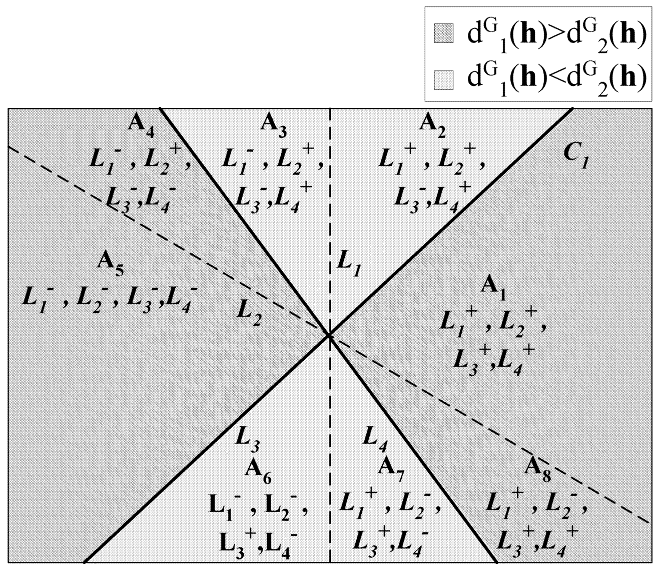

Fig. 22. Illustration for $N = 3, K = 2$ (continued).

The hyperplanes $L_1,\, L_2,\, L_3,\, L_4$ divide C_1 into eight regions (see Fig. 21) $A_i,\, i = 1, \ldots, 8$. This hyperplane L_3 passes in the regions $L_1^+ \bigcap L_2^+,\, L_1^- \bigcap L_2^-$. The hyperplane divides C_1 into two regions: L_3^+ (where $\boldsymbol{p}^T X_1\boldsymbol{h} - \boldsymbol{p}^T X_2\boldsymbol{h} > 0$) and L_3^- (where $\boldsymbol{p}^T X_1\boldsymbol{h} - \boldsymbol{p}^T X_2\boldsymbol{h} < 0$). Clearly, $L_1^+ \bigcap L_2^- \subset L_3^+$ and $L_1^- \bigcap L_2^+ \subset L_3^-$. This way the regions L_3+ and L_3- were determined in Fig. 22. Similarly, L_4^+ and L_4^- were determined.

We need to determine for each of these eight regions whether $d_1^G(\boldsymbol{h}) > d_2^G(\boldsymbol{h}), \forall\, \boldsymbol{h} \in A_i$ or $d_1^G(\boldsymbol{h}) < d_2^G(\boldsymbol{h}), \forall\, \boldsymbol{h} \in A_i$. Region A_1, for example, is given by $L_1^+ \bigcap L_2^+ \bigcap L_3^+ \bigcap L_4^+$. Therefore, $|\boldsymbol{p}^T X_1\boldsymbol{h}| > |\boldsymbol{p}^T X_2\boldsymbol{h}|$ and it follows that $d_1(\boldsymbol{h}) > d_2(\boldsymbol{h})$, $\forall\, \boldsymbol{h} \in A_1$. The same procedure can be carried out for the rest of the regions. We conclude that C_1 is divided into four regions where in two of them $d_1^G(\boldsymbol{h}) > d_2^G(\boldsymbol{h})$ and in the other two $d_1^G(\boldsymbol{h}) > d_2^G(\boldsymbol{h})$. The hyperplanes L_3, L_4 divide C_1 into these four regions; see Fig. 22.

We project both L_3 and L_4 on the GLRT separating surface S^G. The projections are the hyperplanes $L_3' \in S^G$ and $L_4' \in S^G$ that divide S^G into four regions. The subspaces L_3' and L_4' are both of dimension $K - 1$ and, therefore, are hyperplanes in S^G (K-dimensional). Their intersection is D (see (134)). This is so

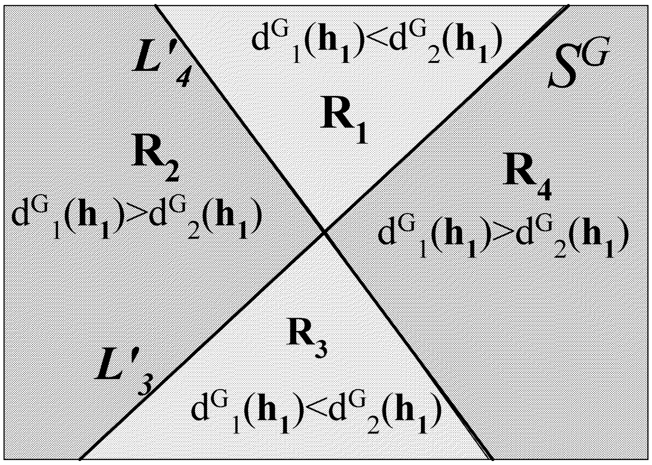

Fig. 23. Illustration for $N = 3, K = 2$ (continued).

since, as we recall, $D \subset L_1$ and, therefore, $D \subset S^G$. Thus, the projection of D on S^G is D itself and since $D = L_3 \cap L_4$ we conclude that $D \subset L'_3 \cap L'_4$. The hyperplanes L'_3 and L'_4 can intersect at most on a $K-2$-dimensional subspace and therefore,

$$D = L'_3 \cap L'_4. \tag{137}$$

See illustration for $N = 3,\ K = 2$ in Fig. 23.

At any point $\boldsymbol{v} \in S^G$ we can construct a normal to S^G. The normal intersects C_1 at $X_1\boldsymbol{h_1}$ for some $\boldsymbol{h_1}$ and C_2 at $X_2\boldsymbol{h_2}$ for some $\boldsymbol{h_2}$. Consider region R_1 (or R_3) in Fig. 23. If we construct a normal to S^G on any point $\boldsymbol{v} \in R_1$ the normal intersects C_1 at $X_1\boldsymbol{h_1}$ for some $\boldsymbol{h_1}$ where $d_1(\boldsymbol{h_1}) < d_2(\boldsymbol{h_1})$. Consider region R_2 (or R_4). If we construct a normal to S^G on any point $\boldsymbol{v} \in R_2$, the normal intersects C_1 at $X_1\boldsymbol{h_1}$ for some $\boldsymbol{h_1}$ where $d_1(\boldsymbol{h_1}) > d_2(\boldsymbol{h_1})$.

The entire procedure above is repeated for C_2. We construct on C_2 the hyperplanes $M_1,\ M_2,\ M_3$, and M_4. We project also M_3 and M_4 on S^G; see illustration for $N = 3,\ K = 2$ in Fig. 24. The regions formed on S^G are denoted by G_i. We assumed that $L'_3,\ L'_4,\ M'_3$, and M'_4 do not overlap, since otherwise the GLRT cannot be uniformly improved as shown by the converse theorem.

Some regions (in our example G_1 and G_5) maintain (43) and (44) and some regions (in our example G_3 and G_7) maintain these assumptions with both inequality signs reversed. Define a region $G_1^\alpha \subset G_1 \subset S^G$ and of finite angle α from the boundaries of G_1; see Fig. 25 for illustration. Similarly, define $G_2^\alpha, \ldots, G_8^\alpha$. We will show that any point $\boldsymbol{v} \in G_1^\alpha,\ G_5^\alpha$ can be mapped to a new point $\boldsymbol{u}$ in the new separating surface according to

$$\boldsymbol{u} = (1 - \epsilon(\boldsymbol{v}))\boldsymbol{v} + \epsilon(\boldsymbol{v})X_1\boldsymbol{h_1} \tag{138}$$

where $\epsilon(\boldsymbol{v}) > 0$ and the projection of $X_1\boldsymbol{h_1}$ on S^G is $\boldsymbol{v}$. Similarly, any point $\boldsymbol{v} \in G_3^\alpha,\ G_7^\alpha$ can be mapped to a new point $\boldsymbol{u}$ in the new separating surface according to

$$\boldsymbol{u} = (1 - \epsilon(\boldsymbol{v}))\boldsymbol{v} + \epsilon(\boldsymbol{v})X_2\boldsymbol{h_2} \tag{139}$$

such that the new separating surface maintains (29) and (30) and, therefore, uniformly improves the GLRT decoder. The even regions of the new separating surface will remain the same as for

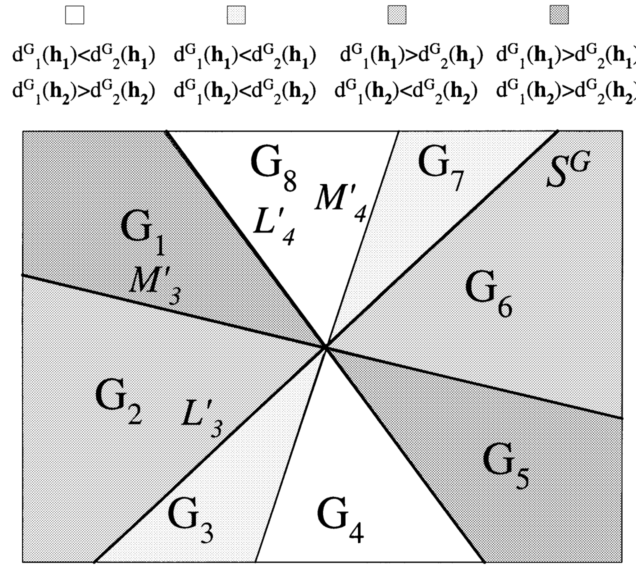

Fig. 24. Illustration for $N = 3, K = 2$ (continued).

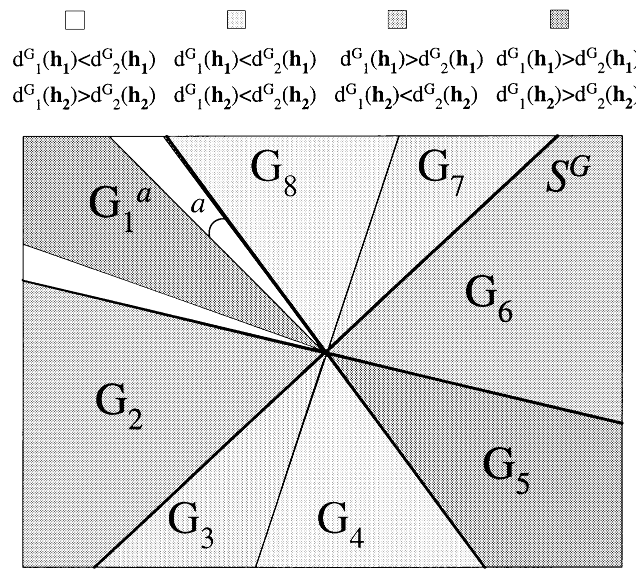

Fig. 25. Illustration for $N = 3, K = 2$ (continued).

the GLRT. We will also show that $\epsilon(\boldsymbol{v})$ does not depend on the magnitude of $\boldsymbol{v}$, but only on its direction. Thus, vectors on S^G with the same direction (linearly dependent) have the same $\epsilon(\boldsymbol{v})$.

We construct the ULRT based on Section V. Construct regions $G_i^1 \in C_1$ such that G_i is the projection of G_i^1 on S^G. Similarly, construct $G_i^2 \in C_2$. Consider, for example, G_1 where $d_1^G(\boldsymbol{h_1}) > d_2^G(\boldsymbol{h_1}),\ d_1^G(\boldsymbol{h_2}) > d_2^G(\boldsymbol{h_2})$. In region G_1^1, since $d_1^G(\boldsymbol{h}) > d_2^G(\boldsymbol{h}),\ S^G$ is strictly outside a ball $\mathcal{B}_1(\boldsymbol{h})$ of radius $d_2(\boldsymbol{h})$ around $X_1\boldsymbol{h}$ (see (53)). Therefore, $\boldsymbol{v}$ is also strictly outside $\mathcal{B}_1(\boldsymbol{h})$ and we can find $\epsilon(\boldsymbol{v})$ so that $\boldsymbol{u}$ is also strictly outside $\mathcal{B}_1(\boldsymbol{h})$.

As for the other regions of C_1 (in our example $G_2^1, \ldots, G_8^1$) we need to show that we can find $\epsilon(\boldsymbol{v})$ so that $\boldsymbol{u}$ is strictly outside $\mathcal{B}_1(\boldsymbol{h})$ of radius $d_1(\boldsymbol{h})$ to around $X_1\boldsymbol{h}$ (see (60)). Denote by $\boldsymbol{v}'$ the projection of some $X_1\boldsymbol{h}$ on S^G (see Fig. 26). Denote by α' the angle between $\boldsymbol{v}$ and $\boldsymbol{v}'$, $\cos(\alpha') = \boldsymbol{v} \cdot \boldsymbol{v}'/\|\boldsymbol{v}\|\|\boldsymbol{v}'\|$.

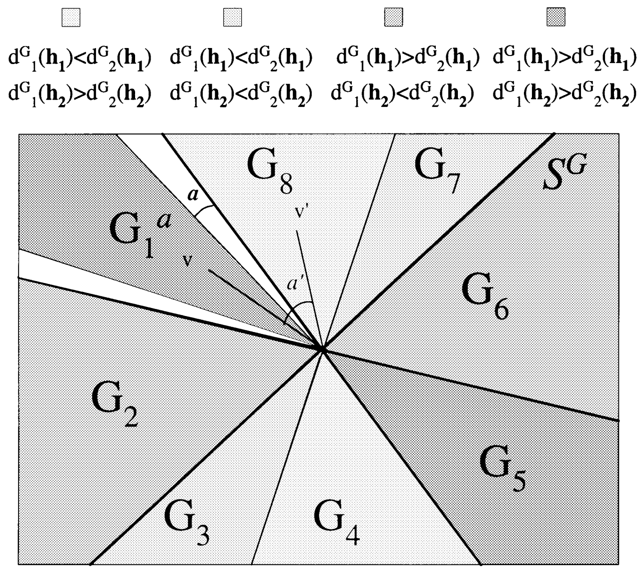

Fig. 26. Illustration for $N = 3, K = 2$ (continued).

According to construction, $\alpha' > \alpha$. The distance between $X_1\boldsymbol{h}$ and $\boldsymbol{v}'$ is $d_1^G(\boldsymbol{h})$. The distance between $\boldsymbol{v}'$ and $\boldsymbol{v}$ is

$$\|\boldsymbol{v} - \boldsymbol{v}'\| = (\|\boldsymbol{v}\|^2 + \|\boldsymbol{v}'\|^2 - 2\boldsymbol{v} \cdot \boldsymbol{v}')^{1/2}$$

(cosine law). It follows that if α and $\|\boldsymbol{v}\|$ are finite then so is $\|\boldsymbol{v} - \boldsymbol{v}'\|$. Denote by d' the distance between $X_1\boldsymbol{h}$ and $\boldsymbol{v}$. According to Pythagoras

$$d' = \left(d_1^G(\boldsymbol{h})^2 + \|\boldsymbol{v} - \boldsymbol{v}'\|^2\right)^{1/2} > d_1^G(\boldsymbol{h}). \tag{140}$$

Therefore, $\boldsymbol{v}$ is strictly outside the ball $\mathcal{B}_1(\boldsymbol{h})$ of radius $d_1^G(\boldsymbol{h})$ around $X_1\boldsymbol{h}$ (see (60)) and we can find $\epsilon(\boldsymbol{v})$ so that $\boldsymbol{u}$ is also outside $\mathcal{B}_1(\boldsymbol{h})$. In other words, according to construction, we know that the union of balls of radius $\min\{d_1^G(\boldsymbol{h}), d_2^G(\boldsymbol{h})\}$ around $X_1\boldsymbol{h}, \forall\, \boldsymbol{h} \in \Re^k$ is not tangent to G_1^α. Therefore, for any $\boldsymbol{v} \in G_1^\alpha$ we can find a suitable $\epsilon(\boldsymbol{v})$ s.t. $\boldsymbol{u}$ is outside the above union of balls.

Define a subset of $G_1^2 \subset C_2$, denoted by $G_1^{2,\alpha}$, where $X_2\boldsymbol{h} \in G_1^{2,\alpha}$ if the projection of $X_2\boldsymbol{h}$ on S^G is within G_1^α. We want to show that error probability will be improved for any $\boldsymbol{h}$ s.t. $X_2\boldsymbol{h} \in G_1^{2,\alpha}$. Since $d_1(\boldsymbol{h}) > d_2(\boldsymbol{h})$ for any $\boldsymbol{h}$ in this region, the union of balls of radius $\min\{d_1^G(\boldsymbol{h}), d_2^G(\boldsymbol{h})\}$ around $X_2\boldsymbol{h}$ is tangent to G_1^α. Therefore, the mapping of the region G_1^α (in the direction of C_1) will improve the error probability for this region.

We turn now to show that the required value of $\epsilon(\boldsymbol{v})$ depends only on the direction of $\boldsymbol{v}$ and not on its magnitude. Suppose that $\boldsymbol{v}$ was mapped to $\boldsymbol{u}$ according to (138) (in some of the regions we know that we can find such $\boldsymbol{u}$). We conclude that $\boldsymbol{u}$ maintains

$$\|X_1\boldsymbol{h} - \boldsymbol{u}\| > \min\{|\boldsymbol{p}^T X_2\boldsymbol{h}|, |\boldsymbol{p}^T X_2\boldsymbol{h}|\}, \qquad \forall\, \boldsymbol{h}. \tag{141}$$

We want to show that the vector $t\boldsymbol{v}$, $t > 0$ can be mapped to $t\boldsymbol{u}$. See Fig. 27 for a cross section for the case $N = 3$ and $K = 2$. Substitute $\boldsymbol{h}/t$ instead of $\boldsymbol{h}$ in (141)

$$\|X_1\boldsymbol{h}/t - \boldsymbol{u}\| > \min\{|\boldsymbol{p}^T X_2\boldsymbol{h}/t|, |\boldsymbol{p}^T X_2\boldsymbol{h}/t|\}, \qquad \forall\, \boldsymbol{h}. \tag{142}$$

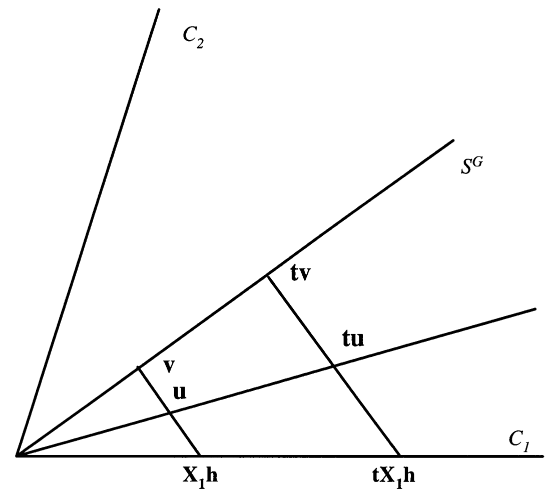

Fig. 27. Illustration for $N = 3, K = 2$ (continued).

Multiplying both sides by $t > 0$ results in

$$\|X_1\boldsymbol{h} - t\boldsymbol{u}\| > \min\{|\boldsymbol{p}^T X_2\boldsymbol{h}|, |\boldsymbol{p}^T X_2\boldsymbol{h}|\}, \qquad \forall\, \boldsymbol{h} \tag{143}$$

which is what we wanted to show. As a result, the relative error probability improvement does not deteriorate for channel parameters with larger magnitude.

REFERENCES

[1] G. D. Forney, "Maximum likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 363–378, May 1972.

[2] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Inform. Theory*, vol. 44, pp. 744–765, Mar. 1998.

[3] A. F. Naguib *et al.*, "A space-time coding modem for high-data-rate wireless communications," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1459–1478, Oct. 1998.

[4] M. Feder and J. A. Catipovic, "Algorithms for joint channel estimation and data recovery—Application to equalization in underwater communications," *IEEE J. Ocean. Eng.*, vol. 16, pp. 42–55, Jan. 1991.

[5] V. Tarokh and H. Jafarkhani, "A differential detection scheme for transmit diversity," *IEEE J. Select. Areas Commun.*, vol. 18, pp. 1169–1174, July 2000.

[6] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1451–1458, Oct. 1998.

[7] B. M. Hochwald and T. L. Marzetta, "Unitary space-time modulation for multiple-antenna communications in Rayleigh flat fading," *IEEE Trans. Inform. Theory*, vol. 46, pp. 543–564, Mar. 2000.

[8] O. Zeitouni, J. Ziv, and N. Merhav, "When is the generalized likelihood ratio test optimal?," *IEEE Trans. Inform. Theory*, vol. 38, pp. 1597–1602, Sept. 1992.

[9] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic, 1981.

[10] A. Lapidoth and P. Narayan, "Reliable communication under channel uncertainty," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2148–2177, Oct. 1998.

[11] N. Merhav, "Universal decoding for memoryless Gaussian channels with a deterministic interference," *IEEE Trans. Inform. Theory*, vol. 39, pp. 1261–1269, July 1993.

[12] M. Feder and N. Merhav, "Universal composite hypothesis testing—A competitive minimax approach," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1504–1517, June 2002.

[13] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1456–1467, July 1999.

[14] E. Erez and M. Feder, "A novel decoder for unknown diversity channels employing space-time codes," *Eurosip J. Appl. Signal Processing*, no. 3, pp. 267–274, Mar. 2002.

[15] H. L. Van Trees, *Detection, Estimation and Modulation Theory, Part I*. New York: Wiley, 1968.
[16] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1998.
[17] E. Erez, "A universal decoder for unknown Gaussian linear channels," M.S. thesis, Tel-Aviv Univ., Tel-Aviv, Israel, 2001.
[18] E. Erez and M. Feder, "Uniformly improving the generalized likelihood ratio test for linear Gaussian channels," in *38th Allerton Conf. Commun., Control, and Comupting*, Urbana, IL, Oct. 4–6, 2000, pp. 498–507.
[19] D. W. Warrier and U. Madhow, "Spectrally efficient noncoherent communication," *IEEE Trans. Inform. Theory*, vol. 48, pp. 651–668, Mar. 2002.
[20] A. Lapidoth and I. E. Telatar, "Gaussian ISI channels and the generalized likelihood ratio test," in *Proc. Int. Symp. Information theory (ISIT 2000)*, Sorrento, Italy, June 25–30, 2000, p. 460.