



**HAL**  
open science

# Constrained Epsilon-Minimax Test for Simultaneous Detection and Classification

Lionel Fillatre

► **To cite this version:**

Lionel Fillatre. Constrained Epsilon-Minimax Test for Simultaneous Detection and Classification. IEEE Transactions on Information Theory, 2011, 57 (12), pp.8055-8071. 10.1109/TIT.2011.2170114 . hal-00749180

**HAL Id: hal-00749180**

**<https://hal.science/hal-00749180>**

Submitted on 6 Nov 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Constrained Epsilon-Minimax Test for Simultaneous Detection and Classification

Lionel Fillatre

**Abstract**—A constrained epsilon-minimax test is proposed to detect and classify nonorthogonal vectors in Gaussian noise, with a general covariance matrix, and in presence of linear interferences. This test is epsilon-minimax in the sense that it has a small loss of optimality with respect to the purely theoretical and incalculable constrained minimax test which minimizes the maximum classification error probability subject to a constraint on the false alarm probability. This loss is even more negligible as the signal-to-noise ratio is large. Furthermore, it is also an epsilon-equalizer test since its classification error probabilities are equalized up to a negligible difference. When the signal-to-noise ratio is sufficiently large, an asymptotically equivalent test with a very simple form is proposed. This equivalent test coincides with the generalized likelihood ratio test when the vectors to classify are strongly separated in term of Euclidean distance. Numerical experiments on active user identification in a multiuser system confirm the theoretical findings.

**Index Terms**—Constrained minimax test, generalized likelihood ratio test, linear nuisance parameters, multiple hypothesis testing, statistical classification, user activity detection.

## I. INTRODUCTION

THE problem of detecting and classifying a vector in noisy measurements under uncertainty of vector presence often appears in engineering applications. This problem has many applications including radar and sonar signal processing [1], image processing [2], speech segmentation [3], [4], integrity monitoring of navigation systems [5], quantitative nondestructive testing [6], network monitoring [7] and digital communication [8] among others. This paper deals with the following detection and classification problem. It is assumed that a measurement vector  $\mathbf{y} \in \mathbb{R}^n$  consists of either a vector of interest (for example, a target or an anomaly) plus an unknown nuisance vector in additive Gaussian noise, or just an unknown nuisance vector in additive Gaussian noise. If present, the vector of interest must be detected and classified. The unknown nuisance vector belongs to the nuisance parameter subspace spanned by the columns of a known  $n \times m$  matrix  $\mathbf{H}$ . Hence, the observation model has the form

$$\mathbf{y} = \boldsymbol{\theta} + \mathbf{H}\mathbf{x} + \boldsymbol{\xi} \quad (1)$$

Manuscript received November 29, 2009; revised June 08, 2011; accepted June 28, 2011. Date of current version December 07, 2011. This work was supported in part by the French National Agency of Research under Grant ANR-08-SECU-013-02.

The author is with the ICD, LM2S, Université de Technologie de Troyes (UTT), UMR STMR, CNRS 6279, BP 2060, 10010, Troyes, France (e-mail: lionel.fillatre@utt.fr).

Communicated by M. Lops, Associate Editor for Detection and Estimation.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIT.2011.2170114

where both the nuisance parameter vector  $\mathbf{x} \in \mathbb{R}^m$  and the vector  $\boldsymbol{\theta} \in \mathbb{R}^n$  are unknown and deterministic. The zero-mean Gaussian noise vector  $\boldsymbol{\xi}$  has the known positive definite general covariance matrix  $\boldsymbol{\Sigma}$ . The vector  $\boldsymbol{\theta}$  belongs to the known set of  $r + 1$  different vectors  $\Theta = \{\boldsymbol{\theta}_0 = \mathbf{0}, \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_r\}$ , also called the vector constellation. The relation between the dimension  $n$  of the observed vector and the number  $r \geq 1$  of non-null vectors is arbitrary but it is assumed that  $n > m$ .

Three objectives are aimed to be achieved: i) vector detection which is to decide if  $\boldsymbol{\theta} \neq \mathbf{0}$  for a given false alarm probability (probability to declare an alarm when the observation vector is anomaly-safe), ii) vector classification which is to specify the actual index  $i$  of the vector  $\boldsymbol{\theta}_i$  and iii) insensitivity to the nuisance parameters which consists of taking the simultaneous detection/classification decision independently from the value of the unknown vector  $\mathbf{x}$ .

### A. Relation to Previous Work

From the statistical point of view, this problem of simultaneous detection/classification can be viewed as an hypotheses testing problem between several composite hypotheses [9], [10]. The goal is to design a statistical test which achieves the above mentioned three objectives according to a prefixed criterion of optimality.

The first approach to the design of statistical detection and classification tests is the uncoupled design strategy where detection performance is optimized under the false alarm constraint and the classification is gated by this optimal detection. On the one hand, the classical Neyman-Pearson criterion of vector detection [10] states that it is desirable to minimize the probability to miss the target subject to a constraint on the false alarm probability. On the other hand, in terms of target classification, it is desirable to minimize the probabilities to badly classify the target. All the above mentioned probabilities generally vary as a function of both the vector  $\boldsymbol{\theta}$  and the nuisance parameter  $\mathbf{x}$ . Hence, the uniform minimization of these probabilities with respect to  $\boldsymbol{\theta}$  and  $\mathbf{x}$  is in general impossible. There is no guarantee that the global performance of this uncoupled strategy will be acceptable. Consequently, a different approach must be taken, namely the coupled design strategies for detection and classification. These strategies have been studied by only a few authors. Pioneering works include the papers [11]–[14]. The common ground in each of these studies is the Bayesian point of view, i.e., prior probabilities are assigned to all the parameters so that average performance can be optimized. The problem of simultaneous detection and classification using a combination of a generalized likelihood ratio test and a maximum-likelihood classifier is studied in [15], [16]. This strategy is optimal only in some cases.

Contrary to a purely Bayesian criterion which needs a complete statistical description of the problem, the minimax criterion is well adapted to detection problems where some parameters are deterministic and unknown (typically the nuisance parameter  $\mathbf{x}$ ) and the appearance probability of each vector  $\theta_i$  is unknown. This criterion consists of minimizing the largest probability to make a decision error (typically a miss detection or a classification error). In [17], the generalized-likelihood ratio test approach is extended to multiple composite hypothesis testing, by breaking the problem into a sequence of binary composite hypothesis tests. In some cases, sufficient conditions for minimax optimality of this strategy are provided. Lastly, a general framework to design minimax tests with a prefixed level of false alarm, namely the constrained minimax tests, between multiple hypotheses composed of a finite number of parameters is established in [18].

It must be mentioned that some interesting papers like [19], [20] study the asymptotic performance of Bayesian tests between multiple hypotheses in absence of constraints on the false alarm probability. When the number of observations is very large, these papers show that the error probabilities depend only on the Kullback-Leibler information between the hypotheses. These results are not yet extended to the case of multiple hypotheses testing with a constraint on the false alarm probability.

### B. Motivation of the Study

The design of the optimal constrained minimax test mainly depends on three major points: 1) the geometric complexity of the vector constellation, 2) the covariance matrix of the Gaussian noise and 3) the presence of nuisance parameters. To underline the importance of these points, the three following cases must be distinguished.

In the simplest case, the vectors  $\theta_i$  are orthogonal and have the same norm (the least complex vector constellation), the covariance matrix  $\Sigma$  is the identity matrix (possibly multiplied by a known scalar) and there is no nuisance parameter  $\mathbf{x}$ . The optimal solution of the problem is given in [9], [21]: this is the so-called  $K$ -slippage problem.

In a more difficult case, the vectors  $\theta_i$  are not orthogonal and/or they have different norms (the most complex vector constellation), the Gaussian noise has a general covariance matrix  $\Sigma$  (not necessarily diagonal) and there is no nuisance parameter  $\mathbf{x}$ . The theoretical optimal solution is given by the constrained minimax test [18] but it is generally intractable. This optimal test compares the maximum of weighted likelihood ratios to a threshold to take its decision. Although the existence of the optimal solution is established, this solution depends on some unknown coefficients, namely the optimal weights and the threshold. Furthermore, the presence of a general covariance matrix  $\Sigma$  plays a significant role in the calculation of the optimal weights, even if the vector constellation is simple. In fact, it is always possible to get a diagonal covariance matrix after prefiltering but this operation may involve that the vector constellation becomes more complex. For example, orthogonal vectors of interest may be no longer orthogonal after prefiltering. For all these reasons, it is often impossible to easily calculate the optimal weights and, even, to reduce

the number of weights to be determined by using invariance principles [10], [22]. The optimization problem to be solved for calculating the optimal weights is highly nonlinear and leads to a combinatorial explosion.

*Example 1 (Slippage Problem With Unstructured Noise):* This example is directly related to the important problem of detecting outliers in multivariate normal data [23]. Let  $\mathbf{y}$  be a Gaussian random vector with zero mean and a known general covariance matrix  $\Sigma$  under hypothesis  $\mathcal{H}_0$ . Under hypothesis  $\mathcal{H}_j$ , the  $j$ th component  $y_j$  of  $\mathbf{y}$  has the known mean  $\Delta > 0$ . Hence, the vector  $\theta_j$  to detect and classify corresponds to  $\Delta\delta_j$  where  $\delta_j$  denotes the vector with a 1 in the  $j$ th coordinate and 0's elsewhere. Since the covariance is known but it differs from the identity matrix, it is no longer possible to use the famous principle of invariance to solve such a slippage problem [9]. The optimal solution is not known up to now. Example 6 shows that the results proposed in this paper can be used to solve this slippage problem.  $\diamond$

In the most difficult case, the vectors  $\theta_i$ 's are not orthogonal and/or they have different norms, the Gaussian noise has a general covariance matrix  $\Sigma$  and there is an unknown nuisance parameter  $\mathbf{x}$ . To our knowledge, the optimal constrained minimax test is unknown in this case. The main reasons which explain this lack of results are the followings. First, the presence of linear nuisance parameters certainly complicates the mutual geometry between the vectors. Next, there is an unavoidable antagonism between the detection and classification performances of the test. For example, to get small classification errors, it is necessary to accept a loss of sensibility for the probability of detection. The tradeoff between these two requirements is essentially based on the worst case of detection and the worst case of classification which are generally difficult to identify. Finally, as underlined in [24]–[27], the analytic calculation of the miss detection probability and the classification error probabilities are intractable, which makes difficult the derivation of an optimal test.

*Example 2 (Integrity Monitoring of Navigation Systems):* Let  $\mathbf{y}$  be a Gaussian random vector with the known covariance matrix  $\Sigma$ . Under hypothesis  $\mathcal{H}_0$ , its mean is  $\mathbf{H}\mathbf{x}$  where  $\mathbf{x}$  is the user unknown parameters and  $\mathbf{H}$  is a matrix describing the measurement system [5]. Under hypothesis  $\mathcal{H}_j$ ,  $y_j$  is contaminated by a scalar error with intensity  $\Delta_j$ . The common solution, namely the parity space approach, involves two steps. First, the user unknown parameters are eliminated by projecting  $\mathbf{y}$  on the null-space of  $\mathbf{H}$ . This null-space is called the parity space in the analytical redundancy literature [28]. Next, the error is detected and classified (isolated) directly in the parity space. Unfortunately, the first step may generate some linear dependencies between the possible error signatures in the parity space. In this case, the problem is not theoretically solved. Example 7 shows that this paper proposes a solution to this problem.  $\diamond$

*Example 3 (New User Identification in a Multiuser System):* In a multiuser system, after chip-matched filtering and chip rate sampling, the received signal vector under hypothesis  $\mathcal{H}_0$  is modeled as  $\mathbf{y} = \mathbf{S}\mathbf{a}\mathbf{x} + \xi$  where the  $j$ th column of the  $n \times m$  matrix  $\mathbf{S}$  is the normalized unit energy signature waveform vector

of user  $j$ ,  $\mathbf{A}$  is the diagonal  $m \times m$  matrix of user amplitudes and  $\mathbf{x}$  is the vector whose  $j$ th component is the antipodal symbol,  $-1$  or  $1$ , transmitted by user  $j$  [29], [30]. The random vector  $\boldsymbol{\xi}$  has zero mean and the known covariance matrix  $\sigma^2 \mathbf{I}_n$  where  $\mathbf{I}_n$  is the identity matrix of size  $n$ . Under hypothesis  $\mathcal{H}_j$ , the vector  $a_{m+1}x_{m+1}\mathbf{s}_j$  is added to  $\mathbf{y}$ , i.e., a new user with the signature  $\mathbf{s}_j$  emits the symbol  $x_{m+1}$  with the amplitude  $a_{m+1}$ . It is assumed that  $\mathbf{s}_j$  belongs to a finite set of predefined nonorthogonal signatures. The multiple-access interferences  $\mathbf{S}\mathbf{A}\mathbf{x}$  can be eliminated by using the above mentioned parity space approach [31]. The goal is to detect the new user arrival and to identify the waveform  $\mathbf{s}_j$ . This is a difficult problem, especially when the common length  $n$  of each user's signature is shorter than the total number of simultaneously active users [32]. Section VI will further elaborate upon this example.  $\diamond$

When the optimal statistical test is unknown or intractable, it is often assumed that the optimal weights are equal (since it is the least informative *a priori* choice) and the threshold is tuned to satisfy the false alarm constraint. The resulting test is called the  $M$ -ary Generalized Likelihood Ratio Test (MGLRT) between equally probable hypotheses. It must be noted that the optimality proof of the MGLRT is still an open problem.

### C. Contribution and Organization of the Paper

The first contribution is the design of a constrained  $\varepsilon$ -minimax detection/classification test solving the detection/classification problem in the case of nonorthogonal vectors with linear nuisance parameters and an additive Gaussian noise with a known general covariance matrix. This test is based on the maximum of weighted likelihood ratios, i.e., it is a Bayesian test associated to some specific weights. It is  $\varepsilon$ -optimal (under mild assumptions) in the sense that it is optimal with a loss of a small part, say  $\varepsilon$ , of optimality with respect to the purely theoretical minimax test. This loss of optimality, which is theoretically bounded, is unavoidable since the purely theoretical minimax test is intractable due to the difficulties above mentioned. This loss is even more negligible as the signal-to-noise ratio is large. It is also shown that this test coincides with a constrained  $\varepsilon$ -equalizer Bayesian test which equalizes the classification error probabilities over the alternative hypotheses up to a constant  $\varepsilon$ .

Secondly, an algorithm is proposed to compute the optimal weights of the proposed test with a reasonable numerical complexity. This algorithm is based on a graph, namely the separability map, describing the mutual geometry between the vectors. This separability map serves to identify the least separable vectors, making possible the design of the associated constrained  $\varepsilon$ -minimax test. This map is also used to calculate in advance the asymptotic maximum classification error probability of the constrained  $\varepsilon$ -minimax test as the Signal-to-Noise Ratio (SNR) tends to infinity. Moreover, in the case of large SNR, an asymptotically equivalent test is proposed whose optimal weights have a very simple form.

Finally, it is shown that the MGLRT is  $\varepsilon$ -optimal when the mutual geometry between the hypotheses is very simple, i.e., when each vector has at most one other vector nearest to it in term of Euclidean distance. In general, the MGLRT is suboptimal and the loss of optimality may be significant.

The paper is organized as follows. Section II starts with the problem statement and introduces the statistical framework that will be used in this paper, including the presentation of the constrained  $\varepsilon$ -minimax criterion. Section III describes the general methodology to reduce the detection/classification problem between multiple hypotheses with nuisance parameters to a detection/classification problem between multiple hypotheses without nuisance parameters. This reduction is based on the fact that it is sufficient to design a constrained  $\varepsilon$ -equalizer Bayesian test to get the constrained  $\varepsilon$ -minimax one. Section IV defines the separability map and proposes the main theorem of this paper which establishes the constrained  $\varepsilon$ -equalizer Bayesian test. The proof of this theorem is given in Appendix A. The false alarm probability of this test is calculated in Appendix B. Section V gives an asymptotically equivalent test to the constrained  $\varepsilon$ -minimax as the SNR tends to infinity. The derivation of this test is given in Appendix C. It is also shown that the MGLRT is asymptotically  $\varepsilon$ -optimal when the mutual geometry between the hypotheses is very simple. Section VI deals with a practical problem, namely the identification of a new active user in a multiuser system, showing the efficiency of the proposed  $\varepsilon$ -optimal test. Finally, Section VII concludes this paper.

## II. PROBLEM STATEMENT

This section presents the multiple hypotheses testing problem which consists in detecting and classifying a vector in the presence of linear nuisance parameters. A new optimality criterion, namely the constrained  $\varepsilon$ -minimax criterion, is introduced and motivated.

### A. Multiple Hypotheses Testing

The observation model has the form (1). Without any loss of generality, it is assumed that the noise vector  $\boldsymbol{\xi}$  follows a zero-mean Gaussian distribution  $\mathcal{N}(\mathbf{0}, \mathbf{I}_n)$ . In fact, it is always possible to multiply (1) on the left by the inverse square-root matrix  $\boldsymbol{\Sigma}^{-\frac{1}{2}}$  of  $\boldsymbol{\Sigma}$  to obtain the linear Gaussian model with the vector  $\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\theta}$ , the nuisance matrix  $\boldsymbol{\Sigma}^{-\frac{1}{2}}\mathbf{H}$  and a Gaussian noise having the covariance matrix  $\mathbf{I}_n$ . It is also assumed that  $\mathbf{H}$  is a full-column rank matrix. If the matrix  $\mathbf{H}$  does not satisfy this assumption, it suffices to keep the maximum number of linear independent columns to get a full-column rank matrix spanning the same linear space. It is then desirable to solve the multiple Gaussian hypotheses testing problem between the statistical hypotheses  $\mathcal{H}_0, \mathcal{H}_1, \dots, \mathcal{H}_r$

$$\mathcal{H}_i : \{ \mathbf{y} \sim \mathcal{N}(\boldsymbol{\psi}, \mathbf{I}_n), \boldsymbol{\psi} \in \Psi_i \} \tag{2}$$

where  $\Psi_i = \{ \boldsymbol{\theta}_i + \mathbf{H}\mathbf{x} : \mathbf{x} \in \mathbb{R}^m \}$ . The following condition of separability is assumed to be satisfied:

$$\forall (\boldsymbol{\theta}, \boldsymbol{\theta}') \in \Theta^2, \boldsymbol{\theta} \neq \boldsymbol{\theta}' \Rightarrow \boldsymbol{\theta} - \boldsymbol{\theta}' \neq \mathbf{H}\mathbf{x} \forall \mathbf{x} \in \mathbb{R}^m. \tag{3}$$

In other words, it is assumed that the intersection of the two linear manifolds  $\Psi_i$  and  $\Psi_j$  (which are parallel to each other) is an empty set for all  $i \neq j$  (two parallel linear manifolds with nonempty intersection are equal). Here, the parameter set  $\Psi_k$  associated to hypothesis  $\mathcal{H}_k$  is not a singleton, hence,  $\mathcal{H}_k$  is called a composite hypothesis [33]. Otherwise, the hypothesis is simple and it is identified by the absence of an underscore, say

$\mathcal{H}_k$ . The set of decision strategies for the  $(r+1)$ -ary hypotheses testing problem (2) is specified by the set of test functions.

*Definition 1:* A test function  $\phi = (\phi_0, \dots, \phi_r)^T$  for the multiple hypotheses  $\mathcal{H}_0, \mathcal{H}_1, \dots, \mathcal{H}_r$  is a  $(r+1)$ -dimensional vector function defined on  $\mathbb{R}^n$  such that  $\phi(\mathbf{y}) \in [0, 1]^{r+1}$  and

$$\sum_{i=0}^r \phi_i(\mathbf{y}) = 1, \quad \forall \mathbf{y} \in \mathbb{R}^n.$$

Given  $\mathbf{y} \in \mathbb{R}^n$ , the test function  $\phi$  decides the hypothesis  $\mathcal{H}_k$  if and only if  $\phi_k(\mathbf{y}) = 1$ . In this case, the other components  $\phi_j(\mathbf{y}), j \neq k$ , are zero. The study of randomized test functions (a random test function satisfies  $0 < \phi_k(\mathbf{y}) < 1$  for some  $k$ ) is not considered in this paper since the probability distribution of  $\mathbf{y}$  is continuous whatever the true hypothesis. The average performance of a particular test function  $\phi$  is determined by the  $r+1$  functions  $\beta_i(\boldsymbol{\psi}) = \mathbb{E}_{\boldsymbol{\psi}}[\phi_i(\mathbf{y})]$  where  $\mathbb{E}_{\boldsymbol{\psi}}[\phi_i(\mathbf{y})]$  stands for the expectation of  $\phi_i$  when  $\mathbf{y}$  follows the distribution  $\mathcal{N}(\boldsymbol{\psi}, \mathbf{I}_n)$ . The false alarm probability function is given by  $\alpha_0(\boldsymbol{\psi}) = 1 - \beta_0(\boldsymbol{\psi})$  when  $\boldsymbol{\psi} \in \Psi_0$ . The function  $\beta_0(\boldsymbol{\psi})$  for  $\boldsymbol{\psi} \notin \Psi_0$  describes the probability of miss detection. When  $\boldsymbol{\psi} \in \Psi_k, \alpha_k(\boldsymbol{\psi}) = 1 - \beta_k(\boldsymbol{\psi})$  is the classification error probability. The maximum classification error probability for the test function  $\phi$  is denoted

$$\alpha_{\max} = \max_{1 \leq k \leq r} \sup_{\boldsymbol{\psi} \in \Psi_k} \alpha_k(\boldsymbol{\psi}).$$

For a test  $\phi^*$  where  $\star$  is a given superscript, all the above mentioned notations are completed by the superscript  $\star$ . Let  $\mathcal{D}_\alpha$  be the set of test functions  $\phi = (\phi_0, \dots, \phi_r)^T$  whose maximum false alarm probability is less or equal to  $\alpha \in [0, 1]$

$$\mathcal{D}_\alpha = \left\{ \phi : \sum_{i=0}^r \phi_i(\mathbf{y}) = 1, \sup_{\boldsymbol{\psi} \in \Psi_0} \alpha_0(\boldsymbol{\psi}) \leq \alpha \right\}.$$

### B. Constrained Epsilon-Minimax Test

As mentioned in the introduction, the constrained minimax criterion given in [18] is a very natural criterion for problem (2). Unfortunately, as underlined in [18], since the geometry of the set  $\Psi = \cup_{i=0}^r \Psi_i$  may be very complex, it is impossible to infer the structure of the minimax test. Hence, to overcome this difficulty, it makes sense to consider constrained  $\varepsilon$ -minimax tests, i.e., tests that approximate optimal minimax test with a small loss, say  $\varepsilon$ , of optimality.

*Definition 2:* A test function  $\phi^* = (\phi_0^*, \dots, \phi_r^*)^T$  is a constrained  $\varepsilon$ -minimax test in the class  $\mathcal{D}_\alpha$  between the hypotheses  $\mathcal{H}_0, \mathcal{H}_1, \dots, \mathcal{H}_r$  if the following conditions are fulfilled :

- i)  $\sup_{\boldsymbol{\psi} \in \Psi_0} \alpha_0^*(\boldsymbol{\psi}) \leq \alpha$ ;
- ii) There exists a positive function  $\varepsilon(x)$  satisfying  $\varepsilon(x) \rightarrow \varepsilon$  as  $x \rightarrow 0$  such that

$$\alpha_{\max}^* (1 - \varepsilon(\alpha_{\max}^*)) \leq \alpha_{\max}$$

for any other test function  $\phi = (\phi_0, \dots, \phi_r)^T \in \mathcal{D}_\alpha$ .

Obviously, Definition 2 assumes that the positive constant  $\varepsilon$  is (very) small. In some cases, it is possible to get  $\varepsilon = 0$  (see Section VI) but, generally,  $\varepsilon > 0$  because of the vector constellation complexity. Contrary to a purely constrained minimax test, the design of a constrained  $\varepsilon$ -minimax test tolerates small errors on the classification error probabilities. Hence, it becomes possible to use lower and upper bounds on these probabilities in order to evaluate the statistical performances of the test. As underlined in [25], the exact calculation of these probabilities is generally intractable.

### III. EPSILON-MINIMAX TEST FOR COMPOSITE HYPOTHESES

This section introduces the constrained  $\varepsilon$ -equalizer Bayesian test of level  $\alpha$ . Proposition 1 shows that such a test is necessarily a constrained  $\varepsilon$ -minimax one. The first step to design the constrained  $\varepsilon$ -equalizer Bayesian test between composite hypotheses in presence of nuisance parameters consists in eliminating these unknown parameters. Proposition 2 shows that this elimination, based on the nuisance parameters rejection, leads to a reduced decision problem between simple statistical hypotheses.

#### A. Constrained Epsilon-Equalizer Test

Let us recall the definition of the constrained Bayesian test before introducing the definition of the constrained  $\varepsilon$ -equalizer test. Let  $\Lambda$  be a probability distribution over  $\Psi$  called the *a priori* distribution. For all  $i \in \{0, \dots, r\}$ , this distribution induces some *a priori* distributions  $\Lambda_i$  on the linear manifolds  $\Psi_i$  and some *a priori* probabilities  $p_i = \Pr(\mathcal{H}_i)$  such that

$$\Lambda(\boldsymbol{\psi}) = \sum_{i=0}^r p_i \Lambda_i(\boldsymbol{\psi}). \quad (4)$$

Let  $f_{\boldsymbol{\psi}}(\mathbf{y})$  be the probability density function (pdf) of the observation vector  $\mathbf{y}$  following the distribution  $\mathcal{N}(\boldsymbol{\psi}, \mathbf{I}_n)$ . To each hypothesis  $\mathcal{H}_i$  is associated the weighted pdf  $h_{\Lambda_i}(\mathbf{y})$  (see details in [10]) defined by

$$h_{\Lambda_i}(\mathbf{y}) = \int_{\Psi_i} f_{\boldsymbol{\psi}}(\mathbf{y}) d\Lambda_i(\boldsymbol{\psi}).$$

Let  $g_i^{(\Lambda)}(\mathbf{y})$  be the weighted log-likelihood ratio defined by

$$g_i^{(\Lambda)}(\mathbf{y}) = \ln \left( p_i \frac{h_{\Lambda_i}(\mathbf{y})}{h_{\Lambda_0}(\mathbf{y})} \right) \quad (5)$$

for  $1 \leq i \leq r$ . The constrained Bayesian test function  $\phi^{(\Lambda)} = (\phi_0^{(\Lambda)}, \dots, \phi_r^{(\Lambda)})^T$  of level  $\alpha$  associated to  $\Lambda$  is given by

$$\phi_0^{(\Lambda)}(\mathbf{y}) = \begin{cases} 1, & \text{if } \max_{i>0} \{g_i^{(\Lambda)}(\mathbf{y})\} \leq \lambda \\ 0, & \text{if } \max_{i>0} \{g_i^{(\Lambda)}(\mathbf{y})\} > \lambda \end{cases} \quad (6)$$

and for  $j = 1, \dots, r$

$$\phi_j^{(\Lambda)}(\mathbf{y}) = \begin{cases} 1, & \text{if } g_j^{(\Lambda)}(\mathbf{y}) = \max_{i>0} \{g_i^{(\Lambda)}(\mathbf{y})\} > \lambda \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

where the threshold  $\lambda = \lambda^{(\Lambda)} \geq 0$  is selected to satisfy the constraint

$$\alpha_0^{(\Lambda)} = \int_{\Psi_0} \mathbb{E}_{\psi}[1 - \phi_0^{(\Lambda)}(\mathbf{y})] d\Lambda_0(\psi) = \alpha. \quad (8)$$

The following  $\varepsilon$ -equalization criterion serves to design a constrained  $\varepsilon$ -minimax test.

*Definition 3:* A test function  $\boldsymbol{\phi}^* = (\phi_0^*, \dots, \phi_r^*)^T$  is a constrained  $\varepsilon$ -equalizer test between the hypotheses  $\mathcal{H}_0, \mathcal{H}_1, \dots, \mathcal{H}_r$  in the class  $\underline{\mathcal{D}}_\alpha$  if the following conditions are fulfilled :

- i)  $\sup_{\psi \in \Psi_0} \alpha_0^*(\psi) \leq \alpha$ ;
- ii) There exists a positive function  $\varepsilon(x)$  satisfying  $\varepsilon(x) \rightarrow \varepsilon$  as  $x \rightarrow 0$  such that

$$\max_{1 \leq k \leq r} \sup_{\psi \in \Psi_k} |\alpha_k^*(\psi) - \alpha_{\max}^*| \leq \alpha_{\max}^* \varepsilon(\alpha_{\max}^*).$$

*Proposition 1:* Let  $\boldsymbol{\phi}^{(\Lambda)}(\mathbf{y})$  be the Bayesian test function of level  $\alpha$  associated to the *a priori* distribution  $\Lambda$  such that  $\boldsymbol{\phi}^{(\Lambda)}(\mathbf{y})$  is a constrained  $\varepsilon$ -equalizer test between  $\mathcal{H}_0, \mathcal{H}_1, \dots, \mathcal{H}_r$ . Then, for any other test function  $\boldsymbol{\phi} = (\phi_0, \dots, \phi_r)^T \in \underline{\mathcal{D}}_\alpha$  for testing among the hypotheses  $\mathcal{H}_0, \mathcal{H}_1, \dots, \mathcal{H}_r$ , it follows that

$$\alpha_{\max}^{(\Lambda)} (1 - \varepsilon(\alpha_{\max}^{(\Lambda)})) \leq \alpha_{\max}$$

where  $\varepsilon(x) \rightarrow \varepsilon$  as  $x \rightarrow 0$ .

*Proof:* Let  $\boldsymbol{\phi}$  be an arbitrary test in  $\underline{\mathcal{D}}_\alpha$  for testing among the hypotheses  $\mathcal{H}_0, \mathcal{H}_1, \dots, \mathcal{H}_r$ . Let  $R^{(\Lambda)}(\boldsymbol{\phi})$  be the Bayesian risk of the test  $\boldsymbol{\phi}$  for the *a priori* distribution  $\Lambda$

$$R^{(\Lambda)}(\boldsymbol{\phi}) \stackrel{\text{def.}}{=} p_0 \int_{\Psi_0} \alpha_0(\psi) d\Lambda_0(\psi) + \sum_{k=1}^r p_k \int_{\Psi_k} \alpha_k(\psi) d\Lambda_k(\psi).$$

From the definition of the Bayesian test [9], [18], we get  $R^{(\Lambda)}(\boldsymbol{\phi}^{(\Lambda)}) \leq R^{(\Lambda)}(\boldsymbol{\phi})$ . Since  $\int_{\Psi_0} \alpha_0(\psi) d\Lambda_0(\psi) \leq \alpha$  and  $\int_{\Psi_0} \alpha_0^{(\Lambda)}(\psi) d\Lambda_0(\psi) = \alpha$ , we get

$$R_i^{(\Lambda)}(\boldsymbol{\phi}^{(\Lambda)}) \leq R_i^{(\Lambda)}(\boldsymbol{\phi}) \quad (9)$$

where

$$R_i^{(\Lambda)}(\boldsymbol{\phi}) = \sum_{k=1}^r p_k \int_{\Psi_k} \alpha_k(\psi) d\Lambda_k(\psi).$$

Since  $\boldsymbol{\phi}^{(\Lambda)}(\mathbf{y})$  is a constrained  $\varepsilon$ -equalizer test, we get for all  $k$  and  $\boldsymbol{\psi}$

$$|\alpha_k^{(\Lambda)}(\boldsymbol{\psi}) - \alpha_{\max}^{(\Lambda)}| \leq \alpha_{\max}^{(\Lambda)} \varepsilon(\alpha_{\max}^{(\Lambda)})$$

where  $\varepsilon(x) \rightarrow \varepsilon$  as  $x \rightarrow 0$ . It follows that

$$|R_i^{(\Lambda)}(\boldsymbol{\phi}^{(\Lambda)}) - (1-p_0)\alpha_{\max}^{(\Lambda)}| \leq (1-p_0)\alpha_{\max}^{(\Lambda)} \varepsilon(\alpha_{\max}^{(\Lambda)}). \quad (10)$$

One also obtain

$$R_i^{(\Lambda)}(\boldsymbol{\phi}) \leq (1-p_0)\alpha_{\max}. \quad (11)$$

Hence, (9), (10) and (11) yield to

$$\alpha_{\max}^{(\Lambda)} (1 - \varepsilon(\alpha_{\max}^{(\Lambda)})) \leq \alpha_{\max}$$

which ends the proof.  $\blacksquare$

### B. Reduction to Epsilon-Equalizer Test for Simple Hypotheses

The presence of linear nuisance parameters complicates the statistical decision problem. The theory of invariance [5], [9], [10] is generally used to obtain a test which is independent of them. It is well-known that the invariance principle leads to some reduced (simplified) decision problems for which the optimal decision test is often a minimax one [10, chap. 8]. Hence, it is proposed to eliminate the nuisance parameters by using the invariance principle and to show that the constrained  $\varepsilon$ -minimax test can be derived from the reduced problem under certain conditions.

The family of distributions  $\mathbf{y} \sim \mathcal{N}(\boldsymbol{\theta} + \mathbf{H}\mathbf{x}, \mathbf{I}_n)$  for  $\boldsymbol{\theta} \in \Theta$  and  $\mathbf{x} \in \mathbb{R}^m$  remains invariant (see [10] for details and definitions) under the group of translations

$$\mathcal{G} = \{g : g(\mathbf{y}) = \mathbf{y} + \mathbf{H}\mathbf{c}, \mathbf{c} \in \mathbb{R}^m\}$$

which induces in the parameter space  $\mathbb{R}^n$  the group

$$\bar{\mathcal{G}} = \{\bar{g} : \bar{g}(\boldsymbol{\theta}) = \boldsymbol{\theta} + \mathbf{H}\mathbf{c}, \mathbf{c} \in \mathbb{R}^m\}$$

that preserves all the sets  $\Psi_i$ , i.e.,  $\bar{g}\Psi_i = \Psi_i$  for all  $i$  and  $\bar{g} \in \bar{\mathcal{G}}$ . Hence, the hypotheses testing problem (2) remains invariant under  $\mathcal{G}$ . It is shown in [5] that  $\mathbf{z} = \mathbf{W}^T \mathbf{y}$  is a maximal invariant to the group of translations  $\mathcal{G}$  where the matrix  $\mathbf{W} = (\mathbf{w}_1, \dots, \mathbf{w}_{n-m})$  of size  $n \times (n-m)$  is composed of the eigenvectors  $\mathbf{w}_1, \dots, \mathbf{w}_{n-m}$  of the projection matrix  $\mathbf{P}_{\mathbf{H}}^\perp = \mathbf{I}_n - \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T$  corresponding to eigenvalue 1. Here,  $\mathbf{A}^{-1}$  denotes the inverse of  $\mathbf{A}$ . The matrix  $\mathbf{W}$  satisfies the following conditions:

$$\mathbf{W}^T \mathbf{H} = \mathbf{0}, \quad \mathbf{W}\mathbf{W}^T = \mathbf{P}_{\mathbf{H}}^\perp, \quad \mathbf{W}^T \mathbf{W} = \mathbf{I}_{n-m}. \quad (12)$$

The multiple hypotheses testing problem between composite hypotheses (2) is reduced to the following multiple Gaussian hypotheses testing problem between simple hypotheses

$$\mathcal{H}_i : \{\mathbf{z} \sim \mathcal{N}(\boldsymbol{\varphi}_i, \mathbf{I}_q)\} \quad (13)$$

for  $0 \leq i \leq r$  where  $\boldsymbol{\varphi}_i = \mathbf{W}^T \boldsymbol{\theta}_i$  and  $q = n - m$ . From (3), it follows that  $\boldsymbol{\varphi}_i \neq \boldsymbol{\varphi}_j$  for all  $i \neq j$ . Let  $\Phi = \{\boldsymbol{\varphi}_0 = \mathbf{0}, \boldsymbol{\varphi}_1, \dots, \boldsymbol{\varphi}_r\}$  be the set of vectors signatures after the nuisance parameter rejection. The Euclidean norm of  $\boldsymbol{\varphi}_i$  is  $\|\boldsymbol{\varphi}_i\|^2 = \boldsymbol{\varphi}_i^T \boldsymbol{\varphi}_i = b_i^2$  where  $b_i > 0$  is known for all  $i$ .

The following proposition shows that a constrained  $\varepsilon$ -equalizer Bayesian test for the reduced problem (13) is a constrained  $\varepsilon$ -minimax test for the initial problem (2). Before presenting this proposition, let us give the simplified form of the constrained Bayesian test function (6)–(8) of level  $\alpha$  applied to the reduced

problem (13). Obviously, this test function is defined on  $\mathbb{R}^q$  instead of  $\mathbb{R}^n$  (see Definition 1) and the  $r + 1$  functions  $\alpha_i(\boldsymbol{\varphi}) = \mathbb{E}_{\boldsymbol{\varphi}}[1 - \phi_i(\mathbf{z})]$  depend on  $\boldsymbol{\varphi} \in \Phi$  instead of  $\boldsymbol{\psi} \in \Psi$ . Define the  $r$ -dimensional unit simplex

$$\mathcal{S}_r = \left\{ \mathbf{p} \in [0, 1]^r : \sum_{i=1}^r p_i = 1 \right\}.$$

For the problem (13), each *a priori* distribution  $\Lambda$  is equivalent to a weight vector  $\mathbf{p} \in \mathcal{S}_r$ . For arbitrary  $\mathbf{p} \in \mathcal{S}_r$ , (5) is rewritten as follows:

$$g_i^{(\mathbf{p})}(\mathbf{z}) = \mathbf{z}^T \boldsymbol{\varphi}_i - \frac{b_i^2}{2} + \log p_i. \quad (14)$$

*Proposition 2:* Let  $\boldsymbol{\phi}^*(\mathbf{z}) = \boldsymbol{\phi}^{(\mathbf{p}^*)}(\mathbf{z})$  be the Bayesian test function of level  $\alpha$  associated to the *a priori* distribution  $\mathbf{p}^* \in \mathcal{S}_r$  such that  $\boldsymbol{\phi}^{(\mathbf{p}^*)}(\mathbf{z})$  is a constrained  $\varepsilon$ -equalizer test between  $\mathcal{H}_0, \mathcal{H}_1, \dots, \mathcal{H}_r$ , i.e., there exists a positive function  $\varepsilon(x)$  satisfying  $\varepsilon(x) \rightarrow \varepsilon$  as  $x \rightarrow 0$  such that

$$\max_{1 \leq k \leq r} \left| \alpha_k^{(\mathbf{p}^*)}(\boldsymbol{\varphi}_k) - \alpha_{\max}^{(\mathbf{p}^*)} \right| \leq \alpha_{\max}^{(\mathbf{p}^*)} \varepsilon(\alpha_{\max}^{(\mathbf{p}^*)}).$$

Then  $\boldsymbol{\phi}^*(\mathbf{y}) \stackrel{\text{def.}}{=} \boldsymbol{\phi}^*(\mathbf{W}^T \mathbf{y})$  is a constrained  $\varepsilon$ -minimax test in the class  $\underline{\mathcal{D}}_\alpha$  between hypotheses  $\underline{\mathcal{H}}_0, \underline{\mathcal{H}}_1, \dots, \underline{\mathcal{H}}_r$ .

*Proof:* Let  $\mathbf{x}_i = -(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \boldsymbol{\theta}_i$  for all  $0 \leq i \leq r$ . Let  $\Lambda^*$  be the *a priori* distribution defined over  $\Psi$  by  $\Lambda^*(\boldsymbol{\psi}) = \sum_{i=0}^r p_i^* \Lambda_i^*(\boldsymbol{\psi})$  where

$$\Lambda_i^*(\boldsymbol{\psi}) = \begin{cases} 1, & \text{if } \boldsymbol{\psi} = \boldsymbol{\theta}_i + \mathbf{H}\mathbf{x}_i = \mathbf{P}_{\mathbf{H}}^\perp \boldsymbol{\theta}_i \\ 0, & \text{else.} \end{cases}$$

Noting that  $\mathbf{z}^T \boldsymbol{\varphi}_i = \mathbf{y}^T \mathbf{P}_{\mathbf{H}}^\perp \boldsymbol{\theta}_i$ , a short algebra immediately shows that  $\boldsymbol{\phi}^*(\mathbf{y}) = \boldsymbol{\phi}^{(\Lambda^*)}(\mathbf{y}) = \boldsymbol{\phi}^*(\mathbf{W}^T \mathbf{y})$  for all  $\mathbf{y} \in \mathbb{R}^n$ . Hence, the test function  $\boldsymbol{\phi}^{(\Lambda^*)}(\mathbf{y})$  is independent of the nuisance parameter  $\mathbf{x}$  since the statistical distribution of  $\mathbf{W}^T \mathbf{y}$  is independent of  $\mathbf{x}$ . Consequently,  $\boldsymbol{\phi}^{(\Lambda^*)}(\mathbf{y})$  is a constrained  $\varepsilon$ -equalizer test in  $\underline{\mathcal{D}}_\alpha$  between  $\underline{\mathcal{H}}_0, \underline{\mathcal{H}}_1, \dots, \underline{\mathcal{H}}_r$ . The proof is ended by using Proposition 1.  $\blacksquare$

According to the proof of Proposition 2, the loss of optimality, say  $\varepsilon_i$ , for the  $\varepsilon$ -equalizer test between  $\underline{\mathcal{H}}_0, \underline{\mathcal{H}}_1, \dots, \underline{\mathcal{H}}_r$  is equal to the loss of optimality, say  $\varepsilon_r$ , for the  $\varepsilon$ -equalizer test between  $\mathcal{H}_0, \mathcal{H}_1, \dots, \mathcal{H}_r$ , i.e.,  $\varepsilon_i = \varepsilon_r$ . This comes from the fact that the rejection principle is equivalent to assume that the parameter  $\boldsymbol{\psi}$  follows under  $\underline{\mathcal{H}}_i$  a degenerate *a priori* distribution  $\Lambda_i^*(\boldsymbol{\psi})$  whose support consists of only one value  $\mathbf{P}_{\mathbf{H}}^\perp \boldsymbol{\theta}_i$  for all  $i=1, \dots, r$ . Strictly speaking, it is certainly possible to reduce  $\varepsilon_i$  by designing and using some nondegenerate *a priori* distributions  $\Lambda_i^*(\boldsymbol{\psi})$  specifically adapted to the statistical performances of the  $\varepsilon$ -equalizer test between  $\mathcal{H}_0, \mathcal{H}_1, \dots, \mathcal{H}_r$ . However, in practice, the expected diminution  $\varepsilon_r - \varepsilon_i$  in the loss of optimality is not sufficiently significant to justify the effort required to look for such specific *a priori* distributions  $\Lambda_i^*(\boldsymbol{\psi})$ .

#### IV. EPSILON-EQUALIZER TEST FOR SIMPLE HYPOTHESES

The main result of this section is Theorem 1. It proposes a constrained  $\varepsilon$ -equalizer Bayesian test for the problem (13),

which means the calculation of the weight vector  $\mathbf{p}^* \in \mathcal{S}_r$  and the threshold  $\lambda^*$  for the test  $\boldsymbol{\phi}^{(\mathbf{p}^*)}(\mathbf{z})$  in Proposition 2.

##### A. Principle of Epsilon-Equalization

Designing a constrained  $\varepsilon$ -equalizer Bayesian test  $\boldsymbol{\phi}^*(\mathbf{z})$  for the problem (13) is difficult for mainly three reasons: i) the common value  $\alpha_{\max}^*$  of the classification error probabilities is unknown, ii) the weight vector  $\mathbf{p}^*$  ensuring the  $\varepsilon$ -equalization of the classification error probabilities is not easily calculable and iii) the threshold  $\lambda^*$  must be chosen accordingly to the prescribed false alarm probability  $\alpha$ .

To overcome these difficulties, this paper proposes an original tool, namely the separability map, which describes the mutual geometry between the vectors. Loosely speaking, the separability map is a graph [34], [35] whose topology is deduced from the Euclidean distances between the vectors to be classified. In this map, the neighborhood of a vector, say  $\boldsymbol{\varphi}_i$ , is composed of the nearest vectors of  $\boldsymbol{\varphi}_i$  according to the Euclidean distance. The neighborhood of  $\boldsymbol{\varphi}_i$  characterizes the maximum classification error probability of  $\boldsymbol{\varphi}_i$ . There is a natural compromise between the number of neighbors of  $\boldsymbol{\varphi}_i$  and their distances to  $\boldsymbol{\varphi}_i$ . Hence, the separability map is used both to identify the least separable vectors and to calculate in advance the asymptotic maximum classification error probability  $\alpha_{\max}^*$  of the constrained  $\varepsilon$ -minimax test. Once  $\alpha_{\max}^*$  is known, the weights  $p_i^*$  associated to the least separable vectors are easily deduced from the separability map. The remaining weights are more difficult to calculate because they depend on the distances between all the remaining vectors. Thus, it is proposed to solve an original linear programming problem whose solution gives the remaining weights. Basically, the principle of this linear programming problem is to equalize the classification error probabilities by exploring the space of possible weights subject to some constraints depending on the distances between all the vectors. Finally, the threshold  $\lambda^*$  is deduced from  $\mathbf{p}^*$ . To the authors's knowledge, such an approach has never been addressed before in the literature.

##### B. Mutual Geometry for Simple Gaussian Hypotheses

The distance between the null hypothesis  $\mathcal{H}_0$  and the alternative one  $\mathcal{H}_i$  is defined as

$$d(\boldsymbol{\varphi}_i, \boldsymbol{\varphi}_0) = d_{i,0} = b_i - u_\alpha \quad (15)$$

where  $u_\alpha$  is a real depending on the prescribed false alarm probability  $\alpha$ . Generally,  $u_\alpha$  is a decreasing function of  $\alpha$  (see Theorem 1) but it is not crucial to derive the following results. The minimum detectability distance between the null hypothesis  $\mathcal{H}_0$  and the alternative hypotheses is given by

$$d_0 = d(\boldsymbol{\varphi}_0) = \min_{1 \leq i \leq r} d_{i,0} = b_{\min} - u_\alpha$$

where  $b_{\min} = \min_{1 \leq i \leq r} b_i$ . Let  $d_{i,j}$  be the distance between two alternative hypotheses  $\mathcal{H}_i$  and  $\mathcal{H}_j$  given by:

$$d_{i,j} = d(\boldsymbol{\varphi}_i, \boldsymbol{\varphi}_j) = \frac{1}{2} \|\boldsymbol{\varphi}_i - \boldsymbol{\varphi}_j\| = d_{j,i}, \quad \forall i, j > 0. \quad (16)$$

Finally, the separability value  $d_i$  of hypothesis  $\mathcal{H}_i$  is given by

$$d_i = d(\boldsymbol{\varphi}_i) = \min_{0 \leq j \neq i \leq r} d_{i,j}$$

and the minimum separability distance between all the alternative hypotheses is

$$d^* = \min_{1 \leq i \leq r} d_i.$$

*Definition 4:* The separability map  $\mathcal{M}$  associated to the set  $\Phi$  is the undirected graph  $\mathcal{M} = (\Phi, \Gamma)$ , where  $\Gamma \subset [\Phi]^2$  and  $[\Phi]^2$  is the set of all 2-subsets of  $\Phi$ , defined by

$$\forall (\varphi, \varphi') \in \Phi^2, \{\varphi, \varphi'\} \in \Gamma \Leftrightarrow d(\varphi, \varphi') = d(\varphi) = d(\varphi').$$

Strictly speaking, it would be more rigorous to precise that the separability map depends on the false alarm probability  $\alpha$ , for instance by using the terminology  $\alpha$ -separability map, but it is not necessary since  $\alpha$  is fixed *a priori*. Two vectors  $\varphi$  and  $\varphi'$  of  $\Phi$  are adjacent if  $\{\varphi, \varphi'\} \in \Gamma$ . Two vectors  $\varphi$  and  $\varphi'$  are connected if there exists a family of  $t$  different vectors  $\{\varphi_1, \varphi_2, \dots, \varphi_t\}$  such that  $\{\varphi_i, \varphi_{i+1}\} \in \Gamma$  for all  $1 \leq i < t$ ,  $\{\varphi, \varphi_1\} \in \Gamma$  and  $\{\varphi_t, \varphi'\} \in \Gamma$ . Since the connection is an equivalence relation [34], there is a partition of  $\Phi$  into  $s$  nonempty subsets  $\dot{\Phi}_1, \dot{\Phi}_2, \dots, \dot{\Phi}_s$  such that two vertices  $\varphi$  and  $\varphi'$  are connected if and only if both  $\varphi$  and  $\varphi'$  belong to the same subset  $\dot{\Phi}_i$ . Let  $\mathcal{C}_i = (\dot{\Phi}_i, \dot{\Gamma}_i)$  be the subgraph of  $\mathcal{M}$  such that  $\dot{\Gamma}_i$  contains all the 2-subsets  $\{\varphi, \varphi'\}$  of  $\Phi$  with  $\varphi, \varphi' \in \dot{\Phi}_i$ . The subgraphs  $\mathcal{C}_i$  are called the components of  $\mathcal{M}$ .

To avoid too complicated vector constellations (see Remark 2), it is assumed that

**A1)** All separability map components are star graphs.

The star graph of order  $m$  is a graph on  $m$  vectors with one vector, called the star's center, having  $m - 1$  adjacent vectors. The other  $m - 1$  vectors have only one adjacent vector, the star's center. The star's center of  $\mathcal{C}_i$  is denoted  $\varphi(\mathcal{C}_i)$ . The standard algorithms of the graph theory [34], [35] can be efficiently used to compute the separability map  $\mathcal{M}$  (representing by an adjacency matrix for example) and its components. Let the well-known function  $Q(\cdot)$  be defined by

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{+\infty} \exp\left(-\frac{u^2}{2}\right) du. \quad (17)$$

*Definition 5:* Consider the vector  $\varphi_i$  with  $i \neq 0$  and assume that it belongs to the component  $\mathcal{C}$  containing  $m + 1$  elements. The critical value  $c_i$  of  $\varphi_i$  is  $c_i = d_i - e_i$  where  $e_i \geq 0$  is the unique solution of one of the following equations:

- if  $\varphi_0 \notin \mathcal{C} : mQ(d_i + e_i) = Q(d_i - e_i)$ ;
- if  $\varphi_0 \in \mathcal{C}$  and  $\varphi_0 \neq \varphi(\mathcal{C})$

$$(m - 1)Q(d_i + e_i) + Q(d_i) = Q(d_i - e_i);$$

- if  $\varphi_0 = \varphi(\mathcal{C})$  or  $m \leq 1 : e_i = 0$ .

By definition of the separability map, all the vectors belonging to the same component have the same critical value. The critical value  $c_i$  characterizes the common maximum classification error  $Q(c_i)$  which can be achieved simultaneously by all the vectors within the component containing  $\varphi_i$ .

*Remark 1:* To justify Definition 5, it is necessary to establish the existence and uniqueness of  $e_i$ . Suppose  $\varphi_0 \notin \mathcal{C}$ . Let  $\tilde{Q}(x) = mQ(d_i + x) - Q(d_i - x)$  be defined for  $x \geq 0$ .

The function  $\tilde{Q}$  is a continuous and strictly decreasing function. The existence and uniqueness of  $e_i$  follows from  $\tilde{Q}(0) = (m - 1)Q(d_i) > 0$  and  $\tilde{Q}(x) \rightarrow -1$  as  $x \rightarrow +\infty$ . The same reasoning shows that the second case also has a unique solution.

The following lemma gives the asymptotic value of  $e_i$  when  $d_i$  becomes arbitrary large. Here, the notation  $f \sim g$  as  $d \rightarrow +\infty$  stands for  $\lim_{d \rightarrow +\infty} f(d)/g(d) = 1$  and  $o(f(d))$  stands for  $\lim_{d \rightarrow +\infty} o(f(d))/f(d) = 0$  as  $d \rightarrow +\infty$ .

*Lemma 1:* Suppose that the component  $\mathcal{C}$  contains  $m + 1$  elements with  $m \geq 2$ . As  $d_i \rightarrow +\infty$ , the real value  $e_i$  associated to  $\varphi_i \in \mathcal{C}$  satisfies the asymptotic relation  $e_i \sim e_i^\infty$  where:

- if  $\varphi_0 \notin \mathcal{C} : e_i^\infty = \frac{\ln m}{2d_i}$ ;
- if  $\varphi_0 \in \mathcal{C}$  and  $\varphi_0 \neq \varphi(\mathcal{C}) : e_i^\infty = \frac{\ln\left(\frac{1}{2} + \sqrt{m - \frac{3}{4}}\right)}{d_i}$ .

The asymptotic behavior  $e_i^\infty$  of  $e_i$  is unique in the sense that  $e_i$  has necessary the form  $e_i^\infty + o(d_i^{-1})$ .

*Proof:* For a large value of  $d$ , it is well-known that [29]

$$Q(d) = \frac{1}{d\sqrt{2\pi}} \exp\left(-\frac{d^2}{2}\right) (1 - d^{-2} + o(d^{-2})).$$

Hence, a bit of algebra yields to

$$\frac{mQ(d_i + e_i)}{Q(d_i - e_i)} = m \exp(-2d_i e_i) (1 + o(d_i^{-2})).$$

It follows that

$$\lim_{d_i \rightarrow +\infty} \frac{mQ(d_i + e_i)}{Q(d_i - e_i)} = 1$$

if and only if  $\lim_{d_i \rightarrow +\infty} \exp(-2d_i e_i + \ln m) = 1$ . This limit is achieved if and only if

$$2d_i e_i - \ln m = o(1).$$

This leads to  $e_i = \frac{\ln m + o(1)}{2d_i}$ .

Following the same reasoning, we get that

$$\begin{aligned} & \frac{(m - 1)Q(d_i + e_i) + Q(d_i)}{Q(d_i - e_i)} \\ &= (m - 1) \exp(-2d_i e_i) + \exp(-d_i e_i) + o(d_i^{-2}). \end{aligned} \quad (18)$$

Hence, (18) tends to 1 if and only if

$$(1 + o(1)) \exp(2d_i e_i) - \exp(d_i e_i) - (m - 1) = 0.$$

This leads to  $e_i = \frac{\ln(1/2 + \sqrt{m-3/4}) + o(1)}{d_i}$ . ■

The critical value  $c^*$  of  $\mathcal{M}$  is defined by

$$c^* \stackrel{\text{def.}}{=} \min_{1 \leq i \leq r} c_i. \quad (19)$$

It is clear that  $c^* \leq d^*$ . A critical component of  $\mathcal{M}$  is a component  $\mathcal{C}_i$  such that the common critical value of its elements is  $c^*$ .

*Lemma 2:* A critical component has at least two elements.

*Proof:* Let  $\mathcal{C}_i$  be a critical component such that all its elements have the common critical value  $c^*$ . Suppose that  $\mathcal{C}_i$  has only one element  $\varphi_j$ . Then, according to the definition of the critical value,  $e_j = 0$  and one obtain  $d_j = c_j$ . Since  $\mathcal{C}_i$  is a critical component, then  $c_j = c^*$ . Hence,  $c^* = d_j \geq d^*$ . Using



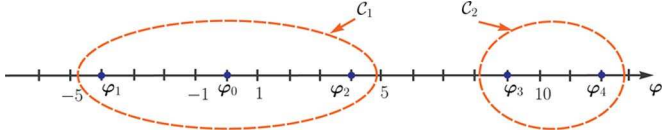


Fig. 1. Vector constellation with two components  $C_1$  and  $C_2$ .

the fact that  $c^* \leq d^*$ , it follows that  $c^* = c_j = d_j = d^*$ . The minimum value  $d^*$  is achieved at  $d_{j,k} = d^*$  for  $k \neq j$ . Hence, there exists an other element  $\varphi_k$  which necessarily belongs to  $C_i$  and, consequently,  $C_i$  has at least two elements. ■

The vectors are separated into three sets according to their role in the separability map. Let  $\mathcal{K}_s$  be the set of strongly critical vectors defined by

$$\mathcal{K}_s = \{i \in \{1, \dots, r\} : c_i = c^* \text{ and } \exists 1 \leq j \leq s, \varphi_i = \varphi(C_j)\} \quad (20)$$

$\mathcal{K}_w$  be the set of weakly critical vectors

$$\mathcal{K}_w = \{i \in \{1, \dots, r\} \setminus \mathcal{K}_s : c_i = c^*\} \quad (21)$$

and  $\mathcal{K}_n$  be the set of noncritical vectors

$$\mathcal{K}_n = \{1, \dots, r\} \setminus (\mathcal{K}_s \cup \mathcal{K}_w). \quad (22)$$

*Remark 2:* Assumption A1) may be not satisfied in practice. In such a case, the approach proposed above still works but it is more difficult to calculate the critical values  $c_i$ . In fact, when the component of  $\varphi_i$  is a star graph, the center is naturally the least separable vector and it is possible to calculate the common maximum classification error probability  $Q(c_i)$  of the component. When the component is not a star graph, the least separable vector is not easily identifiable and it becomes necessary to study in details the internal geometry of the component in order to calculate  $c_i$ .

*Example 4 (Discrete Location Parameter):* To illustrate the above definitions, this toy example considers 5 real 1-D vectors  $\varphi_0 = 0, \varphi_1 = -4, \varphi_2 = 4, \varphi_3 = 9$  and  $\varphi_4 = 12$  (see Fig. 1) with  $\alpha = 10^{-2}$  and  $u_\alpha = Q^{-1}(\alpha/5) = 2.807$ . Computing all the separability values  $d_i$  for  $i = 1, \dots, 5$  is straightforward:  $d_0 = d_1 = d_2 = 1.193$  and  $d_3 = d_4 = 1.5$ . The separability map  $\mathcal{M}$  and its two components  $C_1$  and  $C_2$  are represented in Fig. 2. According to Definition 5,  $e_i = 0$  for  $i = 1, \dots, 5$ . The critical component is  $C_1$  with star's center  $\varphi_0, \mathcal{K}_s = \emptyset, \mathcal{K}_w = \{1, 2\}$  and  $\mathcal{K}_n = \{3, 4\}$ . A solution to this detection and classification problem is discussed in Example 5. ◇

### C. Constrained Epsilon-Equalizer Test for Simple Hypotheses

To derive the constrained  $\varepsilon$ -equalizer test, some bounds on the classification error probabilities and also on the false alarm probability are used. Hence, to ensure the sharpness of these bounds and the relevance of the following results, the false alarm probability must be sufficiently high with respect to the SNR. For this reason, it is assumed that:

**A2)** The prescribed false alarm probability  $\alpha$  satisfies  $r Q(b_{\min}) < \alpha < 1$ .

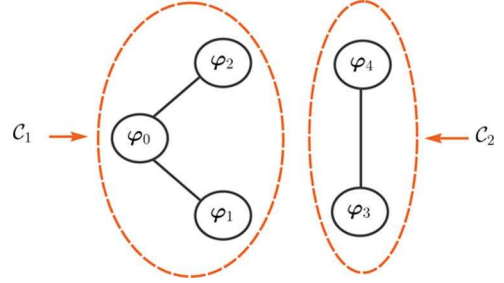


Fig. 2. Separability map  $\mathcal{M}$  and its two components  $C_1$  and  $C_2$ .

Let  $\mathcal{M}$  be the separability map associated to  $\mathcal{H}_0, \mathcal{H}_1, \dots, \mathcal{H}_r$  and  $\mathcal{K}_s, \mathcal{K}_w$  and  $\mathcal{K}_n$  be the sets defined by (20)-(21)-(22). Let  $p^*$  be the weight vector

$$p_i^* = \kappa e^{-z_i^*}, \quad i = 1, \dots, r \quad (23)$$

where  $\kappa = 1 / \sum_{i=1}^r e^{-z_i^*}$  is the normalizing constant. The coefficients  $z_i^*$  are defined as follows:  $z_i^* = 0$  for all  $i \in \mathcal{K}_s, z_i^* = 2d_i e_i$  for all  $i \in \mathcal{K}_w$  and the other positive coefficients  $z_i^*$  for  $i \in \mathcal{K}_n$  are the solutions of the linear programming problem (LP) defined by

$$\max \sum_{i \in \mathcal{K}_n} z_i \quad (24)$$

$$\text{s.t. } z_i \geq 0, \forall i \in \mathcal{K}_n \quad (25)$$

$$z_i \leq \frac{1}{2} d_{i,k} (d_{i,k} - \omega_\eta^*), \forall i \in \mathcal{K}_n, \forall k \in \mathcal{K}_s \cup \mathcal{K}_w \quad (26)$$

$$|z_i - z_k| \leq \frac{1}{2} d_{i,k} (d_{i,k} - \omega_\eta^*), \forall i \neq k \in \mathcal{K}_n \times \mathcal{K}_n \quad (27)$$

$$z_i \leq \frac{1}{2} (b_i^2 + b_{\min}^2) - u_\alpha b_{\min} - \omega_\eta^* b_i, \forall i \in \mathcal{K}_n \quad (28)$$

where  $c^*$  is the critical value of  $\mathcal{M}, \omega_\eta^* = c^* + \frac{\ln \eta}{c^*}$  and  $1 \leq \eta \leq r$ . A bit of calculus immediately shows that  $\kappa = p_{\max}^* = \max_{1 \leq i \leq r} p_i^*$  with

$$p_{\max}^* = \left( |\mathcal{K}_s| + \sum_{i \in \mathcal{K}_w} e^{-2d_i e_i} + \sum_{i \in \mathcal{K}_n} e^{-z_i^*} \right)^{-1} \quad (29)$$

where  $|\mathcal{A}|$  denotes the number of elements in the set  $\mathcal{A}$ .

The interpretation of the constraints is the following. Constraint (25) means that the weight of the noncritical vector  $\varphi_i$  is necessary smaller than the weight of a (strongly or weakly) critical vector. On the contrary, constraint (26) means that the weight of the noncritical vector  $\varphi_i$  does not be too small compared to the weight of a critical vector. Constraint (27) means that the difference between the weights of two noncritical vectors  $\varphi_i$  and  $\varphi_k$  must be balanced in regard to the geometric distances between them. Finally, constraint (28) means that the weight of the noncritical vector  $\varphi_i$  must not exceed a certain level imposed by the prescribed false alarm probability. If this constraint is violated, the maximum classification error probability increases due to the presence of the null hypothesis  $\mathcal{H}_0$ . More generally, if one of these constraints is violated, it is impossible to equalize the classification error probabilities. The value of  $b_{\min}$  is assumed to be large enough to warrant the positivity of constraints (26), (27) and (28). Let  $q_{\eta,i}$  be the number

of active inequality constraints (equality holds) for each coefficient  $z_i^*$  with  $i \in \mathcal{K}_n$ ,  $q_{\eta,\max} = \max_{i \in \mathcal{K}_n} q_{\eta,i}$  and  $q_{\eta,\min} = \min_{i \in \mathcal{K}_n} q_{\eta,i}$ . Let  $Q^{-1}(\cdot)$  be the inverse of the function  $Q(\cdot)$  given in (17).

*Theorem 1:* Under the assumptions A1) and A2), there exist a weight vector  $\mathbf{p}^* \in \mathcal{S}_r$  given by (23) and (29) and a threshold

$$\lambda^* = u_\alpha b_{\min} - \frac{b_{\min}^2}{2} + \ln p_{\max}^* \quad (30)$$

where  $u_\alpha = Q^{-1}(\alpha/r)$ , for which the test  $\phi^*(\mathbf{z}) \stackrel{\text{def.}}{=} \phi^{(\mathbf{p}^*)}(\mathbf{z})$  is a constrained  $\varepsilon$ -equalizer test between  $\mathcal{H}_0, \mathcal{H}_1, \dots, \mathcal{H}_r$  such that  $\alpha_0^*(\varphi_0) \leq \alpha$ . As  $x \rightarrow 0$ , the function  $\varepsilon(x)$  associated to the loss of optimality of  $\phi^*(\mathbf{z})$  converges to

$$\varepsilon_\eta^* = \max \left\{ \frac{|\eta - q_{\eta,\max}|}{\eta}, \frac{|\eta - q_{\eta,\min}|}{\eta}, \frac{q_{\eta,\max} - q_{\eta,\min}}{\eta} \right\}. \quad (31)$$

Finally, as  $b_{\min} \rightarrow +\infty$ , this test satisfies

$$\alpha_{\max}^* \sim Q(c^*) \quad (32)$$

where  $c^*$  is given in (19).

*Proof:* See Appendix A. ■

Let us discuss this theorem. First, the problem (LP) has not necessarily an unique solution; hence,  $\mathbf{p}^*$  may be not unique. The vector  $\mathbf{p}^*$  is called an “optimal equalization weight vector” since it ensures that  $\phi^{(\mathbf{p}^*)}(\mathbf{z})$  is a constrained  $\varepsilon$ -equalizer test. Second, assuming that  $rQ(b_{\min}) < \alpha$  with a small  $\alpha$  is equivalent to assume that  $b_{\min}$  is large. In this case, from (32), it is clear that  $\alpha_{\max}^*$  depends only on  $c^*$ . Third, the loss of optimality  $\varepsilon_\eta^*$  in (31) depends on both the parameters  $\eta, q_{\eta,\max}$  and  $q_{\eta,\min}$ . The integers  $q_{\eta,\max}$  and  $q_{\eta,\min}$  are unknown *a priori*. They depend on the geometry of the separability map. Since  $1 \leq q_{\eta,i} \leq r$  for all  $i \in \mathcal{K}_n$ , it is clear that  $0 < \varepsilon_r^* < 1$ . Hence,  $\eta = r$  can be chosen as a default value. Besides, there exists at least one ideal parameter  $\eta^*$  which minimizes  $\varepsilon_\eta^*$  but it is impossible to guess it easily in the case of a complex separability map. In practice, a possible solution consists of testing several values of  $\eta$  and to choose the best one. As established in Appendix A–C, especially in (49), (50) and (52), the function  $\varepsilon(x)$  decreases exponentially to  $\varepsilon_\eta^*$ . Lastly, the false alarm probability  $\alpha_0^*(\varphi_0)$  is upper bounded by  $\alpha$  but the tightness of this bound depends on the geometry of the separability map. The following proposition shows that it is possible to have a value  $\alpha_0^*(\varphi_0)$  close to  $\alpha$  by keeping unchanged the optimal maximum error probability  $Q(c^*)$ .

*Proposition 3:* Consider the test  $\phi^*(\mathbf{z})$  given in Theorem 1. If there exists  $\ell$  vectors  $\varphi_i$  such that  $c_i = c^*$  and  $b_i = b_{\min}$ , then

$$\alpha_0^*(\varphi_0) \sim \alpha \frac{\ell}{r} \quad (33)$$

as  $b_{\min} \rightarrow +\infty$ . Otherwise, it is possible to find a special  $u_\alpha < Q^{-1}(\alpha/r)$  such that the test  $\phi^*(\mathbf{z})$  keeps the same maximum error probability  $\alpha_{\max}^* \sim Q(c^*)$  and satisfies (33) where  $1 \leq \ell \leq r$  is an integer (unknown *a priori*) depending on the mutual geometry of the vector constellation.

*Proof:* See Appendix 3. ■

The following corollary directly follows from Theorem 1 and Proposition 2.

*Corollary 1:* Under the assumptions A1) and A2), the test  $\phi^*(\mathbf{y}) \stackrel{\text{def.}}{=} \phi^*(\mathbf{W}^T \mathbf{y})$ , where  $\phi^*(\cdot)$  is defined in Theorem 1, is a constrained  $\varepsilon$ -minimax test in the class  $\mathcal{D}_{\alpha_0^*} \subset \mathcal{D}_\alpha$  between hypotheses  $\mathcal{H}_0, \mathcal{H}_1, \dots, \mathcal{H}_r$  where

$$\alpha_0^* \stackrel{\text{def.}}{=} \alpha_0^*(\varphi_0) = \sup_{\psi \in \Psi_0} \mathbb{E}_\psi[1 - \phi_0^*(\mathbf{W}^T \mathbf{y})].$$

According to Proposition 3, the class  $\mathcal{D}_{\alpha_0^*}$  can be very close to the class  $\mathcal{D}_\alpha$  by changing adequately the value of  $u_\alpha$ .

*Example 5 (Discrete Location Parameter—Continued):* The constrained equalizer test solving the simultaneous detection and classification problem of discrete location parameter is proposed in [36]. The optimal weight vector is computed by using an iterative algorithm. The approach proposed in this paper leads to the same solution. However, three main drawbacks of [36] are overcome. First, the proof of existence of the  $\varepsilon$ -equalizer test is established in this paper, not in [36] which considers that the convergence of the algorithm proves the existence of the equalizer test. Second, the numerical solution proposed in [36] involves nonlinear optimization to obtain the optimal weight vector. Third, the classification error probabilities are numerically evaluated in [36] but no closed-form is proposed. ◊

*Example 6 (Slippage Problem—Continued):* Let  $\mathbf{z} = \Sigma^{-\frac{1}{2}} \mathbf{y}$  where  $\Sigma^{-\frac{1}{2}}$  is defined in Section II. Then, the slippage problem with unstructured noise (see example 1) is equivalent to detecting and classifying a vector  $\varphi_j = \Delta \Sigma^{-\frac{1}{2}} \delta_j$  from  $\mathbf{z}$ . The solution of this problem is given by Theorem 1. It is now established that the performances of the test essentially depend on the Euclidean distances between the columns of  $\Sigma^{-\frac{1}{2}}$  or, equivalently, on the correlations between the components of the random vector  $\mathbf{y}$ . ◊

*Example 7 (Integrity Monitoring—Continued):* Example 2 can be solved by using Corollary 1. The performances of the test essentially depend on the structure of the columns of  $\mathbf{W}$ . In practice, the matrix  $\mathbf{W}$  is deduced from the satellite constellation geometry [5]. Hence, the proposed results could be used to evaluate the capacity of the navigation system to preserve its integrity with respect to its geographical position. ◊

## V. ASYMPTOTICALLY EQUIVALENT TEST

This section proposes some asymptotically equivalent tests to the constrained  $\varepsilon$ -minimax one given in Theorem 1. These tests have the same asymptotic maximum classification error but their form is simpler.

### A. Simplified Asymptotic Form of the Epsilon-Minimax Test

Let  $\varrho = \min\{\varrho_d, \varrho_i\}$  be the positive real where

$$\varrho_d = \min_{1 \leq i \leq r} \frac{1}{2} (b_i^2 + b_{\min}^2) - u_\alpha b_{\min} - \omega_r^* b_i$$

with  $u_\alpha = Q^{-1}(\alpha/r)$  and

$$\varrho_i = \min_{i \in \mathcal{K}_n} \min_{k \in \mathcal{K}_s \cup \mathcal{K}_w} \frac{1}{2} d_{i,k} (d_{i,k} - \omega_r^*).$$

Let  $\mathcal{K}_w$  be cut into the three disjoint subsets  $\mathcal{K}_w^{(1)}, \mathcal{K}_w^{(2)}$  and  $\mathcal{K}_w^{(3)} = \mathcal{K}_w \setminus (\mathcal{K}_w^{(1)} \cup \mathcal{K}_w^{(2)})$  such that

$$\begin{aligned}\mathcal{K}_w^{(1)} &= \{i \in \mathcal{K}_w : \varphi_0 = \varphi(\mathcal{C}(\varphi_i))\} \\ \mathcal{K}_w^{(2)} &= \{i \in \mathcal{K}_w : \varphi_0 \in \mathcal{C}(\varphi_i) \text{ and } \varphi_0 \neq \varphi(\mathcal{C}(\varphi_i))\}\end{aligned}$$

where  $\mathcal{C}(\varphi_i)$  denotes the component containing  $\varphi_i$ . The two subsets  $\mathcal{K}_w^{(1)}$  and  $\mathcal{K}_w^{(2)}$  are clearly incompatible: if  $\varphi_0$  is the center of a component then  $\mathcal{K}_w^{(2)}$  is necessarily empty and, reciprocally, if  $\varphi_0$  is not the center of a component then  $\mathcal{K}_w^{(1)}$  is necessarily empty. Let  $w_i^\infty$  be the family of reals

$$w_i^\infty = \begin{cases} 1, & \text{if } i \in \mathcal{K}_s \cup \mathcal{K}_w^{(1)} \\ 4\left(1 + \sqrt{4|\mathcal{C}(\varphi_i)| - 7}\right)^{-2}, & \text{if } i \in \mathcal{K}_w^{(2)} \\ (|\mathcal{C}(\varphi_i)| - 1)^{-1}, & \text{if } i \in \mathcal{K}_w^{(3)} \\ e^{-\varrho}, & \text{if } i \in \mathcal{K}_n. \end{cases}$$

Let  $\mathbf{p}^\infty \in \mathcal{S}_r$  be defined by  $p_i^\infty = p_{\max}^\infty w_i^\infty$  for all  $1 \leq i \leq r$  where

$$p_{\max}^\infty = 1 / \sum_{i=1}^r w_i^\infty$$

and let  $\lambda^\infty$  be the threshold

$$\lambda^\infty = u_\alpha b_{\min} - \frac{b_{\min}^2}{2} + \ln p_{\max}^\infty. \quad (34)$$

*Proposition 4:* Under the assumptions A1) and A2), the test  $\phi^\infty(\mathbf{z}) \stackrel{\text{def.}}{=} \phi^{(\mathbf{p}^\infty)}(\mathbf{z})$  with the threshold  $\lambda^\infty$  (34) satisfies  $\alpha_0^{\infty}(\varphi_0) \leq \alpha$  and is asymptotically equivalent to the constrained  $\varepsilon_r^*$ -minimax test  $\phi^*(\mathbf{z})$  between  $\mathcal{H}_0, \mathcal{H}_1, \dots, \mathcal{H}_r$  as  $b_{\min} \rightarrow +\infty$

$$\max_{1 \leq k \leq r} \alpha_k^\infty(\varphi_k) \sim \max_{1 \leq k \leq r} \alpha_k^*(\varphi_k).$$

*Proof:* See Appendix C. ■

According to Proposition 3, the exact comparison between  $\alpha_0^*(\varphi_0)$  and  $\alpha_0^\infty(\varphi_0)$  is not relevant. It must be noted that  $\phi^\infty(\mathbf{z})$  is not necessarily a constrained  $\varepsilon$ -equalizer test.

### B. Optimality of the MGLRT

Let  $\bar{\mathbf{p}}$  be the uniform weight vector such that  $\bar{p}_i = 1/r$  for all  $1 \leq i \leq r$ . The MGLRT is the Bayesian test  $\bar{\phi}(\mathbf{z}) \stackrel{\text{def.}}{=} \phi^{(\bar{\mathbf{p}})}(\mathbf{z})$  associated to the uniform weight vector  $\bar{\mathbf{p}}$ . Its statistical properties and its optimality are established, respectively, in Proposition 5 and in Corollary 2. Let

$$J = \{i \in \{1, \dots, r\} : b_i = b_{\min}\}.$$

Define  $\bar{q}_i$  as the number of vectors at distance  $d_i$  from  $\varphi_i$  and

$$\bar{q} = \max_{1 \leq i \leq r: d_i = d^*} \bar{q}_i.$$

*Proposition 5:* The test  $\bar{\phi}(\mathbf{z})$ , with the threshold

$$\bar{\lambda} = \lambda(\bar{\mathbf{p}}) = b_{\min} Q^{-1}(\alpha/|J|) - \frac{b_{\min}^2}{2} - \log r$$

satisfies  $\bar{\alpha}_0(\varphi_0) \sim \alpha$  and

$$\bar{\alpha}_{\max} = \max_{1 \leq k \leq r} \bar{\alpha}_k(\varphi_k) \sim \bar{q} Q(d^*) \text{ as } b_{\min} \rightarrow +\infty. \quad (35)$$

*Proof:* From (55) applied to the weight vector  $\bar{\mathbf{p}}$  with the threshold  $\bar{\lambda}$ , we get  $a_k > Q^{-1}(\alpha/|J|)$  for  $k \notin J$  and  $a_k = a = Q^{-1}(\alpha/|J|)$  for  $k \in J$ . Using the lower and upper bounds given in [25] for equally-likely vectors, we get  $\bar{\alpha}_0(\varphi_0) \sim |J| Q(a) \sim \alpha$  and

$$\bar{\alpha}_i(\varphi_i) \sim \bar{q}_i Q(d_i)$$

as  $b_{\min} \rightarrow +\infty$ . The end of the proof is straightforward. ■

Generally, the test  $\bar{\phi}(\mathbf{z})$  is not a constrained  $\varepsilon$ -minimax test because  $\bar{q} Q(d^*) > Q(c^*)$ . However, under the assumption

**A3)** All the components of the separability map have at most two elements, i.e.,  $|\mathcal{C}_i| \leq 2$  for all  $1 \leq i \leq s$ ,

Corollary 2 shows that this test is asymptotically equivalent to the constrained  $\varepsilon$ -minimax test.

*Corollary 2:* Under the assumptions A1), A2) and A3), the test  $\bar{\phi}(\mathbf{z})$  such that  $\bar{\alpha}_0(\varphi_0) \leq \alpha$  is asymptotically equivalent to the constrained  $\varepsilon$ -minimax test  $\phi^*(\mathbf{z})$  as  $b_{\min} \rightarrow +\infty$

$$\max_{1 \leq k \leq r} \bar{\alpha}_k(\varphi_k) \sim \max_{1 \leq k \leq r} \alpha_k^*(\varphi_k). \quad (36)$$

*Proof:* Under the assumptions A1) and A3), we get  $\bar{q} = 1$ . The rest of the proof is straightforward by considering together Proposition 5 and Theorem 1. ■

It is important to note that the assumption A3) is very severe in practice since it imposes very strict requirements on the mutual geometry between the hypotheses. If this assumption is not satisfied (an example is shown in Section VI), then the test  $\bar{\phi}(\mathbf{z})$  is clearly suboptimal.

## VI. APPLICATION TO NEW ACTIVITY DETECTION IN A MULTIUSER SYSTEM

This section is focused on example 3. The goal is to show that the proposed approach could be used to detect the entrance of a new user into a multiuser communication channel and to identify both the transmitted binary digit and the user's signature.

### A. Detection of a New User Entrance

In order to make the simulation free of secondary details, let us consider the plainest, yet general enough, model of Direct-Sequence/Code-Division Multiple-Access (DS/CDMA) involving real signatures and Binary Phase Shift Keying (BPSK) data transmission [29], [30]. The model covers, among others, any system with BPSK signature and data modulation. The simulation considers a fully synchronous case when both chips and borders of data symbols (bits) of all users are strictly

aligned in time. This, along with the assumption of the independence of consecutive data bits of any user, permits limiting the observation interval to a single bit duration  $T_b$ . The group signal of  $m$  users during the bit duration is

$$s(t; \mathbf{x}) = \sum_{k=1}^m a_k x_k s_k(t) \quad (37)$$

where  $a_k > 0$  is the real amplitude of the  $k$ th user signal,  $\mathbf{x} = (x_1, \dots, x_m)^T$  is the vector of data bits of  $m$  users and  $s_k(t)$  is the  $k$ th user's signature. The bit  $x_k$  takes antipodal values, 1 or  $-1$ . For simplicity, it is assumed that the values  $a_k$  are known. It is also assumed that there exists a set of  $m + \bar{r}$  nonorthogonal signal signatures  $s_k(t)$ . This numerical example is specially adapted to the case of an overloaded CDMA system for which  $m + \bar{r}$  is larger than  $n$ .

After chip-matched filtering and signal rate sampling, the received signal vector under hypothesis  $\mathcal{H}_0$  (no new user) is

$$\mathcal{H}_0 : \{ \mathbf{y} \sim \mathcal{N}(\mathbf{H}\mathbf{x}, \mathbf{I}_n), \mathbf{x} \in \mathbb{R}^m \}$$

where  $\mathbf{H} = \mathbf{S}\mathbf{A}$ ,  $\mathbf{A} = \text{diag}(a_1, \dots, a_m)$  is the diagonal matrix of user amplitudes and  $\mathbf{S}$  is the matrix whose  $j$ th column  $\mathbf{s}_j \in \mathbb{R}^n$  is the signature waveform vector obtained from  $s_j(t)$  during the bit duration. Here, to seek simplicity, it is assumed that  $\mathbf{y}$  is Gaussian distributed with an identity covariance matrix. Following [31], the fact that  $\mathbf{x} \in \{-1, 1\}^m$  is not explicitly taken into account. At a certain time, a new user enters in the system with the discrete signature  $\mathbf{s}_{m+j}$  where  $j = 1, \dots, \bar{r}$ . The user known amplitude is  $a_{m+1}$ . The unknown transmitted bit is  $x_{m+1} \in \{-1, 1\}$ . Hence, there are  $r = 2\bar{r}$  alternative hypotheses  $\mathcal{H}_j$  (entrance of a new user)

$$\mathcal{H}_j : \{ \mathbf{y} \sim \mathcal{N}(\mathbf{H}\mathbf{x} + \boldsymbol{\theta}_j, \mathbf{I}_n), \mathbf{x} \in \mathbb{R}^m \}$$

where  $\boldsymbol{\theta}_j = a_{m+1}\mathbf{s}_{m+j}$  and  $\boldsymbol{\theta}_{\bar{r}+j} = -\boldsymbol{\theta}_j$  for  $j = 1, \dots, \bar{r}$ .

Strictly speaking, the goal is to estimate both the entry time, the signature waveform vector and the transmitted bit of the new user. This is a change detection/classification problem between  $r$  hypotheses [31]. As it is stressed in [37], when the monitored (typically large-scale) system has a variable structure, this leads to an extremely complicated sequential strategy which must take into account the structural changes in the system. In fact, the theory of sequential decision is only well-developed in the case of stationary systems in the prechange state. In contrast to the sequential strategy, the repeated Fixed Size Sample (FSS) one is easily applicable to systems with a variable structure for quickest detection and classification of changes. As explained in [37], [38], the repeated FSS strategy is based on the following rule: samples are collected successively and, at the end of each sample, a decision function (the "snapshot test") is computed to test between the  $r$  hypotheses. The detailed relations between the error probabilities of the "snapshot test" and the mean detection delay, the mean time between false alarms and the probabilities of false classification can be found in [37], [38]. For all these reasons, this simulation is focused on the numerical study of the optimal constrained  $\varepsilon$ -minimax test which can be used as the "snapshot test" in the FSS strategy.

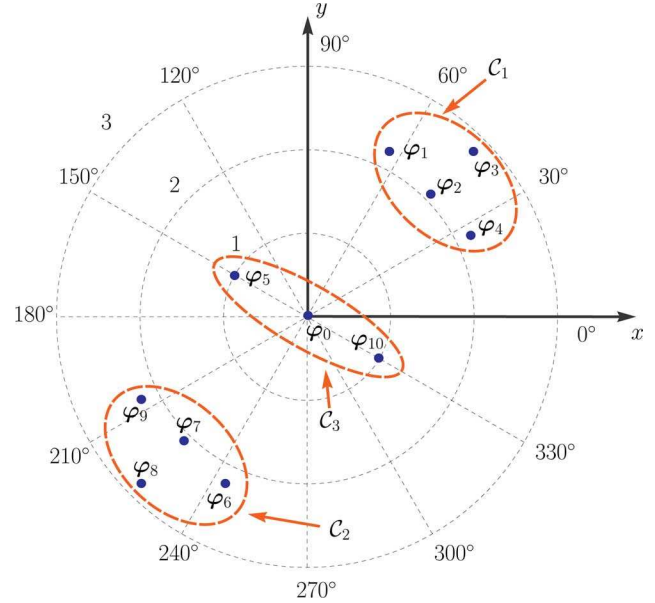


Fig. 3. Vector constellation with the three components  $\mathcal{C}_1, \mathcal{C}_2$  and  $\mathcal{C}_3$  of the associated separability map  $\mathcal{M}$ .

## B. Simulation Results

For simplicity, it is assumed that the signatures  $\mathbf{s}_1, \dots, \mathbf{s}_{m+\bar{r}}$  are chosen such that the rejection mechanism (see Subsection III.B) leads to the reduced model  $\mathbf{z} = \boldsymbol{\varphi} + \boldsymbol{\zeta}$  where  $\boldsymbol{\zeta} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_2)$  and  $\boldsymbol{\varphi}$  is either  $\boldsymbol{\varphi}_0 = (0; 0)^T$  or one of the  $r = 10$  following vectors:

$$\begin{aligned} \boldsymbol{\varphi}_1 &= \mu \begin{pmatrix} 1 \\ 2 \end{pmatrix}, & \boldsymbol{\varphi}_2 &= \mu \begin{pmatrix} 1.5 \\ 1.5 \end{pmatrix}, & \boldsymbol{\varphi}_3 &= \mu \begin{pmatrix} 2 \\ 2 \end{pmatrix} \\ \boldsymbol{\varphi}_4 &= \mu \begin{pmatrix} 2 \\ 1 \end{pmatrix}, & \boldsymbol{\varphi}_5 &= \mu \begin{pmatrix} -\sqrt{3}/2 \\ 0.5 \end{pmatrix}, & \boldsymbol{\varphi}_{5+j} &= -\boldsymbol{\varphi}_j \end{aligned}$$

for  $j = 1, \dots, 5$ . The vectors are plotted in Fig. 3 on a circular grid (see Remark 3). Clearly,  $b_1 = b_4 = \mu\sqrt{5}$ ,  $b_2 = \mu\sqrt{4.5}$ ,  $b_3 = \mu\sqrt{8}$ ,  $b_5 = b_{\min} = \mu$  and  $b_{5+j} = b_j$  for  $j = 1, \dots, 5$ . Here,  $\mu$  is a known scale factor which serves to adjust the Worst Signal-to-Noise Ratio (WSNR) defined by  $\text{WSNR} = 10 \log_{10}(\mu^2)$  in decibels (dB). Under the assumption that  $\alpha > rQ(\mu - \min_{1 \leq i \neq j \leq r} d_{i,j})$ , it is straightforward to verify that the separability map  $\mathcal{M}$  associated to this vector constellation is composed of three components  $\mathcal{C}_1, \mathcal{C}_2$  and  $\mathcal{C}_3$  (see Fig. 3). The first one is a star of order 4 containing  $\boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2, \boldsymbol{\varphi}_3$  and  $\boldsymbol{\varphi}_4$ . Its center is  $\boldsymbol{\varphi}_2$ . All vectors in  $\mathcal{C}_1$  have the same separability value  $d^* = 0.3536\mu$ . The common critical value of these vectors, which coincides with  $c^*$ , is a nonlinear function of  $\mu$ . The second component  $\mathcal{C}_2$  is a star of order 4 containing  $\boldsymbol{\varphi}_6, \boldsymbol{\varphi}_7, \boldsymbol{\varphi}_8$  and  $\boldsymbol{\varphi}_9$ . It has the same separability values and critical values than  $\mathcal{C}_1$ . The third component  $\mathcal{C}_3$  is a star of order 3 containing  $\boldsymbol{\varphi}_0, \boldsymbol{\varphi}_5$  and  $\boldsymbol{\varphi}_{10}$ . Its center is  $\boldsymbol{\varphi}_0$ . The separability value of  $\boldsymbol{\varphi}_5$  and  $\boldsymbol{\varphi}_{10}$  is  $d_5 = d_{10} = \mu - Q^{-1}(\alpha/r)$  and its critical value is  $c_5 = d_5 = c_{10} = d_{10}$ . The separability map  $\mathcal{M}$  leads to  $\mathcal{K}_s = \{2, 7\}$ ,  $\mathcal{K}_w = \{1, 3, 4, 6, 8, 9\}$  and  $\mathcal{K}_n = \{5, 10\}$ .

Numerical results, in term of classification error probabilities, are presented in Table I. Due to the symmetry in the vector constellation, only the estimates of probabilities  $\hat{\alpha}_i^*$  for  $i = 1, \dots, 5$  are presented. The estimates  $\hat{\alpha}_i^*$  for  $i = 6, \dots, 10$  have very

TABLE I  
ESTIMATION, CONFIDENCE INTERVAL AND BOUNDS OF THE  
CLASSIFICATION ERROR PROBABILITIES FOR  $\mu = 8$

Vector index	Numerical values ( $\times 10^{-3}$ )				
$i$	$\alpha_{i,\text{low}}^*$	$\hat{\alpha}_{i,-}^*$	$\hat{\alpha}_i^*$	$\hat{\alpha}_{i,+}^*$	$\alpha_{i,\text{upp}}^*$
1	3.994	3.985	<b>4.002</b>	4.018	4.025
2	3.983	3.975	<b>3.991</b>	4.008	3.994
3	3.994	4.001	<b>4.017</b>	4.034	4.057
4	3.994	3.978	<b>3.995</b>	4.011	4.025
5	3.994	3.979	<b>3.996</b>	4.012	3.994

similar values. The false alarm probability is  $\alpha = 0.01$  and  $\mu = 8$  (WSNR  $\approx 18$  dB). The problem (LP) is solved with  $\eta = 1$  and  $\omega^* = \omega_1^* = c^* = 2.653$ . The optimal solution satisfies  $q_{1,i}^* = 1$  for all  $i \in \mathcal{K}_n$ ; hence,  $\varepsilon_1^* = 0$ . The optimal weight vector  $\mathbf{p}^*$  associated to the constrained  $\varepsilon$ -equalizer test is given by  $p_1^* = p_3^* = p_4^* = 0.0877$ ,  $p_2^* = p_{\text{max}}^* = 0.237$  and  $p_5^* = 3.4 \times 10^{-9}$ . The other weights satisfy  $p_{5+j}^* = p_j^*$  for  $j = 1, \dots, 5$ . The threshold is  $\lambda^* = -8.717$ . The asymptotic maximum classification error probability is  $\alpha_{\text{max}}^* \sim Q(c^*) = 3.994 \times 10^{-3}$ . The classification error rate  $\hat{\alpha}_i^*$  is an estimation of  $\alpha_i^* = \alpha_i^*(\varphi_i)$  obtained from Monte-Carlo simulations with  $10^8$  random samples. This estimate belongs to the 99% confidence interval  $[\hat{\alpha}_{i,-}^*, \hat{\alpha}_{i,+}^*]$ . Moreover, as shown in Appendix A, especially in (41) and (44), each classification error probability  $\alpha_i^*$  is upper bounded by  $\alpha_{i,\text{upp}}^*$  and lower bounded by  $\alpha_{i,\text{low}}^*$ . These two bounds are very close to each other. Clearly, all the classification error probabilities are almost equal (see the column with bold numbers in Table I). Denoting  $\alpha_{\text{max,upp}}^* = \max_{1 \leq i \leq r} \alpha_{i,\text{upp}}^*$  and  $\alpha_{\text{min,low}}^* = \min_{1 \leq i \leq r} \alpha_{i,\text{low}}^*$ , a conservative upper bound  $\tilde{\varepsilon}$  of the loss of optimality  $\varepsilon(\alpha_{\text{max}}^*)$  is

$$\tilde{\varepsilon} = \frac{\alpha_{\text{max,upp}}^* - \alpha_{\text{min,low}}^*}{\alpha_{\text{min,low}}^*}. \quad (38)$$

Here,  $\tilde{\varepsilon} = 0.01$  shows that the loss of optimality is very small.

Fig. 4 compares the performances of the constrained  $\varepsilon$ -minimax test and the MGLRT for  $\mu$  varying from 3 to 10 with a sampling step of 0.25 (WSNR varying from 9.54 to 20 dB). For the step  $i \in \{1, \dots, 29\}$ , the prescribed probability of false alarm  $\alpha(i)$  is fixed at  $\alpha(i) = 0.2 \times 0.8^{(i-1)}$ . To ensure that the MGLRT satisfies the prescribed false alarm, the threshold  $\bar{\lambda}$  is fixed at  $\bar{\lambda} = \bar{\lambda}_{10} = \mu Q^{-1}(\alpha/10) - \frac{\mu^2}{2} - \log 10$  instead of using  $\bar{\lambda}_2 = \mu Q^{-1}(\alpha/2) - \frac{\mu^2}{2} - \log 10$ . This second threshold warrants that  $\bar{\alpha}_0(\varphi_0) \sim \alpha$  as  $\mu \rightarrow +\infty$  but it does not warrant that  $\bar{\alpha}_0(\varphi_0) \leq \alpha$  for a small value of  $\mu$ . In Fig. 4, the maximum classification error zone of the constrained  $\varepsilon$ -minimax test is the region defined by the two following curves. The first one, at the top of the region, is the upper bound  $\alpha_{i_{\text{max}}}^* = \alpha_{\text{max,upp}}^*$ , plotted as a function of  $\mu$ , where  $i_{\text{max}}$  is the index of the vector  $\varphi_{i_{\text{max}}}$  associated to the maximum upper bound  $\alpha_{\text{max,upp}}^*$ . The second one, at the bottom of the region, is the lower bound  $\alpha_{i_{\text{max,low}}}^*$ . These two curves rapidly converge toward the common function  $\delta^* = Q(c^*)$  as  $\mu$  grows. This zone contains the maximum classification error probability  $\alpha_{\text{max}}^*$ . The MGLRT maximum classification error zone is defined similarly. The  $\varepsilon$ -minimax test becomes rapidly more powerful than the MGLRT. The MGLRT

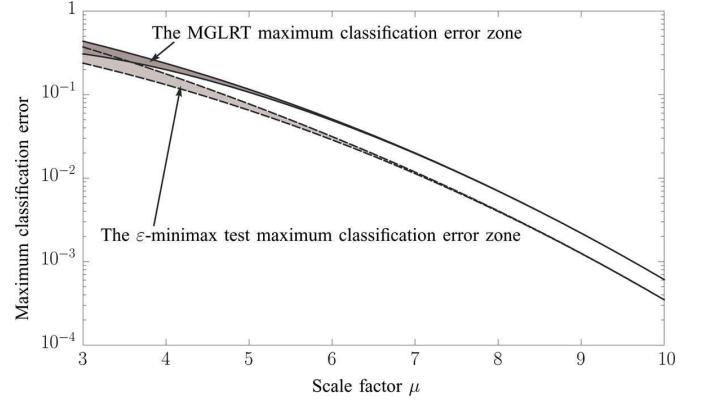


Fig. 4. Comparison of the maximum classification error probability, as a function of  $\mu$ , between the constrained  $\varepsilon$ -minimax test and the MGLRT.

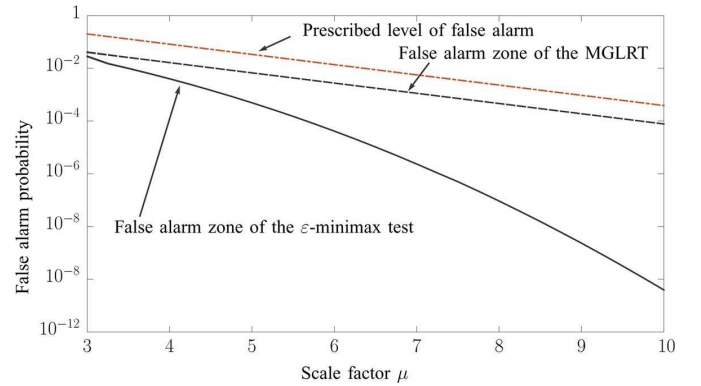


Fig. 5. Comparison of the false alarm probability, as a function of  $\mu$ , between the constrained  $\varepsilon$ -minimax test and the MGLRT.

is not optimal in this example because the vector constellation under consideration does not satisfy the restrictive assumption A3).

The false alarm probability of the tests is studied in Fig. 5. The false alarm zone of the constrained  $\varepsilon$ -minimax test is the region defined by the two following curves. The first one, at the top of the region, is the upper bound  $\alpha_{0,\text{upp}}^*$  given in (57). The second one, at the bottom of the region, is the lower bound  $\alpha_{0,\text{low}}^*$  given in (58). This zone contains the false alarm probability of the test. The false alarm zone of the MGLRT is defined similarly. Visually, the upper bound and the lower bound of each zone coincide (the gap between them is very small). The prescribed false alarm curve is plotted with a dashed-dotted line. As shown in Proposition 3, the false alarm of the constrained  $\varepsilon$ -minimax test tends to zero as  $\mu$  grows. The false alarm of the MGLRT is closer to the prescribed level but does not achieve the prescribed curve since we use the threshold  $\bar{\lambda}_{10}$ .

Fig. 6 shows the evolution of  $\tilde{\varepsilon}$  as a function of  $\mu$ . Clearly, the larger  $\mu$  (or equivalently the WSNR) is, the smaller the loss of optimality is. It must be noted that the loss of optimality is small even for low values of  $\mu$ . The exponential decay of the curve comes from the exponential nature of the  $Q$ -function (17) which appears in the classification error probability bounds (see Appendix A–B). It converges to the theoretical value  $\varepsilon_1^* = 0$ .

Finally, Fig. 7 compares the performance of the constrained  $\varepsilon$ -minimax test and the asymptotically equivalent one given in

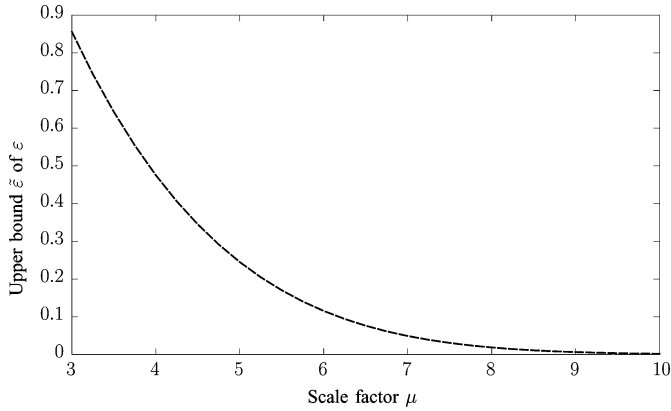


Fig. 6. Upper bound  $\bar{\epsilon}$  of the loss of optimality  $\epsilon$  as a function of  $\mu$ .

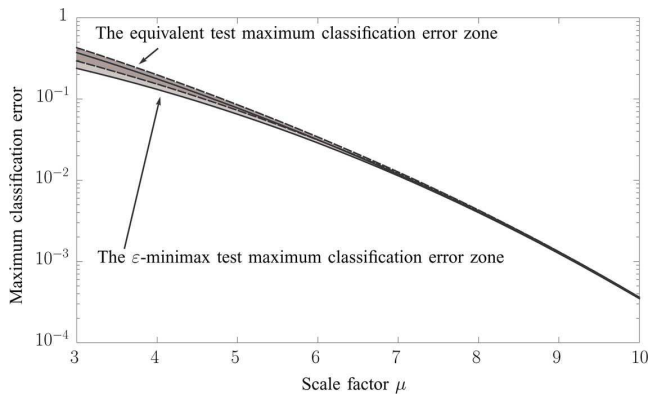


Fig. 7. Comparison of the maximum classification error probability, as a function of  $\mu$ , between the constrained  $\epsilon$ -minimax test and the asymptotically equivalent test.

Proposition 4 as a function of  $\mu$  for  $\alpha = 0.01$ . The weights of the equivalent test are  $p_1^\infty = p_3^\infty = p_4^\infty = 0.083, p_2^\infty = 0.25, p_5^\infty = 7.2 \times 10^{-13}$  and  $p_{5+j}^\infty = p_j^\infty$  for  $j = 1, \dots, 5$ . Its threshold is  $\lambda^\infty = \mu Q^{-1}(\alpha/10) - \frac{\mu}{2} + \ln 0.25$ . The maximum classification error zones of the tests are defined as in Fig. 4. They overlap each other. The performances of the tests are clearly comparable, even for moderate values of  $\mu$ .

*Remark 3:* Several approaches are possible to choose the signal signatures  $s_1(t), \dots, s_{m+\bar{r}}(t)$ , or equivalently the signature waveform vectors  $\mathbf{s}_1, \dots, \mathbf{s}_{m+\bar{r}}$ , in the case of an overloaded CDMA system. Some details and examples are given in [39]–[41]. The goal of this simulation is to illustrate the performances of the proposed  $\epsilon$ -minimax test, not to deal with a real practical situation. For this reason, the simulation is based on an arbitrary vector constellation. It must be noted that a similar constellation can be designed with realistic signature waveform vectors proposed in [39], [41]. If the resulting constellation is simpler, it is still interesting to use the  $\epsilon$ -minimax test but the loss of optimality of the MGLRT may be less significant.

## VII. CONCLUSION

This paper has proposed a constrained  $\epsilon$ -minimax test to detect and classify nonorthogonal vectors in presence of linear nuisance parameters. Proposition 2 shows that these nuisance

parameters can be rejected without any significant loss of optimality. Theorem 1 proposes a test which classifies the nonorthogonal vectors obtained after this rejection by equalizing the classification error probabilities up to a small constant under a constraint on the false alarm probability. The test design is based on some weighting coefficients which are computed by solving a linear programming problem deduced from the separability map. It is established that the critical value of the separability map defines the maximum classification error probability achievable by the constrained  $\epsilon$ -minimax test. When the SNR is large enough, Proposition 4 shows that the optimal weights can be directly calculated by avoiding the resolution of the linear programming problem. Finally, Proposition 5 proves that the proposed test clearly outperforms the famous MGLRT when the vector constellation is too complex.

## APPENDIX A PROOF OF THEOREM 1

The proof of Theorem 1 is composed of four parts. First, the optimal weight vector is calculated. Second, asymptotic tight upper and lower bounds of the classification error probabilities are established. Third, it is shown that the proposed optimal weight vector equalizes the classification error probabilities up to a small difference  $\epsilon$ . Finally, it is shown that the false alarm probability of the proposed test is upper bounded by  $\alpha$ .

### A. Optimal Weight Vector

Let  $\mathcal{K}_s, \mathcal{K}_w$  and  $\mathcal{K}_n$  be the sets defined by (20), (21) and (22). Let us consider the problem (LP) defined by (24)–(26). It is straightforward to verify that

$$\frac{b_k^2 + b_{\min}^2}{2b_k} - u_\alpha \frac{b_{\min}}{b_k} > b_{\min} - u_\alpha \quad (39)$$

for all  $1 \leq k \leq r$ . Hence, noting that  $\omega_\eta^* = c^* + (\ln \eta)/c^*$ , the constraints (28), (27) and (26) have positive right-hand terms provided that  $b_{\min}$  is large enough to ensure that  $(\ln \eta)/c^*$  becomes very small. It follows that  $z_1 = 0, \dots, z_r = 0$  is a feasible solution of (LP). Moreover, the  $z_k$ 's are necessarily bounded since they must satisfy the constraints (25) and (28). Hence, the feasible region of problem (LP) is nonempty and bounded, which implies that the linear programming problem (LP) has an optimal solution, namely  $z_k^*$  for all  $k \in \mathcal{K}_n$ . Each  $z_k^*$  necessarily activates at least one constraint in which it appears. Otherwise, it would be possible to increase the value of  $z_k^*$ , which would lead to a better solution since  $\sum_{k \in \mathcal{K}_n} z_k^*$  would increase. Consequently, it is clear that the coefficient  $z_k^*$ , associated to  $\varphi_k$ , is related to  $q_{\eta,k}$  other anomalies  $\varphi_i$  via the activated constraints where  $1 \leq q_{\eta,k} \leq r$  depends on the vector constellation complexity. Let  $\mathcal{A}_k^* \subset \{0, \dots, r\}$  be the set such that  $i \in \mathcal{A}_k^*$  if and only if  $\varphi_k$  is related to  $\varphi_i$  via an activated constraint (hence,  $|\mathcal{A}_k^*| = q_{\eta,k}$ ). Then, for all  $i \in \mathcal{A}_k^*, z_k^*$  has the form

$$z_k^* = \begin{cases} z_{k,i}^* = 2d_{k,i}(d_{k,i} - \omega_\eta^*), & \text{if } i \in \mathcal{K}_s \\ z_{k,i}^* = z_i + 2d_{k,i}(d_{k,i} - \omega_\eta^*), & \text{if } i \in \mathcal{K}_w \cup \mathcal{K}_n \\ z_{k,0}^* = \frac{1}{2}(b_k^2 + b_{\min}^2) - u_\alpha b_{\min} - \omega_\eta^* b_k, & \text{otherwise.} \end{cases}$$

Let  $\mathbf{p}^* \in \mathcal{S}_r$  be the optimal weight vector (23) associated to the  $z_k^*$ 's. Let  $\phi^*(\mathbf{z}) = \phi^{(\mathbf{p}^*)}(\mathbf{z})$  be the test function defined by (6), (7) and (14) with the threshold  $\lambda^*$  given in (30).

### B. Classification Error Probability

By definition, the  $k$ th classification error  $\alpha_k^* = \alpha_k^{(\mathbf{p}^*)}(\varphi_k)$  of  $\phi^*(\mathbf{z})$  is given by

$$\alpha_k^* = \Pr_{\varphi_k}(\cup_{i=0, i \neq k}^r \{g_k^*(\mathbf{z}) < g_i^*(\mathbf{z})\}) \quad (40)$$

where  $g_k^*(\mathbf{z}) = g_k^{(\mathbf{p}^*)}(\mathbf{z})$ ,  $g_0^*(\mathbf{z}) = \lambda^*$  and  $\Pr_{\varphi}(A)$  stands for the probability of event  $A$  when  $\mathbf{z}$  follows the distribution  $\mathcal{N}(\varphi, \mathbf{I}_q)$ . The most common upper bound of  $\alpha_k^*$  is the union bound  $\alpha_{k, \text{upp}}^*$  given by

$$\begin{aligned} \alpha_{k, \text{upp}}^* &= \sum_{i=0, i \neq k}^r \Pr_{\varphi_k}(\{g_k^*(\mathbf{z}) < g_i^*(\mathbf{z})\}) \\ &= \sum_{i=0, i \neq k}^r Q(\gamma_{k,i}^*) \end{aligned} \quad (41)$$

where  $Q(x)$  is defined in (17) and

$$\gamma_{k,i}^* = d_{k,i} + \frac{\ln p_k^* - \ln p_i^*}{2 d_{k,i}}, \quad 1 \leq i \neq k \leq r \quad (42)$$

$$\gamma_{k,0}^* = \frac{b_k}{2} - \frac{\lambda^* - \ln p_k^*}{b_k}. \quad (43)$$

This bound is asymptotically tight: it approaches the true error probability as  $b_{\min} \rightarrow +\infty$ .

Tight lower bounds are less studied. Here, the proposed lower bound is an extension of Swaszek's one (which is detailed in [25] in the case of equally-likely vectors). The probability  $\alpha_k^*$  is decomposed into a sum of two errors

$$\alpha_k^* = \alpha_{k, \text{low}}^* + \alpha_{k, \text{res}}^*$$

The first term  $\alpha_{k, \text{low}}^*$  corresponds to a subregion of the whole error region composed of several nonoverlapping cylinders with precisely known geometric descriptions. This probability can be easily calculated. The second term is the residual probability calculated on the rest of the whole error region. This second term is generally very difficult to calculate. Fortunately, it is negligible with respect to  $\alpha_{k, \text{low}}^*$  which constitutes the proposed lower bound. By definition, each nonoverlapping cylinder is associated to a single vector  $\varphi_i$ . This cylinder is the cross-product of the semi-infinite interval along the ray from  $\varphi_k$  trough  $\varphi_i$  with a  $(q-2)$ -sphere of radius  $w_k$  to be determined. Following [25, Section II], we obtain the following lower bound:

$$\alpha_{k, \text{low}}^* = \sum_{i=0, i \neq k}^r Q(\gamma_{k,i}^*) P\left(\frac{q-1}{2}, \frac{w_k^2}{2}\right) \quad (44)$$

where  $P(a, x)$  is the (normalized) incomplete gamma function

$$P(a, x) = \frac{1}{\Gamma(a)} \int_0^x e^{-t} t^{a-1} dt, \quad \forall a, x \geq 0.$$

The expression  $P((q-1)/2, w_k^2/2)$  is the probability that  $q-1$  Gaussian variables fall within a  $(q-2)$ -sphere of radius  $w_k$ .

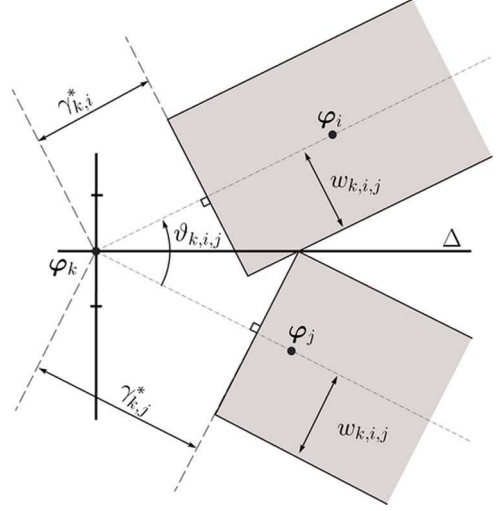


Fig. 8. Plane formed by  $\varphi_k$  and two of its neighbors  $\varphi_i$  and  $\varphi_j$ .

As explained below, this radius is chosen such that the cylinders have no overlaps. Fig. 8 depicts the plane formed by three different vectors  $\varphi_k, \varphi_i$  and  $\varphi_j$ . For simplicity, and due to the spherical symmetry of the Gaussian pdf,  $\varphi_k$  has been placed at the origin. The shaded regions are the intersections of the two cylinders and the plane formed by the three vectors. This figure shows the angle  $\vartheta_{k,i,j}$  between the two vectors  $\varphi_i$  and  $\varphi_j$  with respect to  $\varphi_k$  and the associated maximum admissible radius  $w_{k,i,j}$  ensuring no overlaps. The axis  $\Delta$  is the bisector of the angle  $\vartheta_{k,i,j}$ . Clearly, we must have

$$w_k \leq \min_{i \neq j, i \neq k, j \neq k} w_{k,i,j}. \quad (45)$$

A bit of calculus immediately shows that

$$\min\{\gamma_{k,i}^*, \gamma_{k,j}^*\} \tan(\vartheta_{k,i,j}/2) \leq w_{k,i,j}$$

for all  $i \neq j \neq k$ . Let  $w_k \stackrel{\text{def.}}{=} \gamma_k^* \tan(\vartheta_k/2)$  where  $\gamma_k^* = \min_{0 \leq i \neq k \leq r} \gamma_{k,i}^*$  and  $\vartheta_k = \min_{i \neq j \neq k} \vartheta_{k,i,j}$ . It follows that such a choice of  $w_k$  satisfies (45), which confirms the relevance of the lower bound (44).

### C. Epsilon-Equalizer Test

To calculate the epsilon-approximation of the probabilities  $\alpha_k^*$  for  $k \neq 0$ , we must treat separately the three following cases:  $k \in \mathcal{K}_s, k \in \mathcal{K}_w$  and  $k \in \mathcal{K}_n$ . From (42) and (43), it is easy to verify that

$$\gamma_{k,i}^* = d_{k,i} + \frac{z_i^* - z_k^*}{2 d_{k,i}}, \quad 1 \leq i \neq k \leq r, \quad (46)$$

$$\gamma_{k,0}^* = \frac{b_k^2 + b_{\min}^2}{2 b_k} - u_\alpha \frac{b_{\min}}{b_k} - \frac{z_k^*}{b_k}. \quad (47)$$

1) *Case of  $k \in \mathcal{K}_s$* : Let  $\mathcal{C}(\varphi_k)$  be the component whose  $\varphi_k$  is the star's center. Since  $\mathcal{C}(\varphi_k)$  is a critical component, it has at least two elements according to Lemma 2. Let  $\varphi_i$  be an other element of  $\mathcal{C}(\varphi_k)$ . If  $i > 0$ , we directly get  $\gamma_{k,i}^* = d_k + e_k$ . If

$i = 0$ , it means that  $d_k = d_{k,0} = d_0$ , which leads to  $b_k = b_{\min}$ . From (47), we obtain

$$\gamma_{k,0}^* = b_{\min} - u_\alpha = d_k.$$

If  $\varphi_i \notin \mathcal{C}(\varphi_k)$ , we get  $d_{k,i} > d_k$ . Hence,  $\gamma_{k,i}^* > d_k$  since  $z_i^* \geq 0$  and  $z_k^* = 0$ . From the definition of  $e_k$  and the fact that

$$Q(x) \sim \frac{1}{x\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \quad (48)$$

as  $x \rightarrow +\infty$ , (41) immediately yields to

$$\begin{aligned} \alpha_{k,\text{upp}}^* &= Q(d_k - e_k)(1 + \epsilon_{k,1}(b_{\min})) \\ &= Q(c^*)(1 + \epsilon_{k,1}(b_{\min})) \end{aligned}$$

where  $\epsilon_{k,1}(b_{\min}) \rightarrow 0$  as  $b_{\min} \rightarrow +\infty$ . From (44), we immediately get

$$\begin{aligned} \alpha_{k,\text{low}}^* &= Q(d_k - e_k)(1 - \epsilon_{k,2}(b_{\min})) \\ &= Q(c^*)(1 - \epsilon_{k,2}(b_{\min})) \end{aligned}$$

where  $\epsilon_{k,2}(b_{\min}) \rightarrow 0$  as  $b_{\min} \rightarrow +\infty$ . Finally, we obtain

$$|\alpha_k^* - \delta^*| \leq \delta^* \epsilon_k(b_{\min}) \quad (49)$$

where  $\delta^* = Q(c^*)$  and

$$\epsilon_k(b_{\min}) = \max\{|\epsilon_{k,1}(b_{\min})|, |\epsilon_{k,2}(b_{\min})|\} \rightarrow 0$$

as  $b_{\min} \rightarrow +\infty$ .

2) *Case of  $k \in \mathcal{K}_w$* : Let  $\varphi_j$  be the star's center of  $\mathcal{C}(\varphi_k)$ . If  $j > 0$ , we clearly get  $\gamma_{k,j}^* = d_k - e_k = c^*$ . If  $j = 0$ , we get  $\gamma_{k,j}^* = d_k = c^*$  since  $e_k = 0$  when  $\varphi_0 = \varphi(\mathcal{C}(\varphi_k))$ .

Let  $\varphi_i$  be an other element of  $\mathcal{C}(\varphi_k)$  such that  $i \neq j$  and  $i \neq k$ . If  $i > 0$ , we get  $\gamma_{k,i}^* = d_{k,i}$ . Since  $\mathcal{C}(\varphi_k)$  is a star graph, we necessarily have  $d_{k,i} > d_k \geq c^*$ . Hence,  $\gamma_{k,i}^* > c^*$ . If  $i = 0$ , we necessarily have  $b_k > b_{\min}$  since  $d_{k,0} > d_0 = b_{\min} - u_\alpha = d_j$ . Using (39), it follows from (47) that  $\gamma_{k,0}^* > c^*$  for  $b_{\min}$  sufficiently large since  $z_k^* = 2d_k e_k$  and  $e_k \sim \ln(1/2 + \sqrt{m-3/4})/d_k$  as  $b_{\min} \rightarrow +\infty$  (see Lemma 1).

Finally, let  $\varphi_i \notin \mathcal{C}(\varphi_k)$ . If  $i > 0$ , we get  $d_{k,i} > d_k$ . Since  $z_k^* = 2d_k e_k$  and  $z_i^* \geq 0$ , we get  $\gamma_{k,i}^* \geq d_{k,i} - e_k > d_k - e_k$ . If  $i = 0$ , using again (39), we get  $\gamma_{k,0}^* > c^*$  for  $b_{\min}$  sufficiently large since  $d_0 = b_{\min} - u_\alpha > d_k$ .

Hence, from (41) and (44), we get

$$|\alpha_k^* - \delta^*| \leq \delta^* \epsilon_k(b_{\min}) \quad (50)$$

with  $\epsilon_k(b_{\min}) \rightarrow 0$  as  $b_{\min} \rightarrow +\infty$ .

3) *Case of  $k \in \mathcal{K}_n$* : According to the definition of the constraints (26), (27) and (28) of the problem (LP), we get that  $\gamma_{k,i}^* > \omega_\eta^*$  for all  $i \notin \mathcal{A}_k^*$  and  $\gamma_{k,i}^* = \omega_\eta^*$  for all  $i \in \mathcal{A}_k^*$ . Hence, from (41) and (44), we get

$$\begin{aligned} q_{\eta,k} Q(\omega_\eta^*) (1 + \epsilon_{\eta,k}^{\text{inf}}(b_{\min})) &\leq \alpha_k^* \leq \\ q_{\eta,k} Q(\omega_\eta^*) (1 + \epsilon_{\eta,k}^{\text{sup}}(b_{\min})) &\quad (51) \end{aligned}$$

where  $\epsilon_{\eta,k}^{\text{inf}}$  and  $\epsilon_{\eta,k}^{\text{sup}}$  vanish as  $b_{\min} \rightarrow +\infty$ . Using (48), it must be noted that  $\eta Q(\omega_\eta^*) \sim Q(c^*) = \delta^*$ . Hence, using (51) to bound  $\alpha_k^* - \delta^*$ , we obtain that

$$|\alpha_k^* - \delta^*| \leq \delta^* (\zeta_{\eta,k} + \epsilon_k(b_{\min})) \quad (52)$$

with  $\zeta_{\eta,k} = |(\eta - q_{\eta,k})|/\eta$  and  $\epsilon_k(b_{\min}) \rightarrow 0$  as  $b_{\min} \rightarrow +\infty$ .

For an other  $j \in \mathcal{K}_n$ , (51) can be used to bound  $\alpha_j^*$ . Combining the two inequalities (the one for  $\alpha_k^*$  and the one for  $\alpha_j^*$ ) yields to

$$|\alpha_k^* - \alpha_j^*| \leq \delta^* (\zeta_{\eta,k,j} + \epsilon_{k,j}(b_{\min})) \quad (53)$$

with  $\zeta_{\eta,k,j} = |(q_{\eta,k} - q_{\eta,j})|/\eta$  and  $\epsilon_{k,j}(b_{\min}) \rightarrow 0$  as  $b_{\min} \rightarrow +\infty$ .

4) *Epsilon-Equalization*: Using the triangular inequality with (49), (50), (52) and (53), it follows that

$$\max_{1 \leq k \neq j \leq r} |\alpha_k^* - \alpha_j^*| \leq \delta^* \hat{\epsilon}(b_{\min}) \text{ as } b_{\min} \rightarrow +\infty$$

where  $\hat{\epsilon}(b_{\min}) \rightarrow \epsilon_\eta^*$  as  $b_{\min} \rightarrow +\infty$  with

$$\begin{aligned} \epsilon_\eta^* &= \max \left\{ \max_{k \in \mathcal{K}_n} \{\zeta_{\eta,k}\}, \max_{k \neq j \in \mathcal{K}_n} \{\zeta_{\eta,k,j}\} \right\} \\ &= \max \left\{ \frac{|\eta - q_{\eta,\text{max}}|}{\eta}, \frac{|\eta - q_{\eta,\text{min}}|}{\eta}, \frac{q_{\eta,\text{max}} - q_{\eta,\text{min}}}{\eta} \right\}. \end{aligned}$$

Furthermore, from the above equations, the maximum classification error probability  $\alpha_{\text{max}}^*$  is achieved by  $\alpha_j^*$  for at least one index  $j \in \mathcal{K}_s \cup \mathcal{K}_w$ . Hence

$$\max_{1 \leq k \leq r} |\alpha_k^* - \alpha_{\text{max}}^*| \leq \delta^* \hat{\epsilon}(b_{\min}). \quad (54)$$

Besides, from (49) and (50), we get

$$\delta^* (1 - \max_{k \in \mathcal{K}_s \cup \mathcal{K}_w} \epsilon_k(b_{\min})) \leq \alpha_{\text{max}}^*$$

which leads to

$$\delta^* \leq \alpha_{\text{max}}^* (1 + \bar{\epsilon}(b_{\min}))$$

where  $\bar{\epsilon}(b_{\min}) \rightarrow 0$  as  $b_{\min} \rightarrow +\infty$ . Hence, from (54), we get

$$\max_{1 \leq k \leq r} |\alpha_k^* - \alpha_{\text{max}}^*| \leq \alpha_{\text{max}}^* \epsilon(\alpha_{\text{max}}^*) \text{ as } \alpha_{\text{max}}^* \rightarrow 0$$

where  $\epsilon(\alpha_{\text{max}}^*) \rightarrow \epsilon_\eta^*$ . Replacing the condition  $b_{\min} \rightarrow +\infty$  by the equivalent condition  $\alpha_{\text{max}}^* \rightarrow 0$  is possible because  $\alpha_{\text{max}}^* \rightarrow 0$  if and only if  $b_{\min} \rightarrow +\infty$ .

#### D. False Alarm Probability

By definition, the false alarm probability  $\alpha_0^* = \alpha_0^{(\mathbf{p}^*)}(\varphi_0)$  is given by

$$\begin{aligned} \alpha_0^* &= \text{Pr}_0 \left( \max_{1 \leq k \leq r} \{\mathbf{y}^T \varphi_k - \frac{b_k^2}{2} + \log p_k^* - \lambda^*\} \geq 0 \right) \\ &= \text{Pr}_0 \left( \max_{1 \leq k \leq r} \{\mathbf{y}^T \varphi_k / b_k - a_k^*\} \geq 0 \right) \quad (55) \end{aligned}$$



where

$$\begin{aligned} a_k^* &= \left( \frac{b_k^2}{2} - \log p_k^* + \lambda^* \right) / b_k \\ &= \frac{b_k^2 - b_{\min}^2}{2b_k} + u_\alpha \frac{b_{\min}}{b_k} + \frac{z_k^*}{b_k}. \end{aligned} \quad (56)$$

To find  $\alpha_0^*$ , we need the distribution of the right hand term of (55). This involves to take into account the statistical correlations between the variables  $\mathbf{y}^T \boldsymbol{\varphi}_k$ , which is analytically intractable. Upper and lower limits for the false alarm probability can be obtained by using the methodology given in the previous subsection. It follows that

$$\begin{aligned} \alpha_0^* &\leq \sum_{k=1}^r \Pr_0(\mathbf{y}^T \boldsymbol{\theta}_k / b_k \geq a_k^*) = \sum_{k=1}^r Q(a_k^*) = \alpha_{0,\text{upper}}^* \\ &\quad \text{and} \\ \alpha_0^* &\geq q_0 Q(a_{\min}^*) P\left(\frac{n-1}{2}, \frac{w_0^2}{2}\right) = \alpha_{0,\text{lower}}^* \end{aligned} \quad (57)$$

where  $a_{\min}^* = \min_{1 \leq k \leq r} a_k^*$ ,  $q_0$  is the number of  $a_k^*$  such that  $a_k^* = a_{\min}^*$  and  $w_0$  is a known positive constant satisfying  $w_0 \rightarrow +\infty$  as  $b_{\min} \rightarrow +\infty$ .

Since  $Q(b_{\min}) < \alpha/r$ , it follows that  $u_\alpha \leq b_{\min} \leq \frac{1}{2}(b_k + b_{\min})$  for all  $k$  and, finally,  $\frac{b_{\min} - b_k}{b_k} u_\alpha \geq \frac{b_{\min}^2 - b_k^2}{2b_k}$ . Combining with (56), this leads to  $a_k^* \geq u_\alpha$ . It follows that  $\alpha_0^* \leq \alpha$ , which concludes the proof.

#### APPENDIX B PROOF OF PROPOSITION 3

Suppose there exist  $\ell \geq 1$  vectors  $\boldsymbol{\varphi}_{k_1}, \dots, \boldsymbol{\varphi}_{k_\ell}$  such that  $c_{k_i} = c^*$  and  $b_{k_i} = b_{\min}$ . It follows that  $z_{k_i}^* = 0$  or  $z_{k_i}^* = 2d_{k_i} e_{k_i}$  since  $k_i \in \mathcal{K}_S \cup \mathcal{K}_W$ . The condition  $b_{k_i} = b_{\min}$  implies that  $a_{k_i}^* = u_\alpha + z_{k_i}^*/b_{k_i}$  according to (56). Using Lemma 1, we get

$$a_{k_i}^* \sim u_\alpha.$$

From Subsections A–D, it is clear that  $a_j^* \rightarrow +\infty$  as  $b_{\min} \rightarrow +\infty$  for all other vectors  $\boldsymbol{\varphi}_j \neq \boldsymbol{\varphi}_{k_i}$ . From (57) and (58), it follows that

$$\alpha_0^*(\boldsymbol{\varphi}_0) \sim \ell Q(u_\alpha) = \alpha \frac{\ell}{r}.$$

Otherwise, suppose there is no vector  $\boldsymbol{\varphi}_i$  such that  $c_i = c^*$  and  $b_i = b_{\min}$ . According to (56), for all  $1 \leq i \leq r$ , we get  $a_i^* \rightarrow +\infty$  as  $b_{\min} \rightarrow +\infty$ , which means that  $\alpha_0^*(\boldsymbol{\varphi}_0) \rightarrow 0$ . Besides, the critical components do not contain  $\boldsymbol{\varphi}_0$ . In fact, if a critical component contains  $\boldsymbol{\varphi}_0$ , there exists necessarily a vector  $\boldsymbol{\varphi}_j$  in this component (satisfying  $c_j = c^*$ ) such that  $d_j = d_0 = d_{j,0}$ . This implies that  $b_j = b_{\min}$ , which is impossible. Hence, the critical value  $c^*$  of the separability map does not depend on  $u_\alpha$ . Consider an other real  $u'_\alpha$  such that  $u'_\alpha < u_\alpha$ . Clearly,  $d_0$  increases but not the other  $d_i$ 's. Hence,  $c^*$  remains independent from  $u'_\alpha$ . Replacing  $u_\alpha$  by  $u'_\alpha$  in Theorem 1 gives a test  $\boldsymbol{\phi}^*(\mathbf{z})$  with the same maximum classification error  $\alpha_{\max}^*$  but with a larger threshold. In fact, the right hand term in the constraint (28)

has increased (the other constraints (26) and (27) do not depend on  $u'_\alpha$ ). It means that the  $z_i^*$ 's for  $i \in \mathcal{K}_n$  remain constants or become larger, which implies that  $p_{\max}^*$  increases and, also,  $\lambda^*$ . Hence, we can find a value of  $u'_\alpha < u_\alpha$  (depending on  $b_{\min}$ ) such that at least one coefficient  $a_i^*$  (see (56)) is equal to  $Q^{-1}(\alpha/r)$ . In this case, we get

$$\alpha_0^*(\boldsymbol{\varphi}_0) \sim \alpha \frac{\ell}{r}$$

where  $\ell$  is the number (unknown *a priori*) of coefficients  $a_i^*$  achieving simultaneously the value  $Q^{-1}(\alpha/r)$ .

#### APPENDIX C PROOF OF PROPOSITION 4

Using the definition of  $p_{\max}^*$  in (29) and the asymptotic expression  $e_i^\infty$  of  $e_i$  given in Lemma 1, it is straightforward to verify that  $p_{\max}^* \sim p_{\max}^\infty$  as  $b_{\min} \rightarrow +\infty$ . Hence,  $\lambda^\infty \sim \lambda^*$ .

It is clear that  $z_k^\infty = \varrho$  for all  $k \in \mathcal{K}_n$  is a solution of the linear program (LP). Hence, from (42), (43) and the definition of constraints (26), (27) and (28), we necessarily get  $\gamma_{k,i}^\infty > c^*$  for all  $i \neq k$  since  $\eta = r$ . From the results given in Appendix A–C, we immediately get

$$\alpha_k^\infty(\boldsymbol{\varphi}_k) \leq Q(c^*)(1 - \epsilon_k(b_{\min}))$$

for all  $k \in \mathcal{K}_n$  where  $\epsilon_k(b_{\min}) \geq 0$  and  $\epsilon_k(b_{\min}) \rightarrow 0$  as  $b_{\min} \rightarrow +\infty$ . Using the upper bound (41) and the lower bound (44) with the weights  $p_i^\infty$  leads to

$$\alpha_i^\infty(\boldsymbol{\varphi}_i) \sim Q(c^*)$$

for all  $i \in \mathcal{K}_S \cup \mathcal{K}_W$  since the asymptotic expression of the  $e_i$ 's is used in the definition of the  $p_i^\infty$ 's. From the above results, we get

$$\alpha_{\max}^\infty \sim Q(c^*). \quad (59)$$

Finally, from Appendix A–D and the definition of  $\lambda^\infty$  in (34), it follows that

$$\alpha_0^\infty(\boldsymbol{\varphi}_0) \leq \alpha. \quad (60)$$

From (59) and (60), it follows that  $\boldsymbol{\phi}^\infty(\mathbf{z})$  is asymptotically equivalent to the constrained  $\varepsilon$ -minimax test  $\boldsymbol{\phi}^*(\mathbf{z})$ .

#### ACKNOWLEDGMENT

The author would like to thank Prof. Igor Nikiforov from the Troyes University of Technology (UTT), France, for his very helpful comments. He would also like to thank M. P. Z. for his support to write this paper. Finally, the author is grateful to the anonymous reviewers for their suggestions to improve the quality of the paper.

#### REFERENCES

- [1] H. L. V. Trees, *Detection, Estimation, and Modulation Theory: Radar-Sonar Signal Processing and Gaussian Signals in Noise*. Melbourne, FL: Krieger, 1992.

- [2] A. Frakt, W. Karl, and A. Willksy, "A multiscale hypothesis testing approach to anomaly detection and localization from noisy tomographic data," *IEEE Trans. Image Process.*, vol. 7, no. 6, pp. 825–837, Jun. 1998.
- [3] N. Merhav and Y. Ephraim, "A Bayesian classification approach with application to speech recognition," *IEEE Trans. Signal Process.*, vol. 39, no. 10, pp. 2157–2166, Oct. 1991.
- [4] A. Abramson and I. Cohen, "Simultaneous detection and estimation approach for speech enhancement," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 8, pp. 2348–2359, Aug. 2007.
- [5] M. Fouladirad and I. Nikiforov, "Optimal statistical fault detection with nuisance parameters," *Automatica*, vol. 41, no. 7, pp. 1157–1171, 2005.
- [6] L. Fillatre and I. Nikiforov, "Non-Bayesian detection and detectability of anomalies from a few noisy tomographic projections," *IEEE Trans. Signal Process.*, vol. 55, no. 2, pp. 401–413, Feb. 2007.
- [7] A. Lakhina, M. Crovella, and C. Diot, "Diagnosing network-wide traffic anomalies," presented at the ACM SIGCOMM, 2004.
- [8] J. G. Proakis, *Digital Communications*. New York: McGraw-Hill, 1983.
- [9] T. Ferguson, *Mathematical Statistics: A Decision Theoretic Approach*. New York: Academic, 1967.
- [10] E. Lehman, *Testing Statistical Hypotheses*, 2nd ed. New York: Chapman & Hall, 1986.
- [11] D. Middleton and R. Esposito, "Simultaneous optimum detection and estimation of signals in noise," *IEEE Trans. Inf. Theory*, vol. 14, no. 3, pp. 434–444, Mar. 1968.
- [12] D. Middleton and R. Esposito, "New results in the theory of simultaneous optimum detection and estimation of signals in noise," *Probl. Peredachi Inf.*, vol. 6, no. 2, pp. 3–20, 1970.
- [13] A. Fredriksen, D. Middleton, and D. Vandelinde, "Simultaneous signal detection and estimation under multiple hypotheses," *IEEE Trans. Inf. Theory*, vol. IT-18, no. 5, pp. 760–768, Sep. 1972.
- [14] T. Birdsall and J. Gobien, "Sufficient statistics and reproducing densities in simultaneous sequential detection and estimation," *IEEE Trans. Inf. Theory*, vol. IT-19, no. 6, pp. 760–768, Nov. 1973.
- [15] E. J. Kelly, I. S. Reed, and W. Root, "The detection of radar echoes in noise I," *J. SIAM*, vol. 8, no. 2, pp. 309–341, 1960.
- [16] E. J. Kelly, I. S. Reed, and W. Root, "The detection of radar echoes in noise II," *J. SIAM*, vol. 8, no. 3, pp. 481–505, 1960.
- [17] J. Stuller, "Generalized likelihood signal resolution," *IEEE Trans. Inf. Theory*, vol. IT-21, no. 3, pp. 276–282, May 1975.
- [18] B. Baygün and A. O. Hero, "Optimal simultaneous detection and estimation under a false alarm constraint," *IEEE Trans. Inf. Theory*, vol. 41, no. 3, pp. 688–703, Mar. 1995.
- [19] C. Leang and D. H. Johnson, "On the asymptotics of M-hypothesis Bayesian detection," *IEEE Trans. Inf. Theory*, vol. 43, no. 1, pp. 280–282, Jan. 1997.
- [20] M. B. Westover, "Asymptotic geometry of multiple hypothesis testing," *IEEE Trans. Inf. Theory*, vol. 54, no. 7, pp. 3327–3329, Jul. 2008.
- [21] E. Paulson, "An optimum solution to the k-sample slippage problem for the normal distribution," *Ann. Math. Statist.*, vol. 23, no. 4, pp. 610–616, 1952.
- [22] L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*. New York: Addison-Wesley, 1990.
- [23] A. C. Rencher, *Methods of Multivariate Analysis*. Hoboken, NJ: Wiley, 1995.
- [24] B. Hughes, "On the error probability of signals in additive white Gaussian noise," *IEEE Trans. Inf. Theory*, vol. 37, no. 1, pp. 151–155, Jan. 1991.
- [25] P. F. Swaszek, "A lower bound on the error probability for signals in white Gaussian noise," *IEEE Trans. Inf. Theory*, vol. 41, no. 3, pp. 837–841, Mar. 1995.
- [26] G. E. Séguin, "A lower bound on the error probability for signals in white Gaussian noise," *IEEE Trans. Inf. Theory*, vol. 44, no. 7, pp. 3168–3175, Jul. 1998.
- [27] L. Birgé, "A new lower bound for multiple hypothesis testing," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1611–1615, Apr. 2005.
- [28] P. M. Frank, "Fault diagnosis in dynamic systems using analytic and knowledge-based redundancy—A survey and some new results," *Automatica*, vol. 26, no. 3, pp. 459–474, 1990.
- [29] S. Verdú, *Multuser Detection*. Cambridge, U.K.: Cambridge Univ. Press, 1998.
- [30] V. P. Ipatov, *Spread Spectrum and CDMA: Principles and Applications*. Chichester, U.K.: Wiley, 2005.
- [31] T. Oskiper and H. V. Poor, "Online activity detection in a multiuser environment using the matrix CUSUM algorithm," *IEEE Trans. Inf. Theory*, vol. 48, no. 2, pp. 477–493, Feb. 2002.
- [32] A. Kapur and M. K. Varanasi, "Multiuser detection for overloaded CDMA systems," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1728–1742, Jul. 2003.
- [33] A. A. Borovkov, *Mathematical Statistics*. Amsterdam, The Netherlands: Gordon and Breach, 1998.
- [34] J. A. Bondy and U. S. R. Murty, *Graph Theory With Applications*. London, U.K.: Macmillan, 1976.
- [35] R. Diestel, *Graph Theory*. New York: Springer-Verlag, 2005.
- [36] B. Baygün and A. O. Hero, "An iterative solution to the min-max simultaneous detection and estimation problem," in *Proc. IEEE Workshop on Statistical Signal and Array Processing*, 1996, pp. 8–11.
- [37] L. Fillatre and I. Nikiforov, "A fixed size sample strategy for the sequential detection and isolation of non-orthogonal alternatives," *Seq. Anal.*, vol. 29, pp. 176–192, 2010.
- [38] I. Nikiforov, "Two strategies in the problem of change detection and isolation," *IEEE Trans. Inf. Theory*, vol. 43, no. 2, pp. 770–776, Feb. 1997.
- [39] J. Ross and D. Taylor, "Multiuser signaling in the symbol-synchronous AWGN channel," *IEEE Trans. Inf. Theory*, vol. 41, no. 4, pp. 1174–1178, Apr. 1995.
- [40] H. Sari, F. Vanhaverbeke, and M. Moeneclaey, "Extending the capacity of multiple access channels," *IEEE Commun. Mag.*, vol. 38, no. 1, pp. 74–82, 2000.
- [41] P. Pad, F. Marvasti, K. Alishahi, and S. Akbari, "A class of errorless codes for overloaded synchronous wireless and optical CDMA systems," *IEEE Trans. Inf. Theory*, vol. 55, no. 6, pp. 2705–2715, Jun. 2009.



**Lionel Fillatre** received the M.Sc. degree in decision and information engineering and the Ph.D. degree in systems optimization from the Troyes University of Technology (UTT), France, in 2001 and 2004, respectively.

From 2005 to 2007, he worked at Télécom Bretagne, Brest, France. Since 2007, he is an Associate Professor at the Systems Modelling and Dependability Laboratory, UTT. His current research interests include statistical decision theory, signal and image processing, and network monitoring.