An Iterative Solution to the Min-Max Simultaneous Detection and Estimation Problem

Bülent Baygün
Schlumberger-Doll Research
Old Quarry Road
Ridgefield, CT 06812
baygun@ridgefield.sdr.slb.com

Alfred O. Hero
University of Michigan
Department of EECS
Ann Arbor, MI 48109
hero@eecs.umich.edu

Abstract

Min-max simultaneous signal detection and parameter estimation requires the solution to a nonlinear optimization problem. Under certain conditions, the solution can be obtained by equalizing the probabilities of correctly estimating the signal parameter over the parameter range. We present an iterative algorithm based on Newton's root finding method to solve the nonlinear min-max optimization problem through explicit use of the equalization criterion. The proposed iterative algorithm does not require prior proof of whether an equalizer rule exists: convergence of the algorithm implies existence. A theoretical study of algorithm convergence is followed by an example which has applications in simultaneous detection and power estimation of a signal.

1. Introduction

In practical applications, one frequently needs to design a signal detector or a signal parameter estimator without complete knowledge of the signal or noise model. Several approaches to detector and estimator design exist in the case of incompletely characterized models. Among these are invariance methods, Bayesian methods which use non-informative priors, and min-max methods. Min-max methods form an important solution category because they ensure optimal detector or estimator performance under worst case conditions. Furthermore, min-max solutions give rise to tight performance bounds which can be used to benchmark sub-optimal or ad hoc algorithms. Min-max methods have been applied to problems of adaptive array processing, harmonic retrieval, CFAR detection, and distributed detection.

Signal detection and signal parameter estimation are typically considered as separate problems. In other words, signal parameter estimation methods assume that there is no uncertainty about signal presence. However, there are many applications where signal parameter estimation has to be done under signal presence uncertainty, such as fault detection and estimation in dynamical system control and antenna array processing Such problems are refered to as *simultaneous detection and estimation* problems. A min-max solution to simultaneous detection and estimation was recently given in [2]. The problem considered in [2] is estimation of a discrete parameter under a false alarm constraint. The statistical decision procedure which solves the problem is called the *constrained min-max classifier*. The constrained min-max classifier is characterized by a set of optimal weights. In Bayesian terminology, the optimal weights represent a *least favorable distribution* on the unknown parameter values. Numerical solutions to min-max detection or estimation problems involve nonlinear optimization to obtain the least favorable distribution [3, 1]. On the other hand, under certain assumptions, it is possible to formulate a min-max solution by making explicit use of a simplifying sufficient condition for min-max optimality. In the case of the constrained min-max classifier, this sufficient condition is the equalization of the correct classification probabilities. The purpose of the present work is to present an iterative algorithm for efficiently computing the constrained min-max classifier through the equalization condition. An important attribute of the proposed iterative algorithm is that it does not require prior proof of existence of an equalizer rule. Convergence of the algorithm proves existence, i.e. if we observe convergence, then the associated solution is the constrained min-max classifier.

The correct classification probability of the constrained min-max classifier provides a tight lower bound on the correct classification probability of any similarly constrained detection and classification procedure. By using the proposed algorithm, we can compute both this lower bound and the classification performance of sub-optimal simultaneous detection and classification procedures. Comparison of the performance of sub-optimal procedures with the lower bound allows us to assess the performance loss incurred by employing a sub-optimal approach to simultaneous detection and classification.

2. Problem Formulation

Consider the indexed probability space (Ω, σ, P_μ), where μ is a parameter that lies in a finite discrete parameter space Ξ, σ is a sigma algebra over Ω and P_μ is a probability measure defined on σ. Let $\mathbf{X}$ be a random variable taking values in a sample space Ω. Assume that $\mathbf{X}$ has a probability density function $f_\mu(x)$ with respect to a given measure. We will illustrate our approach for the case of a location parameter, i.e. $f_\mu(x) = f(x-\mu)$ for some fixed probability density function f. Applications of the location parameter case include modeling of a signal of unknown amplitude μ in additive noise whose probability density function is given by f.

Define the hypotheses $H_0, H_1, \ldots, H_n$ by:

$$H_i : \quad \mathbf{X} \sim f_{\mu_i}(x) = f(x-\mu_i) \ , \quad i = 0, \ldots, n \tag{1}$$

Let $R_0, R_1, \ldots, R_n$ be the decision regions for hypotheses $H_0, H_1, \ldots, H_n$, respectively, i.e. the classifier declares $\mu = \mu_i$ if and only if $x \in R_i$, $i = 0, 1, \ldots, n$. The probability of a correct decision under hypothesis H_i, $i = 0, 1, \ldots, n$ is given by

$$\mathrm{P}_{\mu_i}(\text{decide } H_i) \quad = \quad \mathrm{P}_{\mu_i}(\mathbf{X} \in R_i) \tag{2}$$

We will be interested in choosing the decision regions $R_0, R_1, \ldots, R_n$ such that the worst case correct classification probability $\min_i \mathrm{P}_{\mu_i}(\text{decide } H_i)$ is maximized subject to a given upper bound $\alpha \in (0,1]$ on the false alarm probability $1 - \mathrm{P}_{\mu_0}(\text{decide } H_0)$. A decision rule which maximizes the worst case correct classification probability under a false alarm constraint is called a *constrained min-max classifier*. In [2] it was shown that the constrained min-max classifier is a weighted likelihood ratio test:

$$\max_{i>0} \left\{ c_i \frac{f_{\mu_i}(x)}{f_{\mu_0}(x)} \right\} \quad \underset{H_0}{\overset{H_{i_{max}}}{\gtrless}} \quad \gamma \ , \tag{3}$$

i.e. if the maximum weighted likelihood ratio exceeds the threshold γ, then decide $H_{i_{max}}$, where $i_{max} = \arg\max_{i>0} \{c_i f_{\mu_i}(x)/f_{\mu_0}(x)\}$; otherwise decide H_0. The weights $c_1, \ldots, c_N$ are computed as the solution to a nonlinear optimization problem:

$$\min_{c_1,\ldots,c_n} \sum_{i=1}^{n} c_i \mathrm{P}_{\mu_i}(\text{decide } H_i) \ . \tag{4}$$

The threshold γ is determined using the specified bound α. Solution of the nonlinear optimization problem (4) could be computationally expensive. We will outline an alternative solution scheme which characterizes the min-max optimal classifier by means of a sufficient condition.

Suppose that the parameterized density $f_\mu(x) = f(x-\mu)$ has infinite support ($f(x) > 0$ for all x) and has a monotone likelihood ratio. An important class of probability densities that satisfies the monotone likelihood property is the single parameter exponential family. Furthermore, a sufficient condition for $f(x-\mu)$ to have a monotone likelihood ratio is for the function $-\log f(x)$ to be convex in x [4, page 509]. The normal, the double exponential and the logistic distributions all satisfy the convexity condition. Under the monotone likelihood ratio assumption, it can be shown that the constrained min-max classifier (3) gives rise to the following decision regions $R_0, R_1, \ldots, R_n$:

$$\begin{aligned} R_0 &= (-\infty, x_0]; \\ R_i &= (x_{i-1}, x_i], \quad i = 1, \ldots, n-1; \\ R_n &= (x_{n-1}, \infty) \end{aligned} \tag{5}$$

The correct decision probabilities are given by:

$$\begin{aligned} \mathrm{P}_{\mu_0}(\mathbf{X} \in R_0) &= F(x_0 - \mu_0) \\ \mathrm{P}_{\mu_1}(\mathbf{X} \in R_1) &= F(x_1 - \mu_1) - F(x_0 - \mu_1) \\ &\vdots \\ \mathrm{P}_{\mu_n}(\mathbf{X} \in R_n) &= 1 - F(x_{n-1} - \mu_n) \end{aligned} \tag{6}$$

where F is the cumulative distribution function with density f. The acceptance region R_0 for the null hypothesis H_0 can be specified explicitly. For any given value of $\alpha \in (0,1]$, there exists a value of x_0 that satisfies the false alarm constraint: $x_0 = F^{-1}(1-\alpha) + m_0$. The remaining decision boundary values $x_1, \ldots, x_{n-1}$ will be computed by an iterative procedure.

A sufficient condition for min-max optimality is the equalization of the correct classification probabilities $\mathrm{P}_{\mu_i}(\text{decide } H_i)$ for $i = 1, \ldots, n$ [2, Corollary 2]. The equalization condition is represented by the set of equations

$$\mathrm{P}_{\mu_i}(\text{decide } H_i) = p, \quad i = 1, \ldots, n \tag{7}$$

where $p \in (0,1)$ is the unknown common value of the correct classification probabilities. Let $\underline{y} = [x_1, \ldots, x_{n-1}, p]^T$ ("T" denotes matrix transpose) and define the function $G(\underline{y})$ as follows.

$$G(\underline{y}) \stackrel{def}{=} \begin{bmatrix} F(x_1-\mu_1) - F(x_0-\mu_1) - p \\ F(x_2-\mu_2) - F(x_1-\mu_2) - p \\ \vdots \\ F(x_{n-1}-\mu_{n-1}) - F(x_{n-2}-\mu_{n-1}) - p \\ 1 - F(x_{n-1}-\mu_n) - p \end{bmatrix} \tag{8}$$

Then the set of equations (7) is equivalent to

$$G(\underline{y}) = [0, \ldots, 0]^T \tag{9}$$

We propose to solve (9) iteratively using Newton's root finding method. More specifically, we consider the sequence $\underline{y}(k)$ generated through the iterations

$$\underline{y}(k+1) = \underline{y}(k) - J^{-1}(\underline{y}(k))G(\underline{y}(k)) \ , \tag{10}$$

where $J(\underline{y})$ is the Jacobian of the function $G(\underline{y})$, i.e.

$$[J(\underline{y})]_{ij} \stackrel{def}{=} \frac{\partial [G(\underline{y})]_i}{\partial y_j} \ . \tag{11}$$

For $j = 1, \ldots, n-1$, $y_j = x_j$. Therefore, the elements in the first $n-1$ columns of $J(\underline{y})$ are found from (8) to be:

$$[J(\underline{y})]_{ij} = \begin{cases} f(x_j - \mu_j) & \text{, if } i = j \\ -f(x_{j+1} - \mu_j) & \text{, if } i = j+1 \\ 0 & \text{, otherwise} \end{cases} \tag{12}$$

Similarly from (8), since $y_n = p$, the last column of $J(\underline{y})$ is given by all -1's, i.e.

$$[J(\underline{y})]_{in} = -1 \ , \quad i = 1, \ldots, n \tag{13}$$

A few words about the convergence of the iterative algorithm (10) are in order. Assume that there exists a solution $\underline{y}^*$ to the equation (9). If

1. $J^{-1}(\underline{y}^*)$ exists (the Jacobian is invertible); and

2. $||J(\underline{y}^* + \underline{\delta y}) - J(\underline{y}^*)|| \leq \gamma ||\underline{\delta y}||$ for some $\gamma > 0$ and for all sufficiently small perturbations $\underline{\delta y}$ (J is Lipschitz continuous); and

3. $||J^{-1}(\underline{y}^*)|| \leq \beta$ for some $\beta > 0$ (the norm of the Jacobian inverse is bounded from above);

then the sequence $\underline{y}(k)$ generated through (10) is well-defined, converges to $\underline{y}^*$ and has a quadratic rate of convergence with coefficient $\gamma\beta$ [5, Theorem 5.2.1]. Next we provide a sketch of the proof that the three conditions are satisfied in the present problem.

Condition 1: Since $f(x) > 0$ for all x, the columns of J are linearly independent.

Condition 2: The non-zero elements of the difference δJ of two Jacobians evaluated at points $\underline{y} + \underline{\delta y}$ and $\underline{y}$, respectively, are of the form $\pm(f(x_i + \delta x_i - \mu_j) - f(x_i - \mu_j))$. But $f(x_i + \delta x_i - \mu_j) - f(x_i - \mu_j) = \int_{x_i}^{x_i+\delta x_i} f'(t-\mu_j)dt$. Assuming that the derivative f' of the probability density function f is bounded, i.e. $\sup_x |f(x)| \leq M$ for some $M > 0$, it follows that $|f(x_i + \delta x_i - \mu_j) - f(x_i - \mu_j)| \leq M|\delta x_i|$. It can then be shown that the Frobenius norm of δJ, denoted by $||\delta J||_F$ is bounded above by a multiple of the l_2 norm of the vector $\underline{\delta y}$. Since the l_2-induced norm of δJ is smaller than the Frobenius norm of δJ [5, Theorem 3.1.3], Lipschitz continuity is satisfied.

Condition 3: For arbitrary $\underline{z} = [z_1, \ldots, z_n]^T$, consider the linear equation

$$J(\underline{y}(k))\underline{y}(k+1) = \underline{z} \ . \tag{14}$$

For notational simplicity, we will write the Jacobian as J and suppress its dependence on $\underline{y}$. After Gaussian elimination, the equation (14) can be re-written in terms of an upper triangular matrix $\tilde{J}$: $\tilde{J}\underline{y}(k+1) = \underline{\tilde{z}}$. The matrices $\tilde{J}$ and J are related by a non-singular transformation T, i.e. $\tilde{J} = TJ$. It suffices to establish an upper bound on the Frobenius norm $||\tilde{J}^{-1}||_F$ of $\tilde{J}^{-1}$ because $||J^{-1}||_F$ and $||\tilde{J}^{-1}||_F$ are related by $||J^{-1}||_F \leq ||T||_F ||\tilde{J}^{-1}||_F$ and $||T||_F$ is bounded. Suppose that the last column of $\tilde{J}$ is the vector $[-a_1, \ldots, -a_n]^T$, i.e. $[\tilde{J}]_{in} = -a_i$, $i = 1, \ldots, n$. It can be shown that $a_1 = 1$ and $a_i = 1 + a_{i-1}\frac{f(x_{i-1}-\mu_i)}{f(x_{i-1}-\mu_{i-1})}$, $i = 2, \ldots, n$. The Frobenius norm of $\tilde{J}^{-1}$ can be expressed as: $||\tilde{J}^{-1}||_F = [\mathrm{tr}((\tilde{J}^{-1})^T \tilde{J}^{-1})]^{1/2}$, where "tr" denotes matrix trace. After some algebra, we obtain an upper bound:

$$\begin{aligned} ||\tilde{J}^{-1}||_F &= \left(\sum_{i=1}^{n-1} (1 + \frac{a_i^2}{a_n^2}) \frac{1}{f^2(x_i - \mu_i)} + \frac{1}{a_n^2} \right)^{1/2} \\ &\leq ((n-1)L + 1)^{1/2} \ , \end{aligned} \tag{15}$$

where $L = \max_i \{(a_i^2 + a_n^2)/f^2(x_i - \mu_i)\}$, $i = 1, \ldots, n-1$. In finite dimensional spaces all norms are equivalent, therefore there exists some $\beta > 0$ such that $||J|| \leq \beta$.

3. Applications on Simultaneous Detection and Classification in Gaussian Noise

We will illustrate the iterative algorithm (10) for the case of normal densities. Let $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{x^2}{2\sigma^2})$ and $\mu_i = i$ for $i = 0, 1, \ldots, n$. We consider three different simultaneous detection and estimation rules. One of the rules is the constrained min-max classifier described earlier, which maximizes the worst case classification performance under a given false alarm constraint. One can also perform simultaneous detection and estimation by combining a classifier with a separately designed detector. With this strategy, the data are not presented to the classifier unless the detector declares "signal present". In other words, the classifier is gated by the detector.

We consider two gated classifiers and compare their performance to the performance of the constrained min-max classifier. Both of the gated classifiers use a min-max optimal detector for detection, but they differ in the design of their classifier structures. One of them uses an unconstrained min-max classifier designed independently of any detection objective. An unconstrained min-max classifier maximizes the worst case correct classification probability as if signal presence is certain. This classifier is obtained by removing the false alarm constraint ($\alpha = 1$) in the constrained min-max classifier. The other gated classifier uses a conditionally min-max classifier designed with explicit knowledge of the detector decision regions. A conditionally min-max optimal classifier maximizes the worst case correct classification probability conditioned on the detector having declared signal present. The conditionally min-max classifier is obtained by replacing all the densities $f_{\mu_i}(x)$ under the alternative hypotheses $H_1, \ldots, H_n$ with the conditional densities $f_{\mu_i}(x|\mathbf{X} \notin R_0)$ in the analysis of Section 2. Since we are using the min-max detector, $R_0 = (-\infty, x_0]$ as before, and x_0 is specified by the false alarm probability α.

Figure 1 shows the variation of the worst case correct classification probability $\min_i \mathrm{P}_{\mu_i}(\text{decide } H_i)$ for the three simultaneous detection and estimation rules as a function of the false alarm probability α. In this example $\sigma = 0.6$, and there are five alternative hypotheses ($n = 5$). In general, the constrained min-max classifier (solid line) performs best, while the unconstrained min-max classifier gated by the min-max detector (dashed line) gives rise to the lowest performance. The conditionally min-max classifier gated by the min-max detector (dashdot line), although better than the unconstrained min-max classifier, still falls significantly short of the performance of the constrained min-max classifier for small α. On the other hand, as α increases all three curves come together as expected. This is because for high α, the three simultaneous detection and estimation rules degenerate to an unconstrained min-max classifier for the alternative hypotheses $H_1, \ldots, H_n$.

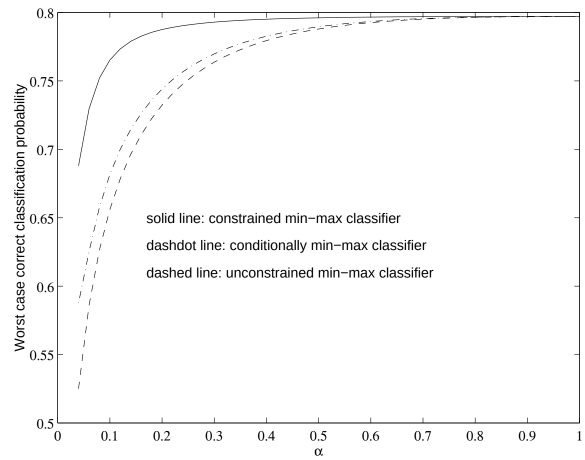

Figure 1. Worst case correct classification probability as a function of α.

References

[1] J. C. Preisig. A minmax approach to adaptive matched field processing in an uncertain propagation environment. *IEEE Trans. on Signal Proc.,* 42(6):1305–1316, June 1994.

[2] B. Baygün and A. O. Hero. Optimal simultaneous detection and estimation under a false alarm constraint. *IEEE Trans. Info. Theory,* 41(3):688-703, May 1995.

[3] C.-I. Chang and L. D. Davisson. Two iterative algorithms for finding minimax solutions. *IEEE Trans. Info. Theory,* 36(1):126–140, January 1990.

[4] E. L. Lehmann. *Testing Statistical Hypotheses.* 2nd ed., Wadsworth & Brooks/Cole, Pacific Grove, 1991.

[5] J. E. Dennis, Jr. and R. B. Schnabel. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations.* Prentice Hall, Englewood Cliffs, 1983.