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

Bülent Baygün and Alfred O. Hero

Summary

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 [1], harmonic retrieval [2], CFAR detection [3], and distributed detection [4].

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 [5]. The problem considered in [5] 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 [6, 1]. On the other hand, under certain assumptions, it is possible to formulate a min-max solution without recourse to the least favorable distribution 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.

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. For the purposes of the summary we limit the analysis to a univariate location parameter. Extensions for scale parameters as well as multivariate location parameters will be discussed in the full paper.

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 classification probability under a false alarm constraint is called a *constrained min-max classifier*. In [5] it was shown that the constrained min-max classifier is a weighted likelihood ratio test:

$$\max_{i>0} \left\{ c_i \frac{f_{\mu_i}(x)}{f_{\mu_0}(x)} \right\} \begin{array}{c} H_{i_{max}} \\ \gtrless \\ H_0 \end{array} \gamma \ , \tag{3}$$

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 [7, 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 $R_0, R_1, \ldots, R_n$:

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

The correct decision probabilities are given by:

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

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 with 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 $i = 1, \ldots, n$ [5, 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 $G(\underline{y})$.

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 $\gamma\beta$ [8, Theorem 5.2.1]. Next we provide a sketch of the proof that the three conditions are satisfied in the present problem. A rigorous proof will be given in the full paper.

For arbitrary $\underline{y}$, the Jacobian $J(\underline{y})$ has the form:

$$J(\underline{y}) = \begin{bmatrix} f(x_1-\mu_1) & 0 & 0 & \dots & 0 & -1 \\ -f(x_1-\mu_2) & f(x_2-\mu_2) & 0 & \dots & 0 & -1 \\ 0 & -f(x_2-\mu_3) & f(x_3-\mu_3) & \dots & 0 & -1 \\ & & \ddots & & 0 & -1 \\ & & \dots & -f(x_{n-2}-\mu_{n-1}) & f(x_{n-1}-\mu_{n-1}) & -1 \\ & & \dots & 0 & -f(x_{n-1}-\mu_n) & -1 \end{bmatrix} \tag{11}$$

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 [8, Theorem 3.1.3], Lipschitz continuity is satisfied.
Condition 3: For arbitrary $\underline{z} = [z_1, \dots, z_n]^T$, consider the linear equation

$$J(\underline{y}(k))\underline{y}(k+1) = \underline{z} \ . \tag{12}$$

For notational simplicity, we will write the Jacobian as J and suppress its dependence on $\underline{y}$. After Gaussian elimination, the equation (12) 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 $[-a_1, \dots, -a_n]^T$, i.e. $[\tilde{J}]_{in} = -a_i, \ i = 1, \dots, n$. It can be shown that $a_1 = 1$ and $a_i = 1 + a_{i-1}\frac{f(x_{i-1}-\mu_i)}{f(x_{i-1}-\mu_{i-1})}, \ i = 2, \dots, n$. 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:

$$\begin{aligned} ||\tilde{J}^{-1}||_F &= \left(\sum_{i=1}^{n-1}(1 + \frac{a_i^2}{a_n^2})\frac{1}{f^2(x_i - \mu_i)} + \frac{1}{a_n^2}\right)^{1/2} \\ &\leq ((n-1)L + 1)^{1/2} \ , \end{aligned} \tag{13}$$

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

Applications on detection and classification in Gaussian noise: We will illustrate the iterative algorithm (10) for the case of normal densities. Further results will be provided in the full paper.

Let $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp(-\frac{x^2}{2\sigma^2})$ and $\mu_i = i$ for $i = 0, 1, \dots, n$. First we consider the coupled detector/classifier scheme, which optimizes classification performance under a given false alarm constraint. Table 1 shows the values of the boundary points $x_0, \dots, x_{n-1}$ and the equalized correct classification probability p obtained by the iterative algorithm for $\sigma = 1$ and $n = 5$ and a false alarm probability of $\alpha = 0.01$. The maximum approximation error $\max_i |\mathrm{P}_{\mu_i}(x_{i-1}(k) < \mathbf{X} \leq x_i(k)) - p(k)| \ , \ i = 1, \dots, n$ falls below 10^{-3} within 9 iterations.

x_0	x_1	x_2	x_3	x_4	p
1.65	2.65	3.65	4.65	5.67	0.17

Table 1: The decision boundaries $x_0, \ldots, x_{n-1}$ and the equalized correct classification probability p for the constrained min-max classifier. ($\sigma = 1$ and $n = 5$ and a false alarm probability of $\alpha = 0.01$)

Results for the unconstrained version of this problem ($\alpha = 1$) are given in Table 2. The unconstrained decision rule would correspond to designing the classifier independently of any detection objective. If both detection and classification were desired, one could design an optimal detector independently and gate the classifier with the detector. With this strategy, the classifier does not make a decision unless the detector declares "signal present". If the same false alarm constraint is to be satisfied, the min-max optimal detector puts a detection threshold at $x_0 = 1.65$, the same value as in the coupled detector/classifier design. The classifier makes a decision only if $\mathbf{X} > x_0$. From Table 1, the right end point x_1 of the decision region R_1 for H_1 is $x_1 = 1.17$. Consequently, the decision region R_1 of the classifier falls entirely in the H_0 region of the detector: the combined detector/classifier can never decide H_1! Therefore, although the unconstrained classifier has a worst case correct classification probability of $p = 0.60$ when signal presence is certain, its worst case correct classification probability is 0 when detection is also required.

The above example clearly illustrates the pitfalls in designing the detector and estimator independently. On the other hand, even when the detector structure is fixed, classification performance can be improved by designing a conditionally optimal classifier which explicitly takes account of the detection step. A conditionally optimal classifier maximizes the worst case classification performance using the conditional densities under the alternative hypotheses $H_1, \ldots, H_n$ conditioned on the signal $\mathbf{X}$ exceeding the detection threshold x_0. This case can be handled by a slightly modified version of our iterative algorithm. A comparison of all three detection/classification approaches will be provided.

x_0	x_1	x_2	x_3	x_4	p
$-\infty$	1.17	2.41	3.59	4.83	0.60

Table 2: The decision boundaries $x_0, \ldots, x_{n-1}$ and the equalized correct classification probability p for the unconstrained min-max classifier. ($\sigma = 1$ and $n = 5$ and a false alarm probability of $\alpha = 0$)

References

[1] Preisig, J. C., "A minmax approach to adaptive matched field processing in an uncertain propagation environment," *IEEE Trans. on Signal Proc.,* vol. 42, no. 6, pp. 1305–1316, June 1994.

[2] Baygün, B. and Hero, A. O., "An order selection criterion via optimal joint estimation/detection theory," in *Proc. IEEE Workshop on Spectrum Estim. and Modeling,* Rochester, NY, 1990.

[3] Schwartz, R. E. "Minimax CFAR detection in additive Gaussian noise of unknown covariance," *IEEE Trans. Info. Theory,* vol. 15, no. 6, pp. 722–725, November 1969.

[4] Gerniotis, E. and Chau, Y. A., "Robust data fusion for multisensor detection systems," *IEEE Trans. on Info. Theory,* vol. 36, no. 6, pp. 1265–1279, November 1990.

[5] Baygün, B. and Hero, A. O., "Optimal simultaneous detection and estimation under a false alarm constraint," *IEEE Trans. Info. Theory,* vol. 41, no. 3, pp. 688-703, May 1995.

[6] Chang, C.-I., Davisson, L. D.,"Two iterative algorithms for finding minimax solutions," *IEEE Trans. Info. Theory,* vol. 36, no. 1, pp. 126–140, January 1990.

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

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

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

Bülent Baygün[1] and Alfred O. Hero[2]

[1] Schlumberger-Doll Research, Old Quarry Road, Ridgefield, CT 06812; ph: (203)-431-5524; e-mail: baygun@ridgefield.sdr.slb.com

[2] Department of EECS, University of Michigan, Ann Arbor, MI 48109; ph: (313)-763-0564; e-mail: hero@eecs.umich.edu

Abstract