AN ADAPTIVE ALGORITHM FOR DISTRIBUTED SIGNAL DETECTION SYSTEM DESIGN

Matthew C. Deans and Rick S. Blum

Electrical Engineering and Computer Science Department
Lehigh University
Bethlehem, PA 18015

ABSTRACT

Combining signal detection decisions from multiple sensors is useful in some practical communications, radar, and sonar applications. The optimum schemes for generating and combining the detector decisions have been studied for cases with independent observations from sensor to sensor. Designing schemes for cases with dependent observations from sensor to sensor is a much more difficult problem and to date very little progress has been made. Design approaches which have been suggested for these cases are quite complicated. Here a simple adaptive design approach is outlined for the important and difficult task of detecting a weak random signal in additive, possibly non-Gaussian noise. The approach is based on considering sensor decision rules and fusion rules which contain some unknown parameters. These rules have previously been shown to be optimum for cases with a large number of observations. These previous results also show that the best parameters minimize the mean square error fit to the best centralized signal detection scheme. Based on these ideas a gradient descent algorithm is proposed for learning the best parameters. Results of the training are compared to known results for multisensor detection schemes.

Keywords: distributed detection, decentralized detection, decision fusion, predetection fusion

1 Introduction

Distributing sensors over a large area is necessary in some signal detection and tracking applications, particularly for surveillance systems. Such arrangements may even provide advantages over single sensor systems in terms of reliability, survivability, and improved signal detection performance. These performance improvements are the result of the inherent spatial diversity combining that occurs in such cases provided the sensors are separated by sufficient distances. In the interest of reduction of communication costs, simplification of processing, and preventing interception of one's communications, it is often advisable to use **distributed detection schemes**, which locate processors directly at each sensor. As illustrated in Figure 1, these processors reduce each sensor's observations to a multi-bit decision, and attempt to retain the essential information in these individual decisions needed to make a final signal detection decision.

Specifying the form of the partial processing at the sensors (the sensor decision rules) and specifying how the partially processed observations will be used in the final signal detection decision (the fusion rule) to obtain best

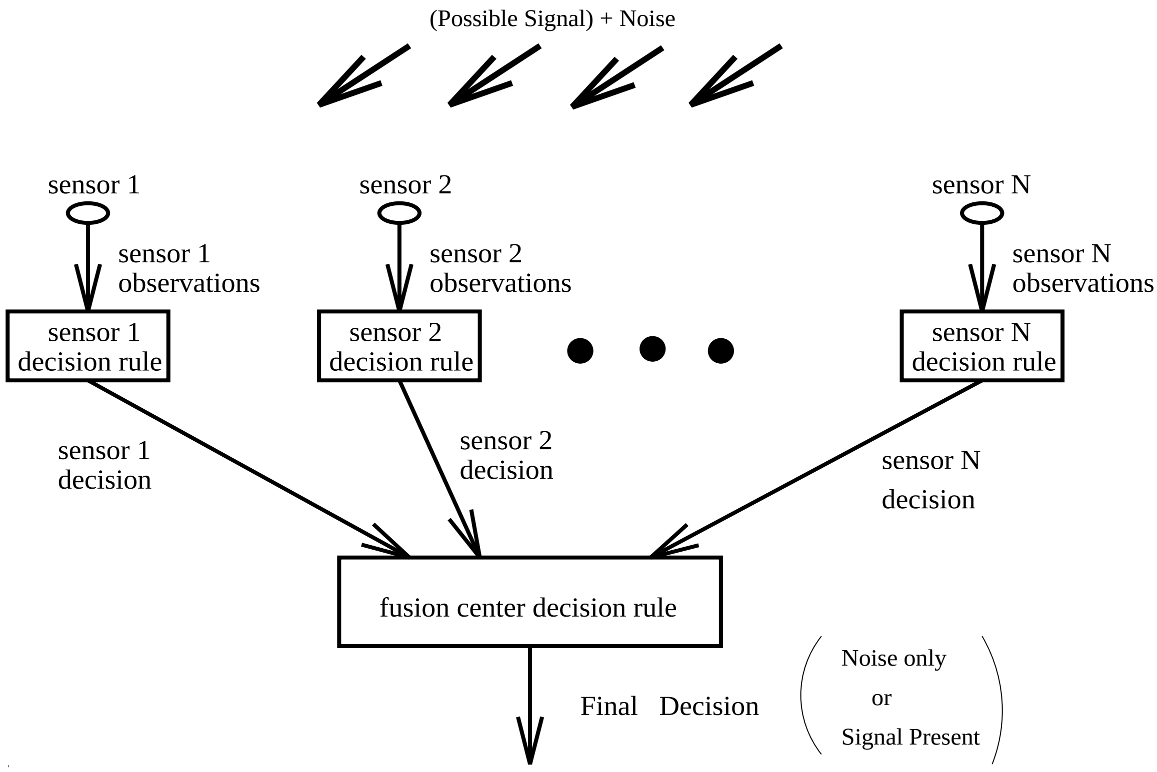


Figure 1: Block diagram of distributed detection scheme.

performance is of fundamental importance. The need to specify each sensor's decision rule and the fusion rule makes distributed detection schemes inherently more complicated to design than the more common **centralized detection schemes**, where all observations are available in their original form at a central location.

There has been significant research on the topic of distributed signal detection lately as described in a recent survey chapter.[1] Several unsolved problems remain and this subject continues to receive attention.[2–20] Designing distributed detection schemes is a difficult problem. This is especially true for cases where observations may be dependent from sensor to sensor. To date there has been very little progress in this area. With the exception of our own work[8–16] and some analysis on using the optimum independent observation schemes,[17,18] we found only one journal paper[19] on this topic, which deals exclusively with using a numerical technique for cases where each sensor decision is based on a single real-valued scalar observation. There has been some research on designing schemes for cases with independent observations from sensor to sensor, see for example,[3,20] but these approaches are generally not applicable to cases with dependent observations from sensor to sensor.[8] The reason is that these approaches capitalize on the optimum distributed sensor test being known to be a scalar quantization of the likelihood ratio, which is not true for cases with dependent observations from sensor to sensor (even if the sensors make hard binary decisions).

Suppose that a distributed sensor provides a vector of observations $\underline{x}$ to its associated processor, which will make a hard binary decision on whether hypothesis H_1 or H_0 is true. Next a final decision is made based on the complete set of sensor decisions. Under some reasonably general conditions, the best sensor test (under several criteria including the Neyman Pearson criterion) compares the test statistic[8]

$$T(\underline{x}) = LR(\underline{x})G(\underline{x}) \tag{1}$$

to a threshold to make a decision, where $LR(\underline{x})$ is the likelihood ratio of the sensor observations and where the function G depends on the design of all the other sensor rules and the fusion rule. If a complete statistical model for all observations is available, then $LR(\underline{x})$ is known. Due to the coupling, obtaining the actual form of G is quite difficult, even if a complete statistical model for all observations is available. However, practical applications require detection schemes that give good performance so we present a simple design approach based on using adaptive processing to minimize a particular mean squared error. While this approach is applicable to a wide range of distributed detection problems we focus here on the detection of a random signal in additive noise.

In some previous work,[14,15] we were able to represent the optimum sensor decision rules (and thus $G(\underline{x})$ in (1)) in terms of some known functions and some unknown scalar parameters for interesting random signal cases with dependent observations from sensor to sensor. To limit complexity, our research focused on an asymptotic criterion which produces optimum results for cases with weak signals and large observation sample sizes. Cases with large sample sizes and weak signals are important in a number of practical applications. The explicit form of the optimum fusion rule was given in our previous work[14,15] in terms of the sensor parameters. Methods for finding the optimum sensor decision rules were described in this previous work, but these methods are very complicated. In our previous work[14,15] we also showed that the best sensor rules will minimize the mean square error between the distributed test statistic and the optimum centralized test statistic for the same problem. This suggests using an adaptive algorithm to find parameters that minimize this mean square error. This is an attractive approach since adaptive algorithms have already been shown to be extremely useful for distributed detection in a related but different context, where one is attempting to learn a set of thresholds for cases with independent observations.[5]

Our training scheme implements the stochastic LMS gradient algorithm, a gradient descent algorithm which has been used successfully in adaptive filter and neural network applications.[21] It is important to note that a gradient descent algorithm does not always find the global minimum. However, here we are interested in looking for any local minimum which gives better performance than the best test under the assumption of independent observations, which others have suggested using.

2 Random Signal Observation Model

Consider the detection of a random signal where the observation at the i^{th} discrete-time instant at the p^{th} sensor is given by

$$X_{p,i} = \theta S_{p,i} + V_{p,i} \tag{2}$$

where $\{V_{p,i}; i = 1, \cdots, n, \quad p = 1, \cdots, N\}$ is a set of independent and identically distributed (iid) zero-mean noise samples with a known common univariate probability density function (pdf) f, θ is a parameter which represents the signal strength, and $\{S_{p,i}; i = 1, \cdots, n, \quad p = 1, \cdots, N\}$ is a set of zero-mean unit variance random signal samples which are generally dependent from sensor to sensor. For simplicity, we assume the signal samples at different time instants are independent and we consider the case where the sensor rules make hard binary decisions. We are also considering cases with multiple bit decisions in our research.[15] We will report on these results at a future conference. We consider the case where each sensor decision is based on the complete set of n observations available at the sensor in question. Under the assumptions outlined, we have shown[14,15] that the best distributed sensor processing schemes perform scalar quantizations of

$$\sum_{i=1}^{n} a_{p,i} \frac{f''(x_{p,i})}{f(x_{p,i})} + \sum_{i=1}^{n} b_{p,i} \frac{f'(x_{p,i})}{f(x_{p,i})} \tag{3}$$

for some set of parameters $a_{p,i}, b_{p,i}, \ i = 1, \cdots, n, p = 1, \cdots, N$ and some thresholds $t_p, \ p = 1, \cdots, N$. In (3) $f'(x)$ and $f''(x)$ denote the first and second derivatives of the noise pdf f.

We propose using an adaptive algorithm to find the values of the parameters $a_{p,i}, b_{p,i}, t_p, \ i = 1, \cdots, n, p = 1, \cdots, N$ which minimize the mean-square error, under $\theta = 0$, between the optimum centralized test statistic and the test statistic for the distributed scheme being designed. For any set of parameters and thresholds, the optimum fusion rule can be found directly from equations obtained from our previous work.[14,15]

Since the statistical description of the observations is known, suitable training data is easily generated and simultaneously input to the distributed scheme under design and the centralized scheme it will be matched to. While an analytical justification for our technique has been given only for asymptotic cases, we believe our technique also has merit for finite sample size cases, which we investigate in this paper.

3 The Distributed Detection Scheme

To simplify matters we consider the case where N individual sensor quantizers each make a single observation and then each sensor quantizer sends a binary decision to a fusion center. This is the case[14] where $n = 1$ in (2) and (3) and the complete set of sensor observations is $\{x_{p,1}, p = 1, \ldots, N\}$. To further simplify our problem we divide equation (3) through by $a_{p,1}$ to yield a p^{th} sensor test which compares

$$\frac{f''(x_{p,1})}{f(x_{p,1})} + \beta_{p,1}\frac{f'(x_{p,1})}{f(x_{p,1})} \tag{4}$$

to an adjusted threshold τ_p where $\beta_{p,1} \equiv \frac{b_{p,1}}{a_{p,1}}$ and $\tau_p \equiv \frac{t_p}{a_{p,1}}$. The division by $a_{p,1}$ is valid under the assumption that $a_{p,1} \neq 0$. If $a_{p,1} = 0$ did provide best performance our techniques should indicate this by producing solutions with extremely large $\beta_{p,1}$ and τ_p.

At sensor p, if the observation $x_{p,1}$ is such that the test statistic in (4) is less than or equal to the threshold τ_p, then the observation is classified as being in the region A_0^p and the sensor quantizer produces the index 0. On the other hand, if the test statistic in (4) is larger than the threshold, then the observation is classified as being in the region A_1^p and the sensor quantizer produces the index 1. The optimum fusion rule[14,15] generates a final decision by comparing the test statistic (the observations are assumed to come from (2))

$$\lambda_{LO} = \sum_{p=1}^{N}\sum_{j=0}^{1} l_{p,j} z_{p,j} + \sum_{p=1}^{N}\sum_{j=0}^{1}\sum_{q=1,q\neq p}^{N}\sum_{r=0}^{1} \tilde{l}_{p,j} z_{p,j} \tilde{l}_{q,r} z_{q,r} \tag{5}$$

to a threshold. The value of the threshold is determined by the required false alarm probability. In (5) we have defined

$$l_{p,j} = \frac{\int_{x\in A_j^p} \frac{f''(x)}{f(x)} f(x)\,dx}{\int_{x\in A_j^p} f(x)dx} \tag{6}$$

and

$$\tilde{l}_{p,j} = \frac{\int_{x\in A_j^p} \frac{f'(x)}{f(x)} f(x)\,dx}{\int_{x\in A_j^p} f(x)\,dx} \tag{7}$$

and $z_{p,j}$ denotes an indicator function such that

$$z_{p,j} = \begin{cases} 1, & \text{if } x_{p,1} \in A_j^p \\ 0, & \text{otherwise} \end{cases} . \tag{8}$$

4 The Training Algorithm

Consider the mean square error between the locally optimum quantized test statistic, λ_{LO}, and the centralized scheme test statistic, λ_{UQ}, as given by

$$J = E\left[\left(\lambda_{UQ} - \lambda_{LO}\right)^2 |\theta = 0\right] \tag{9}$$

where λ_{LO} is given in (5) and λ_{UQ} is given by (the observations are assume to come from (2))

$$\lambda_{UQ} = \sum_{p=1}^{N} \frac{f''(x_{p,1})}{f(x_{p,1})} + \sum_{p=1}^{N} \sum_{q=1,q\neq p}^{N} \frac{f'(x_{p,1})}{f(x_{p,1})} \frac{f'(x_{q,1})}{f(x_{q,1})}. \tag{10}$$

The mean square error J, which is sometimes referred to as a *cost function* in the adaptive signal processing literature, depends on the parameters of the sensor quantizers which are collectively denoted by the vector

$$\vec{W} = \left[\tau_1, \beta_{1,1}, \ldots, \tau_N, \beta_{N,1}\right]^T. \tag{11}$$

In the gradient descent method,[21] the system parameter vector is updated by adjusting it in the direction of steepest descent of the cost function, which is opposite the gradient $\vec{\nabla}_W J$ as in

$$\vec{W}(\ell+1) = \vec{W}(\ell) - \mu\vec{\nabla}_W J, \tag{12}$$

where μ is a small positive constant called the *learning rate* and ℓ is an index which denotes the iteration number in the training. The updated vector $\vec{W}(\ell+1)$ will be closer than $\vec{W}(\ell)$ to the minimum of J provided[21] J is well behaved and has a minimum, μ is sufficiently small, and the initial condition $W(0)$ is in the domain of attraction of the minimum of J. Since λ_{UQ} does not depend on the parameters of the distributed scheme, $\vec{\nabla}_W \lambda_{UQ}$ is zero so that $\vec{\nabla}_W J$ can be written as

$$\vec{\nabla}_W J = -2E\left[\left(\lambda_{UQ} - \lambda_{LO}\right)\vec{\nabla}_W \lambda_{LO} |\theta = 0\right]. \tag{13}$$

For simplicity, the expectation value in the gradient will be approximated during training (with $\theta = 0$), as in

$$\vec{\nabla}_W J \approx -2(\lambda_{UQ} - \lambda_{LO})\vec{\nabla}_W \lambda_{LO}, \tag{14}$$

which is simply the instantaneous value. This approximation reduces the training to the stochastic LMS gradient algorithm which has been previously studied.[21] At each iteration, $\vec{W}$ is updated as

$$\vec{W}(\ell+1) = \vec{W}(\ell) + 2\mu(\lambda_{UQ} - \lambda_{LO})\vec{\nabla}_W \lambda_{LO}. \tag{15}$$

The discontinuity present in λ_{LO} (due to the indicator functions) causes the gradient to be undefined for some parameter settings and inputs. Therefore we have introduced a sigmoid function which smoothes out the discontinuity in the hard decision and gives a continuous and differentiable approximation to $\sum_{j=0}^{1} l_{p,j} z_{p,j}$ as

$$L_p = l_{p,0} + \frac{l_{p,1} - l_{p,0}}{1 + exp\left(-\gamma\left(\frac{f''(x_{p,1})}{f(x_{p,1})} + \beta_{p,1}\frac{f'(x_{p,1})}{f(x_{p,1})} - \tau_p\right)\right)} \tag{16}$$

In a similar way we have defined a continuous, differentiable approximation to $\sum_{j=0}^{1} \tilde{l}_{p,j} z_{p,j}$ as

$$\tilde{L}_p = \tilde{l}_{p,0} + \frac{\tilde{l}_{p,1} - \tilde{l}_{p,0}}{1 + exp\left(-\gamma\left(\frac{f''(x_{p,1})}{f(x_{p,1})} + \beta_{p,1}\frac{f'(x_{p,1})}{f(x_{p,1})} - \tau_p\right)\right)} \tag{17}$$

The scaling parameter γ controls the size of the region over which the discontinuity is smoothed. The test statistic is then formed at the fusion center using the approximate values L_p and $\tilde{L}_p$ as

$$\lambda_{LO} \approx \sum_{p=1}^{N} L_p + \sum_{p=1}^{N} \sum_{q=1, q \neq p}^{N} \tilde{L}_p \tilde{L}_q. \tag{18}$$

The training algorithm is summarized in the pseudocode given below.

1. Generate $x_{p,1}$ for each sensor.
2. Calculate λ_{UQ} from equation (10) and λ_{LO} from equation (18). Find the error $\lambda_{UQ} - \lambda_{LO}$.
3. Calculate the gradient $\vec{\nabla}_W \lambda_{LO}$.
4. Update system parameters according to (15).
5. Repeat.

5 Results

We consider a distributed signal detection scheme as shown in Figure 1 which is designed to detect a weak random signal in noise. For simplicity we consider the case of only two sensors and assume the observations come from the model given in (2) with

$$f(x) = \frac{k}{2\mathcal{A}(k)\Gamma(1/k)} \exp -[|x|/\mathcal{A}(k)]^k \tag{19}$$

where

$$\mathcal{A}(k) = \left(\frac{\Gamma(1/k)}{\Gamma(3/k)}\right)^{\frac{1}{2}} \tag{20}$$

and where $\Gamma(a) = \int_0^\infty x^{a-1}e^{-x}dx$ is the gamma function. f is known as a generalized Gaussian pdf, which has zero mean and unit variance. If $k = 2.0$ then f is the Gaussian pdf and if $k < 2.0$ then f is a pdf with heavy tails. The heavy-tailed pdfs appear to be reasonable models for a number of practical cases[22] including impulsive noise.

The sensor decisions are made using (4) and the fusion center generates a final decision using (5). The parameters in the sensor decision rules are learned using the algorithm outlined in Section 4 with $\mu = 0.00001$ in (15) and the scaling parameter $\gamma = 4$ in (16) and (17).

We have investigated the cases of generalized Gaussian noise with parameter $k = 1.8$ and $k = 2.0$. For $k = 2.0$, the system parameters were initially set to $\tau_1 = 0.0$, $\beta_{1,1} = 1.0$, $\tau_2 = 0.0$, and $\beta_{2,1} = -1.0$. After one million iterations, the parameters converged to values near $\tau_1 = 1.195$, $\beta_{1,1} = 0.0$, $\tau_2 = 1.19$, and $\beta_{2,1} = 0.0$, with some fluctuation around these values. Other initial conditions were tried and similar results were obtained. The time evolution of these parameters is shown in Figure 2 and Figure 3.

For $k = 1.8$, the system parameters were again initially set to $\tau_1 = 0.0$, $\beta_{1,1} = 1.0$, $\tau_2 = 0.0$, and $\beta_{2,1} = -1.0$. After one million iterations, the parameters converged to values near $\tau_1 = 0.76$, $\beta_{1,1} = 0$, $\tau_2 = 0.77$, and $\beta_{2,1} = 0$, with fluctuations around these values. Again, other initial conditions were tried for this case and similar results were obtained. The time evolution of these parameters is shown in Figure 4 and Figure 5.

The values obtained in both cases are close to those which were found to be optimum using a different method in previous work.[14] The time evolution of the system parameters exhibited fluctuations around the optimum values. These fluctuations vary with the learning rate. Smaller learning rates cause less fluctuation but also cause slower convergence. An alternative may be to use a decreasing learning rate so that the parameters converge quickly at first and once in the neighborhood of the optimum value, the fluctuations decrease.

6 Conclusion

We have presented an adaptive algorithm which can be used to train a distributed detection system to learn system parameters which provide good performance for the difficult case of weak signals and dependent observations form sensor to sensor. The sensor test statistics and thresholds are adapted in an attempt to achieve a minimum mean square error match to the best centralized test statistic. The particular training algorithm chosen uses the gradient descent method. In the cases we considered, the system parameters found using this method were close to those found to be optimum in some previous research. The technique used in this paper is easier to use than the techniques suggested in previous research. Extensions to cases with multiple observations and multiple bit sensor decisions are currently under investigation.

Acknowledgments

This paper is based on work partially supported by the National Science Foundation under Grant No. MIP-9211298 and by a contract with the DOD MURI program.

7 REFERENCES

[1] J. N. Tsitsiklis, "Decentralized detection," *Advances in Statistical Signal Processing, vol. 2 - Signal Detection*, H. V. Poor & J. B. Thomas editors, JAI Press: Greenwich, Conn. 1990.

[2] S. Alhakeem and P. K. Varshney, "Unified approach to the design of decentralized detection systems," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 31, No. 1, pp. 9-20, Jan. 1995.

[3] C. W. Helstrom, "Gradient algorithm for quantization levels in distributed detection systems," *IEEE Transactions on Aerospace and Electronic Systems*, AES-31, No. 1, pp. 390-398, Jan. 1995.

[4] W. Wang and P. K. Willett, "Constrained selective communication in real-time information systems," *Information and Decision Technologies*, vol. 19, No. 3, pp. 145-163, 1994.

[5] D. A. Pados, P. Papantoni-Kazakos, D. Kazakos and A. G. Koyiantis, "On-line Threshold Learning for Neyman-Pearson Distributed Detection," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 24, No. 10, pp. 1519-1531, Oct. 1994.

[6] V. A. Aalo and R. Viswanathan, "Multilevel quantization and fusion scheme for decentralized detection of an unknown signal," *IEE Proceedings: Radar, Sonar, and Navigation*, vol. 141, No. 1, pp. 37-44, Feb. 1994.

[7] V. V. Veeravalli, T. Basar and H. V. Poor, "Minimax robust decentralized detection," *IEEE Transactions on Information Theory*, IT-40, No. 1, pp. 35-40, Jan. 1994.

[8] R. S. Blum, "Necessary conditions for optimum distributed sensor detectors under the Neyman-Pearson Criterion," *to appear in IEEE Transactions on Information Theory.*

[9] R. S. Blum, "Locally optimum distributed detection of dependent random signals based on ranks," *to appear in IEEE Transactions on Information Theory.*

[10] R. S. Blum and J. Qiao, "Threshold optimization for distributed order-statistic CFAR signal detection," to appear in *IEEE Transactions on Aerospace and Electronic Systems*, Oct. 1995.

[11] R. S. Blum and S. A. Kassam, "Optimum distributed CFAR detection of weak signals," *Journal of the Acoustical Society of America*, Vol. 98, No. 1, pp 221-229, July 1995.

[12] R. S. Blum, "Distributed detection of narrowband signals," *IEEE Transactions on Information Theory*, Vol. 41, No. 2, pp. 519-522, March 1995.

[13] R. S. Blum and S. A. Kassam, "Distributed cell-averaging CFAR detection in dependent sensors," *IEEE Transactions on Information Theory*, Vol. 41, No. 2, pp. 513-518, March 1995.

[14] R. S. Blum, "Quantization in multisensor random signal detection," *IEEE Transactions on Information Theory*, Vol. 41, No. 1, pp. 204-215, Jan. 1995.

[15] X. Zhang and R. S. Blum, "Distributed quantization for signal detection in dependent sensors," *27th Annual Conference on Information Sciences and Systems*, Princeton University, Princeton, NJ, pp. 726-731, March 1994.

[16] R. S. Blum and S. A. Kassam, "Optimum distributed detection of Weak Signals in dependent sensors," *IEEE Transactions on Information Theory*, IT-38, No. 3, pp. 1066-1079, May 1992.

[17] V. A. Aalo and R. Viswanathan, "On distributed detection with correlated sensors: two examples," *IEEE Transactions on Aerospace and Electronic Systems*, AES-25, No. 3, pp. 414-421, May 1989.

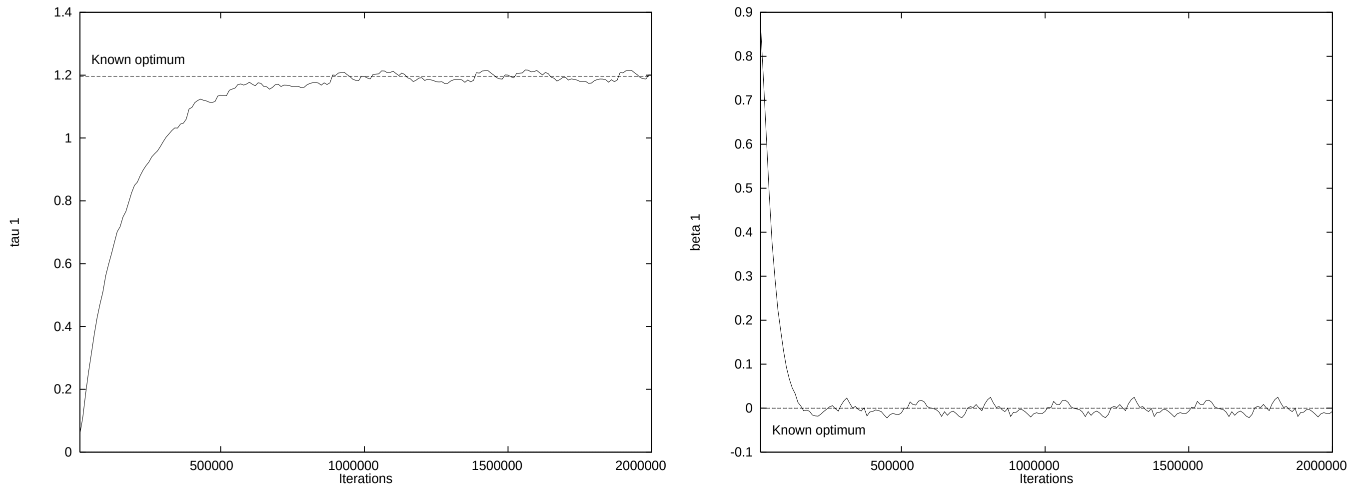
[18] V. Aalo and R. Viswanathan, "Asymptotic performance of a distributed detection system in correlated Gaussian noise," *IEEE Transactions on Signal Processing*, SP-40, No. 1, pp. 211-213, Jan. 1992.

[19] M. Longo, T. Lookabaugh, and R. M. Gray, "Quantization for decentralized hypothesis testing under communication constraints," *IEEE Transactions on Information Theory*, Vol. 36, No. 2, pp. 241-255, March 1990.

[20] Z. Tang, K. R. Pattipati, and D. L. Kleinman, " An algorithm for determining the decision thresholds in a distributed detection problem," *IEEE Transactions on Systems, Man, and Cybernetics* Vol. 21, No. 1, pp. 231-237, Jan 1991.

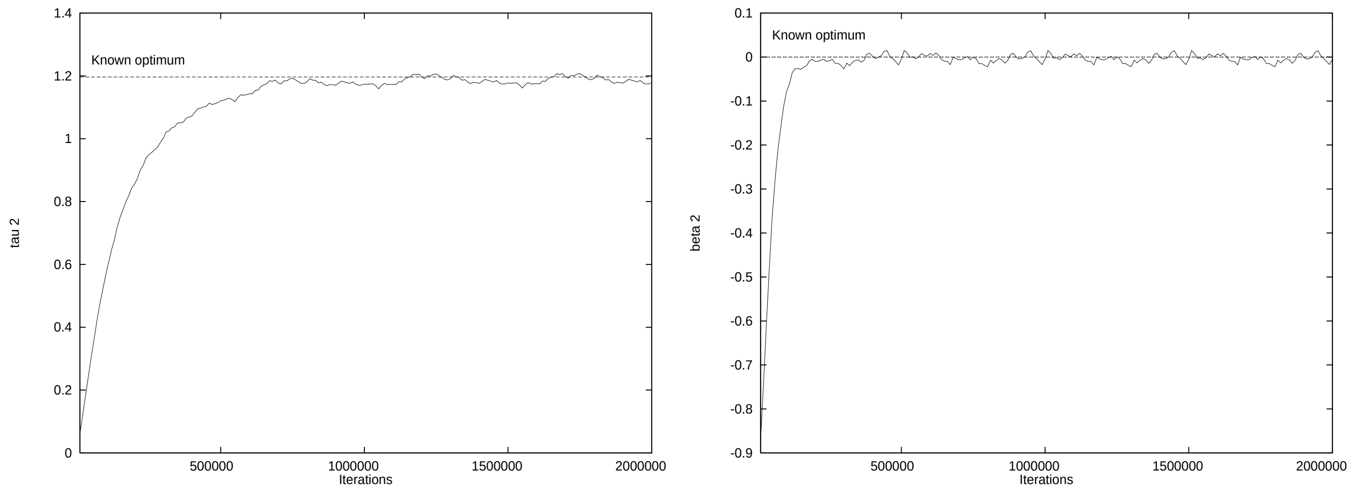
[21] S. Marcos, O. Macchi, and C. Vignat, "A unified framework for gradient algorithms used for filter adaptation and neural network training," *International Journal of Circuit Theory and Applications*, Vol. 20, No. 2, pp. 159-200, Mar 1992.

[22] S. A. Kassam, *Signal Detection in Non-Gaussian Noise,* Springer-Verlag: New York, NY, 1988.



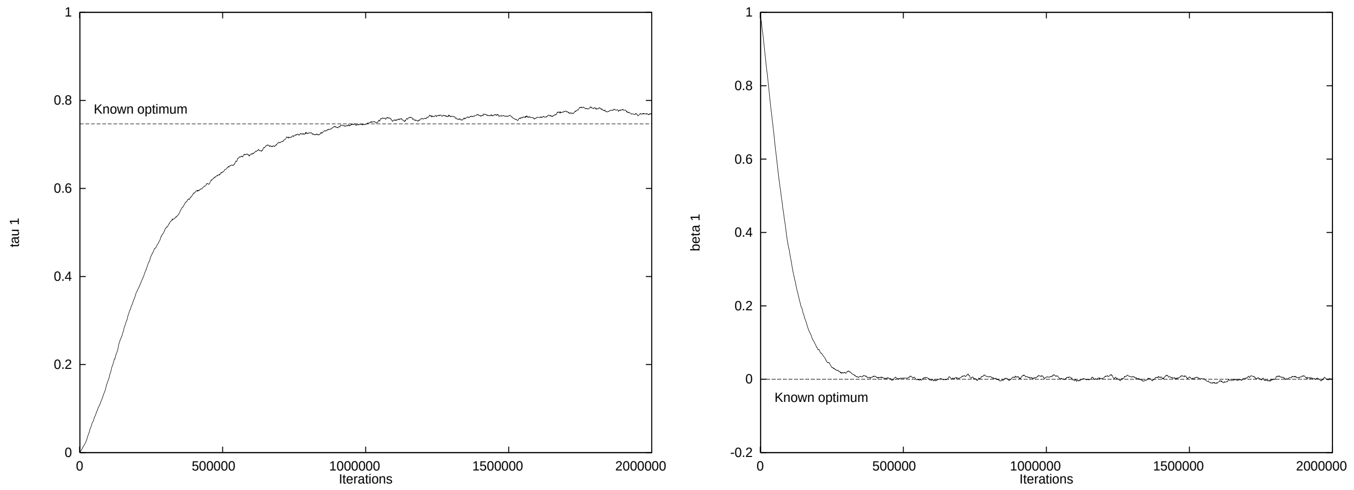
(a) (b)

Figure 2: Time evolution of the parameters for sensor 1 of a two-sensor system and Gaussian noise ($k = 2$). In (a) the threshold τ_1 converges to a value near 1.195, which is close to the known optimum value of 1.1966. In (b) the coefficient $\beta_{1,1}$ converges to a value near 0, which is known to be optimum.



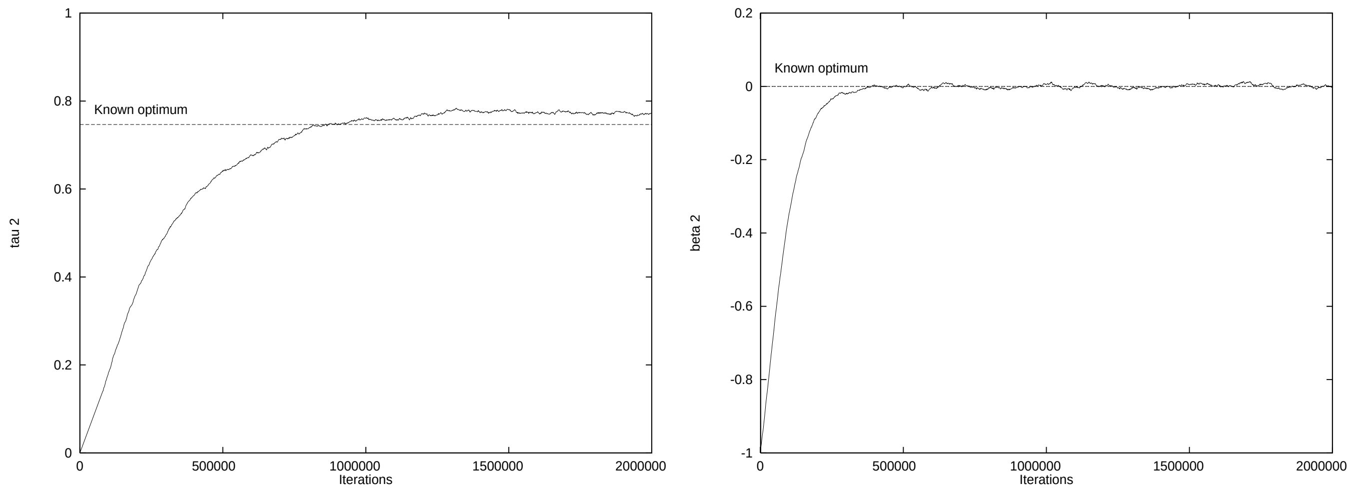
(a) (b)

Figure 3: Time evolution of the parameters for sensor 2 of a two-sensor system and Gaussian noise ($k = 2$). In (a) the threshold τ_2 converges to a value near 1.19, which is close to the known optimum value of 1.1966. In (b) the coefficient $\beta_{2,1}$ converges to a value near 0, which is known to be optimum.



(a) (b)

Figure 4: Time evolution of the parameters for sensor 1 of a two-sensor system and generalized Gaussian noise with parameter $k = 1.8$. In (a) the threshold τ_1 converges to a value near 0.76, which is close to the known optimum value of 0.7467. In (b) the coefficient $\beta_{1,1}$ converges to a value near 0, which is known to be optimum.



(a) (b)

Figure 5: Time evolution of the parameters for sensor 2 of a two-sensor system and generalized Gaussian noise with parameter $k = 1.8$. In (a) the threshold τ_2 converges to a value near 0.77, which is close to the known optimum value of 0.7467. In (b) the coefficient $\beta_{2,1}$ converges to a value near 0, which is known to be optimum.