Detection of Random Signals in Gaussian Mixture Noise

David W. J. Stein
Code 761, NRaD, San Diego, CA 92152-5000

Abstract

A locally optimal detection algorithm for random signals in dependent noise is derived and applied to complex valued Gaussian mixture noise (GMN). The algorithm is modified so that it will detect signals that are not vanishingly small. The resulting detector is essentially a weighted sum of power detectors—the power detector is the locally optimal detector for random signals in Gaussian noise. The performance of the power detector and the locally optimal detector in GMN are compared using simulated and theoretical ROC curves. Additionally, the signal gain of the mixture detector relative to the power detector is calculated, for a fixed false alarm rate, as a function of the mixture parameters. The probability of detection of the mixture detector is also calculated, for fixed parameters and a fixed false alarm rate, as a function of the parameter estimation error.

I. Introduction

Noise, which is the response of a sensor or an array of sensors to a fluctuating number or an insufficient number of identical sources for the Central Limit Theorem to apply, is likely to be non-Gaussian [3,6,8,10,13,14,15]. Gaussian mixture densities have been used to model a variety of non-Gaussian noise environments. If the random variable $X = AU$ where A and U are independent, A is positive and discrete with $p(A = a_i) = p_i$, and U has a Gaussian density with mean μ and covariance matrix Σ, then X is a spherically invariant random variable, and it has the Gaussian mixture density [1,17]

$$p_X(x) = \sum_{i=1}^{m} p_i N(\mu, a_i\Sigma)(x). \tag{1}$$

Thus the Gaussian mixture density applies to data sets for which the variance may vary from sample to sample and for which the data conditioned on the variance are normal. Gaussian mixture densities have been applied to impulsive noise, to various acoustic noise sets, and to sources moving in multi-modal propagation environments [3,10,13,14,15]. Kassam [7] describes other parametric families of densities used to model noise. Noise has also been modeled by using non-parametric techniques [2].

There are various techniques to detect signals in non-Gaussian noise. If the noise-only and signal-plus-noise densities are known, then a likelihood ratio test, using the Neyman-Pearson criterion, provides the greatest probability of detection for a given false alarm rate [12]. If these densities are not known then other approaches may be utilized. If the signal to noise ratio is small, and if the noise density can be approximated by a member of a parametric family of densities for which the parameters can be estimated, or the noise density can be estimated non-parametrically, then locally optimal detection algorithms, which are based on Taylor approximations of the likelihood ratio, may be employed [2,7,9,11,12].

Detection algorithms for a particular application may be selected on the basis of comparative performance analyses and analyses of the sensitivity of the performance to modeling or parameter estimation error. The performance of locally optimal detection algorithms is often studied by using asymptotic relative efficiency (ARE) [7,12]. ARE is most useful if the signal is very weak, and a very large number of samples is available. When these conditions are not met, receiver operating characteristics (ROC), which plot the probability of false alarm versus the probability of detection for fixed values of other parameters, are more revealing. These curves are calculated from probability distributions of the detection statistic conditioned on noise only or signal-plus-noise, and these distributions may be derived or obtained from Monte-Carlo simulations.

II. Locally Optimal Detection of Random Signals

Given data $\{x_1, \ldots, x_n\}$, a detection statistic is used to distinguish between the hypotheses that the data consist of noise only or signal plus noise. The locally optimal detection statistic (LODS) is the first nonvanishing term of the Taylor approximation, at zero signal strength, of the likelihood ratio. Poor and Thomas [11] derive the LODS for zero-mean stochastic signals in independent noise. A formula for the LODS of such signals in dependent noise is stated below—this formula is derived in the complete version of this paper. For $\vec{x} \in \mathbf{R}^n$, let $p_{\mathcal{N}}(\vec{x})$, be the probability density of the noise. Then, under the assumptions that the signal and noise are independent and the probability density of the noise has continuous second-order partial derivatives, the LODS of a zero-mean stochastic signal $\vec{s} = (s_1 \ldots s_n)$, such that $E(s_i^2) = 1$ for $1 \leq i \leq n$, is

$$T(\vec{x}) = \frac{1}{2p_{\mathcal{N}}(\vec{x})} \sum_{i,j=1}^{n} \frac{\partial^2 p_{\mathcal{N}}}{\partial x_i \partial x_j}(\vec{x}) \nu_{ij}, \qquad (2)$$

where $\nu_{ij} = E_s(s_i s_j)$. This detection statistic can be approximated by using either a parametric or a nonparametric model of the probability density of the noise.

This detection statistic is evaluated for independent identically distributed complex Gaussian mixture noise, i.e. $z \in \mathbf{C}$ and

$$p(z) = \sum_{k=1}^{m} p_k \frac{1}{2\pi{\sigma_k}^2} e^{-\frac{\|z\|^2}{2{\sigma_k}^2}}, \qquad (3)$$

where $\sigma_i^2 < \sigma_j^2$ if $i < j$. The detection statistic, applied to a sequence of n complex numbers, $\vec{z} = (z_1 \ldots z_n)$, is

$$T(\vec{z}) = \frac{1}{2} \sum_{i=1}^{n} \sum_{k=1}^{m} \left(\frac{-2}{{\sigma_k}^2} + \frac{\|z_i\|^2}{{\sigma_k}^4} \right) p(k|z_i, H_0), \qquad (4)$$

where $p(z_i|k, H_0)$ is the probability (assuming noise only) of z_i given state k, i.e.

$$p(k|z_i, H_0) = \frac{\frac{p_k}{2\pi\sigma_k^2} e^{-\frac{\|z_i\|^2}{2\sigma_k^2}}}{\sum_{j=1}^{m} \frac{p_k}{2\pi\sigma_j^2} e^{-\frac{\|z_i\|^2}{2\sigma_j^2}}}. \qquad (5)$$

The detection statistic of the power detector, which is the locally optimal detector of random signals in Gaussian noise [12], is equivalent to

$$\mathcal{G}(z) = \sum_{i=1}^{n} \frac{-2}{\sigma^2} + \frac{\|(z_i)\|^2}{\sigma^4}, \qquad (6)$$

where σ^2 is the noise variance. The detection algorithm, which compares (4) to a threshold, is thus a multistate power detector.

If the signal variance is not sufficiently small relative to σ_1^2 the calculation of the conditional probabilities of the noise states under the noise-only hypothesis from signal-plus-noise data may overestimate the probability of the high-noise state. Thus an excess variance term was included in the calculation of the conditional probabilities of the states. From noise-only data one has the parameters $\{m, p_i, \sigma_i^2 | 1 \leq i \leq m\}$, and $T(z_c)$ is to be calculated. Let $\mathcal{W} = \{z_1, \ldots, z_p\}$, be a window of data containing z_c. Define $\sigma_{\mathcal{W}}^2 = .5 \sum_{j=1}^{p} \|z_j\|^2$. If $\sigma_{\mathcal{W}}^2 > \sigma_{\mathcal{N}}^2 = \sum_{i=1}^{m} p_i \sigma_i^2$, then define the excess variance to be $\sigma_e^2 = \sigma_{\mathcal{W}}^2 - \sigma_{\mathcal{N}}^2$. The conditional probabilities of the noise states are then calculated as $p(k|z_c, \vec{\theta}, \sigma_e^2)$, i.e.

$$w_k(z_c) = p(k|z_c, \vec{\theta}, \sigma_e^2) = \frac{\frac{p_k}{2\pi(\sigma_k^2+\sigma_e^2)} e^{\frac{-\|z_c\|^2}{2(\sigma_k^2+\sigma_e^2)}}}{\sum_{j=1}^{m} \frac{p_j}{2\pi(\sigma_j^2+\sigma_e^2)} e^{\frac{-\|z_c\|^2}{2(\sigma_j^2+\sigma_e^2)}}} \qquad (7)$$

The detection statistics studied below are

$$\mathcal{M}(\vec{z}) = \frac{1}{2} \sum_{i=1}^{n} \sum_{k=1}^{m} \left(\frac{-2}{{\sigma_k}^2} + \frac{\|z_i\|^2}{{\sigma_k}^4}\right) p(k|z_i, \vec{\theta}, \sigma_e). \qquad (8)$$

and

$$\hat{\mathcal{M}}(\vec{z}) = \frac{1}{2} \sum_{i=1}^{n} \sum_{k=1}^{m} \left(\frac{-2}{\hat{\sigma}_k^2} + \frac{\|z_i\|^2}{\hat{\sigma}_k^4}\right) p(k|z_i, \hat{\theta}, \hat{\sigma}_e). \qquad (9)$$

where $\hat{}$ indicates that the parameter in question was estimated. In this study, noise parameters were estimated using the EM algorithm [5,18,19].

III. Performance Analysis

For large n the detection statistics have nearly Gaussian distributions, and thus distributions of the statistics under either the noise-only or the signal-plus-noise hypotheses may be obtained from the first two moments. Formulas for the moments of $\mathcal{G}$ and approximate formulas for the moments of $\mathcal{M}$ are given below.

If z has the circular Gaussian density,

$$p(z) = \frac{1}{2\pi\sigma^2} e^{\frac{-\|z\|^2}{2\sigma^2}} \qquad (10)$$

then $u = \|z\|^2$ has the exponential density [35]

$$p(u) = \begin{cases} \frac{1}{2\sigma^2} e^{\frac{-u}{2\sigma^2}} & \text{if } u \geq 0; \\ 0, & \text{otherwise.} \end{cases} \qquad (11)$$

The moments of $\mathcal{G}$ in GMN may be calculated from (6). For $n = 1$,

$$\begin{aligned} E(\mathcal{G}^\ell) &= \sum_{j=1}^{m} p_j E(\mathcal{G}^\ell | j) \\ &= \sum_{j=1}^{m} p_j \sum_{k=0}^{\ell} \binom{\ell}{k} \left(\frac{-2}{\sigma^2}\right)^k \left[\int_0^\infty \left(\frac{t}{\sigma^4}\right)^{\ell-k} \right. \end{aligned}$$

$$\frac{1}{2(\sigma_j^2+\sigma_s^2)} e^{-\frac{u}{2(\sigma_j^2+\sigma_s^2)}} du \Big].$$

$$= \sum_{j=1}^{m} p_j \sum_{k=0}^{\ell} \binom{\ell}{k} \Big[(\frac{-2}{\sigma^2})^k (\frac{2(\sigma_j^2+\sigma_s^2)}{\sigma^4})^{\ell-k} \Gamma(\ell-k+1) \Big], \quad (12)$$

where σ_s^2 is the signal variance and is equal to zero in calculations of the noise-only moments.

Two methods of approximating the moments of $\mathcal{M}$ and $\hat{\mathcal{M}}$ were considered. The clairvoyant detector knows the state that occurs with each datum and is defined by

$$\hat{\mathcal{C}}(\vec{u}) = \sum_{i=1}^{n} \frac{-2}{\hat{\sigma}_{j(i)}^2} + \frac{u_i}{\hat{\sigma}_{j(i)}^4}. \quad (13)$$

where $j(i)$ is the state from which the i^{th} datum arises. $\hat{\mathcal{C}}$ could not be realized because the state corresponding to each datum is generally unknown. However, the first and second conditional moments of $\hat{\mathcal{C}}$ are, for $n = 1$,

$$E(\hat{\mathcal{C}}|j) = \frac{-2}{\hat{\sigma}_j^2} + \frac{\sigma_s^2+\sigma_j^2}{\sigma_j^4}, \quad (14)$$

and

$$E(\hat{\mathcal{C}}^2|j) = \frac{4}{\sigma_j^4} - \frac{8}{\sigma_j^6}(\sigma_s^2+\sigma_j^2) + \frac{8}{\sigma_j^8}(\sigma_s^2+\sigma_j^2)^2. \quad (15)$$

The moments of $\hat{\mathcal{C}}$ are then given by

$$E(\hat{\mathcal{C}}^\ell) = \sum_{j=1}^{m} p_j E(\hat{\mathcal{C}}^\ell|j). \quad (16)$$

Alternatively, $\hat{\mathcal{M}}$ may be approximated by a piecewise polynomial function obtained from piecewise polynomial approximations of the state weighting functions $\hat{w}_j(u) = p(j|u, \hat{\theta}, \sigma_e^2)$, $1 \le j \le m$. One construction of these approximations is described. For an m-state mixture $\lim_{u\to\infty} \hat{w}_m(u) = 1$. Select K so that if $u > K$ then $\hat{w}_m(u) \approx 1$. Select the degree, d, and number p of the polynomials to be used in the approximation. Define a partition $\{0 = u_0 < \cdots < u_{p\cdot d} = K\}$ of $[0, K]$ by $u_r = (r \cdot K)/(d \cdot p)$ for $0 \le r \le p \cdot d$. Then for each $1 \le j \le m-1$ and each $1 \le \ell \le p$ let $P_{\ell j}$ be the Lagrange interpolating polynomial [16] of degree d defined by $[u_{(\ell-1)d}, \ldots, u_{\ell d}]$ and $[\hat{w}_j(u_{(\ell-1)d}, \ldots, \hat{w}_j(u_{\ell d})]$. Other conditions, such as continuity of the derivatives up to certain orders at the points (u_r) or vanishing of certain derivatives at 0 and K, may also be imposed. For a set A, let χ_A be the characteristic function of A, i.e. $\chi_A(u) = 1$ if $u \in A$ and 0 otherwise. Define the piecewise polynomial approximation of the weighting functions by

$$P\hat{w}_j(u) = \sum_{\ell} \chi_{[u_{(\ell-1)d}, u_{\ell d})}(u) P_{\ell j}(u), \quad (17)$$

and the approximate detector by

$$P\hat{\mathcal{M}}(\vec{u}) = \sum_{i=1}^{n}\sum_{j=1}^{m} (\frac{-2}{\sigma_j^2} + \frac{u_i}{\sigma_j^4}) P\hat{w}_j(u_i). \quad (18)$$

Since $P\hat{\mathcal{M}}$ is piecewise polynomial, the conditional moments, for $n = 1$ are easily calculated, and

$$E(P\hat{\mathcal{M}}^\ell) = \sum_{j=1}^{m} p_j E(P\hat{\mathcal{M}}^\ell|j). \quad (19)$$

IV. Numerical Results

Using the above formulas for the moments, approximate probability distributions of the detection statistics were calculated for $n = 65$ by assuming that these distributions are normal. ROC curves for parameter values , $p_1 = .5$, $\sigma_2^2/\sigma_1^2 = 10$, $\sigma_s^2/\sigma_1^2 = 1$, $m = 2$, and $n = 65$ are shown in figure 1. Curves C and E were generated from simulation of $\mathcal{M}$ and $\mathcal{G}$ respectively. Curve D is the ROC curve of the clairvoyant approximation of the mixture detector. Curves A and B are the ROC curves of the polynomial approximations of $\mathcal{M}$ calculated using three polynomials of degree three and two linear polynomials, respectively, to approximate the state 1 weighting function. The piecewise polynomial approximations provided an upper bound on the performance of $\hat{\mathcal{M}}$, while the clairvoyant detector underestimated the simulated performance of the mixture detector. This was due to the higher variance of the clairvoyant detection statistic as compared with the simulated detection statistic and the piecewise polynomial approximations. Curve F is the predicted ROC curve of $\mathcal{G}$.

The performance of $\mathcal{G}$ and $\hat{\mathcal{M}}$ was also compared over a range of values of p_1 and σ_2^2/σ_1^2. Figure 2 is the probability of detection of $\mathcal{G}$ at a fixed probability of false alarm of .001, for $n = 65$, $\sigma_s^2 = .182$, and various values of p_1 and σ_2^2/σ_1^2. The overall noise variance is equal to 1 for all pairs $(p_1, \sigma_2^2/\sigma_1^2)$ Figure 3 shows the theoretical signal gain of the mixture detector relative

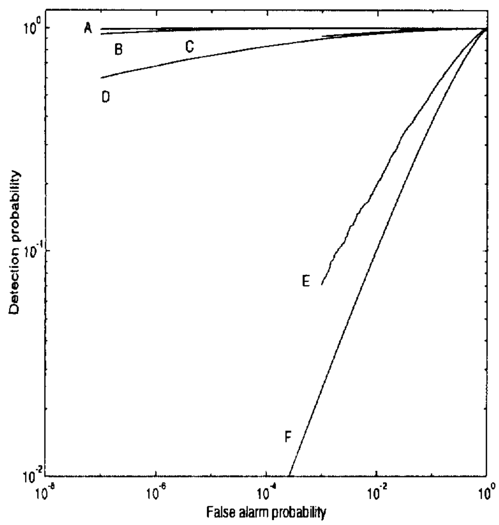

Figure 1: **Simulation generated ROC curves of $\mathcal{G}$ (E) and $\hat{\mathcal{M}}$ (C) are compared with the theoretical ROC curves of $\mathcal{G}$ (F), P$\mathcal{M}$ (A), L$\mathcal{M}$ (B), and $\mathcal{C}$ (D). The parameters used were $p_1 = .5$, $\sigma_1^2 = .182$, $\sigma_2^2 = 1.818$, $\sigma_s^2 = .182$, m = 2,and n = 65.**

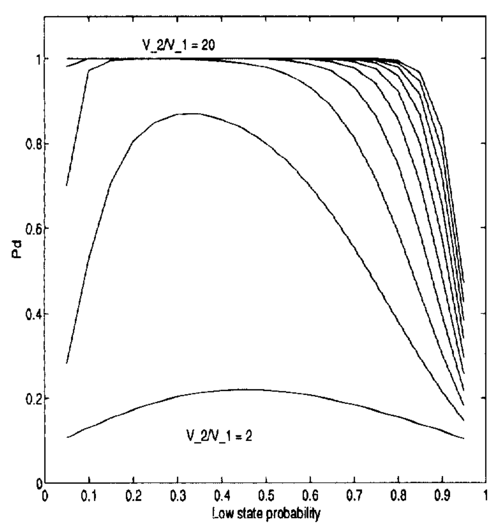

Figure 2: **The probability of detection of $\hat{\mathcal{M}}$ at PFA = .001 for p_1 ranging from .05 to .95, σ_2^2/σ_1^2 = V_2/V_1 ranging from 2 to 20, ($p_1\sigma_1^2 + p_2\sigma_2^2$ was held constant at 1), n = 65, and $\sigma_s^2 = .182$.**

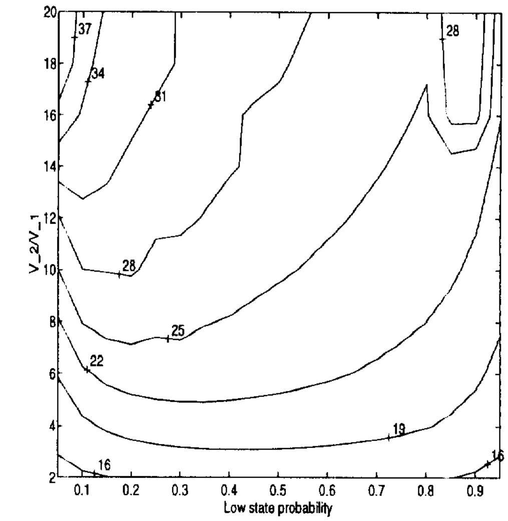

Figure 3: **The signal gain of $\mathcal{M}$ relative to $\mathcal{G}$ (dB) at PFA = .001 for p_1 ranging from .05 to .95, σ_2^2/σ_1^2 = V_2/V_1 ranging from 2 to 20, ($p_1\sigma_1^2 + p_2\sigma_2^2$ was held constant at 1), n = 65, and $\sigma_s^2 = .182$.**

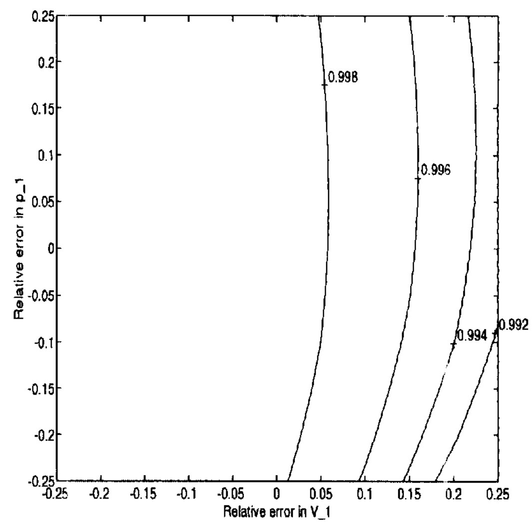

Figure 4: **Contours of the probability of detection of $\hat{\mathcal{M}}$ at PFA = .001, $p_1 = .5$ $\sigma_2^2/\sigma_1^2 = 10$, n = 65, $\sigma_s^2 = .182$, $\sigma_e^2 - \sigma_s^2/\sigma_s^2 = -0.2$, for various relative errors in p_1 and σ_1^2 = V_1 ($p_1\sigma_1^2 + p_2\sigma_2^2$ was held constant at 1).**

to the power detector. If for a fixed false alarm probability σ_g^2 is the signal level which $\mathcal{G}$ requires to achieve the same probability of detection as $\mathcal{M}$ achieves for a signal level σ_s^2 then the signal gain of $\mathcal{M}$ relative to $\mathcal{G}$ is $sg = 10\log_{10}(\sigma_g^2/\sigma_s^2)$. Evidently, for these parameters $\mathcal{M}$ offers significant improvement over $\mathcal{G}$.

The effect of parameter estimation error on the performance of $\hat{\mathcal{M}}$ was also assessed using these methods. Figure 4 is a contour plot of the probability of detection as a function of relative error in p_1 and σ_1^2 for a fixed relative deviation, $(\sigma_e^2 - \sigma_s^2)/\sigma_s^2$. This plot indicates that for the parameters of the calculation $\mathcal{M}$ was rather insensitive to parameter errors.

V. Summary and Conclusions

The locally optimal detection algorithm for random signals in correlated noise was derived and applied to independent circular Gaussian mixture noise. The algorithm was modified so that it could detect signals that are not vanishingly small. The performance of the power detector and this modified locally optimal detector (MLOD) were compared in Gaussian mixture noise using simulation and theoretical performance predictions. It was shown that the simulated performance of the MLOD was close to the performance predicted by using a piecewise polynomial approximation. These approximation techniques also revealed that the mixture detector offers significant gain over the power detector for a broad range of noise parameters, and that the performance of the mixture detector can be fairly insensitive to parameter estimation error.

References

[1] C. R. Baker and A. F. Gualtierotti, "Likelihood Ratios and Signal Detection for Non-Gaussian Processes," *Stochastic Process in Underwater Acoustics*, C. R. Baker, ed. Springer-Verlag, New York, 1986, pp. 154–180.

[2] J. W. Bond, "A Geometric Interpretation of Adaptive Locally Optimum Processing," MILCOM91, IEEE, p. 3.2.1, 1991.

[3] M. Bouvet and S. C. Schwartz, "Underwater Noises: Statistical Modeling, Detection, and Normalization," *J Acoust Soc Am* 83 (3), pp. 1023–1031, 1988.

[4] M. Bouvet and S. C. Schwartz, "Comparison of Adaptive and Robust Receivers," *IEEE Trans ASSP* 37 (5), 1989.

[5] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *J Roy Stat Soc Ser* 39, pp. 1–38, 1977.

[6] J. E. Evans and A. S. Griffiths, "Design of a Sanguine Noise Processor Based Upon World-Wide Extremely Low Frequency (ELF) Recordings," IEEE Trans. Commun., 22 (4), pp. 528–539, 1974.

[7] S. A. Kassam, *Signal Detection in Non-Gaussian Noise*, Springer Verlag, Berlin, 1988.

[8] F. W. Machell, C. S. Penrod, and G. E. Ellis, "Statistical Characteristics of Ocean Noise Processes," in *Topics in Non-Gaussian Signal Processing*, E. J. Wegman, S. C. Schwartz, and J. B. Thomas, eds. Springer-Verlag, New York, 1989.

[9] D. Middleton, "Canonically Optimum Threshold Detection," *Trans Inform Theory* 12 (2), pp. 230–243, 1966.

[10] D. Middleton, "Statistical Physical Models of Electromagnetic Interference," *Trans on EMC* 19 (3), pp. 106–127, 1977.

[11] H. V. Poor and J. B. Thomas, "Locally Optimum Detection of Discrete-Time Stochastic Signals in non-Gaussian Noise," *J Acoust Soc Am* 63 (1), pp. 75–80, 1978.

[12] H. V. Poor, *An Introduction to Signal Detection and Estimation*, Springer-Verlag, New York, 1988.

[13] D. R. Powell and G. R. Wilson, "Class A Modeling of Ocean Acoustic Noise Sources," pp. 17–28, *Topics in Non-Gaussian Signal Processing*, E. J. Wegman, S. C. Schwartz, and J. B. Thomas, eds. Springer-Verlag, New York, 1989.

[14] K. J. Sangston and K. R. Gerlach, "Non-Gaussian Models and Coherent Detection of Radar Targets," NRL/FR/5341-92-9367, 1992.

[15] D. W. Stein, "Statistical Characteristics of Moving Acoustic Sources in Ocean Waveguides," in progress.

[16] J. Stoer and R. Bulirsch, *Introduction to Numerical Analysis*, Springer Verlag, New York, 1993.

[17] D. M. Titterington, A. F. M. Smith, and U. E. Makov, *Statistical Analysis of Finite Mixture Distributions*, John Wiley and Sons, New York, 1985.

[18] C. F. J. Wu, "On the Convergence Properties of the EM Algorithm," Ann Stat, 11 (1), pp. 95–103, 1983.

[19] S. M. Zabin and H. V. Poor, "Efficient Estimation of Class A Noise Parameters Via the EM Algorithm," *Trans Inform Theory* 37 (1), pp. 60–72, 1991.