MAXIMUM ENTROPY AND AVERAGE ERROR RATES IN DIGITAL COMMUNICATION SYSTEMS

Fritz Solms *, Pieter van Rooyen † and Jan Kunicki ‡
* Department of Applied Mathematics, RAU, P.O. Box 524, JHB, 2006
† Alcatel Altech Telecoms, P.O. Box 286, Boksburg, 1460
‡ Cybernetics Lab., RAU, P.O. Box 524, JHB 2006

ABSTRACT

In digital communication systems, the criteria of merit of system performance is usually average probability of error as function of signal-to-noise ratio. The Gauss Quadrature Rule (GQR) formulation and the Maximum Entropy Method (MEM) have been proposed in the literature to determine and calculate an unknown distribution and average error rate from moments. We compare the accuracy of these two methods when a distribution is estimated and the average error rate is calculated. It is shown that the MEM needs significantly less moments when the distribution is estimated from its moments than the GQR formulation, and that the GQR formulation fails under certain conditions when average error rate is calculated. Specifically, the latter is encountered at high signal-to-noise ratios, where the MEM still delivers reliable results.

1. INTRODUCTION

In order to evaluate average bit error rates of digital communication systems (e.g. optical fiber communication systems [1], spread-spectrum multiple access systems [2, 3], digital microwave radio systems [4], ...) due to i.e. noise or intersymbol interference (ISI), one usually has to evaluate integrals of the form

$$\tilde{P}_e = \int_a^b P_e(x)p(x)dx. \qquad (1)$$

Here X is a random interference variable with unknown probability distribution function (pdf) $p(x)$ whose moments can either be calculated or measured. In practical communication systems X could be a random attenuation variable which can take any value within a continuous range. Alternatively X could be interfering messages in a multi-user system with pulse streams of length ℓ which can take any of D^ℓ values, D being the number of levels. The number of such interfering messages is usually very large and direct evaluation (the exhaustive method) is computationally impractical.

One of the most widely used techniques for evaluating the average error probability [5, 6, 7, 8, 2] is based on the Gauss Quadrature Rule (GQR) method of Golub and Welsh [9]. The moments of the pdf

$$\mu_m = \langle x^m \rangle \equiv \int_a^b x^m p(x)dx \quad ; \quad m = 0, \ldots, M \qquad (2)$$

are used to determine a three-term recurrence relation for the sequence of polynomials $\mathcal{P}_n(x)$ orthonormal to the weight function (the pdf) and hence the tridiagonal matrix $\mathbf{J}$. The eigenvalues and the first component of the orthonormal eigenvectors of $\mathbf{J}$ are then used to determine the weights w_n while the roots of the polynomial $\mathcal{P}_{N+1}(x)$ determine the nodes x_n of the quadrature rule $\{w_n, x_n\}_{n=1}^N$ with $M = 2N+1$, and consequently

$$\tilde{P}_e \approx \sum_{n=1}^{N} w_n P_e(x_n). \qquad (3)$$

It can be easily seen [10] that the Gauss quadrature rule is exact if $f(x)$ is a polynomial of degree $2N-1$ or less. The GQR method is computationally reasonably inexpensive and provides accurate results for many problems, provided one has a sufficient number of moments. However, the GQR method fails for certain frequently encountered problems and, in particular, at large signal-to-noise ratios, i.e. when the subsequent moments grow in absolute size. In these cases the algorithm converges extremely slowly (typically more than 80 moments are required) and it becomes numerically unstable before it becomes accurate. Gautshi [11, 8] has developed a more stable algorithm based on modified moments — the latter can be computed directly from the moments μ_m. Although this modified GQR method can treat a much wider class of problems it still fails at high signal-to-noise ratios.

An alternative approach is to use the maximum entropy method (MEM) [15, 16] to determine the unknown pdf from its moments and then to use standard integration methods to calculate the error probabilities. Kavehrad and Joseph [12] have shown that the maximum entropy method yields results that compare well with GQR method but using less moments than the latter.

In this paper we show that the MEM continues to give accurate results for problems where the GQR method fails. In some cases the moments of the pdf of the interference variable are obtained experimentally. In

 0-7803-1998-2/94/$4.00

those cases one is forced to infer the average error probability from a very limited number of moments (typically four) and the GQR method is very likely to fail. The MEM results, on the other hand, are much more reliable. For example, for and AWGN channel with PSK modulation subject to slow Rayleigh fading two moments suffice to infer an analytically exact pdf. Furthermore, the GQR method is restricted to using moments for the information of the pdf while the MEM lends itself to including other useful information if available.

A further advantage of the MEM is that it gives an analytical expression for the inferred pdf, while the GQR method generally only gives an estimate of the average error probability. Attempts to use the GQR method to infer the pdf itself [8] have proven less successful. This method shares the problems of numerical instability of the GQR method. Furthermore, in order to obtain the pdf on a relatively fine grid (for large N) an extremely high number of moments, $M = 2N+1$, is required which in turn often leads to numerical instability. The MEM, on the other hand, gives a full functional form for the inferred pdf for any number of moments, and in general a much smaller number of moments is required to obtain accurate results.

We apply the above formalisms to two model problems where we compare the exact distribution and average error probability to that obtained by the MEM and the GQR methods. In particular we investigate the Exponential pdf and also calculate the average error rate of coherent PSK under Rayleigh fading conditions. We find that the GQR method fails in both cases; the estimated pdf is not very accurate and the error rate is completely inaccurate at large signal-to-noise ratios. On the other hand, the MEM results compare very well with the known analytical results.

The next section will provide us with the necessary background to the MEM, and also discusses the numerical algorithms implemented to obtain results. Sections 3 and 4 discuss, respectively, the estimated pdf and average error rate results, while Section 5 conclude the paper.

2. THE MAXIMUM ENTROPY PRINCIPLE

Shore and Johnson [14] have proven that the maximum entropy method (MEM) [15, 16] is the only method for inferring from incomplete information, that does not lead to logical inconsistencies. This prove has put the maximum entropy principle on a very solid foundation. The MEM has been applied to many inference problems including image reconstruction [18], search theory [19], scattering problems, chaos [20] and many more.

In the MEM the missing information (the information entropy) [24]

$$I(p) = -\int_a^b p(x)\ln p(x)dx \tag{4}$$

is maximized subject to the constraints of the normalization of the pdf and subject to the available information. In our case the expectation values of the moment operators must be equal to the measured or calculated moments, i.e. the pdf must satisfy equations (2). This is a standard maximum entropy moments problem (The MEM moment problem has been studied in detail by Tagliani [23]). The constraints are introduced via Lagrange multipliers and the resulting expression for the inferred pdf is

$$p^*(x) = \frac{1}{Z}\exp\left(-\sum_{m=1}^{M}\lambda_m x^m\right) \tag{5}$$

where the information about the normalization is contained in the partition function

$$Z = \int_a^b \exp\left(-\sum_{m=1}^{M}\lambda_m x^m\right)dx, \tag{6}$$

and the Lagrange multipliers are determined by requiring that

$$\langle x^m\rangle^* = \mu_m \quad ; \quad m = 1,\ldots,M. \tag{7}$$

Agmon, Alhassid and Levine [25] have noted that defining

$$F(\{\lambda_m\}) = \ln Z + \sum_{m=1}^{M}\lambda_m\mu_m \tag{8}$$

yields

$$\frac{\partial F}{\partial\lambda_m} = \mu_m - \langle x^m.\rangle \tag{9}$$

Hence minimizing F is equivalent to solving the set of coupled nonlinear equations (7). Furthermore, the authors have shown that the Hessian matrix $\mathbf{H}$ with

$$H_{mm'} = \frac{\partial^2 F}{\partial\lambda_m\partial\lambda_{m'}} = \langle x^{m+m'}\rangle - \langle x^m\rangle\langle x^{m'}\rangle \tag{10}$$

is positive definite and thus that F is a strictly convex function of the Lagrange multipliers $\{\lambda_m\}$. Consequently F has a unique minimum (i.e. (7) has a unique solution) and a Newton-Raphson minimization procedure [13] is guaranteed to converge. Define an error vector

$$\overline{\epsilon} = (\epsilon_1,\ldots,\epsilon_M)^T \quad : \quad \epsilon_m \equiv \mu_m - \langle x^m\rangle, \tag{11}$$

and let $\overline{\lambda} = (\lambda_1,\ldots,\lambda_M)$. Then the new guess after a Newton Raphson step is

$$\overline{\lambda}' = \overline{\lambda} - \mathbf{H}^{-1} \cdot \overline{\epsilon}. \quad (12)$$

During each iteration we solve a set of coupled linear equations for the Newton step $\overline{\delta} = \overline{\lambda} - \overline{\lambda}'$

$$\mathbf{H} \cdot \overline{\delta} = \overline{\epsilon}. \quad (13)$$

Since $\mathbf{H}$ is positive definite it is also non-singular. We solve equation (13) with a standard LU-decomposition with a back-substitution algorithm coded in C^{++}

To combat spurious divergence problems due to computer round-off errors we translated the "globally convergent" Newton-Raphson method found in *Numerical Recipes in C* [26] into C^{++} using our vector and matrix classes. Doing this allowed us to increase the number of Lagrange multipliers to about 24.

3. ESTIMATED DISTRIBUTION VIA GQR AND MEM

In digital communication problems, we often do not know the distribution of a random variable. The two proposed methods allow us to estimate the desired pdf via its moments. We will examine the accuracy of the two methods by investigating the Exponential distribution, given by

$$p(\gamma_b) = \frac{1}{\tilde{\gamma}_b} \exp\left\{-\frac{\gamma_b}{\tilde{\gamma}_b}\right\} \qquad \forall \qquad \gamma_b \geq 0 \quad (14)$$

with moments

$$\mu_i = \tilde{\gamma}^i i!. \quad (15)$$

This distribution is also used in the next section to calculate average error probabilities.

It is clear from Figure 1 that the GQR algorithm fails again dismally, even for $M = 37$ moments, irrespective of the use of modified moments or the modified Cholesky decomposition.

Figure 2 shows the pdf obtained with MaxEnt; very accurate tail probabilities are obtained with only two moments(μ_0 and μ_1). Table 1 indicates the coefficients, λ_r, when higher order moments are used.

It is quite clear that all the higher order λ_r's are zero for all practical purposes, and therefore very accurate average error probabilities can be obtained.

4. APPLICATION TO DIGITAL COMMUNICATION SYSTEMS

As mentioned earlier we will investigate the average error rate of coherent PSK under slow Rayleigh fading conditions. For BPSK the bit stream is subdivided

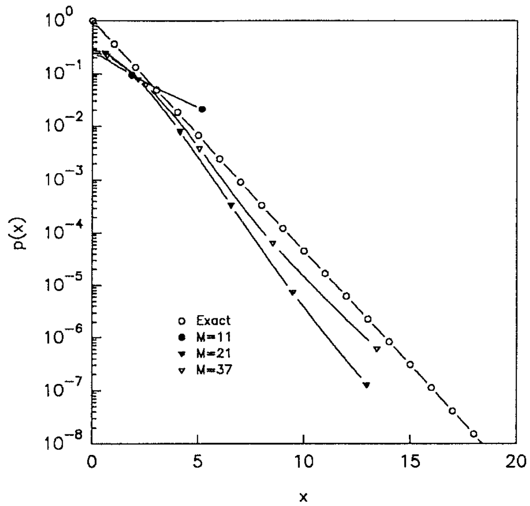

Figure 1: Convergence of GQR estimator for Exponential distribution

into blocks of length ℓ, each block being one of L=2^ℓ messages and each of the these messages is modulated onto L waveforms that differ in phase. The probability P_e of a binary digit error under AWGN conditions is given by [27]

$$P_e(\gamma) = \frac{1}{2}\text{erfc}(\sqrt{\gamma}), \quad (16)$$

where $\gamma = \alpha^2\mathcal{E}/N_0$ is the received signal-to- noise ratio per information bit and

$$\text{erfc(x)} = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-x^2} dx, \quad (17)$$

$\mathcal{E}$ is the average energy per bit and N_0 is the power spectral density of the noise. The attenuation factor α is assumed to be a time independent random variable which is Rayleigh distributed. It therefore follows [27] that the signal-to-noise ratio is exponentially distributed as in (14), moments as in (15) and $\tilde{\gamma}$ is the expectation value of γ, given by

$$\tilde{\gamma} \equiv \langle\gamma\rangle = \int_0^\infty \gamma p(\gamma) d\gamma. \quad (18)$$

The average bit error rate is then obtained via

$$\tilde{P}_e = \int_0^\infty P_e(\gamma) p(\gamma) d\gamma. \quad (19)$$

This integral can be evaluated analytically and the result is [27]

$$\tilde{P}_e = \frac{1}{2}\left\{1 - \sqrt{\frac{\tilde{\gamma}}{1+\tilde{\gamma}}}\right\} \quad (20)$$

N_m	Z	λ_r			
3	1	1	-7.504413e-16	-	-
4	1	1	6.05662e-15	-5.415413e-16	-
5	1	1	1.35561e-19	-1.303146e-15	-1.154032e-17
Exact	1	1	0	0	0

Table 1: Estimated Exponential distribution λ_r's, $\overline{\gamma}_b = 1$

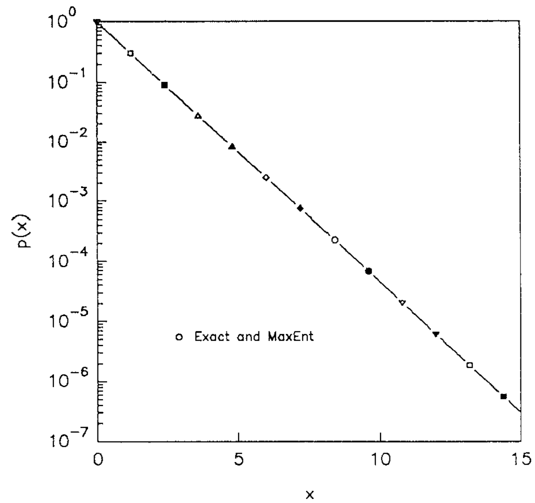

Figure 2: Estimated Exponential distribution obtained with MaxEnt

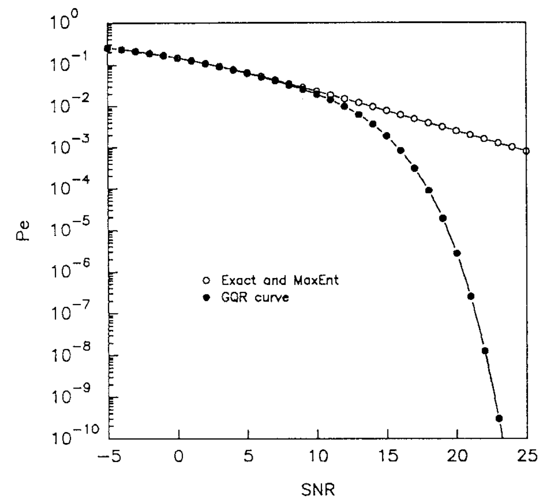

Figure 3: Exact average error rate compared to MEM and GQR results.

Note that the average error rate goes to zero and $\frac{1}{2}$ as the signal to noise ratio (SNR) tends to infinity and zero respectively.

Assume now that we had not known that the SNR is exponentially distributed. Instead we only have the first few moments of the pdf, given by (15), or a small number of measured moments.

In the GQR formulation we perform the quadrature

$$\tilde{P}_e = \frac{1}{2}\sum_{n=1}^{N} W_n \text{erfc}\left\{\sqrt{\gamma_n}\right\} \tag{21}$$

using the quadrature rule (W_n, γ_n) obtained from the method of moments as described in [8]. Figure 3 compares the exact result of the average error rate as a function of the signal to noise ratio with the MEM result (using 2 moments) and the GQR result for 31 moments.

The GQR method becomes numerically unstable before it becomes accurate — the GQR method fails for more than 31 moments. Note that the MEM result coincides with the exact result, while the GQR result becomes completely inaccurate for large values of the signal-to-noise ratio. Furthermore, when only a limited number of moments are available (as is usually the case when the moments are obtained experimentally) the MEM becomes indispensable.

5. SUMMERY AND CONCLUSIONS

In digital communication systems one often has to calculate the error probability averaged over an unknown pdf whose moments and be either measured experimentally or can be calculated. We have found that the widely used GQR based methods tend to fail for large values of the signal to noise ratios and for low error probabilities. The maximum entropy method, on the other hand, continues to yield reliable results.

Although the algorithm to determine a distribution via the GQR formulation are much simpler and computationally quicker (roughly orders quicker) than that of the MEM formulation, it is essentially worthless when accurate estimates of the desired pdf is needed. It is therefore suggested that a new estimator be investigated to determine the distribution via GQR methods. We do not, however, suggest that GQR's can not deliver accurate estimates of a desired distribution, but merely suggests that the estimator as described by Meyer is not sufficient.

Using only a small number of moments has additional benefits in digital communications; it can be shown that in SSMA only a limited number of moments are available under certain conditions. Further, in a practical system only a small number of moments can be measured efficiently.

Applying these methods to AWGN channels with

BPSK signals subject to slow Rayleigh fading we find that the MEM reproduces the analytical result with only two moments. The GQR methods give accurate results in the range of low signal to noise ratio and high error probabilities, but a much higher number of moments (typically about 60).

In general we find that the MEM gives more accurate results than the GQR based methods with fewer moments. This is important for problems where a) the moments of the pdf are obtained experimentally (typically only four moments are measured), b) where the moments are difficult to calculate and c) where the moments grow exponentially causing the GQR methods to become numerically unstable before they become accurate. We find, however, that the MEM requires much more computer time than the GQR methods.

References

[1] R. Dogliotti, A. Luvison and G. Pirani, IEEE Trans. Inform. Theory, **IT-25** (1979) 170.

[2] D. Laforgia, A. Luvison and V. Zingarelli, IEEE Trans. Commun., **COM-32** (1984) 660.

[3] G.R. Cooper and R.W. Nettleton IEEE Trans. Veh. Technol. **VT-27** (1978) 264.

[4] E. Damosso, R. Failli, M. Sentinelli and V. Zingarelli, Proceedings European Conference on Radio Relay Systems, München, 4th–7th Nov. 1986, pp. 82.

[5] S. Benedetto, E. Biglieri, A. Luvison and V. Zingarelli, IEE Proceedings-I **139** (1992) 258.

[6] S. Benedetto, G. De Vincentiis and A. Luvison, IEEE Trans. Comm., **COM-21** (1973) 181.

[7] S. Benedetto, G. De Vincentiis and A. Luvison, IEEE Trans. Comm., **COM-21** (1973) 1159.

[8] M.H. Meyers, The Bell Systems Technical Journal (November 1982) 2245.

[9] G.H. Golub and J.H. Welsh, Math. Comp., **26** (1969) 221.

[10] P.J. Davis and P. Rabinowitz, *Methods of Numerical Integration*, Academic Press, New York (1975).

[11] W. Gautschi, Math. Comp., **24** (1970) 245.

[12] M. Kavehrad and M. Joseph, IEEE Trans. Commun., **COM-34** (1986) 1183.

[13] L.R. Mead and N. Papanicolaou, J. Math. Phys., **28** (1984) 2404.

[14] J.E. Shore and R.W. Johnson, IEEE Trans. Inform. Theory, **IT-26** (1980) 26.

[15] E.T. Jaynes, Phys. Rev., **106** (1957) 620.

[16] E.T. Jaynes, Phys. Rev., **108** (1957) 171.

[17] S. Kullback, *Information Theory and Statistics*, Wiley, New York (1959).

[18] X. Zhuang, R.M. Haralick and Y. Zhao, IEEE Trans. Signal Processing, **39** (1991) 1478.

[19] E.T. Jaynes, in *Maximum-Entropy and Bayesian Methods in Inverse Problems*, Kluwer, Dordrecht (1985) 443.

[20] W.H. Steeb, F. Solms and R. Stoop, *Chaotic Systems and the Maximum Entropy Principle*, accepted for publication in J. Phys. A (1994).

[21] X. Zhuang, L. Chen and S. Chen, IEEE Trans. Signal Processing, **41** (1993) 1730.

[22] J.E. Shore, IEEE Trans. Acoust., Speech, Signal Processing, **ASSP-29** (1981) 230.

[23] A. Tagliani, J. Math. Phys., **34** (1993) 326.

[24] C.E. Shannon, Bell. Syst. Tech. J., **27** (1948) 379, 623.

[25] N. Agmon, Y. Alhassid and R.D. Levine, J. Comput. Phys., **30** (1979) 250.

[26] W.H. Press, B.P. Flannery, S.A. Teukolski and W.T. Vetterling, *Numerical Recipes in C, 2nd edition*, Cambridge University Press (1992).

[27] J.G. Proakis, *Digital Communications*, McGraw-Hill, (1983).