Maximum entropy and minimum relative entropy in performance evaluation of digital communication systems

F. Solms
P.G.W. van Rooyen
J.S. Kunicki

Indexing terms: Digital communication systems, Maximum entropy methods

Abstract: The authors show that the Gauss-quadrature rule (GQR) method, which is widely used in performance evaluation of digital communication systems, fails under certain frequently encountered conditions. The maximum entropy method (MEM), on the other hand, continues to give reliable results for most of these problems. In cases where the MEM becomes numerically unstable before it achieves sufficient accuracy the authors incorporate a prior estimate of the error probability distribution function (PDF) in a minimum relative entropy (MREM) algorithm in order to improve on the accuracy of their results. They compare the MEM and MREM results to those obtained with the GQR method for AWGN channels with single Rayleigh-path-faded BPSK signals to the P-branch diversity systems. In both cases the GQR method fails for large values of the SNR.

1 Introduction

To evaluate average bit error rates for digital communication systems (e.g. optical fibre communication systems [1] or spread-spectrum multiple-access systems [2, 3]) 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 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 multiuser system with pulse streams of length l. The number of such interfering messages is usually very large and direct evaluation (the exhaustive method) is computationally impractical.

Paper 1928I (E7, E8), first received 13th July 1994 and in revised form 18th January 1995

F. Solms is with the Department of Applied Mathematics, Randse Afrikaanse Universiteit, PO Box 524, Auckland Park, 2006, South Africa

P.G.W. van Rooyen is with the Department of Electrical and Electronical Engineering, University of Pretoria, 0002, South Africa

J.S. Kunicki is with the Department of Electrical Engineering, Randse Afrikaanse University

One of the most widely used techniques for evaluating the average error probability [2, 4–6] is based on the Gauss quadrature rule (GQR) method of Golub and Welsh [7]. $M = 2N + 1$ moments

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

of the PDF are used to determine an N-point quadrature rule $\{w_n, x_n\}_{n=1}^N$ and

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

The Gauss quadrature rule is exact if $p(x)$ is a polynomial of degree $2N - 1$ or less. Furthermore, the GQR method is computationally reasonably inexpensive and provides accurate results for many problems provided one has a sufficient number of moments. However, for high signal/noise ratios the subsequent moments grow exponentially in size and the algorithm either converges extremely slowly (typically more than 80 moments are required) or it becomes numerically unstable before it becomes accurate. Although the modified GQR method developed by Gautshi [8, 6] is more stable, it still fails for sufficiently high signal/noise ratios. This drawback is quite significant, since one usually aims to work at low error probabilities and hence necessarily at high SNR.

An alternative approach is to use the maximum entropy method (MEM) [9] to determine the unknown PDF from its moments and then to use some standard integration method to calculate the error probabilities. Kavehrad and Joseph [10] have shown that the maximum entropy method yields results that compare well with the GQR method but using fewer moments than the latter.

In this paper we show that the MEM continues to give accurate results for most problems where the GQR methods fail and also how one can use the minimum relative entropy method (MREM) to improve on the accuracy of the results by incorporating a prior estimate of the PDF.

Furthermore, the MEM and MREM give 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 [6] have proven less successful. This method shares the problems of numerical instability of the GQR method. Furthermore, 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.

We apply the above formalisms to two model problems where we compare the exact average error probability of a known PDF with that obtained by the MEM and the GQR methods, respectively. We find that the GQR method fails in both cases for large SNR (i.e. for low bit error rates). In the first application the MEM provides the analytically exact PDF with only two moments irrespective of the SNR. In the second application, for large SNR, even the MEM becomes numerically unstable before it becomes sufficiently accurate. In this case we use the MREM to improve the accuracy of the results.

Finally we look at a decoding problem where very tight bounds to the average error rate are known. In this case it is found that both the MEM and the GQR method yield results which are either within or very close to these bounds. The MEM, however, requires far fewer moments to achieve these results.

2 MEM and MREM

Shore and Johnson [11] have proven that the maximum entropy method (MEM) [9] (or the minimum relative entropy method (MREM) [12] in the case where the prior distribution is not uniform) is the only method for inferring from incomplete information that does not lead to logical inconsistencies. This proof has put the maximum entropy principle on a very solid foundation. The MEM has been applied to many inference problems including image reconstruction [13], nuclear physics [14, 15] and chaos [16], and the MREM has found some interesting applications in image reconstruction [17] and spectral analysis [18].

In the MEM the missing information (the information entropy) [20]

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

is maximised subject to the constraints of the normalisation 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 eqns. 2. This is a standard maximum entropy moments problem (the MEM moment problem has been studied in detail by Tagliani [19]). 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 normalisation 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 eqn. 2 is satisfied. Alhassid, Agmon and Levine [21] have noted that defining

$$F(\{\lambda_m\}) = \ln Z + \sum_{m=1}^{M} \lambda_m \mu_m \tag{7}$$

yields

$$\frac{\partial F}{\partial \lambda_m} = \mu_m - \langle x_m \rangle \tag{8}$$

Hence minimising F is equivalent to solving the set of coupled nonlinear eqns. 2. The Hessian matrix

$$H_{mm'} = \frac{\partial^2 F}{\partial \lambda_m \partial \lambda_{m'}} = \langle x^{m+m'} \rangle - \langle x^m \rangle \langle x^{m'} \rangle \tag{9}$$

is positive-definite and F is thus a strictly convex function of the Lagrange multipliers $\{\lambda_m\}$. Consequently F has a unique minimum and a Newton–Raphson minimisation procedure [22] is guaranteed to converge. Define an error vector

$$\bar{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_M)^T : \varepsilon_m \equiv \mu_m - \langle x^m \rangle \tag{10}$$

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

$$\bar{\lambda}' = \bar{\lambda} - \boldsymbol{H}^{-1} \cdot \bar{\varepsilon} \tag{11}$$

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

$$H \cdot \bar{\delta} = \bar{\varepsilon} \tag{12}$$

Since $\boldsymbol{H}$ is positive-definite it is also nonsingular. We solve eqn. 12 with a standard LU-decomposition with a backsubstitution algorithm code in C++.

The MREM allows one to include a prior estimation $q(x)$ of the PDF. In the case where the prior distribution is uniform the MREM reduces to the MEM. In the MREM one minimises the discrimination information (also known as the relative entropy, Kulback Leibler distance or the cross entropy),

$$\mathscr{I}(p) = \int_a^b p(x) \ln \frac{p(x)}{q(x)}\, dx \tag{13}$$

subject to the boundary conditions of the available data. Following the same approach as in the preceding Section we obtain the following expression for the inferred PDF p^* in terms of the Lagrange multipliers $\{\lambda_m, m = 1, \ldots, M\}$

$$p^*(x) = \frac{q(x)}{Z} \exp\left(-\sum_{m=1}^{M} \lambda_m x^m\right) \tag{14}$$

with partition function

$$Z = \int_a^b q(x) \exp\left(-\sum_{m=1}^{M} \lambda_m x^m\right) dx \tag{15}$$

The algorithm used to solve for the Lagrange multipliers is virtually identical to that developed for the MEM.

3 Application to digital communication systems

3.1 Binary signalling over a frequency-nonselective slowly fading channel

We first consider a problem which can be solved analytically, namely an AWGN (additive white Gaussian noise) channel with binary phase-shift-keyed (BPSK) signalling subject to slow Rayleigh fading. For BPSK each data bit is transmitted as a bandpass pulse, or as the antipodal version thereof, corresponding to the transmission of a mark or space, respectively. The probability P_e of a binary digit error is given by [23]

$$P_e(\gamma) = \tfrac{1}{2} \operatorname{erfc}(\sqrt{\gamma}) \tag{16}$$

where $\gamma = \alpha^2 \mathscr{E}/N_0$ is the received signal/noise ratio per information bit, erfc is the complementary error function, $\mathscr{E}$ 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 follows [23] that the SNR is exponentially distributed

$$p(\gamma) = \frac{1}{\tilde{\gamma}} \exp(-\gamma/\tilde{\gamma}) \qquad \gamma \geqslant 0 \tag{17}$$

where $\tilde{\gamma}$ is the expectation value of γ

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

The average bit error rate is then obtained via

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

which can be evaluated analytically [23].

Assume now that we had not known that the SNR is exponentially distributed. Instead we only have the first few moments of the PDF (typically four moments are measured experimentally). The moments of the exponential PDF are given by

$$\mu_k = \tilde{\gamma}_b^k k! \tag{20}$$

The maximum entropy method (MEM) gives the exact result with only two moments (μ_0 and μ_1).

Fig. 1 compares the exact result of the average error rate as a function of the signal/noise ratio with the MEM

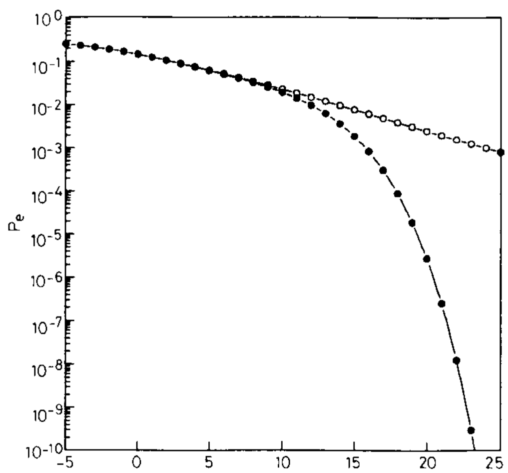

Fig. 1 *Exact average error rate compared to MEM and GQR results*
MEM result coincides with exact results for all SNRs
○ exact curve
● GQR curve

result (using two moments) and the GQR result for 31 moments. For small signal/noise ratios, both the GQR method and the MEM reproduce the exact result, but for high signal/noise ratios the GQR method fails, while the MEM continues to reproduce the exact result.

3.2 Diversity transmission

When the signal is received via multiple paths, the average bit error rate can be reduced by introducing diversity, i.e. by supplying the receiver with several replicas of the same information over independently fading channels. The resulting PDF is given by [23]

$$p(\gamma) = \frac{\gamma^{L-1}}{(L-1)!\tilde{\gamma}^L} \exp(-\gamma/\tilde{\gamma}) \tag{21}$$

where L is the number of paths along which the information is received and γ and $\tilde{\gamma}$ are, as before, the SNR and the average SNR, respectively. The average error rate is given again by eqn. 19 and the analytical result is

$$\tilde{P}_e = \left(\frac{1-\sigma}{2}\right)^L \sum_{l=0}^{L} \binom{L-1+l}{l} \left(\frac{1+\sigma}{2}\right)^l \tag{22}$$

where we have introduced

$$\sigma = \sqrt{\left(\frac{\tilde{\gamma}}{1+\tilde{\gamma}}\right)} \tag{23}$$

Assume again that we only know a few moments of $p(\gamma)$,

$$\mu_m = \frac{(m+L-1)!}{(L-1)!} a^m \tag{24}$$

and consider the case where we transmit over three fading paths ($L = 3$) with an average SNR of ($\tilde{\gamma} = 7$). In this case the GQR method fails completely.

Fig. 2 shows the MEM result given eight moments (curve *a*) compared to the exact result. The inferred PDF

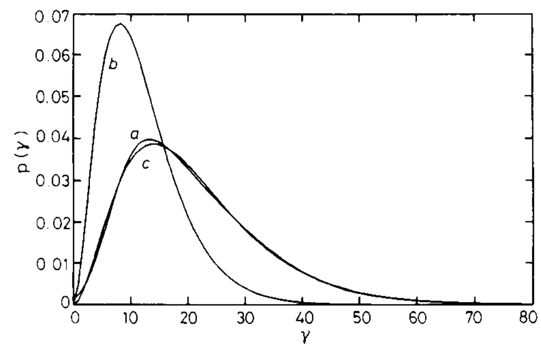

Fig. 2 *PDF for multipath fading*
a MEM result with eight moments
b Prior distribution given by MREM to reproduce exact result with two moments
c Exact result

looks reasonably accurate. However, if we calculate the average bit error rate we find that the MEM result is nearly double the exact result ($\tilde{P}_e^* = 0.00060$ while $\tilde{P}_e = 0.00032$). For more than eight moments our MEM algorithm became numerically unstable because of the exponential increase in the size of the moments ($\mu_0 = 1$ while $\mu_8 = 1.046 \times 10^{13}$). If we had known the functional form of the PDF but not its average SNR $\tilde{\gamma}$, we could use the MREM to improve on the accuracy of the result. Using $p(\gamma)$ with any arbitrary $\tilde{\gamma} > 0$ for the prior PDF, we find that two moments (μ_0 and μ_1) suffice to reproduce the analytically exact result. Curve *b* in Fig. 2 shows the prior PDF ($\tilde{\gamma} = 4$) used by the MREM to infer the exact result ($\tilde{\gamma} = 7$), curve *c*.

Naturally, we do not, in most cases, know the functional form of the PDF. However, one might, for example, have the PDF for a system with diversity for, say, two fading paths (either analytically or numerically) while one wants to obtain the 3-path PDF. One could then make the assumption that one expects the 3-path PDF to be not too difficult from the 2-path result. The 2-path PDF could then be used as a prior PDF for the MREM in order to obtain a more accurate result for the inferred 3-path PDF. Table 1 compares the MEM result with the MREM results given various number of moments. In general the MREM results are much more

Table 1: MEM and MREM inferred average bit error rates for diverse transmission with $L = 3$ and $\bar{\gamma} = 7$

M	MEM		MREM	
	$\tilde{P}_e^*$	N	$\tilde{P}_e^*$	N
1	1.14×10^{-2}	7	1.46×10^{-3}	7
2	3.79×10^{-3}	12	9.11×10^{-4}	7
5	1.20×10^{-3}	14	5.26×10^{-4}	10
8	6.01×10^{-4}	49	4.07×10^{-4}	11
9	No conv.		3.53×10^{-4}	13
Exact result: $\tilde{P}_e = 3.21 \times 10^{-4}$				
GQR result ($M = 31$): $\tilde{P}_e = 4.91 \times 10^{-5}$				

$M + 1$ is the number of moments given and N is the number of iterations required for convergence.

accurate and fewer iterations are needed for convergence. The most accurate result obtained with the MEM is still nearly a factor 2 too large (for more than eight moments the MEM became numerically unstable) while the MREM result was only out by about 27%. Also, the MEM required 49 iterations for convergence, while the MREM converged in only 11 iterations. The most accurate MREM result was obtained with nine moments, where the error in the average bit error rate was only about 10%. The best GQR result is obtained with 31 moments (for more than 31 moments the method becomes numerically unstable), but even for 31 moments the result is still nearly an order of magnitude too small.

3.3 Bounds on error rates of direct-sequence CDMA systems

In a following paper [24] we show how the maximum entropy based methods can be used to obtain the inter-user interference (IUI) PDF for spread-spectrum multiple-access (SSMA) systems and how one can quantify the Gaussian assumption. Here we show only that in a case where there are very tight bounds on the average error rates available [25], both the maximum entropy and the Gauss-quadrature method yield results which are either within or very close to the bounds. This can be seen from Table 2, which compares the upper and lower

Table 2: Average error rates for $K = 2$ users and chip rate $N_c = 31$

E_b/N_0	Lower bound	Upper bound	GQR	MaxEnt
4	1.44×10^{-2}	1.45×10^{-2}	1.442×10^{-2}	1.442×10^{-2}
6	3.35×10^{-3}	3.36×10^{-3}	3.346×10^{-3}	3.345×10^{-3}
8	4.42×10^{-4}	4.24×10^{-4}	4.216×10^{-4}	4.216×10^{-4}
10	2.40×10^{-5}	2.42×10^{-5}	2.399×10^{-5}	2.399×10^{-5}
12	4.81×10^{-7}	4.89×10^{-7}	4.816×10^{-7}	4.816×10^{-7}
14	2.28×10^{-9}	2.34×10^{-9}	2.285×10^{-9}	2.285×10^{-9}

bounds for the error rates of a 2-user SSMA system with chip rate of 31 with the average error rates obtained using the GQR and the MEM, respectively. The MEM, however, required only ten moments to achieve this accuracy, while the GQR method required 23 moments.

4 Conclusions

The maximum entropy based methods are computationally more demanding than the GQR-based methods. The GQR methods require that one Cholesky decomposes an $M \times M$ matrix (M is the number of moments) and then that one finds the eigenvalues and eigenvectors of a tridiagonal M-dimensional matrix. In the case of the maximum entropy methods basically all the computational effort is concentrated in solving a set of M coupled nonlinear equations for the Lagrange multipliers.

The increase computational requirements can, however, be justified as follows. First, the GQR-based methods tend to fail at high signal/noise ratios while the maximum entropy based methods continue to give reliable results. Secondly, the maximum entropy methods generally require considerably fewer moments to yield accurate results. This is especially useful when the moments are obtained experimentally — usually only four moments can be measured accurately.

Furthermore, information other than that supplied by the moments can be incorporated into a maximum entropy formalism. This is not possible for the GQR-based methods. Such information could be in the form of a prior estimate for the PDF, in which case one could use the MREM to improve on the accuracy of the results.

Finally, via the maximum entropy based methods one obtains a functional form for the PDF irrespectively of the number of moments supplied. GQR-based methods, on the other hand, require an extremely high number of moments in order to calculate the PDF on a relatively fine grid. This is often not possible due to numerical instabilities.

5 References

1 DOGLIOTTI, R., LUVISON, A., and PIRANI, G.: 'Error probability in optical fiber transmission systems', *IEEE Trans.*, 1979, **IT-25**, pp. 170–178

2 LAFORGIA, D., LUVISON, A., and ZINGARELLI, V.: 'Bit error rate evaluation for spread spectrum multiple-access systems', *IEEE Trans.*, 1984, **COM-32**, pp. 660–666

3 COOPER, G.R., and NETTLETON, R.W.: 'A spread-spectrum technique for high-capacity mobile communication', *IEEE Trans.*, 1978, **VT-27**, pp. 264–275

4 BENEDETTO, S., BIGLIERI, E., LUVISON, A., and ZINGARELLI, V.: 'Moment-based performance evaluation of digital transmission systems', *IEE Proc. I*, 1992, **139**, pp. 258–266

5 BENEDETTO, S., DE VINCENTIIS, G., and LUVISON, A.: 'Application of Gauss quadrature rules to digital communication problems', *IEEE Trans.*, 1973, **COM-21**, pp. 1159–1165

6 MEYERS, M.H.: 'Computing the distribution of a random variable via Gauss quadrature rules', *Bell Syst. Techn. J.*, 1982, **61**, pp. 2245–2261

7 GOLUB, G.H., and WELSCH, J.H.: 'Calculation of Gauss quadrature rules', *Math. Comput.*, 1969, **26**, pp. 221–230

8 GAUTSCHI, W.: 'On the construction of Gaussian quadrature rules from modified moments', *Math. Comput.*, 1970, **24**, pp. 245–260

9 JAYNES, E.T.: 'Information theory and statistical mechanics I & II', *Phys. Rev.*, 1957, **106**, pp. 620–630; **108**, p. 171

10 KAVEHRAD, M., and JOSEPH, M.: 'Maximum entropy and the method of moments in performance evaluation of digital communication systems', *IEEE Trans.*, 1986, **COM-34**, pp. 1183–1189

11 SHORE, J.E., and JOHNSON, R.W.: 'Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy', *IEEE Trans.*, 1980, **IT-26**, pp. 26–37

12 KULLBACK, S.: 'Information theory and statistics' (Wiley, New York, 1959)

13 ZHUANG, X., HARALICK, R.M., and ZHAO, Y.: 'Maximum entropy and image reconstruction', *IEEE Trans.*, 1991, **SP-39**, pp. 1478–1480

14 MALAZA, E.D., RITCHIE, R.A., SOLMS, F., VON OERTZEN, D.W., and MILLER, H.G.: 'Predictions of LEP hadronic multiplicity distributions using the maximum entropy principle', *Phys. Lett. B*, 1991, **266**, pp. 169–172

15 PEREZ, S.M., QUICK, R.M., DAVIDSON, N.J., MILLER, H.G., and SOLMS, F.: 'Reconstruction of the spin dependence of one nucleon transfer spectroscopic sums from incomplete information', *Phys. Rev. C*, 1992, **45**, pp. 870–872

16 STEEB, W.H., SOLMS, F., and STOOP, R.: 'Chaotic systems and the maximum entropy principle', *J. Phys. A*, 1994, **27**, pp. L399–L402

17 ZHUANG, X., CHEN, L., and CHEN, S.: 'An experimental comparison between maximum entropy and minimum relative-entropy spectral analysis', *IEEE Trans.*, 1993, **SP-41**, pp. 1730–1734

18 SHORE, J.E.: 'Minimum cross-entropy spectral analysis', *IEEE Trans.*, 1981, **ASSP-29**, pp. 230–237
19 TAGLIANI, A.: 'On the application of maximum entropy to the moments problem', *J. Math. Phys.*, 1993, **34**, pp. 326–337
20 SHANNON, C.E.: 'A mathematical theory of communications', *Bell Syst. Tech. J.*, 1948, **27**, pp. 379, 623
21 ALHASSID, Y., AGMON, N., and LEVINE, R.D.: 'An upper bound for the entropy and its applications to the maximal entropy problem', *Chem. Phys. Lett.*, 1972, **53**, pp. 22–26
22 MEAD, L.R., and PAPANICOLAOU, N.: 'Maximum entropy in the problem of moments', *J. Math. Phys.*, 1984, **28**, pp. 2404–2417
23 PROAKIS, J.G.: 'Digital communications' (McGraw-Hill, 1983)
24 SOLMS, F., VAN ROOYEN, P.G.W., and KUNICKI, J.S.: 'Maximum entropy performance analysis of spread-spectrum multiple-access communication systems'. Proceedings of 14th International Workshop on Maximum Entropy and Bayesian Methods, Cambridge, 1994 (Kluwer Academic Press), to be published
25 PURSLEY, M., SERVAT, D., and STARK, W.: 'Error probability for DS-SSMA communications, Part II: Upper and lower bounds', *IEEE Trans.*, 1982, **COM-30**, pp. 975–984