Low-Complexity Approximations to Maximum Likelihood MPSK Modulation Classification

Jon Hamkins*
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109–8099
Email: jon.hamkins@jpl.nasa.gov

***Abstract*— We present a new approximation to the maximum likelihood (ML) classifier to discriminate between M-ary and M'-ary phase-shift-keying (PSK) transmitted on an additive white Gaussian noise (AWGN) channel and received noncoherently, partially coherently, or coherently. When 100 symbols are observed to discriminate between binary PSK (BPSK) and quadrature PSK (QPSK), the misclassification probability of the proposed classifier is indistinguishable from that of the ML classifier for any symbol signal-to-noise ratio (SNR) in the range from -10 dB to 10 dB. A software implementation on a desktop computer requires about 1 millisecond to classify the signal, and a tunable parameter can be used to reduce the complexity further, with a graceful degradation in performance. We also present a new threshold optimization technique that improves several other previously published modulation classifiers. Finally, we derive a lower bound misclassification probability as a function of the number of observed samples, which is an inherent limitation classifiers face even on high SNR channels.**

I. Introduction

We analyze the performance of optimum and sub-optimum modulation classifiers to discriminate M-PSK from M'-PSK. Modulation classification has long been an important component of noncooperative communications in which a listener desires to intercept an unknown signal from an adversary. It is becoming increasingly important in cooperative communications as well, with the advent of the software-defined autonomous radio. Such a radio may know little a priori about an incoming signal but may be required to correctly classify its data rate, modulation type, and forward-error correction code before properly configuring itself into a conventional radio that acquires and tracks the symbol timing, carrier frequency and phase, and ultimately produces decoded bits. This type of radio may be helpful in a Mars orbiter, for example, if it must communicate to multiple landed assets on Mars as it orbits and cannot wait for an uplink command from Earth to reconfigure it for each pass.

II. Preliminaries

A. Signal Model

For ease of exposition, this paper is limited to binary hypothesis testing, although the extension to multiple hypotheses can be done in the usual way [1]. Throughout, we assume $M < M'$ and each is a power of two. Each modulation type is assumed a priori to occur with probability one half and remains the same for N received symbols. Timing is assumed to be known, and the complex baseband representation of each received symbol at the output of a matched filter has the form

$$r_n = Ae^{j(\theta_n+\theta_c)} + n_n, \quad k = 1, \ldots, N \tag{1}$$

where A is a real constant, θ_n is the data modulation for the kth symbol, θ_c is the carrier phase, and $n_n = n_{n,R} + jn_{n,I}$ is a complex Gaussian with mean zero and variance σ^2 in each dimension. For BPSK ($M = 2$), $\theta_n \in \{0, \pi\}$, while for QPSK, $\theta_n \in \{\pi/4, 3\pi/4, 5\pi/4, 7\pi/4\}$, and in either case the θ_n's are assumed to be independent and uniformly distributed from symbol to symbol. Initially, we assume noncoherent reception, in which θ_c is unknown, uniform on $[0, 2\pi)$, and constant over all observed symbols. We also assume that A and σ^2 are known. Later, we will extend the discussion to coherent and partially coherent reception in which something about θ_c is known, and to cases in which the symbol SNR $E_s/N_0 = A^2/(2\sigma^2)$ is known but not A and σ^2 individually.

B. ML modulation classification

Since the modulation types occur with equal a priori probability, the ML classification rule results in the minimum probability of misclassification. It can be implemented by comparing the likelihood ratio (LR) of the N observed samples to a threshold of one. The conditional likelihood function (CLF) given θ_c for M-ary PSK, is [2, Eq. B.3b]

$$CLF_M(\theta_c) = \exp\left[\frac{-NE_s}{N_0} + \sum_{n=1}^{N} \ln\left(\frac{2}{M}\sum_{q=0}^{\frac{M}{2}-1} \cosh[x_n(q;\theta_c)]\right)\right] \tag{2}$$

where $x_n(q;\theta_c) = \frac{E_s}{AN_0}\text{Re}[r_n e^{-j(\theta_c+q\pi/2)}]$. We may rewrite (2) as

$$CLF_M(\theta_c) = e^{-\frac{NE_s}{N_0}}\left(\frac{2}{M}\right)^N \prod_{n=1}^{N}\sum_{q=0}^{\frac{M}{2}-1} \cosh[x_n(q;\theta_c)]. \tag{3}$$

The LR for the M and M' hypotheses is given by averaging the CLF's over θ_c, which is uniform over $(0, 2\pi)$, and forming the ratio:

$$\text{LR} = \frac{LF_M}{LF_{M'}} = \frac{E_{\theta_c}\{CLF_M(\theta_c)\}}{E_{\theta_c}\{CLF_{M'}(\theta_c)\}} \tag{4}$$

To compute the LR, E_s/N_0 is not sufficient; A and σ^2 must be known individually (to compute $x_n(\cdot;\cdot)$).

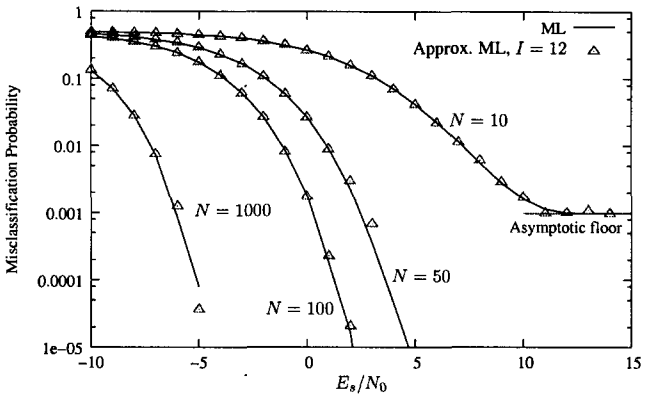

Fig. 1. The approximate ML classifier is nearly indistinguishable from the ML classifier when $I = 12$.

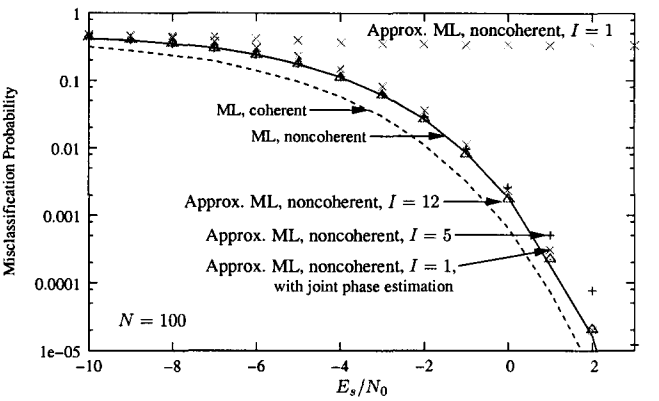

Fig. 2. The approximate ML classifier is near-optimal for small values of I.

C. qLLR approximation to log-LR

At low SNR, one may use the approximations $\cosh(x) \approx 1+x^2$ and $\ln(1+x) \approx x$ in (4), along with the approximation $I_0(x) \approx x$ for large post-detection SNR, which leads to the quasi-LLR or qLLR [2], [3] approximation to the log-LR (LLR):

$$\text{qLLR} = \left\| \sum_{k=1}^{N} r_k^M \right\| \tag{5}$$

where $\|x\|$ denotes the magnitude of the complex value x. In principle, the qLLR metric should be compared to threshold zero, since it is an approximation to the log-LR. We defer a discussion of thresholds to section V.

III. New Approximations to the LR

A. Integral approximation to the LR

For random realizations of $\mathbf{r}$, $CLF_M(\theta_c)$ and $CLF_{M'}(\theta_c)$ frequently differ by more than an order of magnitude. If one of the CLF's is higher that the other for the entire range $\theta_c \in (0, 2\pi)$, then a single test comparing $CLF_M(\theta_c)$ to $CLF_{M'}(\theta_c)$ at any θ_c would produce the correct classification. While this property does not hold with high enough probability to produce near-ML performance, this approach can be refined in a parameterized way, as follows.

One can see from (3) that the period of $CLF_M(\theta_c)$ is $2\pi/M$. Thus, we can write the likelihood function for hypothesis $H \in \{M, M'\}$ as

$$LF_H = \frac{M}{2\pi} \int_0^{\frac{2\pi}{M}} CLF_H(\theta_c) d\theta_c \approx \frac{1}{I} \sum_{i=1}^{I} CLF_H\left(\frac{i\pi}{I}\right) \tag{6}$$

where the last equation becomes equality as $I \to \infty$ by the rectangular rule for integration. The proposed classifier is to substitute (6) into (4), and let I be a tunable parameter.

Fig. 1 illustrates the performance of this classifier compared to ML classification, for the case of noncoherent reception of BPSK/QPSK, with $I = 12$. The approximate method has performance indistinguishable from ML classification.

As can be seen, small values of I in (6) results in excellent classification performance, even though the computation is not necessarily an accurate approximation of LF_H.

B. Hypothesis-dependent carrier-aided LR approximation

When $I = 1$ in the method above, the complexity is low, but the performance is poor (see Fig. 2). However, we can use $I = 1$ and still get near-ML performance by using $\mathbf{r}$ to estimate θ_c in the H-PSK hypothesis by using the Hth power to remove the modulation, averaging over many symbols, and then normalizing the resulting angle:

$$H\text{-ary PSK hypothesis:} \quad \hat{\theta}_c^{(H)} = \frac{1}{H} \angle \sum_{n=1}^{N} r^H \tag{7}$$

When $I = 1$, $CLF_M(\hat{\theta}_c^{(M)})/CLF_{M'}(\hat{\theta}_c^{(M')})$ is a good approximation to (4), as it is an approximation to coherent-reception classification. The carrier estimates in the two hypotheses are used with the same basic reasoning as in per-survivor processing (PSP), the method used to perform joint channel estimation and decoding. The phase ambiguities present in these estimates of θ_c, namely, a $2\pi/M$ ($2\pi/M'$) radian ambiguity for the M-PSK (M'-PSK) hypothesis, do not cause a problem because, as previously mentioned, the CLF's are also periodic with period $2\pi/M$ ($2\pi/M$).

The performance of the joint phase-estimater/classifier for $I = 1$ is shown in Fig. 2. At high SNR, the performance is indistinguishable from the ML noncoherent classifier, while at low SNR, the degraded phase estimation results in only slightly suboptimal performance.

C. Normalized qLLR (nqLLR)

We also propose a lower complexity alternative to the above classifiers. It is simply a normalization of the qLLR classification metric, which we refer to as the nqLLR metric:

$$\text{nqLLR} = \frac{\left\| \sum_{k=1}^{N} r_k^M \right\|}{\sum_{k=1}^{N} \|r_k\|^M} \tag{8}$$

This metric was originally devised as a coarse way to implement the qLLR without needing knowledge of A. Since

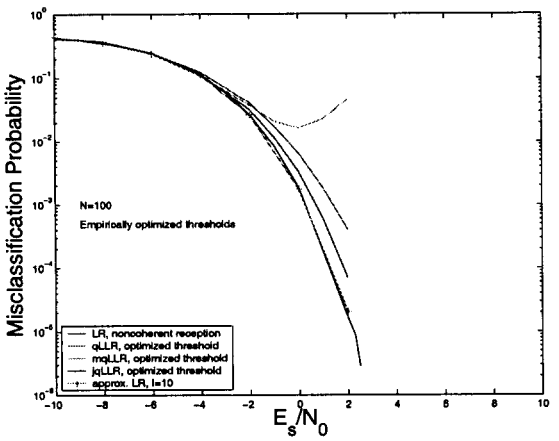

Fig. 3. Performance of the classifier using (6) with $I = 12$, compared to ML, qLLR, nqLLR, and mqLLR noncoherent classifiers, for $N = 100$.

$\sum_{k=1}^{N} \|r_k\|^M$ is a rough estimate of NA^M, nqLLR may be compared to 0.5, instead of comparing qLLR to $\frac{NA^M}{2}$, for example.

Serendipitously, the normalization has a beneficial effect in the classification performance itself. Fig. 1 shows the performance of nqLLR compared to ML, the approximate-ML classifier discussed above, and qLLR with previously suggested thresholds [3].

IV. Coherent/Partially-Coherent Classification

In coherent reception, the carrier phase θ_c is known, and it can be used instead of the estimate in (7). Thus, the expectation in (4) degenerates to an evaluation of the CLF at a single point:

$$\begin{aligned} \mathrm{LR} &= \frac{CLF_2(\theta_c)}{CLF_4(\theta_c)} \\ &= \frac{\prod_{n=1}^{N} \cosh[x_n(0;\theta_c)]}{2^{-N}\prod_{n=1}^{N}(\cosh[x_n(0;\theta_c)] + \cosh[x_n(1;\theta_c)])} \\ &= 2^N \left[\prod_{n=1}^{N}\left(1 + \frac{\cosh[x_n(1;\theta_c)]}{\cosh[x_n(0;\theta_c)]}\right)\right]^{-1} \end{aligned} \tag{9}$$

This is the optimal statistic for ML classification with coherent reception, with performance shown in Fig. 2.

If something is known about the carrier phase, but the randomness is not completely removed, we may account for it in the classifier by using a distribution on θ_c that is different from the uniform distribution assumed in noncoherent reception. We refer to this as partially coherent reception.[1] For example, a phase-locked loop that tracks a residual carrier may produce an error in its estimate that is Tikhonov distributed. Other randomness, due to oscillator phase noise, inter-symbol interference, or phase ambiguities, for example, may also introduce randomness to an estimate of θ_c.

[1]Note, by partially coherent reception we do not mean differentially coherent reception, which also is sometimes referred to by the same term. Rather, we mean that the distribution on θ_c is something other than uniform.

For partially coherent reception, we may use (4), replacing the uniform probability distribution for θ_c with the appropriate distribution. As before, we do not have to compute the numerical integral precisely, but simply sample the CLF functions at a few phase values chosen in accordance with the probability density function for θ_c.

V. Threshold Optimization

A. Suboptimality of previously derived thresholds

A critical limitation of the qLLR metric is that it does not approximate the LLR metric precisely enough for the optimal LLR threshold (zero) to be used with success. Indeed, the qLLR metric is always nonnegative, meaning that with a zero threshold (optimal for LLR), it would produce the same decision regardless of the transmitted modulation.

One can optimize the threshold for the qLLR metric itself, instead of using the optimal threshold (zero) for the LLR metric. This was attempted analytically in [3], using the assumptions that the the real and imaginary parts of the sum in (5) are jointly Gaussian and have equal variance under the two hypotheses. The assumptions lead to the "optimum" threshold (for BPSK/QPSK) [3]:[2]

$$T = A^2 V_2 I_0^{-1}\left(\exp\left[\frac{N}{2V_2}\right]\right) \tag{10}$$

where $V_M = \sum_{l=0}^{M} \frac{(M!)^2}{2l![(M-l)!]^2} \cdot \frac{E_s}{N_0}^{-l}$. However, the variances are not equal, especially at high SNR. For example, in BPSK/QPSK classification, the variance the real part of the sum in (5) is A^2V_2 when BPSK is sent and $A^2(V_2+1)$ when QPSK is sent [3, Eq. (A.12)]. At high SNR, $V_2 \approx \frac{1}{2}$, and thus the variance is nearly three times higher when QPSK is sent. A similar comparison shows the variance of the imaginary part is nearly twice as high when QPSK is sent. Therefore, at low SNR, we expect that the threshold in (10) may be near-optimal, but at high SNR it will be too low. This is confirmed by our numerical results.

A further approximation $I_0(x) \approx e^x$ valid for large x (high post-detection SNR) leads to the approximate threshold [3]

$$T = \frac{NA^2}{2} \tag{11}$$

B. Empirical threshold optimization

An alternative to the analytical derivation of an appropriate thresholds, which in the case of both (10) and (11) involved some approximation, we may optimize the threshold for the metric empirically, using the following process:

1) Generate a large number of received samples according to (1) for each of M-PSK and M'-PSK modulation types.
2) Group the samples into blocks of length N, and compute a set of sample classifier metrics.

[2]The thresholds (10) and (11) differ from the original presentation [3], to account for a difference in scale of the received symbols. The presentation in [3] scales the received signal so that the noise variance is unity, but this presumes that the noise variance is known at the receiver, an assumption we prefer to avoid unless the classifier requires it.

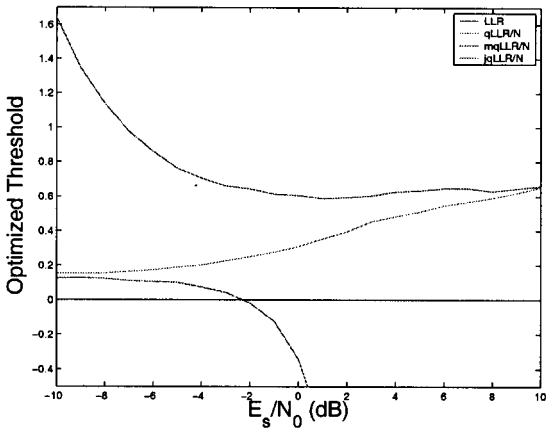

Fig. 4. Empirically optimized thresholds for qLLR, mqLLR, and nqLLR metrics when $N = 100$, compared to the optimal zero-threshold for LLR.

3) Sort the metrics for M-PSK and M'-PSK received symbols separately, in increasing order.
4) For each observed metric α, in increasing order;
   a) Count the number of observed M-PSK metrics less than α
   b) Count the number of observed M'-PSK metrics greater than α
   c) Compute the probability of misclassification when using threshold α, using the sum of the above quantities
5) Report the threshold that minimizes the probability of misclassification.

This procedure may be implemented efficiently enough that a desktop machine can generate about 100 million samples and determine an empirically optimal threshold in less than an hour. Fig. 4 shows the empirically optimized thresholds for the qLLR and nqLLR metrics for $N = 100$, based on 100,000 classifier metrics (10 million received symbols).

Notes:

- Threshold optimization need be performed only once, ahead of time. In a practical implementation, given N and E_s/N_0 a table look-up or function-fit of the appropriate curve in Fig. 4 may be used to determine the threshold.
- The optimal threshold for the LLR metric is zero, and requires no empirical optimization.
- The optimal threshold for the qLLR metric is relatively flat over a broad region of E_s/N_0, being near 0.6 for $E_s/N_0 > -5$ dB. This is a desirable characteristic, because E_s/N_0 may not be known exactly.
- For E_s/N_0 above approximately 2 dB, the classifier performance is so good that many different thresholds could be used with excellent performance. The truly optimal qLLR threshold is likely monotonically decreasing in E_s/N_0.
- The optimal threshold for nqLLR is also quite flat, and ranges only from about 0.15 to 0.6 for $-10 < E_s/N_0 < 10$. At high SNR, it is approximately equal to the optimal qLLR$/(NA^2)$ threshold, because the qLLR$/(NA^2)$ metric and nqLLR metric are nearly identical in that region.

VI. Comparisons and Numerical results

The performance of this classifier, using a threshold of one, is shown in Fig. 2. At high SNR, it is virtually in agreement with the ML performance, because the estimates of θ_c are quite good. At low SNR, the performance degrades slightly. A software implementation required 0.25 milliseconds to perform a classification.

The approximate ML metric may be computed in $O(IN)$ time, because it is the sum of I terms, each of which is a product of N terms, each of which takes $O(1)$ time to compute. Since the integration range is partitioned in the same way under each hypothesis, each hyperbolic cosine evaluation used in the numerator of (4) is also used in the denominator, and thus, only M' hyperbolic cosine evaluations are needed for each (i, n) pair. Also, we may write

$$x_n\left(q; \frac{i\pi}{I}\right) = \mathrm{Re}[r_n]\cos(\alpha(q,i)) - \mathrm{Im}[r_n]\sin(\alpha(q,i)),$$

where $\alpha(q,i) = \pi(i/I + q/2)$. The trigonometric terms for all $\alpha(\cdot,\cdot)$ may be precomputed and stored in a table, since they do not depend on $\mathbf{r}$. Thus, only addition, multiplication, and two hyperbolic cosine evaluations are required for each (i, n) pair.

For the example in which $I = 12$ and $N = 100$, an implementation in C on a linux desktop computer required approximately 1 millisecond to compute the approximate ML classification metric. When $I = 5$, speed approximately doubled, and performance degraded by about 0.5 dB.

. It's performance is illustrated in Fig. 2. There is about a 1 dB difference between coherent and noncoherent ML classification at low SNR, and no difference at high SNR. A software implementation of the coherent classifier required about 0.1 milliseconds.

The approximations used in (11) result in about a 1 dB loss compared to using threshold (10) Note that the performance with threshold (10) begins to deteriorate at $E_s/N_0 > 2$ dB, compared to the ML classifier, as expected because of the suboptimal threshold noted above.

A. Classification performance

Fig. 3 shows the performance of the LLR, qLLR, mqLLR, and nqLLR classifiers using empirically optimized thresholds. In each case, the probability of misclassification is determined from a separate simulation than the one in which the optimal thresholds were determined (i.e., there is no inappropriate reliance on training data in generating the error rates).

As can be seen, when the thresholds are empirically optimized, the qLLR, mqLLR, and nqLRR decision rules perform near the ML-limit for $E_s/N_0 < -4$ dB. As E_s/N_0 increases the decision rules begin to perform differently. The mqLLR metric diverges and is not useful for $E_s/N_0 > 0$ dB. The qLLR metric remains within about 1 dB of the ML limit, and the nqLLR metric remains within about 0.5 of the ML limit throughout the entire range of SNR's.

B. Complexity

It is suggested in [2], [3] that the ML classifier is not practical. However, if computation of $E_{\theta_c}\{CLF(\theta_c)\}$ requires only a constant number of evaluations of $CLF(\theta_c)$, then (4) can be computed in $O(N)$ time. This is because, under either hypothesis (M-ary PSK or M'-ary PSK), the $CLF(\theta_c)$ is product of N items each of which can be computed in $O(1)$ time. In simulations with $N = 100$, $M = 2$, and $M' = 4$, it was found that 100 to 150 CLF evaluations were needed to compute the expectation, or about 100,000 hyperbolic cosine function evaluations overall.

Unfortunately, it is not true that a straightforward computation of $E_{\theta_c}\{CLF(\theta_c)\}$ can be computed in $O(N)$ time, because the $CLF(\theta_c)$ has extrema that are exponentially related to N, and thus the numerical computation of the integral requires finer partitions for higher values of N. This extrema property can be seen as follows. By elementary calculus, the first derivative of $CLF(\theta_c)$ is

$$CLF'(\theta_c) = \sum_{l=1}^{N} \sinh\left[\frac{A}{\sigma^2}\mathrm{Re}(r_l e^{-j\theta_c})\right] \frac{A}{\sigma^2}\mathrm{Im}(r_l e^{j\theta_c})] \prod_{n=1,k\neq l}^{N} \cosh\left[\frac{A}{\sigma^2}\mathrm{Re}(r_n e^{-j\theta_c})\right]. \quad (12)$$

This is can be quite large. For example, if $\theta_c = 0$, the SNR is large, and $\mathbf{r} \approx A(1,\ldots,1)$, then

$$|CLF'(\pi/4)| \approx \frac{NA^2}{\sqrt{2}\sigma^2} \exp\left[\frac{NA^2}{\sqrt{2}\sigma^2}\right]$$

Since $\max|CLF'(\theta_c)| = \Omega(e^N)$, no numerical integration which partitions the domain of the integral into a number of intervals independent of N will produce a correct calculation as N increases. That is, for higher values of N a finer partition of the domain is required and thus an increasing number of evaluations of the CLF. Therefore, computing the LR in this way requires more than $O(N)$ time.

C. Complexity

Nearly all of the classifiers considered here have computational complexity $O(N)$. Linear complexity does not mean it is practical, however. If the ML rule could be implemented, there would be no need for any other decision rule. However, computing the LR as given in (4) requires two numerical integrations of a function that to evaluate one time requires N or $2N$ evaluations of cosh(), N evaluations of ln(), and one evaluation of exp(). This is linear in N, but perhaps still more complicated than a real-time autonomous radio could perform. To reduce the computation, one can rewrite (4) as

$$\mathrm{LLR} = \ln\left[\frac{E_{\theta_c}\left\{\prod_{k=1}^{N}\cosh[x_k(0;\theta_c)]\right\}}{E_{\theta_c}\left\{\prod_{k=1}^{N}\left(\frac{1}{2}\cosh[x_k(0;\theta_c)]\right)\right\}}\right] \quad (13)$$

Assuming a the expectation is accomplished with a numerical integration requiring 100 evaluations of the function, a single LLR metric requires 900N exponential, logarithmic, and trigonometric function evaluations.

The qLLR and nqLLR do not require any exponential, logarithmic, and trigonometric function evaluations. They require only multiplication, division, and magnitude operations.

D. Required Parameter estimation

All other things equal, we would prefer a classifier that requires the least knowledge regarding the signal parameters. The LLR metric requires knowledge of E_s/N_0 to compute, for example, but not E_s or N_0 individually. On the other hand, the qLLR metric with threshold given by (10) requires both E_s/N_0 and E_s. The approximate threshold of (11) performs worse but does not require knowledge of E_s. The mqLLR metric also requires both parameters.

The proposed nqLLR metric requires only E_s/N_0. Furthermore, among all classifiers considered, it is unique in providing acceptable performance with knowledge of neither E_s/N_0 nor E_s. Computing the metric requires neither, and if we set a fixed threshold of 0.4, performance is about within 1 dB at $E_s/N_0 = 1$, and within 3.5 dB at very low E_s/N_0.

VII. A Lower Bound on Misclassification Probability

Theorem 1: If N independent complex baseband symbols from an a priori equiprobable BPSK or QPSK signal are observed at the output of a noncoherent noiseless channel, the minimum probability of modulation misclassification is 2^{-N}.

Proof. r_n is uniformly distributed on the circle of radius A, regardless of θ_n. This is a simple consequence of the fact that θ_c is uniformly distributed. Thus, when $N = 1$ the observed signals for BPSK and QPSK are identically distributed and the minimum misclassification probability is 1/2.

Now suppose $N > 1$. Let $B = (b_1,\ldots,b_N)$, where $b_n = \theta_n + \theta_c$. From $(r_1,\ldots,r_N)$, we may ascertain A and B, and vice versa. Since A does not depend on the modulation type, B is a sufficient statistic for the optimal (minimum probability of misclassification) classifier. If there exist i and j such that $b_i = b_j + \pi/2 \mod 2\pi$, then the signal cannot be BPSK, and the optimal optimal classifier decides that QPSK was sent, with no probability of error. Otherwise, for all $n = 1,\ldots,N$, either $b_n = b_1$ or $b_n = b_1 + \pi \mod 2\pi$. Under the BPSK hypothesis, $P(b_n = b_1 \text{ or } b_n = b_1+\pi \mod 2\pi|BPSK) = 1$. Under the QPSK hypothesis, for $n > 1$, $P(b_n = b_1 \text{ or } b_n = b_1 + \pi \mod 2\pi|QPSK) = 1/2$, since b_n takes on the four values in $\{b_1, b_1+\pi/2 \mod 2\pi, b_1+\pi \mod 2\pi, b_1+3\pi/2 \mod 2\pi\}$ with equal probability. Thus, by the independence of the modulation symbols, Thus, for all $N > 1$, this BPSK-consistent event more probably arises from BPSK than from QPSK and the optimal classifier decides BPSK. In summary, the optimal classifier is incorrect with probability

$$P_e = P(QPSK)P_{e|QPSK} + P(BPSK)P_{e|BPSK} = \frac{1}{2}\cdot 2^{-(N-1)} + \frac{1}{2}\cdot 0 = 2^{-N}.\square \quad (14)$$

Corollary 1: Achieving probability of misclassification below 10^{-6} requires $N \geq 20$ observed samples.

Proof. $2^{-20} \approx 9.54 \times 10^{-7}$. $\square$

This can be generalized as follows.

Theorem 2: If N independent complex baseband symbols from an a priori equiprobable M-PSK or M'-PSK signal, $M < M'$, are observed at the output of a noncoherent noiseless channel, the minimum probability of modulation misclassification is $\frac{1}{2}(\frac{M}{M'})^{-(N-1)}$.

Proof. Using a similar argument as above,

$$P(b_n = b_1 + \frac{2k\pi}{M} \text{ for some integer } k_n | M') = \frac{M}{M'}$$

If this M-PSK-consistent event doesn't occur, the modulation is M'-PSK with certainty. Thus,

$$\begin{aligned} P_e &= P(M' - PSK)P_{e|M'-PSK} + P(M - PSK)P_{e|M-PSK} \\ &= \frac{1}{2} \cdot (\frac{M}{M'})^{-(N-1)} + \frac{1}{2} \cdot 0 \\ &= \frac{1}{2}(\frac{M}{M'})^{-(N-1)}. \square \end{aligned}$$

VIII. Conclusions

For the classifiers considered in this memo, Fig. 2 summarizes relevant attributes: required parameter knowledge, closeness to ML-performance, and complexity. Based on the complexity and performance given here, the nqLLR metric is seen to be a near-optimal, low-complexity modulation classifier appropriate requiring only the SNR but not the signal power.

Acknowledgments

The research described in this publication was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

The author would like to thank Kenneth Andrews, Dariush Divsalar, Sam Dolinar, Fabrizio Pollara, Marvin Simon, and Victor Vilnrotter for their many helpful discussions on this topic.

References

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

[2] A. Polydoros and K. Kim, "On the detection and classification of quadrature digital modulations in broad-band noise," *IEEE Trans. Commun.*, vol. 38, no. 8, pp. 1199–1211, Aug. 1990.

[3] C.-Y. Huang and A. Polydoros, "Likelihood methods for MPSK modulation classification," *IEEE Trans. Commun.*, vol. 43, no. 2/3/4, pp. 1493–1504, Feb/Mar/Apr 1995.

TABLE I

A SUMMARY COMPARISON OF CLASSIFIERS.

Classifier	Eq.	Thresh.	Needed Parameters		dB Gap to ML at $E_s/N_0 =$			Complexity*
			E_s/N_0	E_s	-5	-2	1	
ML	(4)	1	√	√	0	0	0	$\sim 900N$
qLLR	(5)	(10)	√	√	0.3	0.5	2.2	
		(11)		√	1.3	1.4	2.6	
		Emp.	√		?	?	?	
nqLLR	(8)	Emp.	√		0.0	0.2	0.5	0
		0.4			3.5	2.0	~ 1	

* Number of exponential, logarithmic, trigonometric, or Bessel function evaluations.