Combined Likelihood Power Estimation and Multiple Hypothesis Modulation Classification

Keith M. Chugg
Dept. of Electrical & Computer Engineering
The University of Arizona
Tucson, AZ 85718

Chu-Sieng Long and Andreas Polydoros
Communication Sciences Institute
University of Southern California
Los Angeles, CA 90089-2565

Abstract

Previously developed techniques for maximum likelihood (ML) modulation classification have assumed that there are only two possible modulation formats and that both the signal and noise powers are known. In this paper we introduce ML-based techniques for performing autonomous power estimation of a phase-shift-keyed signal and additive white Gaussian noise, and for classifying between OQPSK, BPSK, and QPSK formats. The performance of the ML power estimator is shown to be superior to existing techniques and the false classification rate of the simple, two-stage OQPSK/BPSK/QPSK classification rule is shown to be close to that of the globally optimal classifier. A fully autonomous OQPSK/BPSK/QPSK classifier is demonstrated by combining the two-stage rule, the ML power estimator, and previously developed threshold-setting techniques.

1 Introduction and Background

The problem of classifying between various digital modulation formats has obvious relevance to military communication systems. Identification of the modulation format may be necessary to optimize surveillance or jamming strategies, or may allow the transmitting party to be identified. Such modulation classification (MC) problems may also have commercial applications. For example, in order for a universal modem to monitor an ongoing broadcast message, it may first be necessary to determine the format and parameters of the transmission.

Many of the techniques for performing MC which have been suggested in the literature are based on pattern recognition theory. These techniques typically involve the identification of relevant features of the modulation format, and the design of test statistics which characterize these features. One example feature is the number and magnitude of the phase transitions. Test statistics usually involve constructing an estimate of the spectral density of the received signal or a phase histogram (e.g., [1, 2]). A new approach to the MC problem, based on maximum likelihood (ML) decision theory, was recently proposed and successfully applied [3, 4]. This technique is an attempt to formalize the process of identifying the relevant features of a modulation format, which is heuristic in the previous approaches. Ideally, the ML approach to MC provides a lower bound on the achievable probability of false classification (P_F).

Approximations are required to develop practical quasi-ML MC rules. One result of these approximations is the need to implement tests with non-zero decision thresholds, which, in turn, requires accurate signal and noise power estimates. In the remainder of this section we describe the ML approach and the associated approximations. Previously, the ML approach was applied to “binary” MC problems (i.e., MC problems where the modulation format was one of two possible formats). The difficulties encountered while attempting to extend practical ML approaches to the, often times more realistic, problem of multiple hypothesis MC (MMC) are discussed in this section.

A practical approach to designing ML-inspired MMC tests is described in this paper. As a specific example, the problem of classifying between Binary phase Shift Keying (BPSK), Quadrature PSK (QPSK), and Offset QPSK (OQPSK) is considered. An ML-based power estimation technique is also developed. This power estimation technique, along with a previously developed threshold setting technique, is applied to the OQPSK/BPSK/QPSK (O/B/Q) problem to demonstrate an autonomous 3-way classifier.

1.1 The Likelihood Approach

The MC task may be formulated as a binary hypothesis testing problem between the hypotheses \( H_0 \) and \( H_1 \) given by

\[
H_i : \quad r(t) = s_i(t) + n(t) \quad t \in J \quad i = 0, 1, \quad (1)
\]

with the complex baseband equivalent signal under hypotheses \( H_i \) is expressed as

\[
s_i(t) = \sqrt{S} e^{j \theta} \sum_{n=0}^{N-1} a_n u_T (t - nT) \quad i = 0, 1, \quad (2)
\]

where \( \{a_n\} \) is an independent sequence of unit energy \( \mathbb{E}\{|a_n|^2\} = 1 \) data symbols uniformly distributed over the alphabet \( \mathcal{A}_i \) under \( H_i \), and \( \theta \) is the unknown carrier phase, assumed to be uniformly distributed over \([0, 2\pi]\). The data pulse \( u_T(t) \), is assumed to be unity for \( t \in [0, T] \) and zero otherwise. It follows that \( E = ST \) is the energy per symbol in \( s_i(t) \) under either
hypothesis. The noise \( n(t) \) is a white circular Gaussian process with intensity \( N_0/2 \) in each dimension \( \mathbb{E} \{ n(t+\tau)n^*(t) \} = N_0 \delta(\tau) \), so that the Signal-to-Noise-Ratio (SNR) per bit is \( \gamma = E/N_0 \).

An MC decision may then be made by the Average Likelihood Ratio Test (LRT) \([5]\)

\[
\frac{\Lambda(\mathcal{H}_0; r(t))}{\Lambda(\mathcal{H}_1; r(t))} \gtrsim \mathcal{H}_0 \text{ or } \mathcal{H}_1, \tag{3}
\]

where the average likelihood functional (ALF) for constant envelope signals is

\[
\Lambda(\mathcal{H}_i; r(t)) = \mathbb{E} \left\{ \exp \left[ -\frac{2}{N_0} \Re \left\{ \int f_j(r(t)) s_j^*(t) dt \right\} \right] \right\}, \tag{4}
\]

with the ensemble average being taken over the data and the carrier phase. It has been shown in [4] that the natural log of (4) may be approximated by

\[
\lambda(\mathcal{H}_i; r(t)) \triangleq \ln \left( \Lambda(\mathcal{H}_i; r(t)) \right) \gtrsim t^i, \tag{5}
\]

with

\[
t^i = \sum_{k=1}^{\infty} \left( \frac{\gamma}{\sqrt{STc}} \right)^k \sum_{m=0}^{[k/2]} \frac{\beta_i(k, 2m) N^{-1}}{n!} \left( \sum_{n=0}^{N-1} |r_n|^{2m} r_k^{k-2m} \right), \tag{6}
\]

where \([z]\) represents the integer part of \( z \) and

\[
r_n = c \int_{nT}^{(n+1)T} r(t) dt \tag{7}
\]

\[
\beta_i(k, 2m) = \frac{\epsilon_{k-2m}}{m!(k-m)!} \mathbb{E} \{ |a_n|^{2m} (a_n^*)^{k-2m} \} \tag{8}
\]

with \( \epsilon_m \) denoting the Neumann factor (i.e., \( \epsilon_0 = 1, \epsilon_m = 2, m > 0 \)). The factor \( c \) is an arbitrary constant associated with the receiver gain. The precise value of \( c \) is somewhat irrelevant since the discrete-time model may be expressed as

\[
r_k = c\sqrt{ST} e^{i\theta} a_k + n_k \tag{9}
\]

\[
= \sqrt{E_c} e^{i\theta} a_k + n_k, \tag{10}
\]

where \( E_c = (c\sqrt{ST})^2 \) and \( n_k \) is a sequence of i.i.d. circular Gaussian random variables with \( \mathbb{E} \{ |n_k|^2 \} = N_0 c = N_0 Tc^2 \). In other words, a fixed value of \( c \) doesn’t affect the SNR (\( \gamma \)) and results in an effective symbol energy of \( E_c \) and an effective noise variance of \( N_0 c \).

It follows that (3) is approximately equivalent to the quasi-log-LRT (quasi-LRT)

\[
t^0 - t^1 > Q, \tag{11}
\]

where the approximation of (6), or a truncated version, is used in place of the log-ALF.

The assumption of common signal characteristics, such as \( T, u_T(t) \) and \( E_c \), along with symmetries inherent to many common signal constellations, results in the cancellation of many terms in (11). It has been demonstrated that BPSK/QPSK classification using the test statistic

\[
q_2 = \sum_{n=0}^{N-1} r_n^2 \tag{12}
\]

performs very close to the optimal performance given by the test in (3) over a large range of SNR values [4].

### 1.2 Threshold Settings

If the exact expression for the log-ALF were used in (11), the optimal threshold choice (i.e., the choice which minimizes the error probability assuming equally likely hypotheses) is \( Q = 0 \). However, for truncated versions of the series expansion, a nonzero threshold is required. In general, the optimal choice for \( Q \) depends on \( N, E_c, \) and \( N_0 c \). Two threshold setting techniques will be referred to in this paper. The first is an idealization used for simulation purposes. This ideal threshold is the value of \( Q \) which minimizes \( P_{FC} \) over a large number of data and noise realizations. It is conceivable that such simulations could be run off-line and that the optimal values of \( Q \) could be stored as a function of the signal and noise parameters. A practical receiver would still require estimates of these parameters. A more practical technique was developed in [6] by approximating the density of the test statistic under each hypothesis. When the \( q_2 \) test statistic is used to classify between BPSK and QPSK, the optimal threshold is approximated by

\[
Q_{B/Q} = E_c \left[ N^2(2V + 1)(2V + N) \right]^{1/4}, \tag{13}
\]

where

\[
V = 2\gamma^{-1} + \gamma^{-2}. \tag{14}
\]

Note that this technique requires estimates of \( E_c \) and \( \gamma \).

### 1.3 MMC

ML decision theory for \( M \) hypotheses is well developed [5], and has found application in the optimal reception of \( M \)-ary digital communications signals. Application of this theory to the MMC problem is conceptually straightforward: one must compute the log-ALF of (5) for each of the \( M \) possible signal formats, then select the hypothesis which maximizes this quantity. In fact, by looking at the differences between pairs of log-ALFs this can be reduced to \( M - 1 \) comparisons against a zero threshold.

In practice, approximations to the log-LRT are required via (6), resulting in the need to use nonzero thresholds. In fact, when an approximation is utilized, the best decision region in the \( M \)-dimensional space may be quite complex and not accurately described by a set of threshold comparisons.
For concreteness, consider the $M = 3$ case, for which the optimal test is

$$\lambda(\mathcal{H}_i; r(t)) - \lambda(\mathcal{H}_j; r(t)) \begin{cases} \geq 0 & \mathcal{H}_i \\ < 0 & \mathcal{H}_j \end{cases}$$

(15)

for $(i, j) = (0, 1), (1, 2)$. One could consider the case where the log-ALF is approximated and the decisions are made via two threshold comparisons.\(^1\) In this case, the implicit decision region is linear, which is, in general, suboptimal given the three log-ALF approximations. Even with these two constraints (i.e., approximate test statistics and linear decision regions), one must still estimate a set of $M - 1$ thresholds. Considering the difficulty in determining thresholds for the $M = 2$ MC problem, extension to MMC seems prohibitively complex. In Section 3.1 we demonstrate a practical approach to MMC.

2 Pairwise MC for OQPSK

In this section, we develop quasi-LRT rules for classifying between OQPSK and either BPSK or QPSK. The development of Section 1.1 must be slightly modified to account for an offset modulation format. The complex baseband OQPSK signal may be modeled as

$$s_O(t) = \sqrt{S}e^{j\phi} \sum_{n=0}^{N_O-1} a_n(t)u_{T_O}(t - nT_O - iT_O/2),$$

(16)

with $a_n(t)$ uniformly distributed over $\{-\sqrt{2}, \sqrt{2}\}$, and i.i.d. with respect to index $i$ and $n$. The signal corresponding to $i = 0$ may be thought of as the “nongstered” in-phase component, while the $i = 1$ component is the quadrature “staggered” signal. Note that the symbol-time of this signal is $T_O$ and the energy per symbol is $E_O = ST_C$.

We define the baud-time as the minimum time between data transitions. According to this definition, the baud time of the OQPSK signal defined in (16) is $T_O/2$. It is apparent that, depending on the timing assumptions, there are several cases one may consider for the corresponding QPSK (BPSK) signal. It is reasonable to assume that the QPSK (BPSK) signal should have the same power as the OQPSK signal, otherwise a simple power measurement could provide classification. We characterize the possible cases according to whether the OQPSK and QPSK (BPSK) signals have common symbol-time or baud-time.

2.1 Common Symbol-Time (Case 1)

In this case the baseband equivalent QPSK (BPSK) signal is represented as

$$s_{B/Q}^{(cs)}(t) = \sqrt{S}e^{j\phi} \sum_{n=0}^{N_O-1} a_n(t)u_{T_O}(t - nT_O - iT_O/2),$$

(17)

where all parameters are as described above and $a_n = a_n(0) + ja_n(1)$. For BPSK $a_n(1) = 0$ and $a_n(0)$ is i.i.d.,

\(^1\)A consistent set of three threshold comparisons reduces to two threshold comparisons.

uniformly distributed over $\{-1, +1\}$. For QPSK, $a_n(i)$ is as in the OQPSK signal. The factor $\delta$ is used to determine to sub-cases to the case 1 timing assumption. Specifically, the known epoch sub-case (case 1a), is defined by $\delta = 0$. A more realistic model may be the unknown epoch case, in which $\{\delta = 0\}$ and $\{\delta = 1\}$ each occur with probability $1/2$. The baud-time of the signal of (17) is the same as the symbol-time $(T_O)$.

2.2 Common Baud-Time (Case 2)

In this case the baseband equivalent QPSK (BPSK) signal is represented as

$$s_{B/Q}^{(cb)}(t) = \sqrt{S}e^{j\phi} \sum_{n=0}^{2N_O} a_n(t)u_{T_O/2}(t - nT_O/2),$$

(18)

where $a_n$ is as in case 1. In this case the baud-time and the symbol-time are $T_O/2$. Since the QPSK (BPSK) signal of (18) has a symbol-rate twice that of the OQPSK signal of (16) but equal power, it follows that the SNR per bit is $E_O/(2N_O)$ — half that of the OQPSK signal and the QPSK (BPSK) signal of case 1. We also have a common observation interval of duration $N_O T_O$, so that $N_{B/Q} = 2N_O$ QPSK (BPSK) symbols are observed for case 2.

Since only the output of a baud-rate detector is likely to be available, case 2 may be the most realistic timing assumption. We consider the other two cases for completeness and to facilitate comparison with results obtained elsewhere.

2.3 Log-ALF and Approximations

The same assumptions used to derive (6) can be used to approximate the log-ALF for the OQPSK signal by

$$I_O = \sum_{k=1}^{\infty} \frac{\gamma_O}{\sqrt{ST_c}} [\frac{N_O}{2}]^{2k} \beta_{O}(k; m)$$

$$\cdot \left[ \sum_{n=0}^{N_O-1} |r_n(0)|^{2(k-m)}r_n(0)^{2m} + (-1)^m |r_n(1)|^{2(k-m)}r_n(1)^{2m} \right],$$

(19)

where $r_n(i)$ is the $(T_O)$ sampled output of the bandpass matched filter

$$r_n(i) = c \int_{(n+i/2)T_O}^{(n+i+1/2)T_O} r(t)dt \quad i = 0, 1,$$

(20)

and $\gamma_O = E_O/N_O$. The term $\beta_O(k; m)$ is defined as

$$\beta_{O}(k; m) = \frac{\epsilon_m}{2^k(k + m)!(k - m)!}.$$ 

(21)

Note that the symbol-rate front-end processing consists of two coherent OQPSK receiver front-ends operating in quadrature. The signals for case 2 are more easily described using a baud-rate signal model. This

---
model is obtained by defining the $T_O/2$ integrate-and-dump output $r_k$

$$r_k = c \int_{kT_O/2}^{(k+1)T_O/2} r(t) dt. \quad (22)$$

This baud-rate model is related to the symbol-rate model via

$$r_n(i) = r_{2n+i} + r_{2n+i+1}; \quad n = 0, 1, \ldots, N_O; \quad i = 0, 1. \quad (23)$$

Care must be taken when using either the symbol- or baud-rate model for simulation purposes. For example, the noise components of $r_n(0)$ and $r_n(1)$ are correlated in the symbol-spaced model, while the data transitions of the OQPSK signal in the baud-rate model are restricted.

Expressions for the approximate log-ALF of the QPSK signal of cases 1a and 2 are special cases of (6). An approximate expression for the log-ALF of the QPSK signal of case 1b can be obtained by averaging the results for $\delta = 1$ and $\delta = 0$. In accordance with (11), the quasi-LRT statistic corresponding to the $\gamma^2$-term (first term with non-zero difference) is

$$\text{Case 2: } 2N_O \sum_{n=0}^{2N_O} \Re \{r_n r_{n+1}^* \} + \sum_{n=0}^{2N_O} (-1)^n r_n r_{n+1} \quad (24)$$

for either the OQPSK/BPSK or the OQPSK/QPSK tests under the common baud assumption. Test statistics for the timing assumptions of case 1 were also found for the two pairwise MC problems. Simulation results for each of these cases are plotted in Figure 1.

2Note that the test statistic for case 1a and 1b for the OQPSK/BPSK cases were identical, thus only one curve is shown.

3 A Practical Approach to MMC

In many real-world MMC problems, one is attempting to classify between signals which can be roughly grouped according to their common “macro” characteristics. Macro characteristics include such features as signal power, baud-time, pulse shape, staggered format, etc. Within each of these macro classes, there is often only one or two members. For such cases, we propose an approach to MMC consisting of the following two stages: (i) classify according to macro characteristics, and (ii) within the selected macro group, perform “micro” MC as necessary. This effectively reduces the $M$-ary decision problem to a sequence of 2-way decisions, which can be carried out via decoupled threshold tests.

3.1 MMC: O/B/Q

As a concrete example, consider the O/B/Q ($M = 3$) MMC problem. Making the assumption of common baud-time (case 2 of Section 2) allows us to use the staggered format as a macro characteristic since the case 2 test statistic in (24) holds for both OQPSK/QPSK and OQPSK/BPSK classification. Also, simulation results suggest that the thresholds are approximately the same. This allows the stage-1 test to be conducted to classify between \{OQPSK\} and \{BPSK, QPSK\} using the statistic in (24). If OQPSK is selected, the MMC task is complete; if not, the stage-2 test can be implemented using $q_2$ to decide between QPSK/BPSK.

Simulations showed that each of the three corresponding two-way classification rules (i.e., BPSK/QPSK, QPSK/QPSK, and OQPSK/BPSK) have roughly the same performance. The false classification probability of the two-stage O/B/Q classifier was found to be approximately the sum of the stage-1 and stage-2 error rates, or equivalently, twice the false classification probability of the pairwise tests. Since it has been demonstrated that the $q_2$ rule for QPSK/BPSK classification is near the performance of the exact LRT, it may be concluded that this two-stage O/B/Q test is near optimal in performance. In fact, it can be shown that the $P_{FC}$ of this two-stage test is no more than three times as large as any O/B/Q test.

4 Power Estimation Techniques

An autonomous quasi-LRT classifier requires estimates of $E_c$ and $N_{0,c}$ to set the threshold. The matched filter output in (7) is a sufficient statistic for this power estimation. A simple, but ad hoc, approach is to use moment matching. Assuming the model of (10), it is straightforward to show that

$$\mathbb{E} \{ |r_n|^2 \} = E_c + N_{0,c} \quad (25)$$

$$\text{var} \{ |r_n|^2 \} = N_{0,c}^2 + 2E_cN_{0,c}. \quad (26)$$

This motivates the estimates

$$\hat{E}_{c,MM} = \sqrt{m^2 - \sigma^2} \quad (27)$$

$$\hat{N}_{0,c,MM} = \hat{m} - \hat{E}_{c,MM}, \quad (28)$$
where $\bar{m}$ and $\sigma^2$ are the sample mean and variance of $|r_n|^2$, respectively. As might be expected, this technique works well for large SNR and/or large $N$, otherwise it fails – i.e., $N_{0,c,MM} < 0$.

A more systematic approach to power estimation is based on ML parameter estimation. For the signal model assumed in Section 1.1, ML estimation of $E_c$ and $N_{0,c}$ requires maximizing the quantity

$$N_{0,c}^{-N} e^{-\frac{1}{2} N_{0,c}} \left( \sum_{n=0}^{N-1} |r_n|^2 + N E_c \right) \times \mathbb{E} \left\{ I_0 \left( \frac{2 \sqrt{E_c}}{N_{0,c}} \sum_{n=0}^{N-1} r_n a_n^* \right) \right\}, \quad (29)$$

which is complicated by the averaging of the nonlinear Bessel function over all data sequences.

If the model for the carrier phase is changed from being constant over the entire observation interval, to being a sequence of i.i.d. uniform random variables, then the ML estimation problem requires maximizing of

$$N_{0,c}^{-N} e^{-\frac{1}{2} N_{0,c}} \left( \sum_{n=0}^{N-1} |r_n|^2 + N E_c \right) \prod_{n=0}^{N-1} I_0 \left( \frac{2 \sqrt{E_c}}{N_{0,c}} |r_n| \right) \quad (30)$$

with respect to $E_c$ and $N_{0,c}$. Note that development of (30) makes explicit use of the constant envelope assumption. Differentiation with respect to $E_c$ and $N_{0,c}$ yields the following critical point condition:

$$\sqrt{E_c} = \frac{1}{N} \sum_{n=0}^{N-1} |r_n| I_1 \left( \frac{2 \sqrt{E_c}}{N_{0,c}} |r_n| \right) I_0 \left( \frac{2 \sqrt{E_c}}{N_{0,c}} |r_n| \right) \quad (31a)$$

$$E_c + N_{0,c} = \frac{1}{N} \sum_{n=0}^{N-1} |r_n|^2. \quad (31b)$$

The equations of (31) were solved iteratively. Numerical experiments showed that convergence occurred rapidly and that the fixed point was insensitive to the initialization. This ML approach had significantly lower estimation bias than the moment matching approach and suffered virtually no failures (i.e., negative power estimates).

5 Autonomous O/B/Q MMC

Finally, the ML power estimation algorithm of Section 4 was combined with the threshold setting technique of (13) and the two-stage O/B/Q classification rule of Section 3.1. The results are illustrated in Figure 2. Note that the performance degrades at very low $P_{FC}$ due to extreme sensitivity to the threshold setting. It should be noted that the false error floor in the $N_O = 500$ curve corresponds to a region of large variance in the ML estimate of $N_{0,c}$. For correct

classification rates of 90% and below, a degradation of about 1 dB in SNR is suffered relative to the ideal threshold case.

6 Concluding Remarks

The method of ML modulation classification was extended to include more than two hypothesized modulation formats and to include autonomous power estimation and threshold setting. A 99% correct classification rate for O/B/Q was demonstrated with only 4 dB of OQPSK $E/N_0$ and 500 symbols (i.e., 1 dB of QPSK/BPSK $E/N_0$ and 1000 symbols) using a fully autonomous rule. It appears that the most difficult aspect of designing an effective autonomous classifier is obtaining reliable power estimates when only in-band measurements are available.

References


