Automatic Modulation Detection

Alireza Karami
Electrical Engineering Department
Dalhousie University

Abstract

Automatic modulation detection is a block prior the demodulator. Since in communication, signals are random, we are confronted with unknown parameters such as the carrier frequency and signal phase offset as well as the transmitted symbols. There are challenges to recognize the type of modulation. In this article, we introduce two ways which are used to detect the modulation type: likelihood-based and feature-based methods. In the former, the likelihood function appears as a very important function to constitute the structure of the modulation detection block. The latter takes advantage of time and frequency related features of the signal to detect modulation type.

1 Introduction

Modulation detection in receivers is a step between signal detection and demodulation. It has a wide range of application in military and commercial systems such as eavesdropping enemy signals, spectrum and network traffic management. Nowadays, in many communication systems modulation detection is unavoidable. In these systems, at first, the type of modulation is determined, and then the demodulation process starts to run. In the past, modulation detection was entirely performed by an operator. Thereafter, it was used semi-automatically till recently, when fully automatic ways have been proposed. In its rudimentary version, received signal parameters were calculated by an operator, and then a decision was made on the results. In semi-automatic detection, the operator used a bank of modulators and tried to detect the modulation type observing their outputs. Related parameters include signal waveform and frequency spectrum, signal instantaneous amplitude and frequency. The need for expert operators, long samples of the
signal, capability of detection on low variety of modulations and complexity were foibles of the first versions of modulation detection systems. Hence, the need for more powerful ways led to all-automatic modulation detection systems. Carrier frequency, symbol timing and SNR (Signal to noise ratio) are important parameters, which have a great effect on the modulation detection process. Therefore, to detect a modulation type with high success rate, these parameters must be estimated as precisely as possible. A modulation detection system includes two parts: a pre-processing unit and classification unit. In pre-processing unit actions such as noise cancellation, carrier frequency estimation, symbol period estimation, SNR estimation and equalization or parts of these are carried out. In classification unit, considering the used algorithm, modulation type of the received signal is determined and reported to the demodulation unit. The classification unit can detect the modulation in two ways: LB (likelihood-based) and FB (featured-based). A good classification unit detects among a large number of modulation types in a short observation window.

2 Likelihood Based Method

At first, we must consider a model for the received signal, such as the one in (1).

\[ r(t) = s(t; u_i) + n(t) \quad 0 \leq t \leq kT \]  

where \( s(t; u_i) \) is a baseband signal and \( i = 1, 2, \ldots, N_{mod} \) represents the modulation type.

\[ s(t; u_i) = a_i e^{j2\Delta f t} e^{j\theta} \sum_{k=1}^{K} e^{j\phi_k} s_k^{(i)} g(t - (k - 1)T - \varepsilon T) \]  

In (2) \( a_i = \sqrt{\frac{E_p}{\sigma_{s(i)}^2}} \) and

\[ \sigma_{a(i)}^2 = \frac{1}{M_i} \sum_{m=1}^{M_i} |s_m^{(i)}|^2 \]

where \( \sigma_{a(i)}^2 \) is average energy (variance) of the \( i \)-th modulation type. \( M_i \) is the number of symbols in the \( i \)-th modulation type (here all the points in the constellation are equi-probable), \( E_p \) and \( E_s \) are, respectively, the pulse and baseband signal energies, \( \Delta f \) is the carrier frequency offset and \( \theta \) is the
time invariant carrier phase, \( \{\phi_k\}_{k=1}^K \) represents phase jitter, \( g(t) \) is the pulse shape, affected by the channel, that is if \( h(t) \) is the channel impulse response and \( P_{TX}(t) \) is the pulse shape, then \( g(t) = P_{TX}(t) * h(t) \), \( T \) is the symbol period, \([0, kT]\) represents observed \( k \) symbols and \( \varepsilon \) is the timing offset in the receiver reference frame \((0 < \varepsilon < 1)\). \( s^{(i)}_k = s^{(i)}_{k,I} + js^{(i)}_{k,Q} \) is the transmitted symbol in the \( k \)-th period, and \( n(t) \) is the additive noise in the receiver. \( s(t; u_i) \) includes such unknown parameters as:

\[
u_i = \left[ a_i \Delta f \theta T \varepsilon \ g(t) \ {\phi_k}_{k=1}^K \ {s^{(i)}_k}_{k=1}^K \right],
\]

which is comprised of the signal and the data symbols. There will be nothing lost if we consider data symbols as normalized, i.e., \( \{s^{(i)}_k\}_{k=1}^K = \frac{s^{(i)}_k}{\sigma^{(i)}_k} \).

In likelihood-based methods, LRT (likelihood ratio test) is used, which uses the LF (likelihood function) of the received signal in the time interval \([0, kT]\). Depending on what the unknown parameters are, there are three criteria for discovering the modulation type using LB: ALRT (Average Likelihood Ratio Test), GLRT (Generalized Likelihood Ratio Test) and HLRT (Hybrid Likelihood Ratio Test).

### 2.1 Average Likelihood Ratio Test

In ALRT, the unknown parameters are considered as random variables with certain PDFs (Probability Density Functions), and the likelihood function is calculated using (3). In fact, we obtain likelihood function by averaging over all the unknown parameters.

\[
\Lambda^{(i)}_A[r(t)] = \int \Lambda[r(t)|v_i, H_i] \rho(v_i|H_i) \, dv_i
\]

where \( \Lambda[r(t)|v_i, H_i] \) is the likelihood function of the received signal, \( r(t) \), provided that \( v_i \) and \( H_i \) are given, \( H_i \) represents the modulation type, \( v_i \) is the unknown parameter vector, and \( \rho(v_i|H_i) \) is the probability density function of \( v_i \) for \( H_i \). In a classification unit which should detect two types of modulation, the decision on the modulation type is done as below

\[
\frac{\Lambda^{(1)}_L[r(t)]}{\Lambda^{(2)}_L[r(t)]} < \mu \rightarrow H_1
\]

\[
\frac{\Lambda^{(1)}_L[r(t)]}{\Lambda^{(2)}_L[r(t)]} > \mu \rightarrow H_2
\]
where $\mu$ is a threshold, and the index $L$ represents the type of LRT. Now, Assuming that the noise is zero-mean AWGN (Additive White Gaussian Noise), we try to find a relationship for conditional likelihood function for $i$-th type of modulation. $r(t)$ can be considered with its samples, $r = [r_1 r_2 ... r_K]$. It is clear that $r_k$ are zero-mean gaussian random variables, $r_k = s(k; u_i) + n_k$.

The joint PDF of $r$, provided that $u_i$ is given, can be expressed as

$$p(r|u_i, H_i) = \left(\frac{1}{\sqrt{2\pi N_0}}\right)^K \exp \left\{-\frac{1}{N_0} \sum_{k=1}^{K} [r_k - s(k; u_i)]^2\right\}$$  \hspace{1cm} (4)

where $N_0$ is AWGN two sided PSD (Power Spectral Density). When $N \to \infty$, (4) may be written as

$$p(r(t)|u_i, H_i) = \left(\frac{1}{\sqrt{2\pi N_0}}\right)^K \exp \left\{-\frac{1}{N_0} \int_0^{KT} [r(t) - s(t; u_i)]^2\right\}$$  \hspace{1cm} (5)

The maximization of the related PDF is equivalent to maximization of below, which is likelihood function

$$\Lambda[r(t)|u_i, N_0, H_i] = \exp \left\{-\frac{1}{N_0} \int_0^{KT} [r(t) - s(t; u_i)]^2\right\}$$  \hspace{1cm} (6)

The likelihood function can be rewritten as

$$\Lambda[r(t)|u_i, N_0, H_i] = \exp \left\{-\frac{1}{N_0} \int_0^{KT} |r(t)|^2\right\} + 2N_0^{-1} \int_0^{KT} r(t) s^*(t; u_i) \ dt - N_0^{-1} \int_0^{KT} |s(t; u_i)|^2\right\}$$  \hspace{1cm} (7)

Note that the first term in the exponential function does not depend on
the signal parameters. Therefore, the likelihood function can finally be considered as

$$\Lambda[r(t)|u_1, N_0, H_i] = \exp \left\{ 2N_0^{-1} \text{Re} \left[ \int_0^{K_T} r(t) s^*(t; u_i) \, dt \right] - N_0^{-1} \int_0^{K_T} |s(t; u_i)|^2 \, dt \right\}$$

$$= \exp \left\{ \sum_{k=0}^{K-1} 2N_0^{-1} \text{Re} \left[ \int_{kT}^{(k+1)T} r(t) s^*(t; u_i) \, dt \right] - N_0^{-1} \int_{kT}^{(k+1)T} |s(t; u_i)|^2 \, dt \right\}$$

(8)

An exponential function of summations can be changed to multiplications of the exponential functions, so

$$\Lambda[r(t)|u_1, N_0, H_i] = \prod_{k=1}^{K} \exp \left\{ 2N_0^{-1} \text{Re} \left[ \int_{kT}^{(k+1)T} r(t) s^*(t; u_i) \, dt \right] - N_0^{-1} \int_{kT}^{(k+1)T} |s(t; u_i)|^2 \, dt \right\}$$

(9)

If we consider the unknown parameters of the received signal as $v_i = \left[ \left\{ s_k^{(i)} \right\}_{k=1}^{K} \right]$, since $\rho(v_i|H_i) = \frac{1}{(M_i)^K}$, using (3) and (9)

$$\Lambda_A^{(i)}[r(t)] = \prod_{k=1}^{K} E_{s_k^{(i)}} \exp \left\{ 2\sqrt{S} N_0^{-1} \text{Re}[R_k^{(i)}] - S T N_0^{-1} |s_k^{(i)}|^2 \right\}$$

(10)

where $S = E_s/T$ is the signal power, and $E_{s_k^{(i)}}$ is the average energy of the $i$-th modulation constellation. $R_k^{(i)}$ is obtained from (11).

$$R_k^{(i)} = \int_{(k-1)T}^{kT} r(t) s_k^{(i)*} \, dt$$

(11)

In the case that we consider $v_i = \left[ \theta, \left\{ s_k^{(i)} \right\}_{k=1}^{K} \right]$, that is, the phase $\theta$ and the signal symbols $\left\{ s_k^{(i)} \right\}_{k=1}^{K}$ are considered as the unknown parameters
Λ^{(i)}_A[r(t)] = E_{\{ s_k^{(i)} \}_{k=1}^K} \left[ e^{-STN_0^{-1} \eta_k^{(i)} I_0 (2\sqrt{SN_0^{-1}}|\zeta_k^{(i)}|)} \right] \quad (12)

where $I_0(\cdots)$ is the modified Bessel function of the first kind and order 0. $E_{\{ s_k^{(i)} \}_{k=1}^K} [\cdots]$ is the average of the argument over the $K$ data symbols, $\eta_k^{(i)}$ and $\zeta_k^{(i)}$ are obtained from

$$\eta_k^{(i)} = \sum_{k=1}^K R_k^{(i)} \quad (13)$$
$$\zeta_k^{(i)} = \sum_{k=1}^K |s_k^{(i)}|^2 \quad (14)$$

It is clear that not knowing $\theta$ causes the computations to increase. If we add other parameters into $v_i$, the computational complexity will increase further. For this reason and the need for prior knowledge of some parameters, ALRT is an impractical method for modulation detection. Therefore, another method which originates from it, Q-ALRT (Quasi Average Likelihood Ratio Test) has been proposed.

### 2.1.1 Quasi Average Likelihood Ratio Test

As mentioned before, this method is more practical than ALRT. Considering $v_i$, below algorithms based on QALRT have been proposed to detect linear modulations on the AWGN channel.

$$v_i = \begin{bmatrix} \theta \{ s_k^{(i)} \}_{k=1}^K \end{bmatrix}$$
$$v_i = \begin{bmatrix} \theta \varepsilon \{ s_k^{(i)} \}_{k=1}^K \end{bmatrix}$$

Also, for $v_i$ below other algorithms have been proposed for FSK (Frequency Shift Keying) signal detection on the AWGN channel.

$$v_i = \begin{bmatrix} \{ \phi_k \}_{k=1}^K \{ s_k^{(i)} \}_{k=1}^K \end{bmatrix}$$
$$v_i = \begin{bmatrix} \varepsilon \{ \phi_k \}_{k=1}^K \{ s_k^{(i)} \}_{k=1}^K \end{bmatrix}$$
In (15) considering \( v_i = \left[ \theta \ \{ s_k^{(i)} \}_{k=1}^K \right] \) an approximation of the likelihood function for PSK (Phase Shift Keying) and QAM (Quadrature Amplitude Modulation) signals is provided.

\[
\Lambda_A^{(i)}[r(t)] \approx \exp \left\{ \sum_{n=1}^{\infty} \left( \sqrt{\text{SNR}_0}^{-1} \right)^n K \sum_{q=0}^{\lfloor n/2 \rfloor} [\nu_{n-2q}(q!(n-q)!)^{-1} |m_{s^{(i)},n,q}| |\hat{m}_{r,n,q}|] \right\}
\]

(15)

where \( m_{s^{(i)},n,q} \) is obtained from below

\[
m_{s^{(i)},n,q} = E \left[ (s^{(i)})^{n-q} (s^{(i)\ast})^q \right]
\]

and

\[
\hat{m}_{r,n,q} = K^{-1} \sum_{k=1}^{K} r_k^{n-q} (r_k^{\ast})^q
\]

For \( q = \frac{n}{2} \) and \( q < \frac{n}{2} \), \( \nu_{n-2q} \) is respectively 1 and 2. For symmetric constellations \( n \)-th order moments are zero for odd \( n \).

To differentiate \( M \)-PSK from \( M' \)-PSK \( (M' > M) \), it is enough to calculate the \( M \)-th order moment with \( q = 0 \). Considering \( \varepsilon \) in \( v_i \) as well, the likelihood function can be obtained using (16).

\[
\Lambda_A^{(i)}[r(t)] \approx D^{-1} \sum_{d=0}^{D} \Lambda^{(i)}[r(t)|\varepsilon_d, H_i]
\]

(16)

where \( D \) is the number of levels which \( \varepsilon \) can get and \( \varepsilon_d = \frac{d}{D} \). If \( D \to \infty \), (16) changes into an integral, and as a result the accuracy increases, and so does complexity.

### 2.2 Generalized Likelihood Ratio Test

In the case that the PDF of the signal unknown parameters is not given, assuming \( H_i \) we can estimate related parameters. In this method we try to
find the maximum likelihood function as in (17). In fact, we maximize the likelihood function over $v_i$ assuming that the modulation type is $H_i$.

\begin{equation}
\Lambda_G^{(i)}[r(t)] = \max_{v_i} \Lambda[r(t)|v_i, H_i] 
\end{equation} 

On an AWGN channel for $v_i = \theta \in \left\{ s_k^{(i)} \right\}_{k=1}^{K}$, $\Lambda_G^{(i)}[r(t)]$ is given as

\begin{equation}
\Lambda_G^{(i)}[r(t)] = \max_{\theta} \left\{ \sum_{k=1}^{K} \max_{s_k^{(i)}} \left\{ Re[s_k^{(i)}^* r_k e^{-j\theta}] - \frac{1}{2} \sqrt{ST} |s_k^{(i)}|^2 \right\} \right\} 
\end{equation}

### 2.3 Hybrid Likelihood Ratio Test

This method, HLRT, is a combination of ALRT and GLRT, and the likelihood function is calculated as in (19).

\begin{equation}
\Lambda_H^{(i)}[r(t)] \approx \int_{v_{i1}} \max_{v_{i2}} \Lambda[r(t)|v_{i1}, v_{i2}, H_i] \, dv_{i2} 
\end{equation}

On an AWGN channel for $v_i = \theta \in \left\{ s_k^{(i)} \right\}_{k=1}^{K}$, $\Lambda_H^{(i)}[r(t)]$ is as below

\begin{equation}
\Lambda_H^{(i)}[r(t)] = \max_{\theta} \left\{ \prod_{k=1}^{K} E_{s_k^{(i)}} \left[ exp \left( 2\sqrt{SN_0^{-1}} Re[s_k^{(i)}^* r_k e^{-j\theta}] - STN_0^{-1} |s_k^{(i)}|^2 \right) \right] \right\}
\end{equation}

There are two advantages of GLRT and HLRT in comparison with ALRT. The first is that they include estimation of the unknown parameters, which can be useful as well in the demodulation process. Moreover, not only are GLRT and HLRT useful for AWGN, but they are also used in Rayleigh and Rician fading environments. GLRT seems to be more advantageous because it does not need to know the noise power, and it does not deal with calculating exponential functions. However, it is unsuccessful in detecting nested signal constellations. For example it includes the same likelihood functions for 16QAM and 64QAM. One disadvantage of HLRT is that it needs a lot of time for calculations.
3 Feature Based Method

In this method, modulation detection is done using other features of the signal. The features used, are different in different algorithms depending on what type of modulations are to be detected. The features can be instantaneous amplitude, instantaneous phase or instantaneous frequency. Feature-based methods at first determine the class of modulation, and then degree of that.

3.1 Spectral features

In this part, we introduce the basic features used in modulation detection.

1. $\gamma_{\text{max}}$, derivative from instantaneous amplitude, $A(t)$

$$\gamma_{\text{max}} = \max |DFT(A_{\text{cn}}(i))|^2$$

where $A_{\text{cn}}(i)$ is normalized instantaneous amplitude at $t = \frac{i}{f_s}$, $i = [1, 2, \cdots, N]$. $N$ is number of samples in each frame of the received signal.

$$A_{\text{cn}}(i) = A_n(i) - 1$$

$$A_n(i) = \frac{A(i)}{m_a}$$

where $m_a$ is average of the instantaneous amplitude in each frame.

$$m_a = \frac{1}{N} \sum_{i=a}^{N} A(i)$$

Normalization of instantaneous amplitude is done in order to remove channel gain effects.

2. The standard deviation (Variance) of absolute value of the instantaneous phase nonlinear part, $\sigma_{ap}$, which is obtained from (24).

$$\sigma_{ap} = \sqrt{\frac{1}{C} \left( \sum_{A_n(i) > a_t} \phi_{NL}(i)^2 \right) - \frac{1}{C} \left( \sum_{A_n(i) > a_t} |\phi_{NL}(i)| \right)^2}$$
where $\phi_{NL}(i)$ is normalized instantaneous phase nonlinear part in $t = \frac{i}{fs}$, $i = [1, 2, \ldots, N]$. C is the number of $\phi(i)$ samples for $A_n(i) > a_t$, $a_t$ is a threshold for $A_n(i)$. In fact, those samples of the signal with weak power are not considered.

3. The standard deviation (Variance) of the instantaneous phase nonlinear part, $\sigma_{dp}$, which is obtained from (25).

$$\sigma_{dp} = \sqrt{\frac{1}{C} \left( \sum_{A_n(i) > a_t} \phi_{NL}^2(i) \right) - \frac{1}{C} \left( \sum_{A_n(i) > a_t} \phi_{NL}(i) \right)^2}$$  (25)

4. The standard deviation (Variance) of absolute value of the normalized instantaneous amplitude, $\sigma_{aa}$, which is obtained from (26).

$$\sigma_{aa} = \sqrt{\frac{1}{N} \left( \sum_{i=1}^{N} A_{cn}^2(i) \right) - \frac{1}{N} \left( \sum_{i=1}^{N} |A_{cn}(i)| \right)^2}$$  (26)

5. The standard deviation (Variance) of absolute value of the normalized instantaneous frequency, $\sigma_{fa}$, which is obtained from (27).

$$\sigma_{fa} = \sqrt{\frac{1}{C} \left( \sum_{A_n(i) > a_t} F_N^2(i) \right) - \frac{1}{C} \left( \sum_{A_n(i) > a_t} |F_N(i)| \right)^2}$$  (27)

where

$$F_N(i) = \frac{f_c(i)}{r_b}$$

$$f_c(i) = F(i) - m_f$$  (28)

$$m_f = \frac{1}{N} \sum_{i=1}^{N} f(i)$$

$r_b$ is bit rate, and $f(i)$ is instantaneous frequency in $t = \frac{i}{fs}$, $i = [1, 2, \ldots, N]$.
A feature can be used alone or in cooperation with other ones in order to detect modulation type. Using all the features in an algorithm can provide us with a detector, capable of detecting a large number of modulations with a high certainty rate. There are two approaches to use the mentioned features to detect modulation type. In the first approach, the features can be compared with the thresholds in an algorithm to detect the modulation type. These thresholds and their number depend on the type and the number of modulations as well as SNR. Figure 1 shows a typical decoder of the first approach kind. In the second approach, the features are entered into input layer of a neural network, and the output layer can show what the modulation type is.

4 Conclusion

Two ways of modulation detection were presented. The optimal approach of the detection is included in LB. There are a lot of computations for ALRT, basic kind of LB, which make it impractical. Therefore, other types of LB, which are suboptimal, have been proposed. However, they still suffer from high complexity and being time consuming, while the number of unknown parameters increase. On the other hand, each of them can be useful only for some of modulation types. This has caused the FB approach to be introduced. In this method, some features of received signal are extracted, and then compared with the threshold amounts. Also, neural networks can be used in this regard. Although FB manners are suboptimal, they are less complex than LB and can be implemented to detect a greater number of modulation types.
The results of the method testing in Matlab environment are presented in Tabs. 2 and 3 at SNR of 5 and 0 dB respectively. We have used 200 signal realizations. It is apparent that even at SNR = 0 dB the algorithm recognizes 2FSK, MSK, 2ASK, BPSK, 8PSK and 16QAM modulations with probability at least 95%. 4FSK is recognized with a slightly lower reliability (87.5%). The worst results at SNR = 0 dB are obtained for QPSK signals, but the algorithm still recognizes them correctly in most cases.

Analytic signal
Calculation of $\gamma_{\text{max}}$
Spectrum calculation

$\gamma_{\text{max}} \leq 15$
Calculation of number of carriers

Number of carriers = 1
2ASK

Number of phases = 1

Number of phases = 3
4FSK

Number of phases = 2
2FSK

Calculation of $m_{pd}$

$m_{pd} > 0.05$

MSK

Calculation of instantaneous phase
Calculation of histogram

Calculation of number of phases

Number of phases = 1

Number of phases = 2

Number of phases = 3

Number of phases = 5

BPSK

QPSK

8PSK

16QAM

Figure 1: A typical feature based decoder