A SIMPLE AND ROBUST MODULATION CLASSIFICATION METHOD VIA COUNTING

Xiaoming Huo and David Donoho

Department of Statistics, Stanford University, Stanford, CA94305.
xiaoming@stat.stanford.edu, donoho@stat.stanford.edu

ABSTRACT

Automatic modulation classification (or recognition) is an intrinsically interesting problem with a variety of regulatory and military applications. We developed a method which is simple, fast, efficient and robust. The feature being used is the counts of signals falling into different parts of the signal plane. Compared with the likelihood method and the High Order Correlation method, it is much easier to be implemented, and the execution is much faster. When the channel model is correct, our method is efficient, in the sense that it will achieve the "optimal" classification rate. When unknown contamination is present, our method can automatically overcome to certain degree. At SNR being 10 and 15dB, examples of classifying two modulation types—QAM4 and PSK6—are given. Simulations demonstrate its ability to deal with unknown noises.

1. INTRODUCTION

A lot of methods have been developed for automatic modulation classification. They are either based on likelihood function or high order correlation [2] [6] [9] [13], or based on some special features of digital signal [1] [8] [11] [12]. Our method is similar to those methods in the second category. Its originality is the Hellinger distance parameter estimation method, which is developed in mathematical statistics [3] [4] [5]. It has been known for a while [10] [14] that Hellinger distance estimator is robust and efficient in parameter estimation. In modulation classification, Hellinger distance method seems to be an ideal candidate, because sometimes we do not have the perfect knowledge about the noise model. For example the noise may not be Gaussian, or the estimated SNR may not be accurate, etc. Hellinger distance can automatically overcome moderate degree of model distortion. But the likelihood ratio related method is incapable to do it. As well as achieving robustness, Hellinger distance method also possesses efficiency . Under the right model assumption, at least asymptotically(when the number of received signals is large), the Hellinger distance method is as efficient as likelihood method is.

This work is partially supported by CRASP

Our approach can be divided into three steps: First, the signal plane is partitioned into certain number of cells; Secondly, the number of received symbols is counted in each cell. Obviously the counts are multinomial distributed; Finally the Hellinger distance is calculated for multinomial distribution, which is nothing but a sum of square of the difference of some square roots. Once the partition is fixed, the whole algorithm is extremely easy to implement.

We will describe our formulation of the problem in the next section. In section 3, we describe our methodology and some analysis. Section 4 gives two simulation results. Finally, we conclude and illustrate future work in section 5.

2. PROBLEM FORMULATION

We assume an additive Gaussian noise model. Suppose the signal is 2-D modulated. So each symbol is a point on the signal plane. The received noisy symbols $X_1, X_2, \cdots, X_N$ are samples from a distribution with density

$$f_X(x) = \frac{1}{M}\sum_{i=1}^{M}\psi_{\mu,\sigma}(x),$$

where M is the cardinality of modulation constellation, and $\psi_{\mu,\sigma}$ is a bivariate Normal density with mean μ and covariance matrix $\sigma \cdot I_2$.

The modulation classification problem is to predict from which distribution the noisy signal is sampled.

We gave some analysis of Hellinger metric geometry in term of this setting [7]. Interested readers are referred to that paper.

3. OUR APPROACH

3.1. Multinomial Approximation

Suppose there are K candidate modulation types, and f_1, f_2, $\cdots$, f_K are their corresponding densities. Our approach is as following:

1. Divide the signal plane into L cells, count the number of signals in each cell. $Y_1, Y_2, \cdots, Y_L$ denote counts for all the cells.

2. The Hellinger distance between the empirical distribution and density $f_i, 1 \le i \le K$, is defined as

$$H^2(\hat{f}_N, f_i) = \sum_{j=1}^{L} \left(\sqrt{Y_j/N} - \sqrt{f_i^{(j)}} \right)^2 = 2 - 2\sum_{j=1}^{L} \sqrt{Y_j/N}\sqrt{f_i^{(j)}},$$

where $f_i^{(j)} = \int_{\text{cell}_j} f_i(x)dx$. It is just the Hellinger distance between two multinomial distributions.

3. The MHD classifier chooses minimizer k which satisfies

$$H^2(\hat{f}_N, f_k) = \min_{1 \le k' \le K} H^2(\hat{f}_N, f_{k'}).$$

3.2. Asymptotic Normality

Suppose there are only two candidate modulation types ($K = 2$) with corresponding densities f_0 and f_1. The classification is determined by the sign of the test statistic

$$\gamma_N = H^2(\hat{f}_N, f_0) - H^2(\hat{f}_N, f_1).$$

If $\gamma_N < 0$, the modulation is classified as f_0; Else, it's classified as f_1. The following is true (we state it without giving proof):

Theorem 1 (Asymptotic Normality.) *If the signal is from the density function f_0, then*

$$\sqrt{N}\{\gamma_N + H^2(f_0, f_1)\} \sim \textit{Normal}(0, \sigma^2_{HD}), \quad (1)$$

where $\sigma^2_{HD} = 1 - (\sum_{i=1}^{L} \sqrt{f_0^{(i)} f_1^{(i)}})^2$.

Based on the above result, for two hypothesis with corresponding Hellinger distance $H(f_0, f_1)$, roughly, it takes

$$N \approx \text{Const.} \frac{\sigma^2_{\text{HD}}}{H^4(f_0, f_1)} = (\text{Const.})^2 \cdot \frac{1 - \frac{1}{4}H^2(f_0, f_1)}{H^2(f_0, f_1)},$$

samples to reliably distinguish them.

3.3. Robustness Analysis

Our approach is robust. One way to understand it is to consider the break down point. The break down point of a test is defined as the maximum proportion of contamination that still makes the test consistent (asymptotically). Assume the signal is from f_0, but $\varepsilon(0 < \varepsilon < 1)$ of it is from an unknown distribution H. The limit of test statistic γ_N will be

$$-2\sum_{i=1}^{L} \sqrt{(1-\varepsilon)f_0^{(i)} + \varepsilon H^{(i)}} \left(\sqrt{f_0^{(i)}} - \sqrt{f_1^{(i)}} \right). \quad (2)$$

When it is less than zero, the test is consistent; Otherwise, it's inconsistent.

The following theorem says that the break down point of Hellinger distance test is at least $H^2(f_0, f_1)/\{4+H^2(f_0, f_1)\}$.

Theorem 2 (Break Down Point) *Follow the same notation, the expression in (2) is less than zero as long as*

$$\varepsilon \le \frac{H^2(f_0, f_1)/4}{1 + H^2(f_0, f_1)/4}. \quad (3)$$

This is only a theoretical result. Soon we can see it is very pessimistic. Because in digital communication, those obvious outliers are easy to be excluded by preprocessing. The above theory makes the assumption that the distortion can be infinity. So the break down point given in (3) is very conservative. From the simulations, we will demonstrate the robustness in finite cases. It's much better than we can predict from (3).

3.4. Optimal Partition

All the discussion before is based on one type of partition. A bad partition could make the test undoable. An optimal partition should maximize the Hellinger distance in the multinomial distribution situation. Because from theorem 1, larger $H^2(f_0, f_1)$ will make test easier; from theorem 2, larger $H^2(f_0, f_1)$ will bring a higher break down point. Since

$$H^2(f_0, f_1) = 2 - 2\sum_{i=1}^{L} \sqrt{f_0^{(i)}}\sqrt{f_1^{(i)}},$$

maximize $H^2(f_0, f_1)$ is equivalent to minimize $\sum_{i=1}^{L} \sqrt{f_0^{(i)}} \sqrt{f_1^{(i)}}$. The following three properties will be served as our guidelines in partitioning the signal plane.

Property 1 *Hellinger distance in continuous case is at least no smaller than the corresponding Hellinger distance after discretization.*

Property 2 *In cell $i(1 \le i \le K)$, if the likelihood ratio $f_0(x)/f_1(x)$ is not constant, it will always be better(get a larger Hellinger distance) to split this cell.*

The next one is important in designing the partition of the signal plane.

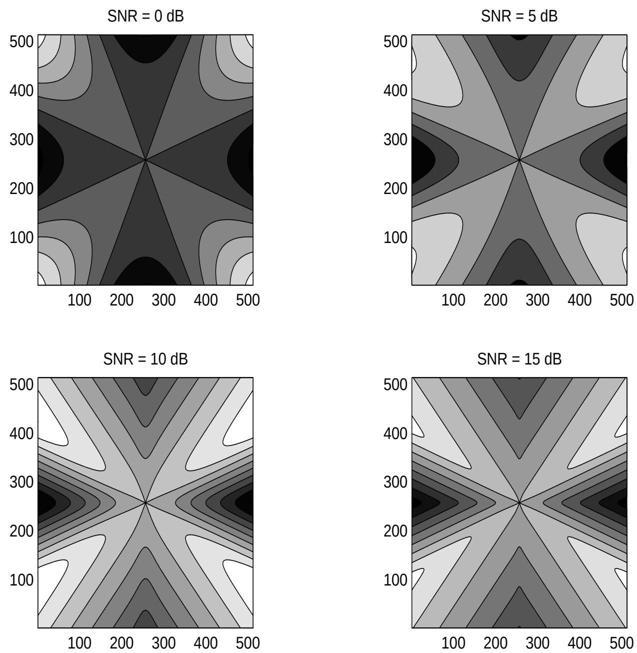

Figure 1: Contour of likelihood ratio-QAM4 vs PSK6.

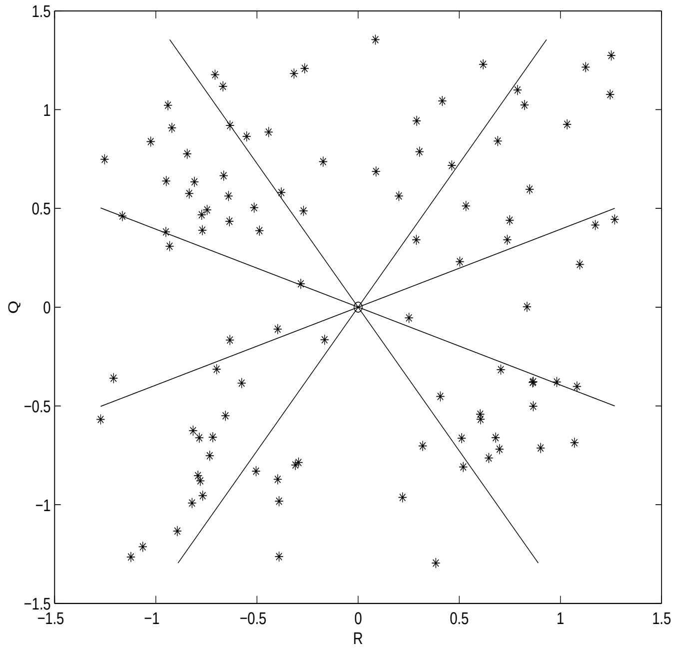

Figure 2: 100 noisy signals from QAM4 at SNR $= 10$dB.

Property 3 *Assume the two density functions, f_0 and f_1, are continuous. Let x be on the boundary between cell i and cell j, and x is not adjacent to any other cells. Suppose the partition minimize $\sum_{i=1}^{L} \sqrt{f_0^{(i)}} \sqrt{f_1^{(i)}}$. The following equality must be satisfied:*

$$\frac{f_1(x)}{\sqrt{f_1^{(i)} f_1^{(j)}}} = \frac{f_0(x)}{\sqrt{f_0^{(i)} f_0^{(j)}}}.$$

So it's always better to partition the signal plane along contour of likelihood ratio f_0/f_1.

Armed with these results, we can proceed to design an optimal partition for a real problem.

4. SIMULATIONS

We consider only two candidate modulation types—QAM4 and PSK6.

From Property 3, the optimal partition is to divide the signal plane along contour of likelihood ratio. Figure 1 shows filled contours of likelihood ratio in the situation where SNR is $0, 5, 10$, and 15dB respectively. Note there are four straight lines going through the origin. This pattern motivates us to partition the signal planes by phase. This way of partition coincides with the heuristic that phase is the most important information to distinguish QAM4 from PSK6.

We can actually compute the Hellinger distance from the continuous density functions and from the multinomial density functions. They turn out to be roughly equal. So the partition, then multinomial approximation, does not shrink the Hellinger distance very much. Following the discussion after theorem 1, the order of the sample size necessary for reliably distinguishing these two modulations can be calculated. For SNR $= 0, 5, 10, 15, 20$dB, the corresponding sample sizes are roughly $60000, 1000, 100, 30, 10$.

A sample size of 60000 implies that it is very difficult to distinguish in applications. A sample size of 10 means too easy to distinguish. We choose to test the method at SNR$=$ 10 and 15dB, which are in the most interested range. The signal length N are chosen to be 100 and 30 respectively. At different SNR, the signal plane is divided into 8 sectors by four straight lines going through the origin. Each line matches or is close to the contour $f_0/f_1 = 1$. Figure 2 gives an example of 100 noisy signals from QAM4 at SNR $= 10$dB. The solid lines shows how to partition the signal plane.

Suppose the signal comes from QAM4. To demonstrate the robustness of this method, we let a fraction of signal generate from a channel suffering 10dB SNR loss. The fraction is called the "contamination percentage". We let the contamination percentage vary from 0 up to $0.45(45\%)$. For comparison, a likelihood method classification is done in the same situation.

Figure 3 illustrates the mis-classification rate based on 1000 simulations. In both 10dB and 15dB situation, As the contamination percentage rise, our method consistently

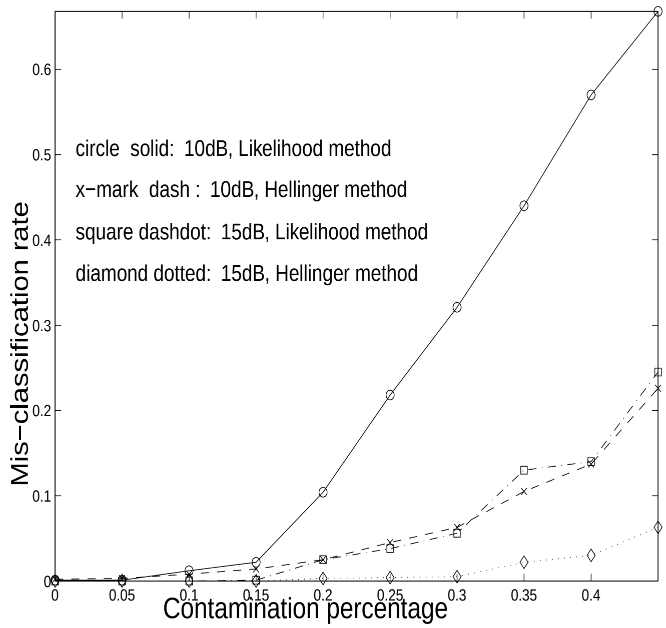

Figure 3: Mis-classification rate in 1000 simulations.

gives smaller mis-classification rates. When there is no contamination, both methods yield a close to zero mis-classification rate.

Another fact noteworthy is that the running time of our method (Hellinger classifier) is about one tenth of the likelihood ratio classifier. It is expected because our method only involves counting and a small amount of arithmetic computation.

5. CONCLUSION AND FUTURE RESEARCH

We propose a new way of performing modulation classification. It's simple, easy to implement, fast(can be done in real time) and robust. Simulations verify our conjecture.

In the future, we will test this method in other situations. For example, testing for other pair of candidate modulation types, testing for multiple hypotheses, etc.

6. REFERENCES

[1] E.E. Azzouz and A.K. Nandi, "Automatic identification of digital modulation types," *Signal Processing*, vol. 47, no. 1, pp. 55–69, November 1995.

[2] B.F. Beidas and C.L. Weber, "Higher-order correlation-based approach to modulation classification of digitally frequency-modulated signals," *IEEE Journal on Selected Areas in Communications*, vol. 13, no. 1, pp. 89–101, January 1995.

[3] Rudolf Beran, "Robust location estimates," *Annals of Statistics*, vol. 5, pp. 431–444, 1977.

[4] Rudolf Beran, "Minimum hellinger distance estimates for parametric models," *Annals of Statistics*, vol. 5, pp. 445–463, 1977.

[5] Rudolf Beran, "An efficient and robust adaptive estimator of location," *Annals of Statistics*, vol. 6, pp. 292–313, 1978.

[6] Char-Dir Chung and A. Polydoros, "Envelope-based classification schemes for continuous-phase binary frequency-shift-keyed modulations," in *Proceedings of MILCOM '94*, Fort Monmouth, NJ, USA, October 1994, vol. 3, pp. 796–800.

[7] David Donoho and Xiaoming Huo, "Large-sample modulation classification using hellinger representation," in *Signal processing advances in wireless communications*, April 1997, pp. 133–7.

[8] S.-Z. Hsue and S.S. Soliman, "Automatic modulation classification using zero crossing," *IEE Proceedings F (Radar and Signal Processing)*, vol. 137, no. 6, pp. 459–64, December 1990.

[9] Chung-Yu Huang and A. Polydoros, "Likelihood methods for mpsk modulation classification," *IEEE Transactions on Communications*, vol. 43, no. 2-4, pp. 1493–504, February 1995.

[10] Bruce G. Lindsay, "Efficiency versus robustness: the case for minimum hellinger distance and related method," *Annals of Statistics*, vol. 22, no. 2, pp. 1081–1114, 1994.

[11] A.K. Nandi and E.E. Azzouz, "Automatic analogue modulation recognition," *Signal Processing*, vol. 46, no. 2, pp. 211–22, October 1995.

[12] A.K. Nandi and E.E. Azzouz, "Modulation recognition using artificial neural networks," *Signal Processing*, vol. 56, no. 2, pp. 165–75, January 1997.

[13] A. Polydoros and K. Kim, "On the detection and classification of quadrature digital modulations in broadband noise," *IEEE Transactions on Communications*, vol. 38, no. 8, pp. 1199–211, August 1990.

[14] Douglas G. Simpson, "Hellinger deviance tests: Efficiency, breakdown points, and examples," *J. of American Statistical Association*, vol. 84, pp. 107–13, 1989.