MODULATION CLASSIFICATION IN MIMO FADING CHANNELS VIA EXPECTATION MAXIMIZATION WITH NON-DATA-AIDED INITIALIZATION

Zhechen Zhu, Asoke K. Nandi

Department of Electronic and Computer Engineering
Brunel University London
Kingston Lane, Uxbridge, Middlesex, UB8 3PH, United Kingdom

ABSTRACT

Non-data aided channel estimation is discussed in this paper to enable blind modulation classification in multiple-input multiple-output fading channels. The channel parameters are jointly estimated via expectation maximization under each modulation hypothesis. Instead of pilot symbols, the initialization of the channel matrix is achieved through a combination of fuzzy c-means clustering and maximum likelihood mapping. The estimated channel matrix and noise power enable the blind classification of modulations using a maximum likelihood classifier. Digital modulations are tested in simulation to validate the proposed classifier. The classifier is able to achieve excellent performance when SNR level is above 5 dB.

Index Terms— modulation classification, channel estimation, fuzzy clustering, Bayesian inference, likelihood classifier, MIMO, Rayleigh fading

1. INTRODUCTION

Modulation classification (MC) has received increasing amount of attention in the last decade or more from emerging intelligent communication systems, such as cognitive radio and software defined radio [1]. The wide application of adaptive modulation and coding provides the opportunity for further improvement of bandwidth efficiency where MC is employed to detect modulation automatically.

Much effect has been dedicated to MC in single-input and single-output systems [2–6]. MIMO systems with associated techniques such as spatial multiplexing (SM) and space-time coding (STC) provides benefits including array gain and spatial gain for improved spectrum efficiency and link reliability. Some recent publications address the issue of blind modulation classification (BMC) for MIMO systems. Choqueuse et al. developed the average likelihood ratio test (ARLT) classifier for MC with perfect channel knowledge [7]. In the same paper, they proposed to use independent component analysis (ICA) with phase correction for channel matrix estimation in order to achieve BMC. The ICA estimator is endorsed by the following publications but accompanied with different classifiers [8, 9]. Muhlhaus et al. proposed high order cumulants based likelihood ratio test classifier for low complexity BMC [8]. Kanterakis and Su suggest complexity reduction to the ALRT classifier by treating ICA recovered signal components at different transmitting antennas as individual processes [9].

In this paper, we propose a BMC solution which is more practical in a blind channel with both unknown channel matrix and unknow noise power. Given that pilot symbols are not available for the classifier, expectation maximization (EM) is adopted for non-data aided blind channel estimation. The initialization of EM is achieved using fuzzy c-means clustering and maximum likelihood mapping. Compared to the ICA estimator, the EM estimator provides the additional estimation of noise variance while not needing the phase correction for the channel matrix. The resulting estimate is used for the maximum likelihood (ML) classifier for decision making.

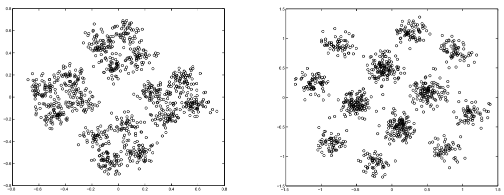
2. SIGNAL MODEL

The MIMO system is composed of N_t transmitting antennas and N_r receiving antennas. A Rayleigh fading channel with time invariant path gains is considered. The resulting channel matrix H is given by a $N_r \times N_t$ complex matrix with the element $h_{j,i}$ representing the path gain between ith transmitting antenna and jth receiving antenna. Assuming perfect synchronization, the nth received MIMO-SM signal sample vector $\boldsymbol{r}_n = [r_n(1), r_n(2), ..., r_n(N_r)]^T$ in a total observation of N samples is expressed as

$$\boldsymbol{r}_n = H\boldsymbol{s}_n + \omega_n \tag{1}$$

where $\boldsymbol{s}_n = [s_n(1), s_n(2), ..., s_n(N_t)]^T$ is the nth transmitted signal symbol vector and $\omega_n = [\omega_n(1), \omega_n(2), ..., \omega_n(N_r)]^T$ is the additive noise observed at the nth signal sample. The

Zhechen Zhu would like to thank the Department of Electronic and Computer Engineering, Brunel University for their financial support. Asoke K. Nandi would like to thank TEKES for their award of the Finland Distinguished Professorship. Professor Nandi is also a Distinguished Visiting Professor of Tongji University, Shanghai, China.



(a) Example 1 (b) Example 2

Fig. 1: Signal received at a receiving antenna in a 2x4 MIMO system.

transmitted symbol vector is assumed to be independent and identically distributed with each symbol assigned from the modulation alphabet with equal probability. The additive noise is assumed to be white Gaussian with zero mean and variance σ^2 which gives $\omega_n \in \mathcal{N}(0, \sigma^2 I_{N_r})$, where I_{N_r} is the identity matrix of size $N_r \times N_r$.

3. NON-DATA-AIDED INITIALIZATION

The channel matrix in a MIMO fading channel is rather complex given the number of paths and different channel effects. When EM was first proposed for MIMO channel estimation, the issue of initialization is achieved with data-aided methods [10]. However, the required pilot symbols may not be available in certain applications. The application of MC in military warfare, for example, requires the channel estimation to be performed in a non-cooperative environment.

3.1. Fuzzy C-means Clustering

For MIMO systems, the received signal at each receiver is a product of multiple transmission streams and the complex channel matrix. While the example in 1a shows clear separation between received symbols, another example of the received signal given in Figure 1b suggest that some of the received symbol states maybe very close to each other. For this reason, the fuzzy c-means algorithm is adopted. As the additive noise is assumed to be Gaussian, we incorporate the likelihood function of Gaussian process into the distance measurement and membership evaluation. The likelihood of sample $r_n(j)$ belonging to the cluster m is given by

$$\mathcal{L}\left(r_n(j), c_m\right) = \frac{1}{2\pi\sigma^2} e^{-\frac{(r_n(j)-c_m)^2}{2\sigma^2}} \tag{2}$$

where c_m is the mean of mth cluster. As the samples are received at the same antenna, the noise variance σ^2 is identical for all clusters. Therefore, we propose the membership calculation of the nth sample in the mth cluster using the following equation

$$u_{n,m} = \frac{1}{\sum_{i=1}^{M} e^{[(r_n(j)-c_m)^2-(r_n(j)-c_i)^2]}} \tag{3}$$

The centroid of the mth cluster is calculated using the mean of the all samples weighted by their membership.

$$c_m = \frac{\sum_n^N u_{n,m}^2 r_n(j)}{\sum_{i=1}^{M} u_{n,m}^2} \tag{4}$$

3.2. Initial Channel Estimate

After clustering, the membership set and clustered centroid set are used to provide the initial estimate of noise variance at the jth receiver and the channel coefficients associated with the jth receiver. The noise variance is calculated as

$$\hat{\sigma}_j^2 = \frac{\sum_{m=1}^{M}\sum_{n=1}^{N} u_{n,m}|r_n(j) - c_m|^2}{\sum_{m=1}^{M}\sum_{n=1}^{N} u_{n,m}} \tag{5}$$

For the initial estimation of the channel coefficient, the matter is a bit more complicated because neither the transmitted symbol vector nor the cluster centroids are ordered or matched up. The relationship between them can be modelled as.

$$c_m = [h_{j,1}, h_{j,2}, ..., h_{j,N_t}] \times {S_k}^T \tag{6}$$

where $S_k \in \mathbf{S}$ is one of the possible sample sets being transmitted with $k = 1, 2, ..., L^{N^t}$. The goal is to find the matching m and k so that the correct channel matrix could be estimated for EM initialization. In this paper, we have taken a semi-exhaustive likelihood based mapping approach to find the matching m and k and the subsequent initial channel matrix estimate. First, N_t number of estimated centroids are selected randomly. Correspondingly, all possible combinations of subset with N_t elements are constructed from **S**. For a MIMO system with N_t transmitter and modulation candidate with L symbol states, there are a total number of $I = L^{N^t}!/[N_t!(L^{N^t} - N_t)!]$ combinations. For each combination, there exist a set of channel matrix $h_{j,\cdot}(i)$ can be estimated from the selected centroids and the transmitted symbol set. To find the best match between the selected the transmitted sample combination subset, the likelihood value for each pairing can be calculated by

$$\mathcal{L}_i(r(j)|\sigma^2, h_{j,\cdot}(i)) = \prod_{n=1}^{N} \frac{1}{M} \sum_{m=1}^{M} \frac{1}{(\pi\sigma^2)^{N_r}} \exp\left(-\frac{\|r_n(j) - h_{j,\cdot}(i)S_m\|_F^2}{2\sigma^2}\right) \tag{7}$$

Using the maximum likelihood criterion, the final estimate of the channel matrix can be determined by find the maximum likelihood from all pairings.

$$\hat{h}(j,\cdot) = \underset{h_{j,\cdot}(i)\in h_{j,\cdot}(I)}{\text{argmax}} \mathcal{L}(R|\sigma^2, h_{j,\cdot}(i)) \tag{8}$$

This process is repeated for each receiver until the entire channel matrix is constructed.

4. EM CHANNEL ESTIMATION

In MIMO systems, we consider the received signal $R = [\boldsymbol{r}_1, \boldsymbol{r}_2 ... \boldsymbol{r}_N]$ as the observed data. Meanwhile, the membership Z of the observed samples is considered as the latent variables. Z is a $M \times N$ matrix with the (m,n)th element being the membership of the nth signal sample $\mathbf{r}_n$, given the transmitted symbol vector S_m. The possible transmitted symbol set $\mathbf{S} = [S_1, S_2 ... S_M]$ gathers all the combinations of transmitted symbols from N_t number of antennas. Given a modulation with L number of states, there exist $M = L^{N_t}$ number of transmitted symbol vectors and a transmitted symbol set of size $N_t \times L^{N_t}$. With $\Theta = \{H, \sigma^2\}$ representing the channel parameters, the expected value of the complete log-likelihood is derived as

$$\begin{aligned} \mathcal{Q}(R,S|\Theta_t) &= \log \prod_{n=1}^{N} \prod_{m=1}^{M} p(r_n, S_m | H_t, \sigma_t^2)^{z_{mn}} \\ &= -\sum_{n=1}^{N} \sum_{m=1}^{M} z_{mn} \left[N_r \log(\pi \sigma_t^2) + \frac{\|r_n - H_t S_m\|_F^2}{\sigma_t^2} \right] \end{aligned} \quad (9)$$

where $p(r_n, S_m | H^t, \sigma^t)$ is the probability of the nth received signal vector being observed given the current estimation of channel matrix H_t and noise variance σ_t^2. $\|\cdot\|_F^2$ is the Frobenius norm. The soft membership z_{mn} is evaluated using the following equation

$$z_{mn} = \frac{p(r_n | S_m, \Theta^t)}{\sum_{m=1}^{M} p(r_n | S_m, \Theta^t)} = \frac{\exp(-\frac{\|r_n - HS_m\|_F^2}{\sigma^2})}{\sum_{m=1}^{M} \exp(-\frac{\|r_n - HS_m\|_F^2}{\sigma^2})}. \quad (10)$$

4.1. Maximization Step

The update of the parameter estimation is achieved through the maximization of the current expected log-likelihood (M-step). To derive the close form update function for the channel matrix and noise variance, we first find the derivatives of $\mathcal{Q}(R,S|\Theta^t)$ with respect to H and σ^2 separately. The derivative of $\mathcal{Q}(R,S|\Theta^t)$ with respect to the individual element $h_(j,i)$ of the channel matrix is given by

$$\begin{aligned} &\frac{\partial \mathcal{Q}(R,S|\Theta^t)}{\partial h_{j,i}} \\ &= -\sum_{n=1}^{N} \sum_{m=1}^{M} z_{mn} \frac{\sum_{i=1}^{N_t} h_{j,i}{}^* |S_m(i)|^2 - r_n(j)^* S_m(i)}{\sigma^2} \end{aligned} \quad (11)$$

In the same way, the derivative of $\mathcal{Q}(R,S|\Theta^t)$ with respect to the noise variance σ^2 is found as

$$\frac{\partial \mathcal{Q}(R,S|\Theta^t)}{\partial \sigma^2} = -\sum_{n=1}^{N} \sum_{m=1}^{M} z_{mn} \left(-\frac{N_r}{\sigma^2} + \frac{\|r_n - HS_m\|_F^2}{\sigma^4} \right) \quad (12)$$

When the derivatives are set to zero, the update functions of $h_{j,i}$ and σ^2 can be derived from Equation (11) and (12). However, it is obvious that different channel parameters are coupled. To simplify the maximization process, the coupled channel parameters are estimated in turns. The path gain $h_{j,i}$ is estimated with the rest of the channel matrix known and represented with the latest estimate for each path gain. The path gains are updated in ascending order with respect to j and i. The resulting update function for $h_{j,i}$ is given by

$$\begin{aligned} &h_{j,i}^{t+1} \\ &= \frac{\sum_{n=1}^{N} \sum_{m=1}^{M} z_{mn} \left[r_n(j) S_m(i)^* - S_m(i)^* \sum_{k=1, k \neq i}^{N_t} h'_{k,i} S_m(k) \right]}{\sum_{n=1}^{N} \sum_{m=1}^{M} z_{mn} |S_m(i)|^2} \end{aligned} \quad (13)$$

where $h'_{k,i}$ is the lasted estimate of path gain $h_{k,i}$. At tth iteration, $h'_{k,i} = h^t_{k,i}$ if it has not been updated or $h'_{k,i} = h^{t+1}_{k,i}$ if it has been updated. After the channel matrix is completely updated, H_{t+1} is used to acquire the noise variance estimation.

$$\sigma_{t+1}^2 = \frac{\sum_{n=1}^{N} \sum_{m=1}^{M} z_{mn} \sum_{j=1}^{N_r} \left| r_n(j) - \sum_{i=1}^{N_t} h_{j,i}^{t+1} S_m(i) \right|^2}{N_r \sum_{n=1}^{N} \sum_{m=1}^{M} z_{mn}} \quad (14)$$

The EM algorithm with such maximization process is known as expectation conditional maximization (ECM). ECM shares the convergence property of EM [11] and can be constructed to converge at similar rate as the EM algorithm [12]. The ECM joint estimation of channel parameters has previously been successfully applied in BMC for SISO systems [13–15].

5. MAXIMUM LIKELIHOOD CLASSIFIER

For classification likelihood evaluation, average likelihood ratio test (ALRT) is adopted [7]. In the case of BMC, the channel matrix and noise variance estimated by EM is used to substitute the known values in the ALRT likelihood evaluation for each modulation hypothesis. The likelihood evaluation of modulation candidate $\mathcal{M}$ is given by

$$\begin{aligned} &\log \mathcal{L}(R|S_{\mathcal{M}} \Theta_{\mathcal{M}}) = -NN_t \log(M) - NN_r log(\pi \sigma_{\mathcal{M}}^2) \\ &+ \sum_{n=1}^{N} \log \left(\sum_{m=1}^{M} \frac{1}{(\pi \sigma_{\mathcal{M}}^2)^{N_r}} \exp(-\frac{\left\| r_n - \hat{H}_{\mathcal{M}} S_m^{\mathcal{M}} \right\|_F^2}{2\sigma_{\mathcal{M}}^2}) \right) \end{aligned} \quad (15)$$

where $S_{\mathcal{M}}$ is the transmitted symbol set defined by modulation $\mathcal{M}$ and $\Theta_{\mathcal{M}}$ is the channel estimation for the same modulation candidate. The resulting classification decision $\hat{\mathcal{M}}$ is

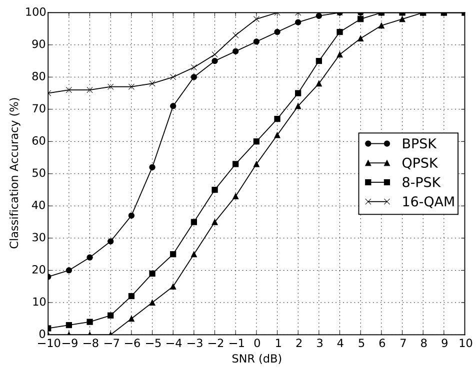


Fig. 2: The classification results of different modulation signals in Rayleigh fading channel with varying AWGN noise level and 512 samples from each transmitter.

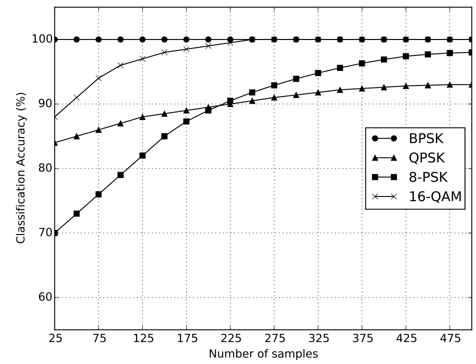


Fig. 3: The classification results of different modulation signals in Rayleigh fading channel with varying observation length and SNR at 5 dB.

found using the maximum likelihood criterion.

$$\hat{\mathcal{M}} = \underset{\mathcal{M}\in\mathfrak{M}}{\operatorname{argmax}}(\log \mathcal{L}(R|S_{\mathcal{M}}, \Theta_{\mathcal{M}})) \tag{16}$$

6. SIMULATION AND RESULTS

To validate the proposed BMC algorithm, MIMO systems in Rayleigh fading channel with AWGN noise is simulated for BMC. Four popular digital modulations are included in the modulation candidate pool: BPSK, QPSK, 8-PSK, 16-QAM.

First, 1,000 testing realizations of modulation signals are generated for each modulation candidate at SNR varying from -10 dB to 10 dB. Each signal realization consists of 512 observed signal samples at each receiving antenna. In Figure 2, classification results averaged over 1,000 realizations are listed. BPSK signals can be correctly classified with SNR above 3 dB. The performance degradation is slow between -3 dB and 3 dB. However, a dramatic decrease in classification accuracy is observed below -3 dB. The QPSK signals require higher SNR (above 8 dB) to achieve perfect classification with. The same pattern is observed for the 8-PSK signals with misclassification for both modulations accounted to 16-QAM. The classification result of 16-QAM concurs the biased behaviour of the classifier. The classification accuracy sees little degradation between -3 dB and 1 dB. Despite the decreasing level of SNR, the classification accuracy of 16-QAM signals remain at around 75%.

Second, 1,000 testing realizations of modulation signals are generated for each modulation candidate with signal length varying from 25 to 500. The SNR level is fixed at 5 dB in all experiments. The classification of BPSK is almost independent of the signal length. With only 25 samples from each receiving antenna, the classification of BPSK signals is able to achieve a 100% accuracy as shown in Figure 3. The robust performance for BPSK signal is mostly due to its lower modulation order. For QPSK, a very slow degradation can be observed with reduced signal length. Meanwhile, the degradation is rather moderate giving 84% classification accuracy with 25 sample at each receiving antenna. It is obvious that limited number of observed samples has a more significant impact on the classification of 8-PSK signal. It is often observed for high order modulations because of their denser symbol population. This, however, is contradicted by the classification accuracy of 16-QAM. The amount of degradation for successful classification of 16-QAM signal with shorter signal is minimal resulting a classification accuracy of 89% when given 25 samples for analysis. Given that the accuracy surpasses the other modulations, it is fair to conclude the classifier biased towards 16-QAM modulations.

7. CONCLUSION

A classifier with fuzzy c-means clustering initial channel estimation, expectation/conditional maximization channel estimation, and maximum likelihood classification is proposed.The employment of expectation maximization provides estimation of noise variance which is not enjoyed by the popular ICA estimator. The likelihood of each modulation candidate is evaluated with channel parameters estimated for the specific candidate. The classification of simulated signals in various settings shows that the classifier is able to provide excellent classification accuracy with SNR above 5 dB for BPSK, QPSK, 8-PSK, and 16QAM. In addition, the classification is robust even with the number of signal samples as low as 25 from each transmitter stream. Meanwhile, the classifier shows a biased character towards high-order modulations, especially 16-QAM modulation. Future research is expected to understand better the biased behaviour of the classifier as well as reducing its complexity for higher-order modulations and high dimension MIMO systems.

8. REFERENCES

[1] Z. Zhu and A. K. Nandi, *Automatic Modulation Classification: Principles, Algorithms, and Applications*. Hoboken, New Jersey: John Wiley & Sons, 2015.

[2] E. E. Azzouz and A. K. Nandi, *Automatic Modulation Recognition of Communication Signals*. Boston, MA: Kluwer, 1996.

[3] A. K. Nandi and E. E. Azzouz, “Algorithms for automatic modulation recognition of communication signals,” *IEEE Transactions on Communications*, vol. 46, no. 4, pp. 431–436, 1998.

[4] M. W. Aslam, Z. Zhu, and A. K. Nandi, “Automatic Modulation Classification Using Combination of Genetic Programming and KNN,” *IEEE Transactions on Wireless Communications*, vol. 11, no. 8, pp. 2742–2750, 2012.

[5] Z. Zhu, M. W. Aslam, and A. K. Nandi, “Genetic Algorithm Optimized Distribution Sampling Test for M-QAM Modulation Classification,” *Signal Processing*, vol. 94, pp. 264–277, Jan. 2014.

[6] Z. Zhu and A. K. Nandi, “Blind Modulation Classification for MIMO Systems using Expectation Maximization,” in *Military Communications Conference*. IEEE, 2014, pp. 1–6 (under review).

[7] V. Choqueuse, S. Azou, K. Yao, L. Collin, and G. Burel, “Blind Modulation Recognition for MIMO Systems,” *Military Technical Academy Review*, vol. XIX, no. 2, pp. 183–196, 2009.

[8] M. S. Mühlhaus, M. Öner, O. A. Dobre, and F. K. Jondral, “A Low Complexity Modulation Classification Algorithm for MIMO Systems,” *IEEE Communications Letters*, vol. 17, no. 10, pp. 1881–1884, 2013.

[9] E. Kanterakis and W. Su, “Modulation Classification in MIMO Systems,” in *Military Communications Conference*, 2013, pp. 35–39.

[10] A. Das and B. D. Rao, “SNR and Noise Variance Estimation for MIMO Systems,” *IEEE Transactions on Signal Processing*, vol. 60, no. 8, pp. 3929–3941, Aug. 2012.

[11] X.-L. Meng and D. B. Rubin, “Maximum Likelihood Estimation via the ECM Algorithm: a General Framework,” *Biometrika*, vol. 80, no. 2, pp. 267–278, 1993.

[12] J. Sexton, “ECM Algorithms That Converge at the Rate of EM,” *Biometrika*, vol. 87, no. 3, pp. 651–662, 2000.

[13] V. G. Chavali and C. R. C. M. da Silva, “Maximum-Likelihood Classification of Digital Amplitude-Phase Modulated Signals in Flat Fading Non-Gaussian Channels,” *IEEE Transactions on Communications*, vol. 59, no. 8, pp. 2051 – 2056, 2011.

[14] E. Soltanmohammadi and M. Naraghi-Pour, “Blind Modulation Classification over Fading Channels Using Expectation-Maximization,” *IEEE Communications Letters*, vol. 17, no. 9, pp. 1692–1695, 2013.

[15] V. G. Chavali and C. R. C. M. da Silva, “Classification of Digital Amplitude-Phase Modulated Signals in Time-Correlated Non-Gaussian Channels,” *IEEE Transactions on Communications*, vol. 61, no. 6, pp. 2408–2419, 2013.