Classification of Digital Linear and Nonlinear Modulations

Nathalie Thomas, Jean-Yves Tourneret, Anchalee Puengnim
Université de Toulouse - INPT - ENSEEIHT - TéSA - IRIT
Equipe Signal et Communications

A. Puengnim, "Classification de modulations linéaires et non-linéaires à l'aide de méthodes Bayésiennes", PhD Thesis, Institut National Polytechnique de Toulouse, France, Sept. 2008.

Introduction

Objective

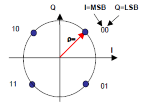
- Classify, or recognize, at the receiver, the type of modulation used by the transmitter.

Considered set of modulations

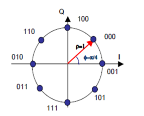
- Linear: Pulse Amplitude Modulation (PAM), Quadrature Amplitude Modulation (QAM), Phase Shift Keying (PSK), Amplitude and Phase Shift Keying (APSK)
- Nonlinear: Gaussian Minimum Shift Keying (GMSK)

Considered channel impairments

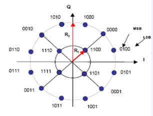
- Carrier phase and frequency offsets,
- Possible ISI coming from a residual channel,
- Rayleigh fading.

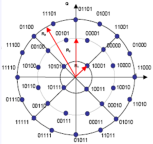


(a) QPSK



(b) 8PSK


(c) 16APSK


(d) 32APSK

Classification Strategies

Two main classes of modulation classification strategies

- **Statistical pattern recognition methods** based on feature extraction from the observations to be used for classification.
- **Decision-theoretic approaches** based on Bayes theory.
  - MAP (Maximum A Posteriori) classifier:

$$\text{assign } \mathbf{x} \text{ to } \lambda_i \text{ if } P(\lambda_i|\mathbf{x}) \geq P(\lambda_j|\mathbf{x}), \forall j = 1, ..., c,$$

$\mathbf{x} = [x(1), ..., x(N)]$: received symbol vector,
$\lambda_1, ..., \lambda_c$: set of possible modulations.

  - ML (Maximum Likelihood) classifier for equally-likely modulations ($P(\lambda_j) = \frac{1}{c}\ \forall j = 1, ..., c$):

$$\text{assign } \mathbf{x} \text{ to } \lambda_i \text{ if } p(\mathbf{x}|\lambda_i) \geq p(\mathbf{x}|\lambda_j), \forall j = 1, ..., c.$$

Two classifiers investigated in the PhD of Anchalee Puegnim

- The two classifiers are approximating the maximum likelihood classifier.
- One uses Markov Chain Monte Carlo (MCMC) methods, the other one relies on the forward/backward Baum-Welch (BW) algorithm.

ML classifier

Ideal case

$$x(k) = d_k + n(k), k = 1, ..., N$$

where $d_k \in \lambda_i = \{S_1^i, ..., S_{M_i}^i\}$, S_j^i being the j^{th} symbol among M_i of modulation λ_i and n is a Gaussian noise. The ML classifier can be rewritten as follows:

$$\text{assign } \mathbf{x} \text{ to } \lambda_i \text{ if } l(\mathbf{x}|\lambda_i) \geq l(\mathbf{x}|\lambda_j) \quad \forall j = 1, ..., c,$$

where

$$l(\mathbf{x}|\lambda_j) = \sum_{k=1}^{N} \ln \left\{ \frac{1}{M_j} \sum_{i=1}^{M_j} \exp\left(-\frac{1}{\sigma_n^2} \| x(k) - S_i^j \|^2 \right) \right\}.$$

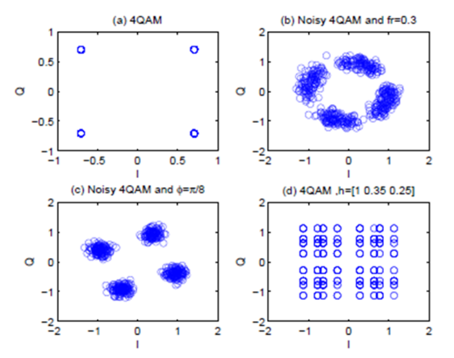
A more realistic case

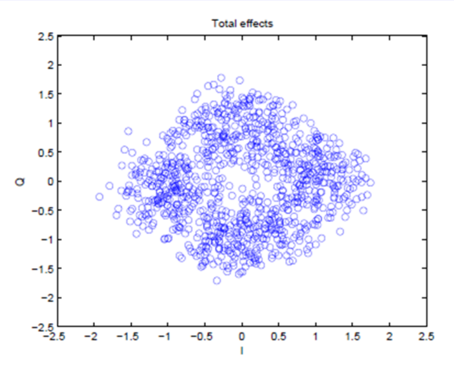
$$x(k) = e^{j(\pi \frac{k}{N} f_r + \phi)} \sum_{l=0}^{q} h_l d(k-l) + n(k), \quad k = 1, ..., N$$

where

- $f_r = 2N\Delta f_c \in [-1/2, 1/2]$ is the normalized residual carrier frequency offset,
- $\mathbf{h} = [1, h_1, ..., h_q]$ is the residual channel coefficient vector,
- ϕ is the carrier phase offset.

Channel Impairments for a QPSK Modulation





ML Classifier for Practical Applications

Channel impairment estimation

- Unknown parameter vector $\theta = [f_r, \mathbf{h}, \phi]$
- Two strategies to obtain $\widehat{\theta}$:
  - Using Markov Chain Monte Carlo (MCMC) methods.
  - Using the Baum Welch (BW) algorithm.

Modified decision rule

$$\text{assign } \mathbf{x} \text{ to } \lambda_i \text{ if } l(\mathbf{x}|\lambda_i, \widehat{\theta}) \geq l(\mathbf{x}|\lambda_j, \widehat{\theta}) \quad \forall j = 1, ..., c,$$

with

$$l(\mathbf{x}|\lambda_j, \widehat{\theta}) = \sum_{k=1}^{N} \ln \left\{ \frac{1}{M_j} \sum_{i=1}^{M_j} \exp\left(-\frac{1}{\sigma_n^2} \parallel x^{\widehat{\theta}}(k) - S_i^j \parallel^2 \right) \right\}.$$

where

$$x^{\widehat{\theta}}(k) = \mathfrak{F}^{-1}\left[x(k) e^{-j\left(\pi \frac{k}{N} \widehat{f}_r + \widehat{\phi}\right)} \right],$$

and $\mathfrak{F}^{-1}$ represents the inverse filter corresponding to $\widehat{\mathbf{h}} = [1, \widehat{h}_1, ..., \widehat{h}_q]$.

MCMC Parameter Estimation

MMSE estimator computed using samples generated with an MCMC method:

$$\widehat{\theta}_{\mathrm{MMSE}} = E\left[\theta|\mathbf{x}\right] = \int \theta p(\theta|\mathbf{x}) d\theta \simeq \frac{1}{L}\sum_{i=1}^{L} \theta^i.$$

where $(\theta^1, ..., \theta^L)$ are samples distributed according to $p(\theta|\mathbf{x})$ generated by running a Markov chain whose stationary distribution is $p(\theta|\mathbf{x})$.

Random-Walk Metropolis-Hasting algorithm

At each iteration, a candidate z is drawn according to an instrumental distribution $q(z|\theta^n)$[a]. This candidate is accepted with the following acceptance rule:

$$\theta^{n+1} = \begin{cases} z & \text{with probability } \alpha(\theta^n, z) \\ \theta^n & \text{with probability } 1-\alpha(\theta^n, z) \end{cases} \quad \text{where } \alpha(\theta^n, z) = \min\left\{1, \frac{p(z|\mathbf{x})q(\theta^n|z)}{p(\theta^n|\mathbf{x})q(z|\theta^n)}\right\},$$

and $\theta^n = (f_r^n, \phi^n, \mathbf{h}^n)$ represents the current Markov chain state.

Random-Walk: $q(z|\theta^n) = \mathcal{N}(\theta^n, \sigma^2) \Leftrightarrow z = \theta^n + \epsilon$, with $\epsilon \sim \mathcal{N}(0, \sigma^2)$.

[a] Any instrumental distribution $q(z|\theta^n)$ can be chosen, provided that the support of $p(z|\mathbf{x})$ is contained in the support of $q(z|\theta^n)$: see W.R. Gilks, S. Richardson and D.J. Spiegelhater, Eds London U.K. Chapman & Hall, 1996

Classification of Linear Modulations using MCMC Methods: Some Simulation Results

Simulation scenario

- Sets of modulations: $\lambda = \{\text{BPSK}, \text{QPSK}, \text{8PSK}, \text{16QAM}\}$ (studied in [Swami00][1]) and $\lambda = \{\text{QPSK}, \text{8PSK}, \text{16APSK}, \text{32APSK}\}$ (DVB-S2 standard),
- Transmission impairments: $\theta = [f_r, \mathbf{h}, \phi]$, with $\mathbf{h} = [1, h_1, h_2]$,
- 1000 trials belonging to each class λ_i and $N = 250$ observed symbols,
- For the MCMC sampler[2]: $L = 1500 +$ (500 burn-in) iterations, $q(z|\theta^n) \sim \mathcal{N}(\theta^n, \sigma^2)$ where $\sigma = 0.03$.

Compared classifiers

- ML classifier (labeled ML) derived assuming $f_r = \phi = 0$ and $\mathbf{h} = [1, 0, 0]$,
- MCMC classifier (labeled MCMC),
- Classifier derived in [Swami00], based on higher-order statistics (labeled HOS).

Classification performance

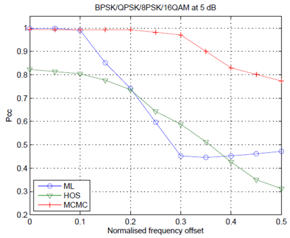
$$P_{\text{cc}} = \frac{1}{c} \sum_{i=1}^{c} P\left[\text{assigning } \mathbf{x} \text{ to } \lambda_i | \mathbf{x} \in \lambda_i\right].$$

---

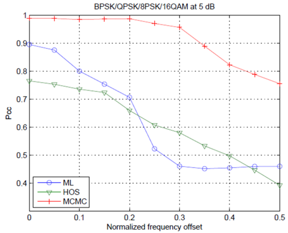
[1] [Swami00]: A. Swami and B. Sadler, Hierarchical Digital Modulation Classification Using Cumulants, IEEE Trans. Commun., vol. 48, no. 3, pp. 416 − 429, March 2000

[2] [Robert98]: C.P. Robert, Discretization and MCMC Convergence Assessment, Berlin: Springer-Verlag, 1998.

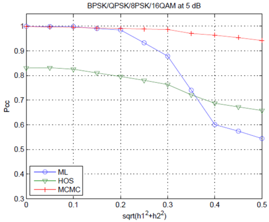
Probability of Correct Classification versus Channel Impairments



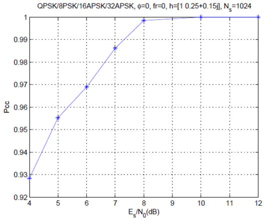
Average probability of correct classification versus f_r ($\mathbf{h} = [1, 0, 0]$, $E_s/N_0 = 5$ dB).



Average probability of correct classification versus f_r ($\mathbf{h} = [1, 0.25, 0.15]$, $E_s/N_0 = 5$ dB).



Average probability of correct classification versus residual channel modulus (only $\mathbf{h}$ is estimated).



Average probability of correct classification for the MCMC classifier versus E_s/N_0 in the presence of a residual channel

BW Parameter Estimation

The BW algorithm requires to associate a first order Hidden Markov Model to the received baseband communication signal: $x(k) = f\left(s(k)\right) + n(k)$

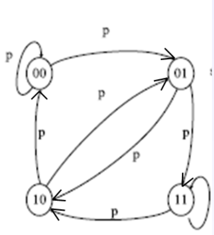
- A state of the HMM at time instant k : $s(k) \in \{s_1, s_2, ..., s_K\}$,
- A state transition probability defined by $d_{ij} = P[s(k+1) = s_j | s(k) = s_i]$
- An initial state distribution vector $\pi = (\pi_1, ..., \pi_K)^T$ defined by $\pi_i = P[s(1) = s_i] = 1/K$ for $i = 1, ..., K$.
- The pdf of the observation $x(k)$ conditioned on state s_i:

$$p_i[x(k)] \equiv p\left[x(k)|s(k) = s_i\right] = \frac{1}{\pi\sigma_n^2} \exp\left(-\frac{|x(k) - m_i|^2}{\sigma_n^2}\right) \text{ for } i = 1, ..., K.$$

where $m_i = f(s_i)$.

Example: BPSK modulation ($d(k) \in \{0, 1\}$, $M = 2$)**, two taps residual channel** ($x(k) = h_0 d(k) + h_1 d(k-1) + n(k), \quad k = 1, ..., N$)

- $s(k) = [d(k)d(k-1)] \in \{s_1 = 00, s_2 = 01, s_3 = 11, s_4 = 10\}$, $K = 4$
- $\mathbf{m} = [0,\ h_1,\ h_0 + h_1,\ h_0]$,
- $p = d_{ij} = P[s(k+1) = s_j | s(k) = s_i] = \frac{1}{M} = \frac{1}{2}$ if all symbols are equally likely.
- $\pi = (\pi_1, ..., \pi_K)^T$ defined by $\pi_i = P[s(1) = s_i] = 1/4$ for $i = 1, ..., 4$.



BW Parameter Estimation

Given the HMM, the BW algorithm[a] can be used:

- to determine the probability of the observation sequence given the modulation,
- to estimate the unknown parameters.

Forward Backward procedure

- Parameter initialization.
- Compute the normalized forward variable $\alpha_i(k) = P\left[\mathbf{x}_{1:k}, s(k) = s_i | \mathbf{m}, \sigma_n^2, \lambda_j\right]$:
  Initialization: $\alpha_i(1) = \pi_i p_i(x(1))$ for $1 \leq i \leq K$ and $c(1) = \left(\sum_{i=1}^{K} \alpha_i(1)\right)^{-1}$
  Induction for $k = 1, ..., N-1$, $j = 1, ..., K$:

$$\alpha_j(k+1) = c(k) p_j \left[x(k+1)\right] \sum_{i=1}^{K} \alpha_i(k) d_{ij}, \;\; c(k+1) = \left(\sum_{i=1}^{K} \alpha_i(k+1)\right)^{-1}$$

- Compute the normalized backward variable $\beta_i(k) = P\left[\mathbf{x}_{k+1:N}, s(k) = s_i | \mathbf{m}, \sigma_n^2, \lambda_j\right]$:
  Initialization: $\beta_i(N) = c(N)$ for $1 \leq i \leq N$
  Induction for $k = 1, ..., N-1$, $j = 1, ..., K$:

$$\beta_j(k) = c(k) \sum_{j=1}^{K} d_{ij} p_j \left[x(k+1)\right] \beta_j(k+1)$$

[a] [Rabiner89]: L. Rabiner, A tutorial on Hidden Markov Models and selected applications in speech recognition, Proc. IEEE, 77(2) : 257 − 286, February 1989

BW Parameter Estimation

- Parameter estimators:

$$\widehat{m}_i = \frac{\sum_{k=1}^{N} \gamma_i(k) x(k)}{\sum_{k=1}^{N} \gamma_i(k)}$$

$$\widehat{\sigma}_n^2 = \frac{1}{N} \sum_{k=1}^{N} \sum_{i=1}^{K} \gamma_i(k) |m_i - x(k)|^2$$

where $\gamma_i(k) = \alpha_i(k)\beta_i(k)$.

- Estimated probability of the observation sequence given the model:

$$\widehat{p}(\mathbf{x}|\boldsymbol{m}, \sigma_n^2, \lambda_j) = \frac{\sum_{i=1}^{K} \alpha_i(N)}{\sum_{i=1}^{N} c(i)}, \quad (1)$$

To be computed for each possible modulation $\lambda_j,\ j = 1, ..., c$ and used in the classification rule.

Classification of Linear Modulations: Some Simulation Results

Simulation scenario 1

- Set of modulations: $\lambda = \{\text{QPSK}, \text{8PSK}, \text{16APSK}\}$,
- Transmission impairments: $f_r = 0$, $\phi = 0$, $\mathbf{h} = [1, 0.35 + 0.33j]$,
- 1000 trials belonging to each class λ_i and $N = 250$ observed symbols.

Classification performance

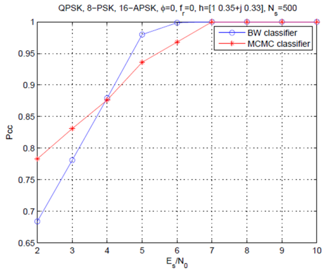


Fig. 14. Average probability of correct classification for MCMC and BW classifiers versus E_s/N_0 in the presence of a residual channel (DVB-S2 standard modulations).

Classification of Linear Modulations: Some Simulation Results

Simulation scenario 2

- Set of modulations: $\lambda = \{\text{QPSK}, \text{8PSK}, \text{16APSK}, \text{32APSK}\}$,
- Transmission impairments: $f_r = 0$, $\phi = 0$, $\mathbf{h} = [1]$,
- 1000 trials belonging to each class λ_i and $N = 250$ observed symbols.
- $E_s/N_0 = 10dB$, QPSK emitted constellation.

Computation time comparison using Matlab 7.4.0.287 (R2007a)

Tested constellation	QPSK	8-PSK	16-APSK	32-APSK
BW classifier	1.47 seconds	4.9 seconds	20.88 seconds	111.9 seconds
MCMC classifier	10.6 seconds	14.4 seconds	23.46 seconds	50.18 seconds

TABLE I

COMPUTATION TIME COMPARISON BETWEEN BW ET MCMC CLASSIFIERS - E_s/N_0=10 dB.

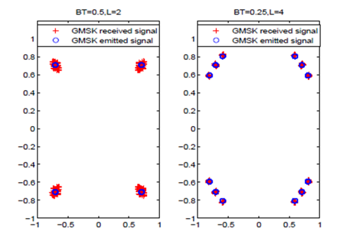
Classification of Linear and Non Linear Modulations using the BW Algorithm: Some Simulation Results

Simulation scenario 3

- Sets of modulations: $\lambda = \{\text{BPSK}, \text{QPSK}, \text{8PSK}, \text{GMSK25}, \text{GMSK50}\}$,

GMSK modulation: $x(k) = e^{j\Phi(kT,\mathbf{d})} + n(k)$

$\Phi(kT, \mathbf{d}) = \pi \sum_{i=k-L+1}^{k} d_i q((k-i)T) + \Phi_k$
$\phi_k = \left[\frac{\pi}{2} \sum_{i=-\infty}^{k-L} d_i\right] \text{mod}(2\pi)$
$q(t) = \int_{-\infty}^{t} g(\tau) d\tau$,
$g(t) = \frac{1}{2T} \left\{ Q\left(2\pi BT \frac{t-\frac{T}{2}}{T\sqrt{\ln 2}}\right) - Q\left(2\pi BT \frac{t+\frac{T}{2}}{T\sqrt{\ln 2}}\right) \right\}$,
where $Q(t) = \int_{t}^{\infty} \frac{1}{\sqrt{2\pi}} \exp(-\frac{\tau^2}{2}) d\tau$.



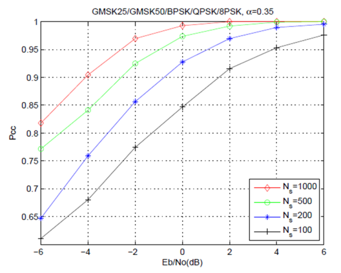
- Transmission impairments: $f_r = 0$, $\phi = 0$, $\mathbf{h} = [1]$,
- 1000 trials belonging to each class λ_i and $N \in \{100, 200, 500, 1000\}$ observed symbols.

Classification of Linear and Non Linear Modulations using the BW Algorithm: Some Simulation Results

Simulation scenario 3

- Sets of modulations: $\lambda = \{\mathrm{BPSK}, \mathrm{QPSK}, \mathrm{8PSK}, \mathrm{GMSK25}, \mathrm{GMSK50}\}$,
- Transmission impairments: $f_r = 0$, $\phi = 0$, $\mathbf{h} = [1]$,
- 1000 trials belonging to each class λ_i and $N \in \{100, 200, 500, 1000\}$ observed symbols.

Classification performance



In/Out	GMSK25	GMSK50	BPSK	QPSK	8PSK
GMSK25	449	51	0	0	0
GMSK50	13	487	0	0	0
BPSK	0	0	500	0	0
QPSK	0	0	0	498	2
8PSK	0	0	0	0	500

TABLE II

BW CLASSIFIER - CONFUSION MATRIX FOR E_b/N_0=0 dB.

In/Out	GMSK25	GMSK50	BPSK	QPSK	8PSK
GMSK25	334	164	1	0	1
GMSK50	123	375	0	1	1
BPSK	0	0	488	4	8
QPSK	0	0	0	313	187
8PSK	0	0	0	81	419

TABLE IV

BW CLASSIFIER - CONFUSION MATRIX FOR E_b/N_0=−6 dB.

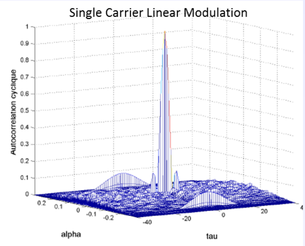
Future Work

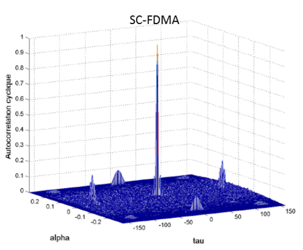
New classes of modulations

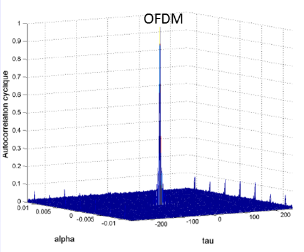
- OFDM, SC-FDMA, GFDM, Dual Talk signals...

New classification methods

- Use cyclostationarity ([Dobre08][3], [Datta14][4]) ?







Robustness with respect to other system parameters

- Symbol rate, synchronization, roll off factor, interferences...

---

[3] O. A. Dobre, Q. Zhang, S. Rajan and R. Inkol: second-order cyclostationnarity of cyclically prefixed single carrier linear digital modulations with applications to signal recognition. Global Telecommunications Conference (IEEE CLOBECOM) 2008.

[4] R. Datta, D. Panaitopol and G. Fettweis, Cyclostationary Detection of 5G GFDM Waveform in Cognitive Radio Transmission, Proc. of the Int. Conf. on Ultra Wideband (ICUWB) 2014.