ON-LINE BLIND EQUALIZATION OF FIR CHANNELS USING A GIBBSIAN TECHNIQUE

Sylvie Perreau
Alcatel CIT France email: perreau@sig.enst.fr

Pierre Duhamel
Telecom Paris 46 rue Barrault 75013 Paris France email:duhamel@sig.enst.fr

ABSTRACT

This paper is concerned with the blind equalisation of a communication system. We propose to realize the identification of the channel impulse response as well as the noise variance estimation and detection of the emitted sequence of symbols. The signal modelisation as a Hidden Markov Model (HMM) has the intrinsic potential for solving such a problem. However, the high performances of such methods are usually obtained at the cost of a high computational complexity and local minima problems. Both issues are addressed in this paper.

1. INTRODUCTION

Several blind equalization methods have been recently proposed in [3], [4],[1], relying on an Expectation-Maximisation (EM) sequential algorithm, classically used to approximate Maximum Likelihood estimates for incomplete data. However, several issues are raised by such a method. In particular, for high SNR levels, the algorithm sometimes converges to local minima. As a consequence, one has to cope with the initialization problem. Here, we propose a stochastic version of the EM algorithm, known as the Monte-Carlo EM algorithm [5]. The interest of a stochastic version for avoiding local minima is discussed, showing that the convergence rate is still very high. The second issue of the optimal methods relying on both EM algorithm and HMM formulation is the computational cost usually exponentionaly increasing with the length of the memory. We use the simplified (suboptimal) methods proposed in [4], [3], but relying on a Gibbs sampler method, classicaly used in image processing for image restoration [6]: the previous EM algorithms using the Gibbs sampler were iterative, and we show here that we can deal as well with a sequential EM algorithm.

2. PROBLEM FORMULATION

let y_t denote the observed signal which is the output of a FIR filter:

$$y_t = H^T X_t \ + \ n_t$$

where $H = (H(0), H(1), ..., H(N-1))^T$ denotes the unknown channel taps,and $X_t = (x_t, x_{t-1}, ...x_{t-N+1})$ is the state vector of the hidden Markov process, this vector containing all symbols stored in the channel memory at time t. x_t are taken from the set $\{-1; 1\}$. (However, generalisation to a different alphabet is straightforward). This state vector follows the state equation, relying on the shift property of the process X_t, A being a shift matrix.

$$X_t = A * X_{t-1} + x_t * [1\, 0 \cdots 0]' \qquad (1)$$

The additive noise process is assumed to be white, zero-mean, Gaussian, with unknown variance σ^2.
The objective is here to recover the transmitted data x_t, from the observations, as well as providing at each step an estimate of the channel impulse response and the noise variance.

3. PREVIOUS WORKS AND THE CONTRIBUTION OF THE PROPOSED METHOD

Few methods ([4], [1]) propose to use the so-called EM algorithm, which maximises the current expectation of the likelihood, given the available observations and the current parameters estimates. The E-step of this algorithm involved taking expectation of the log-likelihood over the stochastic process X_t:
$E\{log(P(y_t|X_t, H))|Y_t, \hat{H}^{t-1}\} =$
$\sum_k |y_t - \hat{H}_{t-1}^T \xi_k|^2 P(X_t = \xi_k|Y_t, \hat{H}^{t-1})$
This requires computing the probabilities
$p_k = P(X_t = \xi_k|y_1, \cdots, y_t)$ for every possible realisation ξ_k of X_t. This method is clearly very computationaly demanding, therefore, in [3], it has been simplified, using the marginal probablities $P(X_t^{(k)}|y_1, , \cdots, y_t, \hat{X}_{t|t-1}^{(j)}, \forall j \neq i)$ as intermediate values to approximate p_i ($X_t^{(k)}$ denoting the $k^{th}+1$ com-

ponent of X_t). Then, we can show ([3]) that for high SNR levels, the function performed during the E-step is approximated by:

$$L_t(H,\sigma^2) = \sum_{p=0}^{n} \lambda^{(n-p)}[-\frac{1}{2}log(\sigma^2) - \frac{1}{2\sigma^2}|y_p - H^T \hat{X}_{p|p}|^2] \quad (2)$$

Note that the above criterion still involves the mathematical expectation of the E-step through the definition of $\hat{X}_{p|p}$ (actually being Conditional Means (CM) estimate of X_p). The M-step consits in taking partial derivatives of $L_t(H,\sigma^2)$ according to H and σ^2, thus providing update equations for parameter estimation.

Our contribution is mainly to change this criterion into a stochastic one leading to the so-called Monte-Carlo EM (MCEM) algorithm, by approximating the expectation of the log-likelihood by a stochastic realization generated thanks to a Gibbs sampler technique ([6]), detailed below:

$$E\{log(P(y_t|X_t,H))|Y_t,\hat{H}^{t-1}\} \approx |y_t - \hat{H}_{t-1}^T X_t(\omega_i)|^2 \quad (3)$$

Where, $X_t(\omega_i)$ is sampled under the *a-posteriori* law $P(X_t|Y_1^t,\hat{H}_{t-1})$.

4. EMITTED SEQUENCE ESTIMATION USING A GIBBS SAMPLER

Assume available at time t:
-the current estimate of the channel impulse response $\hat{H}_t$
-an estimation of the noise variance $\hat{\sigma}_t^2$
-an estimate of the state vector X_{t-1}, denoted by $\hat{X}_{t-1|t-1}$. Therefore, because of the state eq. (1), a first prediction of the state vector X_t is chosen as $\hat{X}_{t|t-1} = A * \hat{X}_{t-1|t-1}$. Then a refined estimate of X_t is obtained as random samples, generated by a Gibbs sampler at time t.

Following the same idea as in ([3]), we approximate, as mentioned above the *a-posteriori* probability $P(X_t|Y_1^t,\hat{H}_{t-1})$ as:

$$P(X_t|Y_1^t,\hat{H}_{t-1}) \approx \prod_{i=1}^{N} P(X_t^{(i)}|y_1,,\cdots,y_t,\hat{X}_{t|t-1}^{(j)},\forall j \neq i) \quad (4)$$

Then, we estimate component by component the state vector $X_t = [x_t\ x_{t-1}\ \cdots x_{t-N+1}]'$. Let $\hat{X}_{t|t}^{(i)}$ denotes the $i^{th}+1$ component of the estimated vector $\hat{X}_{t|t}$. The Gibbs Sampler proceeds as follows:

- We first evaluate the probabilities (assuming the modulation is a $BPSK$:

$$\begin{aligned} P_1^{(0)} &= P(x_t = 1|\hat{x}_{t-i}, y_t, \hat{H}_{t-1}) \\ &= \frac{C}{\sqrt{2}\pi\sigma} \exp(-\frac{(y_t - \hat{H}_{t-1}^T * \mathcal{X}_t^{(0)}(\omega))^2}{2\sigma^2}) \quad (5) \end{aligned}$$

With $\mathcal{X}_t^{(0)}(\omega) = [1, \hat{x}_{t-1|t-1}(\omega), \cdots, \hat{x}_{t-N+1|t-1}]^T$ Where C is a constant normalisation. Also, we easily compute:

$$\begin{aligned} P_{-1}^{(0)} &= P(x_t = -1|\hat{x}_{t-i}, y_t, \hat{H}_{t-1}) \quad (6) \\ &= 1 - P_1^{(0)} \quad (7) \end{aligned}$$

- Let V denote the random variable uniformally distributed on the interval [0 1], and $V(\omega)$ a realisation of V. Then define the random variable $\hat{x}_t$ as:
  - $\hat{x}_t(\omega) = 1$ if $V(\omega) < P1$
  - $\hat{x}_t(\omega) = 1$ otherwise

  Clearly we have:

$$P(\hat{x}_t(\omega) = 1) = P_1^{(0)}$$

- Similarly, the update at time t of the estimation of the symbol x_{t-n+1} is performed after having evaluating:

$$P_1^{(n)} = P(x_{t-n+1} = 1|Y_1^t, \hat{H}_{t-1}) \quad (8)$$

These last quantities are the same ones as those already computed in ([3]) thanks to a sub-optimal Forward recursion of the **HMM formulation**:

$$\begin{aligned} P_1^{(n)} &= P(x_{t-n+1} = 1|Y_1^{t-1}, \hat{H}_{t-2}) \quad (9) \\ &* \frac{C}{\sqrt{2}\pi\sigma} \exp(-\frac{(y_t - \hat{H}_{t-1}^T * \mathcal{X}_t^{(n)}(\omega))^2}{2\sigma^2}) \quad (10) \end{aligned}$$

Where
$\mathcal{X}_t^{(n)}(\omega) = [\hat{x}_t(\omega), \cdots, \hat{x}_{t-n+2|t-1}(\omega), 1, \cdots, \hat{x}_{t-N+1|t-1}]^T$
$\hat{x}_{t-n+1|t}(\omega)$ is then evaluated in the same way as $\hat{x}_{t|t}(\omega)$.

As we draw the vector $\hat{X}_{t|t}$ component by component, the computational cost is linear with the channel memory as in [4], instead of exponential as when performing the optimal forward recursion.

5. INTERPRETATION OF THE STOCHASTIC METHOD

All probabilities mentioned above involve gaussian functions depending on the noise variance value σ^2. For reasonable SNR levels, these gaussian functions are sufficently peaked so that they can almost only take the values 1 or 0. Consequently, the symbol estimates sampled under the these probabilities will almost likely

take the value associated with the probability closed to unity, that is the same one as when performing a Maximum Aposteriori Probablity detection. As a consequence, the behavior of such a stochastic approximation turns out to be closed to the one of a usual EM algorithm.

The particular interest of dealing with a stochastic approximation is its potential ability of solving many local minima problems: the resulting excitment of the stochastic estimated process enables the algorithm to escape from the basin of attraction of a local minimum. However, we observed that the use of the true value of the noise variance leads to an algorithm, behaving like a deterministic one. It would be of interest then to include the noise variance in the parameters to be estimated by the EM algorithm: as long as this variance value is over-estimated, the stochastic feature is enhanced, thus preventing the algorithm to stay in a local minimum. We emphasize now on the estimation of the model parameters.

6. PARAMETERS ESTIMATES

The function to be minimized at time t by the stochastic EM algorithm is written as:

$$L_t(H,\sigma^2) = \sum_{p=0}^{n} \lambda^{(n-p)} [-\frac{1}{2} log(\sigma^2) - \frac{1}{2\sigma^2} |y_p - H^T \hat{X}_{p|p}(\omega)|^2] \quad (11)$$

λ being a forgetting constant. The minimization of this function is performed recursively by calculating its partial derivatives according to H and σ^2. This leads to the following update equations:

$$\hat{\sigma^2}_t = \frac{1}{1-\lambda^{t+1}} (\lambda(1-\lambda^t)\hat{\sigma^2}_{t-1} + (1-\lambda)|y_t - H^T \hat{X}_{t|t}(\omega)|^2 \quad (12)$$

and, for $\hat{H}_t$:

$$\hat{H}_{t+1} = \hat{H}_t + R_t^{-1}(y_t - \hat{H}_t^T \hat{X}_{t|t}(\omega))\hat{X}_{t|t}(\omega)^* \quad (13)$$

$$R_t = \lambda R_{t-1} + \hat{X}_{t|t}(\omega)\hat{X}_{t|t}(\omega)^\dagger \quad (14)$$

Here, the use of the forgetting constant λ is twofold: As far as the update of H is concerned, it allows to track slow variations of the channel as it in the case for mobile-communications. But it has also a particular interest as for estimating the noise variance: one can see in equation (12) that the estimated value performed at time $t-1$ is weighted at time t by λ, thus allowing the estimation process to "forget" the first estimates which are not reliable any more.

7. INTERESTS OF THE METHOD

Advantage of an HMM formulation for equalization: The use of the Forward variable allows the estimate on each symbol to be revisited as long as this symbol is seen by the channel memory, thanks to the informations provided by the whole set of available observations. This results in an improved BER. Note that the suboptimal formulation leading to a reduction of the complexity , does not result in a significant loss of performances as it was already the case in [3].

Advantage of the Gibbs Sampler method As pointed out in the introduction, this stochastic version of the EM algorithm is useful as for solving initialisation and local minima problems: the probabilities used in the Gibbs sampler decribed above are gaussian functions depending on current value of the estimate of σ^2. Note that in the actual algorithm, this noise variance is estimated during the M-step of the MCEM algorithm. As long as this variance estimate is much greater than the true one, the symbols estimates drawn by the Gibbs sampler are not reliable, this fact resulting in a persistent excitation on the adaptation process. As a consequence, the Fisher matrix has a better conditioned number compared the one of a standard EM algorithm as shown on figure 3. This results first in better convergence rate which is easily checked on figure 2. Also, the excitation is such as a local minimum cannot be a satisfactory solution. When the estimated likelihood reaches the bassin of attraction of some significant maximum, the noise variance estimate significantly decreases. Thus, the symbol estimates are similar to CM ones, and concequently, the MCEM algorithm almost behaves like the EM algorithm.

The resulting algorithm works on-line. The algorithm described in the full paper has an adaptive-like structure.

8. SIMULATIONS

The algorithm has been performed on both minimum and non-minimum phase channels, for a SNR of 13db. The first example illustrates the interest of the MCEM over the EM algorithm. We chose the following channel: $H = [1 \ -1 \ 0.5]$. Fig.1.1 depicts the trajectories of the algorithm in the map H(0), H(1). One can see that in this particular case, the EM algorithm converges to a local minimum ($\hat{H} = [1 \ -0.5 \ -0.5]$) for different initialisations (we show however a case where it does converge to the true parameters).

In constrast, one can see on Fig 1.2 that the MCEM algorithm converges to the global maximum, for the same initialisations, even initialised on the local minimum of the EM algorithm.

Fig 1.3 plots the noise variance estimate versus the number of iterations. One can note the accuracy of the final estimate value ($\sigma^2 = 0.0503$).

Fig 1.4 shows the estimate values of the channel param-

eters on a non-minimum phase channel $H = [0.5 - 1 0.3]$ (the 2 zeros of the corresponding transfer function being $z_1 = -1.36$ and $z_2 = 0.3$). Note the fast convergence (less than 500 symbols).

9. CONCLUSION

We proposed here an on-line algorithm, providing at each step an estimate of the impulse response of the channel and the noise variance thanks to a stochastic version of the standard EM algorithm (a Monte-Carlo EM alorithm), as well as a detection of the emitted sequence via a Gibbs sampler approach. The obtained algorithm takes advantage of the HMM properties without the drawback of the high computational complexity, and the stochastic version appears to be usefull for solving initialisation problems. The asymptotic behavior of the proposed algorithm can be shown to be similar to that of the standard EM algorithm. It has to be pointed out that this stochastic version classically used in the off-line contex, is also suitable for on-line processing.

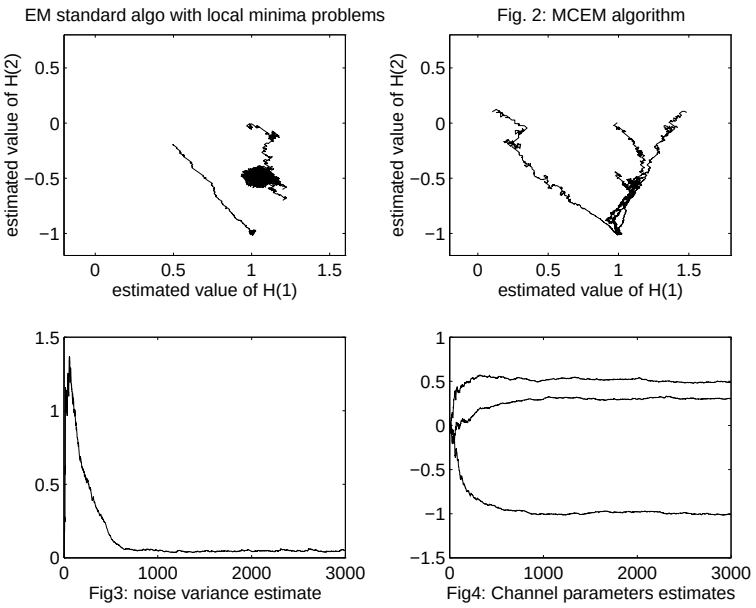

Figure 1. Fig 1.1, Fig 1.2, Fig 1.3, Fig 1.4

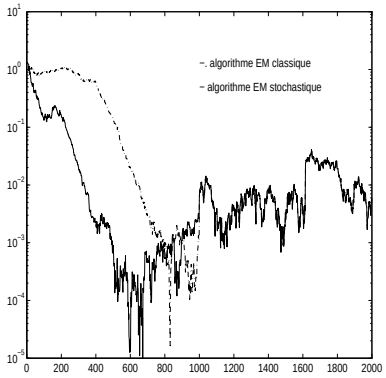

Figure 2. Quadratic Error on parameters, for a SNR of 15db

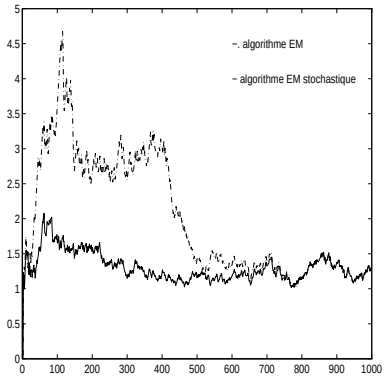

Figure 3. Fisher Matrix Conditioned number

REFERENCES

[1] V. Krishnamurthy and J. B. Moore *On-Line Estimation of hidden Markov Model Parameters Based on the Kullback-Leibler Information Measure* IEEE Trans. SP Vol.41 N08 August 1993

[2] A.P. Dempster, N.M. Laird and D.B Rubin *Maximum Likelihood from incomplete data via the EM algorithm* J.Roy.Soc Vol. 6 1977

[3] L.B. White, S. Perreau, P. Duhamel *Reduced Computation Blind Equalization for FIR channel input Markov Models* proceedings of ICC, June 1995, Seattle, USA

[4] L. B. White and V. Krishnamurthy *Adaptive blind equalization of FIR channels using Hidden Markov Models* IEEE ICC, Geneva 1993

[5] G. Wei and M.Tanner *a Monte-Carlo implementation of the EM algorithm and the poor's man's data augmentation algorithm*, J. amer Statist. Ass. vol 85, pp 699-704, 1990.

[6] Bernard Chalmond *an iterative Gibbsian Technique for reconstruction og m-ary images*, Pattern recognition, vol 22, pp 747-761, 1989.