An approach to blind channel identification by estimating the probability density function of received data

Jugurta Montalvão, Bernadette Dorizzi, and João Cesar M. Mota,

Abstract— **A new blind channel estimation approach is presented in this paper, based on the well known probability density function estimation by Kullback-Leibler's distance minimization. Thanks to a probability approximation, the resulting structure is surprisingly simple. Unfortunately, a complete analytic study of such a structure is quite difficult because of its nonlinear and recursive nature. Some approximated analytic expressions are presented along with some typical simulation results.**

Keywords— **Blind channel estimation, pdf estimation, joint equalization and estimation.**

I. Introduction

In the context of digital communication systems, concerning blind channel estimation by using synchronously sampled data, the *a posteriori* probability density function (pdf) of the sampled observations can be modeled by a parametric function, whose parameters are message and noise pdf and channel model. In the case of a finite alphabet modulation scheme, white Gaussian additive noise and a finite impulse response linear channel, the joint pdf of the sampled observations is in fact a mixture of Gaussians[2].

Then, assuming that the only available information is the message pdf and the noise variance, it is theoretically possible to (blindly) identify the channel model by classical methods such as maximum likelihood or maximum *a posteriori* [3], which in fact is the same as estimating the mixture of Gaussians with a parametric model. Nevertheless, it is well known that the number of Gaussian kernels in such a mixture grows exponentially with the channel memory and message length.

Some strategies have been proposed to cope with this drawback by splitting the entire set of observations into blocks supposedly statistically independent (e.g., Partial Likelihood [4], Split Data Likelihood [5], [6]).

As with the above strategies, the new algorithm presented in this paper also performs an adaptive estimation of a probability density function by means of the sampled observations. However, in contrast with them, we reduce the computational burden by estimating emitted symbols and feeding them back into the algorithm. This procedure is somehow closer to the algorithms proposed by N. Seshadri [7], but our final structure is much simpler, though it has a surprisingly good performance.

Indeed, the parametric pdf estimation approach applied leads us to a particular structure which, in spite of its apparent simplicity, has a good performance concerning the channel estimation task. Simulation results have shown fast convergence even for channels having spectral near-nulls, and fast but slightly biased estimations of channels having spectral nulls. However, since this algorithm involves past decisions, an analytic study of it seems to be quite difficult to obtain. Anyway, we have found some approximated analytic expressions which partially explain some observed results.

In Section II we present this simple channel estimator along with a brief comment on its link to parametric pdf estimation. In Section III we provide some approximate analytic expressions concerning the stable points in the cost function for 2-PSK and 4-QAM modulation schemes. Finally, in Section IV we present some simulation trials with the 4-QAM modulation scheme. These trials were chosen because they illustrate well typical results when the channel presents spectral near-nulls.

II. Channel model and pdf estimation

In this paper, we consider a finite impulse response (FIR) linear channel model, a digital information source, with $a(n)$ standing for a discrete and complex random variable whose variance is σ_a^2, and $a(n), \ldots, a(0)$ being an i.i.d. stream which carries digital data. Furthermore, digital symbols are drawn with equal probability from a finite alphabet $\{a_s : 1 \leq s \leq S\}$. Similarly, let $b(n)$ be an additive and Gaussian noise with variance σ_b^2, and $b(n), \ldots, b(0)$ an i.i.d. set of variables. Finally, $F(z) = \sum_{i=0}^{N-1} f_i z^{-i}$ is the z-transform of the channel impulse response. The channel model is therefore an FIR filter with N taps. We can alternatively represent this filter by the vector: $\mathbf{f} = [f_0 \ f_1 \ \cdots \ f_{N-1}]^T$. The filter output is a random variable $x(n)$, which models observations synchronized to the symbol rate.

In this section, we show that the pdf of $x(k)$ $(0 \leq k \leq n)$, given the past observations, is a mixture of Gaussians, parameterized both by $\mathbf{f}$ and σ_b^2. After that, we show how the application of an equalizer (any equalizer) makes possible an approximation of such a mixture by just one Gaussian at a time. As a consequence, it provides a drastic simplification of the channel estimation algorithm. This algorithm being, in fact, an adaptive matching of the conditional pdf

J. Montalvão and B. Dorizzi are with the EPH Department, "Institut National des Télécommunications (INT)," 91011 Evry, France (E-mail: jugurta.montalvao@int-evry.fr; Bernadette.Dorizzi@int-evry.fr).

Jugurta Montalvão is also with the "Tiradentes" University, Aracaju, Brazil.

J.C. Mota is with the Federal University of Ceará (UFC), Fortaleza, Brazil. E-mail: mota@dee.ufc.br.

of $x(k)$ with another suitable multimodal function, parameterized by estimates of $\mathbf{f}$ and σ_b^2 (i.e. , $\hat{\mathbf{f}}$ and $\hat{\sigma}_b^2$, respectively).

Let $l(0) = p(x(0))$ be the pdf of $x(0)$ and $l(k) = p(x(k)|x(k-1),\ldots,x(0))$ the conditional pdf of $x(k)$. Therefore

$$p\left(x(n),\ldots,x(0)\right) = \prod_{k=0}^{n} l(k) \qquad (1)$$

On the other hand, given an integer $\hat{N}(\geq N)$ and a discrete random vector $\mathbf{a}(n) = [a(n) \ \cdots \ a(n-\hat{N}+1)]^T$, then $l(k)$ can be expanded to:

$$l(k) = \sum_{i=1}^{S^{\hat{N}}} p(x(k)|\mathbf{a}(k)=\mathbf{a}_i)\Pr\left(\mathbf{a}(k)=\mathbf{a}_i|x(k-1),\ldots,x(0)\right)$$

where S is the number of symbols used in the modulation scheme. For instance, our simulations up to now were based on 2-PSK ($S=2$) and 4-QAM ($S=4$) schemes.

Since the noise is Gaussian, it is easy to show that

$$l(k) = \frac{1}{\sqrt{2\pi\sigma_b^2}} \sum_{i=1}^{S^{\hat{N}}} \phi_i\left(x(k)\right)\Pr\left(\mathbf{a}(k)=\mathbf{a}_i|x(k-1),\ldots,x(0)\right) \quad (2)$$

where

$$\phi_i\left(x(k)\right) = \frac{1}{\sqrt{2\pi\sigma_b^2}}\exp\left(-\frac{1}{2\sigma_b^2}\left|x(k)-\sum_{j=0}^{N-1} a_i(n-j)f_j\right|^2\right) \qquad (3)$$

Clearly, the number of terms in the summation exponentially grows with $\hat{N}$ and S, and it could be considered as a serious drawback concerning practical utilization of such an approach.

In order to get round this obstacle, we apply the following strategy: given an equalizer (whichever it may be) providing symbol estimates modeled by the random variable $\hat{a}(n-d)$, where d is a suitable decision delay; we use a vector of such estimates to reduce the number of terms in Eq. 2 to only one term. It is feasible by means of the (arbitrary) approximation

$$A_0\text{: } \Pr\left(\mathbf{a}(n-d)=\mathbf{a}_i|x(n-d),\ldots,x(0)\right) \cong \begin{cases} 1 \text{ if } \mathbf{a}_i = \hat{\mathbf{a}}(n-d) \\ 0 \text{ otherwise} \end{cases}, \forall i \in \left\{1,\cdots,S^{\hat{N}}\right\}$$

where $\hat{\mathbf{a}}(n-d) = [\hat{a}(n-d) \ \cdots \ \hat{a}(n-d-\hat{N}+1)]^T$ is a random vector of past estimates provided by the equalizer[1]. Consequently, applying A_0 in (2), we have

$$l(n-d) \cong \frac{\exp\left(-\frac{1}{2\sigma_b^2}\left|x(n-d)-\sum_{i=0}^{N-1}\hat{a}(n-i)f_i\right|^2\right)}{\sqrt{2\pi\sigma_b^2}}$$

[1]Note that A_0 is somehow related to the clustering procedure applied in the classical K-means algorithm [8]

It is worth noting that the more effective the equalizer, the more A_0 is likely.

Then, we are able to formulate a parametric estimator for $l(n-d)$ as follows:

$$\hat{l}(n-d;\hat{\mathbf{f}},\hat{\sigma}_b^2) = \frac{1}{\sqrt{2\pi\hat{\sigma}_b^2}}\exp\left(-\frac{1}{2\hat{\sigma}_b^2}\left|x(n-d)-\hat{\mathbf{a}}(n-d)^T\hat{\mathbf{f}}\right|^2\right)$$

where $\hat{\mathbf{f}}$ and $\hat{\sigma}_b^2$ are respectively the channel and noise variance estimates. As a consequence of (1), we also have:

$$p\left(x(n-d),\ldots,x(0)\right) \cong \prod_{k=0}^{n-d}\hat{l}(k)$$

A suitable cost function to compare pdfs is the Kullback-Leibler (KL) divergence [9]:

$$J = E_{x(n-d),\ldots,x(0)}\left\{\ln\frac{p(x(n-d),\ldots,x(0))}{\prod_{k=0}^{n-d}\hat{l}(k;\hat{\mathbf{f}},\hat{\sigma}_b^2)}\right\}$$

The gradient of J w. r. t. $\hat{\mathbf{f}}$ is then given by

$$\nabla_{\hat{\mathbf{f}}}J = \frac{1}{\hat{\sigma}_b^2}\frac{\partial J}{\partial\hat{\mathbf{f}}^*}\sum_{k=0}^{n-d}E\left\{\left|x(k)-\hat{\mathbf{a}}(k)^T\hat{\mathbf{f}}\right|^2\right\}$$

where $\hat{\mathbf{f}}^*$ is the complex conjugate of $\hat{\mathbf{f}}$. Note that this result can be directly obtained from A_0 and the well known equivalence between likelihood maximization and squared-error minimization for Gaussian process with uniform parameters priors.

A stochastic minimization of J is probably the simplest way to adapt $\hat{\mathbf{f}}$. Then, a simple channel estimation algorithm is:

$$\hat{\mathbf{f}}(n) = \hat{\mathbf{f}}(n-1)+\alpha_f\left(x(n-d)-\hat{\mathbf{a}}(n-d)^T\hat{\mathbf{f}}(n-1)\right)\hat{\mathbf{a}}(n-d)^*$$

where α_f is a suitable adaptation step.

Finally, we need to choose an equalizer in order to provide the symbol estimates. Clearly, more powerful structures such as Bayesian nonlinear devices should reduce the decision error rate [10], and thus to render A_0 more likely. Nevertheless, in order to keep our structure as simple as possible, we are going to use the Wiener linear transversal equalizer (LTE). Besides, in most realistic cases, where the Wiener equalizer provides a relatively low symbol error rate, the approximation A_0 still holds.

The final estimator structure is shown in Fig. 1, where both $\hat{\mathbf{R}}_x(\hat{\mathbf{f}})$ and $\hat{\mathbf{p}}(\hat{\mathbf{f}})$ are analytically computed from $\hat{\mathbf{f}}$. In other words, the equalizer coefficients are analytically calculated for every new estimate $\hat{\mathbf{f}}$. Furthermore, the estimated noise variance, which is applied to compute $\hat{\mathbf{R}}_x(\hat{\mathbf{f}})$, must be available. However, we have observed that the

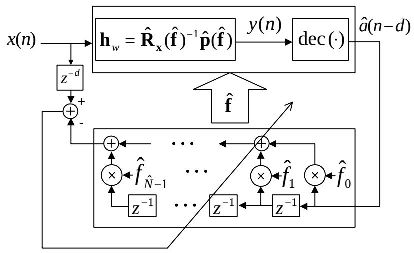

Fig. 1. Channel coefficients estimator

algorithm is almost insensitive even to great deviations between $\hat{\sigma}_b^2$ and σ_b^2. Therefore, in our simulations we have used a moderate but arbitrarily chosen $\hat{\sigma}_b^2$ just to ensure that $\hat{\mathbf{R}}_x(\hat{\mathbf{f}})$ is a well conditioned matrix, whatever the value of $\hat{\mathbf{f}}$. For instance, we may set $\hat{\sigma}_b^2 \approx 0.05$.

III. Convergence study

The resulting structure is surprisingly simple and though it is clear that it can work when past decisions are correct, on the other hand, even a superficial glance at Fig. 1 gives rise to the following question: how does the algorithm work when $\hat{\mathbf{f}}$ is randomly initialized ?

Strictly speaking, this question could be answered analytically but, given the joint effect of recursivity and nonlinearity present in the algorithm, an analytic approach seems to be quite difficult. So, our strategy to get some feeling about this question was two-fold: first we have done hundreds of simulation trials, with different channels and noise variances (we also have done simulations with colored noise), and finally an approximate analytic investigation was done in order to roughly explain the most representative results.

Here we briefly present the main findings of this investigation:

a. Every $\hat{N}$-dimensional vector

$$\hat{\mathbf{f}}_0 = \exp(\phi\sqrt{-1})[0 \;\cdots\; 0 \; f_0 \;\cdots\; f_{N-1} \; 0 \;\cdots\; 0]^T$$

corresponds to a minimum of the cost function J if $\phi \in \{\ldots, -2\pi, -\pi, \pi, 2\pi, \ldots\}$ for 2-PSK schemes, and $\phi \in \{\ldots, -\pi, -\pi/2, \pi/2, \pi, \ldots\}$ for 4-QAM schemes. In fact, these points will be referred to in this text as desired solutions to the blind estimation task.

b. For channels having spectral nulls, the stable equilibrium point corresponding to solutions $\hat{\mathbf{f}}_0$ are slightly biased to $\hat{\mathbf{f}}_b = \hat{\mathbf{f}}_0 + \Delta\hat{\mathbf{f}}_0$. It is possible to show that this bias depends on the modulation scheme and the equalizer performance. For instance, if the symbol error rate (SER) at the Wiener equalizer output is $SER(\hat{\mathbf{f}}_b)$, then $\Delta\hat{\mathbf{f}}_0 \cong -2SER(\hat{\mathbf{f}}_b)\hat{\mathbf{f}}_0$ in the 2-PSK case and $\Delta\hat{\mathbf{f}}_0 \cong -(4/3)\,SER(\hat{\mathbf{f}}_b)\hat{\mathbf{f}}_0$ in the 4-QAM case.

c. In both the above mentioned cases, the cost function roughly lies in $\max\left(\sigma_b^2, 4SER(\hat{\mathbf{f}}_b)\hat{\mathbf{f}}_b^H\hat{\mathbf{f}}_b\right) \leq J(\hat{\mathbf{f}}_b) \leq \sigma_b^2 + 4SER(\hat{\mathbf{f}}_b)\hat{\mathbf{f}}_b^H\hat{\mathbf{f}}_b$.

d. Apart from the desired cost function minima, spurious local minima were also found but only in simulations with 2-PSK. These minima were related to a particular class of channel estimates given by: $\hat{F}_{lm}(z) \cong F(z)\cdot P(z)$ with the coefficients of $P(z)$ having the property: $\sum_i p_i^2 \gg \sum_i p_i p_{i+m}, \forall m \neq 0$. In other words, $P(z)$ is an all-pass filter.

IV. Simulation illustrations

The following simulation trials were done with the same channel found in reference [1]. This channel presents two spectral near nulls, causing difficulties in the estimation task. In [1] the channel estimation was used to compensate for intersymbol interference (ISI) by means of a linear transversal equalizer (LTE). As a result, an ISI reduction was obtained by using a nonlinear least squares estimation algorithm, based on a high-order statistics (HOS) approach.

In contrast to [1], where an ISI reduction of $-17dB$ was obtained by using $20,000$ observed samples, we have obtained the same ISI reduction with fewer than $5,000$ samples (see illustration in Fig. 2).

Further results are presented in the sequel:

- Trial without noise: Fig. 2 shows a typical simulation trial where $\hat{\mathbf{f}}$ converges toward $\hat{\mathbf{f}}_0$. In these cases, the estimator variance asymptotically converges to zero. Note that simulations were done with random initialization of $\hat{\mathbf{f}}$ in order to highlight the self-adaptation capability of the algorithm. Nevertheless, faster convergence has been obtained by initializing $\hat{\mathbf{f}}$ with 1 in the middle vector position, and zeros elsewhere.
- Trial with additive white Gaussian noise: Fig. 3 shows a typical simulation trial with additive white Gaussian noise.
- Trial with additive colored noise: In order to illustrate a case where the noise is no longer white Gaussian, we put together five independent sources, as shown in Fig. 4.
  Strictly speaking, our formulation for $\phi_i\,(x(k))$, in Eq.(3), is no longer valid. However, in almost all trials where the channel corresponding to the user with strongest energy (user number 0) had no deep spectral fading, we had convergence to $\hat{\mathbf{f}}_0$ in fewer than 5000 symbols.

V. Conclusions

We have presented a very simple structure which works surprisingly well, providing fast channel estimation even for arbitrarily chosen initial values of the channel estimated. Though it is not a straightforward matter explaining the transient behavior of such a “nonlinear feedback” structure, we have observed in the majority of the simulation trials a fast convergence toward the desired channel estimated. Moreover, no local minima were found in simulations with 4-QAM. On the other hand, the few undesirable local minima found in simulations with 2-PSK match the case characterized in Section III. Furthermore, we also ob-

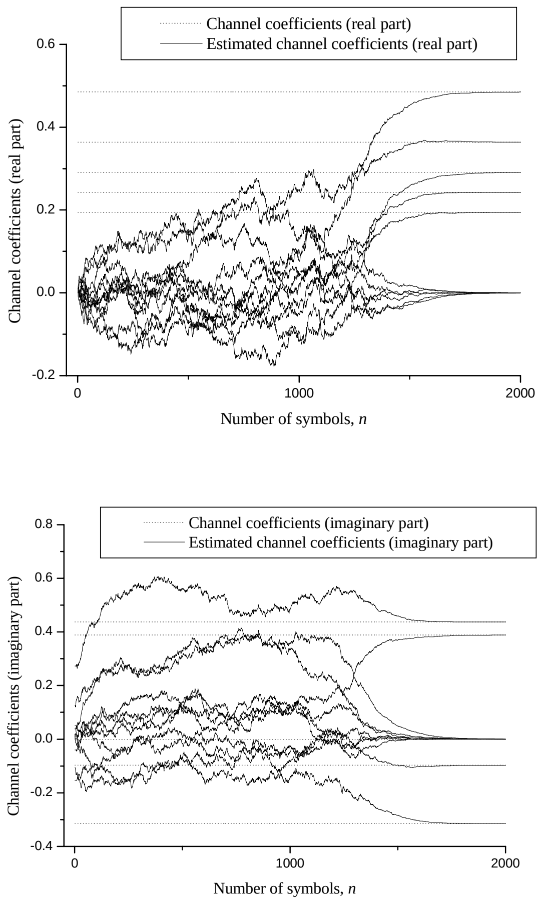

Fig. 2. A single trial with 4-QAM modulation scheme, where $F(z) = (2-0.4j)+(1.5+1.8j)z^{-1}+1z^{-2}+(1.2-1.3j)z^{-3}+(0.8+1.6j)z^{-4}$, 21 taps in the LTE, $N=5$, $\hat{N}=11$, $SNR \rightarrow \infty$ and $\alpha_f = 0.005$.

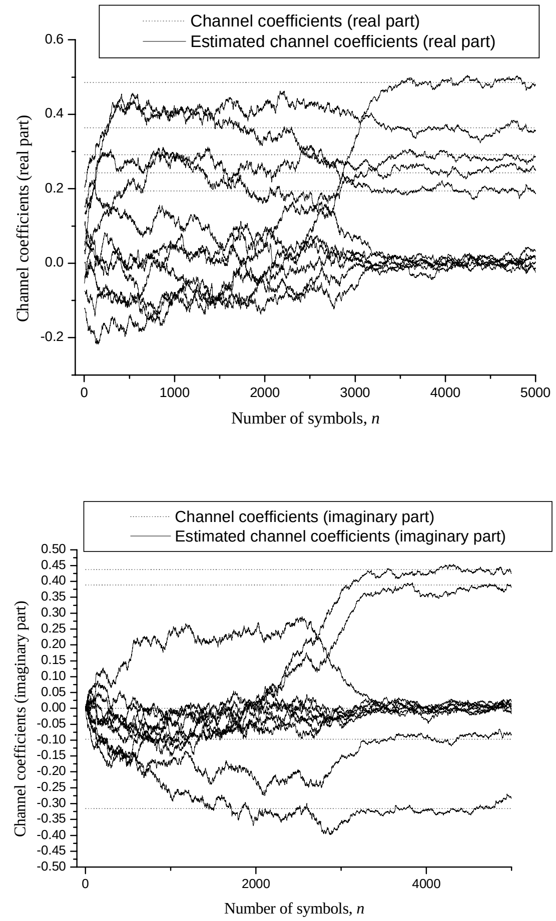

Fig. 3. A single trial with 4-QAM modulation scheme, where 41 taps in the LTE, $N=5$, $\hat{N}=11$, $SNR=12dB$ and $\alpha_f = 0.003$.

served that the stronger the noise, the rarer the local minima incidence, which could suggest some ideas concerning strategies to avoid such a problem.

Concerning the estimation bias referred to in Section III, observed when the channel had spectral nulls, since this bias is related to the SER, it is possible to reduce it by replacing the Wiener LTE by, for example, a nonlinear equalizer with better performance. Specifically, we performed some trials with the suboptimal Bayesian equalizer proposed in [10] and the estimate deviation was reduced, according to the approximate formula presented in Section III, Item b, which associates such a deviation to SER.

VI. Acknowledgment

The authors thank Dr. Eleftherios Kofidis for his valuable suggestions. We also thank the INT-Evry/France and the CNPq/Brazil for the partial financial support.

References

[1] Porat B. and B. Friedlander, "Blind Equalization of Digital Communication Channels using High-order Moments," IEEE Trans. on Signal Processing, VOL. 39, NO. 2, pp. 522-526, 1991.

[2] A.F.M. Smith, D.M. Titterington and U.E. Makov, *Statistical Analysis of Finite Mixture Distributions*, John Wiley & Sons, USA, 1985.

[3] S.M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice Hall, 1993.

[4] T. Adali, X. Liu and M. Kemal Sönmez, "Conditional Distribution Learning with Neural Networks and its Application to Channel Equalization", *IEEE Trans. On Signal Processing* Vol. 45, No. 4, pp. 1051–1064, 1997.

[5] E. Moulines, J. Cardoso, E. Gassiat, "Maximum Likelihood for Blind Separation and Deconvolution of Noise Signals using Mixture Models", *Proc. ICASSP-97*, Munich, Germany, pp. 3617–3620, 1997.

[6] T. Rydén, "Consistent and asymptotically normal parameter estimates for hidden markov models", *The Annals of Statistics*, vol. 22, no. 4, pp. 1884–1895, 1994.

[7] S. Haykin (Editor), *Blind Deconvolution*, Information and systems sciences series. Prentice-Hall, Englewood Cliffs, NJ, USA, 3 edition, 1994.

[8] R.O. Duda and P.E. Hart, *Pattern Classification and Scene Analysis*, Wiley, New York, USA, 1973.

[9] Bishop C.M., *Neural Networks for Pattern Recognition*, Clarendon Press, Oxford, 1995.

[10] Montalvão, J., B. Dorizzi, J. C. M. Mota, "A Family of Nonlinear Equalizers: Sub-Optimal Bayesian Classifiers," IEEE-EURASIP (NSIP'99), Antalya, Turkey, 1999.

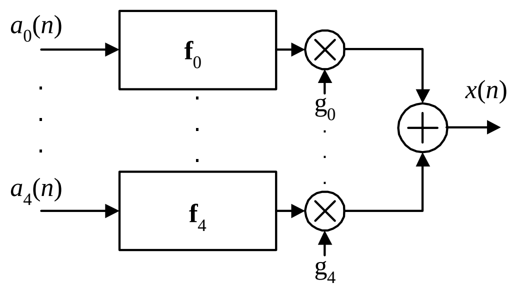

Fig. 4. Multi-user interference model where all sources have the same variance σ_a^2, g_i are real gains and the signal-to-interference ratio (SIR) is defined by $SIR = 10\log\left(g_0^2\mathbf{f}_0^H\mathbf{f}_0/\sum_{i=1}^4 g_i^2\mathbf{f}_i^H\mathbf{f}_i\right)$.

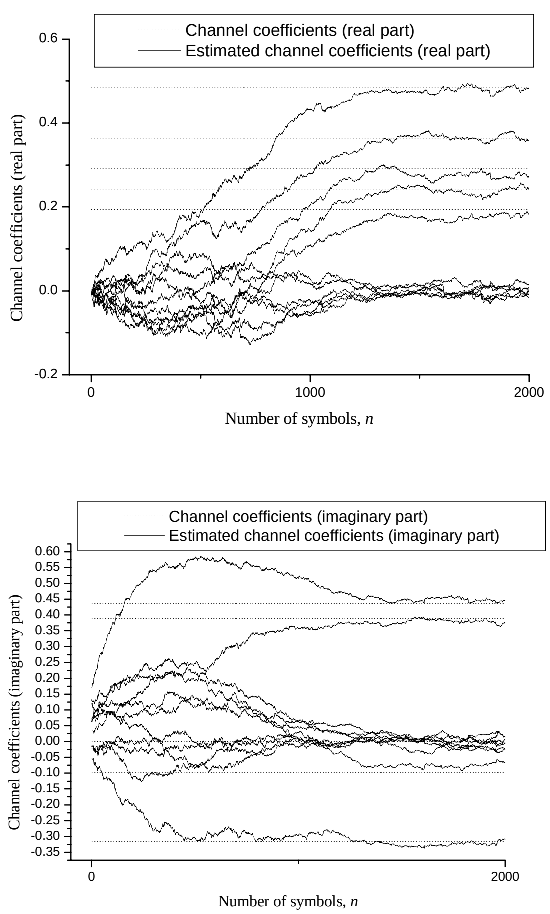

Fig. 5. A single trial with 4 interfering users and a 4-QAM modulation scheme, 41 taps in the LTE, $N = 5$, $\hat{N} = 11$, $SIR = 12dB$ and $\alpha_f = 0.003$.