ON THE PERFORMANCE OF THE VITERBI DECODER WITH TRAINED AND SEMI-BLIND CHANNEL ESTIMATORS

Alexei Gorokhov

CNRS-LSS, École Supérieure d'Electricité
plateau de Moulon, 91192 Gif-sur-Yvette, France
E-mail: gorokhov@lss.supelec.fr

ABSTRACT

Maximum-likelihood sequence estimation is often used to recover digital signals transmitted over finite memory convolutive channels when an estimate of the channel is available. In this letter, we study the impact of channel estimation errors on the quality of sequence detection. The general case of single input multiple output (SIMO) channels is considered. An asymptotic upper bound for the symbol error rate is presented which allows to treat channel estimation errors as equivalent losses in signal-to-noise ratio (SNR). This relationship is studied and numerically validated for the standard least squares channel estimate and for the semi-blind estimator which makes use of the empirical subspace of the observed data.

1. INTRODUCTION

The optimal detection of digital sequences in presence of intersymbol interference (ISI), caused by narrowband pulse shaping and multipath propagation, may be achieved by the recursive nonlinear maximum likelihood (ML) estimator, called the Viterbi algorithm [1]. The use of a ML procedure is necessary in the environments with severe ISI and low SNR, see [2]. The performance of Viterbi detectors in presence of additive white Gaussian noise (AWGN) has been studied by Forney [3]. This and later contributions [4] assume perfect knowledge of the channel between the emitted data and the received signal. The present study of the ML detector with imperfectly known channel has been motivated by the recent emergence of various channel estimation techniques. Although different measures of the channel estimation accuracy have been proposed (*e.g.*, mean square errors of the channel coefficients or linearly reconstructed signals), none of them reflect the impact of estimation errors on ML detection. However, an analysis of ML detectors in the presence of channel estimation errors, which enables comparison of different estimators in the context of data communications, is highly desirable. In this letter, we show that the effect of estimation errors may be accurately approximated as a loss in SNR.

This study was partially supported by the SASPARC project of INTAS.

2. BACKGROUND

Assume that the overall SIMO channel between the emitted data and signals observed at system outputs (*e.g.*, multiple receiving antennas or/and multiple signal polyphases in the case of fractional sampling) admits a rather accurate FIR approximation. Let $x(t)$ be the M-variate complex output noisy series, then

$$x(t) = \sum_{\tau=0}^{L} \mathbf{h}(\tau)\, s(t-\tau) + n(t), \qquad (1)$$

where $s(t)$ is the emitted sequence, $n(t)$ are the samples of an AWGN and $\mathbf{h}(0), \ldots, \mathbf{h}(L)$ is the $M \times 1$ channel impulse response, $M \geq 1$. Let N be the number of samples, $\underline{x} = [x(N)^T, \ldots, x(2)^T, x(1)^T]^T$ be the $MN \times 1$ vector of observations and $\underline{s} = [s(N), \ldots, s(1), \ldots, s(1-L)]^T$ be the $(N+L) \times 1$ vector of input symbols, both stacked in the reverse sense. According to (1), we have

$$\underline{x} = \mathcal{T}(h)\, \underline{s} + \underline{n}, \qquad (2)$$

where $\underline{n}$ is the $MN \times 1$ vector of noise and $\mathcal{T}(h)$ is the convolution matrix associated with the channel *i.e.*, $\mathcal{T}(h)$ is an $MN \times (N+L)$ block-Toeplitz matrix with $M \times 1$ blocks given by $\mathbf{h}(0), \ldots, \mathbf{h}(L)$. The upper block row of $\mathcal{T}(h)$ is $[\mathbf{h}(0), \ldots, \mathbf{h}(L), 0, \ldots, 0]$ and its first column is $[\mathbf{h}(0)^T, 0, \ldots, 0]^T$. Denote by $\mathcal{A}$ the data alphabet, *i.e.*, a discrete set of values taken by $s(t)$. We assume that $s(t)$ is unit variance *i.i.d.* and, for simplicity, that all the symbols are equiprobable. For the *known* channel with AWGN, the ML data estimate is

$$\hat{\underline{s}} = \arg\min_{\underline{u}} \{\, \|\underline{x} - \mathcal{T}(h)\, \underline{u}\| \;\; : \;\; \underline{u} \in \mathcal{A}^{N+L} \,\}. \qquad (3)$$

Suppose that there exists an interval of size m such that *all* the symbols of $\hat{\underline{s}}$ are different from the corresponding

symbols of $\underline{s}$ whereas the preceding/following symbol is the same for $\underline{s}$ and $\hat{\underline{s}}$. Define $\hat{\underline{s}}_m$ and $\underline{s}_m$ to be the vectors of symbols corresponding to this interval and the vector of errors $\underline{e}_m \triangleq \hat{\underline{s}}_m - \underline{s}_m$. Forney called such an occurrence an *error event* of length m. A sub-event $\mathcal{E}_m$ of the error event is that $\hat{\underline{s}}$ is "better" than $\underline{s}$ in the sense of the ML metric :

$$\mathcal{E}_m \;:\quad \|\underline{x}_m - \mathcal{T}_m(h)\,\hat{\underline{s}}_m\| \;\le\; \|\underline{x}_m - \mathcal{T}_m(h)\,\underline{s}_m\|\,, \quad (4)$$

where $\underline{x}_m$ is the sub-vector of $\underline{x}$ and $\mathcal{T}_m(h)$ is the block of $\mathcal{T}(h)$ corresponding to the error interval. The probability $P(\mathcal{E}_m)$ of $\mathcal{E}_m$ allows us to calculate tight bounds for various error statistics such as the bit error rate and mean time between the error events, see *e.g.*, [3]. When the channel is perfectly known, we have,

$$P(\mathcal{E}_m) = \mathcal{Q}[\,\|\,\underline{\epsilon}_m\,\|\,/2\sigma\,], \qquad \underline{\epsilon}_m = \mathcal{T}_m(h)\,\underline{e}_m, \quad (5)$$

$\mathcal{Q}[\,\alpha\,] = (\pi)^{-1/2}\int_\alpha^\infty \exp(-y^2)\,dy$ is the error function and σ^2 is the AGWN variance. Our goal is to approximate to $P(\mathcal{E}_m)$ when the channel is estimated.

3. MAIN RESULT

Let $\mathbf{h} = [\mathbf{h}(0)^T, \ldots, \mathbf{h}(L)^T]^T$ be the vector of *true* channel parameters and $\hat{\mathbf{h}}_T$ be its estimate obtained from T data samples. The use of $\hat{\mathbf{h}}_T$ instead of $\mathbf{h}$ yields

$$\hat{\underline{s}} = \arg\min_{\underline{u}} \{\, \|\underline{x} - \mathcal{T}(\hat{h}_T)\,\underline{u}\| \;:\; \underline{u} \in \mathcal{A}^{N+L} \,\}. \quad (6)$$

The exact evaluation of $P(\mathcal{E}_m)$ is too difficult in cases of practical interest. The following statement gives an asymptotic approximation for large T and L. Denote by $\langle A\rangle_{ij}$ the $M \times M$ block of A with the upper-left corner at $(M(i-1)+1,\, M(j-1)+1)$ and by $\underline{n}_m$ the noise sub-vector corresponding to $\mathcal{E}_m$.

Proposition 1: Assume that $\hat{\mathbf{h}}_T$ is asymptotically unbiased circular Gaussian with the covariance matrix C : $\sqrt{T}\,(\hat{\mathbf{h}}_T - \mathbf{h}) \xrightarrow{\mathrm{d}} \mathcal{N}_c(\,0,\, C\,)$ and that $\hat{\mathbf{h}}_T$ is statistically independent from $\underline{n}_m$. Then, $T, L \to \infty$, $L/T < \infty$, and the technical assumptions $\sup_L \{\mathrm{tr}\,(\,C\,)\,/\|\mathbf{h}\|^2\} < \infty$, $\sup_L \{\|\underline{e}_m\|\,\|\mathbf{h}\|\,/\|\underline{\epsilon}_m\|\,\} < \infty$ yield

$$P(\mathcal{E}_m) \;-\; \mathcal{Q}\left[\frac{\|\underline{\epsilon}_m\|}{2\sigma}\left(1 + \frac{L+1}{T}\frac{\underline{\epsilon}_m{}^H \mathbf{C}_m \underline{\epsilon}_m}{\sigma^2\|\underline{\epsilon}_m\|^2}\right)^{-\frac{1}{2}}\right] \to 0,$$

$$\langle \mathbf{C}_m\rangle_{ij} = \left(\frac{1}{L+1}\sum_{q-p=i-j}\langle C\rangle_{pq}\right) \quad (7)$$

for $|i-j| \le L$ and $\langle \mathbf{C}_m\rangle_{ij} = 0$ for $|i-j| > L$.
See the Appendix for proof. We briefly discuss the above technical conditions. The first one requires that the channel estimation error variance normalized by the output power converges in T uniformly w.r.t. L. The second condition means that the output distance $\|\underline{\epsilon}_m\|$ is comparable to the input distance $\|\underline{e}_m\|$ enhanced by the channel. This condition excludes the case of *catastrophic* channels such that, by analogy with the catastrophic codes, an infinitely long error event corresponds to a finite Euclidean distance at the output.

Strictly speaking, the statistical independence between $\hat{\mathbf{h}}_T$ and $\underline{n}_m$ does not hold for the blind channel identification methods which make use of the Viterbi decoder inputs to calculate $\hat{\mathbf{h}}_T$. However, the duration m of an error event is usually small compared to the data block size T. Hence an estimate $\hat{\mathbf{h}}_T$ which exploits the whole data block $\underline{x}$ is fairly independent of $\underline{n}_m$.

The circularity assumption is introduced for the sake of simplicity; it may be relaxed. The consideration of large T is standard since the distribution of $\hat{\mathbf{h}}_T$ is rarely known for finite T. The most limiting assumption is that of large L. This assumption is, however, required to take into account channel estimation errors in their asymptotic domain. Indeed, for a fixed L, taking $T \to \infty$ gives (5), *i.e.*, the standard result with $\hat{\mathbf{h}}_T = \mathbf{h}$.

The result of Proposition 1 is still difficult to use because of the term $\underline{\epsilon}_m{}^H \mathbf{C}_m\, \underline{\epsilon}_m \;/\; \|\,\underline{\epsilon}_m\,\|^2$, since this depends on $\underline{\epsilon}_m$. To get rid of this dependence, one may take an asymptotic upper bound

$$P(\mathcal{E}_m \;\le\; \mathcal{Q}\left[\frac{\|\,\underline{\epsilon}_m\,\|}{2\sigma}\left(1 + \frac{L+1}{T}\,\frac{\lambda_m}{\sigma^2}\right)^{-\frac{1}{2}}\right] \quad (8)$$

obtained by the relation $\underline{\epsilon}_m{}^H \mathbf{C}_m\, \underline{\epsilon}_m \le \lambda_m\,\|\,\underline{\epsilon}_m\,\|^2$, where λ_m is the maximum eigenvalue of $\mathbf{C}_m$. Comparing (8) with (5), we conclude that the effect of channel estimation error is similar to a loss in SNR. Indeed, the bound (8) may be interpreted as the bound (5) corresponding to the equivalent signal-to-noise ratio

$$\widehat{SNR} = SNR\left(1 + \frac{L+1}{T}\,\frac{\lambda_m}{\sigma^2}\right)^{-1}, \quad (9)$$

where SNR is the true signal-to-noise ratio. This expression is useful in predicting the sample size T which ensures that a certain admissible SNR loss is not exceeded. However a more accurate bound may be required when the performance of competitive estimators is compared. Note also that an improvement in the estimation accuracy at the price of computational complexity or/and sample size makes no sense if the condition $(L+1)\lambda_m\,/(T\,\sigma^2) \ll 1$ is satisfied.

4. LEAST SQUARES ESTIMATES

The commonly used channel estimator makes use of a *training sequence*, *i.e.*, a specific data sequence which is known at the receiver. Denote by $\tilde{\underline{s}} = [s(t_o + T + L), \ldots, s(t_o + 1)]$ the vector that stacks the training

symbols and by $H_L(\underline{\tilde{s}})$ the $T \times (L+1)$ Hankel matrix having the first column $[s(t_o+T+L), \ldots, s(t_o+L+1)]^T$ and the last row $[s(t_o+L+1), \ldots, s(t_o+1)]$. Note that for $M = 1$, the $T \times 1$ vector $\underline{\tilde{x}}$ of the "trained" outputs is given, according to (1), by $\underline{\tilde{x}} = H_L(\underline{\tilde{s}})\,\mathbf{h} + \underline{\tilde{n}}$, where $\underline{\tilde{n}}$ is the corresponding noise vector. In the case of $M > 1$, this expression generalizes to

$$\underline{\tilde{x}} = \mathbf{H}_L(\underline{\tilde{s}})\,\mathbf{h} + \underline{\tilde{n}}, \quad \text{where} \quad \mathbf{H}_L(\underline{\tilde{s}}) = H_L(\underline{\tilde{s}}) \otimes \mathrm{I}_M,$$

where $(\otimes)$ is the Kronecker product and I_M is the $M \times M$ identity matrix. When $T \geq L+1$, the channel is perfectly identifiable from noiseless data. In the presence of AWGN, the minimum variance estimate of $\mathbf{h}$ is given by

$$\hat{\mathbf{h}}_T = \arg\min_{\mathbf{f}} \|\underline{\tilde{x}} - \mathbf{H}_L(\underline{\tilde{s}})\,\mathbf{f}\| \;\Rightarrow\; \hat{\mathbf{h}}_T = \mathbf{H}_L(\underline{\tilde{s}})^{\#}\,\underline{\tilde{x}}, \quad (10)$$

where $(^{\#})$ stands for the Moore-Penrose pseudo-inverse. It is easy to check that $\hat{\mathbf{h}}_T$ is circular Gaussian :

$$\sqrt{T}\,(\hat{\mathbf{h}}_T - \mathbf{h}) \sim \mathcal{N}_c(\,0,\, \sigma^2\,\mathbb{I}^{-1}\,), \quad \mathbb{I} = \frac{1}{T}\,\mathbf{H}_L(\underline{\tilde{s}})^H \mathbf{H}_L(\underline{\tilde{s}}).$$

Denote by γ the maximum eigenvalue of $\mathbb{I}^{-1}$, then $\mathbb{I}^{-1} \leq \gamma\,\mathrm{I}_{M(L+1)}$. One may check that the latter inequality yields $\mathbf{C}_m \leq \sigma^2\,\gamma\,\mathrm{I}_{Mm}$. Now according to (9),

$$\widehat{SNR} = SNR\,\left(1 + \frac{L+1}{T}\,\gamma\right)^{-1}. \quad (11)$$

To minimize the variance of $\hat{\mathbf{h}}_T$, the matrix $\mathbb{I}$ should approach the identity matrix which implies the use of uncorrelated training sequences. When T is large compared to L, we have $\mathbb{I} \approx \mathrm{I}_{M(L+1)}$ and therefore $\gamma \approx 1$. Note that using the minimum training size, *i.e.*, $T = L+1$, leads to the SNR loss of at least 3 dB. In the GSM norm, the standard choice $L = 4$ and $T = 22$ corresponds the loss of about 1 dB.

5. SEMI-BLIND ESTIMATOR

In this section, we study the performance of the Viterbi detector aided by the semi-blind channel estimator recently proposed in [5]. The core idea of this estimator is to enhance the performance of the trained estimator by using the principle of blind subspace based identification [6]. Here we briefly recall the main results presented in [5], [7].

Let $\underline{x}_K(t) = [x(t)^T, \ldots, x(t-K)^T]^T$ be an $M(K+1) \times 1$ vector stacking $(K+1)$ consecutive data samples and let $\hat{\mathbf{R}}_K = (N-K)^{-1} \sum_{t=K+1}^{N} \underline{x}_K(t)\,\underline{x}_K(t)^H$ be the empirical space-time covariance matrix calculated from N received data samples. Define also $\mathbf{\Pi}$ the projector onto the noise subspace of $\mathbb{E}\{\hat{\mathbf{R}}_K\}$ and $\hat{\mathbf{\Pi}}$ the empirical projector calculated from the estimate $\hat{\mathbf{R}}_K$. Due to (2), the space of the observations $\underline{x}_K(t)$ is spanned by the columns of $\mathcal{T}_K(h)$ in the absence of noise and therefore $\mathbf{\Pi}\,\mathcal{T}_K(h) = 0$. Based on this property, the *blind* subspace based estimator in [6] yields the quadratic minimization $\hat{\mathbf{h}}_N = \arg\min_{\mathbf{f}} \{\|\hat{\mathbf{\Pi}}\,\mathcal{T}_K(f)\|_F \;:\; \|\mathbf{f}\| = 1\}$, where $\mathbf{f} = [\mathbf{f}(0)^T, \ldots, \mathbf{f}(L)^T]^T$ is the channel variable. The semi-blind estimator in [5] combines the cost function of this blind estimator with (10) so that

$$\hat{\mathbf{h}}_{T,\,N} = \arg\min_{\mathbf{f}} \{\|\underline{\tilde{x}} - \mathbf{H}_L(\underline{\tilde{s}})\,\mathbf{f}\|^2 \,+\, N\,\|\hat{\mathbf{\Pi}}\,\mathcal{T}_K(f)\|_F^2\},$$

Let us recall the main properties of $\hat{\mathbf{h}}_N$ and $\hat{\mathbf{h}}_{T,\,N}$. The channel transfer function $h(z) = \sum_{\tau=0}^{L} \mathbf{h}(\tau)\,z^{-\tau}$ may be factored as $h(z) = h_o(z)\,c(z)$ with the *prime* $M \times 1$ polynomial $h_o(z) = \sum_{\tau=0}^{L_o} \mathbf{h}_o(\tau)\,z^{-\tau} \neq 0$ and the scalar common factor $c(z) = \sum_{\tau=0}^{L-L_o} \mathbf{c}(\tau)\,z^{-\tau}$. The common factor accounts for a bad diversity of outputs, excessive choice of L and synchronization errors [7]. As shown in [6], the blind estimator $\hat{\mathbf{h}}_N$ collapses when either of these problems is encountered, *i.e.*, when $L > L_o$. Meanwhile the semi-blind estimator $\hat{\mathbf{h}}_{T,\,N}$ remains consistent when $L > L_o$. This estimator admits a relatively simple closed form solution; it also yields a noticeable improvement of the trained estimator in terms of the estimation error variance. In practice, the data block length substantially exceeds the training size : $N \gg T$. According to [5], the asymptotic distribution of $\hat{\mathbf{h}}_{T,\,N}$ at $N \to \infty$ and fixed T is :

$$\begin{aligned} &\sqrt{T}\,(\hat{\mathbf{h}}_{T,\,N} - \mathbf{h}) \xrightarrow{\mathrm{d}} \mathcal{N}_c(\,0,\, C_\star\,), \\ &C_\star = \sigma^2\,B\,(\,B^H\,\mathbb{I}\,B\,)^{-1}\,B^H, \end{aligned} \quad (12)$$

where B is the $M(L+1) \times (L-L_o+1)$ block-Toeplitz matrix with $M \times 1$ blocks. The first column of B is given given by $[\mathbf{h}_o(0)^T, \ldots, \mathbf{h}_o(L_o)^T]^T$ and the first block row $[\mathbf{h}_o(0), 0, \ldots, 0]$. To compare $\hat{\mathbf{h}}_T$ and $\hat{\mathbf{h}}_{T,\,N}$ in terms of decoding performance, we approximate the bound (9) corresponding to the estimation error variance $C_\star$ specified in (12). First, the bound $\mathbb{I}^{-1} \leq \gamma\,\mathrm{I}_{M(L+1)}$ yields $C_\star \leq \sigma^2\gamma B(B^H B)^{-1}B^H$. Note that $B(B^H B)^{-1}B$ is the projector onto the span$\{\,B\,\}$. Consequently, $\mathrm{tr}(\,B(B^H B)^{-1}B^H\,) = L - L_o + 1$.

To derive the bound (9), we will assume that $\mathbf{C}_m$ is approximately block-diagonal. Such an assumption makes sense when (L_o+1) modes $\mathbf{h}_o(\tau)$ are spatially decorrelated. Strictly speaking, $\mathbf{C}_m$ is asymptotically block diagonal when $L_o \to \infty$ and a stochastic channel model is such that different $\mathbf{h}_o(\tau)$ are statistically independent. Under this assumption, λ_m may be calculated as the maximum eigenvalue of the $M \times M$ diagonal blocks of $\mathbf{C}_m$. Taking account (7), we obtain : $\lambda_m \leq (L+1)^{-1}\,\mathrm{tr}(\,C_\star\,) \leq \sigma^2\,\gamma\,(L-L_o+1)\,/\,(L+1)$.

Plugging the upper bound of λ_m into (9), find :

$$\widehat{SNR} = SNR \left(1 + \frac{L - L_o + 1}{T}\gamma\right)^{-1}. \qquad (13)$$

Comparing (13) to (11), we observe that $\hat{\mathbf{h}}_{T,N}$ is statistically equivalent to $\hat{\mathbf{h}}_T$ with the model order L reduced to the difference between L and the order L_o of the channel factor $h_o(z)$ which exhibits a diversity of outputs. The performance gain of $\hat{\mathbf{h}}_{T,N}$ over $\hat{\mathbf{h}}_T$ also depends on the accuracy of the channel modeling. Indeed, choosing $L \gg L_o$ would result in $(L - L_o)/L \approx 1$ and hence comparable performances of (11) and (13).

6. NUMERICAL STUDY

We simulate a communication system with $M = 4$ receiving antennas. Each emitted burst consists of 200 QAM-4 data symbols and $(T + L)$ training symbols. The signal is transmitted over a multipath channel with the delay spread of approximately 2 symbol periods, shaped at the emitter and receiver by the half raised cosine filter with rolloff 0.5 and sampled at the symbol rate. A 5-tap channel estimate ($L = 4$) is calculated according to (10) and used for the ML detection procedure (6). In Fig.1, the symbol error rate, obtained from 10000 Monte-Carlo trials, is plotted against the SNR for different training sequences. The values plotted by $(-\square-)$ are obtained with the *true* channel while $(-\diamond-)$, $(-\triangledown-)$ and $(-\star-)$ correspond to estimates obtained with $T = 8$, $T = 12$ and $T = 22$ respectively. Note that $T = 22$ is the case of the GSM norm. The expected losses in SNR, calculated for the given T and the corresponding γ according to (11), are given by :

T	8	12	22
$SNR/\widehat{SNR}$, dB	2.6324	2.1085	1.1185

The lowest solid line connects the points obtained for the true channel whereas the other three solid lines are obtained by shifting the first one to the right by the values of SNR losses $(SNR/\widehat{SNR})_8$, $(SNR/\widehat{SNR})_{12}$ and $(SNR/\widehat{SNR})_{22}$. We naturally expect that these three curves predict the symbol rates for $T = 8$, $T = 12$ and $T = 22$ respectively. A good agreement of these theoretical curves with the estimated error rates may be observed for the probabilities of error less than 10^{-2}.

In Fig.2, we compare $\hat{\mathbf{h}}_T$ and $\hat{\mathbf{h}}_{T,N}$ at $L = 3, 4$. The simulation environment is as in the previous example, $T = 8$ and $N = 200$. Similarly to the previous figure, the lower pair of curves $(-\square-)$, $(-)$ corresponds to the true channel. The other four solid lines stand for the theoretical performance obtained by shifting the lowest solid line right-wise by the corresponding values of SNR losses calculated via (11) and (13) for $\hat{\mathbf{h}}_T$ and $\hat{\mathbf{h}}_{T,N}$ respectively. Note that in all cases, the performance gain of $\hat{\mathbf{h}}_{T,N}$ against $\hat{\mathbf{h}}_{T,N}$ approaches 1 dB.

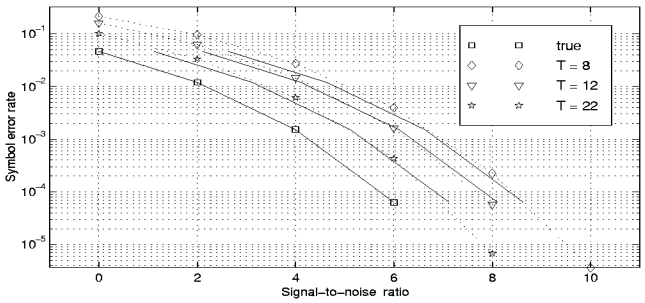

Fig.1. Symbol error rate : $\hat{\mathbf{h}}_T$, different T.

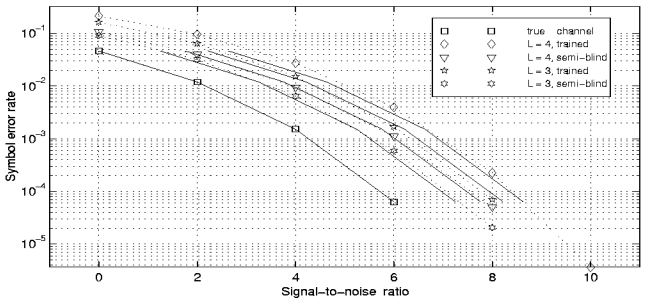

Fig.2. Symbol error rate : $\hat{\mathbf{h}}_T$ and $\hat{\mathbf{h}}_{T,N}$, different L.

7. REFERENCES

[1] G.D. Forney, Jr., "Error bounds for convolutional codes and an asymptotically optimal decoding algorithm," *IEEE Tr. on Info. Theory*, vol. 13, pp. 260–269, Apr. 1967.

[2] J. Proakis, *Digital communications*. New York: Mc Graw-Hill, 1989.

[3] G.D. Forney, Jr., "Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Tr. on Info. Theory*, vol. 18, pp. 363–378, May 1972.

[4] P. Vila, F. Pipon, D. Pirez and L. Féty, "MMSE antenna diversity equalization of a jammed frequency-selective fading channels," in *Proc. EUSIPCO*, pp. 1516–1519, Sept. 1994.

[5] A. Gorokhov and P. Loubaton, "Semi-blind second order identification of convolutive channels," in *Proc. ICASSP*, vol. 5, pp. 3905–3908, May 1997.

[6] E. Moulines, P. Duhamel, J.F. Cardoso and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Tr. on Sig. Proc.*, vol. 43, pp. 516–526, Feb. 1995.

[7] A. Gorokhov, *Blind separation of convolutive mixtures : second order methods*. PhD thesis, Ecole Nationale Supérieure des Télécommunications, 1997.