Unsupervised Bit Error Rate Estimation Using Maximum Likelihood Kernel Methods

Jia Dong[‡], Tarik Ait-Idir[◇], and Samir Saoudi[‡]
[‡] Signal and Communications Department, Institut Telecom / Telecom Bretegne/LabSticc, and UeB Brest, France
[◇] Communications Systems Department, INPT, Rabat, Morocco
Email: samir.saoudi@telecom-bretagne.eu

***Abstract*—This paper addresses the problem of unsupervised Bit Error Rate (BER) estimation for communication systems where no prior information about transmitted bits is available at the receiver. We introduce a new technique using kernel probability density function (pdf) estimation together with Maximum Likelihood-based smoothing parameter computation. Performance is evaluated in the case of CDMA, Turbo and LDPC codes. Compared to Monte Carlo method, we show that the new technique provides good BER estimates even in the region of high SNR while it requires a few number of observations.**

I. Introduction

In many communication systems, bit error rate (BER) is required to be online estimated to perform system-level functions such as scheduling, resource allocation, power control, or link adaptation. BER closed-form expressions are generally difficult to obtain, and only upper and/or lower bounds are used. In [1], an online BER estimation technique where the transmitter sends a fixed information bit value has been proposed. At the receiver side, the BER is computed by estimating the probability density function (pdf) of received soft channel/receiver outputs. This technique is called *soft* BER estimation since it directly uses the channel/receiver outputs to compute the BER without requiring hard decisions about information bits. In this technique, the Kernel method is used for probability density function (pdf) estimation, together with integrated mean square error (IMSE) for the computation of the Kernel smoothing parameter. In [2], this method has been extended to the case of unsupervised BER estimation where the receiver has no prior knowledge about transmitted bits. The proposed technique uses expectation maximization (EM) for estimating *a priori* probabilities of transmitted bit values. Estimation techniques using Gaussian mixtures for pdf computation have introduced in [3]. The number of optimal Gaussian components has been derived using the theory of mutual information, while EM has been used for the estimation of mixture parameters.

In the case of unsupervised estimation with Kernel-based methods, optimal IMSE smoothing parameter is quite difficult to derive, and only approximations are generally used. In this paper, we use the Maximum Likelihood (ML) criterion for the computation of optimal smoothing parameter. BER Performance of the proposed unsupervised ML Kernel technique is evaluated in the frameworks of turbo codes, Code Division Multiple Access (CDMA) systems, and Low Density Parity Check (LDPC) codes, and compared with that of Monte Carlo-based estimation. The results show that the new technique provides similar BER values as Monte Carlo, while it only requires a few number of samples.

II. Kernel Method

In this section, we briefly explain the proposed ML Kernel-based BER estimation method. The entire algorithm will be provided in the full version. Please refer to [1], [3] for more details about the use of Kernel techniques for BER estimation.

A. Brief introduction of kernel method

We consider a digital communication system where a sequence of N independent and identically distributed (i.i.d) information bits $(b_i)_{1\leq i\leq N} \in \{+1, -1\}$ is transmitted using a specific transmission scheme. At the receiver side, the corresponding receiver output sequence is $(X_i)_{1\leq i\leq N}$ which presents realisations of random variable (RV) X. Let $f_X(x)$ denote the pdf of X. We have,

$$f_X(x) \quad = \quad \pi_+ f_X^{b_+}(x) + \pi_- f_X^{b_-}(x), \tag{1}$$

where $f_X^{b_+}(\cdot)$ (respectively, $f_X^{b_-}(\cdot)$) is the conditional pdf of X such that $b_i = +1$ (respectively, $b_i = -1$), $\pi_+ \triangleq P[b_i = +1]$, $\pi_- \triangleq P[b_i = -1]$, and $\pi_+ + \pi_- = 1$. We use Expectation Maximization (EM) techniques iteratively to partition $X_1, \cdots, X_N$ into two classes $\mathcal{C}_+$ and $\mathcal{C}_-$, where $\mathcal{C}_+$ (respectively, $\mathcal{C}_-$) contains all receiver outputs X_i such that the corresponding transmitted bit is $b_i = +1$ (respectively, $b_i = -1$). Let N_+ and N_- denote the cardinals of $\mathcal{C}_+$ and $\mathcal{C}_-$, respectively. The kernel method is used to estimate the conditional pdfs as

$$\hat{f}_{X,N_+}^{b_+}(x) = \frac{1}{N_+ h_{N_+}} \sum_{X_i \in \mathcal{C}_+} K\left(\frac{x - X_i}{h_{N_+}}\right), \tag{2}$$

$$\hat{f}_{X,N_-}^{b_-}(x) = \frac{1}{N_- h_{N_-}} \sum_{X_i \in \mathcal{C}_-} K\left(\frac{x - X_i}{h_{N_-}}\right), \tag{3}$$

where h_{N_+} and h_{N_-} are the smoothing parameters which depend on N_+ and N_-, respectively, and $K(.)$ is an arbitrary pdf called *kernel*. Finding the optimal smoothing parameter according to the IMSE criterion is very difficult (see [1]). In the case of a Gaussian kernel $K(.)$ and a Gaussian RV X, the sub-optimal paramters are given as (see [3])

$$h_{N_+}^{*} = \left(\frac{4}{3N_+}\right)^{\frac{1}{5}} \sigma_+, \tag{4}$$

978-1-4673-0990-5/12/$31.00 ©2012 IEEE

$$h_{N_-}^* = \left(\frac{4}{3N_-}\right)^{\frac{1}{5}} \sigma_-, \tag{5}$$

where σ_+^2 and σ_-^2 are the variances of $\mathcal{C}_+$ and $\mathcal{C}_-$, respectively.

B. Principal of ML criterion for the computation of optimal smoothing parameters

In this sub section, we will show how the optimal smoothing parameter h_{N_+} is derived in the ML criterion sense. The reader can easily derive similar equations for h_{N_-}.

Suppose there is a sample $X_1, X_2..., X_{N_+}$, of N_+ observations coming from a distribution with a pdf $\hat{f}_{X,N_+}^{b+}(.|h_{N_+})$. The value h_{N_+} is unknown. It is demanded to find some estimators $\hat{h}_{N_+}$ which could be as close as possible to the true values. To use the Maximum Likelihood method, we firstly specify the joint density function of all observations. For an independent and identically distributed sample this joint density function will be

$$\hat{f}_{X,N_+}^{b+}(X_1, X_2..., X_{N_+}|h_{N_+}) = \hat{f}_{X,N_+}^{b+}(X_1|h_{N_+})\, \hat{f}_{X,N_+}^{b+}(X_2|h_{N_+}) ... \hat{f}_{X,N_+}^{b+}(X_{N_+}|h_{N_+}), \tag{6}$$

We look at the function above by considering the observed values $X_1, X_2..., X_{N_+}$ to be fixed "parameters" of this function, whereas h_{N_+} will be the function's variable and allowed to vary freely. From this point of view the distribution function will be called the Likelihood:

$$\mathcal{L}(h_{N_+}|X_1, X_2..., X_{N_+}) = \hat{f}_{X,N_+}^{b+}(X_1, X_2..., X_{N_+}|h_{N_+}) \tag{7}$$

$$= \prod_{i=1}^{N_+} \hat{f}_{X,N_+}^{b+}(X_i|h_{N_+}), \tag{8}$$

As log function is a monotone transformation, it's often more convenient to work with the likelihood function $\ln\left(\prod_{i=1}^{N_+} \hat{f}_{X,N_+}^{b+}(X_i)\right)$, called the log-likelihood. The Maximum Likelihood method estimates the smoothing parameter by finding a value of h_{N_+} which maximizes the following function $\ln\left(\prod_{i=1}^{N_+} \hat{f}_{X,N_+}^{b+}(X_i)\right)$

$$h_{N_{+MLE}} = argmax\{ln[\mathcal{L}(h_{N_+}|X_1, X_2..., X_{N_+})]\}, \tag{9}$$

Since the outputs of the receiver are independent, that is equivalent to maximize the function $\ln\left(\prod_{i=1}^{N_+} \hat{f}_{X,N_+}^{b+}(X_i)\right)$ with respect to h_{N_+}. It means that we need to find the value of h_{N_+} that cancels the derivative of the considered function $[\ln\left(\prod_{i=1}^{N_+} \hat{f}_{X,N_+}^{b+}(X_i)\right)]'$. The log-likelihood function $\ln\left(\prod_{i=1}^{N_+} \hat{f}_{X,N_+}^{b+}(X_i)\right)$ can be given as

$$\ln\left(\prod_{i=1}^{N_+} \hat{f}_{X,N_+}^{b+}(X_i)\right) = \sum_{i=1}^{N_+} \ln\left(\frac{1}{N_+ h_{N_+}} \sum_{j=1, j\neq i}^{N_+} K\left(\frac{X_i - X_j}{h_{N_+}}\right)\right) \tag{10}$$

We plot in Fig. 1 this function of h_N for a CDMA system using N=$1,000$ soft observations with MMSE receiver. It is obvious that the maximum value exists. The derivative of this function is given [4] as

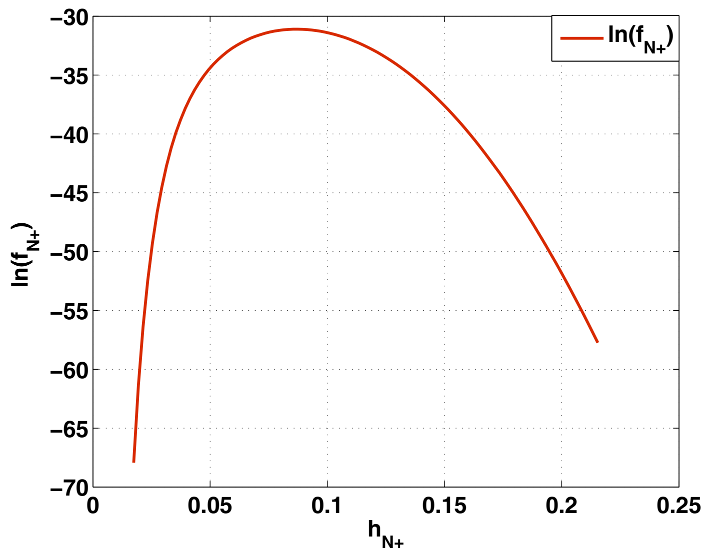

Figure 1. The calculation of $\ln\left(\prod_{i=1}^{N_+} \hat{f}_{X,N_+}^{b+}(X_i)\right)$

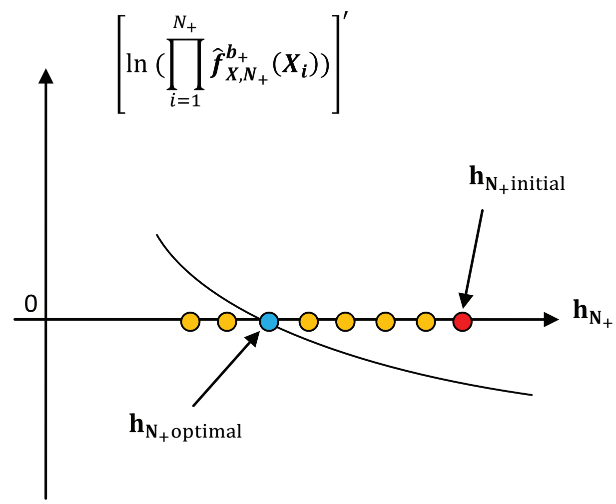

Figure 2. Relationship between the derivative of $\ln\left(\prod_{i=1}^{N_+} \hat{f}_{X,N_+}^{b+}(X_i)\right)$ and h_{N_+}

$$\left[\ln\left(\prod_{i=1}^{N_+} \hat{f}_{X,N_+}^{b+}(X_i)\right)\right]' = -\frac{N_+}{h_{N_+}} - \sum_{i=1}^{N_+} \frac{\sum_{j=1}^{N_+} K'\left(\frac{X_i - X_j}{h_{N_+}}\right)\left(\frac{X_i - X_j}{h_{N_+}^2}\right)}{\sum_{j=1, j\neq i}^{N_+} K\left(\frac{X_i - X_j}{h_{N_+}}\right)} \tag{11}$$

We have simulated the relationship between the derivative of the log-likelihood function and the smoothing parameters h_{N_+}, which has been displayed in Fig. 2. We can prove that the function corresponding to the ML criterion is strictly monotonous.

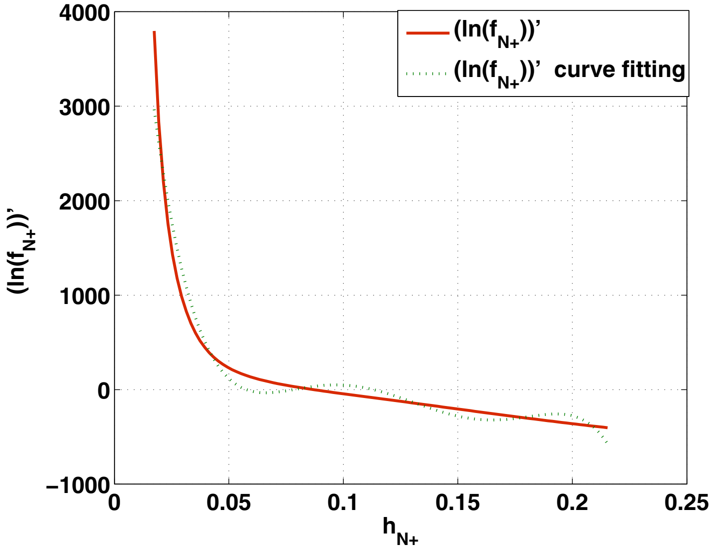

Figure 3. The curve fitting results of $[\ln\left(\prod_{i=1}^{N_+}\hat{f}_{X,N_+}^{b_+}(X_i)\right)]'$

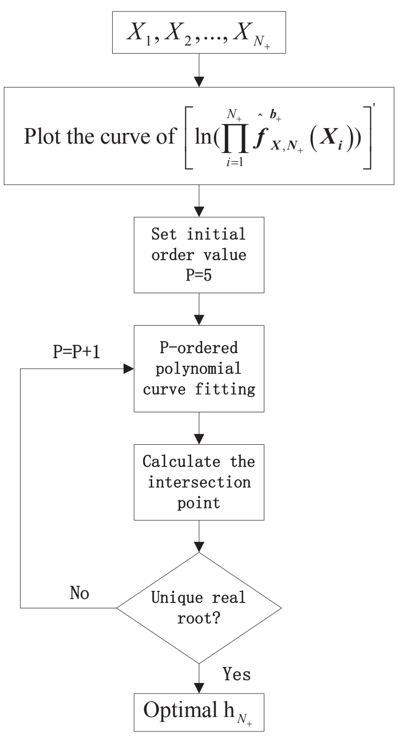

Figure 4. Flow chart for resolution of smoothing parameters h_{N_+} by using the curve fitting method

C. Curve fitting method for the computation of optimal smoothing parameters

By using the initial solution $h^*_{N_+}$ (see equation(4)), we introduce the curve fitting method to compute the optimal value of h_{N_+}. Curve fitting is the process of constructing a curve, or mathematical function, that has the best fit to a series of data points.

In order to derive the optimal value, we can use the polynomial curve fitting method to find the coefficient of a P-order polynomial

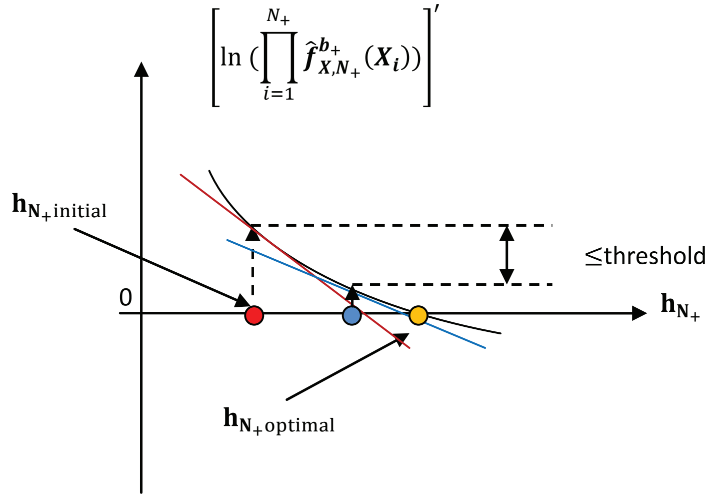

Figure 5. Newton iteration of $[\ln\left(\prod_{i=1}^{N_+}\hat{f}_{X,N_+}^{b_+}(X_i)\right)]'$

that fits the derivative of function $\ln\left(\prod_{i=1}^{N_+}\hat{f}_{X,N_+}^{b_+}(X_i)\right)$. In Fig. 3, curving fitting result of this derivative function is presented by the dotted line for the same CDMA system used in Fig. **??**. It Is plotted by using polynomial method with order P equals to 5. We can observe that this dotted line has at least 3 intersection points with x-axis. Therefore, $h_{N_+optimal}$ can not be calculated uniquely in this case. We need to use a higher-ordered polynomial model. Fig. 4 presents the flow chart of this method. It is clear that if the real root of the polynomial is not unique, we need to increase the order P of the polynomial equation.

To apply this polynomial curve fitting method, we need at first a large number of outputs at the receiver to correctly plot the original function . Another drawback of this solution is that the accuracy of the approximation of the derivative function depends on the chosen method of approximation. For example, the high-order polynomial model can be used to approximate and accurately fit the function, but it increases the complexity and computation time of the program.

D. Newton's method for the computation of optimal smoothing parameters

By using the initial solution $h^*_{N_+}$ which is given by equation 4, another method called Newton's method is also applied. Fig. 5 presents the principal of this solution and Fig. 6 illustrates the flow chart of this algorithm for the computation of h_{N_+}. With the initial value of $h^*_{N_+}$ (red point in Fig. 5), we can calculate the corresponding value of $[\ln\left(\prod_{i=1}^{N_+}\hat{f}_{X,N_+}^{b_+}(X_i)\right)]'$. Then we can find the tangent line (red line) of this point and obtain the first intersection with X axis (blue point). This first intersection is considered as a new updated value of h_{N_+}. By repeating this process, the second intersection point (yellow point) can be found, which is much closer to the optimal value of h_{N_+}. After several times of iterative calculations, with the threshold condition, we can finally find an h_{N_+} which is very close to the real optimal value. Let $g(.)$ be the derivative of the log-likelihood function. To find iteratively the unique root of the g function, Newton's method can be implemented as follows [5]:

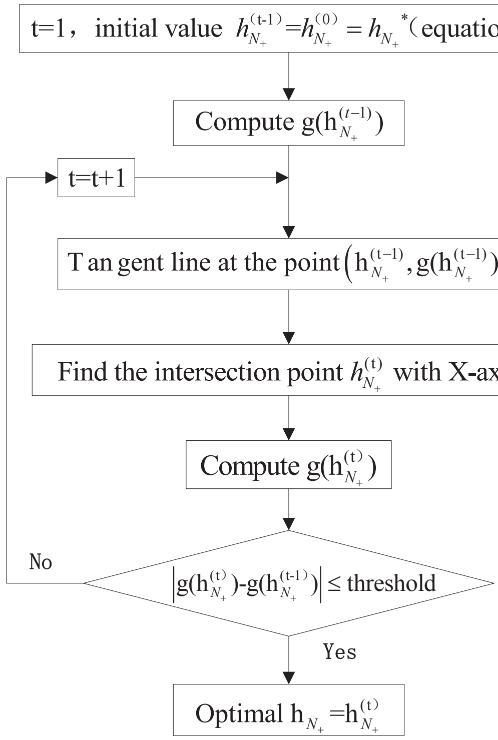

Figure 6. Flow chart for resolution of smoothing parameters h_{N_+} by using Newton's method

We begin with an initial condition $h_{N_+}^{(0)} = h_{N_+}^{*}$ (given by equation 4). Then at each iteration t, a new value of $h_{N_+}^{(t)}$ is given by using the previous value $h_{N_+}^{(t-1)}$ and the g function is presented as

$$h_{N_+}^{(t)} = h_{N_+}^{(t-1)} - \frac{g(h_{N_+}^{(t-1)})}{g'(h_{N_+}^{(t-1)})}, \tag{12}$$

where $g(h_{N_+}) = \left[\ln\left(\prod_{i=1}^{N_+} \hat{f}_{X,N_+}^{b_+}(X_i)\right)\right]'$

III. BER Estimation

The bit error probability can be expressed as

$$p_e = \pi_+ \int_{-\infty}^{0} f_X^{b_+}(x)\,dx + \pi_- \int_{0}^{+\infty} f_X^{b_-}(x)\,dx. \tag{13}$$

Using this expression, and assuming a Gaussian kernel $K(.)$, we can estimate the BER as,

$$\hat{p}_{e,N} = \frac{\pi_+}{N_+} \sum_{i\in\mathcal{C}_+} Q\left(\frac{X_i}{h_{N_+}}\right) + \frac{\pi_-}{N_-} \sum_{i\in\mathcal{C}_-} Q\left(-\frac{X_i}{h_{N_-}}\right) \tag{14}$$

where $Q(\cdot)$ denotes the complementary unit cumulative Gaussian distribution. The details about the derivation of (14) are provided in [2].

IV. Performance Evaluation

In this section, we evaluate the performance of the proposed ML kernel unsupervised BER estimation method. Detailed performance comparisons of the proposed technique with those introduced in [3] and [2] are provided below. In Fig. 7, the CDMA system is adapted to MMSE receiver. We plot the BER performance by using the initial smoothing parameters $h_{initial}$, the optimal smoothing parameters $h_{optimal}$ calculated by the mentioned curve fitting or Newton method. For low signal-noise ratio, the BER performances of different methods are quite similar. For high SNR, the performance of MC method turns to be random.

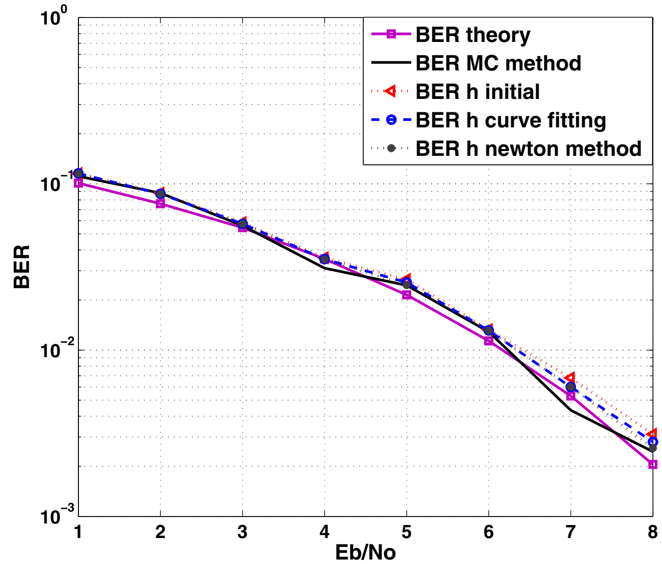

Figure 7. BER performance comparison in the case of a CDMA system using $N = 1,000$ soft observations with MMSE receiver.

Table I presents the initial and optimal values of the smoothing parameter h_{N_+} for SNR from 1 to 6 in the framework of a CDMA system with MMSE receiver using 1,000 samples. The optimal values are calculated by applying newton's method.

Table I
Initial and Optimal smoothing parameter h_{N_+} for different SNR in CDMA system with MMSE receiver

SNR	1	2	3	4	5	6
$h_{N_+-initial}$	0.1545	0.1322	0.1313	0.1274	0.1165	0.1103
$h_{N_+-optimal}$	0.1867	0.1506	0.1354	0.1306	0.1298	0.1140

In Table II, three BER values are demonstrated for SNR from 1 to 6 calculated by the corresponding initial and optimal values of the smoothing parameter h_{N_+}. There are theoretical BER value, estimated BER value by using $h_{N_+initial}$ and estimated BER value by using $h_{N_+optimal}$ which is calculated by newton's method. By comprehensive comparison, we arrive at a conclusion that the estimated BER with $h_{N_+optimal}$ is closer to the theoretical value than the one with $h_{N_+initial}$.

In Fig. 8, we report results in the case of a $1/3$-rate turbo code over a Gaussian channel. To obtain a BER equals to 7×10^{-5}, provided kernel method requires 4×10^5 samples while at the time the traditional Monte Carlo method needs at least 3×10^7 samples. So we demonstrate that the proposed technique provides similar BER results as Monte Carlo while it requires a number of soft outputs seven times less than that of Monte Carlo. We plot the results by the proposed kernel technique with both $h_{N_{initial}}$

Table II
COMPARISON OF THEORETICAL BER AND ESTIMATED BERS BY USING OPTIMAL AND INITIAL VALUE OF h_{N_+} IN CDMA SYSTEM WITH MMSE RECEIVER

SNR	1	2	3	4	5	6
Theoretical BER	0.0759	0.0537	0.0353	0.0212	0.0114	0.0053
BER(h_{N_+-ini})	0.0864	0.0605	0.0433	0.0305	0.0160	0.0077
BER(h_{N_+-opt})	0.0821	0.0582	0.0429	0.0300	0.0139	0.0061

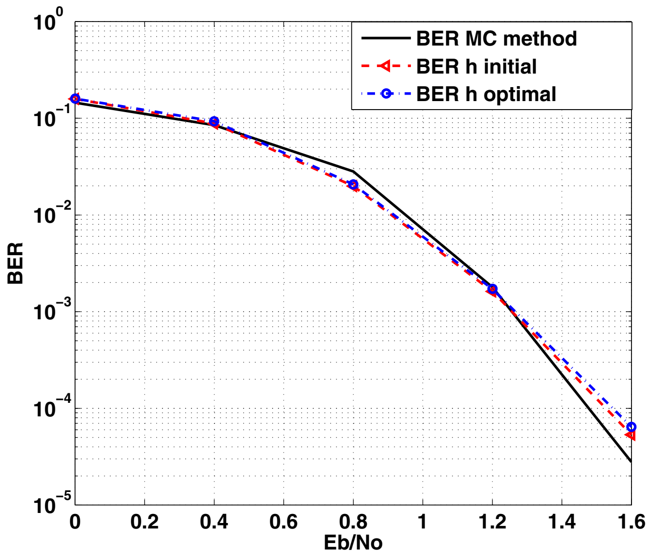

Figure 8. BER performance comparison in the case of a $1/3$-rate turbo code

and $h_{N_{optimal}}$. The optimal smoothing parameter is calculated by newton's method, in the case of turbo code. The BER results are quite similar while the complexity of $h_{N_{optimal}}$ estimation is higher.

In Fig. 9, we plot the obtained BER estimation performance in a 1/2 rate QC-LDPC system. The parity check matrix G is generated as $635\times$ 1270. We use the frame which contains 635 random bits to transmit information. BER equals to 10^{-5} could be achieved by using Monte Carlo method with 500 frames. That means $500\times635 = 317,500$ samples are required by Monte Carlo method while $80\times635 = 50,800$ samples are needed for the same BER value by the proposed kernel method. It is about six times less than that of Monte Carlo Method.

V. CONCLUSIONS

In this paper, we introduced two new ML kernel-based unsupervised BER estimation techniques where the receiver has no prior knowledge about transmitted bits. The proposed methods use new ML algorithms for the computation of smoothing parameters used to estimate conditional pdfs with the aid of the kernel approach. We compare these two methods and show their advantages and limits. For the sake of space limitation, we reported performance in the case of CDMA, LDPC systems and a $1/3$-rate turbo code, and showed that the proposed approach offers good BER estimation compared to Monte Carlo techniques, while it requires only a reduced number of soft observations.

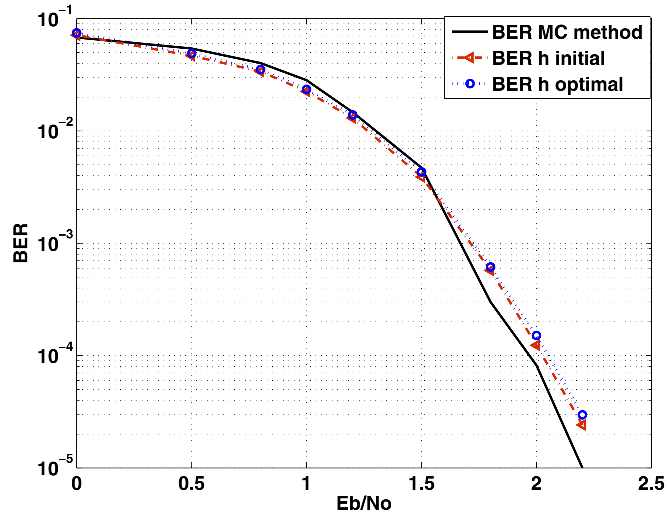

Figure 9. BER performance comparison in the case of a QC-LDPC system

REFERENCES

[1] S. Saoudi, M. Troudi and F. Ghorbel, "An iterative soft bit error rate estimation of any digital communication systems using a non parametric probability density function," *Eurasip Journal on wirless Communications and Networking,* vol. 2009, Article ID512192, 9 pages, 2009.

[2] S. Saoudi, T. Ait-Idir, and Y. Mochida, "A novel non-parametric iterative soft bit error rate estimation technique for digital communication systems," *in Proc. IEEE Intern. Commun. Conf., ICC11,* Kyoto, Japan, Jun. 2011.

[3] S. Saoudi, T. Derham, T. Ait-Idir, and P. Coupe, "A fast soft bit error estimation method," *EURASIP Journal of Wireless Commun. and Networking,* vol. 2010, Article ID 372370, 13 pages, 2010.

[4] Natalia Markovich, "Nonparametric Analysis of Univariate Heavy-Tailed Data," *Institute of Control Science,* 2007.

[5] C.T.Kelley, "Solving Nonlinear Equations with Newton's Method," *Society for Industrial and Applied Mathematics,* 2003.