New criteria for iterative decoding

Florence Alberge, Ziad Naja, Pierre Duhamel

NEW CRITERIA FOR ITERATIVE DECODING

F. Alberge, Z. Naja, P. Duhamel

Laboratoire des Signaux et Systèmes, Univ Paris-Sud, CNRS,
Supélec, Plateau de Moulon, 91190 Gif-sur-Yvette, FRANCE
email :alberge,naja,pierre.duhamel@lss.supelec.fr

ABSTRACT

Iterative decoding was not originally introduced as the solution to an optimization problem rendering the analysis of its convergence very difficult. In this paper, we investigate the link between iterative decoding and classical optimization techniques. We first show that iterative decoding can be rephrased as two embedded minimization processes involving the Fermi-Dirac distance. Based on this new formulation, an hybrid proximal point algorithm is first derived with the additional advantage of decreasing a desired criterion. In a second part, an hybrid minimum entropy algorithm is proposed with improved performance compared to the classical iterative decoding. Even if this paper focus on iterative decoding for BICM, the results can be applied to the large class of turbo-like decoders.

***Index Terms*—** Optimization methods, Iterative methods, Decoding.

1. INTRODUCTION

Bit-Interleaved Coded Modulation (BICM) was first suggested by Zehavi in [1] to improve the Trellis Coded Modulation performance over Rayleigh-fading channels. In BICM, the diversity order is increased by using bit-interleavers instead of symbol interleavers. This improvement is achieved at the expense of a reduced minimum Euclidean distance leading to a degradation over non-fading Gaussian channels [1]. This drawback can be overcome by using iterative decoding (BICM-ID) at the receiver [2]. BICM-ID is known to provide excellent performance for both Gaussian and fading channels.
The iterative decoding scheme used in BICM-ID is very similar to serially concatenated turbo-decoders. Indeed, the serial turbo-decoder makes use of an exchange of information between computationally efficient decoders for each of the component codes. In BICM-ID, the inner decoder is replaced by demapping which is less computationally demanding than a decoding step. Even if this paper focuses on iterative decoding for BICM, the results can be applied to the large class of iterative decoders including serial or parallel concatenated turbo decoders as long as low-density parity-check (LDPC) decoders. Among the different attempts to provide an analysis of iterative decoding, the EXIT chart analysis and density evolution have permitted to make significant progress [3, 4] but the results developed within this setting apply only in the case of large block length. Another tool of analysis is the connection of iterative decoding to factor graphs [5] and belief propagation [6]. Convergence results for belief propagation exist but are limited to the case where the corresponding graph is a tree which does not include turbo code or LDPC. A link between iterative decoding and classical optimization algorithms has been made recently in [7] where the turbo decoding is interpreted as a nonlinear block Gauss Seidel iteration. In parallel, a geometrical approach has been considered and provides an interesting interpretation in terms of projections. The particular case of BICM-decoding has been studied in [8, 9]. In [10], the turbo-decoding is interpreted in a geometric setting as a dynamical system leading to new but incomplete results.
In this paper we reformulate the iterative decoding as two embedded proximal point algorithms involving the Bregman divergence built on the Fermi-Dirac energy. We prove that each iteration of the decoding decreases a certain criterion.We also propose an hybrid minimum entropy algorithm with improved performance compared to the classical BICM.

2. BICM-ID WITH SOFT DECISION FEEDBACK

A conventional BICM system [11] is built from a serial concatenation of a convolutional encoder, a bit interleaver and an M-ary bits-to-symbol mapping (where $M = 2^m$) as shown in fig. 1. The sequence of information bits $\underline{\mathbf{b}}$ is first encoded by a convolutional encoder to produce the output encoded bit sequence $\underline{\mathbf{c}}$ of length L_c which is then scrambled by a bit interleaver (as opposed to the channel symbols in the symbol-interleaved coded sequence) operating on bit indexes. Let $\underline{\mathbf{d}}$ denote the interleaved sequence. Then, m consecutive bits of $\underline{\mathbf{d}}$ are grouped as a channel symbol $\mathbf{d}_k = (d_{km+1}, ...d_{(k+1)m})$. The complex transmitted signal $\mathbf{s}_k = \epsilon(\mathbf{d}_k)$ is then chosen from an M-ary constellation Ψ where ϵ denotes the mapping scheme. For simplicity, we consider transmission over the AWGN channel. The received signal reads:

$$\mathbf{y}_k = \mathbf{s}_k + \mathbf{n}_k \quad 1 \leq k \leq L_c/m \tag{1}$$

where $\mathbf{n}_k$ is a complex white Gaussian noise with independent in-phase and quadrature components having two-sided power spectral density σ_c^2.
Due to the presence of the random bit interleaver, the true maximum

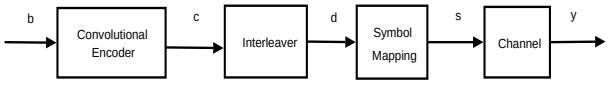

Fig. 1. Transmission model

likelihood decoding of BICM is too complicated to implement in practice. Figure 2 shows the block diagram of the receiver for a BICM-ID system with soft-decision feedback.

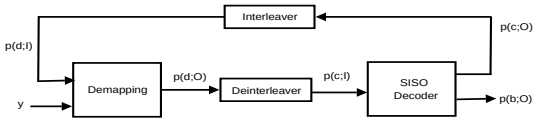

Fig. 2. Receiver for a BICM-ID with soft-decision feedback

In the first iteration, the encoded bits are assumed equally likely. The demapping consists in evaluating *a posteriori* probabilities (APP) for the encoded bits without accounting for the code structure, namely:

$$p_{APP}(d_{km+i}=b) \quad \sim \sum_{\mathbf{s}:\mathbf{s}_k\in\Psi_b^i} p(\mathbf{y}|\mathbf{s})p(\mathbf{s}) \tag{2}$$

$$\sim \sum_{\mathbf{s}_k\in\Psi_b^i} p(\mathbf{y}_k|\mathbf{s}_k)p(\mathbf{s}_k) \tag{3}$$

where $\mathbf{s}=\{\mathbf{s}_1,...,\mathbf{s}_{L_c/m}\}$, $\mathbf{y}=\{\mathbf{y}_1,...,\mathbf{y}_{L_c/m}\}$ and Ψ_b^i, $b\in\{0,1\}$, denotes the subset of Ψ that contains all symbols whose labels have the value b in the i^{th} position. In the turbo decoding process, the quantities exchanged through the blocks are not *a posteriori* probabilities (APP) but extrinsic information [12]. The extrinsic information at the output of the demapping $p(d_{km+i};O)$ is computed as $p_{APP}(d_{km+i})/p(d_{km+i};I)$ where $p(d_{km+i};I)$ is the *a priori* information for the demapping sub-block. Since the bit interleaver makes the bits independent, the extrinsic information $p(d_{km+i};O)$ reads:

$$p(d_{km+i}=b;O)=K_m\sum_{\mathbf{s}_k\in\Psi_b^i}p(\mathbf{y}_k|\mathbf{s}_k)\prod_{j\neq i}p(d_{km+j};I) \tag{4}$$

and the corresponding APP reads:

$$p_{APP}(d_{km+i}=b)=K'_m\sum_{\mathbf{s}_k\in\Psi_b^i}p(\mathbf{y}_k|\mathbf{s}_k)\prod_{j}p(d_{km+j};I) \tag{5}$$

where K_m and K'_m are normalization factors. The extrinsic information $p(d_{km+i};O)$ is de-interleaved and delivered to the SISO decoder [13] as an *a priori* information on the encoded bits. Let $c_l=d_{\sigma^{-1}(km+j)}$ where σ^{-1} is for the permutation on the indexes due to the deinterleaver; $p(c_l;I)$ is the updated input of the Single Input Single Output (SISO) decoder. The extrinsic information at the output of the SISO decoder is obtained through [8, 14]:

$$p(c_l=b;O)=K_c\sum_{\mathbf{c}\in\mathcal{R}_b^l}\mathbf{I}_{\mathcal{C}}(\mathbf{c})\prod_{j\neq l}p(c_j;I) \tag{6}$$

and the corresponding APP is:

$$p_{APP}(c_l=b)=K'_c\sum_{\mathbf{c}\in\mathcal{R}_b^l}\mathbf{I}_{\mathcal{C}}(\mathbf{c})\prod_{j}p(c_j;I) \tag{7}$$

where $\mathbf{I}_{\mathcal{C}}(\mathbf{c})$ stands for the indicator function of the code, i.e. $\mathbf{I}_{\mathcal{C}}(\mathbf{c})=1$ if $\mathbf{c}$ is a codeword and 0 otherwise and $\mathcal{R}_b^l$ denotes the set of binary words of length L_c with value b in the l^{th} position. K_c and K'_c are normalization factors. The extrinsic information $p(c_l;O)$ is interleaved and delivered to the demapping sub-block as a regenerated *a priori* information. If the process converges the APP of the two sub-blocks are the same. The criteria proposed in the following are based on this property and encourage a faster convergence towards this objective.

3. NOTATIONS FROM INFORMATION GEOMETRY

3.1. Basic tools

We first introduce some notations that will be useful in the sequel. Let $\mathbf{B}_i\in\{0,1\}^N$ denote the binary representation of the integer $i, 0\leq i\leq 2^{N-1}$. The binary representation of all the words of length N is gathered into matrix $\mathbf{B}=(\mathbf{B}_0,\mathbf{B}_1,...,\mathbf{B}_{2^N-1})^T$ with dimension $2^N\times N$. Let η be a probability mass function on the outcomes $\chi=\mathbf{B}_i$ then

$$\eta=(\mathbf{Pr}[\chi=\mathbf{B_0}],\mathbf{Pr}[\chi=\mathbf{B_1}],...,\mathbf{Pr}[\chi=\mathbf{B_{2^N-1}}])^{\mathbf{T}}$$

Given a PMF η, its log-coordinates are the vector θ whose i^{th} element is given by $\theta_i=\ln(Pr[\chi=\mathbf{B_i}])-\ln(Pr[\chi=\mathbf{B_0}])$. We can observe that there is a one-to-one mapping between η and θ since the vector η can be written $\eta=\mathbf{exp}(\theta-\psi(\theta))$ where $\psi(\theta)=log(\sum_i exp((\theta)_i))$.We also introduce the bitwise log-probability ratios with elements of the form $\lambda_j=log(\frac{Pr[\chi_j=1]}{Pr[\chi_j=0]})$ where χ_j is the j^{th} bit of the binary word χ and $\lambda\in\mathbb{R}^N$. For factorisable probability measures (*ie* PMF that factors into the product of their bitwise marginals so that $Pr(\chi)=\Pi_j Pr(\chi_j)$), the log-coordinates take the form $\theta=\mathbf{B}\lambda$.

3.2. Link with iterative decoding

Let θ_m denote the log-coordinates vector of the PMF $p(\mathbf{y}|\mathbf{s})$. Let $\lambda_\mathbf{1}$ denote the log-probability ratio corresponding to the prior $p(d_{km+i};I)$ such that:

$$(\lambda_1)_{km+i}=\ln\left(\frac{p(d_{km+i}=1;I)}{p(d_{km+i}=0;I)}\right)$$

Thus, the log-coordinates of $p(\mathbf{y}|\mathbf{s})\Pi_{j,k}p(d_{km+i};I)$ reads $\mathbf{B}\lambda_\mathbf{1}+\theta_\mathbf{m}$.

Let $p_{\mathbf{B}\lambda_\mathbf{1}+\theta_\mathbf{m}}$ represent the vector whose i^{th} element is the probability that the i^{th} bit is 1 according to the measure with log-coordinate $\mathbf{B}\lambda_\mathbf{1}+\theta_\mathbf{m}$. The APP at the output of the demapper merge with $p_{\mathbf{B}\lambda_\mathbf{1}+\theta_\mathbf{m}}$. From eq. (4)-(5), $p_{APP}(d_{km+i}=b)=p(d_{km+i}=b;I)p(d_{km+i}=b;0)$, the log-coordinates of the APP at the output of the demapper also merge with $\mathbf{B}(\lambda_\mathbf{1}+\lambda_\mathbf{2})$ where $(\lambda_2)_{km+i}=\ln\left(\frac{p(d_{km+i}=1;O)}{p(d_{km+i}=0;O)}\right)$. Then the demapper sub-block solves, with respect to $\lambda_\mathbf{2}$, the equation:

$$p_{\mathbf{B}(\lambda_\mathbf{1}+\lambda_\mathbf{2})}=p_{\mathbf{B}\lambda_\mathbf{1}+\theta_\mathbf{m}} \tag{8}$$

Let θ_c denote the log-coordinates of the PMF associated with the indicator function. Then the decoder sub-block solves, with respect to $\lambda_\mathbf{1}$, the equation:

$$p_{\mathbf{B}(\lambda_\mathbf{1}+\lambda_\mathbf{2})}=p_{\mathbf{B}\lambda_\mathbf{2}+\theta_\mathbf{c}} \tag{9}$$

Iterative decoding is thus equivalent to:

$$find\ \lambda_\mathbf{2}^{(k+1)}\ such\ that\ \ p_{\mathbf{B}(\lambda_\mathbf{1}^{(\mathbf{k})}+\lambda_\mathbf{2}^{(\mathbf{k+1})})}=p_{\mathbf{B}\lambda_\mathbf{1}^{(\mathbf{k})}+\theta_\mathbf{m}}$$
$$find\ \lambda_\mathbf{1}^{(k+1)}\ such\ that\ \ p_{\mathbf{B}(\lambda_\mathbf{1}^{(\mathbf{k+1})}+\lambda_\mathbf{2}^{(\mathbf{k+1})})}=p_{\mathbf{B}\lambda_\mathbf{2}^{(\mathbf{k+1})}+\theta_\mathbf{c}}$$

At the convergence, the APP from the two sub-blocks should be in accordance *ie* $p_{\mathbf{B}(\lambda_\mathbf{1}^{(\infty)}+\lambda_\mathbf{2}^{(\infty)})}=p_{\mathbf{B}\lambda_\mathbf{1}^{(\infty)}+\theta_\mathbf{m}}=p_{\mathbf{B}\lambda_\mathbf{2}^{(\infty)}+\theta_\mathbf{c}}$.

4. AN OPTIMIZATION PROBLEM

The Fermi-Dirac divergence is the Bregman divergence built on the Fermi-Dirac entropy $f(\mathbf{p})=\sum_j p_j\ln(p_j)+(1-p_j)\ln(1-p_j)$ with $dom(f)=[0;1]$. The Fermi-Dirac divergence reads

$$D_{FD}(\mathbf{p},\mathbf{q})=\sum_j p_j\ln\left(\frac{p_j}{q_j}\right)+\sum_j(1-p_j)\ln\left(\frac{1-p_j}{1-q_j}\right)$$

and is exactly the Kullback-Leibler distance for bit probabilities. The Fermi-Dirac divergence is a non-symmetric distance. As we can notice, this distance is very convenient for computing distances between bit probabilities.

Proposition 1 *The demapping sub-block solves the minimization problem* $\min_{\lambda_2} D_{FD}(p_{\mathbf{B}\lambda_1+\theta_m}, p_{\mathbf{B}(\lambda_1+\lambda_2)})$

The decoding sub-block solves the minimization problem

$$\min_{\lambda_1} D_{FD}(p_{\mathbf{B}\lambda_2+\theta_c}, p_{\mathbf{B}(\lambda_1+\lambda_2)})$$

Proof: The proof is obvious by noting that $(\lambda_1+\lambda_2)_{km+i} = \ln\left(\frac{p_{B(\lambda_1+\lambda_2)}(d_{km+i}=1)}{p_{B(\lambda_1+\lambda_2)}(d_{km+i}=0)}\right)$ *thus* $p_{\mathbf{B}(\lambda_1+\lambda_2)} = \frac{exp(\lambda_1+\lambda_2)}{1+exp(\lambda_1+\lambda_2)}$.
This proposition illustrates that iterative decoding can be formulated as two embedded minimization steps based on the Fermi-Dirac distance. In the next section, we investigate some modifications of this original criterion.

4.1. An hybrid proximal point algorithm

In the classical iterative decoding, the two minimization steps seem independent meaning that the minimization of one of the criterion does not imply necessarily a decrease of the other criterion at the next iteration. Proximal point methods [15] permit to make the link between the two criteria. These methods are generally used to guarantee the monotonicity of the convergence process often at the cost of a slow convergence speed. Following the proximal point technique we obtain the minimization process:

$$\lambda_2^{(k+1)} = \min_{\lambda_2} J_{\theta_m}(\lambda_1,\lambda_2) = \min_{\lambda_2} D_{FD}(\mathbf{p}_{\mathbf{B}\lambda_1+\theta_m}, \mathbf{p}_{\mathbf{B}(\lambda_1+\lambda_2)}) + \mu_m D_{FD}(\mathbf{p}_{\mathbf{B}(\lambda_1^{(k)}+\lambda_2^{(k)})}, \mathbf{p}_{\mathbf{B}(\lambda_1+\lambda_2)})$$

$$\lambda_1^{(k+1)} = \min_{\lambda_1} J_{\theta_c}(\lambda_1,\lambda_2) = \min_{\lambda_1} D_{FD}(\mathbf{p}_{\mathbf{B}\lambda_2+\theta_c}, \mathbf{p}_{\mathbf{B}(\lambda_1+\lambda_2)}) + \mu_c D_{FD}(\mathbf{p}_{\mathbf{B}(\lambda_1^{(k)}+\lambda_2^{(k+1)})}, \mathbf{p}_{\mathbf{B}(\lambda_1+\lambda_2)})$$

As can be seen, the original criterion is modified through the addition of a penalization term in order to encourage smooth variations of the successive estimates. This minimization process is equivalent to finding $\lambda_2^{(k+1)}$ such that

$$\mathbf{p}_{\mathbf{B}(\lambda_1^{(k)}+\lambda_2^{(k+1)})} = \frac{\mathbf{p}_{\mathbf{B}\lambda_1^{(k)}+\theta_m} + \mu_m \mathbf{p}_{\mathbf{B}(\lambda_1^{(k)}+\lambda_2^{(k)})}}{1+\mu_m} \quad (10)$$

and $\lambda_1^{(k+1)}$ such that

$$\mathbf{p}_{\mathbf{B}(\lambda_1^{(k+1)}+\lambda_2^{(k+1)})} = \frac{\mathbf{p}_{\mathbf{B}\lambda_2^{(k+1)}+\theta_c} + \mu_c \mathbf{p}_{\mathbf{B}(\lambda_1^{(k)}+\lambda_2^{(k+1)})}}{1+\mu_c} \quad (11)$$

Note that this new procedure also converges towards solutions satisfying (8) and (9). A good choice for the parameters μ_m and μ_c permits to ensure that each criterion decreases with the iterations. Actually, we want to enforce $J_{\theta_m}(\lambda_1^{(k)},\lambda_2^{(k+1)}) \leq J_{\theta_c}(\lambda_1^{(k)},\lambda_2^{(k)})$. Since the Fermi-Dirac distance is convex with respect to its second parameter, we have $J_{\theta_m}(\lambda_1^{(k)},\lambda_2^{(k+1)}) \leq \frac{\mu_m}{1+\mu_m}(D_{FD}(\mathbf{p}_{\mathbf{B}\lambda_1^{(k)}+\theta_m}, \mathbf{p}_{\mathbf{B}(\lambda_1^{(k)}+\lambda_2^{(k)})}) + D_{FD}(\mathbf{p}_{\mathbf{B}(\lambda_1^{(k)}+\lambda_2^{(k)})}, \mathbf{p}_{\mathbf{B}\lambda_1^{(k)}+\theta_m}))$. Moreover, we also have $D_{FD}(\mathbf{p}_{\mathbf{B}\lambda_2^{(k)}+\theta_c}, \mathbf{p}_{\mathbf{B}(\lambda_1^{(k)}+\lambda_2^{(k)})}) \leq J_{\theta_c}(\lambda_1^{(k)},\lambda_2^{(k)})$. Connecting the two relations, we obtain an upper bound for μ_m:

$$\mu_m \leq \frac{D_{FD}(\mathbf{p}_{\mathbf{B}\lambda_2^{(k)}+\theta_c}, \mathbf{p}_{\mathbf{B}(\lambda_1^{(k)}+\lambda_2^{(k)})})}{\mathcal{D}_{FD} - D_{FD}(\mathbf{p}_{\mathbf{B}\lambda_2^{(k)}+\theta_c}, \mathbf{p}_{\mathbf{B}(\lambda_1^{(k)}+\lambda_2^{(k)})})}$$

where $\mathcal{D}_{FD}$ is a symmetric distance, namely $\mathcal{D}_{FD} = D_{FD}(\mathbf{p}_{\mathbf{B}\lambda_1^{(k)}+\theta_m}, \mathbf{p}_{\mathbf{B}(\lambda_1^{(k)}+\lambda_2^{(k)})}) + D_{FD}(\mathbf{p}_{\mathbf{B}(\lambda_1^{(k)}+\lambda_2^{(k)})}, \mathbf{p}_{\mathbf{B}\lambda_1^{(k)}+\theta_m})$. The upper bound for μ_c can be obtained in the same way. Iterating (10) and (11) with μ_c and μ_m correctly chosen yields an algorithm that converges towards the same points than the classical iterative decoding with the additional advantage of decreasing at each iteration a desired criterion. In the next section, we propose a new criterion in order to improve the performance of the iterative decoding.

4.2. An hybrid minimum entropy algorithm

The entropy of the vectors of marginals $\mathbf{p}_{\mathbf{B}(\lambda_1+\lambda_2)}$ is defined as

$$E_{B(\lambda_1+\lambda_2)} = -\sum_n p_{B(\lambda_1+\lambda_2)}(n) log2(p_{B(\lambda_1+\lambda_2)}(n)) - \sum_n (1-p_{B(\lambda_1+\lambda_2)}(n)) log2(1-p_{B(\lambda_1+\lambda_2)}(n))$$

The quantity $E_{B(\lambda_1+\lambda_2)}$ gives a measure of the reliability of the decisions. Indeed, $E_{B(\lambda_1+\lambda_2)} \rightarrow 0$ does not always mean that the decisions are correct, but rather that the iterative decoding algorithm is confident about its decisions. Nevertheless, in the iterative decoding, the decisions are in most cases correct when $E_{B(\lambda_1+\lambda_2)} \rightarrow 0$ [16]. In this section, we propose a new criterion that minimizes $E_{B(\lambda_1+\lambda_2)}$ under the constraint $D_{FD}(\mathbf{p}_{\mathbf{B}\lambda_1+\theta_m}, \mathbf{p}_{\mathbf{B}(\lambda_1+\lambda_2)}) \leq \epsilon$ for the demapping and $D_{FD}(\mathbf{p}_{\mathbf{B}\lambda_2+\theta_c}, \mathbf{p}_{\mathbf{B}(\lambda_1+\lambda_2)})$ for the decoding. This is equivalent to:

$$\lambda_2^{(k+1)} = \min_{\lambda_2} D_{FD}(\mathbf{p}_{\mathbf{B}\lambda_1+\theta_m}, \mathbf{p}_{\mathbf{B}(\lambda_1+\lambda_2)}) + \eta_m E_{B(\lambda_1+\lambda_2)} \quad (12)$$

$$\lambda_1^{(k+1)} = \min_{\lambda_1} D_{FD}(\mathbf{p}_{\mathbf{B}\lambda_2+\theta_c}, \mathbf{p}_{\mathbf{B}(\lambda_1+\lambda_2)}) + \eta_c E_{B(\lambda_1+\lambda_2)} \quad (13)$$

By zeroing the gradient of the two criteria in (12) and (13), we obtain the new update equations:

$$\lambda_2{}^{(k+1)}: \quad f_{\eta_m}(p_{B(\lambda_1^{(k)}+\lambda_2^{(k+1)})}(n)) = p_{B\lambda_1^{(k)}+\theta_m}(n) \;\; 1 \leq n \leq L_c$$

$$\lambda_1{}^{(k+1)}: \quad f_{\eta_c}(p_{B(\lambda_1^{(k+1)}+\lambda_2^{(k+1)})}(n)) = p_{B\lambda_2^{(k+1)}+\theta_c}(n) \;\; 1 \leq n \leq L_c$$

where $f_\eta(p_{B(\lambda_1+\lambda_2)}(n)) = p_{B(\lambda_1+\lambda_2)}(n) - \eta p_{B(\lambda_1+\lambda_2)}(n)(1-p_{B(\lambda_1+\lambda_2)}(n)) \log\left(\frac{p_{B(\lambda_1+\lambda_2)}(n)}{1-p_{B(\lambda_1+\lambda_2)}(n)}\right)$. The function f_η is plotted

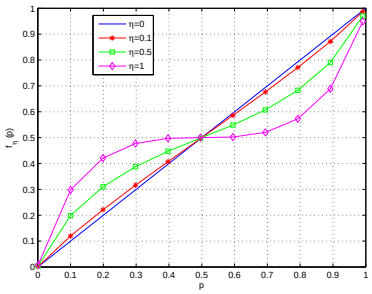

Fig. 3. $f_\eta(p)$ for various values of η

on fig (3). We can notice that: *(i)* the distortion increases with η *(ii)* $f_\eta(p)$ belongs to $[0;1]$ *(iii)* $f_\eta(p)$ is a strictly increasing function. As a consequence each step of the minimization process has a unique solution that can be found using classical techniques.

5. SIMULATION

We compare the performance in terms of bit error rate and iteration number of the classical iterative decoding with the hybrid proximal point algorithm (HPP) and also with the hybrid minimum entropy algorithm (HMEA). Each algorithm stops when the Fermi-Dirac distance between the APP of the two sub-blocks is less than 10^{-3} or when 30 iterations are reached. The generator polynomial of the encoder is $g = [111; 001; 100]$. The bits are mapped using subset partitioning to a 8-PSK modulation. The length of the coded bit sequence is $L_c = 6000$. The step-sizes η_m and η_c in the HMEA are both chosen equal to 0.05. The results are plotted in fig. (4) and (5). We can see that the classical iterative decoding and the HPP exhibits exactly the same performance. This is not surprising concerning the BER since both methods converge towards the same points. We can also notice that these results are obtained with the same number of iterations in both cases meaning that the proximal point technique does not reduce, in this case, the convergence speed. Both methods have almost the same computational complexity with the additional advantage for the HPP to minimize a desired criterion with the iterations. As expected, the HMEA outperforms the others methods in terms of BER in the middle area with a number of iterations at most equal to the number of iterations needed in the classical BICM. However, this last method has a higher computational complexity due to the distortion function f_η.

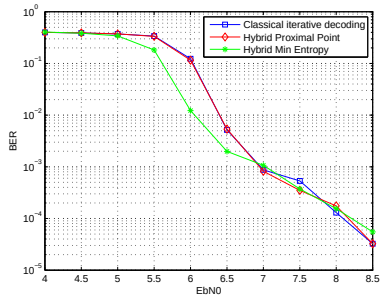

Fig. 4. BER versus EbN0

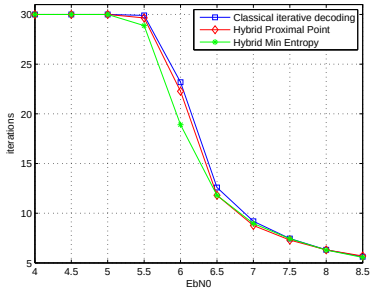

Fig. 5. Iteration number versus EbN0

6. CONCLUSION

In this paper, iterative decoding is rephrased as two embedded minimization processes. From this formulation, we have derived an hybrid proximal point algorithm that exhibits the same performance than the classical iterative decoding. This proximal point algorithm decreases at each step a well identified criterion. We have also built an hybrid minimum entropy algorithm. The minimization of the entropy leads to an improvement of the performance.

7. REFERENCES

[1] E. Zehavi, "8-PSK trellis codes for a Rayleigh fading channel," *IEEE Trans. Commun.*, vol. 40, pp. 873–883, May 1992.

[2] X. Li, A. Chindapol, and J.A. Ritcey, "Bit interleaved coded modulation with iterative decoding and 8-PSK signaling," *IEEE trans Commun.*, vol. 50, pp. 1250–1257, Aug 2002.

[3] S. ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE trans Commun.*, vol. 49, pp. 1727–1737, Oct 2001.

[4] H. El Gamal and A.R. Hammons, "Analysing the turbo decoder using the Gaussian approximation," *IEEE Trans. on Inform. Theory*, vol. 47, pp. 671–686, Feb. 2001.

[5] F.R. Kschischang, B.J. Frey, and H.A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. on Inform. Theory*, vol. 47, pp. 498–519, Feb. 2001.

[6] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Network of Plausible Inference*, San Francisco, CA: Morgan Kaufmann, 1988.

[7] J. M. Walsh, P.A. Regalia, and C. R. Johnson, "Turbo decoding as Iterative Constrained Maximum-Likelihood Sequence Detection," *IEEE Trans. on Inform. Theory*, vol. 52, pp. 5426–5437, Dec. 2006.

[8] B. Muquet, P. Duhamel, and M. de Courville, "A geometrical interpretation of iterative turbo decoding," in *Proc. Int. Symposium on Inform. Theory*, Lausanne, Switzerland, May 2002.

[9] F. Alberge, "Iterative decoding as Dykstra's algorithm with alternate I-projection and reverse I-projection," in *EUSIPCO Proc.*, Lausanne, Switzerland, August 2008.

[10] T. Richardson, "The geometry of turbo-decoding dynamics," *IEEE Trans. on Inform. Theory*, vol. 46, no. 1, pp. 9–23, 2000.

[11] G. Caire, G.Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. on Inform. Theory*, vol. 4, pp. 927–946, May 1998.

[12] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo codes," in *Proc. IEEE Int. Conf. Commun*, 1993, pp. 1064–1070.

[13] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, "A soft-input soft-output APP module for iterative decoding of concatenated codes," *IEEE Commun. Letters*, vol. 1, pp. 22–24, Jan 1997.

[14] M. Moher and T.A. Gulliver, "Cross-entropy and iterative decoding," *IEEE Trans. on Inform. Theory*, vol. 44, no. 7, pp. 3097–3104, Nov. 1998.

[15] F.J. Luque, "Asymptotic convergence analysis of the proximal point algorithm," *SIAM Journal on Control and Optimization*, vol. 22, no. 2, pp. 277–293, 1984.

[16] L. Kocarev, F. Lehmann, G.M. Maggio, B. Scanavino, Z. Tasev, and A. Vardy, "Nonlinear dynamics of iterative decoding systems: analysis and applications," *IEEE Trans. on Infor. Theory*, vol. 52, no. 4, pp. 1366–1384, 2006.