Data-aided Channel Estimation for Turbo-PIC MIMO Detectors

Mohammad-Ali Khalighi, Joseph J. Boutros, and Jean-François Hélard

Abstract— We consider data-aided channel estimation for multiple input multiple output (MIMO) systems when iterative parallel interference cancellation (PIC) is performed for signal detection. We compare some data-aided channel estimation methods based on expectation maximization (EM) algorithm or on hard estimated transmit symbols. In particular, we propose a modified EM-based approach and show that when few iterations are to be performed, it provides considerable performance improvement.

Index Terms— MIMO systems, channel estimation, iterative detection, Expectation Maximization (EM), interference cancellation, turbo-PIC.

I. INTRODUCTION

MULTIPLE input multiple output (MIMO) systems can potentially provide very high data rates in a rich-scattering propagation medium [1]. In the case of coherent signal detection at receiver, channel estimation is necessary. The classical approach of time-multiplexed pilots and data results in an important loss in spectral efficiency for large number of transmit antennas and when channel has rapid variations [2]. Hence, it is of interest to transmit few pilot symbols and to use data-aided (also called semi-blind) channel estimation schemes that make use of data symbols in addition to pilots. In this work, we consider iterative signal detection based on parallel interference cancellation (PIC) [3] and study data-aided channel estimation based on the expectation maximization (EM) algorithm [4]. We first consider the classical implementation of EM, presented in [5], and propose an appropriate formulation of EM under turbo-PIC detection. We next propose a modification to the classical EM algorithm that leads to a better convergence at the receiver. We also study a simple data-aided approach, called Th-HD (for Thresholded Hard Decisions) and initially proposed in [3], that uses reliable detected data symbols in channel estimation in the same way as pilots. Frequency non-selective, uncorrelated quasi-static Rayleigh fading and QPSK modulation are considered.

II. SIGNAL TRANSMISSION AND DETECTION

At the transmitter, bit-interleaved coded modulation is used using a non-recursive non-systematic convolutional (NRNSC) code. After being randomly interleaved, encoded data bits are mapped to symbols x that are power normalized, and then
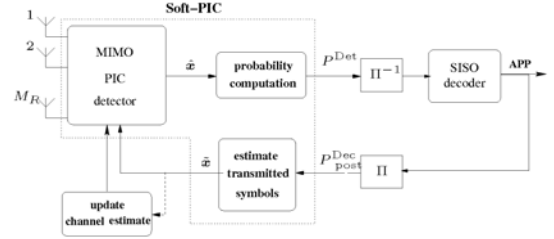
Fig. 1. Block diagram of turbo-PIC detector.

transmitted on M_T antennas. We call a vector $\boldsymbol{x}$ of M_T symbols a *compound symbol*. Let $\boldsymbol{H}$ be the channel matrix of dimension $(M_R \times M_T)$ and n the vector of complex circularly symmetric AWGN of variance σ_n^2. The vector $\boldsymbol{y}$ of received signals on M_R antennas at a time sample n is:

$$\boldsymbol{y}[n] = \boldsymbol{H} \ \boldsymbol{x}[n] + \mathsf{n}[n]. \tag{1}$$

At the receiver, iterative symbol detection and channel decoding is done as shown in Fig.1. The detector is composed of two main blocks of Soft-PIC and maximum a posteriori (MAP) soft-input soft-output (SISO) decoder. Soft-PIC is described in detail in [3], [6]. At the first iteration, PIC detection is equivalent to minimum mean-square error (MMSE) filtering. The detected symbol of the transmit antenna #k is:

$$\hat{x}_k^{(1)} = \boldsymbol{h}_k^{\dagger} \big(\boldsymbol{H} \boldsymbol{H}^{\dagger} + \sigma_n^2 \boldsymbol{I} \big)^{-1} \boldsymbol{y}. \tag{2}$$

$\boldsymbol{h}_k$ is the kth column of $\boldsymbol{H}$, $.^{\dagger}$ denotes conjugate-transpose, and $\boldsymbol{I}$ is the Identity matrix. In next iterations, we estimate the transmitted symbols using the a posteriori probabilities (APP) at the SISO decoder output. These estimates $\tilde{\boldsymbol{x}}$ are used in PIC detector to reduce the co-antenna interference (CAI) before zero-forcing filtering. We use a suboptimal filtering proposed in [3] that assumes almost perfect CAI cancellation. Let $\boldsymbol{H}_k$ of dimension $(M_R \times (M_T - 1))$ be $\boldsymbol{H}$ with its kth column removed, and $\tilde{\boldsymbol{x}}_k^{(m-1)}$ of dimension $((M_T - 1) \times 1)$ the vector of estimated symbols at iteration $(m - 1)$ with its kth entry removed. The detected k^{th} symbol at iteration m is:

$$\hat{x}_k^{(m)} = \big(\boldsymbol{h}_k^{\dagger} \boldsymbol{h}_k \big)^{-1} \boldsymbol{h}_k^{\dagger} \Big(\boldsymbol{y} - \boldsymbol{H}_k \ \tilde{\boldsymbol{x}}_k^{(m-1)} \Big). \tag{3}$$

III. CHANNEL ESTIMATION

Let N_{ps} and N_{ds} be respectively the number of pilot and data compound symbols in a frame, corresponding to N_p and N_d pilot and data bits. For instance, with B bits per symbol, $N_p = N_{ps} B M_T$. Also, let $N_s = N_{ds} + N_{ps}$. The maximum likelihood channel estimate based only on pilot symbols is given in (4) where $\boldsymbol{x}_p$ denotes a compound pilot symbol, i.e., a vector of M_T pilot symbols.

$$\hat{\boldsymbol{H}}^p = \Big(\sum_{n=1}^{N_{ps}} \boldsymbol{y}[n] \ \boldsymbol{x}_p^{\dagger}[n] \Big) \Big(\sum_{n=1}^{N_{ps}} \boldsymbol{x}_p[n] \ \boldsymbol{x}_p^{\dagger}[n] \Big)^{-1} \tag{4}$$

A. Classical EM-based data-aided estimation

The classical formulation of EM that we call Mix-EM (or Mixing-EM) considers the ensemble of pilot and data symbols as *missing data* [5]. In this way, after simplifying the expectation and maximization steps, we come to the following estimation update equation for $\boldsymbol{H}$ at iteration $m > 1$ [5]:

$$\hat{\boldsymbol{H}} = \overline{\boldsymbol{R}}_{yx} \, \overline{\boldsymbol{R}}_x^{-1} \qquad (5)$$

where

$$\overline{\boldsymbol{R}}_{yx} = \sum_{n=1}^{N_s} \sum_{q=1}^{Q} \boldsymbol{y}[n] \, \boldsymbol{x}_q^{\dagger} \, APP_n(\boldsymbol{x}_q) \qquad (6)$$

$$\overline{\boldsymbol{R}}_x = \sum_{n=1}^{N_s} \sum_{q=1}^{Q} \boldsymbol{x}_q \, \boldsymbol{x}_q^{\dagger} \, APP_n(\boldsymbol{x}_q) \qquad (7)$$

$Q = 2^{BM_T}$ and $\boldsymbol{x}_q$ is the qth among Q possible compound symbols whose probability of transmission $APP_n(\boldsymbol{x}_q)$ is calculated using the previous APPs at the SISO decoder output:

$$APP_n(\boldsymbol{x}_q) \propto \prod_{i=1}^{BM_T} P_{\text{post}}^{\text{Dec}}(c_{q,i}) \qquad (8)$$

$P_{\text{post}}^{\text{Dec}}(c_{q,i})$ is the APP corresponding to the ith bit of $\boldsymbol{x}_q$, $c_{q,i}$.[1] For the first iteration, we use $\hat{\boldsymbol{H}} = \hat{\boldsymbol{H}}^p$ from (4). The computational complexity of (6) and (7) increases exponentially with BM_T. However, it can be shown that $\sum_{q=1}^{Q} \boldsymbol{x}_q APP(\boldsymbol{x}_q) = \tilde{\boldsymbol{x}}$. But $\tilde{\boldsymbol{x}}$ is already computed in Soft-PIC and there is no need to compute it again. Now, (6) and (7) can be written as:

$$\overline{\boldsymbol{R}}_{yx} = \sum_{n=1}^{N_s} \boldsymbol{y}[n] \, \tilde{\boldsymbol{x}}^{\dagger}[n] \qquad (9)$$

$$\overline{\boldsymbol{R}}_x(i,j) = \begin{cases} N_s & ; \quad i = j \\ \sum_{n=1}^{N_s} \tilde{x}_i[n] \, \tilde{x}_j^*[n] & ; \quad i \neq j \end{cases} \qquad (10)$$

$\tilde{x}_i[n]$ is the ith entry of the vector $\tilde{\boldsymbol{x}}[n]$. So, EM is implemented in the body of Soft-PIC and in a much simpler way.

B. Modifying EM formulation

Using (1) and (9), we can write $\overline{\boldsymbol{R}}_{yx} = \boldsymbol{H} \, \boldsymbol{R}_x' + \boldsymbol{\eta}$, where

$$\boldsymbol{R}_x' \triangleq \sum_{n=1}^{N_s} \boldsymbol{x}[n] \, \tilde{\boldsymbol{x}}^{\dagger}[n] \quad , \quad \boldsymbol{\eta} = \sum_{n=1}^{N_s} \mathsf{n}[n] \, \tilde{\boldsymbol{x}}^{\dagger}[n]. \qquad (11)$$

$\boldsymbol{\eta}$, the matrix of weighted noise samples, has the auto-covariance matrix $\sigma_n^2 \boldsymbol{R}_x''$, where $\boldsymbol{R}_x'' = \sum_{n=1}^{N_s} \tilde{\boldsymbol{x}}[n] \, \tilde{\boldsymbol{x}}^{\dagger}[n]$. Now from (5) $\hat{\boldsymbol{H}}$ can be written as:

$$\hat{\boldsymbol{H}} = \boldsymbol{H} \, \boldsymbol{R}_x' \, \overline{\boldsymbol{R}}_x^{-1} + \boldsymbol{\eta} \, \overline{\boldsymbol{R}}_x^{-1}. \qquad (12)$$

We see from (10) and (11) that $\boldsymbol{R}_x' \neq \overline{\boldsymbol{R}}_x$, and hence, $\hat{\boldsymbol{H}}$ is a biased estimate. This is due to the fact that $\tilde{\boldsymbol{x}} \neq \boldsymbol{x}$ except at high enough SNR and at concluding iterations where the bias becomes negligible. The bias can degrade the receiver performance by affecting the detector convergence.

[1]Equation (8) gives an approximation of the exact $_n(\boldsymbol{x}_q)$ and is sub-optimal in terms of bit-error-rate and convergence speed, but it is much less complex than the exact calculation of APPs. For more details, see [5].

To remove the bias, we consider separately the unbiased pilot-only-based estimate $\hat{\boldsymbol{H}}^p$ and combine it with $\hat{\boldsymbol{H}}^d$, the data-based estimate via EM, to obtain $\hat{\boldsymbol{H}}$. In the sequel, we consider matrices $\overline{\boldsymbol{R}}_x$, $\boldsymbol{R}_x'$, and $\boldsymbol{R}_x''$ calculated only over data symbols, i.e., with the summations taken on N_{ds} instead of N_s. One solution is to remove the bias of $\hat{\boldsymbol{H}}^d$ by Wiener or inverse filtering, and then to combine the resulting estimate with $\hat{\boldsymbol{H}}^p$. This approach, called RB-EM in [6] (Removed Biased-EM), is computationally complex and does not provide satisfying results either [6]. A more appropriate solution that we present here, assumes large enough N_{ds} and negligible CAI on $\tilde{\boldsymbol{x}}$. Thus, neglecting the off-diagonal terms of $\overline{\boldsymbol{R}}_x$, $\boldsymbol{R}_x'$, and $\boldsymbol{R}_x''$, from (12) we obtain:

$$\hat{\boldsymbol{H}}^d \approx \frac{1}{N_{ds}} \boldsymbol{H} \, \text{diag}\left(\sum_{n=1}^{N_{ds}} x_i[n] \, \tilde{x}_i^*[n] \right) + \boldsymbol{\eta}'' \qquad (13)$$

where diag(.) denotes a diagonal matrix with its (i,i)th entry given, and $\boldsymbol{\eta}'' = \frac{1}{N_{ds}} \boldsymbol{\eta}$. The (i,j)th entry of $\hat{\boldsymbol{H}}^d$ is then,

$$\hat{H}_{ij}^d = \alpha_j H_{ij} + \eta_{ij}'' \text{ where } \alpha_j \triangleq \frac{1}{N_{ds}} \sum_{n=1}^{N_{ds}} x_j[n] \, \tilde{x}_j^*[n]. \quad (14)$$

We deduce from (14) that we should combine $\hat{\boldsymbol{H}}^d$ and $\hat{\boldsymbol{H}}^p$ columnwise by optimal coefficients a_j and b_j:

$$\hat{\boldsymbol{h}}_j = a_j \hat{\boldsymbol{h}}_j^d + b_j \hat{\boldsymbol{h}}_j^p \qquad (15)$$

The optimization criterion is to minimize the mean square estimation error variance, i.e., $\min \left\{ |a_j|^2 \hat{\sigma}_{j,d}^2 + |b_j|^2 \hat{\sigma}_{j,p}^2 \right\}$, subject to unbiased estimation, i.e., $a_j \alpha_j + b_j = 1$. Here $|.|$ stands for complex modulus, $\hat{\sigma}_{j,d}^2 = \frac{1}{N_{ds}^2} R_{xjj}''$ is the variance of $\boldsymbol{\eta}_j''$ entries, and $\hat{\sigma}_{j,p}^2 = \sigma_n^2/N_{ps}$ for mutually orthogonal pilot sequences. Using Lagrange multipliers method we obtain:

$$a_j = \frac{\alpha_j^*}{|\alpha_j|^2 + \frac{N_{ps}}{N_{ds}^2} R_{xjj}''} \ , \ b_j = \frac{\frac{N_{ps}}{N_{ds}^2} R_{xjj}''}{|\alpha_j|^2 + \frac{N_{ps}}{N_{ds}^2} R_{xjj}''} \ . \quad (16)$$

We call this solution MU-EM (*Modified Unbiased*-EM) [6]. To calculate α_j, we replace x_j in (14) by the hard decision on $\tilde{x}_j$, assuming a low enough error probability. Note that at low SNR and in first iterations, we have $\alpha_j \approx 0$ and $R_{xjj}'' \approx 0$, and so, $\hat{\boldsymbol{H}} \approx \hat{\boldsymbol{H}}^p$. On the other hand, at high SNR and in concluding iterations, we have $\alpha_j \approx 1$ and $R_{xjj}'' \approx N_{ds}$.

C. Th-HD estimation method

With Th-HD, the APPs at the SISO decoder output are compared to a threshold P_{TH} in order to choose reliable detected data symbols. If APP $> P_{TH}$ the hard decision 1 is assigned to the corresponding bit, and if APP $< (1 - P_{TH})$, the hard decision 0 is assigned to it. Hard-detected compound symbols are then used in channel estimation like pilots. Simulation results (not presented) show poor dependence of the Th-HD performance to the choice of P_{TH} and an optimal P_{TH} close to 0.5. This comes to using hard estimates of data symbols in channel estimation in the same way as pilots.
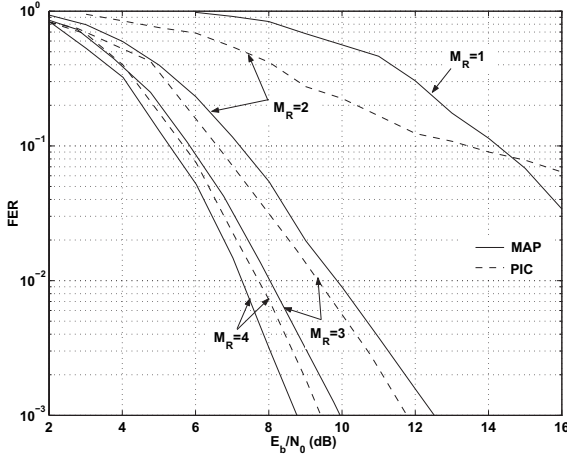
Fig. 2.   Comparing turbo-PIC and turbo-MAP; $M_T = 4$, NRNSC code $(5\ 7)_8$, $N_d = 512$, perfect CSI, fifth iteration of the detector.
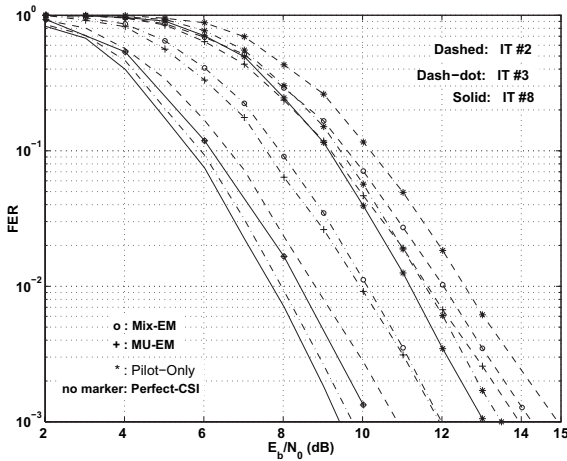

Fig. 3.   Comparing perfect CSI and pilot-only, Mix-EM, and MU-EM estimation methods for second, third, and eighth iteration; $M_T = M_R = 4$, NRNSC code $(5\ 7)_8$, $N_d = 512$, $N_p = 40$.

IV. SIMULATION RESULTS

Almost the same N_{ds} is considered for different M_T. Optimal pilot sequences are used with N_{ps} a little larger than the limit of identifiability BM_T. Let us first compare the performances of the turbo-PIC and turbo-MAP [5] detectors when perfect channel state information (CSI) is available at receiver. Fig.2 contrasts frame error rate (FER) curves for $M_T = 4$ and $M_R \leq M_T$. Receiver array gain is included in E_b/N_0. We see that thanks to iterative detection, our PIC-based detector works efficiently for $M_R = 3$ with a reasonable loss compared to turbo-MAP. For $M_R < 3$, the detector does not provide satisfying performance although it does not diverge. Without coming to an analytical proof, we verified that turbo-PIC works efficiently for $M_R > M_T/2$ where it would be preferred to turbo-MAP due to its considerably lower complexity. We also verified that turbo-PIC is slightly less sensitive to channel estimation errors than turbo-MAP.

Fig.3 compares FERs after two, three, and eight iterations for the cases of perfect CSI and estimation based on pilots only, Mix-EM, and MU-EM. For the two former cases, full convergence is attained after five iterations. We see that after
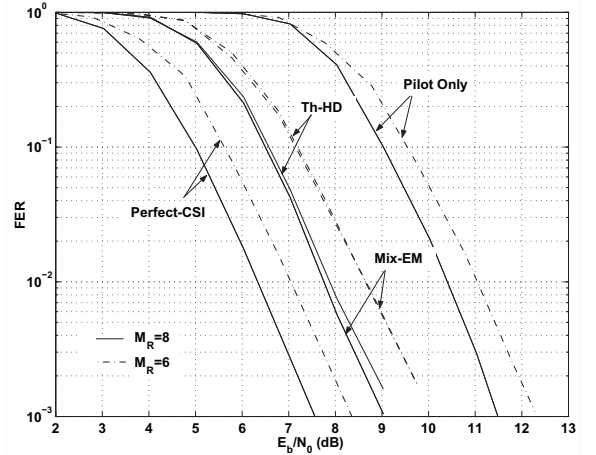

Fig. 4.   Comparing Pilot-Only, Mix-EM, and Th-HD channel estimation methods for $(M_T = 8\ M_R = 8)$ and $(M_T = 8\ M_R = 6)$ systems. Eighth iteration of detector, NRNSC code $(5\ 7)_8$, $N_d = 1024$, $N_p = 160$.

eight iterations, we gain 2.7 dB in E_b/N_0 at FER $= 10^{-2}$ with Mix-EM compared to pilot-only estimation. Comparing Mix-EM and MU-EM, we notice that for the second and third iterations, MU-EM gives respectively a gain of 0.4 dB and 0.18 dB in SNR at FER=10^{-2}. However, if we proceed to the eighth iteration, their performances become equivalent. At the first iterations, MU-EM works better than Mix-EM by imposing a low weight to $\hat{\boldsymbol{H}}^d$ and by privileging $\hat{\boldsymbol{H}}^p$. Fig.4 compares Th-HD and EM-based methods for $M_T = 8$ and $M_R = 6, 8$ after eight iterations. It is seen that using hard estimated data symbols in channel estimation (Th-HD with $P_{TH} = 0.5$) provides almost the same performance as using soft estimates (Mix-EM). However, since soft estimates are already available in turbo-PIC, Mix-EM is preferred.

V. CONCLUSION

We proposed an appropriate formulation for the classical Mix-EM under turbo-PIC detection. We also proposed MU-EM that permits a better convergence of the detector and provides interesting performance improvement for small number of iterations. Use of hard estimates of data symbols in channel estimation was shown to provide the same performance as Mix-EM. As a side contribution, we also showed the interest of using turbo-PIC for $M_R > M_T/2$.

REFERENCES

[1] E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecomm.*, vol. 10, pp. 585–595, Nov./Dec. 1999.
[2] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inform. Theory*, vol. 49, pp. 951–963, Apr. 2003.
[3] M. Sellathurai and S. Haykin, "Turbo-BLAST for wireless communications: theory and experiments," *IEEE Trans. Signal Processing*, vol. 50, pp. 2538–2546, Oct. 2002.
[4] T. K. Moon, "The expectation-maximization algorithm," *IEEE Signal Processing Mag.*, vol. 13, pp. 47-60, Nov. 1996.
[5] J. Boutros, F. Boixadera, and C. Lamy, "Bit-interleaved coded modulations for multiple-input multiple-output channels," in *Proc IEEE ISSSTA'00*, vol. 1, pp. 123-126.
[6] M. A. Khalighi and J. Boutros, "Channel estimation in turbo-BLAST detectors using EM algorithm," in *Proc IEEE ICASSP'05*, vol. 3, pp. 1037–1040.