Self and Turbo Iterations for MIMO Receivers and Large-Scale Systems

Irene Santos and Juan José Murillo-Fuentes

Abstract—We investigate a turbo soft detector based on the expectation propagation (EP) algorithm for large-scale multiple-input multiple-output (MIMO) systems. Optimal detection in MIMO systems becomes computationally unfeasible for high-order modulations and/or large number of antennas and approximations are used instead. We propose a new EP-based detector with enhanced estimation in terms of complexity and performance. Specifically, we improve the convergence of the self-iterated EP stage by re-defining the priors to better characterize the information returned by the decoder once the turbo procedure starts. We also review the EP parameters to avoid instabilities when using high-order modulations and to reduce the computational complexity. Simulation results illustrate the robustness and enhanced performance of this novel detector in comparison to previous approaches found in the literature.

Index Terms—Expectation propagation (EP), MMSE, low-complexity, MIMO, turbo detection, IDD, feedback.

I. INTRODUCTION

MULTIPLE-INPUT multiple-output (MIMO) systems provide spatial multiplexing gain, among other capabilities [1]. This gain increases with the number of antennas. To further improve the data rate, higher-order modulations are also used. In reception, the complexity of the optimal detector grows exponentially with the constellation order and the number of antennas. Hence, approximations are needed [2]–[5]. These approximations should provide probabilistic outputs, from which modern decoders and turbo schemes greatly benefit [6]–[8].

Recently, the expectation propagation (EP) algorithm [9] has been proposed as a better approach to approximate the posterior. This algorithm has already been applied to equalization [10]–[12] and MIMO detection [13], [14]. In these works, it was shown that the EP detector improves the performance of linear minimum mean square error (LMMSE), LMMSE with successive interference cancellation (SIC), Gaussian tree approximation (GTA) [4], GTA-SIC and channel hardening-exploiting message passing (CHEMP) [5]. Also, it was concluded that its computational complexity is proportional to the one of the LMMSE algorithm. The extension to MIMO turbo detection, or iterative detection and decoding (IDD), was introduced in [15], [16]. Approach [15] is based on EP implemented with a damping procedure. In this method, uniform priors are used to approximate the feedback from the decoder within the EP damping steps. In contrast, the authors in [16] assume non-uniform priors and do not include iterations within the EP stage, that it is written in message passing form.

In this letter, we focus on the information exchange between the EP detector and a soft-decision channel decoder. We propose to use non-uniform priors distributed according to the channel decoder output within the EP approach. Then, this procedure is repeated at the EP stage to update its estimation, including a damping procedure. Following the guidelines in [12], we also optimize the EP parameters to avoid instabilities with high-order modulations and reduce the computational complexity of the algorithm. We show experimentally that this novel EP approach greatly improves the convergence and the performance in comparison to previous EP approaches [15], [16] and the LMMSE.

II. SYSTEM MODEL

We consider a MIMO system equipped with $N_t$ transmit and $N_r$ receive antennas, where $N_r \geq N_t$. The information bit vector, $a = [a_1, \ldots, a_K]$\top, is encoded into the codeword $b = [b_1, \ldots, b_V]$\top with a code rate $R = K/V$. This codeword is modulated into complex-valued symbols of an $M$-ary alphabet $A$. The modulated symbols are partitioned into $P$ blocks of length $N_t$, $[u_1^T, \ldots, u_P^T]^T$, where $u_p^T = [u_{p,1}, \ldots, u_{p,N_t}]$. Each block is demultiplexed into $N_t$ substreams and transmitted over the channel. Hereafter, we will omit the index $p$.

The channel is completely specified by the known noise variance, $\sigma_w^2$, and the channel gains, $h_{j,k}$, from transmit antenna $k$ to receive antenna $j$. These gains are the entries of the $N_r \times N_t$ channel matrix, $H$, and are assumed known to the receiver. The received signal for a channel use, $y = [y_1, \ldots, y_{N_r}]^T$, is given by $y = Hu + w$, where $w \sim \mathcal{CN}(w; 0, \sigma_w^2 I)$ is a complex-valued additive white Gaussian noise (AWGN) vector. We assume the coherence time to be larger than the block duration.

The posterior probability of the transmitted symbol vector $u$ given the whole vector of observations $y$ yields

$$p(u|y) \propto \mathcal{CN}(y; Hu, \sigma_w^2 I) \prod_{k=1}^{N_t} p_D(u_k), \quad (1)$$

where the prior, $p_D(u_k)$, is a discrete distribution. In turbo detection, this prior is updated with the output of the channel decoder. The optimal detector solves (1) by checking for all possible combinations, having complexity of $O(M^{N_t})$. For large-scale systems, i.e., large values of $M^{N_t}$, the optimal solution becomes unaffordable.

The extrinsic distribution computed by the decoder is demapped and given to the decoder as extrinsic log-likelihood ratios (LLRs). The channel decoder computes an estimation
of the information bit vector, $\mathbf{a}$, and the extrinsic LLRs of the coded bits, $L_D(b_n)$. These LLRs are again mapped and fed back to the detector as updated priors, $p_D(u)$. This is an iterative process where the steps above are repeated $T$ times, the maximum number of turbo iterations, or until convergence.

III. THE BLOCK-EP DETECTOR

The EP algorithm iteratively provides a feasible Gaussian approximation, $q^{[\ell]}(u)$, to the posterior distribution in (1), $p(u|y)$, where $\ell$ denotes the iteration number. This approximation of the posterior is obtained by replacing in (1) the product of non-Gaussian terms, $p_D(u_k)$, by a product of to be estimated Gaussians, denoted by $t_k^{[\ell]}(u_k) = \mathcal{CN}(u_k; \mu_k^{[\ell]}, \sigma_k^{[\ell]} )$. The approximated posterior factorizes as

$$q^{[\ell]}(u) = \mathcal{CN}(u; \mu_{[\ell]}^q, \Sigma_{[\ell]}^q) \propto \mathcal{CN}(y; \mathbf{H}u, \sigma_w^2 I) \prod_{k=1}^{N_t} t_k^{[\ell]}(u_k)$$

where $\Sigma_{[\ell]} = \text{diag}([\sigma_{t_1}^{[\ell]}; \ldots; \sigma_{t_{N_t}}^{[\ell]}])$, $\mu_{[\ell]}^q = [\mu_1^{[\ell]}, \ldots, \mu_{N_t}^{[\ell]}]^T$. The mean and covariance of $q^{[\ell]}(u)$ can be computed as (see (A.7) in [17]),

$$\Sigma_{[\ell]}^q = \left( \sigma_w^{-2} \mathbf{H}^T \mathbf{H} + \Sigma_{[\ell]}^{-1} \right)^{-1},$$

$$\mu_{[\ell]}^q = \Sigma_{[\ell]}^q \left( \sigma_w^{-2} \mathbf{H}^T y + \Sigma_{[\ell]}^{-1} \mu_{[\ell]} \right).$$

The $k$th marginal of $q^{[\ell]}(u)$ can be easily obtained from this expression as $q^{[\ell]}(u_k) \sim \mathcal{CN}(u_k; \mu_k^{[\ell]}, \sigma_k^{2[\ell]})$, where $\mu_k^{[\ell]}$ is the $k$th entry of $\mu_{[\ell]}^q$ and $\sigma_k^{2[\ell]}$ is the $k$th diagonal entry of matrix $\Sigma_{[\ell]}^q$. Bearing these expressions in mind, we next face the update of the factors $t_k^{[\ell]}(u_k)$ by means of the EP algorithm.

The EP is based on the minimization of the Kullback-Leibler (KL) divergence between the true distribution in (1) and its Gaussian approximation in (2). This is equivalent to matching their moments. Accordingly, along $\ell = 1, \ldots, S$ iterations, we estimate the new values of the means and variances, $\mu_k^{[\ell+1]}$ and $\sigma_k^{2[\ell+1]}$, by

$$q^{[\ell]}(u_k) \overset{\text{moment matching}}{\rightarrow} t_k^{[\ell+1]}(u_k).$$

We define $q_E^{[\ell]}(u_k)$, that plays the role of an extrinsic marginal distribution, as

$$q_E^{[\ell]}(u_k) = q^{[\ell]}(u_k)/t_k^{[\ell]}(u_k) = \mathcal{CN}(u_k; \mu_k^{[\ell]}, \sigma_k^{2[\ell]}).$$

Using (2)-(4) to compute $q^{[\ell]}(u_k)$ and by the definition of the approximating factors, $t_k^{[\ell]}(u_k)$, it follows that (see (A.7) in [17]),

$$\mu_k^{[\ell]} = \frac{\mu_k^{[\ell]} \sigma_k^{2[\ell]} - \mu_k^{[\ell]} \sigma_k^{2[\ell]}}{\sigma_k^{2[\ell]} - \sigma_k^{2[\ell]}}, \quad \sigma_k^{2[\ell]} - \sigma_k^{2[\ell]}.$$

Algorithm 1 Moment Matching and Damping (MMD)

Given inputs: $p_D(u_k)$, $\mu_k^{[\ell]}$, $\sigma_k^{2[\ell]}$, for $k = 1, \ldots, N_t$ and $y$, $\epsilon$, $\beta$.

1. Compute $q^{[\ell]}(u)$ in (2)-(4) and its marginals, $q^{[\ell]}(u_k)$.
2. Compute the extrinsic marginal distributions, $q_E^{[\ell]}(u_k)$ in (6)-(7).
3. Compute the moments of $p^{[\ell]}(u_k)$ in (8), i.e., $\mu_k^{[\ell]}$ and $\sigma_k^{2[\ell]}$. Set a minimum allowed variance as $\sigma_k^{2[\ell]} = \max(\epsilon, \sigma_k^{2[\ell]})$.
4. Match moments: compute new values of the moments of $q_k^{[\ell]}(u_k)$ using (5),

$$\nu_{kq,n}^{[\ell]} = \frac{\mu_k^{[\ell]} \sigma_k^{2[\ell]} - \mu_k^{[\ell]} \sigma_k^{2[\ell]}}{\sigma_k^{2[\ell]} - \sigma_k^{2[\ell]}}, \quad \sigma_k^{2[n+1]} = \frac{\sigma_k^{2[\ell]} \sigma_k^{2[\ell]} - \sigma_k^{2[\ell]} \sigma_k^{2[\ell]}}{\sigma_k^{2[\ell]} - \sigma_k^{2[\ell]}}.$$

5. Run damping: Update the values as

$$\sigma_k^{2[\ell]} = \beta \sigma_k^{2[n+1]} + (1 - \beta) \sigma_k^{2[\ell]};$$

$$\nu_{kq,n}^{[\ell]} = \frac{\mu_k^{[\ell]} \sigma_k^{2[\ell]} - \mu_k^{[\ell]} \sigma_k^{2[\ell]}}{\sigma_k^{2[\ell]} - \sigma_k^{2[\ell]}}, \quad \sigma_k^{2[n+1]} = \frac{\sigma_k^{2[\ell]} \sigma_k^{2[\ell]} - \sigma_k^{2[\ell]} \sigma_k^{2[\ell]}}{\sigma_k^{2[\ell]} - \sigma_k^{2[\ell]}}.$$

Finally, we update the factor $t_k^{[\ell+1]}(u_k)$ with the new values obtained from (5) at every iteration $\ell$, using a damping approach. The full procedure is described in Algorithm 1, where the selection of parameters $\epsilon$ and $\beta$ will be discussed later in this section. The control of negative variances proposed in [14] is also included. Note that to derive the new moments of $t_k^{[\ell+1]}(u_k)$ from (5), we need to compute the moments of

$$p^{[\ell]}(u_k) = q_E^{[\ell]}(u_k)p_D(u_k).$$

We will denote its first and second moments by $\mu_k^{[\ell]}$ and $\sigma_k^{2[\ell]}$, respectively.

A. Turbo Detection

Our proposal for the EP procedure and turbo detector is detailed in Algorithm 2, where the superscript $[\ell]$ with $\ell = 1, \ldots, S$ denotes the iteration number of the EP, that we will refer to as self-iteration, and $(t)$ with $t = 0, \ldots, T$ indicates the turbo iteration number. Unlike [15], at Step 2 of this algorithm the priors used in the moment matching (see (5)) are the non-uniform probability mass function (pmf) at the output of the decoder, $p_D^{(t)}(u)$. For this reason, we refer to this approach as non-uniform block expectation propagation (nuBEP) detector. Note that for $T = 0$ we have a standalone version, with no turbo detection. Also, we may have different values of $\beta$ for each turbo iteration.
Algorithm 2 nuBEP Turbo Decoder for MIMO

Given inputs: \( y, \epsilon, \beta(0), ..., \beta(T) \).

Initialization: Set \( p_D^{(0)}(u_k) = \frac{1}{N} \sum_{u \in A} \delta(u_k - u) \) for \( k = 1, ..., N_t \).

Turbo Iteration:
for \( t = 0, ..., T \) do

1) Compute the mean \( \mu_{tk}^{[1]} \) and variance \( \sigma_{tk}^{[2]} \) of \( p_D^{(t)}(u_k) \).

2) Run the moment matching procedure in Algorithm 1 with inputs \( p_D^{(t)}(u_k), \mu_{tk}^{[\ell]}, \sigma_{tk}^{[2\ell]}(\epsilon), \beta(\epsilon), \beta(t) \), to obtain \( \sigma_{tk}^{[2(\ell+1)]} \) and \( \mu_{tk}^{[(\ell+1)]} \).

3) With the values \( \sigma_{tk}^{[2(S+1)]}, \mu_{tk}^{[S+1]} \) obtained, calculate the extrinsic distribution \( q_E^{(S+1)}(u_k) \) as in (6).

4) Demap the extrinsic distribution and compute the extrinsic LLRs, \( L_E(b_n) \), and deliver it to the channel decoder.

5) Run the channel decoder to output \( p_D^{(t+1)}(u_k) \).

end for

B. EP Parameters

The update of the EP solution is a critical issue due to instabilities, particularly for high-order modulations. In this subsection, we propose new values for the parameters involved: the minimum allowed variance (\( \epsilon \)), a damping factor (\( \beta \)) and the number of EP iterations (\( S \)). The first two parameters determine the convergence speed of the algorithm and control the presence of instabilities. The computational complexity of the algorithm depends linearly on \( S \).

In [14], the authors iterated the EP algorithm over \( S = 10 \) times and set \( \beta = 0.95 \) to speed up convergence. To avoid instabilities, they proposed to start with a high value of the minimum allowed variance (\( \epsilon \)), a damping factor (\( \beta \)) and the number of EP iterations (\( S \)). The first two parameters determine the convergence speed of the algorithm and control the presence of instabilities. The computational complexity of the algorithm depends linearly on \( S \).

In Table I, we describe the values of the EP parameters used in the current proposal (nuBEP) and in the approaches in [14] and [16], hereafter referred to as EPD and MPEP, respectively.

Fig. 1. BER along \( N_t E_s/N_0 \) for the nuBEP (o), EPD [14] (□), MPEP [16] (+) and LMMSE (○) detectors, averaged over 100 randomly channels with 128-QAM and (a) \( 6 \times 6 \), (b) \( 32 \times 32 \) and (c) \( 100 \times 100 \) antennas after \( T = 5 \) turbo iterations.

B. EP Parameters

The update of the EP solution is a critical issue due to instabilities, particularly for high-order modulations. In this subsection, we propose new values for the parameters involved: the minimum allowed variance (\( \epsilon \)), a damping factor (\( \beta \)) and the number of EP iterations (\( S \)). The first two parameters determine the convergence speed of the algorithm and control the presence of instabilities. The computational complexity of the algorithm depends linearly on \( S \).

In [14], the authors iterated the EP algorithm over \( S = 10 \) times and set \( \beta = 0.95 \) to speed up convergence. To avoid instabilities, they proposed to start with a high value of the minimum allowed variance (\( \epsilon \)), a damping factor (\( \beta \)) and the number of EP iterations (\( S \)). The first two parameters determine the convergence speed of the algorithm and control the presence of instabilities. The computational complexity of the algorithm depends linearly on \( S \).

In Table I, we describe the values of the EP parameters used in the current proposal (nuBEP) and in the approaches in [14] and [16], hereafter referred to as EPD and MPEP.
respectively. The computational complexity per turbo iteration of the MPEP [16], EPD [14] and nuBEP is \( S + 1 \) times the complexity of the LMMSE, i.e., \( O((S+1)N^3_t) \) where \( S = 1 \) for the MPEP, \( S = 10 \) for the EPD and \( S = 3 \) for the nuBEP.

IV. SIMULATION RESULTS

In this section we illustrate the performance of the proposed nuBEP turbo detector and compare its performance to the ones of the EPD [14] and the MPEP [16]. We also depict the BER of the LMMSE. We do not include the sphere decoding (SD) [2], Markov chain Monte Carlo (MCMC) [3], LMMSE-SIC, GTA [4] or CHEMP [5] algorithms in the simulations because it has already been shown that EPD quite outperforms these approaches [13]–[15]. The modulator uses a Gray mapping and a 128-QAM constellation. The results are averaged over 100 random channels and \( 10^4 \) random encoded words of length \( V = 4096 \) (per channel realization). Each channel coefficient is independent and identically Gaussian distributed with zero mean and unit variance. A \((3,6)\)-regular low-density parity-check (LDPC) of rate 1/2 is used. The absolute value of LLRs given to the decoder is limited to 5 in order to avoid very confident probabilities. The decoder runs a maximum of 100 iterations. A number of \( T = 5 \) turbo iterations were run.

In Fig. 1 we include the BER obtained for a system with \( N_t = N_r = 6 \) (a), \( N_t = N_r = 32 \) (b) and \( N_t = N_r = 100 \) (c) antennas. It can be observed that the EPD [14] improves the performance of the LMMSE (at low \( E_b/N_0 \) regime) but it is far from the results of the nuBEP and the MPEP. The reason is that the true prior used by the EPD in the moment matching procedure is set to a uniform distribution, while nuBEP and MPEP use a non-uniform one given by the output of the decoder. This choice better characterizes the prior after the turbo feedback. The MPEP approach [16] does not achieve the performance of the nuBEP because it just computes one iteration of the EP algorithm with no damping procedure, exhibiting convergence instabilities at large SNR in Fig. 1 (a). The new proposed approach, nuBEP, exhibits the most accurate and robust performance. More specifically, for moderate systems, see Fig. 1 (a), the nuBEP has gains of 14 dB and 6 dB with respect to the LMMSE and MPEP, respectively, while for a large number of antennas, see Fig. 1 (b)-(c), it has an improvement of 6-7.5 dB with respect to the LMMSE and of 1.5 dB compared with the MPEP algorithm. Note also that, for a large number of antennas, see Fig. 1 (b)-(c), the EPD approach exhibits important instabilities at large \( N_t E_b/N_0 \) since its parameters are not optimized for large-scale constellations.

V. CONCLUSION

We have proposed a new EP-based turbo detector for large-scale MIMO systems, the nuBEP. This detector quite outperforms the LMMSE and previous EP-based detectors for scenarios with large number of antennas and/or high-order modulations. Specifically, it uses a non-uniform prior, rather than a uniform one as in [14]. This prior better characterizes the true prior used in the self-iterations of the EP algorithm, along the moment matching procedure, once the turbo procedure has started. The proposed detector also optimizes its parameters to avoid instabilities at large \( E_b/N_0 \) and to reduce its complexity. Specifically, in turbo detection, it reduces the number of EP iterations from 10 (in [14]) to 3. It also outperforms the EP detector in [16] since we include a self-iterated stage with damping and a different control of negative variances. These features endow the nuBEP approach with a more accurate solution. Simulations results show that the proposed nuBEP turbo detector has gains in the range 5-10 dB with respect to the EPD [14] and 1.5-6 dB compared with the MPEP [16]. LDPC codes were used. Further issues on the optimization of the channel decoder and its outputs or the hardware implementation [7], [8] remain as future research.

REFERENCES

[1] F. Rusek et al., “Scaling up MIMO: Opportunities and challenges with very large arrays,” IEEE Signal Process Mag., vol. 30, no. 1, pp. 40–60, Jan. 2013.
[2] C. Studer, A. Burg, and H. Bolsckai, “Soft-output sphere decoding: Algorithms and VLSI implementation,” IEEE J. Sel. Areas Commun., vol. 26, no. 2, pp. 290–300, Feb. 2008.
[3] T. Datta, N. A. Kumar, A. Chockalingam, and B. S. Rajan, “A novel Monte-Carlo-sampling-based receiver for large-scale uplink multiuser MIMO systems,” IEEE Trans. Veh. Technol., vol. 62, no. 7, pp. 3019–3038, Sep. 2013.
[4] J. Goldberger and A. Lešhem, “MIMO detection for high-order QAM based on a Gaussian tree approximation,” IEEE Trans. Inf. Theory, vol. 57, no. 8, pp. 4973–4982, Aug. 2011.
[5] T. L. Narasimhan and A. Chockalingam, “Channel hardening-exploiting message passing (CHEMP) receiver in large-scale MIMO systems,” IEEE J. Sel. Topics Signal Process., vol. 8, no. 5, pp. 847–860, Oct. 2014.
[6] X. Wang and H. V. Poor, “Iterative (turbo) soft interference cancellation and decoding for coded CDMA,” IEEE Trans. Commun., vol. 47, no. 7, pp. 1046–1061, Jul. 1999.
[7] A. G. D. Uchoa, C. T. Healy, and R. C. de Lamare, “Iterative detection and decoding algorithms for MIMO systems in block-fading channels using LDPC codes,” IEEE Trans. Veh. Technol., vol. 65, no. 4, pp. 2735–2741, Apr. 2016.
[8] W.-C. Sun, W.-H. Wu, C.-H. Yang, and Y.-L. Ueng, “An iterative detection and decoding receiver for LDPC-coded MIMO systems,” IEEE Trans. Circuits Syst. I, Reg. Papers, vol. 62, no. 10, pp. 2512–2522, Oct. 2015.
[9] T. P. Minka, “A family of algorithms for approximate Bayesian inference,” Ph.D. dissertation, Dept. Elect. Eng. Comput. Sci., Massachusetts Inst. Technol., Cambridge, MA, USA, 2001.
[10] I. Santos, J. J. Murillo-Fuentes, R. Boloix-Tortosa, E. Arias-de-Reyna, and P. M. Olmos, “Expectation propagation as turbo equalizer in ISI channels,” IEEE Trans. Commun., vol. 65, no. 1, pp. 360–370, Jan. 2017.
[11] I. Santos, J. J. Murillo-Fuentes, E. Arias-de-Reyna, and P. M. Olmos, “Probabilistic equalization with a smoothing expectation propagation approach,” IEEE Trans. Wireless Commun., vol. 16, no. 5, pp. 2950–2962, May 2017.
[12] I. Santos, J. J. Murillo-Fuentes, E. Arias-de-Reyna, and P. M. Olmos, “Turbo EP-based equalization: A filter-type implementation,” IEEE Trans. Commun., vol. 66, no. 9, pp. 4259–4270, Sep. 2018.
[13] J. Céspedes, P. M. Olmos, M. Sánchez-Fernández, and F. Pérez-Cruz, “Expectation propagation detection for high-order high-dimensional MIMO systems,” IEEE Trans. Commun., vol. 62, no. 8, pp. 2840–2849, Aug. 2014.
[14] J. Céspedes, P. M. Olmos, M. Sánchez-Fernández, and F. Pérez-Cruz, “Probabilistic MIMO symbol detection with expectation consistency approximate inference,” IEEE Trans. Veh. Technol., vol. 67, no. 4, pp. 3481–3494, Apr. 2018.
[15] J. Céspedes, “Approximate inference in massive MIMO scenarios with moment matching techniques,” Ph.D. dissertation, Departamento de Teoría de la Señal y Comunicaciones, Universidad Carlos III de Madrid, Madrid, Spain, Jan. 2017.
[16] M. Senst and G. Ascheid, “How the framework of expectation propagation yields an iterative IC-LMMSE MIMO receiver,” in Proc. IEEE Glob. Telecommun. Conf. (GLOBECOM), Kathmandu, Nepal, Dec. 2011, pp. 1–6.
[17] C. E. Rasmussen and C. K. I. Williams, Gaussian Processes for Machine Learning, Cambridge, MA, USA: MIT Press, 2006.