



HAL
open science

On convergence of approximate message passing

Francesco Caltagirone, Florent Krzakala, Lenka Zdeborová

► **To cite this version:**

Francesco Caltagirone, Florent Krzakala, Lenka Zdeborová. On convergence of approximate message passing. ISIT 2014 - 2014 IEEE International Symposium on Information Theory, IEEE, Jun 2014, Honolulu, United States. pp.1812-1816, 10.1109/ISIT.2014.6875146 . cea-01223403

HAL Id: cea-01223403

<https://cea.hal.science/cea-01223403>

Submitted on 5 Apr 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

On Convergence of Approximate Message Passing

Francesco Caltagirone, Lenka Zdeborová
 Institut de Physique Théorique
 CEA Saclay and URA 2306, CNRS
 91191 Gif-sur-Yvette, France.

Florent Krzakala
 Laboratoire de Physique Statistique, École Normale Supérieure
 and Université Pierre et Marie Curie, Rue Lhomond Paris 75005 France
 ESPCI and CNRS UMR 7083, 10 rue Vauquelin, Paris 75005 France

Abstract—Approximate message passing is an iterative algorithm for compressed sensing and related applications. A solid theory about the performance and convergence of the algorithm exists for measurement matrices having iid entries of zero mean. However, several authors have observed that for more general matrices the algorithm often encounters convergence problems. In this paper we identify the reason of the non-convergence for measurement matrices with iid entries and non-zero mean in the context of Bayes optimal inference. Finally we demonstrate numerically that when the iterative update is changed from parallel to sequential the convergence is restored.

I. INTRODUCTION

Approximate message passing (AMP) [1], [2], [3] is an algorithm derived from belief propagation that has been recently used with success in a number of sparse estimation problems, see e.g. [4], [5]. Highly non-trivial theoretical results were obtained about its performances [6], [7], [8]. Based on these developments and the promising nature of their results we can anticipate that AMP based algorithms will become the state-of-the-art algorithms for many problems of practical interest.

As with any iterative algorithm the main question about AMP, besides its performance, is its convergence. This question is largely open except for the case of compressed sensing, i.e. estimation of a sparse signal \mathbf{s} from noisy linear projections $\mathbf{y} = F\mathbf{s} + \xi$ (1)

with matrices F having iid entries of zero mean, and ξ a white Gaussian noise of variance Δ . This last case has been treated in the rigorous proofs in the very large signal size limit of [6], [9]. However, for many other sparse estimation problems, or for slightly more general matrices F , the basic version of AMP fails to converge (and worst, can diverge violently). Attempts to fix these convergence issues were so far limited to rather basic and empirical strategies such as damping the iterations in various ways, or transforming the matrix by subtracting its mean. Such strategies are rarely discussed in the literature and often appear only in the associated implementations available online. Moreover, they are far from always ensuring the convergence and some of these strategies (e.g. the mean removal) are not usable in some more challenging signal processing settings where approximate message passing can be applied (e.g. the dictionary learning problem [10]). The main motivation of this work is to understand the origin of some of these convergence problems.

The structurally simplest case where AMP fails to converge appears to be when the measurement matrix F has iid entries

of non-zero mean. This problem was noticed by several authors, e.g. [3], [11], and fixed in the implementations by removing the mean of the matrix. Indeed, the average of element of the measurement vector \mathbf{y} reads

$$\bar{y} = \frac{1}{M} \sum_{\mu} y_{\mu} = \sum_i \frac{1}{M} \sum_{\mu} F_{\mu i} s_i. \quad (2)$$

We denote $\bar{F}_i = \frac{1}{M} \sum_{\mu} F_{\mu i}$ the average value of F for column i . One can then work with the modified system $y_{\mu} \bar{y} = \sum_i (\bar{F}_{\mu i} - \bar{F}_i) s_i$ where the mean of the new sensing matrix $\bar{F}_{\mu i} - \bar{F}_i$ is zero. A similar (but different) trick is used in the implementation of [3]. This "remove mean" strategy is, however, not fully satisfactory because it is not understood why it is needed in the first place, nor under what conditions it restores the convergence. Moreover in some more general settings it is not applicable at all.

The goal of this paper is to analyze the origin of the non-convergence for non-zero mean matrices and discuss general strategies to prevent it. Such an understanding is a step towards the design of robustly convergent and hence more efficient AMP-based algorithms. We will hence consider matrices with entries generated as follows

$$F_{\mu i} = \frac{\gamma}{N} + \sqrt{\frac{1}{N}} N(0, 1). \quad (3)$$

The mean and variance need to scale as $1/N$ in order to have a $O(1)$ output given a $O(1)$ input. For $\gamma = 0$ this is the case that has been considered in the literature. Clearly these random non-zero mean matrices do not include a great variety of matrices of interest where convergence issues were observed. What is motivating our study of this case is the fact that the convergence problems can be predicted by the state evolution equations. We anticipate that the source of convergence problems is similar in nature for more general settings, validation of this is left for future work.

To be specific and simple we will consider that the signal \mathbf{s} was generated to have ρN non-zero entries that are iid normally distributed with zero mean and unit variance

$$P(s) = (1 - \rho)\delta(s) + \rho N(0, 1). \quad (4)$$

We will consider the Bayesian version of the AMP algorithm that uses this prior information about the signal. A first observation is that AMP does not depend on γ in an explicit way: this can be checked explicitly by repeating the detailed

derivations of AMP present in the literature for $\gamma > 0$ (follow e.g. the derivation in [11]).

On the other hand the asymptotic analysis of the performance of the algorithm — the state evolution [1], [6]— depends on γ explicitly and hence we have to rederive it. The analysis of the state evolution for $\gamma > 0$ will lead to an understanding of the origin of the convergence problems.

II. THE AMP ALGORITHM

We consider the AMP algorithm in the form that was derived in [2], [3], [11]. The main steps are a) going from belief propagation (BP) to a relaxed BP (r-BP) where only the two first moments of all messages are kept and b) using N sites marginals instead of $N \times M$ messages and adding the compensating Onsager terms [12]. Finally, AMP reads:

$$V_{\mu}^{t+1} = \sum_{i} F_{\mu i}^2 v_i^t, \quad (5)$$

$$\omega_{\mu}^{t+1} = \sum_{i} F_{\mu i} a_i^t - \frac{(y_{\mu} - \omega_{\mu}^t)}{F_{\mu i}} \sum_{i} F_{\mu i}^2 v_i^t, \quad (6)$$

$$(\Sigma_i^{t+1})^2 = \sum_{\mu} \frac{\Delta + V_{\mu}^{t+1}}{\sum_{\mu} F_{\mu i}^2 \frac{(y_{\mu} - \omega_{\mu}^{t+1})^2}{\Delta + V_{\mu}^{t+1}}}, \quad (7)$$

$$R_i^{t+1} = a_i^t + \frac{\sum_{\mu} F_{\mu i} \frac{(y_{\mu} - \omega_{\mu}^{t+1})}{\Delta + V_{\mu}^{t+1}}}{\sum_{\mu} \frac{F_{\mu i}^2}{\Delta + V_{\mu}^{t+1}}}, \quad (8)$$

$$a_i^{t+1} \equiv f_1 \left(\frac{(\Sigma_i^{t+1})^2}{\Sigma_i^2} \right), \quad R_i^{t+1} \equiv f_2 \left(\frac{R_i^{t+1}}{R_i} \right), \quad (9)$$

where $f_k(\Sigma^2, R)$, here and in what follows, are the k -th connected cumulants w.r.t. the probability measure

$$Q(x) = \frac{1}{Z(\Sigma^2, R)} P(x) \frac{e^{-\frac{(x-R)^2}{2\Sigma^2}}}{\sqrt{2\pi\Sigma^2}}, \quad (11)$$

where $Z(\Sigma^2, R)$ is the normalization constant.

The variables a_i and v_i are the AMP estimators for the mean and variance of the component i of the signal. The quality of the reconstruction can be evaluated by computing the mean squared error (MSE) and the average variance

$$E^t = \sum_{i=1}^N \frac{1}{N} (s_i - a_i^t)^2, \quad V^t = \sum_{i=1}^N \frac{1}{N} v_i^t. \quad (12)$$

where s_i is the original signal component. When $\gamma = 0$, the performance of the AMP algorithm was analyzed rigorously in the limit of large system size via the state evolution

in [6], [2], [3], [11]. An important property of the Bayes optimal inference (i.e. when the signal was indeed generated from the assumed prior distribution) is that the two parameters are equal in the large size limit, $E^t = V^t$, and the state evolution hence reduces to an iterative equation of a single real number, which is amenable to rigorous analysis [6]. In statistical physics $E^t = V^t$ is called the Nishimori condition and is discussed in the context of compressed sensing in detail

in [11]. In general, when $\gamma = 0$ we observed by analyzing the state evolution equations that even when at initial times $E^{t=0} = V^{t=0}$ the equality $E^t = V^t$ is restored after a sufficient number of iterations.

III. STATE EVOLUTION WITH NON-ZERO MEAN MATRICES

The state evolution of the AMP algorithm can be derived for measurement matrices with non-zero mean $\gamma > 0$. Here we follow closely the derivation and notation from [11] for zero mean matrices. Among the different variables, the statistical distribution of R_i plays a crucial role in the determination of the state evolution. It can be written as

$$R_i^t = s_i + \frac{1}{\alpha} r_i^t, \quad (13)$$

where $\alpha = M/N$ is the aspect ratio of the matrix and

$$r_i^t = \sum_{\mu} F_{\mu i} \xi_{\mu} + \sum_{\mu} F_{\mu i} \sum_{j=i} F_{\mu j} (s_j - a_{j-\mu}^t) \quad (14)$$

is a Gaussian random variable, and $a_{j-\mu}^t$ is an auxiliary variable related closely to a^t that appears in the derivation of the AMP algorithm. Assumptions used to derive AMP can be used to compute the mean and variance of r^t over realizations

$$\bar{r}^t = \alpha \gamma^2 D^t, \quad (15)$$

$$\text{var}(r^t) = \alpha (E^t + \Delta + \gamma^2 (D^t)^2), \quad (16)$$

where we have defined a new order parameter

$$D = \frac{1}{N} \sum_i (s_i - a_i)^2. \quad (17)$$

The parameter D^t is not needed for zero mean matrices $\gamma = 0$. For $\gamma > 0$, however, the state evolution is written in terms of three parameters E^t , V^t and D^t . The remaining steps in the derivation are basically identical to those for zero mean matrices and following [11] we obtain

$$V^{t+1} = \int ds P(s) \int Dz \times \quad (18)$$

$$f_2 \left(\frac{\Delta + V^t}{\alpha}, s + z A(E^t, D^t) + \gamma^2 D^t \right), \quad (19)$$

$$E^{t+1} = \int ds P(s) \int Dz \times \quad (20)$$

$$f_1 \left(\frac{\Delta + V^t}{\alpha}, s + z A(E^t, D^t) + \gamma^2 D^t \right),$$

where Dz is a Gaussian measure and

$$A(E^t, D^t) = \frac{\sqrt{E^t + \Delta + \gamma^2 (D^t)^2}}{\alpha}. \quad (21)$$

When the mean of the measurement matrix is zero, $\gamma = 0$, these equations clearly reduce to those derived in [3], [11].

Also for $\gamma > 0$ we can identify the Nishimori condition, which reads $E^t = V^t$ (for the same reasons as for the previous case) and $D^t = 0$ (since under Bayes optimal inference the mean of the estimator must be equal to the true mean of the signal). It is a question of simple algebraic verification to see that starting with $E^t = V^t$ and $D^t = 0$ eqs. (18-20) lead to $E^{t+1} = V^{t+1}$ and $D^{t+1} = 0$. Hence if we restrict ourselves to the space on which the Nishimori conditions hold (called the Nishimori line) there is no difference between the $\gamma = 0$ and $\gamma > 0$ case.

IV. INSTABILITY OF THE NISHIMORI LINE

In this Section we analyze the dynamical stability of the Nishimori line (NL) under iterations of eqs. (18-20). We consider the space (K, D) orthogonal to the NL, where $K = V - E$. We know that in this space $(K^*, D^* = 0)$ is a fixed point. We can generically write $(K^*, D^*) = \bar{f}(V^t, K^t, D^t)$,

$D^{t+1} = f_D(V^t, K^t, D^t)$. To analyze the stability we linearize around the fixed point considering the perturbations $\delta K^t = K^t - K^*$ and $\delta D^t = D^t - D^*$. The linearized formula

reads

$$\begin{pmatrix} \delta K^{t+1} \\ \delta D^{t+1} \end{pmatrix} = M \cdot \begin{pmatrix} \delta K^t \\ \delta D^t \end{pmatrix} \quad (22)$$

with

$$M = \begin{pmatrix} \partial_K f_K(V^t, K^*, D^*) & \partial_D f_K(V^t, K^*, D^*) \\ \partial_K f_D(V^t, K^*, D^*) & \partial_D f_D(V^t, K^*, D^*) \end{pmatrix}. \quad (23)$$

It follows from a straightforward algebraic verification that both the off-diagonal terms (the cross derivatives) are zero for the distribution $P(s)$ from eq. (4). The matrix M (23) is hence diagonal. For a more generic prior distribution the situation is slightly more involved, but qualitatively analogous to the one of (4). The diagonal terms read

$$\begin{aligned} \partial_D f_D(V^t) &= -\frac{\alpha \gamma^2}{\Delta + V^t} \int ds P(s) \int Dz f_2(A^2, s + zA) \\ &= -\frac{\alpha \gamma^2 V^{t+1}}{\Delta + V^t}, \end{aligned} \quad (24)$$

$$\begin{aligned} \partial_K f_K(V^t) &= \int \frac{1}{2} \frac{1}{\Delta + V^t} \int ds P(s) \left\{ z f_4(A^2, s + zA) \right. \\ &\quad \left. + 2(f_2(A^2, s + zA) - s f_3(A^2, s + zA))^2 \right\}, \end{aligned} \quad (25)$$

where, as before, the functions $f_k(\Sigma^2, R)$ are the k -th connected cumulants with respect to the measure $Q(\Sigma^2, R)$ (11), and where we denoted

$$A \equiv \frac{\Delta + V^t}{\alpha}. \quad (26)$$

The term $\partial_K f_K(V^t)$ is independent of γ and its modulus is always smaller than one. Hence the Nishimori line is stable in the direction $K = V - E$.

On the other hand the term $\lambda_D = \partial_D f_D(V^t)$ has a non-trivial behavior that we illustrate in Fig. 1 for $\rho = 0.1$, $\alpha = 0.3$, $\Delta = 10^{-10}$ and, respectively, $\gamma = 1.9$, $\gamma = 2.5$, $\gamma = 2.9$ and $\gamma = 3.6$. In the figure we identify three different regimes:

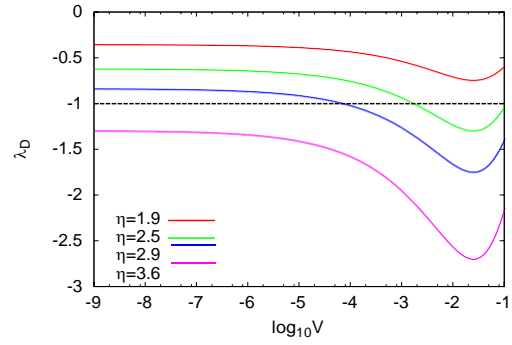


Fig. 1. The term $\lambda_D = \partial_D f_D(V^t)$ associated to the stability of the Nishimori line in the D -direction as a function of the MSE for $\rho = 0.1$, $\Delta = 10^{-10}$ and $\alpha = 0.3$. Three different regimes can be identified, one in which λ_D is always less than 1, the second in which λ_D is larger than 1 in a region, and the third in which λ_D is larger than one along the whole Nishimori line down to the fixed point. The critical values (as defined in the text) for this case are $\gamma_c^{(1)} = 2.197$, $\gamma_c^{(2)} = 3.162$.

- For $|\gamma| < \gamma_c^{(1)}$ the eigenvalue λ_D is always less than 1 in modulus.
- For $\gamma_c^{(1)} < |\gamma| < \gamma_c^{(2)}$ the eigenvalue becomes greater than 1 in modulus in a certain portion of the Nishimori line. In this region the evolution tends to make D larger, while at the same time V and E decrease.
- For $|\gamma| > \gamma_c^{(2)}$ the eigenvalue λ_D is larger than 1 in modulus in the whole range down to the fixed point. This tells us that any fluctuation of D will be progressively enhanced.

We further realize that the expression used to calculate λ_D depends on the value V only through the variable A (26) and not in an explicit way on the parameters α and Δ . This means that the threshold value $\gamma_c^{(1)}$ is from its definition independent of α and Δ . The threshold value $\gamma_c^{(2)}$ is also independent of α for $\Delta = 0$ and only weakly dependent on both α and Δ for small values of Δ . In Fig. 2 we hence plot the two threshold values for $\Delta = 0$ (in which case they are both independent of the undersampling α) as a function of the sparsity ρ .

V. COMPARING STATE EVOLUTION TO AMP

We now discuss how does the instability of the Nishimori line translate into the behavior of the state evolution (SE) initialized usually as $E^{t=0} = V^{t=0} = \rho$ (corresponding to $a_{t=0} = 0$ and $v_{t=0} = \rho$) and $D^{t=0} = 0$. The SE was derived to correspond to the behavior of the AMP algorithm for sufficiently large system sizes N . We observe that

- For $|\gamma| < \gamma_c^{(1)}$ the SE converges to the fixed point with monotonically decreasing $E = V$. There are really infinitesimal fluctuations in D that are due to numerical precision but they are harmless.
- For $\gamma_c^{(1)} < |\gamma| < \gamma_c^{(2)}$ the SE converges to the fixed point with monotonically decreasing $E = V$. In the region of V in which λ_D is larger than one, we observe that the numerical fluctuations of D are slightly increased (especially if we are close to $\gamma_c^{(2)}$), without changing qualitatively the behavior of V , and when $|\lambda_D|$ becomes again smaller than 1 the fluctuations are reabsorbed.

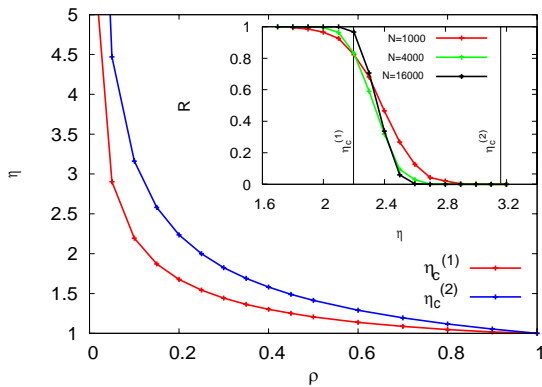


Fig. 2. [Main frame] The threshold values for the mean of the measurement matrix above which the state evolution on the Nishimori line (i.e. $E = V$ and $D = 0$) is not stable. Above $\gamma_c^{(1)}$ only part of the line is unstable, above $\gamma_c^{(2)}$ the full line is unstable. For zero measurement noise these values do not depend on the undersampling rate α . For weak measurement noise only the line $\gamma_c^{(2)}$ depends weakly on both Δ and α . [Inset] The convergence rate R of the AMP algorithm as a function of the mean of the measurement matrix γ with $\rho = 0.1$, $\Delta = 10^{-10}$ and $\alpha = 0.3$ for different values of the signal size N . We can see that the transition is close to the first critical value $\gamma_c^{(1)}$ (marked by the vertical line on the left) and it is smoother for low N and sharper for larger N . For very large N we also expect the transition to move towards the vertical line on the right ($\gamma = \gamma_c^{(2)}$), but this effect is not visible at the N we are able to reach.

- For $\gamma > \gamma_c^{(2)}$ the fluctuations of D are increased along the whole line $E = V$. At some point these fluctuations reach so large values that the difference $K = E - V$ grows and we observe a divergence of both E and V .

Therefore, while with infinite numerical precision the SE should stay on the Nishimori line and converge whatever the value of γ is, from the practical point of view the fluctuations due to numerical precision are sufficient to cause divergence in the third regime. Of course in the AMP algorithm the typical fluctuations are of order $1/\sqrt{N}$ hence relatively large and that is the reason why for $\gamma > \gamma_c^{(2)}$ AMP never converges. In fact these finite size fluctuations are so strong that even in the second regime $\gamma_c^{(1)} < \gamma < \gamma_c^{(2)}$ AMP might have problems. Therefore we observe a smooth transition in the success rate $R = (\text{\#successes}/\text{\#failures})$ between $\gamma_c^{(1)}$ and $\gamma_c^{(2)}$ for finite

N . When N is increased this smooth transition becomes sharper. In the inset of Fig. 2 we show the success rate of the AMP algorithm averaged over 1000 random instances of the measurement matrix for $N = 1000, 4000$ and over 500 instances for $N = 16000$. We see that, even if asymptotically, the reference value for the success/failure transition would be $\gamma_c^{(2)}$, for all practical system sizes, the right threshold to look

at is rather $\gamma_c^{(1)}$. REDUCING THE INSTABILITY

There are at least two strategies that appear in the implementations of the AMP algorithm that improve its convergence. Let us discuss them now in the context of the above analysis.

a) *Damping*: A popular and generic strategy to improve convergence of iterative algorithms is “damping”, i.e. in every new iteration we update the variables only partially. Such damping (with different schemes) appears in basically every

available implementation of AMP. In the view of the preceding analysis a dynamical instability is mitigated by such damping and the eigenvalue $|\lambda_D|$ becomes effectively smaller. Indeed AMP with damping converges well even for matrices with means slightly larger than those corresponding to $\gamma_c^{(2)}$ in Fig. 2.

b) *Expectation maximization learning*: In this paper so far we assumed the prior knowledge of the probability distribution of signal elements as well as of the measurement noise Δ and the sparsity ρ . A classical strategy of expectation maximization was suggested, tested and implemented in [7], [13] in order to learn these parameters when they are not known a priori. A careful investigation of the AMP algorithm with EM learning leads to a conclusion that with the learning the AMP has better convergence properties than without.

This can come as a surprise at first, but in the view of our above investigation it can now be easily explained. The EM update in a sense imposes (in an iterative way) the Nishimori condition, see the derivation of EM in [7], hence it should be expected that it also stabilizes the Nishimori line and consequently improves the convergence of AMP.

VII. THE SEQUENTIAL REDEMPTION

AMP being so sensitive to the mean of the matrix elements is surprising because the standard BP, when applied to discrete random problems, does not experience such problems. In this last section we argue that the convergence problems in the case of CS with non-zero mean measurement matrices are actually specific to the “parallel updates” (involving only matrix multiplications) performed in the AMP algorithm that we presented in Sec. II. Let us recall the so-called relaxed-BP (r-BP) algorithm [14] (for present notations see [11]) where messages are sent on the factor graph:

$$A_{\mu \rightarrow i} = \frac{F^2}{\Delta + \sum_{j=\mu} F_{\mu j}^2 v_{j \rightarrow \mu}}, \quad (27)$$

$$B_{\mu \rightarrow i} = \frac{F_{\mu i} (y_{\mu} - \sum_{j=i} F_{\mu j} a_{j \rightarrow \mu})}{\Delta + \sum_{j=i} F_{\mu j}^2 v_{j \rightarrow \mu}}, \quad (28)$$

$$a_{i \rightarrow \mu} = f_1 \frac{1}{v_{\mu} A_{\mu \rightarrow i}}, \quad \sum_{\nu=\mu} \frac{B_{\nu \rightarrow i}}{A_{\nu \rightarrow i}}, \quad (29)$$

$$v_{i \rightarrow \mu} = f_2 \frac{1}{v_{\mu} A_{\mu \rightarrow i}}, \quad \sum_{\nu=\mu} \frac{B_{\nu \rightarrow i}}{A_{\nu \rightarrow i}}, \quad (30)$$

$$a_{i \rightarrow \mu} = \frac{1}{v_{\mu} A_{\mu \rightarrow i}}, \quad \frac{B_{\nu \rightarrow i}}{v_{\nu} A_{\nu \rightarrow i}} = f, \quad (31)$$

$$v_{i \rightarrow \mu} = \frac{1}{v_{\mu} A_{\mu \rightarrow i}}, \quad \frac{B_{\nu \rightarrow i}}{v_{\nu} A_{\nu \rightarrow i}} = f. \quad (32)$$

We intentionally wrote this algorithm without the time indices, because the update can be performed in two ways. First, in the parallel update all variables are updated at time t given the state at time $t-1$. The state evolution derived above for AMP also applies to the r-BP updated in parallel. Actually for matrices for which fast multiplication algorithms are not

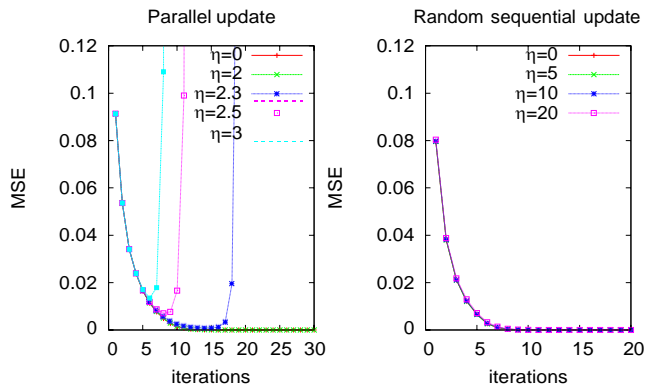


Fig. 3. A comparison of the effect of a non-zero mean γ on the relaxed belief propagation algorithm (27-32) with different update scheme using $\alpha = 0.3$, $\rho = 0.1$, $\Delta = 10^{-10}$ and $N = 10^4$. Left: Parallel update corresponding to (and equivalent to) AMP. For $\gamma > 2.3$, even the damped AMP diverges. Right: Random sequential update. In that case, r-BP converges very fast even for large values of the mean γ .

available there is no crucial advantage of AMP over r-BP, in the limit of large N they are equivalent.

The second update is the random sequential one where one picks a *single* index i and updates all messages corresponding to it. For r-BP, this leads to the same computational complexity, however, it is important to realize that AMP is actually written assuming the r-BP with the parallel update. In Fig. 3 we compare the behavior of parallel and random sequential r-BP: as we see, the sequential update does not seem to be affected by the non-zero mean.

This observation of the parallel update being more problematic than the sequential one is actually not surprising a posteriori. In fact, such a lack of convergence is well known to occur in parallel iterations in many problems that would be classified in statistical physics as antiferromagnets. Antiferromagnetic instabilities, just like the one we have studied here, occur for instance in the hard-core model [15] or coloring [16] problems on random graphs. The insight and evidence collected in studies of those problems also justifies our claim that the use of the sequential update fixes the convergence problems in a wide range of cases. Nevertheless, it is not a *universal* solution since it by no means guarantees convergence for all matrices. Also, the disadvantage of the sequential r-BP update is that it loses the nice property of only involving matrix multiplication that is crucial for scalability for operators such as the fast Fourier transform, for which there exist efficient multiplication methods.

VIII. CONCLUSIONS

We have analyzed the convergence problems of AMP in the specific case of compressed sensing with measurement matrices having iid entries of non-zero mean. Despite the fact that the AMP iterations are not modified w.r.t. the case of zero mean, the state evolution does contain an additional order parameter. The main result of the paper, contained in Sec. IV, is that the presence of this third parameter causes instabilities of the so-called Nishimori line and, therefore in the algorithm itself, if the mean of the matrix elements exceeds some critical

value. In the last section we show that the convergence issue for matrices of non-zero mean are strongly mitigated when random sequential update is used in the message passing instead of the parallel one that is standard to AMP.

This analysis represents a step towards understanding of the nature of convergence issues in message passing algorithms that are ubiquitous in problems ranging from physics to information theory. More complete understanding of these issues is needed before message passing algorithms can become part of standard toolbox to solve a wide range of problems of practical interest.

ACKNOWLEDGMENT

This work has been supported by the ERC under the European Unions 7th Framework Programme Grant Agreement 307087-SPARCS, and by the project TASC of the Labex PALM.

REFERENCES

- [1] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proc. Natl. Acad. Sci.*, vol. 106, no. 45, pp. 18914–18919, 2009.
- [2] D. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing: I. motivation and construction," in *IEEE Information Theory Workshop (ITW)*, 2010, pp. 1–5.
- [3] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," in *IEEE International Symposium on Information Theory Proceedings (ISIT)*, 2011, pp. 2168–2172.
- [4] P. Schniter, "Turbo reconstruction of structured sparse signals," in *Information Sciences and Systems (CISS), 2010 44th Annual Conference on*. IEEE, 2010, pp. 1–6.
- [5] S. Som and P. Schniter, "Compressive imaging using approximate message passing and a markov-tree prior," *Signal Processing, IEEE Transactions on*, vol. 60, no. 7, pp. 3439–3448, 2012.
- [6] M. Bayati and A. Montanari, "The dynamics of message passing on dense graphs, with applications to compressed sensing," *IEEE Transactions on Information Theory*, vol. 57, no. 2, pp. 764–785, 2011.
- [7] F. Krzakala, M. Mézard, F. Sausset, Y. Sun, and L. Zdeborová, "Statistical physics-based reconstruction in compressed sensing," *Phys. Rev. X*, vol. 2, p. 021005, 2012.
- [8] D. L. Donoho, A. Javanmard, and A. Montanari, "Information-theoretically optimal compressed sensing via spatial coupling and approximate message passing," in *Proc. of the IEEE Int. Symposium on Information Theory (ISIT)*, 2012.
- [9] M. Bayati, M. Lelarge, and A. Montanari, "Universality in polytope phase transitions and message passing algorithms," *arXiv preprint arXiv:1207.7321*, 2012.
- [10] F. Krzakala, M. Mezard, and L. Zdeborova, "Phase diagram and approximate message passing for blind calibration and dictionary learning," pp. 659–663, July 2013.
- [11] F. Krzakala, M. Mézard, F. Sausset, Y. Sun, and L. Zdeborová, "Probabilistic reconstruction in compressed sensing: Algorithms, phase diagrams, and threshold achieving matrices," *J. Stat. Mech.*, p. P08009, 2012.
- [12] D. Thouless, P. Anderson, and R. Palmer, "Solution of solvable model of a spin glass," *Philosophical Magazine*, vol. 35, no. 3, pp. 593–601, 1977.
- [13] J. P. Vila and P. Schniter, "Expectation-maximization bernoulli-gaussian approximate message passing," in *Proc. Asilomar Conf. on Signals, Systems, and Computers (Pacific Grove, CA)*, 2011.
- [14] S. Rangan, "Estimation with random linear mixing, belief propagation and compressed sensing," in *Information Sciences and Systems (CISS), 2010 44th Annual Conference on*. IEEE, 2010, pp. 1–6.
- [15] O. Rivoire, G. Biroli, O. C. Martin, and M. Mézard, "Glass models on bethe lattices," *The European Physical Journal B-Condensed Matter and Complex Systems*, vol. 37, no. 1, pp. 55–78, 2004.
- [16] L. Zdeborová and F. Krzakala, "Phase transitions in the coloring of random graphs," *Physical Review E*, vol. 76, no. 3, p. 031131, 2007.