COMPRESSED SENSING UNDER MATRIX UNCERTAINTY:
OPTIMUM THRESHOLDS AND ROBUST APPROXIMATE MESSAGE PASSING

Florent Krzakala[1*], *Marc Mézard*[2] *and Lenka Zdeborová*[3]

[1] CNRS and ESPCI ParisTech 10 rue Vauquelin, UMR 7083 Gulliver Paris 75005 France
[2] Ecole Normale Supérieure, 45 rue d'Ulm, Paris, France, and LPTMS-CNRS, Univ. Paris Sud
[3] IPhT, CEA Saclay, and URA 2306 CNRS, 91191 Gif-sur-Yvette, France

ABSTRACT

In compressed sensing one measures sparse signals directly in a compressed form via a linear transform and then reconstructs the original signal. However, it is often the case that the linear transform itself is known only approximately, a situation called matrix uncertainty, and that the measurement process is noisy. Here we present two contributions to this problem: first, we use the replica method to determine the mean-squared error of the Bayes-optimal reconstruction of sparse signals under matrix uncertainty. Second, we consider a robust variant of the approximate message passing algorithm and demonstrate numerically that in the limit of large systems, this algorithm matches the optimal performance in a large region of parameters.

Index Terms— Compressed sensing, Measurement uncertainty, Belief propagation, Message passing, Performance analysis.

1. INTRODUCTION

Compressed sensing (CS) is designed to measure sparse signals directly in a compressed form by acquiring a small (with respect to the dimension of the signal) number of random linear projections of the signal, and subsequently reconstructing computationally the signal. In particular, it has been shown [1, 2] that this reconstruction is possible and computationally feasible in many cases using ℓ_1-minimization based algorithms. It is often the case in practical applications that the linear transform itself is known only approximately; for instance this is the case if calibration has not been done perfectly. Several works have analyzed the performance of ℓ_1-regularization based algorithm under matrix uncertainty, e.g. [3, 4] and references therein. In this paper we shall use instead a Bayesian approach. In this framework CS under matrix uncertainty was studied recently by [5]. In particular the authors of [5] propose a generalization of the approximate message

passing algorithm (AMP) [6, 7, 8] that treats matrix uncertainty (MU-AMP) and outperforms other methods. Empirically, they show that the MU-AMP algorithm performs near oracle bounds.

Our goal in this paper is twofold. First we compute the best theoretically possible performance of reconstruction algorithms under matrix uncertainty using the replica method [9, 10, 11, 12, 8, 13], and show that in a large region of parameters this performance is indeed matched by MU-AMP. And second, we consider a slightly different algorithm that we refer to as robust-AMP algorithm (using a minimal change first suggested in [6]) and show that it is asymptotically equivalent to MU-AMP and thus also reaches Bayes optimal MSE while not assuming the knowledge of the measurement noise, nor the level of matrix uncertainty.

2. DEFINITIONS

In our analysis we consider a sparse N-dimensional signal **x** having on average K non-zero iid components chosen from a distribution $\phi(x)$. Defining the density $\rho = K/N$ we have

$$P(\mathbf{x}) = \prod_{i=1}^{N} P(x_i) = \prod_{i=1}^{N} [\rho\phi(x_i) + (1-\rho)\delta(x_i)]. \quad (1)$$

Note that our analysis applies as well to non-iid signals with empirical distribution of components converging to $\phi(x)$, as discussed e.g. in [14]. For concreteness, here we shall use a Gaussian $\phi(x)$ of zero mean and unit variance. Although our results depend quantitatively on the form of $\phi(x)$, the overall qualitative picture is robust with respect to this choice. We further assume that the parameters of $P(\mathbf{x})$ are known (if this is not the case they can be learned via expectation-maximization [8, 13, 15]). One could use an approximately sparse signal as in [16] as well. For a discussion about the robustness when the apriori distribution does not match the true distribution of signal components see [8, 13].

The M measurements y_μ are obtained via noisy linear

random projections of the signal

$$y_\mu = \sum_{i=1}^N F_{\mu i}^0 s_i + \xi_\mu\,, \qquad \mu = 1,\dots,M\,. \qquad (2)$$

where we model the noise ξ_μ as a Gaussian random variable with mean 0 and variance Δ, and where $\mathbf{F^0}$ has iid random components of mean 0 and variance $1/N$. The parameter $\alpha = M/N$ is the measurement, or sampling, rate.

Additionally, we consider that the matrix $\mathbf{F^0}$ is not known perfectly, and that we know only its noisy version:

$$F_{\mu i}' = \frac{F_{\mu i}^0 + X_{\mu i}\sqrt{\eta}}{\sqrt{1+\eta}}\,, \qquad (3)$$

where $X_{\mu i}$ is a white noise with mean 0 and variance $1/N$, and where η measures the uncertainty on the matrix. When $\eta \to 0$, the measurement matrix is perfectly known (this is the usual compressed sensing situation) while for $\eta \to \infty$ nothing is known about the measurement matrix. Assuming a Gaussian distribution for the matrix elements and the noise $\mathbf{X}$, we obtain the posterior probability of one matrix element $P(F_{\mu i}^0|F_{\mu i}')$ to be a Gaussian with mean $F_{\mu i} = \frac{F_{\mu i}'}{\sqrt{1+\eta}}$, and variance $S_{\mu i} = \frac{\eta}{N(1+\eta)}$.

Our goal is to reconstruct the signal $\mathbf{s}$ based on the knowledge of $\mathbf{y}$ and $\mathbf{F}'$. The Bayes-optimal estimate of a signal $\mathbf{x}^\star$, that minimizes the mean-squared error (MSE) MSE $= \sum_{i=1}^N (s_i - x_i^\star)^2/N$ with the original signal $\mathbf{s}$, is given by $x_i^\star = \int \mathrm{d}x_i\, x_i\, \nu_i(x_i)$. where $\nu_i(x_i)$ is the marginal distribution of x_i with respect to:

$$P(\mathbf{x}|\mathbf{y},\mathbf{F}') = \frac{P(\mathbf{x})}{Z} \int \mathrm{d}\mathbf{F^0} P(\mathbf{F^0}|\mathbf{F}') e^{-\frac{1}{2\Delta}(\mathbf{y}-\mathbf{F^0}\mathbf{x})^2}\,. \qquad (4)$$

3. REPLICA METHOD: OPTIMUM MSE

The MSE obtained from the Bayes-optimal approach can be computed in the limit of large N via the replica method. Although this method is in general non-rigorous, it is expected that its results on the performance are tight when the apriori distribution matches the true distribution of signal elements, see [9, 10, 11, 12, 8, 13], in some cases the tightness can be proven rigorously, see e.g. [17]. We shall leave out a detailed derivation and refer instead to [8, 13] for a very similar computation. The result is that in the large signal limit, the MSE is given by

$$\mathrm{MSE}(\alpha,\rho,\Delta,\eta) = \operatorname*{argmax}_E \Phi(E)\,, \qquad (5)$$

where the "potential" $\Phi(E)$ is given by

$$\Phi(E) = -\frac{\alpha}{2}\left[\log\left[\Delta+E+(\rho-E)D\right] + \frac{\Delta+\rho}{[\Delta+E+(\rho-E)D]}\right]$$
$$+ \ (1-\rho)\int \mathcal{D}z\, \log\left[1-\rho+\frac{\rho}{\sqrt{m+1}}e^{\frac{z^2 m}{2(m+1)}}\right]$$
$$+ \ \rho\int \mathcal{D}z\, \log\left[1-\rho+\frac{\rho}{\sqrt{m+1}}e^{\frac{z^2 m}{2}}\right]\,, \qquad (6)$$
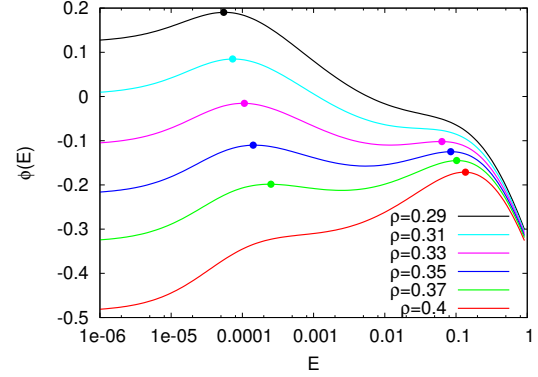

Fig. 1. The potential $\Phi(E)$, eq. (6), for measurement rate $\alpha = 0.5$, matrix uncertainty $\eta = 10^{-4}$, and noise $\Delta = 10^{-10}$, for different values of density ρ. The global maximum $\Phi(E^*)$ determines the values of the Bayes-optimal MSE $= E^*$. As the density varies the value $E^*(\rho)$ has a discontinuity which is usually referred to as a "first-order" or "discontinuous" phase transition. AMP algorithms are performing a steepest ascent of the potential and get trapped at the maximum with the largest value of E, instead of reaching the global one (see e.g. the case $\rho = 0.33$). The appearance of this trapping local maxima is referred to as the "spinodal" transition.
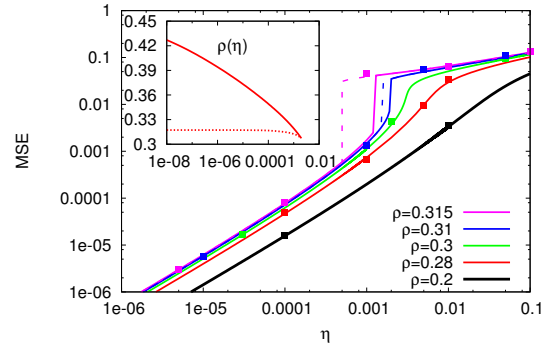

Fig. 2. Bayes-optimal MSE for $\alpha = 0.5$, $\Delta = 10^{-10}$ as a function of matrix uncertainty η for different densities ρ. The MSE of the trapping local maximum is shown in dashed line. The results of the AMP algorithm on some instances are also shown (points, done with $N = 20000$) and agree perfectly with the prediction (lines). In the inset, we show the location of the two phase transitions in the ρ, η plane: the upper line determines the "first-order" threshold beyond which the optimal-MSE suddenly degrades, while the lower line is the "spinodal" transition that marks the degradation of the performance of AMP. Note that the spinodal is almost independent of the noise η in a large range of values.

where the D and m are given as

$$D = \frac{\eta}{1+\eta}\,, \qquad m = \frac{\alpha(1-D)}{\Delta+E+(\rho-E)D}\,. \qquad (7)$$

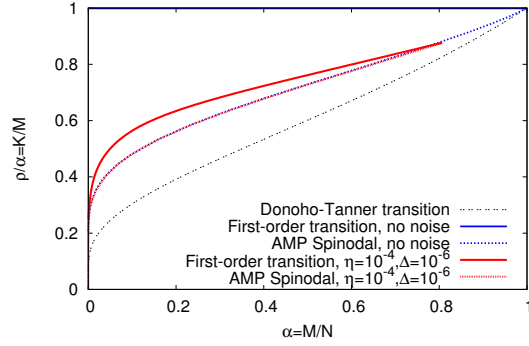The potential for the usual compressed sensing with known measurement matrix [13] is recovered taking $\eta = 0$.

Fig. 3. The phase diagram "a la" Donoho-Tanner [18] in the limit of large signals, $N \to \infty$. The horizontal axis is the sampling ratio M/N and the vertical one measures the gain K/M provided by compressed sensing. The dotted black line is the transition for the ℓ_1 reconstruction [18] in zero noise. The blue lines (from [8, 13]) show the location of the spinodal (dotted blue) and first-order (full blue) phase transitions for the noiseless Bayesian approach. While a perfect reconstruction (MSE=0) can be obtained in principle until $\rho = \alpha$, the AMP algorithm allows such perfect reconstruction only up to the spinodal line (dotted blue). The noisy counterpart of these two transitions is shown in red, using $\Delta = 10^{-4}$ and $\eta = 10^{-6}$. While the region where a good reconstruction is possible (below the full red line) has shrunk a lot, the spinodal line (dotted red line) that marks the onset of good performance of AMP is left almost unchanged below $\rho \approx 0.8$ demonstrating the robustness of the AMP to noises.

The potential (6) is shown in Fig. 1. Maximizing the potential, one obtains the Bayes-optimal MSE which we show in Fig. 2 as a function of η for different densities ρ. We see a phenomenology similar to that described in [13, 16] for other types of noise. Because of a two-maxima shape of the function $\Phi(E)$, the Bayes-optimal MSE displays a sharp transition separating a region of parameters with a small MSE, comparable to $\max(\Delta, \eta)$, from a region with a large $O(1)$ value of the MSE. In this second region of parameters compressed sensing technics –regardless of the reconstruction method– will not be useful. This is a "first order" phase transition.

Another important transition (the "spinodal") that can be studied from the form of the potential function $\Phi(E)$ is the appearance of a high-MSE local maxima. Since the AMP algorithm performs a steepest ascent in the $\Phi(E)$ function (see e.g. [13]) this transition separates the region of parameters in which the performance of AMP will be asymptotically optimal from the one where AMP is suboptimal. An important feature of this transition, shown in Fig. 3, is that its position is only very weakly dependent on the value of the matrix uncertainty. This means that the performance of AMP is extremely robust to matrix uncertainty η. The same result was also obtained for additive measurement noise in [13] and for approximate sparsity in [16].

4. ROBUST MESSAGE-PASSING ALGORITHM

The Bayesian approach to compressed sensing combined with belief propagation based reconstruction algorithm leads to the so-called approximate message passing (AMP) algorithm as first derived in [6] for the minimization of ℓ_1, and subsequently generalized in [19, 7, 13]. This approach was adapted by Parker, Cevher and Schniter to treat the matrix uncertainty MU-AMP in [5]. Here we shall consider a version of the canonical AMP, that we call robust-AMP, that turns out to be robust to noisy measurement and matrix uncertainty, so that it can be used indifferently of the presence or absence of noise and matrix uncertainty. We show in the next section that robust-AMP and MU-AMP are equivalent in the limit of infinite systems.

For every measurement component μ we define one real number ω_μ, for each signal component i we define four real numbers Σ_i, R_i, a_i, v_i. These quantities are updated as follows (for the derivation with these notations see [13]):

$$V^{t+1} = \frac{1}{M} \sum_\mu (y_\mu - \omega_\mu^t)^2 , \quad (8)$$

$$\omega_\mu^{t+1} = \sum_i F_{\mu i} a_i^t - \frac{(y_\mu - \omega_\mu^t)}{V^t} \sum_i F_{\mu i}^2 v_i^t , \quad (9)$$

$$(\Sigma_i^{t+1})^2 = \frac{V^{t+1}}{\sum_\mu F_{\mu i}^2} , \quad (10)$$

$$R_i^{t+1} = a_i^t + \frac{\sum_\mu F_{\mu i}(y_\mu - \omega_\mu^{t+1})}{\sum_\mu F_{\mu i}^2} , \quad (11)$$

$$a_i^{t+1} = f_a\left((\Sigma_i^{t+1})^2, R_i^{t+1}\right) , \quad (12)$$

$$v_i^{t+1} = f_c\left((\Sigma_i^{t+1})^2, R_i^{t+1}\right) . \quad (13)$$

Where only the functions f_a and f_c depend explicitly on the model $P(x)$:

$$f_a(\Sigma^2, R) \equiv \int dx\, x\, \mathcal{M}(\Sigma^2, R, x) , \quad (14)$$

$$f_c(\Sigma^2, R) \equiv \int dx\, x^2\, \mathcal{M}(\Sigma^2, R, x) - f_a^2(\Sigma^2, R) , \quad (15)$$

where $\mathcal{M}(\Sigma^2, R, x)$ is the following probability distribution

$$\mathcal{M}(\Sigma^2, R, x) \equiv \frac{1}{\hat{Z}(\Sigma^2, R)} P(x) \frac{1}{\sqrt{2\pi}\Sigma} e^{-\frac{(x-R)^2}{2\Sigma^2}} . \quad (16)$$

The explicit expression for f_a and f_c for the Gauss-Bernoulli signal is given in [13] while the case of approximate sparsity was considered in [16]. The above equations are initialized with $a_i^{t=0} = 0$, $v_i^{t=0} = \rho$, $\omega_\mu^{t=0} = y_\mu$, then the equations are iterated till convergence.

The difference between the present algorithm with respect to the more common version of AMP is in the way the estimate of the current error on the measurement element is computed in eq. (8). Most previous works (e.g. [7, 13]) used a

μ-dependent vector $V_\mu^{t+1} = \Delta + \sum_i F_{\mu i}^2 v_i^t$ in the case of noisy measurement with a perfectly known matrix, whereas the original paper [6] used a value precomputed by the state evolution. The modification done by the authors of ref. [5] to incorporate the effect of matrix uncertainty is (in our notation) $V_\mu^{t+1} = \Delta + \sum_i F_{\mu i}^2 v_i^t + \sum_i (b_i + a_i^2)\eta/[N(1+\eta)]$. As shown in ref. [5] this leads to a very efficient algorithm with matrix uncertainty. The expression we use instead in eq. (8) was also proposed in [6]. Perhaps surprisingly, it is equivalent to the one of ref. [5] when the system size $N \to \infty$ even in the case of matrix uncertainty. The advantage of expression (8) is that we do not need to know the value of η and Δ^1: the algorithm automatically incorporates the errors coming from the uncertainty on the matrix and the measurement; and hence we are using the same code regardless the presence or absence of noise and/or matrix uncertainty. We thus refer to algorithm (8)-(13) as the "robust"-AMP.

5. DENSITY EVOLUTION

The AMP approach is amenable to asymptotic ($N \to \infty$) analysis in the case of i.i.d. random measurement matrices using a method known as the "cavity method" (in statistical physics) [9], the "density evolution" in coding [20], and the "state evolution" in the context of CS [6, 21]. We shall not detail the computation here, it goes in the same lines as in [13]. Given the parameters $\rho, \alpha, \eta, \Delta$, the MSE follows:

$$E^{t+1} = \int \mathrm{d}x\, P(x) \int \mathcal{D}z f_c(1/m^t, x + z/\sqrt{m^t})\,, \quad (17)$$

where m^t follows eq. (6), with E^t on the right hand side, $E^{t=0} = \rho$, $\mathcal{D}z$ is a Gaussian integral, and f_c is defined by eq. (15). The comparison between the evolution of the MSE in the algorithm and eq. (17) is shown on Fig. 4; the agreement is very good. Note that we have also applied the density evolution to the AMP algorithm with matrix uncertainty of [5], and found that it obeys asymptotically the same equation (17). More specifically, the term V^t from eq. (8) evolves as $\Delta + E^t + (\rho - E^t)D$ in the limit $N \to \infty$ both for the expression in ref. [5] and in eq. (8). From eq. (17), one can also derive that the evolution is equivalent to a steepest ascent of the potential $\phi(E)$ obtained from the replica method in eq. (6). This underlines the importance of the spinodal transition illustrated on Figs. 1, 2, and 3. In particular we see that the region where the AMP converges to the Bayes-optimal value of the MSE is quite large, and notably larger than the region in which the ℓ_1 minimization is able to give precise reconstruction. Another point worth noting is that the location of the spinodal depends only very weakly to the value of the noise, for a large range of matrix and measurement noise (see inset of Fig. 2 and Fig. 3): this affirms the robustness of the algorithm).

---

[1]Or Δ and η can be learned with expectation maximization within the MU-AMP [8, 13, 15], but this adds considerable computational time.
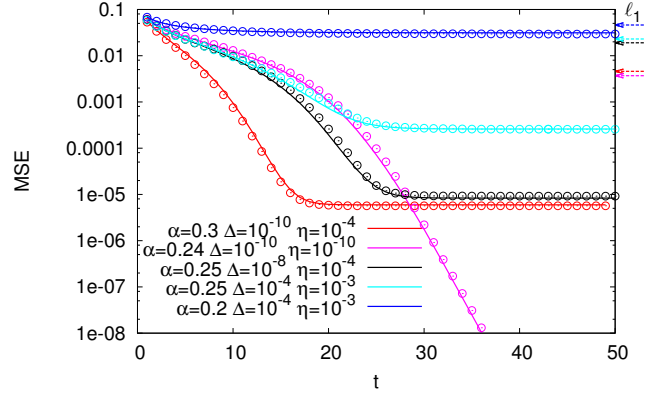

Fig. 4. Comparison of the time evolution of MSE computed with density evolution and the one found numerically using robust-AMP for systems of size $N = 25000$, $\rho = 0.1$. The agreement between theoretical predictions (eq. (17), full line) with the data from the AMP algorithm (points) is excellent. The arrows show the MSE obtained using ℓ_1 minimization on the same instances, for comparison.

α	Δ	η	r-AMP	DE	ℓ_1	c-AMP
0.3	10^{-10}	10^{-4}	-52.5	- 52.2	-23.4	-42.8
0.24	10^{-10}	10^{-10}	-101.0	-101.0	-24.3	-100.3
0.25	10^{-8}	10^{-4}	-50.4	-50.8	-17.2	-43.2
0.25	10^{-4}	10^{-3}	-35.9	-35.9	-16.4	-33.5
0.2	10^{-4}	10^{-3}	-15.3	-15.0	-13.4	-14.0

Table 1. The mean squared error in dB (i.e. $10 \log_{10} \mathrm{MSE}$) for signal of density $\rho = 0.1$. Comparison between robust-AMP, density evolution, ℓ_1 minimization and canonical-AMP that assumes that no noise was present. The agreement between r-AMP and density evolution as well as a better performance of the r-AMP over ℓ_1 and c-AMP is clear.

6. CONCLUSIONS

We have computed the Bayes-optimal value of the MSE for the reconstruction of sparse Gauss-Bernoulli signals in presence of matrix uncertainty with the replica method, and consider a variant of the AMP algorithm robust to such uncertainty and to measurement noise. Performance of this algorithm is compared to ℓ_1 minimization and the canonical-AMP that assumes that no noise was present in Table 1. Finally, we have shown that robust AMP allows one to match the optimum MSE in a large region of parameters, and that the region is very weakly sensitive to measurement or matrix noises. Note that the present analysis applies to random i.i.d. measurement matrices; it is also possible to use the seeded spatially coupled measurement matrices of [8, 13, 14] and this would lead to an even larger region of optimality-matching performance for AMP.

7. REFERENCES

[1] E. J. Candès and T. Tao, "Near-Optimal Signal Recovery From Random Projections: Universal Encoding Strategies?," *IEEE Trans. Inform. Theory*, vol. 52, pp. 5406, 2006.

[2] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, pp. 1289, 2006.

[3] H. Zhu, G. Leus, and G. Giannakis, "Sparsity-cognizant total leastsquares for perturbed compressive sampling," *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 2002–2016, 2011.

[4] M. Rosenbaum and A. Tsybakov, "Sparse recovery under matrix uncertainty," *The Annals of Statistics*, vol. 38, no. 5, pp. 2620–2651, 2010.

[5] J.T. Parker, V. Cevher, and P. Schniter, "Compressive sensing under matrix uncertainties: An approximate message passing approach," in *Conference Record of the Forty Fifth Asilomar Conference on Signals, Systems and Computers*, 2011, pp. 804–808.

[6] David L. Donoho, Arian Maleki, and Andrea Montanari, "Message-passing algorithms for compressed sensing," *Proc. Natl. Acad. Sci.*, vol. 106, no. 45, pp. 18914–18919, 2009.

[7] 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.

[8] F. Krzakala, M. Mézard, F. Sausset, Y.F. Sun, and L. Zdeborová, "Statistical physics-based reconstruction in compressed sensing," *Phys. Rev. X*, vol. 2, pp. 021005, 2012.

[9] M. Mézard, G. Parisi, and M. A. Virasoro, *Spin-Glass Theory and Beyond*, vol. 9 of *Lecture Notes in Physics*, World Scientific, Singapore, 1987.

[10] Sundeep Rangan, Alyson K Fletcher, and Vivek K Goyal, "Asymptotic analysis of map estimation via the replica method and applications to compressed sensing," *IEEE Transactions on Information Theory*, vol. 58, no. 3, pp. 1902–1923, 2012.

[11] Dongning Guo, D. Baron, and S. Shamai, "A single-letter characterization of optimal noisy compressed sensing," in *47th Annual Allerton Conference on Communication, Control, and Computing*, 2009, pp. 52 – 59.

[12] Y. Kabashima, T. Wadayama, and T. Tanaka, "A typical reconstruction limit of compressed sensing based on lp-norm minimization," *J. Stat. Mech.*, p. L09003, 2009.

[13] F. Krzakala, M. Mézard, F. Sausset, Y.F. Sun, and L. Zdeborová, "Probabilistic reconstruction in compressed sensing: Algorithms, phase diagrams, and threshold achieving matrices," *J. Stat. Mech.*, vol. P08009, 2012.

[14] David L. Donoho, Adel Javanmard, and Andrea Montanari, "Information-theoretically optimal compressed sensing via spatial coupling and approximate message passing," in *IEEE International Symposium on Information Theory Proceedings (ISIT)*, 2012, pp. 1231–1235.

[15] Jeremy Vila and Philip Schniter, "Expectation-maximization bernoulli-gaussian approximate message passing," in *Conference Record of the Forty Fifth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, 2011, pp. 799–803.

[16] J. Barbier, F. Krzakala, M. Mézard, and L. Zdeborová, "Compressed sensing of approximately-sparse signals: Phase transitions and optimal reconstruction," in *50th Annual Allerton Conference on Communication, Control, and Computing*, 2012.

[17] Michel Talagrand, "The parisi formula," *Annals of Mathematics*, pp. 221–263, 2006.

[18] David L. Donoho and Jared Tanner, "Sparse non-negative solution of underdetermined linear equations by linear programming," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 27, pp. 9446–9451, 2005.

[19] D.L. 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.

[20] Tom Richardson and Rüdiger Urbanke, *Modern Coding Theory*, Cambridge University Press, 2008.

[21] 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.