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

Florent Krzakala\textsuperscript{1,*}, Marc Mézard\textsuperscript{2} and Lenka Zdeborová\textsuperscript{3}

\textsuperscript{1} CNRS and ESPCI ParisTech 10 rue Vauquelin, UMR 7083 Gulliver Paris 75005 France
\textsuperscript{2} Ecole Normale Supérieure, 45 rue d’Ulm, Paris, France, and LPTMS-CNRS, Univ. Paris Sud
\textsuperscript{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 \cite{david7083} that this reconstruction is possible and computationally feasible in many cases using $\ell_1$-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 $\ell_1$-regularization based algorithm under matrix uncertainty, e.g. \cite{1-2075-9207} and references therein. In this paper we shall use instead a Bayesian approach. In this framework CS under matrix uncertainty was studied recently by \cite{1-2075-9207}. In particular the authors of \cite{1-2075-9207} propose a generalization of the approximate message passing algorithm (AMP) \cite{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 \cite{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 \cite{6}) and show that it is asymptotically equivalent to MU-AMP and thus also it 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

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

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 \cite{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 $P(x)$ are known (if this is not the case they can be learned via expectation-maximization \cite{8-13-15}). One could use an approximately sparse signal as in \cite{16} as well.

The $M$ measurements $y_\mu$ are obtained via noisy linear random projections of the signal

\[ y_\mu = \sum_{i=1}^{N} P_{\mu i} s_i + \xi_\mu, \quad \mu = 1, \ldots, M. \]
where we model the noise $\xi_\mu$ as a Gaussian random variable with mean 0 and variance $\Delta$, and where $F^0$ 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 $F^0$ is not known perfectly, and that we know only its noisy version:

$$F'_{\mu i} = \frac{F^0_{\mu i} + X_{\mu i} \sqrt{\eta}}{\sqrt{1 + \eta}},$$

where $X_{\mu i}$ is a white noise with mean 0 and variance $1/N$, and where $\eta$ 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 $X$, we obtain the posterior probability of one matrix element $P(F^0_{\mu i} | F'_{\mu i})$ to be a Gaussian with mean and variance

$$F_{\mu i} = \frac{F'_{\mu i}}{\sqrt{1 + \eta}}, \quad S_{\mu i} = \frac{\eta}{N(1 + \eta)}.$$  

Our goal is to reconstruct the signal $s$ based on the knowledge of $y$ and $F'$. The Bayes-optimal estimate of a signal $x^*$, that minimizes the mean-squared error (MSE) $\text{MSE} = \sum_{i=1}^{N}(s_i - x_i^*)^2/N$ with the original signal $s$, is given by

$$x_i^* = \int dx_i x_i \nu_i(x_i),$$  

where $\nu_i(x_i)$ is the marginal distribution of $x_i$ with respect to:

$$P(x, F') = \frac{P(x)}{Z} \int dF^0 P(F^0 | F') e^{-\frac{1}{2\eta} (y - F^0 s)^2}. \quad (6)$$

3. Replica Method: Optimum Reconstruction Bounds

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 sometimes possible to prove that the results derived from it are exact, see [2,10,11,12,13]. 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

$$\text{MSE}(\alpha, \rho, \Delta, \eta) = \max_{\Phi(E)} \Phi(E),$$

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

$$\Phi(E) = \frac{-\alpha}{2} \left[ \log [\Delta + E + (\rho - E)D] + \frac{\Delta + \rho}{[\Delta + E + (\rho - E)D]} \right] + (1 - \rho) \int Dz \log \left[ 1 - \rho + \frac{\rho}{\sqrt{m + 1}} \right] e^{\frac{2m}{\sqrt{m + 1}}} + \rho \int Dz \log \left[ 1 - \rho + \frac{\rho}{\sqrt{m + 1}} e^{\frac{2m}{\sqrt{m + 1}}} \right], \quad (8)$$

Note how closely related is this expression to the one derived in [13] for the usual compressed sensing with known measurement matrix (which is recovered taking $\eta = 0$).

The potential $\Phi(E)$ is shown in Fig. 1. Maximizing the potential, one obtains the Bayes-optimal MSE which we show in Fig. 2 as a function of $\eta$ for different densities $\rho$. 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 techniques—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 $\eta$. The same result was also obtained for additive measurement noise in \cite{13} and for approximate sparsity in \cite{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 \cite{6} for the minimization of $\ell_1$, and subsequently generalized in \cite{18,7,13}. This approach was adapted by Parker, Cevher and Schniter to treat the matrix uncertainty MU-AMP in \cite{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 $\mu$ we define one real number $\omega_\mu$, for each signal component $i$ we define four real numbers $\Sigma_i$, $R_i$, $a_i$, $v_i$. These quantities are updated as follows (for the derivation with these notations see \cite{13}):\

\begin{align}
V_i^{t+1} &= \frac{1}{M} \sum_{\mu} (y_\mu - \omega_\mu^t)^2, \quad (10) \\
\omega_\mu^{t+1} &= \frac{1}{\sum_i F_{\mu i}^2} \sum_i F_{\mu i} a_i^t \left( \frac{(y_\mu - \omega_\mu^t)}{V_t} - \sum_i F_{\mu i}^2 v_i^t \right), \quad (11) \\
(\Sigma_i^{t+1})^2 &= \frac{V_i^{t+1}}{\sum_{\mu} F_{\mu i}^2}, \quad (12) \\
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 (13) \\
a_i^{t+1} &= f_a \left( (\Sigma_i^{t+1})^2, R_i^{t+1} \right), \quad (14) \\
v_i^{t+1} &= f_c \left( (\Sigma_i^{t+1})^2, R_i^{t+1} \right). \quad (15)
\end{align}

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

\begin{align}
&f_a(\Sigma^2, R) = \int dx x \mathcal{M}(\Sigma^2, R, x), \quad (16) \\
f_c(\Sigma^2, R) = \int dx x^2 \mathcal{M}(\Sigma^2, R, x) - f_a^2(\Sigma^2, R). \quad (17)
\end{align}
where $\mathcal{M}(\Sigma^2, R, x)$ is the following probability distribution
\[ \mathcal{M}(\Sigma^2, R, x) = \frac{1}{Z(\Sigma^2, R)} P(x) \frac{1}{\sqrt{2\pi\Sigma}} e^{-\frac{(x-R)^2}{2\Sigma}}. \] (18)

The explicit expression for $f_a$ and $f_r$ for the Gauss-Bernoulli signal is given in [13] while the case of approximate sparsity was considered in [16]. Most previous works (e.g. [2, 13]) used a $\mu$-dependent vector $V^t_{\mu} = \Delta + \sum_i F^{2} v^t_i$ 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) $V^t_{\mu} = \Delta + \sum_i F^{2} v^t_i + \sum_i (\theta_1 + \alpha_1^2) \eta/[N(1+\eta)]$. As shown in ref. [5] this leads to a very efficient algorithm with matrix uncertainty. The expression we use instead in eq. (10) 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 (10) is that it is not need to explicitly know the value of $\eta$ and $\Delta$; 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 (10)-(15) 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 [19], and the “state evolution” in the context of CS [6, 20]. We shall not detail the computation here, it goes in the same lines as in [15]. Given the parameters $\rho, \alpha, \eta, \Delta$, the MSE follows:
\[ E^{t+1} = \int dx \, P(x) \int Dz f_c(x/z, 1/m^t, z + z/\sqrt{m^t}), \] (19)

where $m^t$ follows eq. (8), with $E^t$ on the right hand side, $E^{t=0} = \rho$, $Dz$ is a Gaussian integral, and $f_c$ is defined by eq. (17). The comparison between the evolution of the MSE in the algorithm and eq. (19) 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 (19). More specifically, the term $V^t$ from eq. (10) 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. (19). From eq. (19), 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. (8). This underlines the importance of the spinodal transition illustrated on Figs. 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 $\ell_1$ 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. 5); this shows that the robust from of the AMP algorithm is indeed robust to noise(s).

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. Finally, we have shown that 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 least squares 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*, 2011.

[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)*, 31 2011-aug. 5 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] S. Rangan, A.K. Fletcher, and V. Goyal, “Asymptotic analysis of map estimation via the replica method and applications to compressed sensing,” arXiv:0906.3234v2, 2009.

[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 *Proc. of the IEEE Int. Symposium on Information Theory (ISIT)*, 2012.

[15] 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.

[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] 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.

[18] D.L. Donoho, A. Maleki, and A. Montanari, “Message passing algorithms for compressed sensing: I. motivation and construction,” in *Information Theory Workshop (ITW), 2010 IEEE*, 2010, pp. 1–5.

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

[20] 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.