

MIT Open Access Articles

*Extension of replica analysis to MAP estimation
with applications to compressed sensing*

The MIT Faculty has made this article openly available. **Please share**
how this access benefits you. Your story matters.

Citation: Rangan, Sundeeep, Alyson K. Fletcher, and Vivek K Goyal. "Extension of Replica Analysis to MAP Estimation with Applications to Compressed Sensing." IEEE, 2010. 1543–1547. © Copyright 2010 IEEE

As Published: <http://dx.doi.org/10.1109/ISIT.2010.5513520>

Publisher: Institute of Electrical and Electronics Engineers (IEEE)

Persistent URL: <http://hdl.handle.net/1721.1/71885>

Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

Terms of Use: Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.



Extension of Replica Analysis to MAP Estimation with Applications to Compressed Sensing

Sundeep Rangan
Polytechnic Institute of New York University
Email: srangan@catt.poly.edu

Alyson K. Fletcher
University of California, Berkeley
Email: alyson@eecs.berkeley.edu

Vivek K Goyal
Massachusetts Institute of Technology
Email: vgoyal@mit.edu

Abstract—The replica method is a non-rigorous but widely-accepted technique from statistical physics used in the asymptotic analysis of large, random, nonlinear problems. This paper applies the replica method to analyze non-Gaussian maximum a posteriori (MAP) estimation. The main result is a counterpart to Guo and Verdú’s replica analysis of minimum mean-squared error estimation.

The replica MAP analysis can be readily applied to many estimators used in compressed sensing, including basis pursuit, lasso, linear estimation with thresholding, and zero norm-regularized estimation. Among other benefits, the replica method provides a computationally-tractable method for exactly computing various performance metrics including mean-squared error and sparsity pattern recovery probability.

I. INTRODUCTION

Estimating a vector $\mathbf{x} \in \mathbb{R}^n$ from measurements of the form

$$\mathbf{y} = \Phi \mathbf{x} + \mathbf{w}, \quad (1)$$

where $\Phi \in \mathbb{R}^{m \times n}$ represents a known *measurement matrix* and $\mathbf{w} \in \mathbb{R}^m$ represents measurement errors or noise, is a generic problem that arises in a range of circumstances. Even if the priors for \mathbf{x} and \mathbf{w} are separable, optimal estimation with respect to any interesting criterion is generally nonlinear and has behavior that is not easy to characterize. This is because the matrix Φ couples the n unknown components of \mathbf{x} with the m measurements in the vector \mathbf{y} .

Under the assumption of i.i.d. \mathbf{x} and Gaussian i.i.d. \mathbf{w} , Guo and Verdú [1] used the replica method from statistical physics to analyze the class of estimators that are minimum mean-squared error (MMSE) with respect to some prior on \mathbf{x} . The result of this analysis is that for certain large random Φ , there is an *asymptotic decoupling* of the estimation problem into n scalar MMSE estimation problems: the joint distribution of a component of \mathbf{x} and the corresponding component of the MMSE estimate vector $\hat{\mathbf{x}}^{\text{mmse}}(\mathbf{y})$ converges to the joint distribution between signal and estimate in a simple scalar estimation problem. From the joint distribution, various further computations can be made, such as the mean-squared error (MSE) of the estimate or the error probability of a hypothesis test computed from the estimate.

The primary contribution of this paper is an analogue to the Guo and Verdú [1] result for the class of estimators that are MAP (rather than MMSE) with respect to some

prior. Specifically, we derive a simple and asymptotically-exact scalar characterization of the componentwise behavior of MAP estimation from large, random linear measurements.

This paper summarizes the results of [2] and adds a few comments on subsequent work. Many details, including proofs, additional discussion of prior work, additional numerical results, and additional references are given in [2]. Throughout this paper we assume the validity of the main result of [1], even though it relies on certain unproven replica method hypotheses.

A. Applications to Compressed Sensing

As an application of our main result, we will develop a few analyses of estimation problems for sparse \mathbf{x} that arise in compressed sensing [3], [4]. Generically, optimal estimation of \mathbf{x} with a sparse prior is NP-hard [5]. Thus, most attention has focused on greedy heuristics and convex relaxations such as basis pursuit [6] or lasso [7]. While successful in practice, these algorithms are difficult to analyze precisely.

Most analyses of compressed sensing are not just non-Bayesian (i.e., lacking a prior on \mathbf{x}) but fully non-probabilistic, using bounded noise models and deterministic conditions on the transform matrix Φ such as on its coherence or restricted isometry constants. Unfortunately, such worst-case analyses tend to be highly conservative. In contrast, we believe the replica method is the first technique that permits an *exact* performance evaluation. Moreover, the replica method can incorporate arbitrary priors and applies to many common compressed sensing algorithms including lasso, basis pursuit, linear methods and optimal MAP estimation.

B. Related Work

Our work is not the first to use ideas from statistical physics for the study of sparse estimation. Merhav, Guo and Shamai [8] consider the estimation of correlated sparse vectors, and Kabashima, Wadayama and Tanaka [9] consider a related problem in a noise-free setting.

Subsequent to the posting of [2], complementary work was completed by Guo, Barron, and Shamai [10]. Under the measurement model (1), they show convergence to a scalar equivalent of any sufficient statistic for \mathbf{x} rather than for a specific estimator. They also show that asymptotic decoupling can be extended to any finite subset of components of \mathbf{x} .

This material is based upon work supported by the National Science Foundation under CAREER Grant No. 0729069.

It should be noted that the results in this paper only provide an *analysis* of MAP estimation—not a method to actually compute the estimate itself. Recent works [10]–[14] have all presented various belief propagation-based methods that, under certain circumstances, can achieve the performance predicted by the replica analysis.

C. Outline

The precise estimation problem is described in Section II. We then present our main result, the Replica MAP Claim, in Section III. The results are applied to the analysis of compressed sensing algorithms in Section IV, which is followed by numerical simulations in Section V.

II. ESTIMATION PROBLEM AND ASSUMPTIONS

Consider the estimation of a random vector $\mathbf{x} \in \mathbb{R}^n$ from linear measurements of the form

$$\mathbf{y} = \Phi \mathbf{x} + \mathbf{w} = \mathbf{A} \mathbf{S}^{1/2} \mathbf{x} + \mathbf{w}, \quad (2)$$

where $\mathbf{y} \in \mathbb{R}^m$ is a vector of observations, $\Phi = \mathbf{A} \mathbf{S}^{1/2}$, $\mathbf{A} \in \mathbb{R}^{m \times n}$ is a measurement matrix, \mathbf{S} is a diagonal matrix of positive scale factors,

$$\mathbf{S} = \text{diag}(s_1, \dots, s_n), \quad s_j > 0, \quad (3)$$

and $\mathbf{w} \in \mathbb{R}^m$ is zero-mean, white Gaussian noise. We consider a sequence of such problems indexed by n , with $n \rightarrow \infty$. For each n , the problem is to determine an estimate $\hat{\mathbf{x}}$ of \mathbf{x} from the observations \mathbf{y} knowing the measurement matrix \mathbf{A} and scale factor matrix \mathbf{S} .

The components x_j of \mathbf{x} are i.i.d. with some zero-mean prior probability distribution $p_0(x_j)$. The per-component variance of the Gaussian noise is $\mathbf{E}|w_j|^2 = \sigma_0^2$. We use the subscript “0” on the prior and noise level to differentiate these quantities from certain “postulated” values to be defined later. When we develop applications, the prior $p_0(x_j)$ will incorporate presumed sparsity of the components of \mathbf{x} .

In (2), we have factored $\Phi = \mathbf{A} \mathbf{S}^{1/2}$ so that even with the i.i.d. assumption on x_j s above and an i.i.d. assumption on entries of \mathbf{A} , the model can capture variations in powers of the components of \mathbf{x} that are known *a priori* at the estimator. Variations in the power of \mathbf{x} that are not known to the estimator should be captured in the distribution of \mathbf{x} .

We summarize the situation and make additional assumptions to specify the problem precisely as follows:

- (a) The number of measurements $m = m(n)$ is a deterministic quantity that varies with n and satisfies

$$\lim_{n \rightarrow \infty} n/m(n) = \beta$$

for some $\beta \geq 0$.

- (b) The components x_j of \mathbf{x} are i.i.d. with probability distribution $p_0(x_j)$.
- (c) The noise \mathbf{w} is Gaussian with $\mathbf{w} \sim \mathcal{N}(0, \sigma_0^2 \mathbf{I}_m)$.
- (d) The components of the matrix \mathbf{A} are i.i.d. zero mean with variance $1/m$.
- (e) The scale factors s_j are i.i.d. and almost surely positive.
- (f) The scale factor matrix \mathbf{S} , measurement matrix \mathbf{A} , vector \mathbf{x} and noise \mathbf{w} are all independent.

III. REPLICA MAP CLAIM

Let $\mathcal{X} \subseteq \mathbb{R}$ be some (measurable) set and consider an estimator of the form

$$\hat{\mathbf{x}}^{\text{map}}(\mathbf{y}) = \arg \min_{\mathbf{x} \in \mathcal{X}^n} \frac{1}{2\gamma} \|\mathbf{y} - \mathbf{A} \mathbf{S}^{1/2} \mathbf{x}\|_2^2 + \sum_{j=1}^n f(x_j), \quad (4)$$

where $\gamma > 0$ is an algorithm parameter and $f : \mathcal{X} \rightarrow \mathbb{R}$ is some scalar-valued, non-negative cost function. We will assume that the objective function in (4) has a unique essential minimizer for almost all \mathbf{y} .

It can be verified that, for any $u > 0$, (4) is equivalent to a MAP estimator with the *postulated prior*

$$p_u(\mathbf{x}) = \left[\int_{\mathbf{x} \in \mathcal{X}^n} \exp(-uf(\mathbf{x})) d\mathbf{x} \right]^{-1} \exp(-uf(\mathbf{x})). \quad (5)$$

and *postulated noise level*

$$\sigma_u^2 = \gamma/u. \quad (6)$$

To analyze this MAP estimator, we consider a sequence of MMSE estimators. For each u , let

$$\hat{\mathbf{x}}^u(\mathbf{y}) = \mathbf{E}(\mathbf{x} | \mathbf{y}; p_u, \sigma_u^2), \quad (7)$$

which is the MMSE estimator of \mathbf{x} under the postulated prior p_u in (5) and noise level σ_u^2 in (6). Using a standard large deviations argument, one can show that under suitable conditions

$$\lim_{u \rightarrow \infty} \hat{\mathbf{x}}^u(\mathbf{y}) = \hat{\mathbf{x}}^{\text{map}}(\mathbf{y})$$

for all \mathbf{y} . Under the assumption that the behaviors of the MMSE estimators are described as in [1], we can then extrapolate the behavior of the MAP estimator. This will yield our main result, the Replica MAP Claim.

To state the claim, define the scalar MAP estimator

$$\hat{x}_{\text{scalar}}^{\text{map}}(z; \lambda) = \arg \min_{x \in \mathcal{X}} F(x, z, \lambda) \quad (8)$$

where

$$F(x, z, \lambda) = \frac{1}{2\lambda} |z - x|^2 + f(x). \quad (9)$$

The Replica MAP Claim pertains to the estimator (4) applied to the sequence of estimation problems defined in Section II. Our assumptions are as follows:

Assumption 1: For u sufficiently large, the replica assumptions apply to the MMSE estimator $\hat{x}^u(\mathbf{y})$. See [2] for more details on this assumption.

Assumption 2: Suppose for each n , $j \in \{1, \dots, n\}$ is some index and $\hat{x}_j^u(n)$ is the MMSE estimate of the component x_j based on the postulated prior p_u and noise level σ_u^2 . Then, assume that limits can be interchanged to give the following equality:

$$\lim_{u \rightarrow \infty} \lim_{n \rightarrow \infty} \hat{x}_j^u(n) = \lim_{n \rightarrow \infty} \lim_{u \rightarrow \infty} \hat{x}_j^u(n),$$

where the limits are in distribution.

Assumption 3: For every n , \mathbf{A} , and \mathbf{S} , assume that for almost all \mathbf{y} , the minimization in (4) achieves a unique essential minimum. Here, essential should be understood in

the standard measure theoretic sense in that the minimum and essential infimum agree.

Assumption 4: Assume that $f(x)$ is non-negative and satisfies

$$\lim_{|x| \rightarrow \infty} f(x)/\log|x| = \infty,$$

where the limit must hold over all sequences $x \in \mathcal{X}$ with $|x| \rightarrow \infty$. If \mathcal{X} is compact, this limit is automatically satisfied.

Assumption 5: For all $\lambda \in \mathbb{R}$ and almost all z , the minimization in (8) has a unique, essential minimum. Moreover, for all λ and almost all z , there exists a $\sigma^2(z, \lambda)$ such that

$$\lim_{x \rightarrow \hat{x}} \frac{|x - \hat{x}|^2}{2(F(x, z, \lambda) - F(\hat{x}, z, \lambda))} = \sigma^2(z, \lambda), \quad (10)$$

where $\hat{x} = \hat{x}_{\text{scalar}}^{\text{map}}(z; \lambda)$.

We can now state our main result.

Replica MAP Claim: Consider the estimation problem in Section II. Let $\hat{\mathbf{x}}^{\text{map}}(\mathbf{y})$ be the MAP estimator (4) defined for some $f(x)$ and $\gamma > 0$ satisfying Assumptions 1–5. For each n , let $j = j(n)$ be some deterministic component index with $j(n) \in \{1, \dots, n\}$. Then:

- (a) As $n \rightarrow \infty$, the random vectors $(x_j, s_j, \hat{x}_j^{\text{map}})$ converge in distribution to the random vector (x, s, \hat{x}) for some limiting *effective noise levels* σ_{eff}^2 and γ_p . Here, x , s , and v are independent with $x \sim p_0(x)$, $s \sim p_S(s)$, $v \sim \mathcal{N}(0, 1)$, and

$$\hat{x} = \hat{x}_{\text{scalar}}^{\text{map}}(z, \lambda_p) \quad (11a)$$

$$z = x + \sqrt{\mu}v, \quad (11b)$$

where $\mu = \sigma_{\text{eff, map}}^2/s$ and $\lambda_p = \gamma_p/s$.

- (b) The effective noise levels $\sigma_{\text{eff, map}}^2$ and γ_p satisfy the equations

$$\sigma_{\text{eff, map}}^2 = \sigma_0^2 + \beta \mathbf{E}[s|x - \hat{x}|^2] \quad (12a)$$

$$\gamma_p = \gamma + \beta \mathbf{E}[s\sigma^2(z, \lambda_p)], \quad (12b)$$

where the expectations are taken over $x \sim p_0(x)$, $s \sim p_S(s)$, and $v \sim \mathcal{N}(0, 1)$, with \hat{x} and z defined in (11).

Analogously to the main result of [1], the Replica MAP Claim asserts that asymptotic behavior of the MAP estimate of any single component of \mathbf{x} is described by a simple equivalent scalar estimator. In the equivalent scalar model, the component of the true vector \mathbf{x} is corrupted by Gaussian noise and the estimate of that component is given by a scalar MAP estimate of the component from the noise-corrupted version.

IV. ANALYSIS OF COMPRESSED SENSING

Our results thus far hold for any separable distribution for \mathbf{x} and under mild conditions on the cost function f . In this section, we describe how the choice of f yields MAP estimators relevant to compressed sensing; many more details appear in [2]. We reiterate that the role of f is to determine the estimator; it is not the same as specifying the true prior for \mathbf{x} in the problem of interest. In the subsequent section,

the numerical evaluations and simulations are given for sparse priors for \mathbf{x} .

Choosing the cost function $f(x) = \frac{1}{2}|x|^2$, which corresponds to the negative log of a Gaussian prior with zero mean and unit variance, allows us to analyze linear estimation. Of course, this is not the easiest route to analyzing linear estimation, and the full analysis merely recovers known results.

We next consider the lasso estimate, which is given by

$$\hat{\mathbf{x}}^{\text{lasso}}(\mathbf{y}) = \arg \min_{\mathbf{x} \in \mathbb{R}^n} \frac{1}{2\gamma} \|\mathbf{y} - \mathbf{A}\mathbf{S}^{1/2}\mathbf{x}\|_2^2 + \|\mathbf{x}\|_1, \quad (13)$$

where $\gamma > 0$ is an algorithm parameter. The estimator is essentially a least-squares estimator with an additional $\|\mathbf{x}\|_1$ regularization term to encourage sparsity in the solution. The parameter γ is selected to trade off the sparsity of the estimate with the prediction error.

The lasso estimator (13) is identical to the MAP estimator (4) with the cost function $f(x) = |x|$. With this cost function, $F(x, z, \lambda)$ in (9) is given by

$$F(x, z, \lambda) = \frac{1}{2\lambda}|z - x|^2 + |x|,$$

and therefore the scalar MAP estimator in (8) is given by

$$\hat{x}_{\text{scalar}}^{\text{map}}(z; \lambda) = T_\lambda^{\text{soft}}(z), \quad (14)$$

the soft thresholding operator with threshold λ .

There does not exist a postulated prior such that the lasso estimate is the MMSE estimate. Thus, the Replica MAP Claim is newly providing the ability to determine the asymptotic joint distribution between a component of \mathbf{x} and the corresponding component of $\hat{\mathbf{x}}^{\text{lasso}}(\mathbf{y})$. To our knowledge, this is the first technique to provide asymptotically-exact performance characterization for this commonly-used estimator. This applicability of the Replica MAP Claim along with inapplicability of the MMSE result of [1] holds also for the intractable estimator that we describe next.

Lasso can be regarded as a convex relaxation of zero norm-regularized estimation

$$\hat{\mathbf{x}}^{\text{zero}}(\mathbf{y}) = \arg \min_{\mathbf{x} \in \mathbb{R}^n} \frac{1}{2\gamma} \|\mathbf{y} - \mathbf{A}\mathbf{S}^{1/2}\mathbf{x}\|_2^2 + \|\mathbf{x}\|_0, \quad (15)$$

where $\|\mathbf{x}\|_0$ is the number of nonzero components of \mathbf{x} . To apply the Replica MAP Claim to the zero norm-regularized estimator (15), we observe that the zero norm-regularized estimator is identical to the MAP estimator (4) with the cost function

$$f(x) = \begin{cases} 0, & \text{if } x = 0; \\ 1, & \text{if } x \neq 0. \end{cases} \quad (16)$$

Technically, this cost function does not satisfy the conditions of the Replica MAP Claim. Heuristically, we can avoid this problem by considering the limit of a sequence of cost functions that do satisfy the conditions [2].

With $f(x)$ given by (16), the scalar MAP estimator in (8) is given by

$$\hat{x}_{\text{scalar}}^{\text{map}}(z; \lambda) = T_t^{\text{hard}}(z), \quad t = \sqrt{2\lambda}, \quad (17)$$

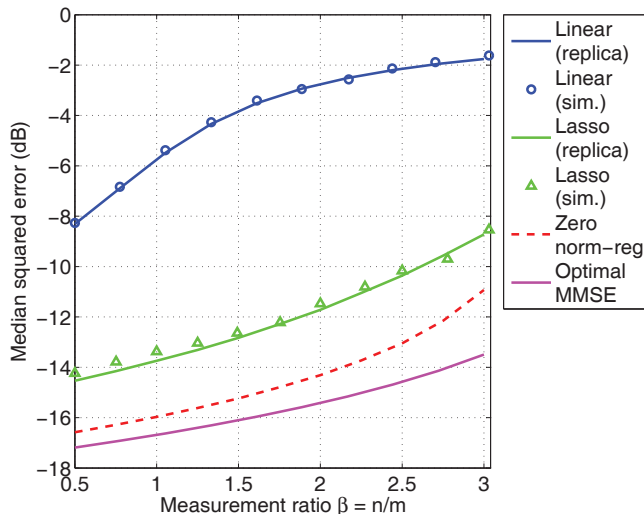


Fig. 1. MSE performance prediction with the Replica MAP Claim. Plotted is the median normalized SE for various sparse recovery algorithms: linear MMSE estimation, lasso, zero norm-regularized estimation, and optimal MMSE estimation. Solid lines show the asymptotic predicted MSE from the Replica MAP Claim. For the linear and lasso estimators, the circles and triangles show the actual median SE over 1000 Monte Carlo simulations. The unknown vector has i.i.d. Bernoulli-Gaussian components with a 90% probability of being zero. The noise level is set so that $\text{SNR}_0 = 10$ dB. See text for details.

the hard thresholding operator with threshold t .

The lasso estimator (13) and zero norm-regularized estimator (15) require the setting of a regularization parameter γ . Qualitatively, the parameter provides a mechanism to trade off the sparsity level of the estimate with the prediction error. One of the benefits of the replica analysis is that it provides a simple mechanism for optimizing the parameter level given the problem statistics. This is detailed in [2].

V. NUMERICAL SIMULATION

A. Bernoulli-Gaussian Mixture Distribution

To validate the predictive power of the Replica MAP Claim for finite dimensions, we first performed numerical simulations where the components of \mathbf{x} are distributed as

$$x_j \sim \begin{cases} \mathcal{N}(0, 1), & \text{with prob. } \rho; \\ 0, & \text{with prob. } 1 - \rho, \end{cases}$$

where ρ represents a sparsity ratio. In the experiments, $\rho = 0.1$. This is one of many possible sparse priors.

We took the vector \mathbf{x} to have $n = 100$ i.i.d. components with this prior, and we varied m for 10 different values of $\beta = n/m$ from 0.5 to 3. For the measurements (2), we took a measurement matrix \mathbf{A} with i.i.d. Gaussian components and a constant scale factor matrix $\mathbf{S} = I$. The noise level σ_0^2 was set so that signal-to-noise ratio is 10 dB.

We simulated various estimators and compared their performances against the asymptotic values predicted by the replica analysis. For each value of β , we performed 1000 Monte Carlo

trials of each estimator. For each trial, we measured the normalized squared error (SE) in dB $10 \log_{10} (\|\hat{\mathbf{x}} - \mathbf{x}\|^2 / \|\mathbf{x}\|^2)$, where $\hat{\mathbf{x}}$ is the estimate of \mathbf{x} . The results are shown in Fig. 1, with each set of 1000 trials represented by the median normalized SE in dB.

The top curve shows the performance of the linear MMSE estimator. This performance is predicted very well.

The next curve shows the performance of the lasso estimator (13). To compute the MSE predicted by the Replica MAP Claim, we numerically solve the required fixed-point equations to obtain the effective noise levels $\sigma_{\text{eff, map}}^2$ and γ_p . We then use the scalar MAP model with the estimator (14) to predict the MSE. The regularization parameter γ is determined analytically to minimize the effective noise level as described in [2, Sect. V-A], and the same regularization is used for simulations. We see from Fig. 1 that the predicted MSE matches the median SE within 0.3 dB over a range of β values.

Fig. 1 also shows the minimum MSE and the MSE from the zero norm-regularized estimator; again, the choice of regularization parameter follows [2, Sect. V-A]. For these two cases, the estimators cannot be simulated since they involve NP-hard computations. But we have included these curve to show that the replica method can be used to calculate the gap between practical and impractical algorithms. Interestingly, we see that there is about a 2 to 2.5 dB gap between lasso and zero norm-regularized estimation, and another 1 to 2 dB gap between zero norm-regularized estimation and optimal MMSE.

B. Support Recovery with Thresholding

In estimating vectors with strictly sparse priors, one important problem is to detect the *locations* of the nonzero components in the vector \mathbf{x} . Several works have attempted to find conditions under which the support of a sparse vector \mathbf{x} can be fully detected [15]–[17] or partially detected [18]–[20]. Unfortunately, with the exception of [15], the only available results are bounds that are not tight.

One of the uses of the Replica MAP claim is to *exactly* predict the fraction of support that can be detected correctly. To see how to predict the support recovery performance, observe that the Replica MAP Claim provides the asymptotic joint distribution for the vector (x_j, s_j, \hat{x}_j) , where x_j is the component of the unknown vector, s_j is the corresponding scale factor and \hat{x}_j is the component estimate. Now, in support recovery, we want to estimate θ_j , the indicator function that x_j is nonzero

$$\theta_j = \begin{cases} 1, & \text{if } x_j \neq 0; \\ 0, & \text{if } x_j = 0. \end{cases}$$

One natural estimate for θ_j is to compare the magnitude of the component estimate \hat{x}_j to some scale-dependent threshold $t(s_j)$. This idea of using thresholding for sparsity detection has been proposed in [21] and [22]. Using the joint distribution (x_j, s_j, \hat{x}_j) , one can then compute the probability of sparsity misdetection, $p_{\text{err}} = \Pr(\hat{\theta}_j \neq \theta_j)$. The probability of error can be minimized over the threshold levels $t(s)$.

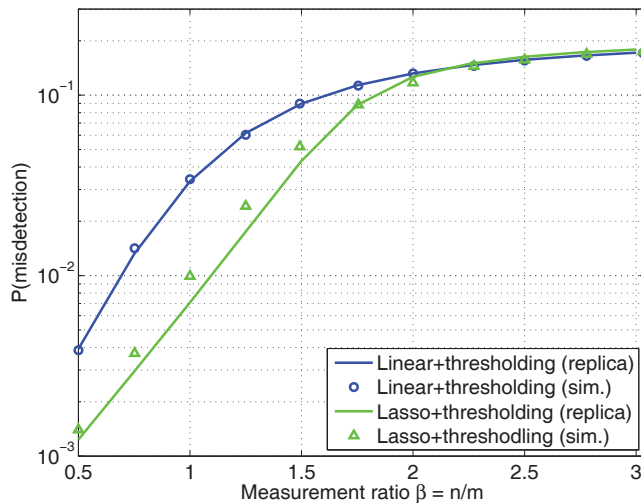


Fig. 2. Support recovery performance prediction with the replica method. The solid lines show the theoretical probability of error in sparsity misdetection using linear and lasso estimation followed by optimal thresholding. The circles and triangles are the corresponding mean probabilities of misdetection over 1000 Monte Carlo trials.

To verify the accuracy of this calculation, we generated random vectors \mathbf{x} with $n = 100$ i.i.d. components taking values in $\{0, \pm 1\}$, scaled to have SNR of 10 dB and having sparsity fraction of 0.2. Fig. 2 compares the probability of sparsity misdetection predicted by the replica method against the actual probability of misdetection based on the average of 1000 Monte Carlo trials. We tested two algorithms: linear MMSE estimation and lasso estimation. For lasso, the regularization parameter was selected for minimum MMSE. The results show a good match.

VI. CONCLUSIONS AND FUTURE WORK

We have applied the replica method from statistical physics for computing the asymptotic performance of MAP estimation of non-Gaussian vectors with large random linear measurements. The method can be readily applied to problems in compressed sensing. Moreover, we believe that the availability of a simple scalar model that exactly characterizes certain sparse estimators opens up numerous avenues for analysis. For one thing, it would be useful to see if the replica analysis of lasso can be used to recover the scaling laws of Wainwright [15] and Donoho and Tanner [23] for support recovery and to extend the latter to the noisy setting. Also, the best known bounds for MSE performance in sparse estimation are given by Haupt and Nowak [24] and Candès and Tao [25]. Since the replica analysis is asymptotically exact, we may be able to obtain much tighter analytic expressions. In a similar vein, several researchers have attempted to find information-theoretic lower bounds with optimal estimation. Using the replica analysis of optimal estimators, one may be able to improve these scaling laws as well.

REFERENCES

- [1] D. Guo and S. Verdú, "Randomly spread CDMA: Asymptotics via statistical physics," *IEEE Trans. Inform. Theory*, vol. 51, no. 6, pp. 1983–2010, Jun. 2005.
- [2] S. Rangan, A. K. Fletcher, and V. K. Goyal, "Asymptotic analysis of MAP estimation via the replica method and applications to compressed sensing," arXiv:0906.3234v1 [cs.IT], Jun. 2009.
- [3] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
- [4] E. J. Candès and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. Inform. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.
- [5] B. K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM J. Computing*, vol. 24, no. 2, pp. 227–234, Apr. 1995.
- [6] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comp.*, vol. 20, no. 1, pp. 33–61, 1999.
- [7] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Royal Stat. Soc., Ser. B*, vol. 58, no. 1, pp. 267–288, 1996.
- [8] N. Merhav, D. Guo, and S. Shamai, "Statistical physics of signal estimation in Gaussian noise: Theory and examples of phase transitions," arXiv:0812.4889v1 [cs.IT], Dec. 2008.
- [9] Y. Kabashima, T. Wadayama, and T. Tanaka, "Typical reconstruction limit of compressed sensing based on l_p -norm minimization," arXiv:0907.0914 [cs.IT], Jun. 2009.
- [10] D. Guo, D. Baron, and S. Shamai, "A single-letter characterization of optimal noisy compressed sensing," in *Proc. 47th Ann. Allerton Conf. on Commun., Control and Comp.*, Monticello, IL, Sep.–Oct. 2009.
- [11] A. Montanari and D. Tse, "Analysis of belief propagation for non-linear problems: The example of CDMA (or: How to prove Tanaka's formula)," arXiv:cs/0602028v1 [cs.IT], Feb. 2006.
- [12] D. Guo and C.-C. Wang, "Asymptotic mean-square optimality of belief propagation for sparse linear systems," in *Proc. IEEE Inform. Theory Workshop*, Chengdu, China, Oct. 2006, pp. 194–198.
- [13] D. L. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing," arXiv:0907.3574v1 [cs.IT], Jul. 2009.
- [14] S. Rangan, "Estimation with random linear mixing, belief propagation and compressed sensing," arXiv:1001.2228v1 [cs.IT], Jan. 2010.
- [15] M. J. Wainwright, "Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1 -constrained quadratic programming (lasso)," *IEEE Trans. Inform. Theory*, vol. 55, no. 5, pp. 2183–2202, May 2009.
- [16] —, "Information-theoretic limits on sparsity recovery in the high-dimensional and noisy setting," Univ. of California, Berkeley, Dept. of Statistics, Tech. Rep. 725, Jan. 2007.
- [17] A. K. Fletcher, S. Rangan, and V. K. Goyal, "Necessary and sufficient conditions for sparsity pattern recovery," *IEEE Trans. Inform. Theory*, vol. 55, no. 12, pp. 5758–5772, Dec. 2009.
- [18] M. Akçakaya and V. Tarokh, "Shannon-theoretic limits on noisy compressive sampling," *IEEE Trans. Inform. Theory*, vol. 56, no. 1, pp. 492–504, Jan. 2010.
- [19] G. Reeves, "Sparse signal sampling using noisy linear projections," Univ. of California, Berkeley, Dept. of Elec. Eng. and Comp. Sci., Tech. Rep. UCB/EECS-2008-3, Jan. 2008.
- [20] S. Aeron, M. Zhao, and V. Saligrama, "On sensing capacity of sensor networks for the class of linear observation, fixed SNR models," arXiv:0704.3434v3 [cs.IT], Jun. 2007.
- [21] H. Rauhut, K. Schnass, and P. Vanderghelynst, "Compressed sensing and redundant dictionaries," *IEEE Trans. Inform. Theory*, vol. 54, no. 5, pp. 2210–2219, May 2008.
- [22] V. Saligrama and M. Zhao, "Thresholded basis pursuit: Support recovery for sparse and approximately sparse signals," arXiv:0809.4883v2 [cs.IT], Mar. 2009.
- [23] D. L. Donoho and J. Tanner, "Counting faces of randomly-projected polytopes when the projection radically lowers dimension?" *J. Amer. Math. Soc.*, vol. 22, no. 1, pp. 1–53, Jan. 2009.
- [24] J. Haupt and R. Nowak, "Signal reconstruction from noisy random projections," *IEEE Trans. Inform. Theory*, vol. 52, no. 9, pp. 4036–4048, Sep. 2006.
- [25] E. J. Candès and T. Tao, "The Dantzig selector: Statistical estimation when p is much larger than n ," *Ann. Stat.*, vol. 35, no. 6, pp. 2313–2351, Dec. 2007.