

Optimal incorporation of sparsity information by weighted ℓ_1 optimization

Toshiyuki Tanaka
Graduate School of Informatics
University of Kyoto
Kyoto, Japan
Email: tt@i.kyoto-u.ac.jp

Jack Raymond
Department of Physics
Hong Kong University of Science and Technology
Hong Kong, China
Email: jackraymond@physics.org

Abstract—Compressed sensing of sparse sources can be improved by incorporating prior knowledge of the source. In this paper we demonstrate a method for optimal selection of weights in weighted ℓ_1 norm minimization for a noiseless reconstruction model, and show the improvements in compression that can be achieved.

I. INTRODUCTION

Compressed sensing research has the aim of improving the efficiency and reliability of source estimation when the number of measurements taken is smaller than the number of degrees of freedom in the source. Although somewhat counterintuitive, it turns out that, given some special structure in the source, it is possible to reconstruct accurately all source components [1].

Owing to the fundamental nature of the problem, it is not surprising that research and methodology span a number of fields such as image processing, topology, multi-user detection, and convex optimization [2], [3], [4], [5], [6]. In this article we also apply methods developed initially within physics, that have recently been applied to compressed sensing [7], [8], [9].

In the canonical model, a real-valued vector \mathbf{y} constitutes observations of a real-valued source vector \mathbf{x}^0 via a measurement matrix A as

$$\mathbf{y} = A\mathbf{x}^0. \quad (1)$$

The dimension of \mathbf{y} is M and the dimension of \mathbf{x}^0 is $N(\geq M)$. The parameter $\alpha = M/N$ quantifies the number of observations per the degree of freedom of the source, and we call it the compression rate. The problem of compressed sensing is to reconstruct the source \mathbf{x}^0 from the observations \mathbf{y} and the measurement matrix A .

When the compression rate α is smaller than one, the problem amounts to solving an ill-conditioned set of linear equations and hence its solution is not unique. However, if the source is suitably structured, then such structural information may be utilized to overcome the ill-conditionedness. One example is sparsity, whereby the source has many zero components. For reconstruction of sparse sources it is now well known that provided α is not too small an accurate reconstruction of the source \mathbf{x}^0 may be achieved by minimizing the ℓ_1 norm

$$\|\mathbf{x}\|_1 = \sum_i |x_i| \quad (2)$$

subject to the constraint $\mathbf{y} = A\mathbf{x}$. This minimization problem can be solved efficiently by casting it into a problem of linear programming. It has recently been shown that ℓ_1 minimization also allows perfect reconstruction of sparse sources [4], [7] with high probability for large systems provided that the compression rate α is larger than a certain threshold value α_c . Furthermore a sharp transition between perfect and imperfect reconstruction is observed at $\alpha = \alpha_c$, which we call the compression threshold.

We shall consider weighted norms, in which each component in the sum (2) has a different coefficient (weight). The consideration of non-uniform weights has been motivated by Bayesian and topological arguments, algorithmic studies and much empirical evidence. Numerous results indicate that use of non-uniform weights can reduce the compression threshold, guarantee optimality in a probabilistic sense, or otherwise improve algorithmic performance [10], [11], [12], [13]. A topology based method able to place bounds on weighted reconstruction was recently proposed, and demonstrated for the case of a source with two blocks of distinct density of non-zero components [3].

II. PROBLEM

A. Reconstruction with marginal density information

The problem we address is the following: given a mean density of non-zero components, and some additional prior information on the distribution of non-zero components, how can the weights be chosen so as to minimize the number of measurements required to reliably reconstruct the source using weighted ℓ_1 norm (w- ℓ_1) minimization.

We assume a standard model framework in which components of A are independent and identically-distributed random variables; drawn from a symmetric distribution of finite variance and higher-order moments without finite probability mass at the origin, the Gaussian distribution being one intuitive case. The observations \mathbf{y} are also random variables, determined as functions of the measurement matrix and a random source through (1).

In the standard sparse problem a fraction $0 < \bar{\rho} < 1$ of all source components are non-zero, this subset being selected uniformly at random. It is instead assumed in this article that a (potentially distinct) marginal probability is known for

every component in the source, so that x_i^0 is non-zero with probability $0 \leq \rho_i \leq 1$ independently of other components. We call the parameter ρ_i the density of component i . The prior is therefore

$$P(\mathbf{x}^0) = \prod_{i=1}^N [\rho_i S(x_i^0) + (1 - \rho_i) \delta(x_i^0)] , \quad (3)$$

where $S(x_i^0)$ is the conditional prior of x_i^0 given that it is non-zero. The conditional prior $S(x_i^0)$ is assumed unknown to the observer, but it should have finite moments and should not have finite probability weight at $x_i^0 = 0$. The densities $\{\rho_i\}$ encode prior information about sparsity of the source, are assumed available to the observer, and will be called marginal density information. This definition covers the case of block-structured sources, wherein a source is divided into sets of components according to their densities, including two blocks as a special case.

To each source component is also associated a weight $w_i > 0$. The estimate of the source is obtained as a solution to an optimization problem: minimization of the weighted ℓ_1 norm

$$\min_{\mathbf{x}} \left(\sum_i w_i |x_i| \right) \quad \text{subject to} \quad \mathbf{y} = \mathbf{A}\mathbf{x} . \quad (4)$$

We note irrespective of weighting a trivial lower bound on the compression required for perfect reconstruction: $\alpha_c \geq \bar{\rho}$. The lower bound is easily achieved, by a matrix inverse operation, when the positions of all zeros in the source are known – or equivalently all the densities ρ_i are extremal, 0 or 1. Any variation from these extremal values for fixed $\bar{\rho}$ increases the uncertainty in the source values, and so will increase the compression threshold.

The principal results in this article are regarding:

- (1) The compression threshold α_c for perfect reconstruction, with high probability, given a set of weights;
- (2) The assignment of weights, as a function of the marginal densities, that achieves a minimal α_c , and the value of this minimum.

III. ANALYSIS FOR w - ℓ_1 MINIMIZATION

A. Overview

Rather than directly finding and classifying minima of the w - ℓ_1 minimization problem, an auxiliary probability distribution may be introduced, for which the *maximum a posteriori* (MAP) solutions are equivalent to the solutions of the optimization problem. A parameter β is introduced to soften the probability distribution and to allow application of analytic methods, and this is finally taken to be large so as to recover the posterior mean coincident with the MAP solution as well as the solution of the w - ℓ_1 minimization. The probability is the model posterior

$$P(\mathbf{x}|\mathbf{y}, A) = \frac{1}{Z} \delta(\mathbf{A}\mathbf{x} - \mathbf{y}) \exp \left(-\beta \sum_i w_i |x_i| \right) \quad (5)$$

where Z is the normalization for the probability distribution. In the limit of large β the probability measure will be

concentrated on values \mathbf{x} that solve the weighted minimization problem (4). We further define a generating function in the large-system limit and its expectation value with respect to measurements

$$f = f(\beta, N) = -\frac{1}{\beta N} \log Z, \quad \bar{f} = \lim_{\beta \rightarrow \infty} \lim_{N \rightarrow \infty} \mathbb{E}_{A, \mathbf{y}}[f] . \quad (6)$$

The function $f(\beta, N)$ is linearly related to a generalized version of the mutual information between the source and the observations [14], [15].

Statistics on w - ℓ_1 solutions are extracted from f , or \bar{f} , by adding small perturbations to the model posterior and taking derivatives. Consider adding a small perturbation $-\beta h \|\mathbf{x} - \mathbf{x}^0\|_2^2$ to the exponent of (5), whereby the normalization Z now depends on h as well. Here, $\|\cdot\|_2$ denotes the ℓ_2 norm. The normalized mean square error (MSE) with respect to the model posterior (5) is given by

$$N^{-1} \mathbb{E}[\|\mathbf{x} - \mathbf{x}^0\|_2^2] = \lim_{h \rightarrow 0} \frac{\partial}{\partial h} f(\beta, N) , \quad (7)$$

which in the limit $\beta \rightarrow \infty$ gives the normalized MSE of the w - ℓ_1 solution. If MSE is evaluated to be zero, this implies that the w - ℓ_1 minimization is sufficient for perfect reconstruction.

An analysis of f is problematic owing to its definition in terms of random measurement parameters, but we mitigate for this by studying the asymptotics. Many interesting statistics, such as MSE, are found to be *self-averaging* – the statistics derived from f for almost all \mathbf{y} and A converge to those derived from \bar{f} for large N . We thus analyze the expectation value \bar{f} in place of the random object f , expressing the former in an analytic form at leading orders in N and β , thereby demonstrating conditions for perfect reconstruction.

In our analysis we assume, besides the self-averaging property, the validity of exchanging order of limits. We will solve the problem by the replica method, which introduces further heuristic assumptions – the *replica symmetric* (RS) solution is developed. We are unable to verify rigorously each assumption, but the methodology is a standard one [7], [8], [9], and a catalogue of successes indicates the results herein should be treated as exact. Some closely related results in compressed sensing [7] are already known to coincide with those of rigorous methodologies [4].

B. The replica method

The expectation with respect to the random variables (6) is taken by a standard approach. The generating function may finally be expressed in an extremization form, as

$$\begin{aligned} \bar{f} = \text{Ext}_{\mathbf{r}, m, \chi, q, \hat{m}, \hat{\chi}, \hat{q}} \left\{ \alpha \frac{\bar{\rho} - 2m + q}{\chi} + m\hat{m} - \frac{q\hat{q}}{2} + \frac{\chi\hat{\chi}}{2} \right. \\ \left. + \mathbb{E}_{\{\rho_i, w_i\}} \left[\int \left((1 - \rho_i) \delta(x^0) w_i^2 \epsilon(\lambda \sqrt{\hat{\chi}}/w_i, \hat{q}) \right. \right. \right. \\ \left. \left. \left. + \rho_i S(x^0) w_i^2 \epsilon((\sqrt{\hat{\chi}}\lambda + \hat{m}x^0)/w_i, \hat{q}) \right) D\lambda dx^0 \right] \right\} , \quad (8) \end{aligned}$$

where $D\lambda = (2\pi)^{-1/2} e^{-\lambda^2/2} d\lambda$ is a standard Gaussian measure, where $\mathbb{E}_{\{\rho_i, w_i\}}[\cdot \cdot \cdot] = N^{-1} \sum_{i=1}^N (\cdot \cdot \cdot)$ is an empirical

average with respect to the densities and weights of source components, and where the function $\epsilon(\cdot, \cdot)$ is defined as

$$\epsilon(a, b) = -\frac{(|a| - 1)^2}{2b} \Theta(|a| > 1). \quad (9)$$

A solution is the extremum where the partial derivatives with respect to the six order parameters are zero, and the set of self-consistent equations describing the order parameters are called the saddle-point equations. Note that for the special case of uniform weights and Gaussian $S(x_i^0)$, (8) becomes identical to results derived in the absence of marginal prior knowledge [7].

The order parameters for the correct solution of (8) coincide with informative quantities: q is the mean-squared estimate $N^{-1} \mathbb{E}[\|\mathbf{x}\|_2^2]$; m is the overlap of source and estimate $N^{-1} \mathbb{E}[\mathbf{x} \cdot \mathbf{x}^0]$; and χ is a statistic on the pair correlation functions. The mean square error is $N^{-1} \mathbb{E}[\|\mathbf{x} - \mathbf{x}^0\|_2^2] = \bar{\rho} - 2m + q$.

The six saddle-point equations are easily reduced to three equations on the conjugate parameters (denoted by hat) \hat{m} , \hat{q} and $\hat{\chi}$, and further to two equations by identifying $\hat{q} = \hat{m}$. Given a specific parameterization of the weights and measurement process these might be solved numerically. However, given our noiseless model (1) it is known that the mean square error will be zero in some range of large α , and reasonable weights. This perfect reconstruction solution was found in the unweighted case with $1/\hat{m} = 0$ [7], and generalizes to the weighted case.

We can solve the equations to leading order in $1/\hat{m}$. At leading order many features of $S(x_i^0)$, the prior on non-zero components, are inconsequential. It is convenient to define two functions of $\hat{\chi}$

$$g_1(\hat{\chi}) = \bar{\rho} + \mathbb{E}[(1 - \rho_i)2\mathcal{Q}(u_i)], \quad (10)$$

where $u_i = w_i/\sqrt{\hat{\chi}}$ is a rescaled weight, and

$$g_2(\hat{\chi}) = \bar{\rho} + \mathbb{E}[\rho_i(u_i)^2] - \mathbb{E}\left[(1 - \rho_i)u_i \frac{2 \exp(-u_i^2/2)}{\sqrt{2\pi}}\right] + \mathbb{E}\left[(1 - \rho_i)(1 + (u_i)^2)2\mathcal{Q}(u_i)\right], \quad (11)$$

where $\mathcal{Q}(z) = \int_z^\infty D\lambda$ is the conventional Q-function. The two saddle-point equations may be expressed at leading order as

$$\hat{m}^{-1} = \hat{m}^{-1} \left(\frac{\alpha}{g_1(\hat{\chi})} \right)^{-1}, \quad (12)$$

and

$$\hat{\chi} = \hat{\chi} \frac{\alpha g_2(\hat{\chi})}{(g_1(\hat{\chi}))^2}, \quad (13)$$

where the latter equation can be solved numerically.

C. Stability criteria

For the RS solution of the replica method a local stability analysis gives necessary criteria for the validity of any solutions found, and can be studied by three eigenvalues derived from the full, rather than RS (8), saddle-point form [16]. Two eigenvalues relate to instabilities consistent with the RS saddle-point equation (8), and can be evaluated within the RS

framework, by analysis of (12) and (13). The final 'replicon' eigenvalue can indicate a failure of the RS assumption yielding (8), but cannot be derived from (8). When the RS assumption fails in this sense, a symmetry breaking approximation should be required.

Stability within the RS framework requires that a small fluctuation in the order parameters of the right-hand side of (12) and (13) decays to zero under iteration of the equations. Small fluctuations in $\hat{\chi}$ and \hat{m}^{-1} about the perfect reconstruction solution are stable provided that

$$\alpha > g_1(\hat{\chi}). \quad (14)$$

It can be seen that for large α and small $\bar{\rho}$ these conditions will be met. The replicon eigenvalue is consistent with the stability of the solution.

IV. MAIN RESULTS

A. Threshold equations

Stability and uniqueness of the perfect reconstruction solution implies success of w - ℓ_1 reconstruction up to errors of order $1/N$, with high probability. The instability criterion is met as α is decreased from one so that there exists a compression threshold $\alpha_c \leq 1$. A pair of equations describe the threshold in terms of only one order parameter $\hat{\chi}$

$$\alpha_c = g_1(\hat{\chi}) \quad \text{and} \quad g_3(\hat{\chi}) = g_2(\hat{\chi}) - g_1(\hat{\chi}) = 0. \quad (15)$$

The latter equality is a consequence of (13) combined with (14), and is independent of α_c . Aside from the explicit dependence on $\hat{\chi}$, there is a dependence on the densities and weights. The two equations (15) provide implicit expressions for the dependences of the compression threshold α_c on the set $\{\rho_i, w_i\}$ of densities and weights.

If one assumes uniform weights, the compression threshold α_c is described as an implicit function of $\bar{\rho}$ by

$$\alpha_c^{-1} = 1 + \sqrt{\frac{\pi}{2\hat{\chi}}} \exp\left(\frac{1}{2\hat{\chi}}\right) \left[1 - 2\mathcal{Q}\left(\frac{1}{\sqrt{\hat{\chi}}}\right)\right], \quad (16)$$

and

$$\frac{\bar{\rho}}{1 - \bar{\rho}} = \sqrt{\frac{2\hat{\chi}}{\pi}} \exp\left(-\frac{1}{2\hat{\chi}}\right) - 2\mathcal{Q}\left(\frac{1}{\sqrt{\hat{\chi}}}\right). \quad (17)$$

For an inhomogeneous system of fixed $\bar{\rho}$, any knowledge of the marginal densities constitutes additional information. Incorporation of non-uniform weights introduces additional degrees of freedom in the model that can be exploited to reduce α_c . It also means that it is necessary to specify the weights according to the given set of densities in order to find the optimum value of α_c .

B. Optimization of α_c by weight selection

In order to select optimally the weights, we consider the optimization problem of minimizing the compression threshold with respect to the weights. For this purpose we can take $\hat{\chi}$ and $\{\rho_i\}$ to be fixed parameters and minimize $\alpha_c = g_1$ with respect to the weights $\{w_i\}$ subject to the constraint that g_3

is zero (see (15)). The constraint can be dealt with via the Lagrange multiplier method as

$$\min_{\{w_i\}, \lambda} (g_1 - \lambda g_3). \quad (18)$$

The derivative with respect to w_i leads to the following criteria

$$u_i \exp(u_i^2/2) \left(\frac{\rho_i}{1 - \rho_i} + 2\mathcal{Q}(u_i) \right) = \frac{1 + \lambda}{\lambda \sqrt{2\pi}}, \quad (19)$$

combined with the derivative with respect to λ , which reads

$$\frac{1}{\sqrt{2\pi}} \mathbb{E} [u_i(1 - \rho_i) \exp(-u_i^2/2)] \left(\frac{1 + \lambda}{\lambda} - 2 \right) = 0, \quad (20)$$

yielding $\lambda = 1$. Therefore the set of equations (19), each of which determines one weight, are independent given $\hat{\chi}$ and ρ_i .

Using this independence it is possible to optimally set the weights for an arbitrary density distribution $\{\rho_i\}$ by the following procedure. Solve (19) for u_i , and the solution is a function of ρ_i only. The set of solutions $\{u_i\}$ defines the optimal weights up to the overall scaling. Observing that the w - ℓ_1 minimization problems are invariant under the overall scaling, every weight can be assigned straightforwardly as $w_i = u_i$. This set of weights will minimize α_c .

To evaluate the minimum of α_c , one observes from (10) and (15) that $\alpha_c = \mathbb{E}[\alpha(\rho_i)]$ holds, where $\alpha(\rho_i) = \rho_i + (1 - \rho_i)2\mathcal{Q}(u_i(\rho_i))$. This expression can be understood as if each component of the source, with density ρ , would require a compression rate at least $\alpha(\rho)$ for its perfect reconstruction, and the lower bound of the total compression rate is the empirical average of this componentwise bound over all components. This interpretation would further suggest a possibly fundamental importance of the quantity $\alpha(\rho)$ as a measure of some sort of information associated with a random variable that takes non-zero values with probability ρ . One can also confirm, by particularizing our results to the unweighted system, that $\alpha(\rho)$ gives the compression threshold of the unweighted system with density ρ . Thus, evaluation of α_c for the weighted system only requires the densities $\{\rho_i\}$ of the system and the curve $\alpha(\rho)$ for the unweighted system. Moreover, the curve for the unweighted system is convex upward, so that Jensen's inequality tells us that α_c is below $\alpha(\bar{\rho})$, except when the density distribution is concentrated at a single point, implying that the w - ℓ_1 minimization improves the compression threshold over the unweighted counterpart.

It should also be noted that, when a block structure can be assumed for the source, the minimum α_c is equal to the compression rate that would be achievable if each block were measured separately with the optimum compression rate for that block and then reconstructed individually from other blocks. This observation implies that introduction of optimal weights successfully compensates degradation due to intermixing of blocks of different densities in measurements.

Finally we can note an interesting asymptotic in the result applicable in the case that $\bar{\rho}$ is close to one ($\hat{\chi}$ small). In this case the optimal weights are assigned according to a simplified

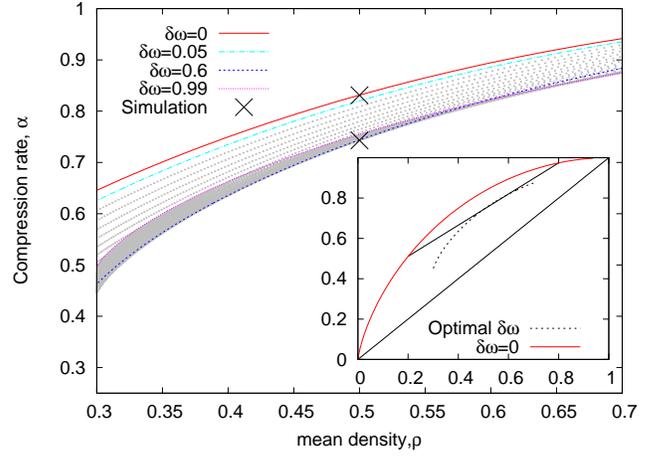


Fig. 1. (color online) Compression threshold α_c as a function of density $\{0.3 < \bar{\rho} < 0.7, \delta\rho = 0.3\}$ for various δw . Four curves are highlighted from a range $0 \leq \delta w \leq 0.99$. The lower hull of all curves indicates the minimum α_c achievable by optimal weight selection. Inset: The hull of all curves (corresponding to the maximum compression) can be constructed from the unweighted ($\delta w = 0$) curve by a simple linear construction.

form of (19)

$$\rho_i/(1 - \rho_i) \sim \sqrt{\frac{2}{\pi}} \frac{\exp(-u_i^2/2)}{u_i^3}. \quad (21)$$

V. A TWO-BLOCK EXAMPLE

A. Setting

As a simple case we consider the source consisting of two equally-sized blocks labeled $b = \pm 1$, such that components in the block -1 are non-zero with some uniform lower probability than components of block $+1$. A corresponding asymmetry in the weights can be assumed, but weights must also be uniform within any block. This case allows a deal of intuition and has been studied previously owing to its simple structure [3].

B. The large-system limit

Let the density and the weight of block $b = \pm 1$ be $\rho_b = \bar{\rho} + b\delta\rho$ and $w_b = 1 - b\delta w$, respectively, with $\delta w, \delta\rho \geq 0$. Figure 1 demonstrates the result for a variety of weight asymmetry δw given a density asymmetry $\delta\rho = 0.3$. Since the density in each block must be in $(0, 1)$ the relevant values of $\bar{\rho}$ are confined to the interval $(0.3, 0.7)$. Curves for $0 \leq \delta w \leq 0.99$ are plotted, with four labeled cases, the uppermost curve being the unweighted case. It can be seen that the compression threshold is reduced by allowing asymmetric weights, and that the lower envelope of all curves indicates the achievable performance by optimal selection amongst the weights. For the system $(\rho_{+1}, \rho_{-1}) = (0.8, 0.2)$ the midpoint of the line connecting $(\rho_{+1}, \alpha(\rho_{+1}))$ and $(\rho_{-1}, \alpha(\rho_{-1}))$ coincides with the lower envelope at $\bar{\rho} = 0.5$ corresponding to the optimal compression rate for the two-block model.

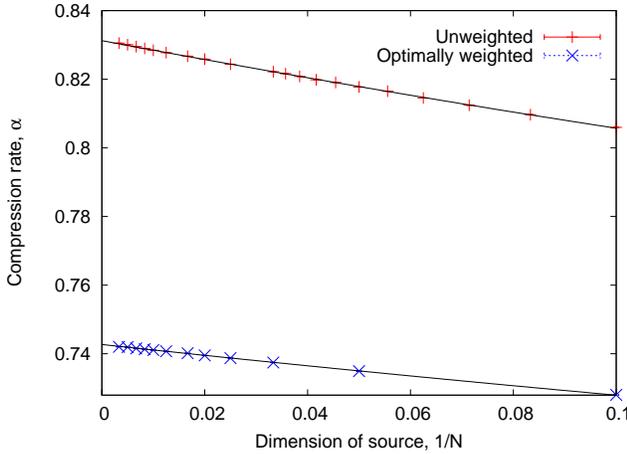


Fig. 2. (color online) Compression threshold α_c for unweighted ($\delta w = 0$) versus optimally weighted ($\delta w = 0.684$) reconstruction for a two-block model ($\rho_{+1} = 0.8, \rho_{-1} = 0.2$) for various system sizes. Each symbol represents the mean threshold from 10^6 trials each with independently generated measurements, error bars are small by comparison with symbol size. The data is fitted by χ^2 regression.

C. Numerical verification

To test the theory we have generated a number of instances of the problem for various system sizes and solved these by linear programming. Components of A and non-zero components of \mathbf{x}^0 were sampled from normal distributions. Optimally weighted, and unweighted, ℓ_1 minimization problems were solved to determine the compression threshold where optimal reconstruction failed. Figure 2 demonstrates result for a two-block system with a uniform weighting $w_{+1} = w_{-1}$ and optimized weighting. It can be seen that the thresholds are distinct with a significant improvement in performance for the optimally weighted case.

Results of fitting these numerical data with second-order polynomials in $1/N$, obtained by χ^2 regression, are also plotted in Fig. 1. Extrapolation to $1/N \rightarrow 0$ yields an estimate of the compression threshold $\alpha_c = 0.74273 \pm 0.00006$ for the optimally weighted case, which is in agreement with the analytical result $\alpha_c = 0.74272$. The extrapolated result for the unweighted case $\alpha_c = 0.83132 \pm 0.00006$ is also in agreement with the analytical result $\alpha_c = 0.83130$.

VI. CONCLUSION

This paper has demonstrated a method for optimal selection of weights in the w - ℓ_1 minimization utilizing prior knowledge about densities, and thereby providing the optimal threshold for compression. The result is a simple one with decoupling structure for source components, and described by a single order parameter in the case of perfect reconstruction. The threshold in the compression rate for which perfect reconstruction is possible in a system of known marginal densities can be straightforwardly derived from the threshold curve for the unweighted system with a simple graphical procedure. This work should in future be extended to consider the effect of noise and of correlations between the source components.

The method relied on the replica method and a saddle-point formulation, which although complicated in origin provides a concise and intuitive saddle-point framework from which to derive results.

The analysis presented has been verified in experiment and provides a mechanism that may be immediately incorporated in practical problems where marginal density information is available. A complete description of the replica analysis and various extensions will be forthcoming in an article under preparation.

ACKNOWLEDGMENT

Support from the Grant-in-Aid for Scientific Research on Priority Areas “Deepening and Expansion of Statistical-Mechanical Informatics” by the Ministry of Education, Culture, Sports, Science and Technology, Japan (No. 18096010), is acknowledged. J.R. is supported in part by Research Grants Council of Hong Kong (Grant No. HKUST 604008, 603607).

REFERENCES

- [1] D. Donoho, “Compressed sensing,” *IEEE Trans. Info. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [2] F. Parvaresh, H. Vikalo, S. Misra, and B. Hassibi, “Recovering sparse signals using sparse measurement matrices in compressed dna microarrays,” *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 3, pp. 275–285, 2008.
- [3] W. Xu, A. Amin Khajehnejad, S. Avestimehr, and B. Hassibi, “Breaking through the thresholds: an analysis for iterative reweighted ℓ_1 minimization via the Grassmann angle framework,” preprint, [online] arXiv:0904.0994v1 [math.PR], 2009.
- [4] D. Donoho and J. Tanner, “Counting faces of randomly projected polytopes when the projection radically lowers dimensions,” *J. Amer. Math. Soc.*, vol. 22, no. 1, pp. 1–53, 2009.
- [5] S. Fletcher, A.K. Rangan and V. Goyal, “On-off random access channels: A compressed sensing framework,” preprint, [online] arXiv:0903.1022v2 [cs.IT], 2009.
- [6] E. Candes and T. Tao, “Decoding by linear programming,” *IEEE Trans. Info. Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.
- [7] Y. Kabashima, T. Wadayama, and T. Tanaka, “A typical reconstruction limit for compressed sensing based on lp-norm minimization,” *J. Stat. Mech.*, vol. 2009, no. 09, p. L09003, 2009.
- [8] S. Rangan, A. Fletcher, and V. Goyal, “Asymptotic analysis of map estimation via the replica method and applications to compressed sensing,” preprint, [online] arXiv:0906.3234v2 [cs.IT], 2009.
- [9] H. Nishimori, *Statistical Physics of Spin Glasses and Information Processing*. Oxford, UK: Oxford Science Publications, 2001.
- [10] E. J. Candes, M. Wakin, and S. Boyd, “Enhancing sparsity by reweighted ℓ_1 minimization,” *J. Fourier Anal. Appl.*, vol. 14, nos. 5-6, pp. 877–905, 2007.
- [11] V. Cevher, “Learning with compressible priors,” in Y. Bengio et al. (eds.), *Advances in Neural Information Processing Systems 22*, 2009, pp. 261–269.
- [12] D. Wipf and S. Nagarajan, “Sparse estimation using general likelihoods and non-factorial priors,” in Y. Bengio et al. (eds.), *Advances in Neural Information Processing Systems 22*, 2009, pp. 2071–2079.
- [13] D. Wipf and S. Nagarajan, “Iterative reweighted ℓ_1 and ℓ_2 methods for finding sparse solutions,” *IEEE J. Select. Topics Signal Process.*, vol. 4, no. 2, pp. 317–329, 2010.
- [14] T. Tanaka, “A statistical-mechanics approach to large-system analysis of CDMA multiuser detectors,” *IEEE Trans. Info. Theory*, vol. 48, no. 11, pp. 2888–2910, 2002.
- [15] D. Guo and T. Tanaka, “Generic multiuser detection and statistical physics,” in M. L. Honig (ed.), *Advances in Multiuser Detection*, Wiley, 2009, Chapter 5.
- [16] J. d. Almeida and D. Thouless, “Stability of the Sherrington-Kirkpatrick solution of a spin glass model,” *J. Phys. A*, vol. 11, no. 5, pp. 983–990, 1978.