Statistical Mechanics of Compressed Sensing

Surya Ganguli*
Sloan Swartz Center for Theoretical Neurobiology, UCSF, San Francsico, California 94143, USA

Haim Sompolinsky†
Interdisciplinary Center for Neural Computation, Hebrew University, Jerusalem 91904, Israel and Center for Brain Science, Harvard University, Cambridge, Massachusetts 02138, USA

(Received 10 December 2009; published 7 May 2010)

Compressed sensing (CS) is an important recent advance that shows how to reconstruct sparse high dimensional signals from surprisingly small numbers of random measurements. The nonlinear nature of the reconstruction process poses a challenge to understanding the performance of CS. We employ techniques from the statistical physics of disordered systems to compute the typical behavior of CS as a function of the signal sparsity and measurement density. We find surprising and useful regularities in the nature of errors made by CS, a new phase transition which reveals the possibility of CS for nonnegative signals without optimization, and a new null model for sparse regression.

DOI: 10.1103/PhysRevLett.104.188701 PACS numbers: 89.20.−a, 05.70.Fh, 89.75.−k

The recovery of high dimensional signals from limited numbers of measurements is central to many fields, from signal processing and machine learning, to neuroscience and systems biology. For example, one can speed up functional magnetic resonance imaging (fMRI) by minimizing the number of Fourier measurements needed to reconstruct each image frame. Also, using microarrays, one would like to infer regulatory interactions among genes by measuring expression levels of thousands of genes at once, but only under a limited number of experimental conditions, i.e., $O(100)$. And, in neural systems, a high dimensional afferent signal is often compressed by a downstream system, as is the case with the output of the retina, motivating the study of efficient methods that might be used in the brain to recover the original signal.

The new field of compressed sensing (CS), which has attracted much attention recently [1,2], demonstrates that if the high dimensional signal to be recovered is sparse, one can accurately recover the signal from a number of measurements much smaller than the dimensionality of the signal. Indeed many high dimensional signals of interest are sparse. For example, natural images, human speech, and fMRI data, when presented in a basis of wavelet coefficients, typically only have significant power in a very small fraction of the coefficients. Similarly, of the thousands of genes measured in a microarray, only a very small fraction actually regulate a gene of interest.

More formally, let $\mathbf{x}^0$ be an unknown sparse T dimensional vector which has only a fraction f of its elements nonzero. Suppose we are given a vector $\mathbf{y}$ of $N < T$ measurements, which is linearly related to $\mathbf{x}^0$ by an N by T measurement matrix $\mathbf{A}$, i.e., $\mathbf{y} = \mathbf{A}\mathbf{x}^0$. Thus each measurement y_μ, for $\mu = 1, \ldots, N$ is a linear function $\mathbf{a}_\mu \cdot \mathbf{x}^0$ of the unknown signal $\mathbf{x}^0$, where $\mathbf{a}_\mu$ is the μ'th row of $\mathbf{A}$. In the context of signal processing, $\mathbf{x}^0$ could be a temporal signal, and the $\mathbf{a}_\mu$ could be a set of N temporal filters. In the context of genetic network inference, $\mathbf{x}^0$ could be a vector of regulatory interactions governing the linear response of a single gene y_u to a pattern of expression of T genes $\mathbf{a}_\mu$, all measured in experimental condition μ.

One approach to CS involves recovering $\mathbf{x}^0$ by solving

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \sum_{i=1}^{T} V(x_i) \quad \text{subject to } \mathbf{y} = \mathbf{A}\mathbf{x}, \qquad (1)$$

where $V(x)$ is any sparsity promoting function. A natural choice is $V(x) = 1$ if $x = 0$ and $V(x) = 0$ otherwise, so that (1) yields the $\mathbf{x}$ consistent with the measurements $\mathbf{y}$ with the minimum number of nonzero elements. However, this is in general a hard combinatorial optimization problem. Here we consider the choice $V(x) = |x|$ thereby minimizing the L_1 norm of $\mathbf{x}$, which is a convex problem.

Much of the seminal theoretical work in CS [3–5] has focused on sufficient conditions on $\mathbf{A}$ to guarantee perfect signal recovery, so that $\hat{\mathbf{x}} = \mathbf{x}^0$ in (1), in the case of L_1 minimization. But often, large random measurement matrices $\mathbf{A}$ which violate these sufficient conditions nevertheless typically yield good signal reconstruction performance. Important work on this typical behavior [5,6] has revealed a phase transition in the performance of CS as a function of the signal sparsity, controlled by f, and the degree of sampling, $\alpha = N/T$, for large random matrices $\mathbf{A}$, using the methods of convex geometry. Here we take an alternate approach to understanding the typical behavior of CS using replica theory [7,8]. This approach yields new features in the phase diagram for CS and new insights into the nature of CS in the error regime.

Statistical mechanical framework.—We wish to understand properties of the solution $\hat{\mathbf{x}}$ to the optimization problem in (1). To do so, we define an energy function on the residual $\mathbf{u} = \mathbf{x} - \mathbf{x}^0$ given by

$$E(\mathbf{u}) = \frac{\lambda}{2T}\mathbf{u}^T\mathbf{A}^T\mathbf{A}\mathbf{u} + \sum_{i=1}^{T}|u_i + x_i^0|, \quad (2)$$

and analyze the statistical mechanics of the Gibbs distribution $P_G(\mathbf{u}) = \frac{1}{Z}e^{-\beta E(\mathbf{u})}$. We take the limit $\lambda \to \infty$ to enforce the constraint $\mathbf{y} = \mathbf{A}\mathbf{x}$, and then take $\beta \to \infty$ which yields information about the solution to (1).

The free energy $-\beta F = \ln Z$ of P_G depends on the measurement matrix $\mathbf{A}$ and on the signal $\mathbf{x}^0$. We take these to be random variables; the matrix elements $A_{\mu i}$ are drawn independently from a standard normal distribution, while $\mathbf{x}^0$ has fT randomly chosen nonzero elements each drawn independently from a distribution $P(x^0)$. Thus $\mathbf{A}$ and $\mathbf{x}^0$ play the role of quenched disorder in the thermal distribution P_G. For large N and T, we expect self-averaging, i.e., the properties of P_G for any typical realization of $\mathbf{A}$ and $\mathbf{x}^0$ coincide with the properties of P_G averaged over $\mathbf{A}$ and $\mathbf{x}^0$. Therefore we compute the average free energy $-\beta\bar{F} \equiv \langle\langle \ln Z\rangle\rangle_{\mathbf{A},\mathbf{x}^0}$ using the replica method [8,9].

We find [10] the u_i's are distributed as $P(u) \propto \exp(-H_{\text{eff}})$, with the mean-field effective Hamiltonian,

$$H_{\text{eff}} = \frac{\alpha\beta\lambda}{2(1+\beta\lambda\Delta Q)}(u - z\sqrt{Q_0/\alpha})^2 + \beta|u + x^0|, \quad (3)$$

and z is a quenched random variable with a standard normal distribution. The order parameters obey

$$Q_0 = \langle\langle\langle u\rangle^2\rangle\rangle_{z,x^0}, \quad (4)$$

$$\Delta Q = \langle\langle\langle \delta u^2\rangle\rangle\rangle_{z,x^0}, \quad (5)$$

where the thermal average of u and its variance $\langle\delta u^2\rangle$ are defined with respect to the Gibbs distribution with Hamiltonian H_{eff}. The double angular average $\langle\langle\cdot\rangle\rangle_{z,x^0}$ refers to integrating over the Gaussian distribution of z and over the distribution of x^0.

Perfect reconstruction regimes.—Here we search for solutions to (4) and (5) which reflect perfect reconstruction performance in (1). First, for any finite β, we take $\lambda \to \infty$, eliminating λ from H_{eff}. Next, as $\beta \to \infty$, we search for solutions in which $\Delta Q = \alpha\Delta q/\beta^2$, and $Q_0 = \alpha q_0/\beta^2$ where both Δq and q_0 are $O(1)$. This scaling ensures that as the temperature is lowered, the Gibbs measure $P_G(\mathbf{u})$ concentrates in a region of size $O(1/\beta)$ containing the residual vector $\mathbf{u} = 0$, i.e., $\hat{\mathbf{x}} = \mathbf{x}^0$. Using this scaling, we find that (4) and (5) admit finite solutions for any f and $\alpha > \alpha_c(f)$. As α decreases towards the phase boundary $\alpha_c(f)$, Δq and q_0 diverge, signaling a transition to the error regime, with $\Delta q/q_0^2 \equiv x^2$, remaining $O(1)$. In this limit, $\alpha_c(f)$ is given by the solution to

$$\alpha = 2(1-f)\int_x^\infty \mathcal{D}z(z-x)^2 + f(1+x^2), \quad (6)$$

$$\alpha = 2(1-f)H(x) + f. \quad (7)$$

Here $H(x) \equiv \int_x^\infty \mathcal{D}z$, and $\mathcal{D}z$ is the standard Gaussian measure. Note that the phase boundary does not depend on the distribution of nonzeros, $P(x^0)$. As shown in Fig. 1(a), $\alpha_c(f)$ predicts well the transition between perfect and imperfect reconstruction in (1). In the sparse limit, $f \to 0$, $\alpha_c(f) \to f\log 1/f$.

Nonnegative signal reconstruction: compressed sensing without sparseness optimization.—An interesting case is when the signal is known to be constrained in its sign. Such a constraint modifies the phase boundary of the regime of perfect reconstruction. Following the same analysis as above, we derived the phase boundary for this case [10] and the result is shown in Fig. 1(b).

Can a nonnegative signal be fully reconstructed from a limited number of measurements without optimizing sparsity? To address this, we reexamine the statistical mechanics of (3) as $\beta \to \infty$ but without the L_1 potential term, or equivalently keeping β finite and taking $\lambda \to \infty$. Then the relevant scaling of the order parameters is $\Delta Q = \Delta q/\beta\lambda$ and $Q_0 = q_0/\beta\lambda$. This scaling implies a solution with zero error as the linear equality constraints, together with the nonnegativity constraints, are strictly enforced. The resulting order parameter equations [10] admit finite solutions for $\alpha > \frac{1+f}{2}$. As α approaches this lower bound, both q_0 and Δq diverge, marking the phase transition to a nonzero error. Below this value of α the solution to the constraints on $\mathbf{x}$ is nonunique and sparsity optimization is required for perfect reconstruction, which is possible at lower values of α until the lower phase boundary in Fig. 1(b) is reached.

The above remarkable results can also be derived using the theory of perceptron capacity [11,12] by the following argument. Let $\mathbf{C}$ be a $T \times (T-N)$ matrix whose columns form a basis for the null space of $\mathbf{A}$. A nonnegative solution $\mathbf{x}$ to the linear equations $\mathbf{y} = \mathbf{A}\mathbf{x}^0 = \mathbf{A}\mathbf{x}$ is equivalent to a solution $\mathbf{v}$ of the set of inequalities $\mathbf{C}\mathbf{v} \geq -\mathbf{x}^0$. Let us focus on the $T(1-f)$ inequalities for which $x_i^0 = 0$. From perceptron theory, with high probability there is no nonzero solution to these inequalities if the number of inequalities $T(1-f)$ exceeds twice the number of unknowns $T-N$, which yields the bound $\alpha \geq (1+f)/2$ for $\mathbf{x}^0$ to be the

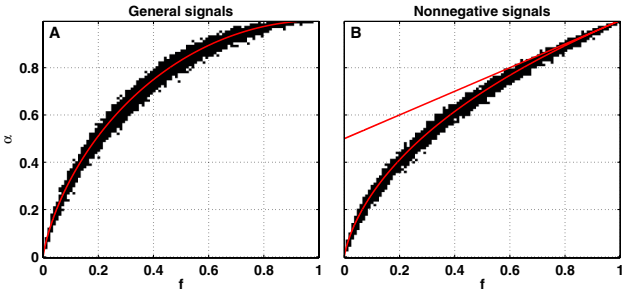

FIG. 1 (color online). (a) We use linear programming to solve (1) 50 times for each value of α and f in increments of 0.01, with $T = 500$. The black transition region shows when the fraction of times perfect recovery occurs is neither 0 nor 1. The red curve is the theoretical phase boundary $\alpha_c(f)$ obtained by solving (6) and (7). (b) Similar to (a) except now $\mathbf{x}^0$ and $\mathbf{x}$ are constrained to be nonnegative. The additional red line is $\alpha = (1+f)/2$, above which the solution space for $\mathbf{x}$ condenses to a single point, $\mathbf{x}^0$.

unique nonnegative solution to $\mathbf{y} = \mathbf{A}\mathbf{x}$. To show that this is also a tight bound, note that once there is a solution to the set of inequalities with zero x_i^0, this solution can be multiplied by a positive scaling factor, which can always be chosen to satisfy the set of inequalities $\mathbf{C}\mathbf{u} \geq -\mathbf{x}^0$ with positive x_i^0.

The error regime.—Here we search for solutions to (4) and (5) which describe the performance of CS based on (1) in the error regime. As above, we first take $\lambda \to \infty$, but now, as $\beta \to \infty$, the order parameters scale as $\Delta Q = \alpha \Delta q/\beta$ and $Q_0 = \alpha q_0$ where both Δq and q_0 are $O(1)$. Under this ansatz, in the case of nonnegativity constraints on $\mathbf{x}$, we obtain (see [10])

$$\alpha = (1-f)G(z^0, z^0) + f\langle\langle G(z^+, z^0) + \frac{(x^0)^2}{q_0} H(-z^+)\rangle\rangle_{x^0}$$

$$\alpha = (1-f)H(z^0) + f\langle\langle H(z^+)\rangle\rangle_{x^0}.$$

Here, $z^0 = \frac{\Delta q}{\sqrt{q_0}}$, $z^\pm = \frac{-x^0 \pm \Delta q}{\sqrt{q_0}}$, and $G(a,b) = \int_a^\infty \mathcal{D}z(z-b)^2$. The corresponding equations for general $\mathbf{x}$ can be found in [10]. Note that unlike the zero error equations, the equations in the error phase depend on the distribution of nonzeros $P(x^0)$. To explore this dependence, in the following we focus on 3 classes of signals: 01 signals in which $P(x^0) = \delta(x^0 - 1)$, plus-minus (PM) signals in which $P(x^0) = \frac{1}{2}\delta(x-1) + \frac{1}{2}\delta(x+1)$, and Gaussian signals in which $P(x^0)$ is the standard normal distribution.

We solved the error equations numerically and found q_0 and Δq as a function of f and α. The predicted reconstruction error is $Q_0 = \alpha q_0$. As an example, the agreement of the prediction with direct numerical optimization of (1) is shown for 2 types of signals in Figs. 2(a) and 2(b).

The rise of the error near the phase transition depends on the distribution of nonzeros $P(x^0)$ only through its behavior near the origin [10]. Let $\delta\alpha = \alpha_c - \alpha$ be the distance into the error phase. If $P(x^0)$ has a gap so that $P(|x^0|) = 0$ when $|x^0| < \Delta$ for some $\Delta > 0$, then for small $\delta\alpha$ the error rises sharply as $O(1/\log\frac{1}{\delta\alpha})$ [Fig. 2(a)]. Otherwise, if $P(x^0) \propto (x^0)^\nu$ near the origin, with $\nu > -1$, then the error rises as $O(\delta\alpha^{2/(1+\nu)})$. See Fig. 2(b) for an example of Gaussian signals where $\nu = 0$. Overall, the more strongly the nonzeros are confined to the origin, the shallower the rise of the error after the transition.

The error also displays an interesting nonmonotonic dependence for small α. In the absence of measurement constraints, when $\alpha = 0$, $\hat{\mathbf{x}} = \mathbf{0}$ and the error is f [in units of the second moment of $P(x^0)$]. A small number of measurements then leads to a sharp rise in the error, which behaves near $\alpha = 0$ as $f(1 + O(1/\log\frac{1}{\alpha}))$, independent of the distribution $P(x^0)$ [see Figs. 2(a) and 2(b)]. The error then falls below f only at larger values of α.

The replica method allows us to compute not only the typical reconstruction error, but also the distribution of a reconstructed signal component $\hat{x}_i$ conditioned on the true value x_i^0. For general signals, we find [10] $P(\hat{x}_i|x_i^0) = \frac{1}{\sqrt{2\pi q_0}} \exp(-\frac{(\hat{x}_i - x_i^0 + \mathrm{sgn}(\hat{x}_i)\Delta q)^2}{2q_0}) + (H(z^-) - H(z^+))\delta(x)$.

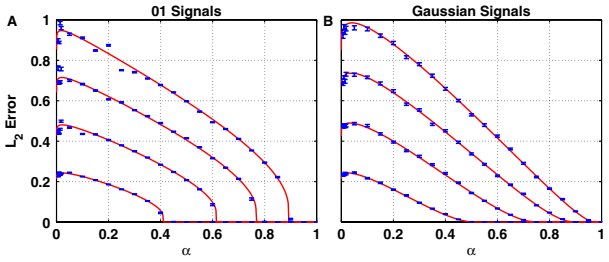

FIG. 2 (color online). (a) and (b) Blue points are the average L_2 reconstruction error obtained by solving (1) 100 times for each for 4 values of $f = 0.2, 0.4, 0.6, 0.8$, and various α. Here $T = 500$. Error bars reflect the standard error. Red curves are plots of Q_0 obtained by solving the appropriate order parameter equations in the error phase.

From this distribution, we can compute the typical L_p norm of the reconstructed signal $\hat{\mathbf{x}}$ in (1), defined as $L_p(\hat{\mathbf{x}}) \equiv \frac{1}{T}\sum_{i=1}^T |\hat{x}_i|^p$. The match between theory and numerics for the L_1 norm is shown in Fig. 3(a). Similar behavior is observed for the other signal ensembles. As expected, because the reconstruction algorithm is minimizing L_1, as the number of constraints α decreases, the L_1 norm monotonically decreases to 0. The onset of the error phase occurs when $\mathbf{x}^0$ is no longer the ground state of P_G, and $\mathbf{x}$ satisfying $\mathbf{y} = \mathbf{A}\mathbf{x}$ can be found with L_1 norm smaller than that of $\mathbf{x}^0$, which is f.

The original motivation behind CS was to minimize the L_0 norm in (1). L_1 minimization is used instead as a practical, convex alternative to the hard L_0 problem. We can nevertheless ask, how $L_0(\hat{\mathbf{x}})$ behaves under L_1 minimization. The theory gives a remarkably simple, universal prediction for this quantity that is independent of $P(x^0)$: in the perfect reconstruction phase, $L_0(\hat{\mathbf{x}}) = f$ independent of α, and in the error phase, $L_0(\hat{\mathbf{x}}) = \alpha$ independent of f [10]. Between these two phases, $L_0(\hat{\mathbf{x}})$ displays a sharp first order transition [Fig. 3(b)].

We give a direct argument that $L_0(\hat{\mathbf{x}})$ cannot exceed α for a typical realization of $\mathbf{A}$. Consider the Lagrangian $\mathcal{L}(\lambda, \mathbf{x}) = \sum_{\mu=1}^N \lambda_\mu (y_\mu - (\mathbf{A}\mathbf{x})_\mu) + \sum_{i=1}^T |x_i|$, which implements the constrained optimization (1). Let $(\hat{\lambda}, \hat{\mathbf{x}})$ be a critical point of this Lagrangian, and let i be an index such that $\hat{x}_i \neq 0$. A necessary condition for $(\hat{\lambda}, \hat{\mathbf{x}})$ to be a critical

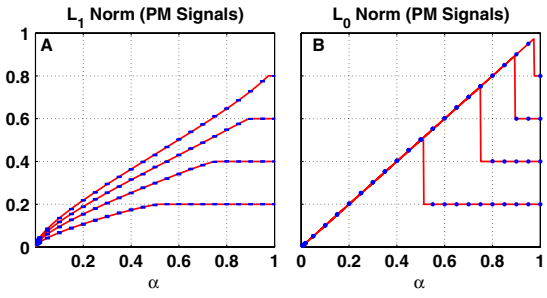

FIG. 3 (color online). (a) and (b) The blue points represent the average L_1 and L_0 norms of the reconstruction $\hat{\mathbf{x}}$ obtained by solving (1), as in Fig. 2, while the red curves are the theoretical predictions.

point is $\frac{\partial \mathcal{L}}{\partial x_i}|_{\hat{\mathbf{x}}} = 0 \Rightarrow \sum_{\mu=1}^{N} \hat{\lambda}_\mu A_{\mu i} = \mathrm{sgn}(\hat{x}_i)$. There are $L_0(\hat{\mathbf{x}})T$ such linear constraints on the $N = \alpha T$ Lagrange multipliers. Thus if $L_0(\hat{\mathbf{x}}) > \alpha$, these equations for generic $\mathbf{A}$ will be overconstrained and have no solution. Therefore any L_1 minimizer $\hat{x}$ will with high probability have $L_0(\hat{\mathbf{x}}) \leq \alpha$. The surprising result then is that as soon as the error phase is entered, $L_0(\hat{\mathbf{x}})$ makes a discontinuous first order jump to this maximal value despite the fact that the L_1 norm and L_2 reconstruction errors display a continuous second order transition.

Sparse regression.—Until now we have focused on the problem of signal reconstruction, in which $\mathbf{A}$ is a known measurement matrix acting on an unknown sparse signal $\mathbf{x}^0$ to yield a set of measurements $\mathbf{y}$, from which $\mathbf{x}^0$ must be recovered. An alternate scenario is that of regression, in which $\{y_\mu, \mathbf{a}_\mu\}$ constitute a set of N data points (here $\mathbf{a}_\mu$ are the rows of $\mathbf{A}$) and the goal is to find an unknown sparse vector of regression coefficients, or rule, $\mathbf{x}^0$ that explains the data, i.e., $y_\mu = \mathbf{a}_\mu \cdot \mathbf{x}^0 \ \forall \ \mu$. In applications, the true level of sparseness of $\mathbf{x}^0$ is unknown, and so if (1) were applied to find a candidate rule $\hat{\mathbf{x}}$, it is important to know whether the resulting sparsity of $\hat{\mathbf{x}}$ reflects the true nature of the data $\{y_\mu, \mathbf{a}_\mu\}$ as arising from a sparse rule, or whether this sparsity arises simply due to chance.

To address this question, here we analyze (1) in a "null" scenario where $\mathbf{y}$ is generated by a random process, independent of $\mathbf{A}$. The relevant energy function is

$$E(\mathbf{x}) = \frac{\lambda}{2T}(\mathbf{y} - \mathbf{A}\mathbf{x})^T(\mathbf{y} - \mathbf{A}\mathbf{x}) + \sum_{i=1}^{T} |x_i|, \qquad (8)$$

with $\mathbf{A}$ a random $N \times T$ matrix as before, and y_μ drawn independently from a standard normal. We used replica theory to analyze the statistical mechanics of (8) in the λ, $\beta \to \infty$ limit, thereby computing theoretical predictions for various L_p norms of $\hat{\mathbf{x}}$ in (1) for typical realizations of $\mathbf{A}$ and $\mathbf{y}$ (see [10] for details). Results are shown in Fig. 4. In particular, Fig. 4 allows one to test the significance of a given level of sparsity of $\hat{\mathbf{x}}$ obtained from (1) in any application of CS to regression. If either the L_1 or L_0 norm of $\hat{\mathbf{x}}$ falls significantly below the corresponding curve at the appropriate value of α (usually known in applications), then it is highly likely that the data can be explained by a sparse rule. Interestingly, the L_0 norm takes its maximal possible value α, just as in the case of CS applied to signal reconstruction in the error regime [see Fig. 3(b)].

Discussion.—In summary, by exploiting statistical mechanics, we have obtained a deeper understanding of the behavior of CS. In particular, our finding of a universal, first order transition in the L_0 norm of the reconstruction [Fig. 3(b)] has important practical implications. It provides a simple rule to detect perfect recovery in CS even when $\mathbf{x}^0$ is unknown: if the number of nonzeros in $\hat{\mathbf{x}}$ in (1) is less than the number of measurements, then $\hat{\mathbf{x}}$ is with high probability the true signal $\mathbf{x}^0$. An extension of this work is the analysis of L_p minimization, where $V(x) = |x|^p$. We have analyzed the replica symmetric theory and shown that it predicts perfect reconstruction for all $\alpha > f$ and $0 \leq p < 1$ [10]. However, numerically we have shown that while $\mathbf{x}^0$ is a local minimum of L_p in this regime, it is not a global minimum [10], indicating that replica symmetry breaking is required for a correct analysis of L_p minimization. Overall, given the promise of CS [1,2], we believe a statistical mechanics approach to CS will be of broad relevance to a variety of fields, from the design and analysis of algorithms for rapid signal acquisition to the interpretation of high dimensional biological data.

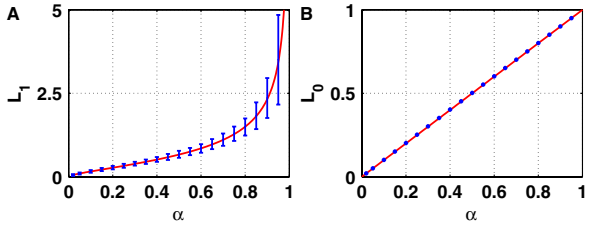

FIG. 4 (color online). Here (1) is solved 100 times for various α when $\mathbf{y}$ is a dense random Gaussian vector uncorrelated with $\mathbf{A}$, and $T = 500$. The average L_1 (a) and (b) L_0 norms of $\hat{\mathbf{x}}$ are shown (blue points). Red curves are the theoretical predictions.

S. G. and H. S. thank the Swartz, Burroughs Wellcome, and Israeli Science foundation for support, and Daniel Lee for useful discussions.

Note added.—Recently, we became aware of [13] which partially overlaps with our work.

*surya@phy.ucsf.edu
†haim@fiz.huji.ac.il

[1] A. Bruckstein, D. Donoho, and M. Elad, SIAM Rev. **51**, 34 (2009).
[2] E. Candes and M. Wakin, IEEE Signal Process. Mag. **25**, 21 (2008).
[3] D. Donoho and M. Elad, Proc. Natl. Acad. Sci. U.S.A. **100**, 2197 (2003).
[4] E. Candes, J. Romberg, and T. Tao, IEEE Trans. Inf. Theory **52**, 489 (2006).
[5] E. Candes and T. Tao, IEEE Trans. Inf. Theory **51**, 4203 (2005).
[6] D. Donoho and J. Tanner, Proc. Natl. Acad. Sci. U.S.A. **102**, 9446 (2005).
[7] D. Donoho and J. Tanner, Proc. Natl. Acad. Sci. U.S.A. **102**, 9452 (2005).
[8] M. Mezard, G. Parisi, and M. Virasoro, *Spin Glass Theory and Beyond* (World Scientific, Singapore, 1987).
[9] A. Engel and C. V. den Broeck, *Statistical Mechanics of Learning* (Cambridge Univ. Press, Cambridge, 2001).
[10] See supplementary material at http://link.aps.org/supplemental/10.1103/PhysRevLett.104.188701 for derivations of results.
[11] T. Cover, IEEE Trans. Electron. Comput. **14**, 326 (1965).
[12] E. Gardner, J. Phys. A **21**, 257 (1988).
[13] Y. Kabashima, T. Wadayama, and T. Tanaka, J. Stat. Mech. (2009) L09003.