

# The phase transition of matrix recovery from Gaussian measurements matches the minimax MSE of matrix denoising

David L. Donoho<sup>a,1</sup>, Matan Gavish<sup>a</sup>, and Andrea Montanari<sup>a,b</sup>

Departments of <sup>a</sup>Statistics and <sup>b</sup>Electrical Engineering, Stanford University, Stanford, CA 94305

Contributed by David L. Donoho, April 3, 2013 (sent for review February 5, 2013)

Let  $X_0$  be an unknown  $M$  by  $N$  matrix. In matrix recovery, one takes  $n < MN$  linear measurements  $y_1, \dots, y_n$  of  $X_0$ , where  $y_i = \text{Tr}(A_i^T X_0)$  and each  $A_i$  is an  $M$  by  $N$  matrix. A popular approach for matrix recovery is nuclear norm minimization (NNM): solving the convex optimization problem  $\min \|X\|_*$  subject to  $y_i = \text{Tr}(A_i^T X)$  for all  $1 \leq i \leq n$ , where  $\|\cdot\|_*$  denotes the nuclear norm, namely, the sum of singular values. Empirical work reveals a phase transition curve, stated in terms of the undersampling fraction  $\delta(n, M, N) = n/(MN)$ , rank fraction  $\rho = \text{rank}(X_0)/\min\{M, N\}$ , and aspect ratio  $\beta = M/N$ . Specifically when the measurement matrices  $A_i$  have independent standard Gaussian random entries, a curve  $\delta^*(\rho) = \delta^*(\rho; \beta)$  exists such that, if  $\delta > \delta^*(\rho)$ , NNM typically succeeds for large  $M, N$ , whereas if  $\delta < \delta^*(\rho)$ , it typically fails. An apparently quite different problem is matrix denoising in Gaussian noise, in which an unknown  $M$  by  $N$  matrix  $X_0$  is to be estimated based on direct noisy measurements  $Y = X_0 + Z$ , where the matrix  $Z$  has independent and identically distributed Gaussian entries. A popular matrix denoising scheme solves the unconstrained optimization problem  $\min \|Y - X\|_F^2/2 + \lambda \|X\|_*$ . When optimally tuned, this scheme achieves the asymptotic minimax mean-squared error  $\mathcal{M}(\rho; \beta) = \lim_{M, N \rightarrow \infty} \inf_x \sup_{\text{rank}(X) \leq \rho \cdot M} \text{MSE}(X, \hat{X}_\lambda)$ , where  $M/N \rightarrow \beta$ . We report extensive experiments showing that the phase transition  $\delta^*(\rho)$  in the first problem, matrix recovery from Gaussian measurements, coincides with the minimax risk curve  $\mathcal{M}(\rho) = \mathcal{M}(\rho; \beta)$  in the second problem, matrix denoising in Gaussian noise:  $\delta^*(\rho) = \mathcal{M}(\rho)$ , for any rank fraction  $0 < \rho < 1$  (at each common aspect ratio  $\beta$ ). Our experiments considered matrices belonging to two constraint classes: real  $M$  by  $N$  matrices, of various ranks and aspect ratios, and real symmetric positive-semidefinite  $N$  by  $N$  matrices, of various ranks.

compressed sensing | matrix completion

Let  $X_0$  be an unknown  $M$  by  $N$  matrix. How many measurements must we obtain to “completely know”  $X_0$ ? Although it seems that  $MN$  measurements must be necessary, in recent years intense research in applied mathematics, optimization, and information theory has shown that, when  $X_0$  is of low rank, we may efficiently recover it from a relatively small number of linear measurements by convex optimization (1–3). Applications have been developed in fields ranging widely, for example from video and image processing (4) to quantum state tomography (5) to collaborative filtering (2, 6).

Specifically, let  $\mathcal{A} : \mathbb{R}^{M \times N} \rightarrow \mathbb{R}^n$  be a linear operator and consider measurements  $y = \mathcal{A}(X_0)$ . If  $n < MN$ , the problem of inferring  $X_0$  from  $y$  may be viewed as attempting to solve an underdetermined system of equations. Under certain circumstances, it has been observed that this (seemingly hopeless) task can be accomplished by solving the so-called nuclear norm minimization problem:

$$(P_{nuc}) \min \|X\|_* \text{ subject to } y = \mathcal{A}(X). \quad [1]$$

Here, the nuclear norm  $\|X\|_*$  is the sum of singular values of  $X$ . For example, it was found that if  $X_0$  is sufficiently low rank, with a principal subspace in a certain sense incoherent to the

measurement operator  $\mathcal{A}$ , then the solution  $X_1 = X_1(y)$  to  $(P_{nuc})$  is precisely  $X_0$ . Such incoherence can be obtained by letting  $\mathcal{A}$  be random, for instance if  $\mathcal{A}(X_0)_i = \text{Tr}(A_i^T X_0)$  with  $A_i \in \mathbb{R}^{m \times n}$  having independent and identically distributed (i.i.d) Gaussian entries. In this case we speak of “matrix recovery from Gaussian measurements” (1).

A key phrase from the previous paragraph is “if  $X_0$  is sufficiently low rank.” Clearly, there must be a quantitative tradeoff between the rank of  $X_0$  and the number of measurements required, such that higher-rank matrices require more measurements. In the Gaussian measurements model, with  $N, M$  sufficiently large, empirical work by Recht et al. (1, 7, 8), Tanner and Wei (9), and Oymak and Hassibi (10) documents a phase transition phenomenon. For matrices of a given rank, there is a fairly precise number of required samples, in the sense that a transition from nonrecovery to complete recovery takes place sharply as the number of samples varies across this value. For example, in Table S1 we present data from our own experiments, showing that, for reconstructing matrices of size 60 by 60 that are of rank 20, 2,600 Gaussian measurements are sufficient with very high probability, but 2,400 Gaussian measurements are insufficient with very high probability.

In this paper, we present a simple and explicit formula for the phase transition curve in matrix recovery from Gaussian measurements. The formula arises in an apparently unrelated problem: matrix denoising in Gaussian noise. In this problem, we again let  $X_0$  denote an  $M$  by  $N$  matrix, and we observe  $Y = X_0 + Z$ , where  $Z$  is Gaussian i.i.d noise  $Z_{ij} \sim \mathcal{N}(0, 1)$ . We consider the following nuclear norm denoising scheme:

$$(P_{nuc, \lambda}) \min \left\{ \frac{1}{2} \|Y - X\|_F^2 + \lambda \|X\|_* \right\}. \quad [2]$$

In this problem the measurements  $Y$  are direct, so in some sense complete, but noisy. The solution  $\hat{X}_\lambda(Y)$  can be viewed as a shrinkage estimator. In the basis of the singular vectors  $U_Y$  and  $V_Y$  of  $Y$ , the solution  $\hat{X}_\lambda(Y)$  is diagonal, and the diagonal entries are produced by soft thresholding of the singular values of  $Y$ .

Because the measurements  $y$  in the matrix recovery problem are noiseless but incomplete, whereas the measurements  $Y$  in the matrix denoising problem are complete but noisy, the problems seem quite different. Nevertheless, we show here that there is a deep connection between the two problems.

Let us quantify performance in the denoising problem by the minimax MSE, namely

Author contributions: D.L.D., M.G., and A.M. designed research; D.L.D. and M.G. performed research; D.L.D., M.G., and A.M. contributed new reagents/analytic tools; D.L.D., M.G., and A.M. analyzed data; and D.L.D., M.G., and A.M. wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

Data deposition: Data for this article have been deposited in the Stanford Digital Repository, <http://purl.stanford.edu/tz124hw0000>.

<sup>1</sup>To whom correspondence should be addressed. E-mail: donoho@stat.stanford.edu.

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1306110110/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1306110110/-DCSupplemental).

$$\mathcal{M}(\rho; M, N) = \min_{\lambda} \max_{\text{rank}(X) \leq \rho M} \text{MSE}(X_0, \hat{X}_\lambda(Y)),$$

where MSE refers to the dimension-normalized mean-squared error  $(MN)^{-1}E\|X - \hat{X}_\lambda\|_F^2$  and subscript  $F$  denotes Frobenius norm. The asymptotic minimax MSE  $\mathcal{M}(\rho; \beta) = \lim_{N \rightarrow \infty} \mathcal{M}(\rho; \beta N, N)$  has been derived in ref. 11. Explicit formulas for the curve  $\rho \mapsto \mathcal{M}(\rho; \beta)$  appear in *Appendix*. A parametric form is given for the case of asymptotically square matrices,  $\beta = 1$ . Fig. 1 depicts the various minimax MSE curves.

We can now state our main hypothesis for matrix recovery from Gaussian measurements.

**Main Hypothesis: Asymptotic Phase Transition Formula.** Consider a sequence of matrix recovery problems with parameters  $\{(r, n, M, N)\}_{N \geq 1}$  having limiting fractional rank  $\rho = \lim_{N \rightarrow \infty} r / \min\{M, N\}$ , limiting aspect ratio  $\beta = \lim_{N \rightarrow \infty} M/N$ , and limiting incompleteness fraction  $\delta = \lim_{N \rightarrow \infty} n/(MN)$ . In the limit of large problem size  $N$ , the solution  $X_1(y)$  to the nuclear norm minimization problem ( $P_{nuc}$ ) is correct with probability converging to one if  $\delta > \mathcal{M}(\rho; \beta)$  and incorrect with probability converging to one if  $\delta < \mathcal{M}(\rho; \beta)$ . In short, the asymptotic phase transition  $\delta^*(\rho; \beta)$  in Gaussian matrix recovery is equal to the asymptotic minimax MSE  $\mathcal{M}(\rho; \beta)$ .

In particular, for the case of small rank  $r$  and large  $N, M$ , by studying the small  $\rho$  asymptotics of Eq. 14, we obtain that reconstruction is possible from  $n \geq 2r(M + N + \sqrt{MN})(1 + O(r/N))$  measurements, but not from substantially less.

This brief announcement tests this hypothesis by conducting a substantial computer experiment generating large numbers of random problem instances. We use statistical methods to check for disagreement between the hypothesis and the predicted phase transition. To bolster the solidity of our results, we conduct the experiment in two different settings: (i) the matrix  $X_0$  is a general  $M$  by  $N$  matrix, for various rank fractions  $\rho$  and aspect ratios  $\beta$  and (ii)  $X_0$  is a symmetric positive-semidefinite matrix, for various rank fractions  $\rho$ . In the latter case the positive-semidefinite constraint is added to the convex program ( $P_{nuc}$ ). As described below, there are different asymptotic MSE curves for the two settings. We demonstrate an empirically accurate match in each of the cases, showing the depth and significance of the connection we expose here.

In the discussion and conclusion we connect our result with related work in the field of sparsity-promoting reconstructions,

where the same formal identity between a minimax MSE and a phase transition boundary has been observed, and in some cases even proved. We also discuss recent rigorous evidence toward establishing the above matrix recovery phase transition formula.

## Methods

We investigated the hypothesis that the asymptotic phase transition boundary agrees with the proposed phase transition formula to within experimental error.

For notational simplicity we will focus here on the case  $M = N$ , and defer the case of nonsquare matrices to *Supporting Information*. Hence, we will drop throughout the main text the argument  $\beta = 1$ . The asymptotic phase plane at point  $(\rho, \delta)$  is associated to triples  $(r, n, N)$ , where  $\rho = r/N \in [0, 1]$  is the rank fraction, and  $\delta = \delta(n, N/X) = n/\text{dim}(X)$  is the undersampling ratio, where  $\text{dim}(X)$  is the dimension of the underlying collection of matrices  $X$ . We performed a sequence of experiments, one for each tuple, in which we generated random rank- $r$   $N$  by  $N$  matrices  $X_0 \in X$ , random measurement matrices  $A = A$  of size  $n \times N^2$ , and obtained random problem instances  $(y, A)$ . We then applied a convex optimization procedure, obtaining a putative reconstruction  $\hat{X} = \hat{X}(y, A)$ . We declared a reconstruction successful when the Frobenius norm of the error was smaller than a threshold. Our raw empirical observations consist of a count of empirical successes and sample sizes at a selected set of positions  $(\rho, \delta)$  and a selected set of problem sizes  $N$ . From these raw counts we produce fitted success probabilities  $\hat{\pi}(r, n, N, X)$ ; the finite- $N$  phase transition is the place where the true underlying probability of successful reconstruction takes the value 50%. We tested the hypothesis that the finite- $N$  transition was consistent with the proposed asymptotic phase transition formula.

This section discusses details of data generation and data analysis.

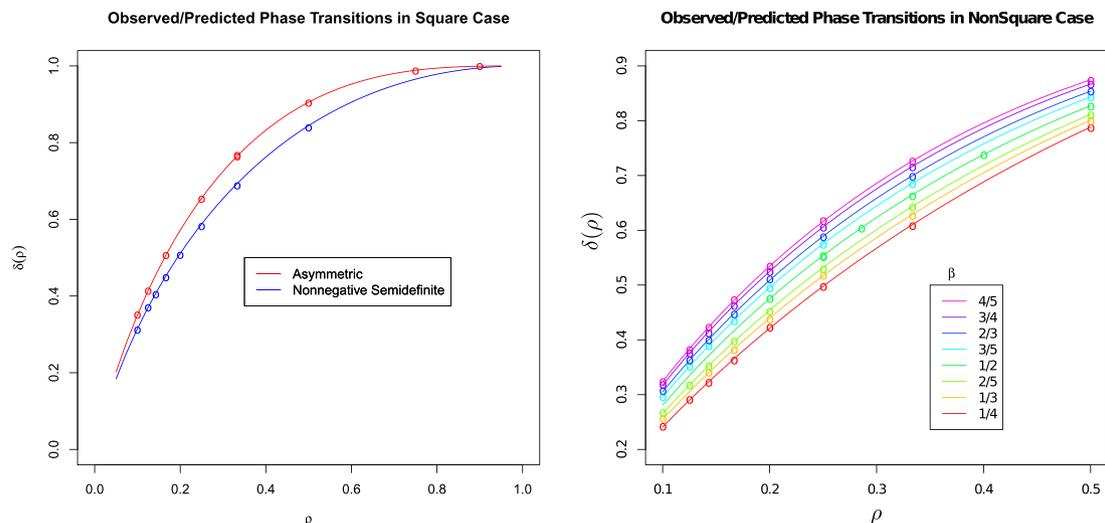
**Generation of Problem Instances.** Each problem instance  $(y, A)$  was generated by, first, generating a random rank- $r$  matrix  $X_0$ , then, generating a random measurement matrix  $A = A_{n, N^2}$  and then applying  $y = A \cdot \text{vec}(X_0)$ .

We considered problem instances of two specific types, corresponding to matrices  $X_0 \in X$ , with  $X$  one of two classes of matrices:

- $Mat_N$ : all  $N \times N$  matrices with real-valued entries and
- $Sym_N$ : all  $N \times N$  real symmetric matrices which are positive-semidefinite.

In the case  $X = Mat_N$ , we consider low-rank matrices  $X_0 = UV'$ , where  $U$  and  $V$  are each  $N$  by  $r$  partial orthogonal matrices uniformly distributed in the Stiefel manifold  $St(N, r)$ . In the case  $X = Sym_N$ , we consider low-rank matrices  $X_0 = UU'$ , where  $U$  is an  $N$  by  $r$  partial orthogonal matrix in  $St(N, r)$ , and again the matrix is uniformly distributed.

For measurement matrices  $A$ , we use Gaussian random matrices satisfying  $A_{i,j} \sim N(0, 1/n)$ .



**Fig. 1.** Observed/expected phase transitions. (Left) Square matrices; (Right) nonsquare matrices. Curves, asymptotic minimax MSEs  $\mathcal{M}(\rho|\beta, X)$ . Points, empirical phase transition locations for matrix recovery from incomplete measurements (Tables S2–S5). Curves in left panel, square matrices  $M = N$ ,  $\beta = 1$ :  $X = Mat_N$  (blue),  $X = Sym_N$  (red), as defined in text. Curves in right panel, nonsquare matrix settings  $X = Mat_{M,N}$ , varying aspect ratio factor  $\beta = M/N$ .

**Convex Optimization.** For a given problem instance  $(y, A)$ , we attempt to recover the underlying low-rank object  $X_0$  from the measurements  $y$  by convex optimization. Each of our choices  $X$  gives rise to an associated optimization problem:

$$(P_{nuc}^X) \min \|X\|_* \text{ subject to } y = A \cdot \text{vec}(X), \quad X \in \mathbf{X}.$$

Here,  $\mathbf{X}$  is one of the two classes of matrices,  $Mat_N$  or  $Sym_N$ . The two resulting optimization problems can each be reduced to a so-called semi-definite programming problem (see refs. 12, 13).

**Probability of Exact Recovery.** Because both the measurement matrix  $A$  and the underlying low-rank object  $X_0$  are random,  $(y, A)$  is a random instance for  $(P_{nuc}^X)$ . The probability of exact recovery is defined by

$$\pi(r|n, N, \mathbf{X}) = \text{Prob}\{X_0 \text{ is the unique solution of } (P_{nuc}^X)\}.$$

Clearly,  $0 \leq \pi \leq 1$ ; for fixed  $N$ ,  $\pi$  is monotone decreasing in  $r$  and monotone increasing in  $n$ . Also  $\pi(r|n, N, Mat_N) < \pi(r|n, N, Sym_N)$ .

**Estimating the Probability of Exact Recovery.** Our procedure follows refs. 14 and 15. For a given matrix type  $\mathbf{X}$  and rank  $r$  we conduct an experiment whose purpose is to estimate  $\pi(r|n, N, \mathbf{X})$  using  $T$  Monte Carlo trials. In each trial we generate a random instance  $(y, A)$ , which we supply to a solver for  $(P_{nuc}^X)$ , obtaining the result  $X_1$ . We compare the result  $X_1$  to  $X_0$ . If the relative error  $\|X_0 - X_1\|_F / \|X_0\|_F$  is smaller than a numerical tolerance, we declare the recovery a success; if not, we declare it a failure. (In this paper, we used an error tolerance of 0.001.) We thus obtain  $T$  binary measurements  $l_i$  indicating success or failure in reconstruction. The empirical success fraction is then calculated as

$$\hat{\pi}(r|n, N, T, \mathbf{X}) = \frac{\#\{\text{successes}\}}{\#\{\text{trials}\}} = \frac{1}{T} \sum_{i=1}^T l_i.$$

These are the raw observations generated by our experiments.

**Asymptotic Phase Transition Hypothesis.** Consider a sequence of tuples  $(r, n, N)$  with  $r/N \rightarrow \rho$  and  $n/N \rightarrow \delta$ . We assume that there is an asymptotic phase transition curve  $\delta^*(\rho|\mathbf{X})$ , i.e., a curve obeying

$$\pi(r|n, N, \mathbf{X}) \rightarrow \begin{cases} 1 & \delta < \delta^*(\rho|\mathbf{X}) \\ 0 & \delta > \delta^*(\rho|\mathbf{X}) \end{cases}. \quad [3]$$

For many convex optimization problems the existence of such an asymptotic phase transition is rigorously proven; see *Discussion: Existing Literature and Our Contributions*.

The hypothesis we investigate concerns the value of  $\delta^*(\rho|\mathbf{X})$ ; specifically, whether

$$\delta^*(\rho|\mathbf{X}) = \mathcal{M}(\rho|\mathbf{X}). \quad [4]$$

Here,  $\mathcal{M}(\rho|Mat)$ , respectively  $\mathcal{M}(\rho|Sym)$ , is the minimax MSE for Eq. 2 for general matrices (respectively, positive-semidefinite ones). Formulas for  $\mathcal{M}$  were derived by the authors in ref. 11; computational details are provided in *Appendix*.<sup>†</sup>

**Empirical Phase Transitions.** The empirical phase transition point is estimated by fitting a smooth function  $\hat{\pi}(n/N)$  (in fact a logistic function) to the empirical data  $\hat{\pi}(r|n, N, \mathbf{X})$ , where  $r$  is held fixed, using the `glm()` command in the R computing environment. In fact we fit the logistic model that  $\text{logit}(\hat{\pi}) \equiv \log\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right) = a + b\Delta$ , where  $\Delta(\delta|\rho) = \delta - \mathcal{M}(\rho)$  is the offset between  $\delta$  and the predicted phase transition. The coefficients  $a$  and  $b$  are called the intercept and slope, and will be tabulated below. The intercept gives the predicted logit exactly at  $\Delta=0$ , i.e.,  $\delta = \mathcal{M}(\rho)$ . The empirical phase transition is located at  $\hat{\delta}(r, N, M, \mathbf{X}) = \mathcal{M}(\rho|\mathbf{X}) - a/b$ . This is the value of  $\delta = \delta(n, N|\mathbf{X})$  solving

$$\hat{\pi}(\delta) = 1/2.$$

Under the hypothesis of Eq. 4 we have

<sup>†</sup>Previous empirical work (9) used the minimum degrees of freedom  $r(M+N-r)/MN$  rather than  $r$  as the x-coordinate for phase transition curves; however, for present purposes,  $r/\min(M, N)$  seems more natural, because of the way the formulas for  $\mathcal{M}(\rho)$  work out.

$$\lim_{N \rightarrow \infty} \lim_{r/N \rightarrow \rho} \lim_{T \rightarrow \infty} \hat{\delta}(r, N, T, \mathbf{X}) = \mathcal{M}(\rho|\mathbf{X}).$$

Consequently, in data analysis we will compare the fitted values  $\hat{\delta}(r, N, T, \mathbf{X})$  with  $\mathcal{M}(r/N|\mathbf{X})$ .<sup>‡</sup>

**Experimental Design.** To address our main hypothesis regarding the agreement of phase transition boundaries, we measure  $\hat{\pi}(r|n, N, \mathbf{X})$  at points  $\delta = n/N$  and  $\rho = r/N$  in the phase plane  $(\delta, \rho)$ , which we expect to be maximally informative about the location of the phase transition. In fact, the informative locations in binomial response models correspond to points where the probability of response is in the middle range (1/10, 9/10) (16). As a rough approximation to such an optimal design, we sample at equispaced  $\delta \in [\mathcal{M}(\rho|\mathbf{X}) - 0.05, \mathcal{M}(\rho|\mathbf{X}) + 0.05]$ .

**Computing.** We primarily used the MATLAB computing environment, and the popular CVX convex optimization package (17), a modeling system for disciplined convex programming by Boyd, Grant, and others supporting two open-source interior-point solvers: SeDuMi and SDPT3 (18, 19).

We also studied the robustness of our results across solvers. Zulfikar Ahmed translated our code into Python and used the general-purpose solver package CVXOPT by Anderson et al. (20).

## Results

Our experimental data have been deposited (21). The data are contained in a text file with more than 100,000 lines, each line reporting one batch of Monte Carlo experiments at a given  $r, n, M, N$ , and  $\mathbf{X}$ . Each line also documents the number of Monte Carlo trials  $T$ , and the observed success fraction  $\hat{\pi}$ . The file also contains metadata identifying the solver and the researcher responsible for the run.

In all cases, we observed a transition from no observed successes at  $\delta = \mathcal{M}(\rho) - 0.05$  to no observed failures at  $\delta = \mathcal{M}(\rho) + 0.05$ . Fig. 1, *Right* shows results we obtained at nonsquare matrix ensembles, with varying  $\beta = M/N$ . The minimax MSE curves  $\mathcal{M}(\rho|\beta, \mathbf{X})$  vary widely, but the observed phase transitions track the MSE curves closely.

Fig. 2 shows a small subset of our results in the square case  $\mathbf{X} = Mat_N$ , to explain our empirical results; the full tables are given in *Supporting Information* for the square, nonsquare, and symmetric positive-semidefinite cases. In the square case, the empirical phase transition agrees in all cases with the formula  $\mathcal{M}(\rho)$  to two-digit accuracy. Table S3 shows that, in the symmetric positive-semidefinite case  $\mathbf{X} = Sym_N$ , the empirical phase transition falls within  $[\mathcal{M}(\rho) - 0.01, \mathcal{M}(\rho) + 0.01]$ .

Previous empirical studies of phase transition behavior in sparse recovery show that, even in cases where the asymptotic phase transition curve is rigorously proven and analytically available, such large- $N$  theory cannot be expected to match empirical finite- $N$  data to within the usual naive SEs. Instead, one observes a finite transition zone of width  $\approx c_1/N^{1/2}$  and a small displacement of the empirical phase transition away from the asymptotic Gaussian phase transition, of size  $\approx c_2/N$  (14, 15). Hence, the strict literal device of testing the hypothesis that  $E\hat{\delta} = \mathcal{M}(\rho)$  is not appropriate in this setting.<sup>§</sup>

A precise statement of our hypothesis uses the fitted logistic parameters  $\hat{a}(r, N, M, \mathbf{X})$  and  $\hat{b}(r, N, T, \mathbf{X})$ , estimating the model parameters  $a$  and  $b$  defined above. The asymptotic relation<sup>¶</sup>

<sup>‡</sup>Tanner and Wei (9) used a “horizontal” approach to phase transition measurement in contrast to the “vertical” approach used here. They hold  $\delta$  constant and vary  $r$ , observing very sharp transitions of a practically all-or-nothing nature.

<sup>§</sup>As shown in Fig. 2, and in Tables S2–S5, our results in many cases do generate Z scores for the intercept term in the logistic regression that are consistent with traditional acceptance of this hypothesis. However, traditional acceptance is not needed, in order for the main hypothesis to be valid, and because of finite- $N$  scaling effects indicated above would not ordinarily be expected to hold.

<sup>¶</sup> $\hat{\pi} \rightarrow \pi$  denotes convergence in probability.

$\rho$	$N$	$T$	$\mathcal{M}(\rho)$	$\hat{\delta}(\rho)$	$a$	$b$	$Z$
1/10	40	400	0.351	0.352	-0.128	182.978	-0.581
	50	400	0.351	0.350	0.282	200.131	1.213
	60	400	0.351	0.352	-0.096	221.212	-0.398
	80	400	0.351	0.350	0.415	295.049	1.452
	100	400	0.351	0.349	0.641	383.493	1.900

**Fig. 2.** Typical results.  $\mathbf{X} = \text{Mat}_N$ , with  $\rho = 1/10$  and varying  $N$ .  $T$ , total number of reconstructions;  $a$ ,  $b$ , fitted logistic regression parameters;  $Z$ , traditional Z-score of logistic regression intercept. Extracted from Table S2.

$$\lim_{N \rightarrow \infty} \lim_{T \rightarrow \infty} \frac{\hat{a}(r, N, T, \mathbf{X})}{\hat{b}(r, N, T, \mathbf{X})} = \rho, \quad r = \lfloor \rho N \rfloor \quad [5]$$

implies  $\delta(\rho|\mathbf{X}) = \mathcal{M}(\rho|\mathbf{X})$ . Now note that in Fig. 2 the coefficient  $b$  scales directly with  $N$  and takes values of order several hundred for large  $N$ . This means that, in these experiments, the transition from complete failure to complete success happens in a zone of width  $< 1/100$ . Notice also that  $a$  stays bounded, for the most part between 0 and 2. This means that the response probability evaluated exactly at  $\mathcal{M}(\rho)$  obeys typically

$$\text{logit}(\hat{\pi}(\mathcal{M}(\rho))) \in [0, 2].$$

Fig. 3, *Left* presents the fitted slopes  $b$  and problem sizes  $N$ , as well as an empirical fit. All of the data came from Table S2, but we omitted results for  $\rho \in \{3/4, 9/10\}$  because those slopes were large multiples of all other slopes. Similarly, Fig. 3, *Right* presents the slopes from Table S3. In each plot, the line  $b = 6 + 3N$  is overlaid for comparison. Both plots show a clear tendency of the slopes to increase with  $N$ .

The combination of linear growth of  $b$  with  $N$  and nongrowth of  $a$  implies Eq. 5, and acceptance of our main hypothesis.

**Non-Gaussian Measurements.** This paper studies matrix recovery from Gaussian measurements of the unknown matrix and specifically does not study recovery from partial entry-wise measurements, known in the literature as “matrix completion.” Entry-wise measurements yield a phase transition at a different location (9). Our conclusions do extend to certain non-Gaussian measurements based on  $\mathcal{A}$  with independent and identically distributed entries that are equiprobable  $\pm 1$  (a.k.a. Rademacher matrices). See Table S6.

### Discussion: Existing Literature and Our Contributions

Phase transitions in the success of convex optimization at solving nonconvex problems have been observed previously. Donoho and Tanner considered linear programs useful in the recovery of sparse vectors, established the existence of asymptotic phase transitions, and derived exact formulas for the asymptotic phase transitions (14, 22–25) as well as finite  $N$  exponential bounds. Their work considered the recovery of  $k$ -sparse vectors with  $N$  entries from  $n$  Gaussian measurements. The phase diagram in those papers could be stated in terms of variables  $(\kappa, \delta)$ , where, as in this paper,  $\delta = n/N$  is the under sampling fraction, and  $\kappa = k/N$  is the sparsity fraction, which plays a role analogous to the role played here by the rank fraction  $\rho$ . The proofs in those papers were obtained by techniques from high-dimensional combinatorial geometry, and the formulas for the asymptotic phase transition were implicit, written in terms of special functions.

Donoho et al. (26) later developed a new so-called Approximate Message Passing (AMP) approach to the sparse recovery problem, which gave new formulas for phase boundaries, confirmed rigorously by Bayati and Montanari (27, 28). Whereas the previous formulas involved combinatorial geometry, the new (necessarily equivalent) ones involved minimax decision theory

instead. An extensive literature on AMP algorithms has since been developed (see, e.g., refs. 29 and 30), most notably implying universality in the sparse recovery phase transition (31).

Donoho, Johnstone, and Montanari (DJM) (32) generalized previous AMP-based phase transition results to block-sparse, monotone, and piecewise constant signals. They provided evidence that the observed phase transition  $\delta^*(\cdot)$  of the associated convex optimization reconstruction methods obeyed formulas of the form

$$\delta^*(\kappa) = \mathcal{M}(\kappa), \quad 0 < \kappa < 1. \quad [6]$$

Here,  $\kappa$  is a variable measuring generalized sparsity and  $\mathcal{M}(\kappa)$  the minimax MSE of an appropriate denoiser based on direct noisy measurements. The main result in this brief report fits in this general framework whereby the sparsity  $\kappa$  is identified with the fractional rank  $\rho$ , and the minimax MSE symbol  $\mathcal{M}$  applies to the singular value thresholding denoiser. Our main result is therefore an extension of DJM-style formulas from the sparse recovery setting problem to the low rank recovery setting.<sup>††</sup>

Much mathematical study of matrix recovery (1, 2, 7, 33) has focused on providing rigorous bounds that show the existence of a region of success, without, however, establishing a phase transition phenomenon, or determining its exact boundary. A relatively accurate result by Candès and Recht (CR) for the case of Gaussian measurements (34) implies that  $n \geq 3r(M+N)[1 + O(r/N)]$  measurements are sufficient. Our formulas for the square case  $\beta = 1$  (or equivalently  $M = N$ ) show that

$$\mathcal{M}(\rho|\text{Mat}_N) \sim 6\rho, \quad \rho \rightarrow 0.$$

This agrees with the CR bound in the very low-rank case. However, our relation  $\delta^*(\rho) = \mathcal{M}(\rho)$  is apparently noticeably more accurate than the CR formula at finite  $N$ . Table S5 presents experiments in which the rank is fixed at  $r = 1, 2, 3$ , or 4, and  $N$  varies between 40 and 90. Even though in such cases the corresponding  $\rho = r/N$  is rather small, for example  $1/90$  in the case  $r = 1$  and  $N = 90$ , the empirical phase transition in our experiments agrees much more closely with the nonlinear formula  $\mathcal{M}(\rho)$  than it does with the linear formula  $6\rho$ . Also, in the nonsquare case  $\beta \neq 1$ ,  $\mathcal{M} \sim 2\rho(1 + \beta + \sqrt{\beta})$ , which is strictly smaller than  $6\rho$  for  $\beta < 1$ , so the CR formula is noticeably less accurate than the  $\delta^* = \mathcal{M}$  formula in the nonsquare case.

Of the mathematical methods developed to identify phase transitions but that are not based on combinatorial geometry or approximate message passing, the most precise exploit the “Escape Through the Mesh” (ETM) bounds of Gordon (10, 35, 36). ETM was used to prove upper bounds on the number of Gaussian measurements needed for reconstruction of sparse signals by Stojnic (36) and for low-rank matrices by Oymak and Hassibi (10). In particular, they (10) studied one of the two cases studied here and observed in passing that in the square case ETM gives bounds that seemingly agree with actual phase transition measurements. Very recently, building on the same approach, a DJM-style inequality  $\delta^*(\kappa) \leq \mathcal{M}(\kappa)$  has been announced by Oymak and Hassibi for a wide range of convex optimization problems, including nuclear norm minimization (37); they also presented empirical evidence for  $\delta^*(\rho) = \mathcal{M}(\rho)$  in the square case  $\mathbf{X} = \text{Mat}_N$ .

**Our Contributions.** This paper presents explicit formulas for the minimax MSE of singular value soft thresholding in various cases and shows that these formulas match the appropriate empirical phase transitions in a formal comparison. Compared with earlier work, we make here the following contributions:

- A broad range of phase transition studies, covering nonsquare, square, and symmetric positive-semidefinite matrix recovery

<sup>††</sup>Actually, even sublinear growth of  $a$  implies the result.

<sup>††</sup>The same DJM heuristics based on state evolution apply in both settings.



$$\theta + \cot(\theta) \cdot \left(1 - \frac{1}{3} \cos^2(\theta)\right) = \frac{\pi(1 + \alpha^{-1}\rho - \rho)}{2(1 - \rho)}, \quad [12]$$

which is a simplified version of Eq. 9.

**Parametric Representation of the Minimax Curves.** For square matrices ( $\rho = \tilde{\rho}$ ) the minimax curves  $\mathcal{M}(\rho, 1|Mat)$  and  $\mathcal{M}(\rho|Sym)$  admit a parametric representation in the  $(\rho, \mathcal{M})$  plane using elementary trigonometric functions (see ref. 11). As  $\theta$  ranges over  $[0, \pi/2]$ , the curve  $\theta \mapsto (\rho(\theta|Mat), \mathcal{M}_1(\theta))$  is a parametric representation of  $\mathcal{M}(\rho, 1|Mat)$ , where

$$\rho(\theta|Mat) = 1 - \frac{\pi/2}{\theta + \left(\cot(\theta) \cdot \left(1 - \frac{1}{3} \cos^2(\theta)\right)\right)}$$

$$\mathcal{M}_\alpha(\theta) = 2\rho(\theta) - \rho^2(\theta) + 4\rho(\theta)(1 - \rho(\theta))\sin^2(\theta)$$

$$+ \frac{4\alpha}{\pi}(1 - \rho)^2 \left[ (\pi - 2\theta) \left( \frac{5}{4} - \cos(\theta)^2 \right) + \frac{\sin(2\theta)}{12} (\cos(2\theta) - 14) \right]$$

- Recht B, Fazel M, Parrilo PA (2010) Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Rev* 52(3):471–501.
- Candès EJ, Recht B (2008) Exact matrix completion via convex optimization. *Found Comput Math* 9(6):717–772.
- Gross D (2011) Recovering low-rank matrices from few coefficients in any basis. *IEEE Trans Inf Theory* 57(3):1548–1566.
- Candès EJ, Li X, Ma Y, Wright J (2009) Robust principal component analysis? *JACM* 58(3), 10.1145/1970392.1970395.
- Gross D, Liu YK, Flammia ST, Becker S, Eisert J (2010) Quantum state tomography via compressed sensing. *Phys Rev Lett* 105(15):150401.
- Keshavan RH, Montanari A, Oh S (2010) Matrix completion from noisy entries. *The Journal of Machine Learning* 11:2057–2078.
- Recht B, Xu W, Hassibi B (2010) Null space conditions and thresholds for rank minimization. *Math Program* 127(1):175–202.
- Recht B, Xu W, Hassibi B (2008) Necessary and sufficient conditions for success of the nuclear norm heuristic for rank minimization. *Proceedings of the 47th IEEE Conference on Decision and Control* (IEEE, New York), pp 3065–3070.
- Tanner J, Wei K (2012) Normalized iterative hard thresholding for matrix completion. Available at [http://people.maths.ox.ac.uk/tanner/papers/TaWei\\_NIHT.pdf](http://people.maths.ox.ac.uk/tanner/papers/TaWei_NIHT.pdf). Accessed April 15, 2013.
- Oymak S, Hassibi B (2010) New null space results and recovery thresholds for matrix rank minimization. Available at <http://arxiv.org/abs/1011.6326>. Accessed April 15, 2013.
- Donoho DL, Gavish M (2013) *Minimax Risk of Matrix Denoising by Singular Value Thresholding* (Stanford Univ Dept. of Statistics, Stanford, CA). Available at <http://arxiv.org/abs/1304.2085>. Accessed April 15, 2013. Tech Rep 2013-03.
- Fazel M, Hindi H, Boyd SP (2001) A rank minimization heuristic with application to minimum order system approximation. *Proceedings of the 2001 American Control Conference* (IEEE, New York), pp 4734–4739.
- Fazel M (2002) Matrix rank minimization and applications. PhD thesis (Stanford Univ, Stanford, CA).
- Donoho DL, Tanner J (2009) Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing. *Philos Transact A Math Phys Eng Sci* 367(1906):4273–4293.
- Monajemi H, Jafarpour S, Gavish M, Donoho DL; Stat 330/CME 362 Collaboration (2013) Deterministic matrices matching the compressed sensing phase transitions of Gaussian random matrices. *Proc Natl Acad Sci USA* 110(4):1181–1186.
- Kalish LA (1990) Efficient design for estimation of median lethal dose and quantal dose-response curves. *Biometrics* 46(3):737–748.
- Grant M, Boyd SP (2010) *CVX: Matlab Software for Disciplined Convex Programming* (CVX Research Inc.). Available at <http://cvxr.com/cvx>. Accessed April 15, 2013.
- Toh KC, Todd MJ, Tutuncu RH (1999) SDPT3 - A Matlab software package for semi-definite programming. *Optimization Methods Software*.
- Sturm JF (1999) Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. *Optimization Methods Software*. Available at <http://www.tandfonline.com/doi/abs/10.1080/10556789908805766>. Accessed April 15, 2013.
- Andersen D, Dahl J, Vandenberghe L (2012) CVXOPT: Python software for convex optimization. Available at <http://abel.ee.ucla.edu/cvxopt>. Accessed April 15, 2013.
- Donoho DL, Gavish M, Montanari A (2013) Data for the article ‘The phase transition of matrix recovery from gaussian measurements matches the minimax MSE of matrix

is a parametric representation of  $\mathcal{M}(\rho, 1|Mat)$ , and similarly  $\theta \mapsto (\rho(\theta|Sym), \mathcal{M}_{1/2}(\theta))$  is a parametric representation of  $\mathcal{M}(\rho|Sym)$ , where

$$\rho(\theta|Sym) = 1 - \frac{\theta + \left(\cot(\theta) \cdot \left(1 - \frac{1}{3} \cos^2(\theta)\right)\right) - \pi/2}{\theta + \left(\cot(\theta) \cdot \left(1 - \frac{1}{3} \cos^2(\theta)\right)\right) + \pi/2}$$

**Note Added in Proof.** After this paper was refereed, a preprint by Amelunxen and coauthors appeared (40), which claims to prove  $\delta^*(\rho) = \mathcal{M}(\rho)$  in cases including the present ones.

**ACKNOWLEDGMENTS.** We thank Iain Johnstone for advice at several crucial points, Ewout van den Berg for helpful comments, and Brian Wandell for leadership in scientific computing at Stanford. This work was partially supported by National Science Foundation (NSF) Division of Mathematical Sciences Grant 0906812 (American Recovery and Reinvestment Act of 2009). M.G. was supported by a William R. and Sara Hart Kimball Stanford Graduate Fellowship and a Technion EE Sohnis Promising Scientist Award. A.M. was partially supported by NSF Faculty Early Career Development Award CCF-0743978 and Defense Advanced Research Projects Agency/Air Force Office of Scientific Research Grant FA9550-12-1-0411.

- denoising.’ Available at <http://purl.stanford.edu/tz124hw0000>. Accessed February 15, 2013.
- Donoho DL, Tanner J (2005) Neighborliness of randomly projected simplices in high dimensions. *Proc Natl Acad Sci USA* 102(27):9452–9457.
- Donoho DL (2005) High-dimensional centrally symmetric polytopes with neighborliness proportional to dimension. *Discrete Comput Geom* 35(4):617–652.
- Donoho DL, Tanner J (2009) Counting faces of randomly projected polytopes when the projection radically lowers dimension. *J Am Math Soc* 22(1):1–53.
- Donoho DL, Tanner J (2010) Precise undersampling theorems. *Proc IEEE* 98(6):913–924.
- Donoho DL, Maleki A, Montanari A (2009) Message-passing algorithms for compressed sensing. *Proc Natl Acad Sci USA* 106(45):18914–18919.
- Bayati M, Montanari A (2011) The dynamics of message passing on dense graphs, with applications to compressed sensing. *IEEE Trans Inf Theory* 57(2):764–785.
- Bayati M, Montanari A (2012) The LASSO risk for Gaussian matrices. *IEEE Trans Inf Theory* 58(4):1997–2017.
- Vila J, Schniter P (2011) Expectation-maximization Bernoulli-Gaussian approximate message passing. *Conference Record of the Forty-Fifth ASILOMAR Conference on Signals, Systems and Computers* (IEEE, New York), pp 799–803.
- Rangan S (2011) Generalized approximate message passing for estimation with random linear mixing. *2011 IEEE International Symposium on Information Theory Proceedings* (IEEE, New York), pp 2168–2172.
- Bayati M, Lelarge M, Montanari A (2012) Universality in polytope phase transitions and message passing algorithms. Available at <http://arxiv.org/abs/1207.7321>. Accessed April 15, 2013.
- Donoho DL, Johnstone IM, Montanari A (2011) Accurate prediction of phase transitions in compressed sensing via a connection to minimax denoising. Available at <http://arxiv.org/abs/1111.1041>. Accessed April 15, 2013.
- Oymak S, Mohan K, Fazel M, Hassibi B (2011) A simplified approach to recovery conditions for low rank matrices. *2011 IEEE International Symposium on Information Theory Proceedings* (IEEE, New York), pp 2318–2322.
- Candès EJ, Recht B (2011) Simple bounds for recovering low-complexity models. *Math Program*, 1–11. Available at [www.springerlink.com/index/655751h7568x3621.pdf](http://www.springerlink.com/index/655751h7568x3621.pdf). Accessed April 15, 2013.
- Gordon Y (1988) On Milman’s inequality and random subspaces which escape through a mesh in RN. *Lecture Notes in Mathematics* (Springer, Berlin), Vol 1317, pp 84–106.
- Stojnic M (2009) Various thresholds for l1-optimization in compressed sensing. Available at <http://arxiv.org/abs/0907.3666>. Accessed April 15, 2013.
- Oymak S, Hassibi B (2012) On a relation between the minimax denoising and the phase transitions of convex functions. Available at [www.its.caltech.edu/~soymak/Relation.pdf](http://www.its.caltech.edu/~soymak/Relation.pdf). Accessed April 15, 2013.
- Donoho DL, Gavish M (2013) Companion website for the article ‘The phase transition of matrix recovery from Gaussian measurements matches the minimax MSE of matrix denoising.’ Available at [www.runmycode.org/CompanionSite/Site265](http://www.runmycode.org/CompanionSite/Site265). Accessed February 15, 2013.
- Marcenko V, Pastur L (1967) Distribution of eigenvalues for some sets of random matrices. *Mathematics USSR Sbornik* 1(4):457–483.
- Amelunxen D, Lotz M, McCoy MB, Tropp JA (2013) Living on the edge: A geometric theory of phase transitions in convex optimization. Available at <http://arxiv.org/abs/1303.6672>. Accessed April 15, 2013.