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

David L. Donoho\textsuperscript{a,1}, Matan Gavish\textsuperscript{b}, and Andrea Montanari\textsuperscript{a,b}

Departments of *Statistics and \textsuperscript{b}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 $y_1, \ldots, y_n$ of $X_0$, where $y_i = \text{Tr}(AX_0)$ and each $A$ is an $M \times N$ matrix. A popular approach for matrix recovery is nuclear norm minimization (NMM): solving the convex optimization problem

\[ \min \|X\|_{\text{nuc}}, \quad \text{subject to} \quad y_i = \text{Tr}(AX), \quad 1 \leq i \leq n, \]

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 $\delta = n/(MN)$, 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 $\delta(\rho)$ exists such that, if $\delta > \delta(\rho)$, NMM typically succeeds for large $MN$, whereas if $\delta < \delta(\rho)$, it typically fails. An apparently 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

\[ \min \|Y - X\|_2^2 + \lambda \|X\|_{\text{nuc}}, \]

where $\lambda$ is a parameter. 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 $\delta(\rho) = \text{MSE}(\rho)$ in the second problem, matrix denoising in Gaussian noise: $\delta(\rho) = \text{MSE}(\rho)$, 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

In this problem the measurements $Y$ are direct, so in some sense complete, but noisy. The solution $X_0(Y)$ can be viewed as a shrinkage estimator. In the basis of the singular vectors $U_i$ and $V_i$ of $Y$, the solution $X_0(Y)$ 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

\[ (P_{\text{nuc}}) \min \frac{1}{2} \|Y - X\|_2^2 + \frac{\lambda}{2} \|X\|_{\text{nuc}}. \]

Here, the nuclear norm $\|X\|_{\text{nuc}}$ 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 $A$, then the solution $X_1 = X_1(y)$ to $(P_{\text{nuc}})$ is precisely $X_0$. Such incoherence can be obtained by letting $A$ be random, for instance if $A(X_0) = \text{Tr}(A_0^T X_0)$ with $A_0 \in \mathbb{R}^{M \times N}$ 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 $Z_i \sim \mathcal{N}(0, 1)$. We consider the following nuclear norm denoising scheme:

\[ (P_{\text{nuc}}) \min \left\{ \frac{1}{2} \|X - Y\|_2^2 + \frac{\lambda}{2} \|X\|_{\text{nuc}} \right\}. \]

In this problem the measurements $Y$ are direct, so in some sense complete, but noisy. The solution $X_1(Y)$ can be viewed as a shrinkage estimator. In the basis of the singular vectors $U_i$ and $V_i$ of $Y$, the solution $X_1(Y)$ is diagonal, and the diagonal entries are produced by soft thresholding of the singular values of $Y$.

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/zt2424w0000.

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.
\[ M(\rho; M, N) = \min_{\beta} \max_{\text{rank}(X) \leq \beta} \text{MSE}(X_0, \hat{X}(Y)), \]

where MSE refers to the dimension-normalized mean-squared error \((MN)^{-1} E[||X - \hat{X}||_F^2] \) and subscript \(F \) denotes Frobenius norm. The asymptotic minimax MSE \( M(\rho; \beta) = \lim_{N \to \infty} M(\rho; \beta/N) \) has been derived in ref. 11. Explicit formulas for the curve \( \rho \to M(\rho; \beta) \) appear in Appendix A. 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 \((r, n, M, N)\) having limiting fractional rank \( \rho = \lim_{N \to \infty} \min(M, N) / \max(M, N) \), limiting aspect ratio \( \beta = \lim_{N \to \infty} M/N \), and limiting inappropriateness fraction \( \delta = \lim_{N \to \infty} n/MN \). In the limit of large problem size \( N \), the solution \( X_1(y) \) to the nuclear norm minimization problem \((P_{\text{nuc}})\) is correct with probability converging to one if \( \delta > M(\rho; \beta) \) and incorrect with probability converging to one if \( \delta < M(\rho; \beta) \). In short, the asymptotic phase transition \( \delta(\rho; \beta) \) in Gaussian matrix recovery is equal to the asymptotic minimax MSE \( M(\rho; \beta) \).

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 \( n \geq 2r(M + N + \sqrt{MN})/(1 + O(N)) \) 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 \( M \times N \) 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 \((P_{\text{nuc}})\). 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 \( \delta = \delta(n, N) = n/\text{dim}(X) \) 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 \( X_0 = X \), random measurement matrices \( A = A \times n \times N, \) and obtained random problem instances \((y, A)\). We then applied a convex optimization procedure, obtaining a putative reconstruction \( \hat{X}(y, A) \). 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 \( \hat{\alpha}(r, n, N, \rho, \delta) \); 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 \( A = A_0 + \eta \) 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:

- \( \text{Mat}_{\text{nuc}} \): all \( N \times N \) matrices with real-valued entries and
- \( \text{Sym}_{\text{nuc}} \): all \( N \times N \) real symmetric matrices which are positive-semidefinite.

In the case \( X = \text{Mat}_{\text{nuc}} \), 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 \( X = \text{Sym}_{\text{nuc}} \), we consider low-rank matrices \( X_0 = UU^T \), 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 \( A_{ij} \sim N(0, 1/n) \).

---

**Fig. 1.** Observed/expected phase transitions. (Left) Square matrices; (Right) nonsquare matrices. Curves, asymptotic minimax MSEs \( M(\rho; \beta, X) \). Points, empirical phase transition locations for matrix recovery from incomplete measurements (Tables S2–S5). Curves in left panel, square matrices \( M = N, \beta = 1: \)

\( X = \text{Mat}_{\text{nuc}} \) (blue), \( X = \text{Sym}_{\text{nuc}} \) (red), as defined in text. Curves in right panel, nonsquare matrices \( X = \text{Mat}_{\text{nuc}} \), varying aspect ratio factor \( \beta = M/N \).

2 of 6 | www.pnas.org/cgi/doi/10.1073/pnas.1306110110

Donoho et al.
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:

\[
(P_{\text{rec}}) \min |X|, \quad \text{subject to } y = A \cdot \text{vec}(X). \quad X \in \mathbb{X}.
\]

Here, \(X\) is one of the two classes of matrices, \(M_{\text{dn}}\) or \(S_{\text{ym}}\). 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 \((P_{\text{rec}})\). The probability of exact recovery is defined by

\[
\pi(r, N, X) = \text{Prob}(X_0 = \text{unique solution of } (P_{\text{rec}})).
\]

Clearly, \(0 \leq \pi \leq 1\); for fixed \(N, \pi\) is monotone decreasing in \(r\) and monotone increasing in \(n\). Also \(x(r, N, M_{\text{dn}}) = \pi(r, N, S_{\text{ym}})\).

Estimating the Probability of Exact Recovery. Our procedure follows refs. 14 and 15. For a given matrix type \(X\) and rank \(r\) we conduct an experiment whose purpose is to estimate \(x(r, N, X)\) using \(T\) Monte Carlo trials. In each trial we generate a random instance \((y, A)\), which we supply to a solver for \((P_{\text{rec}})\), obtaining the result \(X_i\). We compute the result \(X_i\) to \(X_0\). If the relative error \(|X_0 - X_i|/|X_0|\) 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 \(\delta\), indicating success or failure in reconstruction. The empirical success fraction is then calculated as

\[
\hat{x}(r, N, T, X) = \frac{\# \text{ successes}}{\# \text{ trials}} = \frac{1}{T} \sum_{i=1}^{T} \delta_i.
\]

These are the raw observations generated by our experiments.

Asymptotic Phase Transition Hypothesis. Consider a sequence of tuples \((r, n, N)\) with \(r/N - \rho\) and \(n/N - \delta\). We assume that there is an asymptotic phase transition curve \(\delta'(\rho)\), i.e., a curve obeying

\[
x(r, n, N, X) = \begin{cases} 1 & \delta < \delta'(\rho) \\ 0 & \delta > \delta'(\rho) \end{cases}.
\]

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 \(\delta'(\rho)\); specifically, whether

\[
\delta'(\rho) = \psi(\rho).
\]

Here, \(\psi(\rho, \text{Mat})\), respectively \(\psi(\rho, \text{Sym})\), is the minimax MSE for Eq. 2 for general matrices (respectively, positive-semidefinite ones). Formulas for \(\psi\) were derived by the authors in ref. 11; computational details are given in Appendix.

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

\[
\hat{x}(\delta) = 1/2.
\]

Under the hypothesis of Eq. 4 we have

\[
\lim_{N \to \infty} \lim_{r \to M_{\text{dn}} \over N} \hat{\delta}(r, N, T, X) = \mathcal{M}(\psi(\rho)).
\]

Consequently, in data analysis we will compare the fitted values \(\hat{\delta}(r, N, T, X)\) with \(\mathcal{M}(\psi(\rho))\).

Experimental Design. To address our main hypothesis regarding the agreement of phase transition boundaries, we measure \(x(r, n, N, X)\) at points \(\delta = r/N\) 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 \((1/10, 9/10)\). As a rough approximation to such an optimal design, we sample at equispaced \(\delta \in [\mathcal{M}(\psi(\rho)) - 0.05, \mathcal{M}(\psi(\rho)) + 0.05]\).

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 \(r, n, M_{\text{dn}}\), and \(X\). Each line also documents the number of Monte Carlo trials \(T\), and the observed success fraction \(\hat{x}\). 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 \(\delta = \mathcal{M}(\psi) \pm 0.05\) to no observed failures at \(\delta = \mathcal{M}(\psi) \pm 0.05\). Fig. 1, Right shows results we obtained at nonsquare matrix ensembles, with varying \(\beta = M_{\text{dn}}/N\). The minimax MSE curves \(\mathcal{M}(\psi(\beta, X))\) 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 \(X = S_{\text{ym}}\), 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 \(\mathcal{M}(\psi)\) to two-digit accuracy. Table S3 shows that, in the symmetric positive-semidefinite case \(X = S_{\text{ym}}\), the empirical phase transition falls within \([\mathcal{M}(\psi) - 0.01, \mathcal{M}(\psi) + 0.01]\).

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 \(E\hat{\delta} = \mathcal{M}(\psi)\) is not appropriate in this setting.

A precise statement of our hypothesis uses the fitted logistic parameters \(\hat{\alpha}(r, N, M_{\text{dn}})\) and \(\hat{\beta}(r, N, T, X)\), estimating the model parameters \(\alpha\) and \(\beta\) defined above. The asymptotic relation \(\hat{\delta}(r, N, T, X) = \mathcal{M}(\hat{\psi}(\rho))\).

\[\text{ Tanner and Wei (9) used a "horizontal" approach to phase transition measurement in contrast to the "vertical" approach used here. They hold } \rho \text{ constant and vary } r, \text{ observing very sharp transitions of a practically all-or-nothing nature.} \]

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

\[\text{ \(\nu_{CP}\) denotes convergence in probability.} \]
imply Eq. \( \rho \in [0, 1] \) 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.

\[
\lim_{N \to \infty} \frac{\sum_{i=1}^{N} f_i(T, X)}{N} = \rho, \quad r = \lfloor \rho N \rfloor \tag{5}
\]

implies \( \delta(\rho) = \mathcal{M}(\rho|X) \). Now note that in Fig. 2 the coefficient \( b \) scales directly with \( \delta \rightarrow s \) and \( \frac{1}{s} \cdot \frac{1}{s} \rightarrow 1 \). This agrees with the CR bound in the very low-rank case. However, our relation \( \delta(\rho) = \mathcal{M}(\rho|X) \) is noticeably less accurate than the CR formula at \( \rho \). This agrees with the CR bound in the very low-rank case. However, our relation \( \delta(\rho) = \mathcal{M}(\rho|X) \) is noticeably less accurate than the CR formula at \( \rho \). This agrees with the CR bound in the very low-rank case. However, our relation \( \delta(\rho) = \mathcal{M}(\rho|X) \) is noticeably less accurate than the CR formula at \( \rho \).

This agrees with the CR bound in the very low-rank case. However, our relation \( \delta(\rho) = \mathcal{M}(\rho|X) \) is noticeably less accurate than the CR formula at \( \rho \).

![Fig. 2](image-url) Typical results: \( X = \text{Mat}_N \), 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.

Here, \( x \) 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.\(^{11}\)

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 \( \delta(\rho) = \mathcal{M}(\rho|X) \) is noticeably less accurate than the CR formula at \( \rho \).

This agrees with the CR bound in the very low-rank case. However, our relation \( \delta(\rho) = \mathcal{M}(\rho|X) \) is noticeably less accurate than the CR formula at \( \rho \). This agrees with the CR bound in the very low-rank case. However, our relation \( \delta(\rho) = \mathcal{M}(\rho|X) \) is noticeably less accurate than the CR formula at \( \rho \).

\[ \mathcal{M}(\rho|\text{Mat}_N) \sim 6\rho, \quad \rho \to 0. \]

This agrees with the CR bound in the very low-rank case. However, our relation \( \delta(\rho) = \mathcal{M}(\rho|X) \) is noticeably less accurate than the CR formula at \( \rho \).

**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 \( (\kappa, \delta, \gamma) \), where, as in this paper, \( \delta = \sqrt{n}/N \) is the under sampling fraction, and \( \gamma = k/N \) 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, piecewise-linear, and piecewise-constant signals. They provided evidence that the observed phase transition \( \delta(\rho) \) of the associated convex optimization reconstruction methods obeyed formulas of the form

\[
\delta(\rho) = \mathcal{M}(\kappa), \quad 0 < \kappa < 1.
\]

![Fig. 2](image-url) Typical results: \( X = \text{Mat}_N \), 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.

Here, \( x \) 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.\(^{11}\)

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 \( \delta(\rho) = \mathcal{M}(\rho|X) \) is noticeably less accurate than the CR formula at \( \rho \).

This agrees with the CR bound in the very low-rank case. However, our relation \( \delta(\rho) = \mathcal{M}(\rho|X) \) is noticeably less accurate than the CR formula at \( \rho \).

\[ \mathcal{M}(\rho|\text{Mat}_N) \sim 6\rho, \quad \rho \to 0. \]

This agrees with the CR bound in the very low-rank case. However, our relation \( \delta(\rho) = \mathcal{M}(\rho|X) \) is noticeably less accurate than the CR formula at \( \rho \).

This agrees with the CR bound in the very low-rank case. However, our relation \( \delta(\rho) = \mathcal{M}(\rho|X) \) is noticeably less accurate than the CR formula at \( \rho \).

This agrees with the CR bound in the very low-rank case. However, our relation \( \delta(\rho) = \mathcal{M}(\rho|X) \) is noticeably less accurate than the CR formula at \( \rho \).

This agrees with the CR bound in the very low-rank case. However, our relation \( \delta(\rho) = \mathcal{M}(\rho|X) \) is noticeably less accurate than the CR formula at \( \rho \).

This agrees with the CR bound in the very low-rank case. However, our relation \( \delta(\rho) = \mathcal{M}(\rho|X) \) is noticeably less accurate than the CR formula at \( \rho \).

This agrees with the CR bound in the very low-rank case. However, our relation \( \delta(\rho) = \mathcal{M}(\rho|X) \) is noticeably less accurate than the CR formula at \( \rho \).

**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

\(^{11}\)The same DJM heuristics based on state evolution apply in both settings.
from Gaussian measurements. We also made certain non-Gaussian measurements and observed similar results.

- A broad range of prediction formulas. We here make available explicit formulas for minimax MSE in the square asymmetric, and nonsquare cases, as well as in the symmetric square positive-semidefinite case.
- Careful empirical technique, including the following.

**Reproducibility.** Publication of the code and data underlying our conclusions.

**Validation.** Matlab/CVX results were reimplemented in Python/CVXOPT, with similar results. Code was executed on three different computing clusters, with similar results.

**Study of finite-N scaling.** We studied tendencies of $a, b$ as $N$ varies and observed behavior consistent with our asymptotic phase transition hypothesis. Studies at a single $N$ could only have shown that an empirical phase transition was near to a given theoretical curve, at a given $N$, but not shown the scaling behavior with $N$ that the main hypothesis properly involves.

**Conclusions**

For the problem of matrix recovery from Gaussian measurements, our experiments, as well as those of others, document the existence of a finite-$N$ phase transition. We compared our measured empirical phase transition curve with a formula from the theory of matrix denoising and observed a compelling match. This connection suggests both new limits on what is possible in the matrix recovery problem and also new ways of trying to reach those limits.

**Appendix**

**Asymptotic Minimax MSE Formula.** The following provides explicit formulas for the matrix denoising minimax curves $M(\rho, \beta|\text{Mat})$ and $M(\rho|\text{Sym})$ used above. Please see ref. 11 for the derivations. A simple web-based calculator, and its underlying source code implementing an efficient computation of these quantities, are provided in ref. 38. Let

$$P_j(x; k) = \frac{1}{2\pi k} \int_0^\infty \frac{t^{k-1}}{\sqrt{\gamma_+(t)}} dt,$$  

where $\gamma_+ = (1 + \sqrt{\gamma})^2$, denote the complementary incomplete moments of the Marcenko–Pastur distribution (39). Define

$$M_j(\Lambda; \rho, \beta) = \rho + \rho - \rho \beta + (1 - \rho)$$

$$\left[\rho \Lambda^2 + (1 - \rho) \left(P_j(\Lambda^2; 1) - 2 \Lambda P_j(\Lambda^2; \frac{1}{2}) + \Lambda^2 P_j(\Lambda^2; 0)\right)\right],$$  

with $\gamma = \gamma(\rho, \beta) = (\rho - \rho \beta)/(\rho - \rho \beta)$.

**Case $X = \text{Mat}_{M,N}$.** It is convenient to orient the $M$ by $N$ matrix so that $M \leq N$, i.e., $\beta \leq 1$. The minimax curve is given by $M(\rho, \beta|\text{Mat}) = \inf_\Lambda M_j(\Lambda; \rho, \beta)$. The left-hand side of Eq. 8 from $\rho$-b to $\rho$ is decreasing in $\Lambda$ and the solution is $\Lambda_0$ we get that $\argmin_\Lambda M_j(\Lambda; \rho, \beta)$ is the unique root of the equation

$$\Lambda^{-1}P_j\left(\Lambda; \frac{1}{2}\right) - P_j\left(\Lambda; 0\right) = \frac{\rho}{a(1 - \rho)}.$$  

The left-hand side of Eq. 9 is decreasing in $\Lambda$ and the solution is determined numerically by binary search.

For square matrices ($\rho = \beta$) Eq. 8 can be expressed using elementary trigonometric functions. In ref. 11 it is shown that

$$M_\alpha(\Lambda; \rho, \rho) = \rho(2 - \rho) + (1 - \rho)[\alpha \Lambda^2 + + a(1 - \rho)]Q_1(\Lambda) - 2 \Lambda Q_2(\Lambda) + \Lambda^2 Q_0(\Lambda)],$$

where

$$Q_0(x) = \frac{1}{\pi} \int_x^2 \frac{1}{\sqrt{4 - t^2}} = 1 - x - \frac{\pi}{4} \sqrt{4 - x^2} = \frac{2}{\pi} \arctan\left(\frac{x}{\sqrt{4 - x^2}}\right)$$

$$Q_1(x) = \frac{1}{\pi} \int_x^2 \frac{t}{\sqrt{4 - t^2}} = \frac{1}{3\pi} (4 - x^2)^{3/2}$$

$$Q_2(x) = \frac{1}{\pi} \int_x^2 \frac{t^2}{\sqrt{4 - t^2}} = \frac{1}{4\pi} \sqrt{4 - x^2} (x^2 - 2) = \frac{2}{\pi} \arcsin\left(\frac{x}{2}\right)$$

are the complementary incomplete moments of the Quarter Circle law. Moreover,

$$\argmin_\Lambda M_j(\Lambda; \rho, \rho) = 2 \sin(\theta_\rho(\rho)),$$

where $\theta_\rho(\rho) \in [0, \pi/2]$ is the unique solution to the transcendental equation

![Fig. 3. Finite-N scaling effects. Fitted slopes for Tables S1 and S2 versus matrix sizes N. (Left) Asymmetric square matrices $X = \text{Mat}_{M,N}$. (Right) Symmetric positive-semidefinite matrices $X = \text{Sym}_{M,N}$. Superimposed lines are $b = 10 + 3\pi$. In left panel, data for $\rho \in [3/4, 9/10]$ were excluded because the slopes $b$ were very large (700 and 2,700, respectively).](image-url)
which is a simplified version of Eq. 9.

### Parametric Representation of the Minimax Curves.

For square matrices ($\rho = \rho$) the minimax curves $\mathcal{M} (\rho, 1 | \text{Mat})$ and $\mathcal{M} (\rho, \text{Sym})$ admit a parametric representation in the $(\rho, M)$ plane using elementary trigonometric functions (see ref. 11). As $\theta$ ranges over $[0, \pi/2]$, the curve $\theta \mapsto (\rho(\theta | \text{Mat}), M_1(\theta))$ is a parametric representation of $\mathcal{M} (\rho, 1 | \text{Mat})$, where

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

$$\mathcal{M}_1(\theta) = 2 \rho(\theta) - \rho^2(\theta) + 4 \rho(\theta) \cos^2(\theta) - 1 + \sin^2(\theta)$$

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

### Note Added in Proof.

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

### ACKNOWLEDGMENTS.

We thank lain 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 Sohni’s 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.

---