SEE COMMENTARY

Check for updates

Deterministic matrices matching the compressed sensing phase transitions of Gaussian random matrices

Hatef Monajemi[a], Sina Jafarpour[b], Matan Gavish[c], Stat 330/CME 362 Collaboration[1], and David L. Donoho[c,2]

Departments of [a]Civil and Environmental Engineering and [c]Statistics, Stanford University, Stanford, CA 94305-4065; and [b]Bina Technologies, Redwood City, CA 94065

Contributed by David L. Donoho, November 12, 2012 (sent for review August 9, 2012)

In compressed sensing, one takes $n < N$ samples of an N-dimensional vector x_0 using an $n \times N$ matrix A, obtaining undersampled measurements $y = Ax_0$. For random matrices with independent standard Gaussian entries, it is known that, when x_0 is k-sparse, there is a precisely determined phase transition: for a certain region in the (k/n,n/N)-phase diagram, convex optimization min $||x||_1$ subject to $y = Ax, x \in X^N$ typically finds the sparsest solution, whereas outside that region, it typically fails. It has been shown empirically that the same property—with the same phase transition location—holds for a wide range of non-Gaussian random matrix ensembles. We report extensive experiments showing that the Gaussian phase transition also describes numerous deterministic matrices, including Spikes and Sines, Spikes and Noiselets, Paley Frames, Delsarte-Goethals Frames, Chirp Sensing Matrices, and Grassmannian Frames. Namely, for each of these deterministic matrices in turn, for a typical k-sparse object, we observe that convex optimization is successful over a region of the phase diagram that coincides with the region known for Gaussian random matrices. Our experiments considered coefficients constrained to X^N for four different sets $X \in \{[0,1], R_+, R, C\}$, and the results establish our finding for each of the four associated phase transitions.

sparse recovery | universality in random matrix theory equiangular tight frames | restricted isometry property | coherence

Compressed sensing aims to recover a sparse vector $x_0 \in \mathbf{X}^N$ from indirect measurements $y = Ax_0 \in \mathbf{X}^n$ with $n < N$, and therefore, the system of equations $y = Ax_0$ is underdetermined. Nevertheless, it has been shown that, under conditions on the sparsity of x_0, by using a random measurement matrix A with Gaussian i.i.d entries and a nonlinear reconstruction technique based on convex optimization, one can, with high probability, exactly recover x_0 (1, 2). The cleanest expression of this phenomenon is visible in the large n, N asymptotic regime. We suppose that the object x_0 is k-sparse—has, at most, k nonzero entries—and consider the situation where $k \sim \rho n$ and $n \sim \delta N$. Fig. 1*A* depicts the phase diagram $(\rho, \delta,) \in (0,1)^2$ and a curve $\rho^*(\delta)$ separating a success phase from a failure phase. Namely, if $\rho < \rho^*(\delta)$, then with overwhelming probability for large N, convex optimization will recover x_0 exactly; however, if $\rho > \rho^*(\delta)$, then with overwhelming probability for large N convex optimization will fail. [Indeed, Fig. 1 depicts four curves $\rho^*(\delta|\mathbf{X})$ of this kind for $\mathbf{X} \in \{[0,1], \mathbf{R}_+, \mathbf{R}, \mathbf{C}\}$—one for each of the different types of assumptions that we can make about the entries of $x_0 \in \mathbf{X}^N$ (details below).]

How special are Gaussian matrices to the above results? It was shown, first empirically in ref. 3 and recently, theoretically in ref. 4, that a wide range of random matrix ensembles exhibits precisely the same behavior, by which we mean the same phenomenon of separation into success and failure phases with the same phase boundary. Such universality, if exhibited by deterministic matrices, could be very important, because certain matrices, based on fast Fourier and fast Hadamard transforms, lead to fast and practical iterative reconstruction algorithms, even in the very large N setting, where these results would have greatest impact. In certain fast algorithms like FISTA (5) and AMP (6), such matrices are simply applied implicitly and never need to be stored explicitly, saving space and memory accesses; the implicit operations often can be carried out in order $N\log(N)$ time rather than the naive order N^2 time typical with random dense matrices. Also, certain deterministic systems (7) have special structures that enable especially fast reconstruction in especially large problems.

In this paper, we show empirically that randomness of the matrix A is not required for the above phase transition phenomenon.* Namely, for a range of deterministic matrices, we show that the same phase transition phenomenon occurs with the same phase boundary as in the Gaussian random matrix case. The probability statement is now not on the matrix, which is deterministic, but instead, on the object to be recovered; namely, we assume that the positions of the nonzeros are chosen purely at random. Our conclusion aligns with theoretical work pointing in the same direction by Tropp (8), Candès and Plan (9), and especially, Donoho and Tanner (10) discussed below; however, the phenomenon that we document is both much broader and more precise and universal than what currently available theory could explain or even suggest. The deterministic matrices that we study include many associated with fast algorithms, and therefore, our results can be of real practical significance. The section *Surprises* also identifies two anomalies uncovered by the experiments.

Methods

For each of the deterministic matrix sequences $(A_{n,N})$ under study and each choice of coefficient set $\mathbf{X} \in \{[0,1], \mathbf{R}_+, \mathbf{R}, \mathbf{C}\}$, we investigated the hypothesis that the asymptotic phase transition boundary is identical to the known boundary for Gaussian random matrices. To measure the asymptotic phase plane at a point (δ, ρ), we chose a sequence of tuples (k, n, N) such that $k/n = \rho$ and $n/N = \delta$, and we performed a sequence of experiments, one for each tuple. In each experiment, we performed Monte Carlo draws of random k-sparse objects $x_0 \in X^N$, attempted to recover $x_0 \in X^N$ from $y = A_{n,N}x_0$, and documented the ratio $\hat{\pi}$ of successes to trials. Our raw empirical observations, thus, consist of a list of entries of the form $\hat{\pi}(k|A_{n,N}, X)$ associated with carefully chosen locations (δ, ρ) in the phase plane. This section discusses details of creation and subsequent analysis of these empirical observations.

Deterministic Matrices Under Study. The different matrices that we studied are listed in Table 1. A simple example is the $n \times 2n$ Spikes and Sines matrix (11–13), $A = [I_n F_n]$, where F_n is usual discrete Fourier transform matrix. Our list includes numerous tight frames, obeying $||A^T x||_2 = c||x||_2$; several are equiangular tight frames, with the smallest possible coherence $1/\sqrt{n}$ [i.e., the maximal inner product between normalized columns (12)] (14–16). Not

Author contributions: D.L.D. designed research; H.M., S.J., M.G., S.3.C.3.C., and D.L.D. performed research; S.J. and D.L.D. contributed new reagents/analytic tools; H.M. and D.L.D. analyzed data; and H.M., M.G., and D.L.D. wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

Data deposition: The data reported in this paper are available at http://purl.stanford.edu/wp335yr5649.

See Commentary on page 1146.

[1]A complete list of the Stat 330/CME 362 Collaboration can be found in *SI Text*.

[2]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.1219540110/-/DCSupplemental.

*This research began as a class project at Stanford University by students of Stat 330/CME 362, which was taught by D.L.D. in Fall of 2011 (Teaching Assistant: M.G.). The basic discovery was a joint effort of all of the participants. Independently, S.J. made a similar discovery in studying Delsarte-Goethals frames for his thesis.

Downloaded from https://www.pnas.org by Cornell University Library on August 16, 2022 from IP address 157.139.19.51.

APPLIED MATHEMATICS

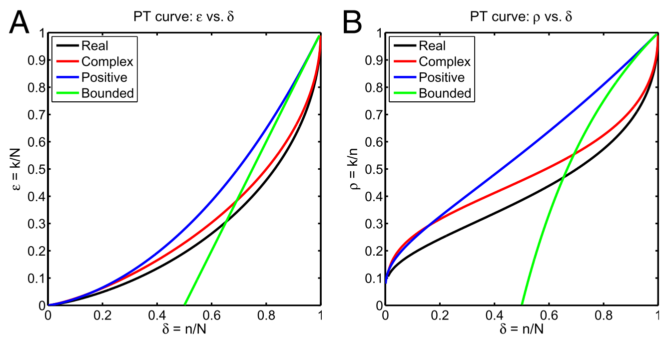

Fig. 1. The four fundamental phase transitions for compressed sensing with Gaussian matrices in (*A*) ϵ–δ and (*B*) ρ–δ coordinates: **R** (black), **C** (red), $\mathbf{R}_+$ (blue), and $[0,1]$ (green).

all of our frames are maximally incoherent; we consider a so-called Affine Chirp frame with coherence $n^{-1/4}$, which is neither tight nor equiangular. Our experiments began as assigned coursework for a course on compressed sensing at Stanford University; they considered the matrices labeled Spikes and Sines (SS), Spikes and Hadamard (SH), and Grassmannian Frame (GF), which served as simple examples of deterministic measurement matrices in that course. After our initial experiments made the basic discovery reported here, we extended our experiments to other matrices. The second stage of experiments involved the deterministic matrices labeled Paley Tight Frame (PETF), Delsarte-Goethals (DG), and Linear Chirp Frame (LC), which were extensively studied in recent years by researchers at Princeton University, especially Robert Calderbank and others. The Spikes and Noiselets (SN), Affine Plane Chirps (AC), and Cyclic matrices were added in a third stage of refinement; their selection was based on personal interests and knowledge of the authors. We emphasize that there are no unreported failures. Namely, we report results for all ensembles that we studied; we are not limiting our reports to those results that seem to support our claims, while holding back information about exceptions. The observed exceptions are disclosed in the section *Surprises* below.

Generation of Pseudorandom Sparse Objects. For $\mathbf{X}=\mathbf{R}_+$, **R**, and **C**, a vector $x_0\in\mathbf{X}^N$ is called k-sparse if $0\le k\le N$ and $\#\{i: x_0(i)\ne 0\}=k$. For case $\mathbf{X}=[0,1]$, $x_0\in\mathbf{X}^N$ is k-sparse if $\#\{i: x_0(i)\notin\{1,0\}\}=k$; this notion was called k-simple in the work by Donoho and Tanner (10). We abuse language in the case $X=[0,1]$ by saying that entries of $x_0(i)$ not in $\{0,1\}$ are nonzeros, whereas entries in $\{0,1\}$ are zeros.

In our experiments, pseudorandom k-sparse objects $x_0\in X^N$ were generated as follows.

Random positions of nonzeros. The positions of the k nonzeros are chosen uniformly random without replacement.

Values at nonzeros. For $X\in\{\mathbf{R}_+,\mathbf{R},\mathbf{C}\}$, the nonzero entries have values i.i.d uniformly distributed in the unit amplitude set $\{|z|=1\}\cap\mathbf{X}$. Thus, for $\mathbf{X}=\mathbf{R}_+$, the nonzeros are all equal to one; for **R**, they are ±1 with signs chosen by i.i.d fair coin tossing, and for **C**, the nonzeros are uniformly distributed on the unit circle. For $\mathbf{X}=[0,1]$, the values not equal to zero or one are uniformly distributed in the unit interval $(0,1)$.

Values at zeros. For $X\in\{\mathbf{R}_+,\mathbf{R},\mathbf{C}\}$, the zeros have value zero. For $\mathbf{X}=[0,1]$, the zeros have values chosen uniformly at random from $\{0,1\}$.

Convex Optimization Problems. Each Monte Carlo iteration in our experiments involves solving a convex optimization problem, in which we attempt to recover a given k-sparse object from $y=Ax_0$. Each of the four specific constraint sets $\mathbf{X}\in\{[0,1],\mathbf{R}_+,\mathbf{R},\mathbf{C}\}$ leads to a different convex program for sparse recovery of $x_0\in X^N$. For each specific choice of **X**, we solve

$$(P^1_{\mathbf{X}})\quad \min\|x\|_1 \text{ subject to } y=Ax,\ x\in\mathbf{X}^N,$$

where $y\in\mathbf{R}^n$ and $A\in\mathbf{R}^{n\times N}$.

When **X** is one of $[0,1]$, $\mathbf{R}_+$, or **R**, the corresponding $(P^1_{\mathbf{X}})$ can be reduced to a standard linear program and solved by a simplex method or an interior point method; in case **R**, the problem $(P^1_{\mathbf{R}})$ is sometimes called Basis Pursuit (17). The case **C** is a so-called second-order cone problem (18).

Probability of Exact Recovery. For a fixed matrix A, let $x_0\in\mathbf{X}^N$ be a random k-sparse object. Let $y=Ax_0$, and therefore, notwithstanding the deterministic nature of A, (y,A) is a random instance of $(P^1_{\mathbf{X}})$. Define

$$\pi(k|A,\mathbf{X}) = \text{Prob}\{x_0 \text{ is the unique solution of } (P^1_{\mathbf{X}})\}.$$

This quantity is the same for a wide range of random k-sparse x_0, regardless of details of their amplitude distributions, provided that they obey certain exchangeability and centrosymmetry properties.

Estimating the Probability of Exact Recovery. Our procedure follows the procedure in ref. 3. For a given matrix A, coefficient type X, and sparsity k, we conduct an experiment with the purpose of estimating $\pi(k|A,\mathbf{X})$ using M Monte Carlo trials. In each trial, we generate a pseudorandom k-sparse vector $x_0\in X^N$ as described above and compute the indirect underdetermined measurements $y=Ax_0$. (y,A) gives an instance of $(P^1_{\mathbf{X}})$, which we supply to a solver, and obtains the result x_1. We compare the result x_1 with x_0. If the relative error $\|x_0-x_1\|_2/\|x_0\|_2$ 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 threshold of 0.001.) We, thus, obtain M binary measurements Y_i, indicating success or failure in reconstruction. The empirical success fraction is then calculated as

$$\hat\pi(k|A,\mathbf{X}) = \frac{\#\{\text{successes}\}}{\#\{\text{trials}\}} = M^{-1}\sum_{i=1}^{M} Y_i.$$

These raw observations are generated by our experiments.

Asymptotic Phase Transition. Let $A_{n,N}$ be an $n\times N$ random matrix with i.i.d Gaussian entries, and consider a sequence of tuples (k,n,N) with $k/n\to\rho$ and $n/N\to\delta$. Then (**1**),

$$\pi(k|A_{n,N},\mathbf{X})\to\begin{cases}1 & \rho<\rho^*(\delta|\mathbf{X})\\ 0 & \rho>\rho^*(\delta|\mathbf{X})\end{cases}, \qquad [1]$$

where the convergence is almost sure (10, 19–22).

Now let $(A_{n,N})$ denote a sequence of deterministic matrices under study, with the same shape/sparity tuples (k,n,N) as in the Gaussian case just mentioned. The hypothesis that we investigate is that expression **1** still holds. There is precedent: the theorem below shows, for the case $\mathbf{X}=[0,1]$, that if each $A_{n,N}$ is a matrix with its columns in general position in $\mathbf{R}^n$, expression **1** holds.

Table 1. Matrices considered here and their properties

Label	Name	Natural δ	Coherence	Tight frame	Equiangular	Definition	Refs.
SS	Spikes and Sines	1/2	$\{1,2\}/\sqrt{n}$	Yes	No	Eq. **S1**	12, 13, 17
SH	Spikes and Hadamard	1/2	$1/\sqrt{n}$	Yes	No	Eq. **S2**	17
SN	Spikes and Noiselets	1/2	$\{1,2\}/\sqrt{n}$	Yes	No	Eq. **S3**	43
PETF	Paley Tight Frame	1/2	$1/\sqrt{n}$	Yes	No	*SI Appendix*	16
GF	Grassmannian Frame	$1/L$	$1/\sqrt{n}$	Yes	Yes	Eq. **S4**	14
DG	Delsarte-Goethals	$1/L$	$1/\sqrt{n}$	Yes	Yes	Eq. **S6**	15, 36
LC	Linear Chirp Frame	$1/L$	$1/\sqrt{n}$	Yes	Yes	Eq. **S5**	15, 44
AC	Affine Plane Chirps	$1/L$	$n^{-1/4}$	No	No	Eq. **S7**	45
CYC	Cyclic	$\in(0,1)$	$1+o(1), N\to\infty$	No	No	Eq. **2**	41, 42

Downloaded from https://www.pnas.org by Cornell University Library on August 16, 2022 from IP address 157.139.19.51.

Downloaded from https://www.pnas.org by Cornell University Library on August 16, 2022 from IP address 157.139.19.51.

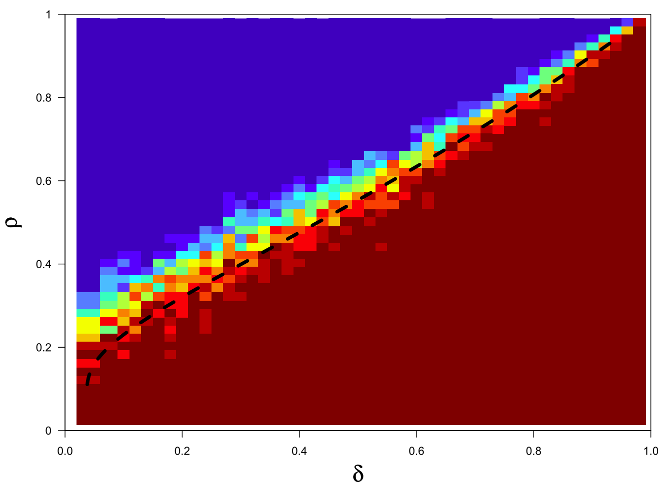

Fig. 2. Empirical (δ, ρ)-phase diagram for the cyclic ensemble. Vertical axis: $\rho = k/n$. Horizontal axis: $\delta = n/N$. Shaded attribute gives fraction of successful reconstructions. Red, 100%; blue, 0%. Dashed line, asymptotic Gaussian prediction $\rho^*(\delta|\mathbf{R})$. In this experiment, $n = 256$, $\delta = 0.04(0.04)0.98$.

Empirical Phase Transitions. The hypothesis that ρ^* marks the large N-phase transition boundary of each deterministic matrix sequence under study is investigated as follows. The empirical phase transition point is obtained by fitting a smooth function $\hat{\pi}(k/n)$ (e.g., a probit function) to the empirical data $\hat{\pi}(k|A, X)$ and finding the empirical 50% response point $\hat{\rho}(n, N, M, \mathbf{X})$—the value of ρ solving:

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

Examples are in Fig. S1. Under the hypothesis that expression **1** holds not only for Gaussian matrices but also for our sequence of deterministic matrices $A_{n,N}$, we have

$$\lim_{N\to\infty, n/N\to\delta} \lim_{M\to\infty} \hat{\rho}(n, N, M, X) =_{a.s.} \rho^*(\delta|\mathbf{X}).$$

Consequently, in data analysis, we will compare the fitted values $\hat{\rho}(n, N, M, \mathbf{X})$ with $\rho^*(\delta|\mathbf{X})$.

Computing. We used a range of convex optimizers available in the MATLAB environment:

CVX: A modeling system for disciplined convex programming by Grant and Boyd (23), Grant et al. (24), and Grant and Boyd (25) supporting two open-source interior-point solvers: SeDuMi and SDPT3.

ASP: A software package for sparse solutions by Friedlander and Saunders (26); its main solver BPdual (27) uses an active-set method to solve the dual of the regularized basis pursuit denoise (BPDN) problem based on dense QR factors of the matrix of active constraints, with only the R factor being stored and updated.

FISTA: A fast iterative soft-thresholding algorithm for solving the BPDN problem. A MATLAB implementation of the algorithm is available from the authors (5).

SPGL1: A solver by van den Berg and Friedlander (28, 29) for the large-scale BPDN problem based on sampling the so-called Pareto curve; it uses the Spectral Gradient Projection method.

Mosek: A commercial optimization toolbox that offers both interior-point and primal simplex solvers (30).

Zulfikar Ahmed also translated our code into python and used the general purpose solver package CVXOPT by Anderson and Vandeberghe (31). We verified the robustness of our results across solvers. We found that SPGL1, with the settings that we used, did not match the other solvers, giving consistently lower phase transitions; therefore, we did not use it in the results reported here. In practice, most of our results were obtained using CVX by participants in the Stanford University graduate class Stat 330/CME 362.

Results

The data that we obtained in our experiments have been deposited (32, 33); they are contained in a text file with more than 100,000 lines, each line reporting one batch of Monte Carlo experiments at a given $k, n, N, A_{n,N}$, and X. Each line documents the coefficient field $\mathbf{X}$, the type of matrix ensemble, the matrix size, the sparsity level, the number of Monte Carlo trials, and the observed success fraction. The file also contains metadata identifying the solver and the researcher responsible for the run. In all, more than 15 million problem solutions were obtained in this project.

Our overall database can be partitioned into several subsets, which address three general questions.

Broad Phase Diagram Survey. In such a survey, we systematically sample the empirical success frequency $\hat{\pi}$ over a 49×49 grid covering the full-phase diagram $0 \le \rho, \delta \le 1$, including regions where we already know that we will see either all failures or all successes (an example is given in Fig. 2). For a matrix type definable only when $N = 2n$ (i.e., undersampling rate $\delta = 1/2$), we considered 49 equispaced ρ-values in $\rho \in \{.02, \ldots, .98\}$ (Fig. S1 shows an example). The GF, DG, LC, and AC frames naturally allow N of the form $N = Ln$ for whole-number L; in such cases, we sampled $\delta = 1/2$, 1/3, etc. (Eq. **S2**). In Figs. 2–4, the two-phase structure is evident.

Precise Positioning of Phase Transition. To address our main hypothesis regarding the agreement of phase transition boundaries, we measure $\hat{\pi}$ at points $\delta = n/N$ and $\rho = k/n$ in the phase plane (δ, ρ), which we expect to be maximally informative about the

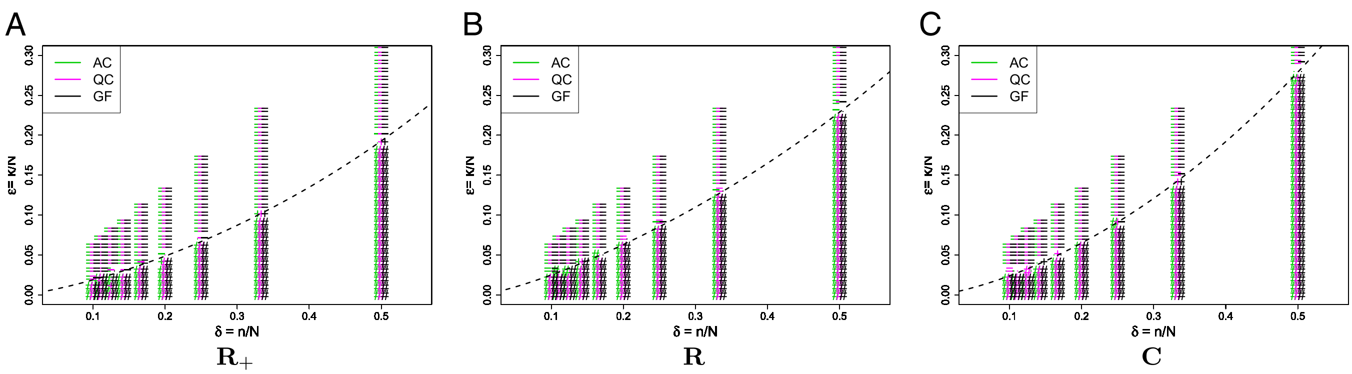

Fig. 3. Partial (δ, ϵ)-phase diagram for chirping frames AC (green), LC (blue), and GF (black). Panels show different coefficient fields: (*A*) real ($n_{AC} = 578$, $n_{LC} = 514$, $n_{GF} = 514$), (*B*) complex ($n_{AC} = 289$, $n_{LC} = 257$, $n_{GF} = 257$), and (*C*) positive ($n_{AC} = 578$, $n_{LC} = 514$, $n_{GF} = 514$). In each panel, symbols (=, –, ., +, #) correspond to probability range $(x - 0.1, x + 0.1)$ for $x = (0.1, 0.3, 0.5, 0.7, 0.9)$, respectively. Symbols are shifted horizontally to minimize overstrike. For example, all three columns on the x axis near $\delta = 1/2$ actually refer to data located on the vertical $\delta = 1/2$ line.

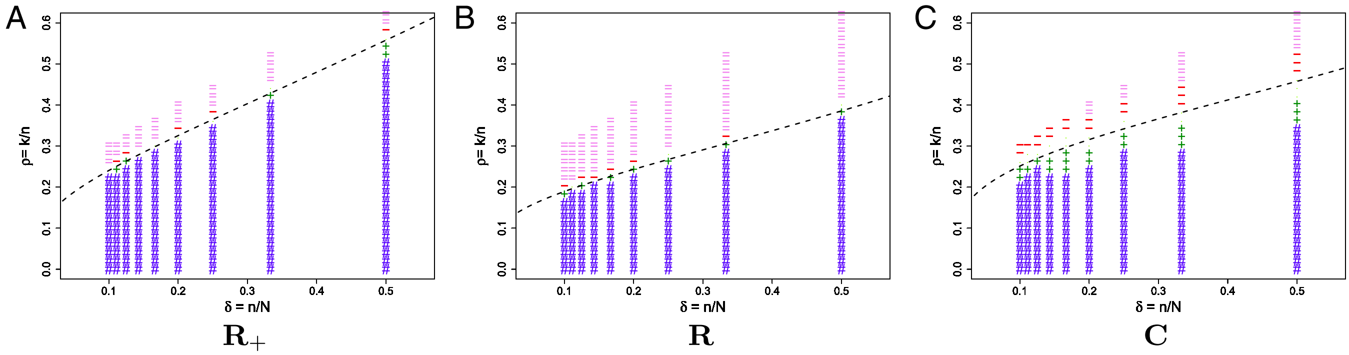

Fig. 4. Partial (δ, ρ)-phase diagram for DG frames. Panels show different coefficient fields: (*A*) positive ($n = 256$), (*B*) real ($n = 256$), and (*C*) complex ($n = 128$). In each panel, symbols (=, –, ., +, #)/colors (violet, red, yellow, green, blue) correspond to probability range $(x - 0.1, x + 0.1)$ for $x = (0.1, 0.3, 0.5, 0.7, 0.9)$, respectively.

location of the phase transition. In fact, the informative locations in binomial response models correspond to points where the probability of response is nearly 50%; hence, we sample heavily for $\rho = k/n \approx \rho^*(\delta|\mathbf{X})$. Figs. S1–S3 show examples comparing the fitted phase transition with the Gaussian theoretical ones. In brief, the results are broadly consistent in all cases with the Gaussian theory, with deviations following a familiar and expected pattern.

Finite-*N* Scaling. The Gaussian theory is asymptotic as $(N \to \infty)$ (19–21, 34), and in general, even with Gaussian random matrices, the large N theory cannot be expected to match empirical data to within the usual naive SEs (21). Instead, one observes, at finite problem sizes, 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 $\approx c_2/N$. Analysis techniques used in ref. 21 motivated Figs. S1 and S2. Visual inspection shows that, as N increases, the data become increasingly consistent with the Gaussian $N \to \infty$ prediction. In brief, the results are consistent with a transition zone width that tends to zero and a transition offset that also tends to zero as N increases.

Discussion

Rigorous Analysis Justifying Our Approach. In one of four coefficient situations that we study—the case where coefficients in x_0 are real and bounded: $\mathbf{X} = [0, 1]$—Donoho and Tanner (10) have proven that, for every reasonable matrix, the same probability distribution holds in finite samples. In consequence, the Gaussian theory, which describes one specific random matrix ensemble, actually describes all reasonable deterministic matrices.

Theorem (10). *Suppose that A is a fixed matrix with its N columns in general position*[†] *in* $\mathbf{R}^n$. *Then*

$$\pi\left(k\middle|A_{n\times N}, [0,1]\right) = 1 - 2^{-(N-k-1)} \sum_{\ell=0}^{N-n-1} \binom{N-k-1}{\ell}$$

$$= P_{N-n,N-k}, \textit{say}.$$

This probability is independent of the matrix A for A ranging through an open dense set in the space of $n \times N$ matrices. In addition to motivating our conclusion, it gives a valuable check on our analysis techniques, because it provides the exact expression

$$E\hat{\pi}\left(k\middle|A_{n\times N}, [0,1]\right) = P_{N-n,N-k},$$

and the exact distribution of $S = M \cdot \hat{\pi}$ as binomial $S \sim bin(M, P_{N-n,N-k})$.

It also motivates our analysis technique. From the binomial form of $P_{N-k,N-n}$, one can see that

$$\pi\left(k\middle|A_{n\times N}, [0,1]\right) = \begin{cases} \geq 1/2 & k \leq 2n - N + 1 \\ \leq 1/2 & k \geq 2n - N + 1 \end{cases},$$

and therefore, for any sequence of matrices $(A_{n\times N})$ all in general position, our experimental method will recover the correct phase transition:

$$\lim_{n/N\to\delta} \lim_{M\to\infty} \hat{\rho}(n, N, M, [0,1]) =_{a.s} (2 - 1/\delta)_+ = \rho^*(\delta; [0,1]).$$

The theorem also motivates the finite-N scaling analysis. Using the exact binomial law $S \sim bin(M, P_{N-n,N-k})$, we conclude that, for large N, the 97.5% success point $k_{97.5}$ satisfies $k_{97.5} \approx \rho^*(\delta; [0,1])n - z_{97.5}\sqrt{2(N-n)}$, whereas the 2.5% success point $k_{2.5}$ satisfies $k_{2.5} \approx \rho^*(\delta|[0,1])n + z_{97.5}\sqrt{2(N-n)}$, where z_p denotes the pth percentile of the standard normal ($z_{97.5} \approx 2$). Hence, one sees that the transition zone between complete failure and complete success has a width roughly $4\sqrt{2(N-n)}$, thereby justifying our fitting power laws in n to the observed transition width.

Finally, the universality of the phase transition $\rho^*(\delta|[0,1]) = (2-1/\delta)_+$ across all deterministic matrices with columns in general position motivates the thrust of this whole project.

Asymptotic Analysis. Tropp (8) and Candès and Plan (9) obtained initial theoretical results on the problem of a single large matrix. They consider a sequence of matrices $A_{n,N}$ and for each fixed problem size (n, N), a random k_n-sparse x_0, which they try to recover from measurements $y = Ax_0$. Their methods apply to all of the matrix families that we have considered here, because our matrices all have low coherence. They give conditions on the aspect ratio n/N, the coherence of $A = A_{n,N}$, and k_n, such that, with high probability, ℓ_1 minimization will correctly recover x_0. Their results are ultimately qualitative in that, for a wide variety of matrices, they predict that there will be a success phase for small k_n/n. However, their results are unable to shed light on the size or shape of the success region. In contrast, we show here that the region is asymptotically the same for certain deterministic matrices as the region for random Gaussian measurement matrices. Separately, Howard et al. (35) and Calderbank et al. (36) pointed to the analogy between compressed sensing and random coding in information theory and observed that average-case reconstruction performance of a deterministic sensing matrix can be expected to be very good, even for some matrices with poor worst-case performance.

Restricted Isometry Properties of Deterministic Matrices. In recent work, several authors have constructed deterministic matrices

[†]A collection of vectors in $\mathbf{R}^n$ is said to be in general position if no subcollection of, at most, n vectors is linearly dependent.

Downloaded from https://www.pnas.org by Cornell University Library on August 16, 2022 from IP address 157.139.19.51.

SEE COMMENTARY

Table 2. Evidence for the positive adaptivity property

Label	Name	δ	$\rho^*(1/2\vert\mathbf{R})$	$\hat{\rho}(\mathbf{R},\mathbf{R})$	$\hat{\rho}(\mathbf{R},\mathbf{R}_+)$	$\hat{\rho}(\mathbf{R}_+,\mathbf{R}_+)$	$\rho(1/2\vert\mathbf{R}_+)$
SS	Spikes and Sines	1/2	0.386	0.3914	0.5621	0.5624	0.558
SH	Spikes and Hadamard	1/2	0.386	0.3942	0.5716	0.5687	0.558
SN	Spikes and Noiselets	1/2	0.386	0.3677	0.5531	0.5605	0.558
LC	Linear Chirp Frame	1/2	0.386	0.3865	0.5614	0.5576	0.558
AC	Affine Plane Chirps	1/2	0.386	0.3866	0.5609	0.5582	0.558

Note that $\hat{\rho}^*(\mathbf{X}_1,\mathbf{X}_2)$ means the phase transition observed when the solver assumes $x \in \mathbf{X}_1^N$, whereas the object actually obeys $x_0 \in \mathbf{X}_2^N$. Thus, the notation $\hat{\rho}^*(\mathbf{R},\mathbf{R}_+)$ means the empirical phase transition that we observed when we ran the solver for signed objects, but in fact, the object was nonnegative. Data were taken at $N=256$ and 25 Monte Carlo repetitions at each grid point. SEs of estimated phase transition yield 2 SE error bars of width ~ 0.01 (compare with Fig. S2).

obeying the so-called restricted isometry property (RIP) of Candès and Tao (2). Such matrices, if they could be built for sufficiently high k (relative to n, N) would guarantee sparse recovery in a particularly strong sense: every k-sparse x_0 would be recoverable. This statement would be a deterministic and not a probabilistic one. Bourgain et al. (37) are the current record holders in deterministic RIP constructions; they have constructed matrices obeying RIP for k slightly[§] larger than $\sqrt{n}$ (37). This result is still far too weak to imply anything close to the empirically observed phase transitions or the known strong neighborliness phase transition for the case of Gaussian matrices.

Calderbank et al. (36) introduced a notion of statistical RIP (StRIP) and constructed a number of deterministic matrices with StRIP. Such matrices guarantee sparse recovery in the same sense as used in this paper (i.e., according to statistical statements across random k-sparse objects with random positions for the nonzeros). However, to our knowledge, existing arguments are not able to derive precise phase transitions from StRIP; they only show that there is some region with high success but do not delineate precisely the regions of success and failure.

Table 1 lists several deterministic matrices obeying StRIP. Indeed, Calderbank et al. (36) used group theory to provide sufficient conditions for a sensing matrix to satisfy StRIP. Subsequently, Jafarpour (15) showed that the coherence property introduced in ref. 38 in conjunction with the tightness of the frame honors the requirements given in ref. 36 and therefore, provides sufficient conditions for StRIP. Combining these arguments, the matrices in Table 1 labeled SS, SH, SN, PETF, GF, DG, and LC obey StRIP for sufficiently large problem sizes. Refs. 15, 38, and 39 have additional discussion on the conditioning and null-space structure of these matrices. We show here that, for all these ensembles, the success region is empirically consistent with the theoretical region for random Gaussian matrices.

Other Ensembles. We studied several matrix ensembles not mentioned so far. For example, we used the same simulation framework and software to study random Gaussian matrices, partial Fourier matrices, and partial Hadamard matrices; our results are in line with earlier reports of Donoho and Tanner (3). We also considered several variations of the SS example based on Discrete Cosine Transforms (DCTs) of types I, II, III, and IV and the discrete Hartley transform. Finally, because of their importance to the theory of convex polytopes, we also considered the cyclic matrices when $\mathbf{X}=\mathbf{R}_+$. All of the deterministic matrices that we considered yielded experimental data consistent with asymptotic agreement of the empirical phase transition and the theoretical phase transition in the Gaussian case—with exceptions noted immediately below.

Surprises. We were surprised by two anomalies.

Positive coefficients, solver for signed coefficients. We can apply the signed solver [i.e., the solver for $(P^1_{\mathbf{R}})$] even when the coefficients are nonnegative. For several of the matrices that we considered, the phase transition that will be observed is the one that is universal for signed coefficients $\rho^*(\delta|\mathbf{R})$. However, in several cases, we observed instead a phase transition at $\rho^*(\delta|\mathbf{R}_+)$, although the problem solved did not assume nonnegative structure. Table 2 presents a list of matrices with this positive adaptivity property. We learned from this phenomenon that, to observe universal signed behavior in the signed case, for some matrices A, it was necessary to make sure that the object x_0 did not always obey $x_0 \geq 0$. For some other matrices, the signed solver gave the same phase transition regardless of whether the nonzeros contained both positive and negative values or only positive values. For conditions under which the sign pattern does not affect reconstruction, see ref. 40.

Cyclic matrix, positive coefficients. Assuming n is even and $j=1,2,\cdots,N$, the cyclic matrix is defined as (41, 42) (Eq. **2**)

$$A_{ij}=\begin{cases}\cos\left(\dfrac{\pi(i+1)(j-1)}{N}\right) & i=1,3,\cdots,n-1\\[2ex] \sin\left(\dfrac{\pi i(j-1)}{N}\right) & i=2,4,\cdots,n.\end{cases} \qquad [2]$$

As discussed in ref. 42, in the case of nonnegative coefficients $\mathbf{X}=\mathbf{R}_+$ and assuming $k \leq n/2$ nonzeros in x_0, there is a unique solution to $(P^1_{\mathbf{R}_+})$: x_0. Consequently, the phase transition associated with this matrix must obey $\rho^* \geq 1/2$ for every δ. Empirically, we do not observe $\hat{\rho} \geq 1/2$; we, instead, observe $\hat{\rho} \approx \rho^*(\delta|\mathbf{R}_+)$ exactly as with other matrices! (See Fig. 1.)

In short, although the theory of cyclic polytopes seemingly forbids it, our observations are consistent with large-N universality of the Gaussian-based formula. Remember that we are considering the behavior of numerical algorithms and that the cyclic matrix contains many very poorly conditioned subsets of k-columns. Possibly, numerical ill-conditioning is responsible for the failure of the predictions from polytope theory.

Limits to Universality. We emphasize that, although the prediction from Gaussian theory applies to many matrices, we do not expect it to apply to all matrices, the theorem quoted from ref. 10 notwithstanding.

Conclusions

For an important collection of large deterministic matrices, the behavior of convex optimization in recovering random k-sparse objects is accurately predicted by the theoretical expressions that are known for the case of Gaussian random matrices. Evidence is presented for objects with coefficients over each of the sets $\{[0,1]^N, \mathbf{R}_+^N, \mathbf{R}^N, \mathbf{C}^N\}$ when the convex optimization problem is appropriately matched and the positions and signs of the nonzeros are randomly assigned.

Standard presentations of compressed sensing based on RIP suggest to practitioners that good deterministic matrices for compressed sensing are as yet unavailable, and possibly will

[§]I.e., $k \sim n^{\alpha}$, with $1/2 < \alpha << 1$.

Downloaded from https://www.pnas.org by Cornell University Library on August 16, 2022 from IP address 157.139.19.51.

be available only after much additional research. We use instead the notion of phase transition, which measures, in a straightforward way, the probability of exact reconstruction. We show here that reconstruction of sparse objects by convex optimization works well for certain deterministic measurement matrices—in fact, just as well as for true random matrices. In particular, our demonstration covers several explicit deterministic matrices for which fast transforms are known.

ACKNOWLEDGMENTS. We would like to thank Michael Saunders and Michael Friedlander for providing us with a prerelease version of the optimization software ASP. This work was partially supported by National Science Foundation Grant DMS 0906812 (American Reinvestment and Recovery Act). Part of the work was done when S.J. was a research scientist at Yahoo! Research. Sam Gross held an NSF-DMS VIGRE fellowship, and Young Choi was supported by Army Grant W911NF-07-2-0027-1. Kevin S. Raines was supported by National Science Foundation Grant DGE 1147470 as well as the Stanford Humanities and Sciences Fellowship. Mainak Choudhury was supported by a 3Com Corp. Stanford Graduate Fellowship. Carlos Sing-Long was partially supported by a Fulbright-CONICYT scholarship.

1. Donoho DL (2006) Compressed sensing. *IEEE Trans Inf Theory* 52(4):1289–1306.
2. Candès EJ, Tao T (2005) Decoding by linear programming. *IEEE Trans Inf Theory* 51(1): 4203–4215.
3. Donoho D, 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.
4. Bayati M, Lelarge M, Montanari A (2012) Universality in polytope phase transitions and message passing algorithmsarXiv:1207.7321.
5. Beck A, Teboulle M (2009) Fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J Imaging Sci* 2(1):183–202.
6. Donoho DL, Maleki A, Montanari A (2009) Message-passing algorithms for compressed sensing. *Proc Natl Acad Sci USA* 106(45):18914–18919.
7. Howard S, Calderbank R, Searle S (2008) Fast reconstruction algorithm for deterministic compressive sensing using second order reed-muller codes. *Conference on Information Sciences and Systems (CISS)* (IEEE, Piscataway, NJ), pp 11–15.
8. Tropp J (2008) On the conditioning of random subdictionaries. *Appl Comput Harmon Anal* 25(1):1–24.
9. Candès EJ, Plan Y (2009) Near-ideal model selection by $\ell 1$ minimization. *Ann Stat* 37(5A):2145–2177.
10. Donoho D, Tanner J (2010) Counting the faces of randomly-projected hypercubes and orthants, with applications. *Discrete Comput Geom* 43(3):522–541.
11. Donoho DL, Stark PB (1989) Uncertainty principles and signal recovery. *SIAM J Appl Math* 49(3):906–931.
12. Donoho DL, Huo X (2001) Uncertainty principles and ideal atomic decomposition. *IEEE Trans Inf Theory* 47(7):2845–2862.
13. Tropp J (2008) On the linear independence of spikes and sines. *J Fourier Anal Appl* 14(5-6):838–858.
14. Strohmer T, Heath R (2003) Grassmannian frames with applications to coding and communication. *Appl Comput Harmon Anal* 14(3):257–275.
15. Jafarpour S (2011) Deterministic compressed sensing. PhD thesis (Princeton University, Princeton).
16. Bandeira A, Fickus M, Mixon D, Wong P (2012) The road to deterministic matrices with the restricted isometry property. *arXiv*:12021234.
17. Chen S, Donoho D, Saunders M (1999) Atomic decomposition by Basis Pursuit. *SIAM J Sci Comput* 20(1):33–61.
18. Boyd S, Vandenberghe L (2004) *Convex Optimization* (Cambridge Univ Press, Cambridge, United Kingdom).
19. Donoho D (2006) High-dimensional centrally symmetric polytopes with neighborliness proportional to dimension. *Discrete Comput Geom* 35(4):617–652.
20. Donoho DL, Tanner J (2005) Neighborliness of randomly projected simplices in high dimensions. *Proc Natl Acad Sci USA* 102(27):9452–9457.
21. Donoho D, Tanner J (2009) Counting faces of randomly-projected polytopes when the projection radically lowers dimension. *J Amer Math Soc* 22(1):1–53.
22. Maleki A, Anitoriy L, Yang Z, Baraniuk R (2011) Asymptotic analysis of complex lasso via complex approximate message passing (camp)arXiv:11080477v1.
23. Grant M, Boyd S (2010) *CVX: Matlab Software for Disciplined Convex Programming, Version 1.21*. Available at http://cvxr.com/cvx. Accessed December 12, 2012.
24. Grant M, Boyd S, Ye Y (2006) Disciplined convex programming. *Global Optimization: From Theory to Implementation, Nonconvex Optimization and Its Applications*, eds Liberti L, Maculan N (Springer, New York), pp 155–210.
25. Grant M, Boyd S (2012) *CVX Users' Guide for CVX Version 1.22*. Available at http://cvxr.com/cvx/cvx_usrguide.pdf. Accessed December 12, 2012.
26. Friedlander M, Saunders M (2010) *ASP: A Set of Matlab Functions for Solving Basis Pursuit-Type Problems*. Available at http://www.cs.ubc.ca/~mpf/asp. Accessed December 14, 2012.
27. Friedlander M, Saunders M (2011) *A Dual Active-Set Quadratic Programming Method for Finding Sparse Least-Squares Solutions, Department of Computer Science, University of British Columbia*. Available at http://www.cs.ubc.ca/~mpf/asp. Accessed December 15, 2012.
28. van den Berg E, Friedlander MP (2008) Probing the pareto frontier for basis pursuit solutions. *SIAM J Sci Comput* 31(4):890–912.
29. van den Berg E, Friedlander MP (2007) *SPGL1: A Solver for Large-Scale Sparse Reconstruction*. Available at http://www.cs.ubc.ca/labs/scl/spgl1. Accessed December 12, 2012.
30. Mosek APS (2012) Mosek Optimization Software. Available at http://www.mosek.com. Accessed December 12, 2012.
31. Andersen M, Vandeberghe L (2012) *CVXOPT*. Available at http://abel.ee.ucla.edu/cvxopt/index.html. Accessed December 27, 2012.
32. Monajemi H, Jafarpour S, Gavish, M {Stat 330/CME 362} Collaboration, Donoho DL (2012) Data for the article Deterministic Matrices Matching the Compressed Sensing Phase Transitions of Gaussian Random Matrices. Available at http://purl.stanford.edu/wp335yr5649. Accessed December 14, 2012.
33. Monajemi H, Jafarpour S, {Stat 330/CME 362} Collaboration, Gavish M, Donoho D (2012) RunMyCode Companion Website for Deterministic Matrices Matching the Compressed Sensing Phase Transitions of Gaussian Random Matrices. Available at http://www.runmycode.org/CompanionSite/Site190. Accessed December 27, 2012.
34. Donoho D, Tanner J (2010) Precise undersampling theorems. *Proc IEEE* 98(6):913–924.
35. Howard S, Calderbank A, Searle S (2008) A fast reconstruction algorithm for deterministic compressive sensing using second order reed-muller codes. *Proceedings of 42nd Annual IEEE Conference on Information Sciences and Systems* (IEEE, Piscataway NJ), pp 11–15.
36. Calderbank AR, Howard S, Jafarpour S (2010) Construction of a large class of deterministic matrices that satisfy a statistical restricted isometry property. *IEEE J Sel Top Signal Process* 4(2):358–374.
37. Bourgain J, Dilworth S, Ford K, Konyagin S, Kutzarova D (2011) Explicit constructions of RIP matrices and related problems. *Duke Math J* 159(1):145–185.
38. Bajwa W, Calderbank R, Jafarpour S (2010) Revisiting model selection and recovery of sparse signals using one-step thresholding. *Proceedings of the 48th Annual Allerton Conference on Communication, Control, and Computing* (IEEE Piscataway, NJ), pp 977–984.
39. Jafarpour S, Duarte MF, Calderbank AR (2012) Beyond worst-case reconstruction in deterministic compressed sensing. *IEEE International Symposium on Information Theory (ISIT)* (IEEE, Cambridge, MA), pp 1852–1856.
40. Bajwa W, Calderbank R, Jafarpour S (2010) Why Gabor frames? Two fundamental measures of coherence and their role in model selection. *arXiv*:1006.0719.
41. Grünbaum B (2003) *Convex Polytopes*, Graduate Texts in Mathematics, Vol 221 (Springer, New York), 2nd Ed.
42. Donoho DL, Tanner J (2005) Sparse nonnegative solution of underdetermined linear equations by linear programming. *Proc Natl Acad Sci USA* 102(27):9446–9451.
43. Coifman R, Meyer Y, Geschwind F (2001) Noiselets. *Appl Comput Harmon Anal* 10(1): 27–44.
44. Applebaum L, Howard S, Searle S, Calderbank R (2009) Chirp sensing codes: Deterministic compressed sensing measurements for fast recovery. *Appl Comput Harmon Anal* 26(2):283–290.
45. Gurevich S, Hadani R, Sochen N (2008) The finite harmonic oscillator and its associated sequences. *Proc Natl Acad Sci USA* 105(29):9869–9873.

Downloaded from https://www.pnas.org by Cornell University Library on August 16, 2022 from IP address 157.139.19.51.