DETERMINISTIC MATRICES MATCHING THE
COMPRESSED SENSING PHASE TRANSITIONS
OF GAUSSIAN RANDOM MATRICES

By

Hatef Monajemi
Sina Jafarpour
Matan Gavish
David L. Donoho

Technical Report No. 2012-14
December 2012

Department of Statistics
STANFORD UNIVERSITY
Stanford, California 94305-4065
DETERMINISTIC MATRICES MATCHING THE COMPRESSED SENSING PHASE TRANSITIONS OF GAUSSIAN RANDOM MATRICES

By

Hatef Monajemi
Stanford University

Sina Jafarpour
Bina Technologies, Redwood City, CA

Matan Gavish
Stanford University

David L. Donoho
Stanford University

Technical Report No. 2012-14
December 2012

This research was supported in part by National Science Foundation grant DMS 0906812.

Department of Statistics
STANFORD UNIVERSITY
Stanford, California 94305-4065

http://statistics.stanford.edu
Deterministic Matrices Matching the Compressed Sensing Phase Transitions of Gaussian Random Matrices

Hatef Monajemi*, Sina Jafarpour†
Matan Gavish*, Stat 330/CME 362 Collaboration*
David L. Donoho*

December 2012

Abstract

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 Gaussian i.i.d 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, while 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.

**keywords:** Compressed sensing, Sparse Recovery, Grassmannian Frames, Paley Frames, Delsarte-Goethals Frames, Noiselets

*Stanford University, Stanford, CA
†Bina Technologies, Redwood City, CA


1 Introduction

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 $n < N$, so that 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 $x_0$; see [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 $n \sim \delta N$. Figure 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; while on the other hand, if $\rho > \rho^*(\delta)$, then with overwhelming probability for large $N$ convex optimization will fail. (Indeed, Figure 1 depicts four curves $\rho^*(\delta)_{X}$ of this kind, for $X \in \{[0,1], \mathbf{R}_+, \mathbf{R}, \mathbf{C}\}$ one for each of the different types of assumptions we can make about the entries of $x_0 \in \mathbf{X}^N$; see details below).

How special are Gaussian matrices to the above results? It was shown, first empirically in [3] and, recently, theoretically in [4], that a wide range of random matrix ensembles exhibit 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\footnote{This research began as a class project at Stanford University by students of Stat 330/CME 362, taught by Donoho in Fall 2011 (TA: Matan Gavish). The basic discovery was a joint effort of all the participants. Independently, Jafarpour made a similar discovery in studying Delsarte-Goethals frames for his thesis.} \footnote{The researchers in Stat 330/CME 362 who ran experiments for this collaboration include: Sivaram Ambikasaran, Sergio Bacallado, Dinesh Bharadia, Yuxin Chen, Young Choi, Mainak Chowdhury, Soham Chowdhury, Anil Dogle, Will Fithian, Georges Goetz, Logan Grosenick, Sam Gross, Gage Hills, Michael Horstein, Milinda Lakham, Jason Lee, Jian Li, Linxi Liu, Carlos Long, Mike Marx, Akshay Mittal, Albert No, Reza Omran, Leonid Pekelis, Junjie Qin, Kevin Raines, Ernest Ryu, Andrew Saxe, Dai Shi, Keith Silits, David Strauss, Gary Tang, Chaojun Wang, Zoley Zhou, and Zhen Zhu. Zulfikar Ahmed translated our code into python and replicated our experiments on PiCloud/Amazon Web Services. All these participants should be considered co-authors of this paper. The named authors designed the experiments, amalgamated the data, prepared the analysis in this paper, and either framed the discussion of the final manuscript or wrote the manuscript.} show empirically that randomness of the matrix $A$ is
Figure 1: The four fundamental phase transitions for compressed sensing with Gaussian matrices in (a) $\epsilon$-$\delta$, and (b) $\rho$-$\delta$ coordinates: $R$ (black), $C$ (red), $R_+$ (blue), $[0, 1]$ (green).

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 non zeros 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; but the phenomenon we document is both much broader and more precise and universal than what currently available theory could explain or even suggest. The deterministic matrices we study include many associated with fast algorithms, and so our results can be of real practical significance. The section labelled Surprises also identifies two anomalies uncovered by the experiments.

2 Methods

For each of the deterministic matrix sequences $(A_{n,N})$ under study, and each choice of coefficient set $X \in \{[0, 1], R_+, R, C\}$, 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 $(\delta, \rho)$, we chose a sequence of tuples $(k,n,N)$ such that $k/n = \rho$ and $n/N = \delta$ and 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 to carefully chosen locations $(\delta, \rho)$ in the phase plane. The following discusses details of creation and subsequent analysis of these empirical observations.
Table 1: Matrices considered here, and their properties

<table>
<thead>
<tr>
<th>Label</th>
<th>Name</th>
<th>Natural #</th>
<th>Coherence</th>
<th>Tight Frame</th>
<th>Equiangular</th>
<th>Definition</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SS</td>
<td>Spikes &amp; Sines</td>
<td>1/2</td>
<td>$[1, 2]/\sqrt{n}$</td>
<td>Yes</td>
<td>No</td>
<td>Eq. 4</td>
<td>[17]</td>
</tr>
<tr>
<td>SH</td>
<td>Spikes &amp; Hadamard</td>
<td>1/2</td>
<td>$1/\sqrt{n}$</td>
<td>Yes</td>
<td>No</td>
<td>Eq. 4</td>
<td>[17]</td>
</tr>
<tr>
<td>SN</td>
<td>Spikes &amp; Noiselets</td>
<td>1/2</td>
<td>$[1, 2]/\sqrt{n}$</td>
<td>Yes</td>
<td>No</td>
<td>Appendix</td>
<td>[16]</td>
</tr>
<tr>
<td>PETF</td>
<td>Paley Tight Frame</td>
<td>1/2</td>
<td>$1/\sqrt{n}$</td>
<td>Yes</td>
<td>No</td>
<td>Eq. 5</td>
<td>[43]</td>
</tr>
<tr>
<td>GF</td>
<td>Grassmannian Frame</td>
<td>1/2</td>
<td>$1/\sqrt{n}$</td>
<td>Yes</td>
<td>Yes</td>
<td>Eq. 6</td>
<td>[14]</td>
</tr>
<tr>
<td>DG</td>
<td>Delsarte-Goethals</td>
<td>1/2</td>
<td>$1/\sqrt{n}$</td>
<td>Yes</td>
<td>Yes</td>
<td>Eq. 8</td>
<td>[36, 15]</td>
</tr>
<tr>
<td>LC</td>
<td>Linear Chirp Frame</td>
<td>1/2</td>
<td>$1/\sqrt{n}$</td>
<td>Yes</td>
<td>Yes</td>
<td>Eq. 7</td>
<td>[44, 15]</td>
</tr>
<tr>
<td>AC</td>
<td>Affine Plane Chirps</td>
<td>1/2</td>
<td>$n^{-1/4}$</td>
<td>No</td>
<td>No</td>
<td>Eq. 9</td>
<td>[45]</td>
</tr>
<tr>
<td>CYC</td>
<td>Cyclic</td>
<td>$\in (0, 1)$</td>
<td>$1 + o(1), N \to \infty$</td>
<td>No</td>
<td>No</td>
<td>Eq. 2</td>
<td>[41, 42]</td>
</tr>
</tbody>
</table>

2.1 Deterministic Matrices Under Study

The different matrices we studied are listed in Table 1. A simple example is the $n \times 2n$ “Spikes and Sines” matrix [11, 12, 13], $A = [I \ F]$, where $F$ 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]), see [14, 15, 16]. Not 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; they considered the matrices labelled SS, SH, and 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 labelled PETF, DG, and LC, which were extensively studied in recent years by researchers at Princeton, especially Robert Calderbank and co-authors. The SN, AC, and CYC 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 which seem to support our claims while holding back information about exceptions. The observed exceptions are disclosed in the section *Surprises* below.

2.2 Generation of Pseudo-Random Sparse Objects

For $X = \mathbb{R}_+$, $\mathbb{R}$ and $\mathbb{C}$, a vector $x_0 \in \mathbb{X}^N$ is called $k$-sparse if $0 \leq k \leq N$ and $\#\{i : x_0(i) \neq 0\} = k$. For case $X = [0, 1]$, $x_0 \in \mathbb{X}^N$ is $k$-sparse if $\#\{i : x_0(i) \notin \{0, 1\}\} = k$; This notion was called $k$-simple in 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 ‘non-zeros’, while entries in $\{0, 1\}$ are ‘zeros’.

In our experiments, pseudo-random $k$-sparse objects $x_0 \in \mathbb{X}^N$ were generated as follows.

- **Random Positions of Non-Zeros**. The positions of the $k$ non-zeros are chosen at uniformly random without replacement.
• **Values at Non-Zeros.** For \( X \in \{ \mathbb{R}_+, \mathbb{R}, \mathbb{C} \} \), the non-zero entries have values i.i.d uniformly distributed in the unit amplitude set \( \{|z| = 1\} \cap X \). Thus for \( X = \mathbb{R}_+ \) the non zeros are all equal to 1, while for \( \mathbb{R} \) they are \( \pm 1 \) with signs chosen by i.i.d fair coin tossing and for \( \mathbb{C} \) the non zeros are uniformly distributed on the unit circle. For \( X = [0, 1] \), the values not equal to 0 or 1 are uniformly distributed in the unit interval \((0, 1)\).

• **Values at Zeros.** For \( X \in \{ \mathbb{R}_+, \mathbb{R}, \mathbb{C} \} \), the zeros have value 0. For \( X = [0, 1] \), the zeros have values chosen uniformly at random from \( \{0, 1\} \).

### 2.3 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 \( X \in \{ [0, 1], \mathbb{R}_+, \mathbb{R}, \mathbb{C} \} \) lead to a different convex program for sparse recovery of \( x_0 \in X^N \). For each specific choice of \( X \) we solve,

\[
(P_X^1) \quad \min \|x\|_1 \quad \text{subject to} \quad y = Ax, \ x \in X^N,
\]

where \( y \in \mathbb{R}^n \) and \( A \in \mathbb{R}^{n \times N} \).

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

### 2.4 Probability of Exact Recovery

For a fixed matrix \( A \), coefficient type \( X \) and sparsity \( k \) we conduct an experiment whose purpose is to estimate \( \pi(k|A, X) \) using \( M \) Monte Carlo trials. In each trial we generate a pseudo-random \( k \)-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 \((P_X^1)\), which we supply to a solver, and obtain the result \( x_1 \). We compare the result \( x_1 \) to \( x_0 \). If the relative error \( \|x_0 - x_1\|_2/\|x_0\|_2 \) is smaller than a numerical tolerance, we declare the

### 2.5 Estimating the Probability of Exact Recovery

Our procedure follows [3] For a given matrix \( A \), coefficient type \( X \) and sparsity \( k \) we conduct an experiment whose purpose is to estimate \( \pi(k|A, X) \) using \( M \) Monte Carlo trials. In each trial we generate a pseudo-random \( k \)-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 \((P_X^1)\), which we supply to a solver, and obtain the result \( x_1 \). We compare the result \( x_1 \) to \( x_0 \). 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

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

These are the raw observations generated by our experiments.

### 2.6 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

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

where the convergence is almost sure. See [19, 20, 21, 10, 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 we investigate is that (1) still holds. There is precedent for this: Theorem 1 below shows for the case $X = [0, 1]$, that if each $A_{n,N}$ is a matrix with its columns in general position in $\mathbb{R}^n$, (1) holds.

### 2.7 Empirical Phase Transitions

The hypothesis that $\rho^*$ 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 $\hat{\rho}(n, N, M, X)$ - the value of $\rho$ solving

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

For examples, see Figure 2. Under the hypothesis that (1) holds not only for Gaussian matrices, but also for our sequence of deterministic matrices $A_{n,N}$, we have

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

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

### 2.8 Computing

We used a range of convex optimizers available in the MATLAB environment. These include:
+ **CVX** [23]: A modeling system for disciplined convex programming by Boyd, Grant and others, supporting two open source interior-point solvers: SeDuMi and SDPT3 [24, 25].

+ **ASP** [26]: A software package for sparse solutions by M. Friedlander and M. Saunders; 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** [28]: A fast iterative soft-thresholding algorithm for solving the BPDN problem. A MATLAB implementation of the algorithm is available from the authors.

+ **SPGL1** [29, 30]: A solver by Friedlander and van den Berg for the large-scale BPDN problem based on sampling the so-called Pareto curve; it uses the Spectral Gradient Projection (SPG) method.

+ **Mosek** [31] A commercial optimization toolbox that offers both interior-point and primal simplex solvers.

Zulfikar Ahmed also translated our code into Python and used the general purpose solver package CVXOPT [32] by Vanderberghe and others. We verified the robustness of our results across solvers. We found that SPGL1, with the settings we used, did not match the other solvers, giving consistently lower phase transitions; so we did not employ it in the results reported here. In practice, most of our results were obtained using CVX, by participants in Stanford graduate class Stat 330/CME 362.

### 3 Results

The data we obtained in our experiments have been deposited at [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 \( k, n, N, A_{n,N} \text{ and } X \). Each line documents the coefficient field \( X \), 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 \( 49 \times 49 \) grid covering the full phase diagram \( 0 \leq \rho, \delta \leq 1 \), including regions where we already know we will see either all failures or all successes; see Figure 7 for an example. For a matrix type definable only when \( N = 2n \) (i.e. undersampling rate
δ = 1/2) we considered 49 equispaced ρ values in ρ ∈ {.02, . . . , .98}. See Figure 2 for 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 δ = 1/2, 1/3, etc. See Figure 4. In all these plots, 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 ˆπ at points δ = n/N and ρ = k/n in the phase plane (δ, ρ) 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 nearly 50%; hence we sample heavily for ρ = k/n ≈ ρ∗(δ|X). Figures 2 and 6 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 → ∞) [20, 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 standard errors [21]. Instead, one observes, at finite problem sizes, a finite transition zone of width ≈ c1/N1/2 and a small displacement of the empirical phase transition away from the asymptotic Gaussian phase transition, of size ≈ c2/N. Analysis techniques used in [21] motivated Figures 2 and 3. Visual inspection shows that, as N increases the data become increasingly consistent with the Gaussian N → ∞ 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.

4 Discussion

Rigorous analysis justifying our approach. In one of the four coefficient situations we study - the case where coefficients in x0 are real and bounded: 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.
Figure 2: Fitted success probabilities \( \hat{\pi}(\rho|A_{n,N}, R) \) for various matrices and problem sizes; all at undersampling ratio \( \delta = 1/2 \). In each panel, color encodes problem size: Green, smallest; Blue, midmost; Black, largest. Problem sizes: USE (256, 512, 1024); SS (514, 802, 1202); SH (256, 1024, 2048); SN (256, 512, 1024); PETF (258, 510, 1022); GF (388, 1028, 2036); DG (512, 2048, 8192); LC (194, 514, 1018); AC (484, 2116, 3364). The vertical dashed line locates the asymptotic phase transition for Gaussian ensembles. The horizontal dashed line locates the 50% success probability. In all cases the tendency is for the curves associated with larger problem sizes to cross the 50% probability line near the theoretical phase transition for Gaussian matrices, and for the steepness of the transition to increase with \( N \).
Figure 3: Offset from asymptotic Gaussian phase transition, offset = \( \hat{\rho}(n, N, M, R) - \rho^*(\frac{1}{2}R) \), as a function of problem size \( N \) for coefficient set \( X = R \) and \( \delta = n/N = 1/2 \). The x axis shows \( 1/N \) and the y axis shows offset \( \pm 2SE(\text{offset}) \). In each case as \( N \) increases, we see that the interval gets closer to the asymptotic result indicated by the dashed line. Note: the label USE indicates behavior with initially Gaussian random matrices which are then normalized so all column norms are 1; by mathematical theorems, results must agree in the \( N \to \infty \) limit with the dashed line, in finite problem sizes there are empirical differences. The sizes of those empirical differences are consistent with empirical differences seen in all other cases.
**Theorem:** [10] Suppose that $A$ is a fixed matrix with its $N$ columns in general position\(^3\) in $\mathbb{R}^n$.

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

This 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, since it provides the exact expression

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

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

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

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

and so 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 $S \sim \text{bin}(M, P_{N-n,N-k})$ we conclude that, for large $N$, the 97.5% success point $k_{97.5}$ satisfies $k_{97.5} \approx \rho^*(\delta; [0, 1])n - z_{97.5}\sqrt{2(N-n)}$, while 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 $p$-th 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 $4\sqrt{2(N-n)}$. This justifies 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.

\textbf{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 the matrix families we have considered.

\footnote{A collection of vectors in $\mathbb{R}^n$ is said to be in general position if no subcollection of at most $n$ vectors is linearly dependent}
here, because our matrices all have low coherence. They give conditions on the aspect ratio $n/N$ and on the coherence of $A = A_{n,N}$ and on $k_n$ such that, with high probability, $\ell_1$ 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 $k_n/n$ small enough. 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, Calderbank and co-authors [35, 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.

RIP properties of deterministic matrices. In recent work, several authors have constructed deterministic matrices 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 would be a sure statement, and not a probabilistic one. Bourgain and co-authors are the current record holders in deterministic RIP constructions [37]; they have constructed matrices obeying RIP for $k$ slightly\(^4\) larger than $\sqrt{n}$. 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, Howard, and Jafarpour [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 without delineating precisely the regions of success and failure.

Our 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

\[^4\text{i.e. } k \sim n^{\alpha}, \text{ with } 1/2 < \alpha < 1\]
Figure 7: Empirical $(\delta, \rho)$-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, $n = 256$, $\delta = 0.04(0.04)0.98$.

that the coherence property introduced in [38] in conjunction with the tightness of the frame honors the requirements of [36] and therefore provides sufficient conditions for StRIP. Combining these arguments, the matrices in Table 1 labelled SS, SH, SN, PETF, GF, DG, and LC obey StRIP for sufficiently large problem sizes. See [38, 15, 39] for further 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 in [3]. We also considered several variations of the spikes and sines example, based on Discrete Cosine Transforms of types I, II, III, 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 the deterministic matrices 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 $(P^0_{\mathbf{R}})$, even when the coefficients are nonnegative. For several of the matrices we considered, the phase transition that will be observed is the one which is universal for signed
coefficients $\rho^*(\delta|\mathbf{R})$. However, in several cases, we observed instead a phase transition at $\rho^*(\delta|\mathbf{R}_+)$, even though 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 in order to observe universal ‘signed’ behavior in the signed case, for some matrices $\mathbf{A}$ it was necessary to make sure that the object $\mathbf{x}_0$ did not always obey $\mathbf{x}_0 \geq 0$. For some other matrices, the signed solver gave the same phase transition whether or not the nonzeros contained both positive and negative values, or only positive values. For conditions under which the sign pattern doesn’t affect reconstruction, see [40].

- Cyclic Matrix, Positive Coefficients. Assuming $n$ is even and $j = 1, 2, \cdots, N$, the cyclic matrix is defined as follows [41, 42]:

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

As pointed out in [42], in the case of nonnegative coefficients $\mathbf{X} = \mathbf{R}_+$ and assuming $k \leq n/2$ nonzeros in $\mathbf{x}_0$, there is a unique solution to $(P_{\mathbf{R}_+})$: $\mathbf{x}_0$. Consequently, the phase transition associated to this matrix must obey $\rho^* \geq 1/2$ for every $\delta$. Empirically, we do not observe this; we instead observe $\hat{\rho} \approx \rho^*(\delta|\mathbf{R}_+)$, exactly as with other matrices!

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 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 all matrices, the Theorem quoted from [10] notwithstanding.

5 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 which are known for the case of Gaussian random matrices. This is true for objects with coefficients over any of the sets $\{|0, 1|^N, \mathbf{R}_N^N, \mathbf{R}^N, \mathbf{C}^N\}$ when the convex optimization problem is appropriately matched, and the positions and signs of the nonzeros are randomly assigned.
Table 2: Evidence for the Positive Adaptivity Property. Note: in this table $\hat{\rho}^*(X_1, X_2)$ means: the phase transition observed when the solver assumes $x \in X_1^N$ while the object actually obeys $x_0 \in X_2^N$. Thus the notation $\hat{\rho}^*(R, R_+) \) means: the empirical phase transition 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. Standard errors of estimated phase transition yield 2SE error bars of width approximately 0.01; compare Figure 3.

| Label       | Name                  | $\delta$ | $\hat{\rho}^*(1/2|R)$ | $\hat{\rho}(R, R)$ | $\hat{\rho}(R, R_+)$ | $\hat{\rho}(R_+, R_+)$ | $\rho(1/2|R_+)$ |
|-------------|-----------------------|----------|------------------------|---------------------|----------------------|------------------------|----------------|
| SS          | Spikes & Sines        | 1/2      | 0.386                  | 0.3914              | 0.5621               | 0.5624                 | 0.558          |
| SH          | Spikes & Hadamard     | 1/2      | 0.380                  | 0.3942              | 0.5716               | 0.5687                 | 0.558          |
| SN          | Spikes & Noiselets    | 1/2      | 0.386                  | 0.3977              | 0.5541               | 0.5685                 | 0.558          |
| LC          | Linear Chirp Frame    | 1/2      | 0.380                  | 0.3865              | 0.5614               | 0.5576                 | 0.558          |
| AC          | Affine Plane Chirps   | 1/2      | 0.386                  | 0.3866              | 0.5609               | 0.5582                 | 0.558          |

Standard presentations of compressed sensing based on RIP suggest to practitioners that deterministic matrices for compressed sensing are still distant possibilities. 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 via convex optimization works well for certain deterministic measurement matrices – in fact just as well as for true random matrices. We show this for simple explicit deterministic matrices for which fast transforms are known.

Acknowledgments

This work was partially supported by NSF DMS 0906812 (ARRA). Part of the work was done when S. Jafarpour was a research scientist at Yahoo! Research. Sam Gross was on a VIGRE fellowship, and Youngsoo Choi was supported by the Army grant W911NF-07-0027-1. Kevin S. Raines was supported by NSF DGE 1147470 as well as the Stanford H&S Fellowship. We would also like to thank Michael Saunders for providing us with a pre-release version of the optimization software ASP.

[Matrix Specifications]

Spikes and Sines

The SS frame in the complex case is simply the $n$ by $2n$ matrix

$$A = [I_n F_n]$$

where $I_n$ the $n$ by $n$ identity matrix, $F_n$ is the $n$ by $n$ unitary discrete Fourier transform matrix, with entries $F_{i,j} = \exp\{2\pi \sqrt{-1} ij / n\} / \sqrt{n}$. In the real case, $A_{n,2n} = [IR_n]$ where $R_n$ is a real orthogonal $n$ by $n$ matrix whose columns
are each a normalized version of the real part or imaginary part of some corresponding column of $F_n$. The frame can be defined for any $n$. As noted in the text, we also considered these variants of real sinusoids: Hartley transform, DCT I,II,III,IV, with little change in results.

**Spikes and Hadamard**

The SH frame in all cases is simply the $n$ by $2n$ matrix

$$A = [I_n, H_n]$$

(4)

where $I_n$ the $n$ by $n$ identity matrix, and $H_n$ is the $n$ by $n$ real orthogonal discrete Hadamard transform matrix. In principle, a wide range of $n$ is possible for the Hadamard transform; in practice we used only dyadic $n$, $n = 2^j$, and Hadamard matrices defined recursively by

$$H = 2^{-1/2} \begin{bmatrix} H_{n/2} & H_{n/2} \\ H_{n/2} & -H_{n/2} \end{bmatrix}.$$  

**Spikes and Noiselets**

The SN frame in the complex case is the $n$ by $2n$ matrix

$$A = [I_n, W_n]$$

(5)

where $I_n$ the $n$ by $n$ identity matrix, and $W_n$ is the $n$ by $n$ unitary matrix representing the discrete Noiselet transform [43]. In the real case $W_n$ is replaced by an $n$ by $n$ real matrix $R_n$ whose columns are the non-redundant and normalized versions of the real and the imaginary parts of columns of the complex matrix $W_n$.

**Paley Tight Frame**

The Paley frame was defined by [16] et al. as follows. Let $p$ be an odd prime and $N = p + 1$. From the usual orthonormal discrete Fourier transform matrix $F_N$, number the rows starting at 0 and select only the rows corresponding to quadratic residues mod $p$. This will select $(p - 1)/2$ rows deterministically. Append to this matrix a row of constants.

**Grassmannian Frames**

The Grassmannian frames we use are defined in [14] as follows. Let $n \geq 5$ be a prime integer and let $a_t = \exp{2\pi\sqrt{-1}t^4/n}$ be the Alltop sequence, a quadratic-phase chirp. Let $A(\ell)$ denote the diagonal matrix with the $A(\ell)_{ii} = a_{i-\ell}$. Let $N = n \cdot L$ and let $F_n$ denote the usual orthonormal discrete Fourier transform matrix. Then $A$ is an $n \times N$ matrix defined by concatenating together the $L$ block matrices $A(\ell)F_n$:

$$A = [A(0)F_n | A(1)F_n | A(2)F_n | ... | A(L-1)F_n].$$

(6)

This frame is only considered for $n$ prime and $0 < L < n$; it is then equiangular.
Linear Chirp Frame
The linear chirp matrix we use was defined in [44]. Let $n$ be prime and let $c_ℓ^t = \exp\left\{2\pi\sqrt{-1}tℓ^2/n\right\}$ be the linear Chirp sequence with chirp rate $ℓ$. Let $C(ℓ)$ denote the diagonal matrix with the $C(ℓ)_{tt} = c_ℓ^t$. Let $N = n \cdot L$ and let $F_n$ denote the usual orthonormal discrete Fourier transform matrix. Then $A$ is an $n \times N$ matrix defined by concatenating together the $L$ block matrices $C(ℓ)F_n$:

$$A = [C(0)F_n|C(1)F_n|C(2)F_n| \ldots |C(L-1)F_n].$$

This frame is only considered for $n$ prime and $1 < L < n$; it is then tight and equiangular.

Delsarte-Goethals Frame
The Delsarte-Goethals (DG) frames we use are defined in [36, 15], who called them $DG(m, 0)$ or Kerdock frames. Let $m$ be an odd integer, $n = 2^m$ be dyadic and let $d_ℓ^t = (\sqrt{-1})^{i(ℓ)t}p_{i(ℓ)}$ be the ‘binary chirp’ based on the $m$ by $m$ binary symmetric matrix $P_t$ in the $DG(m, 0)$ set, where $i(ℓ)$ is the binary $m$-tuple bit vector encoding the value of $t \in \{0, \ldots, n-1\}$. Let $D(ℓ)$ denote the diagonal matrix with the $D(ℓ)_{tt} = d_ℓ^t$. Let $N = n \cdot L$ and let $H_n$ denote the usual orthonormal discrete Hadamard transform matrix. Then the complex-valued DG frame $A$ is an $n \times N$ matrix defined by concatenating together the $L$ block matrices $D(ℓ)H_n$:

$$A = [D(0)H_n|D(1)H_n|D(2)H_n| \ldots |D(L-1)H_n].$$

The real-valued DG frames are also defined in [15]. These are $2n$ by $N$ matrices that are obtained by applying the Gray map to the corresponding complex-valued DG frames. Both real and complex DG frames are only considered for odd integer $m$, and $1 < L < n$; They are equiangular tight frames.

Affine Plane Chirp Frame
Let $n = p^2$ where $p$ is prime and consider the affine plane $\mathbb{Z}_p^2$, with affine lines $L_{a,b} = \{(i, j) : j = a + bi \mod p\}$ for $a, b \in \mathbb{Z}_p$, with indicators $χ_{a,b}(i, j)$. Let $a_ℓ(i, j) = \exp\left\{\frac{2\pi\sqrt{-1}tℓ}{p}\right\}$ denote a spatially chirping sinusoid. Let $L_n$ denote the $n$ by $n-1$ matrix whose columns are all the distinct $vec(χ_{a,b})$ excluding $a = b = 0$, and let $A(ℓ)$ denote the $n$ by $n$ diagonal matrix with $vec(a_ℓ)$ along the diagonal. Construct $A$ via

$$A = [A(0)L_n|A(1)L_n|A(2)L_n| \ldots |A(L-1)L_n].$$

This frame is neither tight nor equiangular. It has sparse columns and rows, with only $p = \sqrt{n}$ non zeros in each, and coherence $n^{-1/4}$, asymptotically much larger than the optimal value $n^{-1/2}$. We made this construction independently, but perhaps the paper [45] constructs this frame as well; we are not able to decipher the discussion in that reference.
References


