COMPRESSED SENSING: HOW SHARP IS THE RESTRICTED ISOMETRY PROPERTY?

JEFFREY D. BLANCHARD*, CORALIA CARTIS†, AND JARED TANNER‡

Abstract. Compressed sensing is a recent technique by which signals can be measured at a rate proportional to their information content, combining the important task of compression directly into the measurement process. Since its introduction in 2004 there have been hundreds of manuscripts on compressed sensing, a large fraction of which have focused on the design and analysis of algorithms to recover a signal from its compressed measurements. The Restricted Isometry Property (RIP) has become a ubiquitous property assumed in their analysis. We present the best known bounds on the RIP, and in the process illustrate the way in which the combinatorial nature of compressed sensing is controlled. Our quantitative bounds on the RIP allow precise statements as to how aggressively a signal can be undersampled, the essential question for practitioners.

Key words. Compressed sensing, sparse approximation, restricted isometry property, phase transitions, convex relaxation, Gaussian matrices, singular values of random matrices.

AMS subject classifications. Primary: 41A46, 94A12, 94A20. Secondary: 15A52, 60F10, 90C25.

1. Introduction. Consider the task of measuring an unknown vector $x \in \mathbb{R}^N$ by taking inner products with vectors of one's choosing. The obvious choice would be to ask for the inner product of x with respect to each of the N canonical unit vectors e_j (the j^{th} entry of e_j being one and all others zero). But what if it is known a priori that x has only $k < N$ nonzero entries (referred to, henceforth, as k-sparse)? If the nonzero entries of x are indexed by the set K ($x(j) \neq 0$ if $j \in K$ and $x(j) = 0$ for $j \in K^c$), then only k inner products are needed, those with the canonical unit vectors e_j for $j \in K$. However, what if K is unknown? Is it still possible to make fewer than N measurements of x?

The affirmative resolution of this question dates back at least to the early 1940s with group testing of large populations for medical diseases with infection rates substantially below one half, [25, 32]. This early resolution is based on adaptive measurements, where the next measurement vector is selected after viewing all previous measurements. In fact, it has recently been shown that adaptivity is unnecessary. Namely, a matrix A of size $n \times N$ is fixed and then all n measurements are taken in parallel via $y = Ax$; further, x can be recovered from y and A using computationally efficient and stable algorithms. This realization is the essential core of the recent topic Compressed Sensing. Compressed Sensing (CS) proposes that one makes considerably fewer measurements than N by assuming that the signals of interest are compressible. Since the release of the seminal CS papers in 2004, [10, 6, 17], a great deal of excitement has been generated in signal processing and applied mathematics research, with hundreds of papers on the theory, applications, and extensions of compressed sensing (more than 400 of these are collected at Rice's online Compressive Sensing Resources archive `dsp.rice.edu/cs`). For a recent review of CS see [5].

In CS the matrix A and reconstruction algorithm are referred to as an encoder/decoder pair and much of the research has focused on their construction; that is, how should the measurement matrix A be selected and what are the most computationally efficient and robust algorithms for recovering x given y and A? The two most prevalent encoders in the literature are to construct A by populating its entries independently and identically from a Gaussian normal distribution, or to construct A by having its rows populated by random rows of a Fourier matrix. The most prevalent

*Department of Mathematics and Statistics, Grinnell College, Grinnell, Iowa, USA (`jeff@math.utah.edu`). JDB acknowledges support from NSF DMS (VIGRE) grant number 0602219 while a VIGRE postdoctoral fellow at Department of Mathematics, University of Utah.

†School of Mathematics, University of Edinburgh, Edinburgh, UK (`coralia.cartis@ed.ac.uk`).

‡School of Mathematics, University of Edinburgh, Edinburgh, UK (`jared.tanner@ed.ac.uk`). JT acknowledges support from the Alfred P. Sloan Foundation and thanks John E. and Marva M. Warnock for their generous support in the form of an endowed chair.

decoder has been ℓ^1-minimization,

$$\min_{z\in\mathbb{R}^N} \|z\|_1 \text{ subject to } Az = y, \tag{1.1}$$

which is the convex relaxation of the computationally intractable decoder, [36], which seeks the sparsest solution in agreement with the measurements

$$\min_{z\in\mathbb{R}^N} \|z\|_0 \text{ subject to } Az = y; \tag{1.2}$$

following the usual convention in the CS community, $\|z\|_0$ counts the number of nonzero entries in z. Many other encoder/decoder pairs are also being actively studied, with new alternatives being proposed regularly, see Section 3.

Here we do not review these exciting activities, but focus our attention on how to interpret the existing theoretical guarantees; in particular, for what (k, n, N) will an encoder/decoder pair recover the measured signal, or a suitable approximation thereof. To exemplify this, we restrict our attention to the encoder/decoder pair of A Gaussian and ℓ^1-minimization. These are selected due to both their prevalence in the literature and the existing wealth of knowledge in random matrix theory concerning Gaussian random matrices. Even further, we restrict our attention almost exclusively to the most widely used tool for analyzing the performance of encoder/decoder pairs, the *Restricted Isometry Property* (RIP) introduced by Candès and Tao [11].

Definition 1.1 (Restricted Isometry Property). *A matrix A of size $n \times N$ is said to satisfy the RIP with RIP constant $R(k, n, N; A)$ if, for every $x \in \chi^N(k) := \{x \in \mathbb{R}^N : \|x\|_0 \le k\}$,*

$$R(k, n, N; A) := \min_{c\ge 0} c \text{ subject to } (1-c)\|x\|_2^2 \le \|Ax\|_2^2 \le (1+c)\|x\|_2^2. \tag{1.3}$$

The RIP constant $R(k, n, N; A)$ is the maximum distance from 1 of all the eigenvalues of the $\binom{N}{k}$ submatrices, $A_K^T A_K$, derived from A, where K is an index set of cardinality k which restricts A to those columns indexed by K.

For many CS decoders it has been shown that if the RIP constants for the encoder remain bounded as n and N increases with $n/N \to \delta$, then the decoder can be guaranteed to recover the sparsest x for k up to some constant multiple of n, $\rho(\delta)\cdot n$; see [30, 37] and references therein. However, little more than $\rho(\delta) > 0$ is generally known. In this paper, we endeavour to be as precise as possible about the value of $\rho(\delta)$, because the range of allowable k tells practitioners how aggressively they may undersample.

To quantify the sparsity/undersampling trade off, we adopt a *proportional-growth* asymptotic, in which we consider sequences of triples (k, n, N) where all elements grow large in a coordinated way, $n \sim \delta N$ and $k \sim \rho n$ for some constants $\delta, \rho > 0$. This defines a two dimensional phase space (δ, ρ) in $[0, 1]^2$ for asymptotic analysis.

Definition 1.2 (Proportional-Growth Asymptotic). *A sequence of problem sizes (k, n, N) is said to follow* proportional growth *if, for $(\delta, \rho) \in [0, 1]^2$, $\frac{n}{N} \to \delta$ and $\frac{k}{n} \to \rho$ as $n \to \infty$.*

Ultimately, we want to determine, as precisely as possible, which subset of this phase space corresponds to successful recovery and which subset corresponds to unsuccessful recovery. This is the phase-transition framework advocated by Donoho et al [16, 18, 19, 20, 23]; see Section 3 for a precise definition. By translating the sufficient RIP conditions into the proportional-growth asymptotic, we find lower bounds on the phase-transition for (δ, ρ) in $[0, 1]^2$. An answer to this question plays the role of an *undersampling theorem*: to what degree can we undersample a signal and still be able to reconstruct it?

In this paper, we sharpen the use of the RIP and squeeze the most out of it, quantifying what can currently be said in the proportional-growth asymptotic and thereby making precise the undersampling theorems the RIP implies. We proceed in Section 2 along two main avenues. First, we concentrate on Gaussian matrices, using bounds on their singular values we develop the sharpest

known bounds on their RIP constants; in fact, these are the the best known bounds of any class of matrices in the proportional-growth asymptotic with $n < N$. These bounds are explicit and quantitative. Second, we use an asymmetric definition of the RIP where the lower and upper eigenvalues are treated separately, and in so further improve the conditions in which the RIP implies CS decoders recover the measured signal. In Section 3 we combine these two improvements to exhibit a region of the (δ, ρ) phase space where RIP analysis shows that undersampling will be successful for the ℓ^1-minimization decoder (1.1).

The RIP is not the only tool used to analyze the performance of compressed sensing decoders. The different methods of analysis lead to results that are rather difficult to compare. In Section 3.2, we describe the proportional-growth asymptotic, with A Gaussian and the ℓ^1-minimization decoder, to two alternative methods of analysis: Donoho's polytope analysis [16, 18, 22] and the geometric functional analysis techniques of Rudelson and Vershynin [39]. By translating these two methods of analysis and the RIP analysis into the proportional-growth asymptotic, we can readily compare the results obtained by these three techniques by examining the regions of the (δ, ρ) phase space where each method of analysis was shown to guarantee successful recovery.

2. Bounds on RIP for Gaussian Random Matrices. Let $K \subset \{1, \ldots, N\}$ be an index set of cardinality k which specifies the columns of A chosen for a submatrix, A_K, of size $n \times k$. Explicitly computing $R(k, n, N; A)$ would require enumerating all $\binom{N}{k}$ subsets K of the columns of A, forming each matrix $G_K = A_K^T A_K$, and calculating their largest and smallest eigenvalues. We have never seen this done except for small sizes of N and k, so not much is known about the RIP constants of deterministic matrices. Fortunately, analysis can penetrate where computation becomes intractable. Associated with a random matrix ensemble is an, as of yet unknown, probability density function for $R(k, n, N)$. Let us focus on the Gaussian ensemble where much is already known about its eigenvalues. We say that an $n \times N$ random matrix A is drawn from the *Gaussian ensemble* of random matrices if the entries are sampled independently and identically from the standard normal distribution, $\mathcal{N}(0, n^{-1})$. We say that a $k \times k$ matrix $W_{n,k}$ is a *Wishart* matrix if it is the Gram matrix $X^T X$ of an $n \times k$ matrix X from the Gaussian ensemble. The largest and smallest eigenvalues of a Wishart matrix are random variables, denoted here $\Lambda_{n,k}^{max} = \lambda^{max}(W_{n,k})$ and $\Lambda_{n,k}^{min} = \lambda^{min}(W_{n,k})$. These random variables tend to defined limits, in expectation, as n and k increase in a proportional manner. With $\frac{k}{n} \to \rho$ as $n \to \infty$, we have $\mathcal{E}(\Lambda_{n,k}^{max}) \to (1 + \sqrt{\rho})^2$ and $\mathcal{E}(\Lambda_{n,k}^{min}) \to (1 - \sqrt{\rho})^2$; [31, 41], see Figure 2.1. Explicit formulas bounding $\Lambda_{n,k}^{max}$ and $\Lambda_{n,k}^{min}$ are available [28]. An empirical approximation of the probability density functions of $\Lambda_{n,k}^{max}$ and $\Lambda_{n,k}^{min}$ is shown in Figure 2.2.

The asymmetric deviation from 1 of the expected eigenvalues of $W_{n,k} = A_K^T A_K$ suggests that the symmetric definition of the RIP is unnecessarily restrictive. We generalize the RIP to an asymmetric form in order to more precisely quantify the RIP and derive the most favorable recovery conditions for compressed sensing decoders.

Definition 2.1 (Asymmetric Restricted Isometry Property). *For a matrix A of size $n \times N$, the* asymmetric RIP constants *$L(k, n, N; A)$ and $U(k, n, N; A)$ are defined as:*

$$L(k, n, N; A) := \min_{c \geq 0} c \ \textit{subject to} \ (1 - c)\|x\|_2^2 \leq \|Ax\|_2^2, \quad \textit{for all} \quad x \in \chi^N(k); \tag{2.1}$$

$$U(k, n, N; A) := \min_{c \geq 0} c \ \textit{subject to} \ (1 + c)\|x\|_2^2 \geq \|Ax\|_2^2, \quad \textit{for all} \quad x \in \chi^N(k). \tag{2.2}$$

(A similar change in the definition of the RIP constants was used independently by Foucart and Lai in [30], motivated by different concerns.)

Remark 1. *Although both the smallest and largest singular values of $A_K^T A_K$ affect the stability of the reconstruction algorithms, the smaller eigenvalue is dominant for compressed sensing in that it allows distinguishing between sparse vectors from their measurement by A. In fact, it is often incorrectly stated that $R(2k, n, N) < 1$ is a necessary condition where it is actually $L(2k, n, N) < 1$ that is a necessary condition to ensure that there are no two k-sparse vectors, say x and x', with $Ax = Ax'$.*

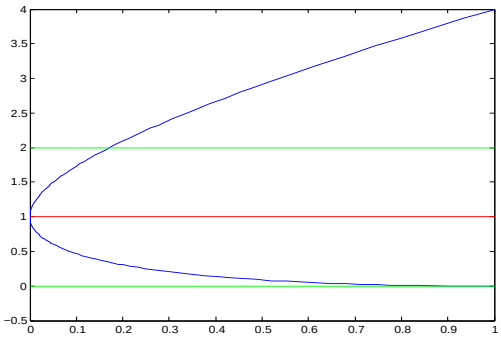

FIG. 2.1. *Expected values of the largest and smallest eigenvalues of a Wishart matrix $W_{n,k}$ with $\rho = \frac{k}{n}$. Note the asymmetry with respect to 1.*

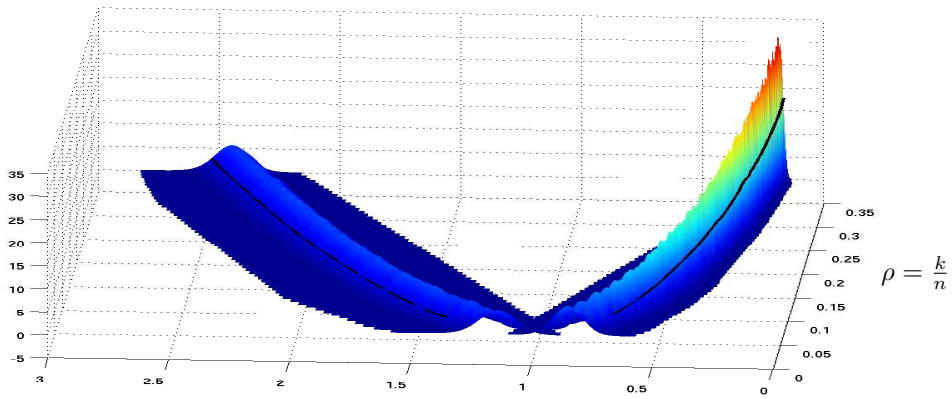

FIG. 2.2. **Empirical Distributions of the Largest and Smallest Eigenvalues of a Wishart Matrix.** *A collection of frequency histograms of $\Lambda_{n,k}^{max}$ and $\Lambda_{n,k}^{min}$: x-axis – size of the eigenvalue; y-axis – number of occurrences; z-axis – ratio $\rho = \frac{k}{n}$ of the Wishart parameters. Overlays: curves depicting the expected values $(1 \pm \sqrt{\rho})^2$ of $\Lambda_{n,k}^{max}$ and $\Lambda_{n,k}^{min}$. Here $n = 200$. At this value of n it is evident that $\Lambda_{n,k}^{max}$ and $\Lambda_{n,k}^{min}$ lie near, but not on curves. For larger n, the concentration would be tighter.*

We see from (2.1) and (2.2) that $(1 - L(k,n,N)) = \min_K \lambda^{min}(G_K)$ and $(1 + U(k,n,N)) = \max_K \lambda^{max}(G_K)$ with $G_K = A_K^T A_K$. A standard large deviation analysis of bounds on the probability density functions for the $\Lambda_{n,k}^{max}$ and $\Lambda_{n,k}^{min}$ allows us to establish upper bounds on the typical size of $L(k,n,N)$ and $U(k,n,N)$ for the Gaussian ensemble in the proportional-growth asymptotic.

DEFINITION 2.2 (Asymptotic RIP Bounds). *Let A be a matrix of size $n \times N$ drawn from the Gaussian ensemble and consider the proportional-growth asymptotic ($\frac{n}{N} \to \delta$ and $\frac{k}{n} \to \rho$ as $n \to \infty$). Let $H(p) := p\log(1/p) + (1-p)\log(1/(1-p))$ denote the usual Shannon Entropy with base e logarithms, and let*

$$\psi_{min}(\lambda, \rho) := H(\rho) + \frac{1}{2}\left[(1-\rho)\log\lambda + 1 - \rho + \rho\log\rho - \lambda\right], \tag{2.3}$$

$$\psi_{max}(\lambda, \rho) := \frac{1}{2}\left[(1+\rho)\log\lambda + 1 + \rho - \rho\log\rho - \lambda\right]. \tag{2.4}$$

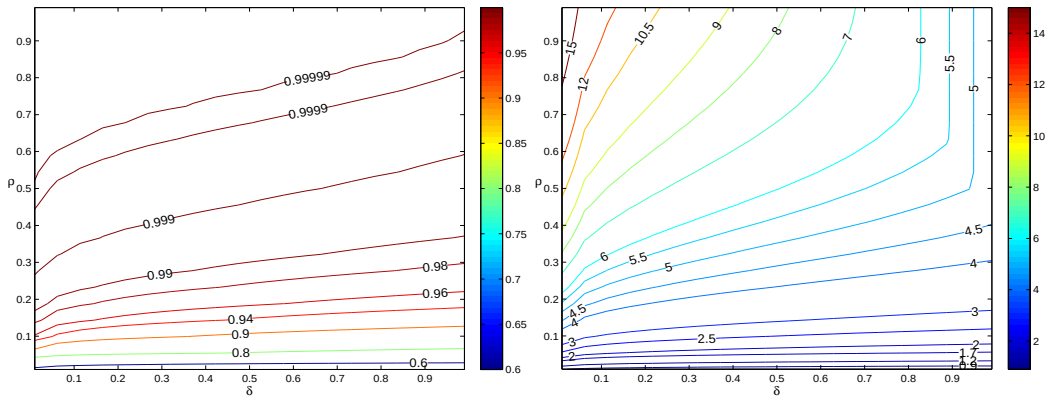

FIG. 2.3. **The RIP bounds of Eq.** (2.7)**.** *Level sets of $L(\delta,\rho)$ (left panel) and $U(\delta,\rho)$ (right panel) over the phase space $(\delta,\rho)\in[0,1]^2$. For large matrices from the Gaussian ensemble, it is overwhelmingly unlikely that the RIP constants $L(k,n,N;A)$ and $U(k,n,N;A)$ will be greater than these values.*

Define $\lambda^{min}(\delta,\rho)$ and $\lambda^{max}(\delta,\rho)$ as the solution to (2.5) *and* (2.6)*, respectively:*

$$\delta\psi_{min}(\lambda^{min}(\delta,\rho),\rho)+H(\rho\delta)=0 \quad \textit{for} \quad \lambda^{min}(\delta,\rho)\le 1-\rho \tag{2.5}$$

$$\delta\psi_{max}(\lambda^{max}(\delta,\rho),\rho)+H(\rho\delta)=0 \quad \textit{for} \quad \lambda^{max}(\delta,\rho)\ge 1+\rho. \tag{2.6}$$

Define $L(\delta,\rho)$ and $U(\delta,\rho)$ as

$$L(\delta,\rho):=1-\lambda^{min}(\delta,\rho) \quad \textit{and} \quad U(\delta,\rho):=\min_{\nu\in[\rho,1]}\lambda^{max}(\delta,\nu)-1. \tag{2.7}$$

The asymptotic bounds $L(\delta,\rho)$ and $U(\delta,\rho)$ are explicit and control the size of the asymmetric RIP constants $L(k,n,N)$ and $U(k,n,N)$ when A is drawn from the Gaussian ensemble. In the proportional growth asymptotic, the probability that these bounds are valid tends to 1 as $n\to\infty$. In fact, all probabilities presented in this manuscript converge to their limit "exponentially in n"; that is, the probability for finite n approaches its limit as n grows with discrepancy bounded by a constant multiple of $e^{-n\beta}$ for some fixed $\beta>0$.

THEOREM 2.3 (Validity of RIP Bounds). *Fix $\epsilon>0$. Under the proportional-growth asymptotic, Definition 1.2, sample each $n\times N$ matrix A from the Gaussian ensemble. Then*

$$Prob\left(L(k,n,N;A)<L\left(\delta,\rho\right)+\epsilon\right)\to 1 \quad \textit{and} \quad Prob\left(U(k,n,N;A)<U\left(\delta,\rho\right)+\epsilon\right)\to 1$$

exponentially in n.

Extensive empirical estimates of $L(k,n,N)$ and $U(k,n,N)$ show that these bounds are rather sharp, in fact they are at least within 2 times the actual upper bounds on $L(k,n,N)$ and $U(k,n,N)$, see Figure 2.4, and are much closer for the region applicable for CS decoders, $\rho\ll 1$.

2.1. Proof of Theorem 2.3. To prove Theorem 2.3, we first establish some useful lemmas concerning the extreme eigenvalues of Wishart matrices. The matrix A generates $\binom{N}{k}$ different Wishart matrices $G_k=A_K^TA_K$. Exponential bounds on the tail probabilities of the largest and smallest eigenvalues of such Wishart matrices can be combined with exponential bounds on $\binom{N}{k}$ to control the chance of large deviations using the union bound. This large deviation analysis technique is characteristic of proofs in compressed sensing. By using the exact probability density functions on the tail behavior of the extreme eigenvalues of Wishart matrices the overestimation of the union bound is dramatically reduced. We focus on the slightly more technical results for the

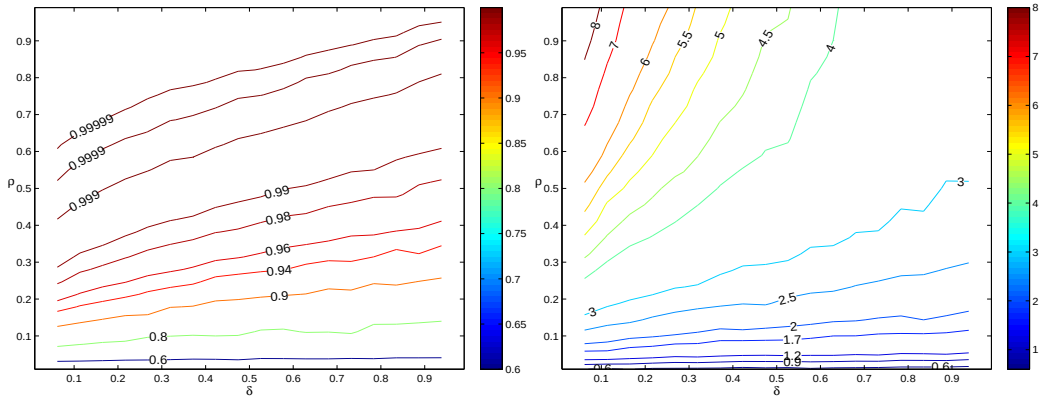

FIG. 2.4. **Empirically observed lower estimates of RIP bounds of RIP constants.** *Although there is no computationally tractable method for calculating the RIP constants of a matrix, there are efficient algorithms which perform local searches for extremal eigenvalues of submatrices; allowing for observable* lower *bounds on the RIP constants. Algorithm for observing $L(k,n,N)$, [26], and $U(k,n,N)$, [35], were applied to dozens of A drawn Gaussian $\mathcal{N}(0,n^{-1})$ with $n = 400$ and N increasing from 420 to 8000. Level sets of the observed $L(k,n,N;A)$ (left panel) and $U(k,n,N;A)$ (right panel).*

bound on the most extreme of the largest eigenvalues, $U(\delta,\rho)$, and prove these statements in full detail. Corresponding results for $L(\delta,\rho)$ are stated with their similar proofs omitted.

The probability density function, $f_{max}(k,n;\lambda)$, for the largest eigenvalue of the $k \times k$ Wishart matrix $A_K^T A_K$ was determined by Edelman in [27]. For our analysis, a simplified upper bound suffices.

LEMMA 2.4 (Lemma 4.2, pp. 550 [27]). *Let A_K be a matrix of size $n \times k$ whose entries are drawn i.i.d from $\mathcal{N}(0,n^{-1})$. Let $f_{max}(k,n;\lambda)$ denote the probability density function for the largest eigenvalue of the Wishart matrix $A_K^T A_K$ of size $k \times k$. Then $f_{max}(k,n;\lambda)$ satisfies:*

$$f_{max}(k,n;\lambda) \leq \left[(2\pi)^{1/2}(n\lambda)^{-3/2}\left(\frac{n\lambda}{2}\right)^{(n+k)/2}\frac{1}{\Gamma(\frac{k}{2})\Gamma(\frac{n}{2})}\right]\cdot e^{-n\lambda/2} =: g_{max}(k,n;\lambda). \tag{2.8}$$

For our purposes, it is sufficient to have a precise characterization of $g_{max}(k,n;\lambda)$'s exponential (with respect to n) behavior.

LEMMA 2.5. *Let $k/n = \rho \in (0,1)$ and define*

$$\psi_{max}(\lambda,\rho) := \frac{1}{2}\left[(1+\rho)\log\lambda + 1 + \rho - \rho\log\rho - \lambda\right].$$

Then

$$f_{max}(k,n;\lambda) \leq p_{max}(n,\lambda)\exp(n\cdot\psi_{max}(\lambda,\rho)) \tag{2.9}$$

where $p_{max}(n,\lambda)$ is a polynomial in n,λ.

Proof. Let $g_{max}(k,n;\lambda)$ be as defined in (2.8) and let $\rho_n = k/n$. To extract the exponential behavior of $g_{max}(k,n;\lambda)$ we write $\frac{1}{n}\log(g_{max}(k,n;\lambda)) = \Phi_1(k,n;\lambda)+\Phi_2(k,n;\lambda)+\Phi_3(k,n;\lambda)$ where

$$\begin{aligned}
\Phi_1(k,n;\lambda) &= \frac{1}{2n}\log(2\pi) - \frac{3}{2n}\log(n\lambda)\\
\Phi_2(k,n;\lambda) &= \frac{1}{2}\left[(1+\rho_n)\log\left(\frac{\lambda n}{2}\right) - \lambda\right]\\
\Phi_3(k,n;\lambda) &= -\frac{1}{n}\log\left(\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{n}{2}\right)\right).
\end{aligned}$$

Clearly, $\lim_{n\to\infty}\Phi_1(k,n;\lambda)=0$ and can be subsumed as part of $p_{max}(n,\lambda)$. To simplify Φ_3, we apply the second of Binet's log gamma formulas [44], namely $\log(\Gamma(z))=(z-1/2)\log z - z + \log\sqrt{2\pi} + I$ where I is a convergent, improper integral. With $c(n,\rho)$ representing the constant and integral from Binet's formula we then have

$$\Phi_2(k,n;\lambda)+\Phi_3(k,n;\lambda)=\frac{1}{2}\left[(1+\rho_n)\log\lambda-\left(\rho_n-\frac{1}{n}\right)\log\rho_n+\frac{2}{n}\log\frac{n}{2}+\rho_n+1-\lambda+\frac{1}{n}c(n,\rho_n)\right].$$

As $\lim_{n\to\infty} n^{-1}c(n,\rho_n)=0$ it can be absorbed into $p_{max}(n,\lambda)$ and we have

$$\psi_{max}(\lambda,\rho):=\lim_{n\to\infty}\frac{1}{n}\log\left[g_{max}(k,n;\lambda)\right]=\frac{1}{2}\left[(1+\rho)\log\lambda-\rho\log\rho+\rho+1-\lambda\right]$$

and the conclusion follows. □

To bound $U(k,n,N)$, we must simultaneously account for all $\binom{N}{k}$ Wishart matrices $A_K^TA_K$ derived from A. Using a union bound this amounts to studying the exponential behavior of $\binom{N}{k}g_{max}(k,n;\lambda)$. In the proportional-growth asymptotic this can be determined by characterizing $\lim_{N\to\infty}N^{-1}\log\left[\binom{N}{k}g_{max}(k,n;\lambda)\right]$, which from Lemma 2.5 is given by

$$\begin{aligned}\lim_{N\to\infty}\frac{1}{N}\log\left[\binom{N}{k}g_{max}(k,n;\lambda)\right]&=\lim_{N\to\infty}\frac{1}{N}\log\left[\binom{N}{k}\right]+\lim_{N\to\infty}\frac{1}{N}\log\left[g_{max}(n,k;\lambda)\right]\\&=H\left(\frac{k}{N}\right)+\delta\lim_{n\to\infty}\frac{1}{n}\log\left[g_{max}(n,k;\lambda)\right]\\&=H(\rho\delta)+\delta\psi_{max}(\lambda,\rho)=:\delta\psi_U(\delta,\rho;\lambda).\end{aligned}\tag{2.10}$$

Recall that $H(p):=p\log(1/p)+(1-p)\log(1/(1-p))$ is the usual Shannon Entropy with base e logarithms.

Equipped with Lemma 2.5 and (2.10), Proposition 2.6 establishes $\lambda^{max}(\delta,\rho)-1$ as an upper bound on $U(k,n,N)$ in the proportional-growth asymptotic.

PROPOSITION 2.6. *Let $\delta,\rho\in(0,1)$, and A be a matrix of size $n\times N$ whose entries are drawn i.i.d. from $\mathcal{N}(0,n^{-1})$. Define $\tilde{U}(\delta,\rho):=\lambda^{max}(\delta,\rho)-1$ where $\lambda^{max}(\delta,\rho)$ is the solution to (2.6). Then for any $\epsilon>0$, in the proportional-growth asymptotic*

$$Prob\left(U(k,n,N)>\tilde{U}(\delta,\rho)+\epsilon\right)\to 0$$

exponentially in n.

Proof. Throughout this proof δ and ρ are fixed, and we focus our attention on λ, often abbreviating $\psi_U(\delta,\rho;\lambda)$ in (2.10) as $\psi_U(\lambda)$. We first verify that (2.6) has a unique solution. Since

$$\frac{d}{d\lambda}\psi_U(\lambda)=\frac{1}{2}\left(\frac{1+\rho}{\lambda}-1\right),$$

$\psi_U(\lambda)$ is strictly decreasing on $[1+\rho,\infty)$ and is strictly concave. Combined with

$$\psi_U(1+\rho)=\delta^{-1}H(\rho\delta)+\frac{1}{2}\left[(1+\rho)\log(1+\rho)+\rho\log\frac{1}{\rho}\right]>0$$

and $\lim_{\lambda\to\infty}\psi_U(\lambda)=-\infty$, there is a unique solution to (2.6), namely $\lambda^{max}(\delta,\rho)$.

Select $\epsilon>0$ and let (k,n,N) be such that $\frac{n}{N}=\delta_n$, $\frac{k}{n}=\rho_n$. First, we write the probability

statement in terms of $\lambda^{max}(\delta_n, \rho_n)$:

$$\begin{aligned} Prob\left[U(k,n,N) > \tilde{U}(\delta_n,\rho_n)+\epsilon)\right] &= Prob\left[U(k,n,N) > \lambda^{max}(\delta_n,\rho_n) - 1 + \epsilon)\right] \\ &= Prob\left[1 + U(k,n,N) > \lambda^{max}(\delta_n,\rho_n) + \epsilon)\right] \\ &= \binom{N}{k} \int_{\lambda^{max}(\delta_n,\rho_n)+\epsilon}^{\infty} f_{max}(k,n;\lambda) d\lambda \\ &\leq \binom{N}{k} \int_{\lambda^{max}(\delta_n,\rho_n)+\epsilon}^{\infty} g_{max}(k,n;\lambda) d\lambda. \end{aligned} \tag{2.11}$$

To bound the integral in (2.11) in terms of $g_{max}(\delta,\rho;\lambda^{max}(\delta_n,\rho_n))$ we write $g_{max}(k,n;\lambda)$ in terms of n, ρ_n, and λ as $g_{max}(k,n;\lambda) = \varphi(n,\rho_n)\lambda^{-\frac{3}{2}}\lambda^{\frac{n}{2}(1+\rho_n)}e^{-\frac{n}{2}\lambda}$ where

$$\varphi(n,\rho_n) = (2\pi)^{\frac{1}{2}} n^{-\frac{3}{2}} \left(\frac{n}{2}\right)^{\frac{n}{2}(1+\rho_n)} \frac{1}{\Gamma\left(\frac{n}{2}\rho_n\right)\Gamma\left(\frac{n}{2}\right)}.$$

Since $\lambda^{max}(\delta_n,\rho_n) > 1+\rho_n$, the quantity $\lambda^{\frac{n}{2}(1+\rho_n)}e^{-\frac{n}{2}\lambda}$ is strictly decreasing in λ on $[\lambda^{max}(\delta,\rho_n),\infty)$. Therefore we have

$$\begin{aligned} \int_{\lambda^{max}(\delta_n,\rho_n)+\epsilon}^{\infty} g_{max}(k,n;\lambda)d\lambda &\leq \varphi(n,\rho_n)\left(\lambda^{max}(\delta_n,\rho_n)+\epsilon\right)^{\frac{n}{2}(1+\rho_n)} e^{-\frac{n}{2}(\lambda^{max}(\delta_n,\rho_n)+\epsilon)} \int_{\lambda^{max}(\delta_n,\rho_n)+\epsilon}^{\infty} \lambda^{-\frac{3}{2}} d\lambda \\ &= \left(\lambda^{max}(\delta_n,\rho_n)+\epsilon\right)^{\frac{3}{2}} g_{max}\left(k,n;\lambda^{max}(\delta_n,\rho_n)+\epsilon\right) \int_{\lambda^{max}(\delta_n,\rho_n)+\epsilon}^{\infty} \lambda^{-\frac{3}{2}} d\lambda \\ &= 2\left(\lambda^{max}(\delta_n,\rho_n)+\epsilon\right) g_{max}\left(k,n;\lambda^{max}(\delta_n,\rho_n)+\epsilon\right). \end{aligned} \tag{2.12}$$

Therefore, combining (2.11) and (2.12) we obtain

$$\begin{aligned} Prob\left[U(k,n,N) > \tilde{U}(\delta_n,\rho_n)+\epsilon)\right] &\leq 2\left(\lambda^{max}(\delta_n,\rho_n)+\epsilon\right)\binom{N}{k} g_{max}\left(k,n;\lambda^{max}(\delta_n,\rho_n)+\epsilon\right) \\ &\leq p_{max}\left(n,\lambda^{max}(\delta_n,\rho_n)\right)\exp\left[n\cdot\psi_U\left(\lambda^{max}(\delta_n,\rho_n)+\epsilon\right)\right] \\ &\leq p_{max}\left(n,\lambda^{max}(\delta_n,\rho_n)\right)\exp\left[n\epsilon\cdot\frac{d}{d\lambda}\psi_U(\lambda)|_{\lambda=(\lambda^{max}(\delta_n,\rho_n))}\right], \end{aligned} \tag{2.13}$$

with the last inequality following from the strict concavity of $\psi_U(\lambda)$. Since $\frac{d}{d\lambda}\psi_U\left(\lambda^{max}(\delta,\rho)\right) < 0$ is strictly bounded away from zero and $\lim_{n\to\infty}\lambda^{max}(\delta_n,\rho_n) = \lambda^{max}(\delta,\rho)$, we arrive at, for any $\epsilon > 0$

$$\lim_{n\to\infty} Prob\left[U(k,n,N) > \tilde{U}(\delta,\rho)+\epsilon)\right] \to 0.$$

□

The term $H(\rho\delta)$ in (2.10), from the union bound over all $\binom{N}{k}$ matrices $A_K^T A_K$, results in an overly pessimistic bound in the vicinity of $\rho\delta = 1/2$. As we are seeking the least upper bound on $U(k,n,N)$ we note that any upper bound for $U(j,n,N)$ for $j > k$ is also an upper bound for $U(k,n,N)$, and replace the bound $\tilde{U}(\delta,\rho)$ with the minimum of $\tilde{U}(\delta,\nu)$ for $\nu \in [\rho,1]$.

PROPOSITION 2.7. *Let $\delta,\rho \in (0,1)$, and define $U(\delta,\rho) := \min_{\nu\in[\rho,1]}\tilde{U}(\delta,\nu)$ with $\tilde{U}(\delta,\nu)$ defined as in Proposition 2.6. For any $\epsilon > 0$, in the proportional-growth asymptotic*

$$Prob\left(U(k,n,N) > U(\delta,\rho)+\epsilon\right) \to 0$$

exponentially in n.

Proof. By the definition of $\chi^N(k)$ in Definition 1.1, $U(j,n,N) \geq U(k,n,N)$ for $j = k+1, k+2,\ldots,n$; combined with Proposition 2.6 for $\frac{j}{n} \to \nu$ as $n\to\infty$

$$Prob\left(U(j,n,N) > U(\delta,\nu)+\epsilon\right) \to 0$$

exponentially in n, and taking a minimum over the compact set $\nu \in [\rho, 1]$ we arrive at the desired result. □

A similar approach leads to corresponding results for $L(\delta, \rho)$. Edelman also determined the probability density function, $f_{min}(k, n; \lambda)$, for the smallest eigenvalue of the $k \times k$ Wishart matrix $A_K^T A_K$ [27]. Here again, a simplified upper bound suffices:

LEMMA 2.8 (Prop. 5.2, pp. 553 [27]). *Let A_K be a matrix of size $n \times k$ whose entries are drawn i.i.d. from $\mathcal{N}(0, n^{-1})$. Let $f_{min}(k, n; \lambda)$ denote the probability density function for the smallest eigenvalue of the Wishart matrix $A_K^T A_K$ of size $k \times k$. Then $f_{min}(k, n; \lambda)$ satisfies:*

(2.14)
$$f_{min}(k, n; \lambda) \leq \left(\frac{\pi}{2n\lambda}\right)^{1/2} \cdot e^{-n\lambda/2} \left(\frac{n\lambda}{2}\right)^{(n-k)/2} \cdot \left[\frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{k}{2})\Gamma(\frac{n-k+1}{2})\Gamma(\frac{n-k+2}{2})}\right] =: g_{min}(k, n; \lambda).$$

With Lemma 2.8, we establish a bound on the asymptotic behavior of the distribution of the smallest eigenvalue of Wishart matrix of size $k \times k$.

LEMMA 2.9. *Let $k/n = \rho \in (0, 1)$ and define*

$$\psi_{min}(\lambda, \rho) := H(\rho) + \frac{1}{2}\left[(1-\rho)\log\lambda + 1 - \rho + \rho\log\rho - \lambda\right].$$

Then

$$f_{min}(k, n; \lambda) \leq p_{min}(n, \lambda) \exp(n \cdot \psi_{min}(\lambda, \rho)) \tag{2.15}$$

where $p_{min}(n, \lambda)$ is a polynomial in n, λ.

With Lemma 2.9, the large deviation analysis yields

$$\lim_{N\to\infty} \frac{1}{N} \log\left[\binom{N}{k} g_{min}(k, n; \lambda)\right] = H(\rho\delta) + \delta\psi_{min}(\lambda, \rho). \tag{2.16}$$

Similar to the proof of Proposition 2.6, Lemma 2.9 and (2.16) are used to establish $L(\delta, \rho)$ as an upper bound on $L(k, n, N)$ in the proportional-growth asymptotic.

PROPOSITION 2.10. *Let $\delta, \rho \in (0, 1]$, and A be a matrix of size $n \times N$ whose entries are drawn i.i.d. from $\mathcal{N}(0, n^{-1})$. Define $L(\delta, \rho) := 1 - \lambda^{min}(\delta, \rho)$ where $\lambda^{min}(\delta, \rho)$ is the solution to (2.5). Then for any $\epsilon > 0$, in the proportional-growth asymptotic*

$$Prob\left(L(k, n, N) > L(\delta, \rho) + \epsilon\right) \to 0$$

exponentially in n.

The bound $L(\delta, \rho)$ is strictly increasing in ρ for any $\delta \in (0, 1)$, and as a consequence no tighter bound can be achieved by minimizing over matrices of larger size as was done in Proposition 2.7.

3. RIP Undersampling Theorems. The high level of interest in compressed sensing is due to the introduction of computationally efficient and stable algorithms which provably solve the seemingly intractable (1.2) even for k proportional to n. New compressed sensing decoders are being introduced regularly; broadly speaking, each fall into one of two categories: greedy algorithms and regularizations. Greedy algorithms are iterative, with each step selecting a locally optimal subset of entries in x which are adjusted to improve the desired error metric. Examples of greedy algorithms include Orthogonal Matching Pursuit (OMP) [42], Regularized OMP (ROMP) [38], Stagewise OMP (StOMP) [24], Compressive Sampling MP (CoSaMP) [37], Subspace Pursuit (SP) [15], and Iterated Hard Thresholding (IHT) [4]. Regularization formulations for sparse approximation began with the relaxation of (1.2) to the now ubiquitous (convex) ℓ^1-minimization [14], (1.1), and has since been extended to non-convex ℓ^q-minimization for $q \in (0, 1)$, [30, 13, 12, 40]. Although general-purpose convex optimization solvers may be employed to solve ℓ^1-minimization (1.1), highly-efficient software

has been recently designed specifically for ℓ^1-minimization in the context of compressed sensing, see [14, 29, 43, 45]. Non-convex formulations offer substantial improvements, but at the cost of limited guarantees that the global minima can be found efficiently.

One of the central aims of this article is to advocate a unifying framework for the comparison of results in compressed sensing. Currently there is no general agreement in the compressed sensing community on such a framework, making it difficult to compare results obtained by different methods of analysis or to identify when new results are improvements over existing ones. Donoho has put forth the phase transition framework borrowed from the statistical mechanics literature and used successfully in a similar context by the combinatorial optimization community, see [33, 34]. This framework has been successfully employed in compressed sensing by Donoho et al, [19, 20, 23].

Fortunately, every compressed sensing algorithm that has an optimal recovery order of n proportional to k can be cast in the phase transition framework of Donoho et al., parametrized by two inherent problem size parameters[1]:

- the *undersampling rate* of measuring x through n inner products with the rows of A, as compared to directly sampling each element of $x \in \mathbb{R}^N$:

$$\delta_n = n/N \in (0,1)$$

- the *oversampling rate* of making n measurements as opposed to the optimal *oracle* rate of making k measurements when the oracle knows the support of x:

$$\rho_n = k/n \in (0,1).$$

For each value of $\delta_n \in (0,1)$ there is a largest value of ρ_n which guarantees successful recovery of x.

We now formalize the phase transition framework described above.

DEFINITION 3.1 (Strong Equivalence). *The event StrongEquiv(A, alg) denotes the following property of an $n \times N$ matrix A: for* every *k-sparse vector x, the algorithm "alg" exactly recovers x from the corresponding measurements $y = Ax$.*

For most compressed sensing algorithms and for a broad class of matrices, under the proportional-growth asymptotic there is a strictly positive function $\rho_S(\delta) > 0$ defining a region of the (δ, ρ) phase space which ensures successful recovery of every k-sparse vector $x \in \chi^N(k)$. This function, $\rho_S(\delta)$, is called the *Strong phase transition* function [10, 16, 18].

DEFINITION 3.2 (Region of Strong Equivalence). *Consider the proportional-growth asymptotic with parameters $(\delta, \rho) \in (0,1) \times (0, 1/2)$. Draw the corresponding $n \times N$ matrices A from the Gaussian ensemble and fix $\epsilon > 0$. Suppose that we are given a function $\rho_S(\delta)$ with the property that, whenever $0 < \rho < (1-\epsilon)\rho_S(\delta)$, Prob(StrongEquiv(A, alg)) $\to 1$ as $n \to \infty$. We say that $\rho_S(\delta)$ bounds a* region of strong equivalence.

REMARK 2. *The subscript S emphasizes that the phase transition function $\rho_S(\delta)$ will define a region of the (δ, ρ) phase space which guarantees that the event StrongEquiv(A) is satisfied with high probability on the draw of A. This notation has been established in the literature by Donoho [16, 20] to distinguish strong equivalence (that every k-sparse vector x is successfully recovered) from weaker notions of equivalence (such as all but a small fraction of k-sparse vectors are successfully recovered). For example, [16, 20] study the event where ℓ^1-minimization* (1.1) *exactly recovers x from the corresponding measurements $y = Ax$, except for a fraction $(1-\epsilon)$ of the support sets.*

For the remainder of this section, we translate existing guarantees of $StrongEquiv(A, \ell^1)$ into bounds on the region of strong equivalence in the proportional-growth asymptotic. A similar presentation of other CS decoders is available in [3]. In order to make quantitative statements, the matrix, or random matrix ensemble must first be specified, [2], we again consider A drawn from

[1] For some algorithms, such as ℓ^1-regularization, these two parameters fully characterize the behavior of the algorithm for a particular matrix ensemble, whereas for other algorithms, such as OMP, the distribution of the nonzero coefficients also influences the behavior of the method.

the Gaussian ensemble.[2] In Section 3.1 we demonstrate how to incorporate the RIP bounds from Section 2 into results obtained from an RIP analysis. In Section 3.2 we compare bounds on the region of $StrongEquiv(A,\ell^1)$ proven by three distinct methods of analysis: eigenvalue analysis and the RIP [30], geometric functional analysis [39], and convex polytopes [16].

3.1. Region of $StrongEquiv(A,\ell^1)$ implied by the RIP. In this section, we incorporate the bounds on RIP constants established in Section 2 into a known condition implying $StrongEquiv(A,\ell^1)$ obtained from an RIP analysis. Following the pioneering work of Candès, Romberg, and Tao [8, 11], many different conditions on the RIP constants have been developed which ensure recovery of every k-sparse vector via ℓ^1-minimization, [6, 7, 9, 10, 39] to name a few. The current state of the art RIP conditions for ℓ^1-minimization were developed by Foucart and Lai [30].

THEOREM 3.3 (Foucart and Lai [30]). *For any matrix A of size $n \times N$ with RIP constants $L(2k,n,N)$ and $U(2k,n,N)$, for $2k \le n < N$. Define*

$$\mu^{FL}(k,n,N) := \frac{1+\sqrt{2}}{4}\left(\frac{1+U(2k,n,N)}{1-L(2k,n,N)} - 1\right). \tag{3.1}$$

If $\mu^{FL}(k,n,N) < 1$, then there is $StrongEquiv(A,\ell^1)$.

To translate this result into the phase transition framework for matrices from the Gaussian ensemble, we employ the RIP bounds (2.7) to the asymmetric RIP constants $L(2k,n,N)$ and $U(2k,n,N)$. It turns out that naively inserting these bounds into (3.1) yields a bound on $\mu^{FL}(k,n,N)$, see Lemma 3.6, and provides a simple way to obtain a bound on the region of strong equivalence.

DEFINITION 3.4 (RIP Region of $StrongEquiv(A,\ell^1)$). *Define*

$$\mu^{FL}(\delta,\rho) := \frac{1+\sqrt{2}}{4}\left(\frac{1+U(\delta,2\rho)}{1-L(\delta,2\rho)} - 1\right) \tag{3.2}$$

and $\rho_S^{FL}(\delta)$ as the solution to $\mu^{FL}(\delta,\rho) = 1$.

The function $\rho_S^{FL}(\delta)$ is displayed as the red curve in Figure 3.1.

THEOREM 3.5. *Fix $\epsilon > 0$. Consider the proportional-growth asymptotic, Definition 1.2, with parameters $(\delta,\rho) \in (0,1)\times(0,1/2)$. Draw the corresponding $n \times N$ matrices A from the Gaussian ensemble. If $\rho < (1-\epsilon)\rho_S^{FL}(\delta)$, then $Prob(StrongEquiv(A,\ell^1)) \to 1$ as $n \to \infty$.*

Therefore the function $\rho_S^{FL}(\delta)$ bounds a region of strong equivalence for ℓ^1-minimization.

Theorem 3.5 follows from Theorem 3.3 and the validity of the probabilistic bounds on the RIP constants, Theorem 2.3. In particular, Lemma 3.6 bounds $\mu^{FL}(k,n,N)$ in terms of the asymptotic RIP bounds $L(\delta,2\rho)$ and $U(\delta,2\rho)$, by the quantity $\mu^{FL}(\delta,(1+\epsilon)\rho)$ defined in (3.3). If $\rho_\epsilon(\delta)$ is the solution to $\mu^{FL}(\delta,(1+\epsilon)\rho) = 1$, then for $\rho < \rho_\epsilon(\delta)$ we achieve the desired bound, $\mu^{FL}(k,n,N) < 1$, to ensure $StrongEquiv(A,\ell^1)$. The statement of Theorem 3.5 follows from relating $\rho_\epsilon(\delta)$ to $\rho_S^{FL}(\delta)$, the solution to $\mu^{FL}(\delta,\rho) = 1$.

LEMMA 3.6. *Fix $\epsilon > 0$. Consider the proportional-growth asymptotic with parameters $(\delta,\rho) \in (0,1)\times(0,1/2)$. Draw the corresponding $n \times N$ matrices A from the Gaussian ensemble. There is an exponentially high probability on the draw of A that*

$$\mu^{FL}(k,n,N) < \frac{1+\sqrt{2}}{4}\left(\frac{1+U(\delta,2(1+\epsilon)\rho)}{1-L(\delta,2(1+\epsilon)\rho)} - 1\right) = \mu^{FL}(\delta,(1+\epsilon)\rho). \tag{3.3}$$

Proof. Theorem 2.3 and the form of $\mu^{FL}(\delta,\rho)$ imply a similar bound to the above with a modified dependence on ϵ. For any $c\epsilon > 0$, with $n/N \to \delta \in (0,1)$ and $k/n \to \rho \in (0,1/2]$ as $(k,n,N) \to \infty$, there is an exponentially high probability on the draw of A from the Gaussian ensemble that

$$\mu^{FL}(k,n,N) < \frac{1+\sqrt{2}}{4}\left(\frac{1+U(\delta,2\rho)+c\epsilon}{1-L(\delta,2\rho)-c\epsilon} - 1\right). \tag{3.4}$$

[2]Similar results have been proven for other random matrix ensembles, but they are even less precise than those for the Gaussian distribution.

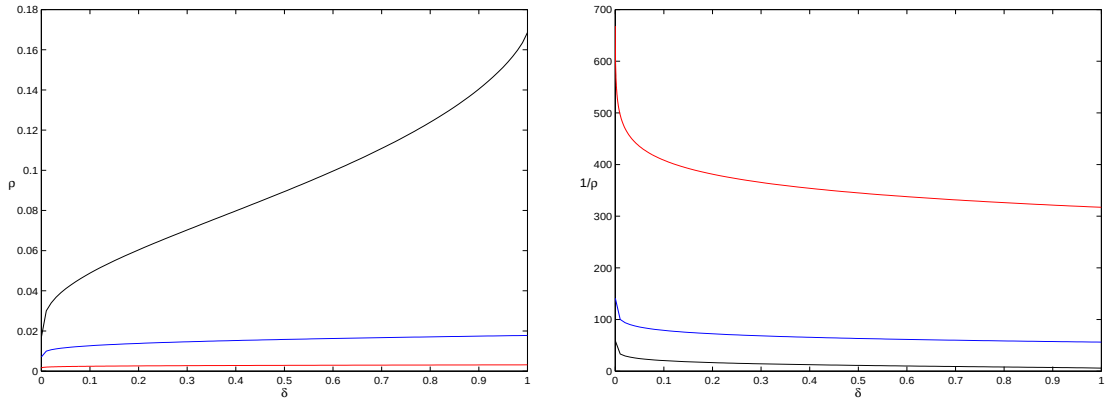

FIG. 3.1. *Left panel: Three lower bounds on the Strong ℓ^1/ℓ^0-equivalence phase transition, $\rho_S(\delta)$, for Gaussian random matrices from Theorem 3.7 ($\rho_S^D(\delta)$, black), Theorem 3.9 ($\rho_S^{RV}(\delta)$, blue), and Theorem 3.5 ($\rho_S^{FL}(\delta)1$, red). Right panel: The inverse of the ℓ^1/ℓ^0-equivalence phase transition lower bounds in the left panel.*

Since $U(\delta,\rho)$ is non-decreasing in ρ and $L(\delta,\rho)$ is strictly increasing in ρ for any δ and $\rho \in (0,1)$, it follows that the right-hand side of (3.4) can be bounded by the right-hand side of (3.3) for any fixed ϵ satisfying $0 < \epsilon < \frac{1}{2\rho} - 1$, by setting

$$c := \frac{\rho}{2} \left. \frac{\partial L(\delta, z)}{\partial z} \right|_{z=2(1+\epsilon)\rho} > 0.$$

(The upper bound on ϵ is imposed so that the second argument of $U(\delta,\cdot)$ and $L(\delta,\cdot)$, $2(1+\epsilon)\rho$, is in the admissible range of $(0,1)$.) That the bound (3.3) is satisfied for all $\epsilon > 0$ sufficiently small, and that the right hand side of (3.3) is strictly increasing in ϵ establishes that (3.3) is in fact satisfied with exponentially high probability on the draw of A for *any* $\epsilon \in \left(0, \frac{1}{2\rho} - 1\right)$. □

Proof. [Theorem 3.5] Let $\rho_\epsilon(\delta)$ be the solution of $\mu^{FL}(\delta, (1+\epsilon)\rho) = 1$. Then, for any $\rho < \rho_\epsilon(\delta)$, Lemma 3.6 implies that $\mu^{FL}(k,n,N) < 1$, which by Theorem 3.3, ensures $StrongEquiv(A,\ell^1)$. To remove the dependence on the level curve $\rho_\epsilon(\delta)$, note that $\rho_\epsilon(\delta)$ is related to $\rho_S^{FL}(\delta)$, the solution of $\mu^{FL}(\delta,\rho) = 1$, by $(1+\epsilon)\rho_\epsilon(\delta) \equiv \rho_S^{FL}(\delta)$. Since $(1-\epsilon) < (1+\epsilon)^{-1}$ for all $\epsilon > 0$, we have $(1-\epsilon)\rho_S^{FL}(\delta) < \rho_\epsilon(\delta)$. Thus, provided $\rho < (1-\epsilon)\rho_S^{FL}(\delta)$, the statement of Theorem 3.5 is satisfied. □

3.2. Comparison of bounds on $StrongEquiv(A,\ell^1)$. In this section we use the phase transition framework to readily compare bounds on the region of $StrongEquiv(A,\ell^1)$ obtained from vastly different methods of analysis. In Section 3.1, we have already determined the region of strong equivalence for ℓ^1-minimization obtained by using the RIP. Here we look at two other examples, namely Donoho's polytope results [16, 18] and the sufficient condition of Rudelson and Vershynin [39] obtained from geometric functional analysis. We do not go into great details about how the results were obtained, but simply point out that the methods of analysis are rather different. As a result, the original statements of the theorems take drastically different forms and are therefore difficult to compare even qualitatively. Translating the results into the phase transition framework, however, offers a direct, quantitative, and simple method of comparison.

Using polytope theory and the notion of central-neighborliness, Donoho [16] defined a function $\rho_S^D(\delta)$ which defines a region of the (δ,ρ) phase space ensuring StrongEquiv(A,ℓ^1) with high probability on the draw of A. The phase transition function $\rho_S^D(\delta)$ is displayed as the black curve in Figure 3.1.

THEOREM 3.7 (Donoho [16]). *Fix $\epsilon > 0$. Consider the proportional-growth asymptotic, Definition 1.2, with parameters $(\delta,\rho) \in (0,1)\times(0,1/2)$. Sample each $n \times N$ matrix A from the Gaussian ensemble. Suppose $\rho < (1-\epsilon)\rho_S^D(\delta)$. Then Prob$(StrongEquiv(A,\ell^1)) \to 1$ as $n \to \infty$.*

Therefore $\rho_S^D(\delta)$ bounds a region of strong equivalence for ℓ^1-minimization.

Rudelson and Vershynin [39] used an alternative geometric approach from geometric functional analysis (GFA) to determine regions of $StrongEquiv(A, \ell^1)$ for Gaussian and random partial Fourier matrices. For Gaussian matrices their elegantly simple proof involves employing Gordon's "escape through the mesh theorem" on the nullspace of A. Their lower bound on the region of $StrongEquiv(A, \ell^1)$ is larger for the Gaussian ensemble than for the Fourier ensemble. We restate their condition for the Gaussian ensemble in the proportional growth asymptotic.

DEFINITION 3.8 (GFA Region of $StrongEquiv(A, \ell^1)$). *Define*

$$\gamma(\rho\delta) := \exp\left(\frac{\log(1 + 2log(e/\rho\delta))}{4\log(e/\rho\delta)}\right),$$

$$\mu^{RV}(\delta, \rho) := \rho\left(12 + 8\log(1/\rho\delta) \cdot \gamma^2(\rho\delta)\right), \tag{3.5}$$

and $\rho_S^{RV}(\delta)$ as the solution to $\mu^{RV}(\delta, \rho) = 1$.

The function $\rho_S^{RV}(\delta)$ is displayed as the blue curve in Figure 3.1.

THEOREM 3.9 (Rudelson and Vershynin [39]). *Fix $\epsilon > 0$. Consider the proportional-growth asymptotic, Definition 1.2, with parameters $(\delta, \rho) \in (0, 1) \times (0, 1/2)$. Sample each $n \times N$ matrix A from the Gaussian ensemble. Suppose $\rho < (1 - \epsilon)\rho_S^{RV}(\delta)$. Then Prob$(StrongEquiv(A, \ell^1)) \to 1$ as $n \to \infty$.*

Therefore $\rho_S^{RV}(\delta)$ bounds a region of strong equivalence for ℓ^1-minimization.

Versions of Theorems 3.7 and 3.9 exist for finite values of (k, n, N), [21, 39], but in each case the recoverability conditions rapidly approach the stated asymptotic limiting functions $\rho_S(\delta)$ as (k, n, N) grow; we do not further complicate the discussion with their rates of convergence.

Since Theorems 3.5, 3.7, and 3.9 provide a region of $StrongEquiv(A, \ell^1)$, we now have three subsets of the exact region of $StrongEquiv(A, \ell^1)$. Although Theorems 3.5, 3.7, and 3.9 each have the same goal of quantifying the exact boundary of $StrongEquiv(A, \ell^1)$ for Gaussian random matrices, they are arrived at using substantially different methods of analysis. The efficacy of the bounds from the largest region of $StrongEquiv(A, \ell^1)$ to the smallest region are $\rho_S^D(\delta)$ of Donoho, $\rho_S^{RV}(\delta)$ of Rudelson and Vershynin, and $\rho_S^{FL}(\delta)$ of Foucart and Lai, see the left panel of Figure 3.1. From the inverse of $\rho_S(\delta)$, see the right panel of Figure 3.1, we can read the constant of proportionality where the associated method of analysis guarantees $StrongEquiv(A, \ell^1)$: from Theorems 3.7, 3.9, and 3.3 they are bounded below by: $n \geq 5.9k$, $n \geq 56k$, and $n \geq 317k$ respectively.

The phase transition framework can also be used to quantify what has been proven about an encoder/decoder pair's speed of convergence, its degree of robustness to noise, and to make comparisons of these properties between different algorithms. A general framework for translating RIP based results into the phase transition framework is presented in [3], where it is also applied to three exemplar greedy algorithms CoSaMP [37], Subspace Pursuit [15], and Iterated Hard Thresholding [4]. Bounds on regions of $StrongEquiv(A, \ell^q)$ for ℓ^q-minimization for $q \in (0, 1]$ implied by the RIP are available in an extended technical report [1], where the effects of noise are also considered. Through these objective measures of comparison we hope to make clear the proven efficacy of sparse approximation algorithms and allow for their transparent comparison.

REFERENCES

[1] J. D. Blanchard, C. Cartis, and J. Tanner. Compressed sensing: how sharp is the restricted isometry property. extended preprint, 2009.

[2] J. D. Blanchard, C. Cartis, and J. Tanner. Decay properties for restricted isometry constants. *IEEE Signal Proc. Letters*, 16(7):572–575, 2009.

[3] J. D. Blanchard, C. Cartis, J. Tanner, and A. Thompson. Phase transitions for greedy sparse approximation algorithms. submitted, 2009.

[4] T. Blumensath and M. E. Davies. Iterative hard thresholding for compressed sensing. *Appl. Comp. Harm. Anal.*, 27(3):265–274, 2009.

[5] A. M. Bruckstein, David L. Donoho, and Michael Elad. From sparse solutions of systems of equations to sparse modeling of signals and images. *SIAM Review*, 51(1):34–81, 2009.
[6] E. J. Candès. Compressive sampling. In *International Congress of Mathematicians. Vol. III*, pages 1433–1452. Eur. Math. Soc., Zürich, 2006.
[7] E. J. Candès. The restricted isometry property and its implications for compressed sensing. *C. R. Math. Acad. Sci. Paris*, 346(9-10):589–592, 2008.
[8] E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52(2):489–509, 2006.
[9] E. J. Candès, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.*, 59(8):1207–1223, 2006.
[10] E. J. Candès and T. Tao. Decoding by linear programming. *IEEE Trans. Inform. Theory*, 51(12):4203–4215, 2005.
[11] E. J. Candès and T. Tao. Near-optimal signal recovery from random projections: universal encoding strategies? *IEEE Trans. Inform. Theory*, 52(12):5406–5425, 2006.
[12] R. Chartrand. Exact reconstructions of sparse signals via nonconvex minimization. *IEEE Signal Process. Lett.*, 14:707–710, 2007.
[13] R. Chartrand and V. Staneva. Restricted isometry properties and nonconvex compressive sensing. *Inverse Problems*, 24(035020):1–14, 2008.
[14] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM Rev.*, 43(1):129–159 (electronic), 2001. Reprinted from SIAM J. Sci. Comput. **2**0 (1998), no. 1, 33–61.
[15] W. Dai and O. Milenkovic. Subspace pursuit for compressive sensing signal reconstruction. submitted, arXiv:0803.0811v3, 2008.
[16] D. L. Donoho. Neighborly polytopes and sparse solution of underdetermined linear equations. Technical Report, Department of Statistics, Stanford University, 2005.
[17] D. L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, 2006.
[18] D. L. Donoho. High-dimensional centrally symmetric polytopes with neighborliness proportional to dimension. *Discrete Comput. Geom.*, 35(4):617–652, 2006.
[19] D. L. Donoho and V. Stodden. Breakdown point of model selection when the number of variables exceeds the number of observations. In *Proceedings of the International Joint Conference on Neural Networks*, 2006.
[20] D. L. Donoho and J. Tanner. Sparse nonnegative solutions of underdetermined linear equations by linear programming. *Proc. Natl. Acad. Sci. USA*, 102(27):9446–9451, 2005.
[21] D. L. Donoho and J. Tanner. Exponential bounds implying construction of compressed sensing matrices, error-correcting codes and neighborly polytopes by random sampling. submitted, 2008.
[22] D. L. Donoho and J. Tanner. Counting faces of randomly projected polytopes when the projection radically lowers dimension. *J. AMS*, 22(1):1–53, 2009.
[23] D. L. Donoho and Y. Tsaig. Fast solution of l1 minimization problems when the solution may be sparse. *IEEE Trans. Inform. Theory*, 54(11):4789–4812, 2008.
[24] D. L. Donoho, Y. Tsaig, I. Drori, and J.-L. Stark. Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit. *IEEE Trans. Inform. Theory*, submitted.
[25] R. Dorfman. The detection of defective members of large populations. *Ann. Math. Statist.*, 14(4):436–440, 1943.
[26] C. Dossal, G. Peyré, and J. Fadili. A numerical exploration of compressed sampling recovery. submitted, 2009.
[27] A. Edelman. Eigenvalues and condition numbers of random matrices. *SIAM J. Matrix Anal. Appl.*, 9(4):543–560, 1988.
[28] A. Edelman and N. R. Rao. Random matrix theory. *Acta Numer.*, 14:233–297, 2005.
[29] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright. Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. *IEEE J. Sel. Topics Signal Process.*, 1(4):586–597, 2007.
[30] S. Foucart and M.-J. Lai. Sparsest solutions of underdetermined linear systems via ℓ_q-minimization for $0 < q \leq 1$. *Appl. Comput. Harmon. Anal.*, 26(3):395–407, 2009.
[31] S. Geman. A limit theorem for the norm of random matrices. *Ann. Probab.*, 8(2):252–261, 1980.
[32] A. C. Gilbert, M. A. Iwen, and M. J. Strauss. Group testing and sparse signal recovery. In *42nd Asilomar Conference on Signals, Systems, and Computers*, 2008.
[33] A. K. Hartmann and H. Rieger. *New Optimization Algorithms in Physics*. Wiley VCH, Cambridge, 2006.
[34] A. K. Hartmann and M. Weight. *Phase Transitions in Combinatorial Optimization Problems*. Wiley VCH, Cambridge, 2005.
[35] M. Journée, Y. Nesterov, P. Richtárik, and R. Sepulchre. Generalized power method for sparse principal component analysis. submitted, 2008.
[36] B. K. Natarajan. Sparse approximate solutions to linear systems. *SIAM J. Comput.*, 24(2):227–234, 1995.
[37] D. Needell and J. Tropp. Cosamp: Iterative signal recovery from incomplete and inaccurate samples. *Appl. Comp. Harm. Anal.*, 26(3):301–321, 2009.
[38] D. Needell and R. Vershynin. Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit. *Foundations of Comp. Math.*, 9(3):317–334, 2009.
[39] M. Rudelson and R. Vershynin. On sparse reconstruction from Fourier and Gaussian measurements. *Comm. Pure Appl. Math.*, 61(8):1025–1045, 2008.
[40] R. Saab and O. Yilmaz. Sparse recovery by non-convex optimization–instance optimality. *Appl. Comp. Harm.*

Anal., in press.

[41] J. W. Silverstein. The smallest eigenvalue of a large-dimensional Wishart matrix. *Ann. Probab.*, 13(4):1364–1368, 1985.

[42] J. A. Tropp. Greed is good: algorithmic results for sparse approximation. *IEEE Trans. Inform. Theory*, 50(10):2231–2242, 2004.

[43] E. van den Berg and M. P. Friedlander. Probing the pareto frontier for basis pursuit solutions. *SIAM Journal on Scientific Computing*, 31(2):890–912, 2008.

[44] E. T. Whittaker and G. N. Watson. *A course of modern analysis*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 1996. An introduction to the general theory of infinite processes and of analytic functions; with an account of the principal transcendental functions, Reprint of the fourth (1927) edition.

[45] W. Yin, S. Osher, D. Goldfarb, and J. Darbon. Bregman iterative algorithms for ℓ^1-minimization with applications to compressed sensing. *SIAM Journal on Imaging Science*, 1(1):143–168, 2008.