COMPRESSED SENSING: HOW SHARP IS THE RIP?

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

Abstract. Consider a measurement matrix A of size $n \times N$, with $n < N$, y a signal in $\mathbb{R}^N$, and $b = Ay$ the observed measurement of the vector y. From knowledge of (b, A), compressed sensing seeks to recover the k-sparse x, $k < n$, which minimizes $\|b - Ax\|$. Using various methods of analysis — convex polytopes, geometric functional analysis, and the restricted isometry property (RIP) — it has been proven that x can be reconstructed via ℓ^q-regularization ($q \in (0, 1]$) provided A satisfies conditions dictated by the method of analysis. This article focuses on the RIP approach and A with entries drawn i.i.d. from the Gaussian distribution $\mathcal{N}(0, 1/\sqrt{n})$, developing more precise bounds on the restricted isometry constants, and using these bounds in an asymmetric RIP formulation to quantify the region of $(\frac{n}{N}, \frac{k}{n})$ in which RIP implies that ℓ^q-regularization will typically recover all k-sparse signals. Letting $\frac{n}{N} \to \delta$ and $\frac{k}{n} \to \rho$ as $n \to \infty$, the aforementioned recoverability region is characterized by all $\rho < (1 - \epsilon)\rho_S^{RIP}(\delta; q)$ for any $\epsilon > 0$, where $\rho_S^{RIP}(\delta; q)$ is a lower bound of the true phase transition below which ℓ^q-regularization will typically recover all k-sparse signals. This phase transition framework, proposed in this context by Donoho (2005), is applied to compressed sensing results obtained by the analysis techniques of centro-symmetric polytope theory (Donoho), geometric functional analysis (Rudelson and Vershynin), and the RIP (Candès, Romberg and Tao; Foucart and Lai; Chartrand). Recasting the results from different methods of analysis into a common phase transition framework allows for the direct comparison of the efficacy of the respective results.

Key words. Compressed sensing, sparsity, sparse approximation, sparse solutions to underdetermined systems, restricted isometry property, restricted isometry constants, phase transitions, convex relaxation, random matrices, Gaussian matrices, Wishart matrices, singular values of random matrices, eigenvalues of random matrices

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

1. Introduction.

The introduction of random sampling and nonlinear reconstruction algorithms in the context of sparse approximation and Compressed Sensing [10, 16] has fundamentally changed signal processing. Predominantly studied in its discrete form, an unknown vector of interest, $y \in \mathbb{R}^N$, is measured via $n < N$ inner products through the action of a matrix $\tilde{A}$, so that $b := \tilde{A}y \in \mathbb{R}^n$. From the knowledge of $\tilde{A}$ and b, one seeks to determine y. Despite the system being underdetermined ($n < N$), and as a result, typically having infinitely many solutions, y or an approximation thereof can often be recovered provided y has some underlying simplicity and $\tilde{A}$ is chosen appropriately. The most commonly exploited form of simplicity in y is *compressibility* or *sparsity*; in its simplest form, there exists a known $N \times N$ matrix Φ and a vector x with only k nonzeros such that $\|\Phi x - y\| \ll 1$. The optimal k-sparse vector x in the compressed sensing framework is the solution to

$$\min_{\|x\|_{\ell^0} \le k} \|Ax - b\| \tag{1.1}$$

with $A := \tilde{A}\Phi$, where following the usual convention in the community $\|\cdot\|_{\ell^0}$ counts the number of nonzeros. However, (1.1) is computationally intractable, being NP-hard for general A, [35].

The success of compressed sensing can be traced to the ability of computationally tractable algorithms, such as greedy methods and regularization techniques, to approximately, and sometimes even exactly, solve (1.1). The central question in the theory of compressed sensing is to determine for a given measurement matrix A and parameter values (k, n, N), the performance of the reconstruction algorithms in terms of computational efficiency and accuracy of the recovered solution. The most widely used approach for analyzing the performance of sparse approximation algorithms is currently

*Department of Mathematics, University of Utah, Salt Lake City, Utah, USA (jeff@math.utah.edu). JDB acknowledges support from NSF DMS (VIGRE) grant number 0602219.

†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.

the Restricted Isometry Property (RIP) introduced by Candès and Tao [11]. A matrix A of size $n \times N$ is said to have the RIP property with RIP constant $R(k, n, N)$ if

$$R(k, n, N) = \delta_k := \min_{c \geq 0} c \text{ subject to } (1-c)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1+c)\|x\|_2^2, \ \forall x \in \chi^N(k), \tag{1.2}$$

where

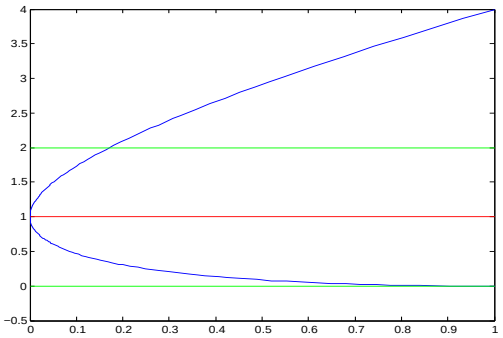
$$\chi^N(k) := \left\{ x \in \mathbb{R}^N : \ \|x\|_{\ell^0} = k \right\} \tag{1.3}$$

is the family of vectors with k or fewer nonzero entries, geometrically represented by the union of j dimensional faces of a scaled ℓ^1-ball for $j = 0, 1, \ldots, k-1$. In the following, we deviate from the standard notation of δ_k for the RIP constant, and instead use $R(k, n, N)$ in order to make explicit the dependence on all three problem parameters, (k, n, N), and to avoid conflict with another standard notation, namely $\delta = \frac{n}{N}$.

A significant advance in compressed sensing has been the development of conditions under which various computationally tractable algorithms perform nearly optimally in the sense that the regime in which these methods will approximately recover the solution of (1.1) is as high as k proportional to n in the limit, when $n, N \to \infty$ and $n/N \to \delta \in (0, 1)$. There exist conditions on the value of the RIP constant which are sufficient to ensure this near optimal performance of reconstruction algorithms, see [30, 36] and references therein. Currently, there is no known matrix A of size $n \times N$ which satisfies the sufficient conditions for RIP matrices to ensure success rates for k proportional to n. However, for many ensembles of random matrices it has been shown that, with high probability on the draw of a matrix from these ensembles, the desired conditions on the RIP constants are satisfied, [10]. Despite these successes, only perfunctory attempts have been made to determine the quantitative behavior of the RIP constants for classes of random matrices, and in particular, the behavior of the proportionality constant of k/n.

The central aim of this article is twofold; firstly, to develop more precise bounds on the RIP constants for a particular random matrix ensemble, and secondly, to quantify for this ensemble, the proportionality constant, k/n, for which regularization algorithms have been proven — using RIP and other methods of analysis — to recover a suitable approximation to (1.1). In doing so, we hope to elucidate the current state of knowledge, indicate its successes and shortfalls, and point to potentially productive open problems in the area. We focus our attention on the class of random matrices whose entries are drawn i.i.d. at random from a Gaussian distribution, quantifying the typical behavior of RIP constants in the limiting case when $n, N \to \infty$ while $n/N \to \delta$. From these results, and by using more flexible notions of the RIP that better capture the behavior of these random matrices, we are able to specify regions of the parameters $(n/N, k/n)$ under which sparse approximation regularization algorithms will typically behave nearly optimally. This follows the phase transition framework for comparing the behavior of sparse approximation algorithms advocated by Donoho et. al, [15, 18, 19, 21, 25].

2. Bounds on RIP for Gaussian Random Matrices. Let $\Lambda \subset \{1, \ldots, N\}$ be an index set of cardinality k which specifies the columns of A chosen for a submatrix, A_Λ, of size $n \times k$. The RIP constant $R(k, n, N)$, (1.2), measures the extreme deviation from unity of the largest and smallest eigenvalue of all $\binom{N}{k}$ different $A_\Lambda^T A_\Lambda$ derived from submatrices of A (or equivalently, of the most extreme singular values of all A_Λ). Associated with a random matrix ensemble is an, as of yet unknown, probability distribution function for $R(k, n, N)$. Bounds on the upper and lower tails of the probability distribution function for the largest and smallest eigenvalues of $A_\Lambda^T A_\Lambda$ can be used to bound the probability that a matrix drawn from an ensemble will have an RIP constant exceeding a specified value. This focus on the extreme tail behavior of a probability distribution function is the classical theory of large deviations, and is appropriate here due to the exponentially growing number of k-dimensional submatrices, $\binom{N}{k}$. Let us now focus on the ensemble of Gaussian random matrices.



$\rho = k/n$

FIG. 2.1. *The deviation from 1 of the expected largest and smallest eigenvalues of* $A_\Lambda^T A_\Lambda$ *with* $\rho = \frac{k}{n}$.

The inherent combinatorial nature of the RIP makes it impossible to directly compute the RIP constant of a matrix, and no other method is currently known by which it can be calculated. Despite this inaccessibility of directly calculating RIP constants, fundamental results in random matrix theory tell us that for many ensembles, the eigenvalues concentrate at known values. In particular, explicit formulae for the probability distribution functions of the largest and smallest singular values of Gaussian and other random matrix ensembles have been determined; see [28] and the references therein. Letting A_Λ be a matrix of size $n \times k$ whose entries are drawn i.i.d. from the normal distribution with variance $n^{-1/2}$, $\mathcal{N}(0, 1/\sqrt{n})$, an empirical approximation of the probability distribution functions of the largest and smallest eigenvalues of the Wishart matrices $A_\Lambda^T A_\Lambda$ is shown in Figure 2.2. The probability distribution functions of the extreme eigenvalues of the Wishart matrix $A_\Lambda^T A_\Lambda$ become exponentially localized as $n \to \infty$, with the expected values of the largest and smallest eigenvalues converging to $(1 + \sqrt{k/n})^2$ and $(1 - \sqrt{k/n})^2$ respectively as k and n go to infinity, [31, 39], see Figure 2.1.

Let us now consider in this context, the limit when the problem parameters (k, n, N) go to infinity, but their ratios,

$$\delta_n := \frac{n}{N} \qquad \rho_n := \frac{k}{n},$$

converge to nonzero values (δ, ρ) in the unit square, $\delta, \rho \in (0, 1)$. In this regime, the largest and smallest eigenvalues of a random Wishart matrix $A_\Lambda^T A_\Lambda$ converge to $(1 + \sqrt{\rho})^2$ and $(1 - \sqrt{\rho})^2$ as $(k, n, N) \to \infty$. The asymmetric deviation from unity of the expected eigenvalues of $A_\Lambda^T A_\Lambda$ suggest that the symmetric definition of the RIP is unnecessarily restrictive, and would likely be controlled by the largest eigenvalues, which is in fact the less important[1] bound for the RIP in the context of sparse approximation. For this reason, we generalize the RIP to an asymmetric form.

DEFINITION 1. *For a matrix* A *of size* $n \times N$, *the* asymmetric RIP constants $L(k, n, N)$ *and* $U(k, n, N)$ *are defined as:*

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

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

[1] Although both the smallest and largest singular values of $A_\Lambda^T A_\Lambda$ affect the stability of the reconstruction algorithms, the smaller eigenvalue is more crucial in that it determines the ability to distinguish 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 $Ax = Ax'$.

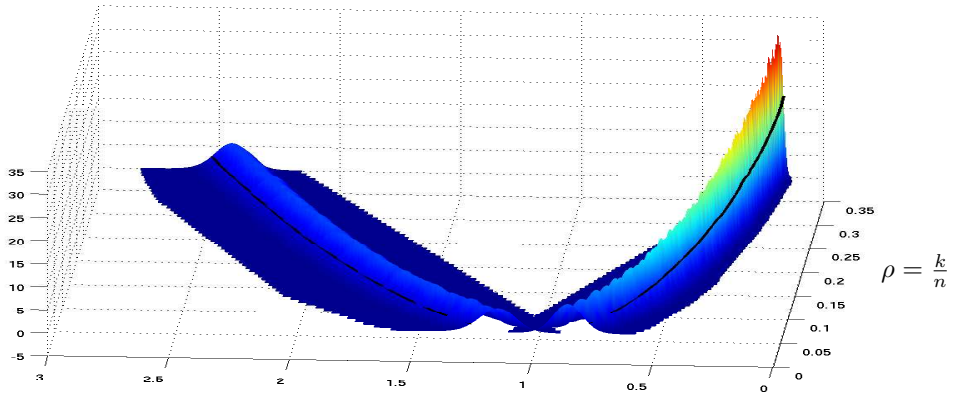


FIG. 2.2. *An empirical approximation of the probability distribution functions for the largest and smallest eigenvalues of random Wishart matrices $A_\Lambda^T A_\Lambda$ with entries drawn $\mathcal{N}(0, 1/\sqrt{n})$ and $\rho = \frac{k}{n}$. Here calculated with $n = 200$. The expected value of the largest and smallest eigenvalues converge to $(1+\sqrt{\rho})$ and $(1-\sqrt{\rho})$ respectively, overlaid on the empirical probability distribution function as thick black lines.*

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

Let λ_{max}^Λ and λ_{min}^Λ denote the largest and smallest eigenvalues for the Wishart matrix $A_\Lambda^T A_\Lambda$. We see from (2.1) and (2.2) that $(1-L(k,n,N)) = \min_\Lambda \lambda_{min}^\Lambda$ and $(1+U(k,n,N)) = \max_\Lambda \lambda_{max}^\Lambda$. A standard large deviation analysis of bounds on the probability distribution function for the largest and smallest eigenvalues of random Wishart matrices allows us to establish upper bounds on the typical size of $L(k,n,N)$ and $U(k,n,N)$ for this ensemble.

THEOREM 1. *Let A be a matrix of size $n \times N$ whose entries are drawn i.i.d. from $\mathcal{N}(0, 1/\sqrt{n})$ and let $n \to \infty$ with $\frac{k}{n} \to \rho$ and $\frac{n}{N} \to \delta$. Let $H(p) := p\log(1/p) + (1-p)\log(1/(1-p))$ denote the usual Shannon Entropy with base e logarithms, and let*

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

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

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

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

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

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

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

For any $\epsilon > 0$, as $n \to \infty$,

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

The details of the proof of Theorem 1 are deferred to Section 5.1, but a sketch is as follows. The bounds are derived using a simple union bound over all $\binom{N}{k}$ of the $k \times k$ Wishart matrices $A_\Lambda^T A_\Lambda$ that

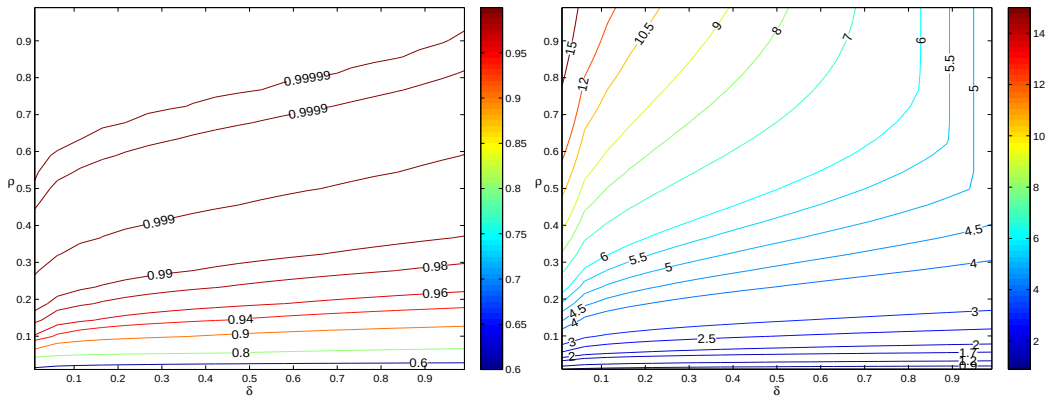


FIG. 2.3. *Bounds, $L(\delta,\rho)$ and $U(\delta,\rho)$ (left and right respectively), above which it is exponentially unlikely that the RIP constants $L(k,n,N)$ and $U(k,n,N)$ exceed, with entries in A drawn i.i.d. $N(0,1/\sqrt{n})$ and in the limit as $\frac{k}{n}\to\rho$ and $\frac{n}{N}\to\delta$ as $n\to\infty$, see Theorem 1.*

can be formed from columns of A. Bounds on the tail behavior of the probability distribution function for the largest and smallest eigenvalues of $A_\Lambda^T A_\Lambda$ can be expressed in the form $p(n,\lambda)\exp(n\psi(\lambda,\rho))$ with ψ defined in (2.3) and (2.4). Following standard practices in large deviation analysis, the tails of the probability distribution functions are balanced against the exponentially large number of Wishart matrices (2.5) and (2.6) to define upper and lower bounds on the largest and smallest eigenvalues of all $\binom{N}{k}$ Wishart matrices, with bounds $\lambda_{min}(\delta,\rho)$ and $\lambda_{max}(\delta,\rho)$, respectively. Overestimation of the union bound over the combinatorial number of $\binom{N}{k}$ Wishart matrices causes the bound $\lambda_{max}(\delta,\rho)$ to not be strictly increasing in ρ for δ large; to utilize the best available bound on the extreme of the largest eigenvalue, we note that any bound $\lambda_{max}(\delta,\nu)$ for $\nu\in[\rho,1]$ is also a valid bound for submatrices of size $n\times k$. The asymptotic bounds of the asymmetric RIP constants, $L(\delta,\rho)$ and $U(\delta,\rho)$, follow directly.

3. Phase Transitions for Reconstruction Algorithms. The success of sparse approximation as a practical tool is a consequence of algorithms which are able to approximately solve the seemingly intractable (1.1) in a computationally efficient manner for some problem classes. With the current high level of interest in the topic, new algorithms are introduced continuously, with each algorithm having different characteristics and areas of anticipated applicability. Despite these differences, broadly speaking, the algorithms fall into two categories: greedy algorithms and regularizations. Greedy algorithms are iterative, with each step selecting a subset of entries in x which are adjusted to improve the desired error metric. Examples of greedy algorithms include Orthogonal Matching Pursuit (OMP) [40], Regularized OMP (ROMP) [37], Stagewise OMP (StOMP) [26], and Compressive Sampling MP (CoSaMP) [36]. Regularization formulations for sparse approximation began with the now ubiquitous (convex) ℓ^1-regularization [14], and has since been extended to consider non-convex ℓ^q-regularizations,

$$\min_x \|x\|_q \quad \text{subject to} \quad \|Ax-b\|_2\le\tau, \tag{3.1}$$

for $q\le 1$ [12, 13, 30], and to include additional constraints on x such as nonnegativity [20, 21, 24] or structure in the pattern of nonzeros [2, 3]. General-purpose convex optimization solvers may be employed for the solution of the various (exact and inexact) ℓ^1-regularization formulations. Recently, highly-efficient software has been designed specifically for these problems, such as [14, 29, 41, 43], a discussion of which is beyond the scope of this paper. Solving the ℓ^q, $q<1$, instances is more difficult, however, due to the nonconvexity and nonsmoothness of the problem. Generic (nonconvex) optimization software may be applied to the ℓ^q problem or some smoothed variant of it, with the

expectation that only a local solution will in general be found. The above-cited papers on ℓ^q also propose algorithms for these problems, with some numerical evidence of efficiency.

Although each of these algorithms have different characteristics, the general belief in the sparse approximation community has been that greedy algorithms have lower computational complexity than algorithms for ℓ^q-regularization, and that regularization formulations have a larger region of near optimal recoverability than the greedy methods. This question is far from resolved, and with the development of optimization algorithms specifically designed for sparse approximation, the distinction between greedy and regularized algorithms is becoming increasingly blurred. At the time of writing this article, it seems that greedy algorithms such as StOMP and CoSaMP are the method of choice when the optimal x has nearly equally weighted non-zero coefficients, while ℓ^1-regularization is preferred when computational speed is needed and the coefficients in the optimal x are of substantially different magnitude. Lastly, ℓ^q-regularization with $q < 1$ (non-convex) is the method of last resort when other algorithms are not sufficiently successful. We restrict our further discussion here to quantifying the region of successful recovery for ℓ^q-regularization, $q \leq 1$, with analogous statements for the greedy algorithm CoSaMP and ROMP in [4]. We do not discuss the particular algorithms for solving the regularization formulations (3.1), but point the interested reader to [29, 41, 43].

One of the central aims of this article is to advocate a unifying framework for the comparison of results in sparse approximation. A paradigm for the comparison of results in sparse approximation has already been put forth by Donoho [15, 18] and used by Donoho et al, [19, 21, 25], borrowed from the statistical mechanics literature and used successfully in a similar context by the combinatorial optimization community, see [33, 34], and references therein. However, there has been no general agreement upon the employment of this or any other framework in the context of sparse approximation, and hence it is neither clear when new results are improvements on existing ones, nor is it possible to know which methods of analysis are achieving the best results. Fortunately, every sparse approximation algorithm can be cast in the above-mentioned paradigm of Donoho et al., parameterized by two inherent problem size parameters:[2] the *undersampling* rate of measuring x through n inner products with the rows of A, as compared to directly sampling each element of $x \in \mathbb{R}^N$, namely $\delta_n = n/N \in (0, 1)$, and secondly, the *oversampling* rate of making n measurements as opposed to the optimal *oracle* rate of making k measurements, that is $\rho_n = k/n \in (0, 1)$. For a given value of $\delta_n \in (0, 1)$, and metric of success, there is a largest value of ρ_n such that the metric of success is satisfied. The interest in ℓ^1-regularization follows from its optimal rate, namely, that for a broad class of matrices, as $\delta_n \to \delta$ and $\rho_n \to \rho$ with $\delta, \rho \in (0, 1)$, there is a strictly positive function $\rho_S(\delta) > 0$ for which if $\rho < \rho_S(\delta)$ then ℓ^1-regularization will exactly recover every k sparse vector; this function is referred to as the Strong phase transition $\rho_S(\delta)$, [10, 15, 18]. Moreover, there is another (lower) function of δ for which this reconstruction is stable to noise, or will recover a vector $\hat{x}$ for which $\|A\hat{x} - b\|_q$ is nearly as small as would be obtained from the optimal k-sparse x, [17]. The phase transitions functions $\rho_S(\delta)$ for $(k, n, N) \to \infty$, and when available, their finite dimensional analogues, allow for a simple quantitative comparison between different algorithms, matrices, and success criteria.

In order to make quantitative statements, the matrix, or random matrix ensemble must first be specified. Here we restrict our attention to A whose entries are drawn i.i.d. from a Gaussian distribution. Similar results have been proven for other random matrix ensembles, but they are even less precise than those for the Gaussian distribution. In Section 3.1 we consider ℓ^1-regularization in the simplest case of x being strictly k sparse, and cast in term of phase transitions results proven by three distinct methods of analysis: eigenvalue analysis and the RIP [30], geometric functional analysis [38], and convex polytopes [15]. In Section 3.2 we consider the more general case of ℓ^q-regularization with $q \in (0, 1]$ and x not strictly k sparse, again casting these results in terms of

[2]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.

phase transitions.

3.1. ℓ_1/ℓ_0-equivalence Phase Transitions. In the case where the measurement vector b can be expressed exactly using $k < n$ columns of A, (1.1) is recast as (3.1) with $\tau = 0$ and $q = 0$ (the counting measure). When the solution to (3.1) with $\tau = 0$ is the same for $q = 0$ and $q = 1$, we refer to this as ℓ_1/ℓ_0-*equivalence*; these are problems in which the seemingly combinatorial search for the sparsest solution, ℓ^0, can be solved computationally efficiently using standard linear programming algorithms. Recall the family $\chi^N(k)$ of k-sparse vectors defined in (1.1). Here we consider the most strict condition, requiring that, with high probability on the random draw of A, *every* $x \in \chi^N(k)$ is a point of ℓ_1/ℓ_0-equivalence. This Strong condition motivates the subscript S in our notation to distinguish ℓ_1/ℓ_0-equivalence for *all* $x \in \chi^N(k)$ from weaker notions of ℓ_1/ℓ_0-equivalence [15, 21]. Donoho [15] provided a necessary and sufficient condition on the matrix A such that every $x \in \chi^N(k)$ is a point of ℓ_1/ℓ_0-equivalence; namely the projection of the unit ℓ_1 ball, C^N, under A preserves all k-faces, i.e. AC^N and C^N have the same number of $k-1$ dimensional faces. This is the notion of k *central-neighborliness*. Donoho computed a lower bound on the ℓ_1/ℓ_0-equivalence phase transition, $\rho_S(\delta; C^N)$, below which Gaussian matrices are typically k central-neighborly [18] and therefore exactly recover every $x \in \chi^N(k)$ via (3.1) with $\tau = 0$ and $q = 1$.

THEOREM 2. *(Donoho [15]) For any $\epsilon > 0$, 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 with Gaussian i.i.d. entries that every $x \in \chi^N(k)$ is a point of ℓ_1/ℓ_0-equivalence if $\rho < (1-\epsilon)\rho_S(\delta; C^N)$. The phase transition function $\rho_S(\delta; C^N)$ is displayed as the black curve in Figure 3.1.*

Rudelson and Vershynin [38] used an alternative geometric approach from geometric functional analysis to determine regions of ℓ_1/ℓ_0-equivalence for Gaussian and partial Fourier random matrices. For Gaussian matrices their elegantly simple proof is arrived at by employing Gordon's escape through the mesh theorem on the null space of A. Their lower bound on the region of ℓ_1/ℓ_0-equivalence is larger for the Gaussian ensemble than for the Fourier ensemble; the lower bound for the Gaussian ensemble is restated in Theorem 3 in the asymptotic limit of $n \to \infty$ and in the language of phase transitions.

THEOREM 3. *(Rudelson and Vershynin [38]) For any $\epsilon > 0$, 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 with Gaussian i.i.d. entries that every $x \in \chi^N(k)$ is a point of ℓ_1/ℓ_0-equivalence if $\rho < (1-\epsilon)\rho_S^{RV}(\delta)$. The phase transition function $\rho_S^{RV}(\delta)$ is defined as the solution of*

$$\rho = \frac{1}{12 + 8\log(1/\rho\delta) \cdot \gamma^2(\rho\delta)} \quad \text{with} \quad \gamma(\rho\delta) := \exp\left(\frac{\log(1 + 2\log(e/\rho\delta))}{4\log(e/\rho\delta)}\right). \tag{3.2}$$

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

Following the pioneering work of Candès, Romberg, and Tao [8, 11], many different conditions on the RIP have been developed which if satisfied imply some form of sparse approximation rate for ℓ^1-regularization, [5, 6, 9, 10, 38] to name a few; more recently these results have been extended to ℓ^q-regularization and other greedy algorithms [13, 36, 37]. The current state of the art RIP related conditions for ℓ^q regularization were developed by Foucart and Lai.

THEOREM 4. *(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 $2k \le n < N$, if $\mu(2k, n, N) < 1$ where*

$$\mu(2k, n, N) := \frac{1+\sqrt{2}}{4}\left(\frac{1 + U(2k, n, N)}{1 - L(2k, n, N)} - 1\right) \tag{3.3}$$

then every every $x \in \chi^N(k)$ is a point of ℓ_1/ℓ_0-equivalence.

Substituting bounds on the asymmetric RIP constants $L(2k, n, N)$ and $U(2k, n, N)$ from Theorem 1 results in a lower bound on the ℓ_1/ℓ_0-equivalence phase transition $\rho_S(\delta)$ for Gaussian random matrices implied by the RIP, the proof is delayed to Section 5.2.

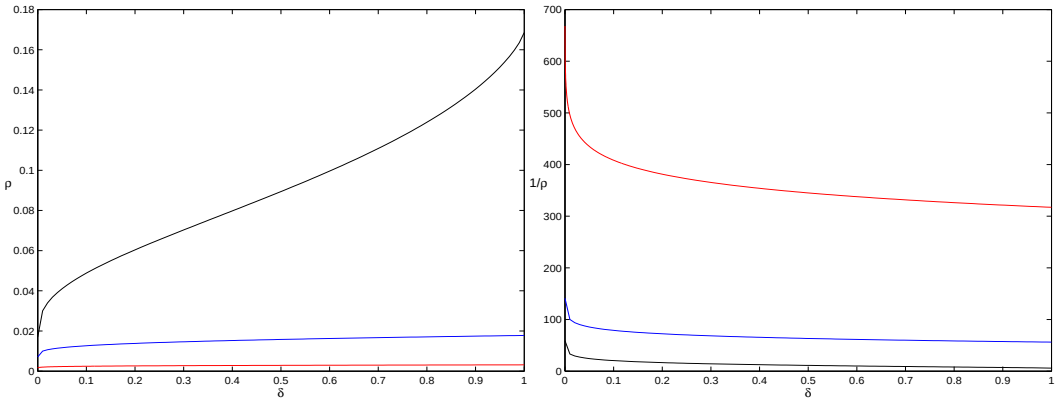


FIG. 3.1. *Left panel: Three lower bounds on the Strong ℓ_1/ℓ_0-equivalence phase transition, $\rho_S(\delta)$, for Gaussian random matrices from Theorem 2 ($\rho_S(\delta; C^N)$, black), Theorem 3 ($\rho_S^{RV}(\delta)$, blue), and Theorem 5 ($\rho_S^{RIP}(\delta)$, red). Right panel: The inverse of the ℓ_1/ℓ_0-equivalence phase transition lower bounds in the left panel.*

THEOREM 5. *For any $\epsilon > 0$, 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 with Gaussian i.i.d. entries that every $x \in \chi^N(k)$ is a point of ℓ_1/ℓ_0-equivalence if $\rho < (1-\epsilon)\rho_S^{RIP}(\delta)$. The function $\rho_S^{RIP}(\delta)$ is defined as the solution of $\mu(\delta, 2\rho) = 1$ where*

$$\mu(\delta, 2\rho) := \frac{1+\sqrt{2}}{4}\left(\frac{1+U(\delta,2\rho)}{1-L(\delta,2\rho)} - 1\right). \tag{3.4}$$

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

The lower bounds on the ℓ_1/ℓ_0-equivalence phase transition $\rho_S(\delta)$ in Theorems 2, 3, and 5 are discussed in Section 4. No attempt has of yet been made to extend the geometric approaches of Donoho, as well as of Rudelson and Vershynin, to the more general case of x not being strictly k sparse. One of the major contributions of the RIP approach for sparse approximation is that stability corollaries follow naturally. Theorems 4 and 5 follow from more general theorems for ℓ^q-regularization for $q \in (0,1]$, which also allow for an inexact fit between b and Ax.

3.2. ℓ^q-regularization Phase Transitions for $q \in (0,1]$ Implied by RIP Constants. Foucart and Lai improved on the previously best known RIP bounds of Candès ($q = 1$) and Chartrand ($q \in (0,1)$) [6, 12] for ℓ^q-regularization. Theorem 4 is the simplest case of Foucart and Lai's results, for ℓ^1-regularization and x exactly k sparse. More generally, they considered the family of x_θ which satisfy a scaled approximate fit to b,

$$\|Ax_\theta - b\|_2 \leq (1 + U(2k,n,N)) \cdot \theta \quad \text{for} \quad \theta \geq 0. \tag{3.5}$$

Letting $x_\theta^\star$ be the argmin for the ℓ^q-regularized constrained problem

$$\min_z \|z\|_q \quad \text{subject to} \quad \|Az - b\|_2 \leq (1 + U(2k,n,N)) \cdot \theta, \tag{3.6}$$

Foucart and Lai bounded the discrepancy between $x_\theta^\star$ and any x_θ satisfying (3.5) in terms of the discrepancy between x_θ and its best k sparse approximation,

$$\sigma_k(x_\theta)_q := \inf_{\|z\|_{\ell^0} \leq k} \|x_\theta - z\|_q. \tag{3.7}$$

THEOREM 6. *(Foucart and Lai [30]) Given $q \in (0,1]$, for any matrix A of size $n \times N$ with $n < N$ and with RIP constants $L(2k,n,N)$ and $U(2k,n,N)$ and $\mu(2k,n,N)$ defined as (3.3), if*

$$\alpha^{1/2-1/q}\mu(2k\alpha, n, N) < 1 \quad \text{for any} \quad 1 \leq \alpha \leq \frac{n}{2k}, \tag{3.8}$$

then a solution $x_\theta^\star$ of (3.6) approximates any x_θ satisfying (3.5) within the bounds

$$\|x_\theta - x_\theta^\star\|_q \le C_1 \cdot \sigma_k(x_\theta)_q \ + \ D_1 \cdot k^{1/q-1/2} \cdot \theta \tag{3.9}$$

$$\|x_\theta - x_\theta^\star\|_2 \le C_2 \cdot \sigma_k(x_\theta)_q \cdot (\alpha k)^{1/2-1/q} \ + \ D_2 \cdot \theta \tag{3.10}$$

with C_1, C_2, D_1, and D_2 functions of q, α, and $\frac{1+U(2\alpha k,n,N)}{1-L(2\alpha k,n,N)}$.

The parameter α in Theorem 6 is a free parameter from the method of proof, and should be selected so as to maximize the region where (3.8) and/or other conditions are satisfied. For brevity we do not state the formulae for C_1, C_2, D_1, and D_2 as functions of (k, n, N), but only state them in Theorem 7 in terms of their bounds for Gaussian random matrices as $(k, n, N) \to \infty$.

Although the solution of (3.6), $x_\theta^\star$, has unknown sparsity, Theorem 6 ensures that if there is a solution of (3.5), x_θ, which can be well approximated by a k sparse vector, i.e. if $\sigma_k(x_\theta)_q$ is small, then if (3.8) is satisfied the discrepancy between $x_\theta^\star$ and x_θ will be similarly small. For instance, if the sparsest solution of (3.6), x_θ, is k sparse, then (3.10) implies that $\|x_\theta - x_\theta^\star\|_2 \le D_2 \cdot \theta$; moreover, if $\theta = 0$ then $x_\theta^\star$ will be k sparse and satisfy $Ax_\theta^\star = b$ (in the case $q = 1$ this result is summarized as Theorem 4). Substituting bounds on the asymmetric RIP constants $L(2\alpha k, n, N)$ and $U(2\alpha k, n, N)$ from Theorem 1 we arrive at a quantitative version of Theorem 6 for Gaussian random matrices.

THEOREM 7. *Given $q \in (0, 1]$, for any $\epsilon > 0$, as $(k, n, N) \to \infty$ with $n/N \to \delta \in (0, 1)$ and $k/n \to \rho \in (0, 1/2]$, if $\rho < (1-\epsilon)\rho_S^{RIP}(\delta; q)$ where $\rho_S^{RIP}(\delta; q)$ is the maximum over $1 \le \alpha \le 1/2\rho$ of the solutions, $\rho(\delta; q, \alpha)$, of $\mu_\alpha(\delta, 2\alpha\rho) := \alpha^{1/2-1/q}\mu(\delta, 2\alpha\rho) = 1$ with $\mu(\delta, 2\rho)$ defined as in (3.4), there is an exponentially high probability on the draw of A with Gaussian i.i.d. entries that a solution $x_\theta^\star$ of (3.6) approximates any x_θ satisfying (3.5) within the bounds*

$$\|x_\theta - x_\theta^\star\|_q \le C_1(\delta, 2\alpha\rho) \cdot \sigma_k(x_\theta)_q \ + \ D_1(\delta, 2\alpha\rho) \cdot k^{1/q-1/2} \cdot \theta \tag{3.11}$$

$$\|x_\theta - x_\theta^\star\|_2 \le C_2(\delta, 2\alpha\rho) \cdot \sigma_k(x_\theta)_q \cdot (\alpha k)^{1/2-1/q} \ + \ D_2(\delta, 2\alpha\rho) \cdot \theta. \tag{3.12}$$

The multiplicative "stability factors" are defined as:

$$C_1(\delta, 2\alpha\rho) := \frac{2^{2/q-1}(1 + \mu_\alpha(\delta, 2\alpha\rho)^q)^{1/q}}{(1 - \mu_\alpha(\delta, 2\alpha\rho)^q)^{1/q}}, \qquad D_1(\delta, 2\alpha\rho) := \frac{2^{2/q-1}\beta(\delta, 2\alpha\rho)}{(1 - \mu_\alpha(\delta, 2\alpha\rho)^q)^{1/q}}$$

$$C_2(\delta, 2\alpha\rho) := \frac{2^{2/q-2}(\beta(\delta, 2\alpha\rho) + 1 - \sqrt{2})}{(1 - \mu_\alpha(\delta, 2\alpha\rho)^q)^{1/q}},$$

$$D_2(\delta, 2\alpha\rho) := \frac{2^{1/q-2}\beta(\delta, 2\alpha\rho)(\beta(\delta, 2\alpha\rho) + 1 - \sqrt{2})}{(1 - \mu_\alpha(\delta, 2\alpha\rho)^q)^{1/q}} + 2\beta(\delta, 2\alpha\rho) \tag{3.13}$$

with $\beta(\delta, \rho) := (1 + \sqrt{2})\frac{1+U(\delta,\rho)}{1-L(\delta,\rho)}$.

Unlike Theorem 5 which specifies one function $\rho_S^{RIP}(\delta)$ which bounds from below the phase transition for ℓ_1/ℓ_0-equivalence, Theorem 7 specifies a multiparameter family of threshold functions depending on q and possibly with further dependence on bounds on the multiplicative stability factors, such as $C_1(\delta, \rho)$. The function $\rho_S^{RIP}(\delta)$ in Theorem 5 corresponds to the case $\theta = 0$, $q = 1$, and no bounds on the stability parameters. The function $\rho_S^{RIP}(\delta; q)$ in Theorem 7 corresponds to ℓ^q regularization with unbounded stability coefficients, and as a result, it is only meaningful for ρ strictly below $\rho_S^{RIP}(\delta; q)$ or in the case where there exists a k sparse solution to $Ax_\theta = b$ and $\theta = 0$. More generally, specifying a bound on one or more of the multiplicative stability factors determines functions $\rho_S^{RIP}(\delta; q, Cond)$. For instance, imposing a bound of Υ on the stability factor $C_1(\delta, \rho)$ generates a function $\rho_S^{RIP}(\delta; q, C_1(\delta, \rho) \le \Upsilon)$; Figure 3.2 shows $\rho_S^{RIP}(\delta; q, C_1(\delta, \rho) \le \Upsilon)$ for $q = 1$ and $q = 1/2$ in panels (a-b) and (c-d) respectively. Software is available upon request which will generate functions with these and other choices of parameters in Theorem 7.

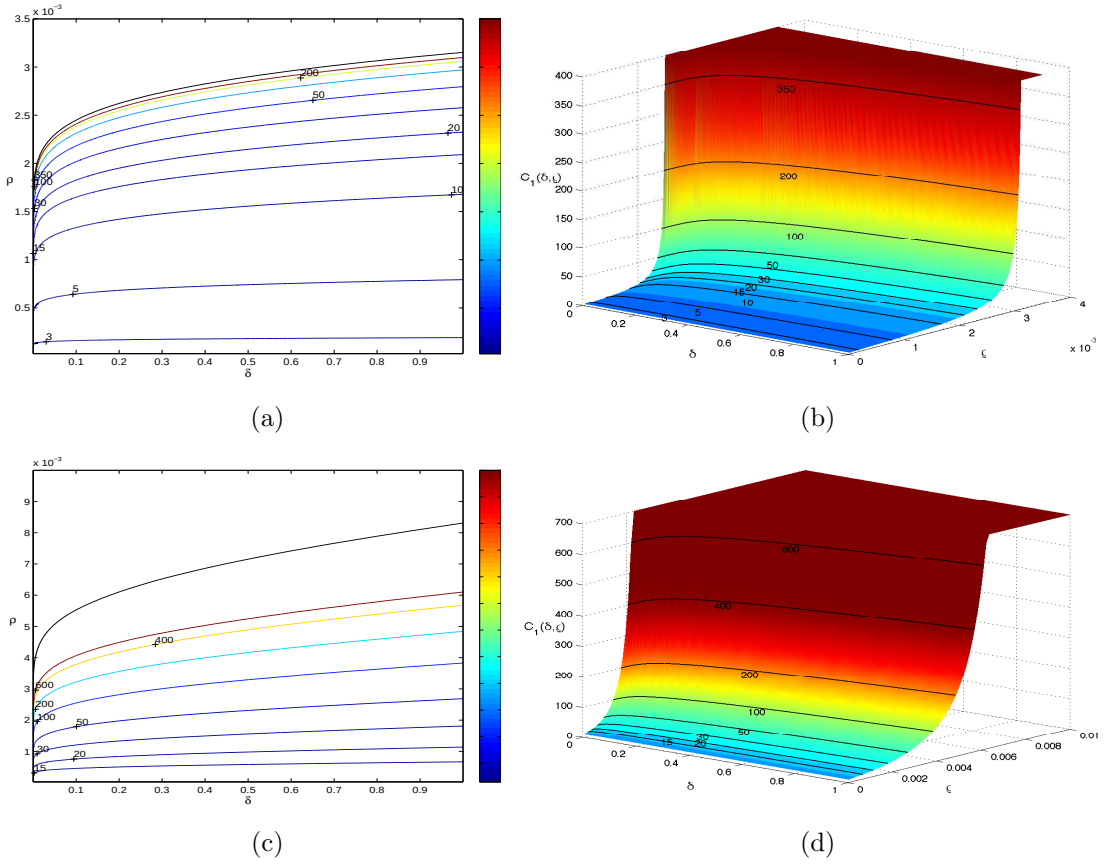


FIG. 3.2. *The surface whose level curves specify* $\rho_S^{RIP}(\delta; q, C_1(\delta,\rho) \leq \Upsilon)$ *for* $q = 1$ *and* $q = 1/2$ *in Panels (b) and (d) respectively, with level curves for specific values of* Υ *shown in Panels (a) and (c) for* $q = 1$ *and* $q = 1/2$ *respectively*

4. Discussion. Section 3.1 states three lower bounds on the phase transition curve of ℓ_1/ℓ_0-equivalence for *all* $x \in \chi^N(k)$, with high probability on the draw of A with Gaussian i.i.d. entries. Although Theorems 2, 3, and 4 each have the same goal of quantifying the ℓ_1/ℓ_0-equivalence phase transition for Gaussian random matrices, they are arrived at using substantially different methods of analysis. The efficacy of the three phase transitions going from the largest lower bound to the least lower bound are $\rho_S(\delta; C^N)$ of Donoho, $\rho_S^{RV}(\delta)$ of Rudelson and Vershynin, and $\rho_S^{RIP}(\delta)$ of Foucart and Lai, see Figure 3.1 (a). As we are seeking the largest lower bound, the results of Donoho supplant the other two. In fact, Donoho and Tanner showed that in the region most applicable for compressed sensing, namely, small n/N, the ℓ_1/ℓ_0-equivalence phase transition of Donoho approaches the exact ℓ_1/ℓ_0-equivalence phase transition [24] and gave a simple expression for the ℓ_1/ℓ_0-equivalence phase transition in this limit,

$$\rho_S(\delta; C^N) \to [2e \log(N/n)]^{-1} \quad \text{as} \quad n \to \infty \quad \text{and} \quad n/N \to 0.$$

That the phase transitions of Donoho are the most accurate is not surprising, as they were arrived at by making a precise "if and only if" formulation of ℓ_1/ℓ_0-equivalence in terms of randomly projected polytopes, which is amenable to a precise analysis by extending the work of Affentranger and Schneider [1]. In contrast, Rudelson and Vershynin derived $\rho_S^{RV}(\delta)$ in order to show that existing techniques of geometric functional analysis suffice to prove results in sparse approximation. Lastly, $\rho_S^{RIP}(\delta)$ is unsurprisingly the least accurate of all, as it only employs a generic property of sparse

vectors, the RIP, which has no inherent connection with ℓ^1-regularization. Moreover, the way in which $\rho_S^{RIP}(\delta)$ is derived is principally designed to ensure stability of the regularization method (3.1); although it is interesting to note that the original RIP based proof of ℓ_1/ℓ_0-equivalence by Candès and Tao [10] employed a separating hyperplane argument analogous to the geometric derivation of Donoho [15].

The phase transitions indicate the number of measurements needed as a multiple of the sparsity, $n > (\rho_S(\delta))^{-1} \cdot k$; the inverse of $\rho_S(\delta; C^N)$, $\rho_S^{RV}(\delta)$, and $\rho_S^{RIP}(\delta)$ are shown in the right panel of Figure 3.1 . As no computationally feasible method is known by which we can obtain this information for a given matrix A, it is crucial to have accurate analytical estimates of $\rho_S(\delta)$ in order for compressed sensing applications to benefit from the fewest number of measurements necessary. For any fixed ratio of n/N, the function $(\rho_S(\delta))^{-1}$ has a finite value independent of the magnitude of n and N. This is in stark contrast to the usual colloquial statement in compressed sensing that ℓ^1-regularization performs within a logarithm of the optimal rate; in fact, ℓ^1 regularization performs at the optimal rate of n proportional to k. This common misconception comes from the observation that as δ approaches zero, the exact phase transition $\rho_S(\delta)$ approaches $|2e\log(\delta)|^{-1}$; in other words, the necessary and sufficient number of measurements to ensure ℓ_1/ℓ_0-equivalence for all $x \in \chi^N(k)$ using Gaussian random matrices approaches $n \geq 2e \cdot k \cdot \log(N/n)$ as $\delta \to 0$ and $n \to \infty$, [24]. Despite this optimal rate for fixed n/N, unless the constant of proportionality of n compared to k is moderate, compressed sensing will only be of practical interest for N extremely large compared to k; from Theorems 2, 3, and 4 they are bounded below by: $n \geq 5.9k$, $n \geq 56k$, and $n \geq 317k$ respectively.

Although the geometric techniques of Donoho as well as Rudelson and Vershynin have substantially larger lower bounds on the phase transition for ℓ_1/ℓ_0-equivalence than implied by the current results following from the RIP, the geometric approaches have not as of yet been extended to ensure stability analogous to that implied by RIP in Theorem 6. The lower bound on the ℓ_1/ℓ_0-equivalence phase transition implied by the RIP for strictly k sparse signals, $\rho_S^{RIP}(\delta)$ of Theorem 4, does not have any implied stability. In order to ensure stability, Theorem 4 requires further restrictive bounds on the *stability factors* in Theorem 6, further reducing the lower bound on the phase transition. For example, $C_1(\delta, \rho)$ for $q = 1$ is shown in Figure 3.2 (b), with level curves of $C_1(\delta, \rho)$ corresponding to fixed stability factors proceeding (3.7) in Theorem 6; phase transitions below which specified bounds on $C_1(\delta, \rho)$ can be ensured are shown in Figure 3.2 (a-b). The stability factor becomes unbounded at finite ρ as $\rho \uparrow \rho_S^{RIP}(\delta)$. A similar behavior is expected for stability bounds that could be derived by extending the work of Donoho or Rudelson and Vershynin, although the rate at which the phase transition decreases with respect to bounds on stability factors is unknown.

A current trend in sparse approximation is to consider ℓ^q-regularization for $q \in (0, 1)$, with the aim of increasing the recoverability region [12, 13]. Existing results have shown that indeed the region where ℓ^q-regularization successfully recovers k sparse vectors at least does not decrease as q decreases [32], though little is known as to the rate, if any, at which it increases. Theorem 6 gives lower bounds on these regions where ℓ^q-regularization is guaranteed to have specified recoverability properties, and in fact for any strictly k sparse vector it implies that if (3.3) is finite, there is a small enough q such every k sparse vector can be recovered exactly from (b, A) by solving (3.6) with $\theta = 0$. Despite this and other encouraging results, many fundamental questions about ℓ^q-regularization remain, in particular how to find the global minimizer of (3.6). Moreover, it is unknown if ℓ^q-regularization remains stable as q decreases. In order to ensure stability, Theorem 4 requires further restrictive bound on the stability factor in Theorem 6, further reducing the lower bound on the phase transition. For example, $C_1(\delta, \rho)$ for $q = 1/2$ is shown in Figure 3.2 (d), with level curves of $C_1(\delta, \rho)$ corresponding to fixed stability factors preceeding (3.7) in Theorem 6; phase transitions below which specified bounds on $C_1(\delta, \rho)$ can be ensured are shown in Figure 3.2 (c-d). Decreasing q from 1 to $1/2$ does increase the value of ρ at which the stability factors in Theorem 6 become unbounded; however, comparing Figure 3.2 (d) and (b) it is apparent that this elevating of the unstable phase transition comes at the price of also elevating $C_1(\delta, \rho)$ for small values of ρ. In particular, the region

where $C_1(\delta,\rho) \leq 50$ is, in fact, larger for $q = 1$ than for $q = 1/2$.

Finite dimensional analogues of Theorems 2 and 3 exist, [23, 38], but in each case they 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.

It should be noted that the analysis presented here concerns properties that are guaranteed to hold for *all* $x \in \chi^N(k)$, whereas practitioners observe the average behavior over the set of $x \in \chi^N(k)$. This adherence to considering the worst case behavior over all $x \in \chi^N(k)$ is in stark contrast to the probabilistic nature of the same statements when considering the matrix A. Average case analysis of ℓ_1/ℓ_0-equivalence has been studied by Candès, Romberg and Tao using eigen-analysis similar to the RIP [7, 8], and by Donoho [15], as well as Donoho and Tanner [21, 22, 24] using geometric arguments similar to those used to derive Theorem 2. The average case ℓ_1/ℓ_0-equivalence phase transitions of Donoho and Tanner are markedly higher than those discussed in this article, and are in fact consistent with empirical observation. It is also worth noting that there are matrix ensembles which are observed to have good average case complexity, yet their RIP constants diverge much more rapidly from 1 than do those of Gaussian matrices. In particular, consider a Bernoulli random matrix B whose entries are drawn i.i.d. from -1,1 with equal probability, and $\tilde{B} := B + 1_n 1_N^T$ with entries drawn i.i.d. from 0,2 with equal probability. From a compressed sensing perspective B and $\tilde{B}$ differ only by adding $\sum_{i=1}^N x_i$ to the measurements b, which is observed to make no difference in the average behavior of ℓ^1-regularization, but the RIP constants of $\tilde{B}$ diverge from zero much more rapidly than do those of B. Such examples are important reminders that RIP based results are sufficient, but not necessary.

5. Proof of Main Results.

5.1. Proof of Theorem 1. The probability distribution functions for the eigenvalues of various random matrix ensembles have been extensively studied, and in many instances precise formulae are known, see [28] and references therein. However, the probability distribution functions associated with $L(k, n, N)$ and $U(k, n, N)$ are only of recent interest, and their precise formulae are not known. Results in sparse approximation which rely on the knowledge of the RIP constants make due with relatively crude bounds on the underlying probability distribution functions. These bounds of the RIP constants of a matrix A are typically constructed using concentration of measure bounds on the singular values of submatrices of A of size $n \times k$ and employing a simple union bound over all $\binom{N}{k}$ such submatrices, [10]. Theorem 1 follows in a similar way, but with more accurate bounds on the probability distribution functions for the largest and smallest singular values of Gaussian random matrices and an asymmetric formulation of the RIP in Definition 1 which yield more accurate quantitative estimates. As indicated in Section 2, the asymmetric RIP constants $L(k, n, N)$ and $U(k, n, N)$ are determined by the most extreme of the smallest and largest singular values of the $\binom{N}{k}$ submatrices A_Λ of size $n \times k$ derived by taking k columns from A of size $n \times N$. Here we focus on A with entries drawn i.i.d. from the Gaussian Normal distribution $\mathcal{N}(0, 1/\sqrt{n})$, and the largest and smallest eigenvalues of the $\binom{N}{k}$ derived Wishart matrices $A_\Lambda^T A_\Lambda$ of size $k \times k$.

We focus on the slightly more technical results for the 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 very similar proofs omitted. The probability distribution function, $f_{max}(k, n; \lambda)$, for the largest eigenvalue of the $k \times k$ Wishart matrix $A_\Lambda^T A_\Lambda$ was determined by Edelman in [27]. For our purposes, a simplified upper bound suffices:

LEMMA 5.1 (Lemma 4.2, pp. 550 [27]). *Let A_Λ be a matrix of size $n \times k$ whose entries are drawn i.i.d. from $\mathcal{N}(0, 1/\sqrt{n})$. Let $f_{max}(k, n; \lambda)$ denote the probability distribution function for the largest eigenvalue of the derived Wishart matrix $A_\Lambda^T A_\Lambda$ of size $k \times k$. Then $f_{max}(k, n; \lambda)$ satisfies:*

$$f_{max}(k, n; \lambda) \leq \left[(2\pi)^{1/2} (n\lambda)^{-3/2} \left(\frac{n\lambda}{2}\right)^{(n+k)/2} \frac{1}{\Gamma(\frac{k}{2})\Gamma(\frac{n}{2})} \right] \cdot e^{-n\lambda/2} =: g_{max}(k, n; \lambda). \tag{5.1}$$

With Lemma 5.1, we establish a bound on the asymptotic behavior of the distribution of the largest eigenvalue of a Wishart matrix of size $k \times k$.

LEMMA 5.2. *Let A_Λ be matrix of size $n \times k$ whose entries are drawn i.i.d. from $\mathcal{N}(0, 1/\sqrt{n})$. As $n \to \infty$ with $k/n \to \rho \in (0,1)$*

$$f_{max}(k,n;\lambda) \leq p_{max}(n,\lambda)\exp(n \cdot \psi_{max}(\lambda,\rho)) \tag{5.2}$$

where

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

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

Proof. Let $g_{max}(k,n;\lambda)$ be as defined in (5.1). Then with $\rho_n = \frac{k}{n}$ 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, $\lim_{n\to\infty}\Phi_1(k,n;\lambda) = 0$. To simplify Φ_3, we apply the second of Binet's log gamma formulas [42], 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

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

Therefore 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. □

In our pursuit of a bound on $U(k,n,N)$, we must simultaneously account for all $\binom{N}{k}$ Wishart matrices $A_\Lambda^T A_\Lambda$ derived from A. As such, we perform a large deviation analysis and study the asymptotic behavior of the distributions as $(k,n,N) \to \infty$, $\frac{k}{n} \to \rho$, and $\frac{n}{N} \to \delta$, by investigating $\lim_{N\to\infty} N^{-1}\log\left[\binom{N}{k} g_{max}(k,n;\lambda)\right]$. Using Lemma 5.2, we observe

$$\begin{aligned}
\lim_{N\to\infty}\frac{1}{N}\log\left[\binom{N}{k}g_{max}(k,n;\lambda)\right] &= \lim_{N\to\infty}\frac{1}{N}\log\left[\binom{N}{k}\right] + \lim_{N\to\infty}\frac{1}{N}\log\left[g_{max}(n,k;\lambda)\right] \\
&= H\left(\frac{k}{N}\right) + \delta\lim_{n\to\infty}\frac{1}{n}\log\left[g_{max}(n,k;\lambda)\right] \\
&= H(\rho\delta) + \delta\psi_{max}(\lambda,\rho). \qquad (5.3)
\end{aligned}$$

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 5.2 and (5.3), Proposition 5.3 establishes $\lambda_{max}(\delta,\rho) - 1$ as an upper bound on $U(k,n,N)$ as $n \to$ with $\frac{k}{n} \to \rho$, $\frac{n}{N} \to \delta$ and with $\rho, \delta \in (0,1)$.

PROPOSITION 5.3. *Let $\delta, \rho \in (0,1)$, and A be a matrix of size $n \times N$ whose entries are drawn i.i.d. from $\mathcal{N}(0, 1/\sqrt{n})$. Define $\tilde{U}(\delta,\rho) := \lambda_{max}(\delta,\rho) - 1$ where $\lambda_{max}(\delta,\rho)$ is the solution to (2.6). Let $(k,n,N) \to \infty$ with $\frac{k}{n} \to \rho$ and $\frac{n}{N} \to \delta$, then for any $\epsilon > 0$,*

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

exponentially in n.

Proof. Fix δ, ρ. From (5.3), define $h(\lambda) = \delta^{-1}H(\rho\delta) + \psi_{max}(\lambda, \rho)$. Since

$$h'(\lambda) = \frac{d}{d\lambda}\psi_{max}(\lambda,\rho) = \frac{1}{2}\left(\frac{1+\rho}{\lambda} - 1\right),$$

h is strictly decreasing on $[1+\rho, \infty)$; moreover, $h(\lambda)$ is strictly concave. By definition of $\lambda_{max}(\delta,\rho)$, $h(\lambda_{max}(\delta,\rho)) = 0$. Observe that $\lim_{\lambda\to\infty} h(\lambda) = -\infty$,

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

Therefore there is a unique solution to (2.2), $\lambda_{max}(\delta,\rho) \in (1+\rho, \infty)$.

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 $\lambda_{max}(\delta_n, \rho_n)$:

$$\begin{aligned} Prob\left[U(k,n,N) > \tilde{U}(\delta_n,\rho_n) + \epsilon)\right] &= Prob\left[U(k,n,N) > \lambda_{max}(\delta_n,\rho_n) - 1 + \epsilon)\right] \\ &= Prob\left[1 + U(k,n,N) > \lambda_{max}(\delta_n,\rho_n) + \epsilon)\right] \\ &= \binom{N}{k}\int_{\lambda_{max}(\delta_n,\rho_n)+\epsilon}^{\infty} f_{max}(k,n;\lambda)d\lambda \\ &\le \binom{N}{k}\int_{\lambda_{max}(\delta_n,\rho_n)+\epsilon}^{\infty} g_{max}(k,n;\lambda)d\lambda. \end{aligned} \tag{5.4}$$

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

$$\begin{aligned} \int_{\lambda_{max}(\delta_n,\rho_n)+\epsilon}^{\infty} g_{max}(k,n;\lambda)d\lambda &\le \varphi(n,\rho_n)\left(\lambda_{max}(\delta_n,\rho_n)+\epsilon\right)^{\frac{n}{2}(1+\rho_n)}e^{-\frac{n}{2}(\lambda_{max}(\delta_n,\rho_n)+\epsilon)}\int_{\lambda_{max}(\delta_n,\rho_n)+\epsilon}^{\infty}\lambda^{-\frac{3}{2}}d\lambda \\ &= \left(\lambda_{max}(\delta_n,\rho_n)+\epsilon\right)^{\frac{3}{2}} g_{max}\left(k,n;\lambda_{max}(\delta_n,\rho_n)+\epsilon\right)\int_{\lambda_{max}(\delta_n,\rho_n)+\epsilon}^{\infty}\lambda^{-\frac{3}{2}}d\lambda \\ &= 2\left(\lambda_{max}(\delta_n,\rho_n)+\epsilon\right) g_{max}\left(k,n;\lambda_{max}(\delta_n,\rho_n)+\epsilon\right). \end{aligned} \tag{5.5}$$

Therefore, combining (5.4) and (5.5) we obtain

$$\begin{aligned} Prob\left[U(k,n,N) > \tilde{U}(\delta_n,\rho_n)+\epsilon)\right] &\le 2\left(\lambda_{max}(\delta_n,\rho_n)+\epsilon\right)\binom{N}{k} g_{max}\left(k,n;\lambda_{max}(\delta_n,\rho_n)+\epsilon\right) \\ &\le p_{max}\left(n,\lambda_{max}(\delta_n,\rho_n)\right)\exp\left[n\cdot h\left(\lambda_{max}(\delta_n,\rho_n)+\epsilon\right)\right] \\ &\le p_{max}\left(n,\lambda_{max}(\delta_n,\rho_n)\right)\exp\left[n\epsilon\cdot h'\left(\lambda_{max}(\delta_n,\rho_n)\right)\right], \end{aligned} \tag{5.6}$$

with the last inequality following from $h(\lambda)$ being strictly concave. Since $h'\left(\lambda_{max}(\delta,\rho)\right) < 0$ is strictly bounded away from zero and all limits of (δ_n,ρ_n) are smoothly varying functions 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 (5.3), from the union bound over all $\binom{N}{k}$ matrices $A_\Lambda^T A_\Lambda$, results in an overly pessimistic bound for $\rho\delta \approx 1/2$. 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 5.4. *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 5.3. Let $(k,n,N) \to \infty$ with $\frac{k}{n} \to \rho$ and $\frac{n}{N} \to \delta$, then for any $\epsilon > 0$,*

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

exponentially in n.

Proof. By the definition of $\chi^N(k)$ in (1.3), $U(j,n,N) \geq U(k,n,N)$ for $j = k+1, k+2, \ldots, n$; combined with Proposition 5.3 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 distribution function, $f_{min}(k,n;\lambda)$, for the smallest eigenvalue of the $k \times k$ Wishart matrix $A_\Lambda^T A_\Lambda$ [27]. Here again, a simplified upper bound suffices:

LEMMA 5.5 (Prop. 5.2, pp. 553 [27]). *Let A_Λ be a matrix of size $n \times k$ whose entries are drawn i.i.d. from $\mathcal{N}(0,1/\sqrt{n})$. Let $f_{min}(k,n;\lambda)$ denote the probability distribution function for the smallest eigenvalue of the derived Wishart matrix $A_\Lambda^T A_\Lambda$ of size $k \times k$. Then $f_{min}(k,n;\lambda)$ satisfies:*

(5.7)

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

With Lemma 5.5, we establish a bound on the asymptotic behavior of the distribution of the smallest eigenvalue of Wishart matrix of size $k \times k$.

LEMMA 5.6. *Let A_Λ be matrix of size $n \times k$ whose entries are drawn i.i.d. from $\mathcal{N}(0,1/\sqrt{n})$. As $n \to \infty$ with $k/n \to \rho$*

$$f_{min}(k,n;\lambda) \leq p_{min}(n,\lambda) \exp(n \cdot \psi_{min}(\lambda,\rho)) \tag{5.8}$$

where

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

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

With Lemma 5.6, the large deviation analysis yields

$$\lim_{N\to\infty} \frac{1}{N} \log\left[\binom{N}{k} g_{min}(k,n;\lambda)\right] = H(\rho\delta) + \delta\psi_{min}(\lambda,\rho). \tag{5.9}$$

Similar to the proof of Proposition 5.3, Lemma 5.6 and (5.9) are used to establish $L(\delta,\rho)$ as an upper bound on $L(k,n,N)$ for large problem instances.

PROPOSITION 5.7. *Let $\delta,\rho \in (0,1]$, and A be a matrix of size $n \times N$ whose entries are drawn i.i.d. from $\mathcal{N}(0,1/\sqrt{n})$. Define $L(\delta,\rho) := 1 - \lambda_{min}(\delta,\rho)$ where $\lambda_{min}(\delta,\rho)$ is the solution to (2.5). Let $(k,n,N) \to \infty$ with $\frac{k}{n} \to \rho$, $\frac{n}{N} \to \delta$, then for any $\epsilon > 0$,*

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

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 5.4.

5.2. Proof of Theorem 5. Theorem 5 follows from Theorem 4 and the probabilistic bounds on the RIP constants in Theorem 1. In particular, Lemma 5.8 bounds $\mu(2k,n,N)$ in terms of the asymptotic RIP bounds $L(\delta,\rho)$ and $U(\delta,\rho)$, by the quantity $\mu(\delta,2(1+\epsilon)\rho)$ defined below (5.10). If $\rho_\epsilon(\delta)$ is the solution to $\mu(\delta,2(1+\epsilon)\rho)=1$, then for $\rho<\rho_\epsilon(\delta)$ we achieve the desired bound, $\mu(2k,n,N)<1$, to ensure ℓ_1/ℓ_0-equivalence for all $x\in\chi^N(k)$. The statement of Theorem 5 follows from relating $\rho_\epsilon(\delta)$ to $\rho_S^{RIP}(\delta)$, the solution to $\mu(\delta,2\rho)=1$.

LEMMA 5.8. *For any $\epsilon>0$, 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 with Gaussian i.i.d. entries that*

$$\mu(2k,n,N)<\frac{1+\sqrt{2}}{4}\left(\frac{1+U(\delta,2(1+\epsilon)\rho)}{1-L(\delta,2(1+\epsilon)\rho)}-1\right)=\mu(\delta,2(1+\epsilon)\rho). \tag{5.10}$$

Proof. Theorem 1 and the form of $\mu(\delta,2\rho)$ 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 with Gaussian i.i.d. entries that

$$\mu(2k,n,N)<\frac{1+\sqrt{2}}{4}\left(\frac{1+U(\delta,2\rho)+c\epsilon}{1-L(\delta,2\rho)-c\epsilon}-1\right). \tag{5.11}$$

Since $U(\delta,\rho)$ is nondecreasing 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 (5.11) can be bounded by the right-hand side of (5.10) for any fixed ϵ satisfying $0<\epsilon<\frac{1}{2\rho}-1$, by setting

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

(The upper bound on ϵ is imposed so that the second argument of $\mu(\delta,\cdot)$, $2(1+\epsilon)\rho$, is in the admissible range of $(0,1)$.) That the bound (5.10) is satisfied for all $\epsilon>0$ sufficiently small, and that the right hand side of (5.10) is strictly increasing in ϵ establishes that (5.10) 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 5] Let $\rho_\epsilon(\delta)$ be the solution of $\mu(\delta,2(1+\epsilon)\rho)=1$. Then, for any $\rho<\rho_\epsilon(\delta)$, Lemma 5.8 implies that $\mu(2k,n,N)<1$, which by Theorem 4, provides the desired ℓ_1/ℓ_0-equivalence. To remove the dependence on the level curve $\rho_\epsilon(\delta)$, note that $\rho_\epsilon(\delta)$ is related to $\rho_S^{RIP}(\delta)$, the solution of $\mu(\delta,2\rho)=1$, by $(1+\epsilon)\rho_\epsilon(\delta)\equiv\rho_S^{RIP}(\delta)$. Since $(1-\epsilon)\le(1+\epsilon)^{-1}$ for all $\epsilon\in[0,1]$, we have $(1-\epsilon)\rho_S^{RIP}(\delta)<\rho_\epsilon(\delta)$. Thus, provided $\rho<(1-\epsilon)\rho_S^{RIP}(\delta)$, the statement of Theorem 5 is satisfied. □

Theorem 7 follows similarly from Theorem 6 and an analogous argument to that of Theorem 5; we state the essential intermediate Lemma required, but omit its proof for brevity.

LEMMA 5.9. *For any $\epsilon>0$, with $n/N\to\delta\in(0,1)$ and $k/n\to\rho\in(0,1/2]$ as $(k,n,N)\to\infty$, and for $1\le\alpha\le1/2\rho$, there is an exponentially high probability on the draw of A with Gaussian i.i.d. entries that*

$$\mu(2k\alpha,n,N)<\mu(\delta,2\alpha(1+\epsilon)\rho). \tag{5.12}$$

REFERENCES

[1] F. Affentranger and R. Schneider. Random projections of regular simplices. *Discrete Comput. Geom.*, 7(3):219–226, 1992.

[2] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde. Model-based compressive sensing. *IEEE Trans. Inform. Theory*, submitted.

[3] D. Baron, M. B. Wakin, M. F. Duarte, S. Sarvotham, and R. G. Baraniuk. Distributed compressed sensing. *IEEE Trans. Inform. Theory*, submitted.
[4] J. D. Blanchard, C. Cartis, J. Tanner, and A. Thompson. Sparse approximation recoverability regions for cosamp and romp with gaussian random matrices. submitted, 2008.
[5] E. J. Candès. Compressive sampling. In *International Congress of Mathematicians. Vol. III*, pages 1433–1452. Eur. Math. Soc., Zürich, 2006.
[6] 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.
[7] E. J. Candès and J. Romberg. Quantitative robust uncertainty principles and optimally sparse decompositions. *Foundations of Comp. Math.*, 6(2):227–254, 2006.
[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. K. 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. Submitted, 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] D. L. Donoho. Neighborly polytopes and sparse solution of underdetermined linear equations. Technical Report, Department of Statistics, Stanford University, 2005.
[16] D. L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, 2006.
[17] D. L. Donoho. For most large underdetermined systems of equations, the minimal l_1-norm near-solution approximates the sparsest near-solution. *Comm. Pure Appl. Math.*, 59(7):907–934, 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. Neighborliness of randomly-projected simplices in high dimensions. *Proc. Natl. Acad. Sci. USA*, 102(27):9452–9457, 2005.
[21] 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.
[22] D. L. Donoho and J. Tanner. Counting the faces of randomly-projected hypercubes and orthants, with applications. submitted, 2008.
[23] 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.
[24] 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.
[25] D. L. Donoho and Y. Tsaig. Fast solution of l1 minimization problems when the solution may be sparse. submitted, 2006.
[26] 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.
[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. Gradier projection for sparse reconstruction: Application to compressed sensing and other inverse problems. *IEEE Trans. Inform. Theory*, to appear.
[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.*, 2008. In press.
[31] S. Geman. A limit theorem for the norm of random matrices. *Ann. Probab.*, 8(2):252–261, 1980.
[32] R. Gribonval and M. Nielsen. Sparse representations in unions of bases. *IEEE Trans. Inform. Theory*, 49(12):3320–3325, 2003.
[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] B. K. Natarajan. Sparse approximate solutions to linear systems. *SIAM J. Comput.*, 24(2):227–234, 1995.
[36] D. Needell and J. Tropp. Cosamp: Iterative signal recovery from incomplete and inaccurate samples. *Appl. Comp. Harm. Anal.*, in press.
[37] D. Needell and R. Vershynin. Uniform uncertainty principle and signal recovery via regularized orthogonal

matching pursuit. *Foundations of Comp. Math.*, in press.

[38] M. Rudelson and R. Vershynin. On sparse reconstruction from Fourier and Gaussian measurements. *Comm. Pure Appl. Math.*, 61(8):1025–1045, 2008.

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

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

[41] 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.

[42] 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.

[43] 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.