On Support Sizes of Restricted Isometry Constants

Jeffrey D. Blanchard[*,a,1], Andrew Thompson[b]

[a]*Department of Mathematics and Statistics, Grinnell College, Grinnell, Iowa 50112-1690, USA.*
[b]*School of Mathematics and the Maxwell Institute, University of Edinburgh, King's Buildings, Mayfield Road, Edinburgh EH9 3JL, UK.*

Abstract

A generic tool for analyzing sparse approximation algorithms is the restricted isometry property (RIP) introduced by Candès and Tao. If $R(k,n,N)$ is the RIP constant with support size k for an $n \times N$ measurement matrix, we investigate the trend of reducing the support size of the RIP constants for qualitative comparisons between sufficient conditions. For example, which condition is easier to satisfy, $R(4k,n,N) < 0.1$ or $R(2k,n,N) < 0.025$? Using a quantitative comparison via phase transitions for Gaussian measurement matrices, three examples from the literature of such support size reduction are considered. In each case, utilizing a larger support size for the RIP constants results in a sufficient condition for exact sparse recovery that is satisfied by a significantly larger subset of Gaussian matrices.

Key words: Compressed sensing, restricted isometry constants, restricted isometry property, sparse approximation, sparse signal recovery

1. Introduction

In sparse approximation and compressed sensing, [8, 11, 14], one seeks to recover a compressible, or simply sparse, signal from a limited number of linear measurements. This is generally modeled by applying an underdetermined measurement matrix A of size $n \times N$ to a signal $x \in \mathbb{R}^N$ known to be compressible or k-sparse. Having obtained the measurements $y = Ax$, a nonlinear reconstruction technique is applied which seeks a sparse signal returning these measurements. Numerous reconstruction algorithms have been analyzed using a generic tool introduced by Candès and Tao [11], namely the *restricted isometry property* (RIP). To account for the asymmetry about 1 of the singular values of submatrices of the measurement matrix A, the *asymmetric restricted isometry property* (ARIP) has recently been introduced [1, 21]. We denote the set of all k-sparse signals by $\chi^N(k) = \{x \in \mathbb{R}^N : \|x\|_0 \le k\}$, where $\|x\|_0$ counts the number of nonzero entries of x.

Definition 1 (RIP [11] and ARIP [1])**.** For an $n \times N$ matrix A, the *asymmetric RIP constants* $L(k,n,N)$ and $U(k,n,N)$ are defined as:

$$L(k,n,N) := \min_{c \ge 0} c \;\text{ subject to }\; (1-c)\|x\|_2^2 \le \|Ax\|_2^2, \;\text{ for all }\; x \in \chi^N(k); \tag{1}$$

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

The symmetric (standard) *RIP constant* $R(k,n,N)$ is then defined by

$$R(k,n,N) := \max\{L(k,n,N), U(k,n,N)\}. \tag{3}$$

*Corresponding author
Email addresses: `jeff@math.grinnell.edu` (Jeffrey D. Blanchard), `a.thompson-8@sms.ed.ac.uk` (Andrew Thompson)
[1]JDB was partially supported by NSF DMS grant 0602219 while a VIGRE postdoctoral fellow at Department of Mathematics, University of Utah.

We refer to the first argument of the RIP constants, k, as the *support size* of the RIP constant. In the analysis of sparse approximation algorithms, intuition and aesthetics have led to the desire to reduce the support size of the RIP constants to $2k$. Two important facts have played a substantial role in motivating the search for RIP statements with support size $2k$. First, in order to correctly recover a k-sparse signal, the measurement matrix A must be able to distinguish between any two signals in $\chi^N(k)$, therefore, necessitating[2] that $L(2k,n,N) < 1$. Secondly, the early sufficient conditions [10, 11] for successful k-sparse recovery via ℓ_1-regularization involved various support sizes. The results were eventually overshadowed by Candès's elegant sufficient RIP condition with support size of $2k$, i.e. $R(2k,n,N) < \sqrt{2}-1$, [9]. When analyzing alternative sparse approximation algorithms, qualitative comparisons to Candès's result motivate a desire to state results in terms of RIP constants with support size $2k$.

However, reducing the support size of an RIP constant is not necessarily quantitatively advantageous. Typically, sufficient conditions appear in the literature in the form $R(bk,n,N) < \alpha$ implies success of exact sparse recovery for a certain reconstruction algorithm. Since the RIP and ARIP constants are increasing as a function of their support size, $R(ak,n,N) \le R(bk,n,N)$ for $a \le b$, it is clear that there are two ways to weaken this condition. First, one could increase α, the bound on $R(bk,n,N)$. Second, one could decrease the support size while keeping α fixed: $R(ak,n,N) < \alpha$ for $a < b$. However, reducing the support size while simultaneously reducing the bound α is not necessarily quantitatively advantageous and is completely dependent on the growth rate of the RIP constants (as the support size grows) for a chosen matrix ensemble. Moreover, having a matrix ensemble where the RIP constants grow rapidly is certainly not desirable when trying to satisfy conditions of the form $R(bk,n,N) < \alpha$ for large values of k in proportion to n.

As the behavior of the RIP constants is dependent on the matrix ensemble, we focus on matrices from the *Gaussian ensemble*, i.e. matrices whose entries are selected i.i.d. from the normal distribution $\mathcal{N}(0,1/n)$. In this article, we employ the phase transition framework advocated by Donoho et al. [16, 17, 18, 19] and subsequently applied to the RIP [1, 2, 3]. The phase transition framework provides a method for quantitative comparison of results involving the RIP. The quantitative comparisons demonstrate that the desire for qualitative comparisons obtained by reducing the support size often leads to smaller regions where recovery can be guaranteed.

1.1. The Phase Transition Framework

Computing the RIP constants for a specific matrix is a combinatorial problem and therefore intractable for large matrices. In order to make quantitative comparisons, bounds on the probability density function for the RIP constants have been derived for various random matrix ensembles [1, 11, 12, 15]. The current best known bounds[3] for Gaussian matrices were derived in [1] and are denoted $\mathcal{L}(\delta,\rho), \mathcal{U}(\delta,\rho), \mathcal{R}(\delta,\rho)$ where δ and ρ define a proportional growth among the problem parameters (k,n,N).

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

The following is an adaptation of [1, Thm. 1].

Theorem 1 (Blanchard, Cartis, Tanner [1])**.** *Fix $\epsilon > 0$. Under the proportional-growth asymptotic, Definition 2, sample each $n \times N$ matrix A from the Gaussian ensemble. Let $\mathcal{L}(\delta,\rho)$ and $\mathcal{U}(\delta,\rho)$ be defined as in [1, Thm. 1]. Define $\mathcal{R}(\delta,\rho) = \max\{\mathcal{L}(\delta,\rho), \mathcal{U}(\delta,\rho)\}$. Then for any $\epsilon > 0$, as $n \to \infty$,*

$$\mathit{Prob}\,[L(k,n,N) < \mathcal{L}(\delta,\rho) + \epsilon] \to 1, \tag{4}$$

$$\mathit{Prob}\,[U(k,n,N) < \mathcal{U}(\delta,\rho) + \epsilon] \to 1, \tag{5}$$

$$\text{and} \quad \mathit{Prob}\,[R(k,n,N) < \mathcal{R}(\delta,\rho) + \epsilon] \to 1. \tag{6}$$

[2]An advantage of the asymmetric formulation of the RIP is that this is truly a necessary condition as opposed to the often stated, but not necessary requirement $R(2k,n,N) < 1$.

[3]Extensive empirical testing show these bounds are no more than twice the empirically observed RIP constants for all values of δ. For a detailed discussion, see [1].

The proof of this result appears in [1] with a more thorough explanation of its application to the phase transition framework. Briefly, the phase transition framework is applied to results obtained via the RIP in the following manner. For a given sparse approximation algorithm, for example ℓ_1-regularization, CoSaMP, or Iterative Hard Thresholding (IHT), a sufficient condition is derived from an RIP analysis of the measurement matrix A. This sufficient condition can be arranged to take the form $\mu(k, n, N) < 1$ where $\mu(k, n, N)$ is a function of the ARIP constants, $L(\cdot, n, N)$ and $U(\cdot, n, N)$, or, in a symmetric RIP analysis, it is a function of $R(\cdot, n, N)$. As the RIP constants are bounded by Theorem 1, one obtains a bound, $\mu(\delta, \rho)$ for the function $\mu(k, n, N)$ as $n \to \infty$ with $\frac{k}{n} \to \rho$ and $\frac{n}{N} \to \delta$. The lower bound on the phase transition for the associated sufficient condition for exact recovery of every $x \in \chi^N(k)$ is then determined by a function $\rho_S(\delta)$ which is the solution to the equation $\mu(\delta, \rho) = 1$. The curve defined by $\rho_S(\delta)$ graphically displays the lower bound on the strong (exact recovery of all k-sparse signals) phase transition defined by the sufficient condition $\mu(k, n, N) < 1$. When A is a Gaussian matrix of size $n \times N$ and the ordered pair $(\frac{n}{N}, \frac{k}{n})$ falls below the phase transition curve, then with overwhelming probability on the draw of A, the sufficient condition is satisfied and the algorithm will exactly recover every $x \in \chi^N(k)$, i.e. the algorithm exactly recovers all k-sparse signals.

In this letter, we utilize the phase transition framework to compare existing sufficient conditions for exact k-sparse recovery using matrices drawn from the Gaussian ensemble. We say a condition is *weaker* when it is satisfied with overwhelming probability (i.e. probability of the complementary event vanishing exponentially in n) by a larger portion of matrices drawn from the Gaussian ensemble. This is represented by a higher phase transition curve associated with the condition as a higher phase transition curve carves out a larger region of the phase space where exact k-sparse recovery is guaranteed with overwhelming probability. Likewise, we say a condition is *stronger* if the associated phase transition curve is lower and therefore defines a smaller region where exact k-sparse recovery is guaranteed with overwhelming probability.

1.2. Organization and Notation

In the following, we present three instances from the literature where reducing the support sizes of the RIP constants results in a stronger sufficient condition for sparse signal recovery. By using the quantitative comparisons available through the phase transition framework, outlined in Sec. 1.1, we examine three cases where larger RIP support sizes yield weaker sufficient conditions for exact k-sparse recovery with Gaussian measurement matrices. These three examples are certainly not exhaustive, but suffice in conveying the idea.

(i) For *Compressive Sampling Matching Pursuit* (CoSaMP) [22], Needell and Tropp apply a bound on the growth rate of RIP constants to reduce the support size of the RIP constants from $4k$ to $2k$ resulting in a significantly stronger sufficient condition. (Sec. 2)

(ii) The currently accepted state of the art sufficient condition for ℓ_1*-regularization* obtained by Foucart and Lai [21] involves RIP constants with support size $2k$. However, a sufficient condition involving RIP constants with support sizes $11k$ and $12k$ yields a weaker sufficient condition for exact k-sparse recovery via ℓ_1-regularization. (Sec. 3)

(iii) A technique of splitting support sets introduced by Blumensath and Davies [5] allows a reduction of the support size of RIP constants in the analysis of *Iterative Hard Thresholding* (IHT). In this case, the sufficient conditions for exact k-sparse recovery via IHT are again weaker with the larger support size of the RIP constants. (Sec. 4)

In the following, if S is an index set, then $|S|$ denotes the cardinality of S, A_S represents the submatrix of A obtained by selecting the columns indexed by S, and x_S is the set of entries of x indexed by S. Finally, even when not explicitly stated, it is assumed throughout that the support size of an RIP constant is no larger than the number of measurements, e.g. if A is of size $n \times N$ with RIP constant $R(mk, n, N)$, we implicitly assume $mk \leq n < N$.

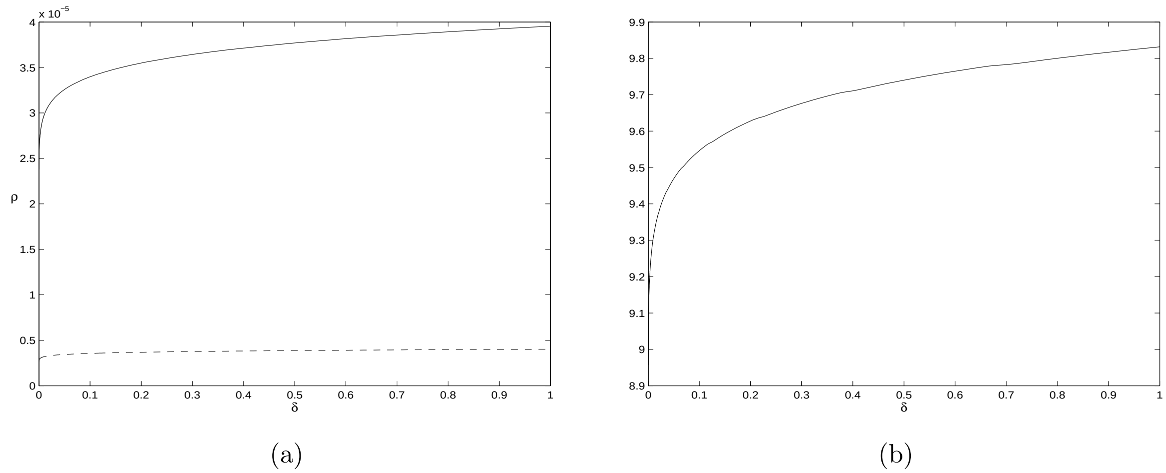


Figure 1: CoSaMP: (a) lower bounds on the exact k-sparse recovery phase transition for Gaussian random matrices; $\mathcal{R}(\delta, 4\rho) < .1$ (solid), $\mathcal{R}(\delta, 2\rho) < .025$ (dash-dash). (b) the ratio of the two curves in (a).

2. A Simple Example

A straightforward and dramatic example of a weaker condition with a larger RIP support size is found in the analysis of the greedy algorithm CoSaMP [22]. In this work, Needell and Tropp provide a sufficient condition for guaranteed recovery of k-sparse vectors, namely

$$R(4k, n, N) < 0.1. \tag{7}$$

The authors also provide a bound on the growth rate of RIP constants, $R(ck, n, N) \leq c \cdot R(2k, n, N)$, [22, Cor. 3.4]. For qualitative comparison to current results for ℓ_1-regularization, such as Candès's $R(2k, n, N) < \sqrt{2} - 1$, [9], Needell and Tropp apply their bound on the growth of RIP constants to obtain the condition

$$R(2k, n, N) < 0.025, \tag{8}$$

which is therefore also sufficient for the exact recovery of k-sparse signals [22, Remark 2.2]. Lacking a quantitative comparison of the two conditions, the authors do not claim that reducing the support size is advantageous, only that such a reduction is still sufficient for CoSaMP to exactly recover k-sparse signals. However, the condition involving the support size $4k$, (7), is considerably weaker than the condition with support size $2k$, (8), when applied to Gaussian random matrices. By employing a bound (described in Sec. 1.1) $\mathcal{R}(\delta, \rho)$ as $n \to \infty$ with $\frac{n}{N} \to \delta$, $\frac{k}{n} \to \rho$, lower bounds on the phase transition for exact k-sparse recovery via CoSaMP are obtained.

In Fig. 1(a), lower bounds on the phase transition are displayed for the two conditions. With overwhelming probability Gaussian matrices A of size $n \times N$ will satisfy the sufficient conditions for CoSaMP to exactly recover every $x \in \chi^N(k)$ provided the ordered pair $(\delta, \rho) \equiv (\frac{n}{N}, \frac{k}{n})$ falls below the associated phase transition curve in the phase space $[0, 1]^2$. For Gaussian matrices, $R(4k, n, N) < 0.1$ is a superior bound to $R(2k, n, N) < 0.025$ as the region of the phase space representing matrices satisfying (7) with high probability has greater than 9.7 times the area of the phase space region determined by (8). Moreover, Fig. 1(b) shows that the phase transition curve for the weaker ($4k$) condition is between 8.96 and 9.83 times higher depending on δ. This implies that the condition with larger support size guarantees CoSaMP recovery of signals with roughly nine times the number of nonzero entries as the signals guaranteed to be recovered by the sufficient condition with the reduced support size.

3. The RIP for ℓ_1-regularization

In this section we examine the current knowledge obtained from an RIP analysis for exact recovery of k-sparse signals via ℓ_1-regularization. The problem of finding the sparsest signal x equipped only with the measurement matrix A and the measurements $y = Ax$ is, in general, a combinatorial problem. It is now well understood that, under the right conditions, the solution to the tractable ℓ_1-regularization,

$$\min \|x\|_1 \text{ subject to } y = Ax, \tag{9}$$

is the unique, sparsest signal satisfying $y = Ax$.

Currently, it is generally accepted that the state of the art sufficient condition[4] obtained by RIP analysis for ℓ_1-regularization was proven by Foucart and Lai [21].

Theorem 2 (Foucart, Lai [21]). *For any matrix A of size $n \times N$ with ARIP constants $L(2k, n, N)$ and $U(2k, n, N)$, for $2k \leq n < N$, if $\mu^{fl}(k, n, N) < 1$ where*

$$\mu^{fl}(k, n, N) := \frac{1+\sqrt{2}}{4}\left(\frac{1+U(2k,n,N)}{1-L(2k,n,N)} - 1\right), \tag{10}$$

then ℓ_1-regularization will exactly recover every $x \in \chi^N(k)$.

Motivated by the fact that the symmetric RIP constants are not invariant to scaling the matrix, Theorem 2 is proven with an asymmetric RIP analysis. Their method of proof also resulted in a slight improvement to the symmetric RIP result of Candès mentioned above. With $R(2k, n, N)$ defined as in (3), a sufficient condition for exact recovery of every $x \in \chi^N(k)$ from ℓ_1-regularization is $R(2k, n, N) < \frac{2}{3+\sqrt{2}} \approx 0.4531$.

Note that Theorem 2 (and the result for $R(2k, n, N)$) involve RIP support sizes of $2k$. One of the early sufficient conditions for exact recovery of every $x \in \chi^N(k)$ by solving (9), namely $3R(4k, n, N)+R(3k, n, N) < 2$, was obtained by Candès, Romberg, and Tao [10] from an RIP analysis. Chartrand [13] extended this result to essentially arbitrary support sizes and to ℓ_q-regularization for $q \in (0, 1]$. Chartrand's extension was further studied by Saab and Yilmaz [23, 24]. Here, the result is stated with an asymmetric RIP analysis following the proof in [13] which, in turn, is an adaptation of Candès, Romberg, and Tao's proof [10].

Theorem 3. *Suppose $k \in \mathbb{N}^+$ and $b > 2$ with $bk \in \mathbb{N}^+$. For any matrix A of size $n \times N$ with ARIP constants $L([b + 1]k, n, N)$ and $U(bk, n, N)$ for $[b + 1]k \leq n < N$, if $\mu^{bt}(k, n, N; b) < 1$ where*

$$\mu^{bt}(k, n, N; b) := \frac{bL([b+1]k,n,N) + U(bk,n,N)}{b-1}, \tag{11}$$

then ℓ_1-regularization will exactly recover every $x \in \chi^N(k)$.

Proof. Let $x \in \chi^N(k)$ and $y = Ax$. Suppose z is a solution to (9). Define $h = z - x$. We demonstrate that $h = 0$. Let $T_0 = \text{supp}(x)$ and arrange the elements of $|h|$ on the complement, T_0^c, in decreasing order using the partition $T_0^c = T_1 \cup T_2 \cup \cdots \cup T_J$ with $|T_j| = bk$ for $j = 1, \ldots, J-1$ and $|T_J| \leq bk$. Denote $T_{01} = T_0 \cup T_1$. Counting arguments and norm comparisons taken directly from [13] provide the following inequalities,

$$\left\|h_{T_0^c}\right\|_1 \leq \|h_{T_0}\|_1\,, \tag{12}$$

$$\|A_{T_{01}} h_{T_{01}}\|_2 \leq \sum_{j=2}^{J} \left\|A_{T_j} h_{T_j}\right\|_2\,, \tag{13}$$

$$\sum_{j=2}^{J} \left\|h_{T_j}\right\|_2 \leq \frac{1}{\sqrt{bk}} \left\|h_{T_0^c}\right\|_1\,. \tag{14}$$

[4]This sufficient condition on $R(2k, n, N)$ was subsequently improved by Foucart[20] and by Cai et al.[7].

Now since $|T_{01}| = |T_0| + |T_1| = [b+1]k$ and $|T_j| \leq bk$ for each $j \geq 1$, then by Def. 1,

$$\|A_{T_{01}} h_{T_{01}}\|_2 \geq \sqrt{1 - L([b+1]k, n, N)}\, \|h_{T_{01}}\|_2 \tag{15}$$

$$\text{and} \quad \left\|A_{T_j} h_{T_j}\right\|_2 \leq \sqrt{1 + U(bk, n, N)}\, \left\|h_{T_j}\right\|_2. \tag{16}$$

Therefore, inserting (15) and (16) into (13) yields

$$\|h_{T_{01}}\|_2 \leq \sqrt{\frac{1 + U(bk, n, N)}{1 - L([b+1]k, n, N)}} \sum_{j=2}^{J} \left\|h_{T_j}\right\|_2. \tag{17}$$

Combining (12), (14), (17), and the standard relationship between norms, we then have

$$\begin{aligned} \|h_{T_{01}}\|_2 \leq \sqrt{\frac{1 + U(bk, n, N)}{1 - L([b+1]k, n, N)}} \frac{1}{\sqrt{bk}} \|h_{T_0}\|_1 &\leq \sqrt{\frac{1 + U(bk, n, N)}{1 - L([b+1]k, n, N)}} \sqrt{\frac{k}{bk}} \|h_{T_0}\|_2 \\ &\leq \sqrt{\frac{1 + U(bk, n, N)}{b - bL([b+1]k, n, N)}} \|h_{T_{01}}\|_2. \end{aligned} \tag{18}$$

Squaring and rearranging (18),

$$\big([b-1] - [bL([b+1]k, n, N) + U(bk, n, N)] \big) \|h_{T_{01}}\|_2^2 \leq 0. \tag{19}$$

The hypothesis $\mu(k, n, N; b) < 1$ ensures the left hand side of (19) is nonnegative and zero only when $\|h_{T_{01}}\|_2 = 0$. Therefore, $h = 0$ implying $z = x$. □

Following the framework described in Sec. 1.1 and more formally developed in [1], we restate Theorems 2 and 3 in the language of phase transitions for Gaussian matrices. Clearly, (11) defines a family of functions indexed by $b > 2$ with $bk \in \mathbb{N}^+$. Applying the bounds defined in Theorem 1 to (10) and (11), we obtain the functions

$$\mu^{fl}(\delta, \rho) := \frac{1 + \sqrt{2}}{4} \left(\frac{1 + \mathcal{U}(\delta, 2\rho)}{1 - \mathcal{L}(\delta, 2\rho)} - 1 \right) \quad \text{and} \quad \mu^{bt}(\delta, \rho; b) := \frac{b\mathcal{L}(\delta, [b+1]\rho) + \mathcal{U}(\delta, b\rho)}{b - 1}. \tag{20}$$

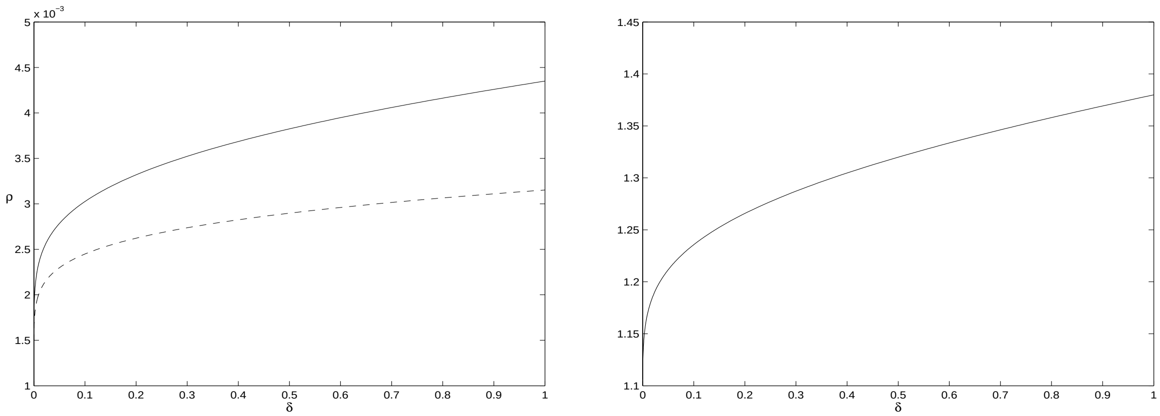
Now define $\rho_S^{fl}(\delta)$ as the solution to $\mu^{fl}(\delta, \rho) = 1$. Similarly, define $\rho_S^{bt}(\delta; b)$ as the solution to $\mu^{bt}(\delta, \rho; b) = 1$. The functions $\rho_S^{fl}(\delta)$ and $\rho_S^{bt}(\delta; b)$ define regions of the phase space which guarantee sparse recovery. We collect the phase transition formulation of Thms. 2 and 3 in Thm. 4 (i) and (ii), respectively.

Theorem 4. *Fix $\epsilon > 0$. Under the proportional-growth asymptotic, $\frac{n}{N} \to \delta$ and $\frac{k}{n} \to \rho$ as $n \to \infty$, sample each $n \times N$ matrix A from the Gaussian ensemble. Then with overwhelming probability every $x \in \chi^N(k)$ is exactly recovered by ℓ_1-regularization* (9)*, provided one of the following conditions is satisfied:*

(i) $\rho < (1 - \epsilon)\rho_S^{fl}(\delta)$;

(ii) $\rho < (1 - \epsilon)\rho_S^{bt}(\delta; b)$.

$\rho_S^{fl}(\delta)$ is displayed as the dash-dash curve in Fig. 2(a). The highest phase transition curves are obtained with $b \approx 11$. No single value of b provides a phase transition curve that is highest for all values of δ. For example, $\rho_S^{bt}(\delta; 10) > \rho_S^{bt}(\delta; 11)$ for $\delta \in [0, .44]$ and $\rho_S^{bt}(\delta; 10) < \rho_S^{bt}(\delta; 11)$ for $\delta \in [.45, 1]$. $\rho_S^{bt}(\delta; 11)$ is displayed as the solid curve in Fig. 2(a). The heavier weighting of the ARIP bound, $\mathcal{L}(\delta, [b+1]\rho)$, which for Gaussian matrices is less than 1 with probability 1, permits b to grow well beyond 2. Although intuitively murky on the surface, support sizes of $11k$ and $12k$ provide a larger region of Gaussian matrices which provably guarantee exact k-sparse recovery from an RIP analysis. Figure 2(b) shows an improvement by a factor



(a) (b)

Figure 2: ℓ_1-regularization: (a) lower bounds on the exact k-sparse recovery phase transition for Gaussian random matrices; $\rho_S^{bt}(\delta; 11)$ (solid) and $\rho_S^{fl}(\delta)$ (dash-dash). (b) The improvement ratio $\frac{\rho_S^{bt}(\delta;11)}{\rho_S^{fl}(\delta)}$.

ranging from 1.11 to 1.38. By using a quantitative comparison, we see that the weakest RIP condition is not Thm. 2, rather a weaker RIP sufficient condition for exact recovery of every $x \in \chi^N(k)$ via ℓ_1-regularization, at least for Gaussian matrices commonly used in compressed sensing, is

$$11L(12k, n, N) + U(11k, n, N) < 10. \tag{21}$$

In a related direction, Saab, Chartrand and Yilmaz [23] observed that, for ℓ_q-regularization with $q \in (0, 1]$, larger support sizes provide improved constants amplifying the error in the noisy or compressible setting. Saab and Yilmaz [24] further discuss a sufficient condition for ℓ_q-regularization, $q \in (0, 1]$, which is weaker than Thm. 2 as the support size of the RIP constants increases.

4. Splitting Support Sets

In [5], Blumensath and Davies demonstrate how the introduction of an adaptive step-size into the Iterative Hard Thresholding algorithm leads to better guarantees of stability. A further qualitative contribution of the paper is to reduce the support size of the RIP constants in the convergence condition from $3k$ to $2k$. We might also then ask whether the new condition represents a quantitative improvement.

This reduction in support size is essentially achieved by a single step in the proof of [5, Thm. 4]. Given the previous iterate x^l, we seek to identify some constant η such that

$$\|A_S^* A_T (x - x^l)_T\|_2 \leq \eta \|(x - x^l)_T\|_2,$$

where S and T are disjoint subsets of maximum cardinality $2k$ and k respectively. It is a straightforward consequence of the RIP that η can be taken to be $R(3k, n, N)$, and this choice of η is used by Blumensath and Davies in [4, Lemma 2] to derive for IHT the convergence condition $R(3k, n, N) < 1/\sqrt{8}$. However, the authors observe in [5] that one may split the set S into two disjoint subsets each of size k, and subsequently apply the triangle inequality. In the symmetric setting this leads to the alternative choice of $\eta = \sqrt{2}R(2k, n, N)$. Though the method of proof in [5] varies significantly from that in [4] in other ways, such as the switch to an adaptive step-size and a generalization to the asymmetric setting, we can nonetheless examine the effect of the support set splitting alone. In this case, it is easy to adapt the proof of [4, Corollary 4] to show that the alternative convergence condition is $R(2k, n, N) < 1/4$. We have thus reduced the RIP support from $3k$ to $2k$ at the expense of decreasing the bound on $R(2k, n, N)$ by $1/\sqrt{2}$.

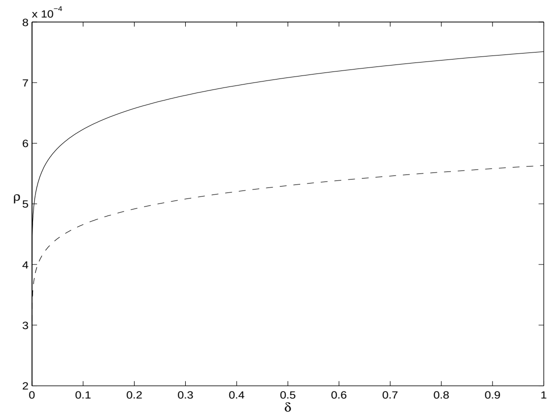


Figure 3: Lower bounds on the exact k-sparse recovery phase transition for Gaussian random matrices via IHT: $\mathcal{R}(\delta, 3\rho) < 1/\sqrt{8}$ (solid), $\mathcal{R}(\delta, 2\rho) < 1/4$ (dash-dash).

We may compare the two conditions for Gaussian random matrices by means of the bounds defined in Thm. 1. The resulting lower bounds on the exact k-sparse recovery phase transition are displayed in Fig. 3. For Gaussian matrices, the convergence condition derived by support set splitting is in fact stronger, showing that there is no quantitative advantage gained by splitting the support set in this manner.

In order to see why this is likely to be the case, let us define $\eta(s, t, n, N)$ to be the smallest η such that

$$\|A_S^* A_T v_T\|_2 \leq \eta \|v_T\|_2 \tag{22}$$

holds for all disjoint sets S and T with cardinality s and t, respectively. The desire to state conditions in terms of RIP constants then typically leads to the use of the bound

$$\eta(s, t, n, N) \leq R(s + t, n, N), \tag{23}$$

even though this bound is not necessarily sharp.

Now consider the situation for IHT where the set S is of cardinality ms so that we may split S into m equally-sized, disjoint subsets S_i such that $S = \bigcup_{i=1}^m S_i$. Then, since the S_i are disjoint, we have

$$\|A_S^* A_T v_T\|_2^2 = \sum_{i=1}^m \left\|A_{S_i}^* A_T v_T\right\|_2^2 \leq m \cdot \left(\eta(s, t, n, N)\|v_T\|_2\right)^2. \tag{24}$$

By (22), $\eta(ms, t, n, N)$ is the smallest number satisfying $\|A_S^* A_T v_T\|_2 \leq \eta\|v_T\|_2$ since $|S| = ms$ and $|T| = t$. Therefore, $\eta(ms, t, n, N) \leq \sqrt{m} \cdot \eta(s, t, n, N)$. In other words, splitting the support so as to replace $\eta(ms, t, n, N)$ with the upper bound $\sqrt{m} \cdot \eta(s, t, n, N)$ can only make the condition stronger, and certainly never give a quantitative improvement. It comes as no surprise, then, when we employ the bound (23) and compare the RIP conditions by means of Gaussian RIP bounds, that we obtain a lower phase transition. To summarize, while this technique achieved Blumensath and Davies's desired goal of reducing the support sizes, it could never be expected to offer a real quantitative advantage.

5. Conclusion

Although a strict condition, the RIP is a versatile tool for analyzing sparse approximation algorithms. The desire for qualitative comparison of various results obtained from an RIP analysis motivates a reduction in the support size of the RIP constants as evidenced by the literature. However, following the same methods of proof and quantitatively comparing the resulting sufficient conditions within the phase transition framework

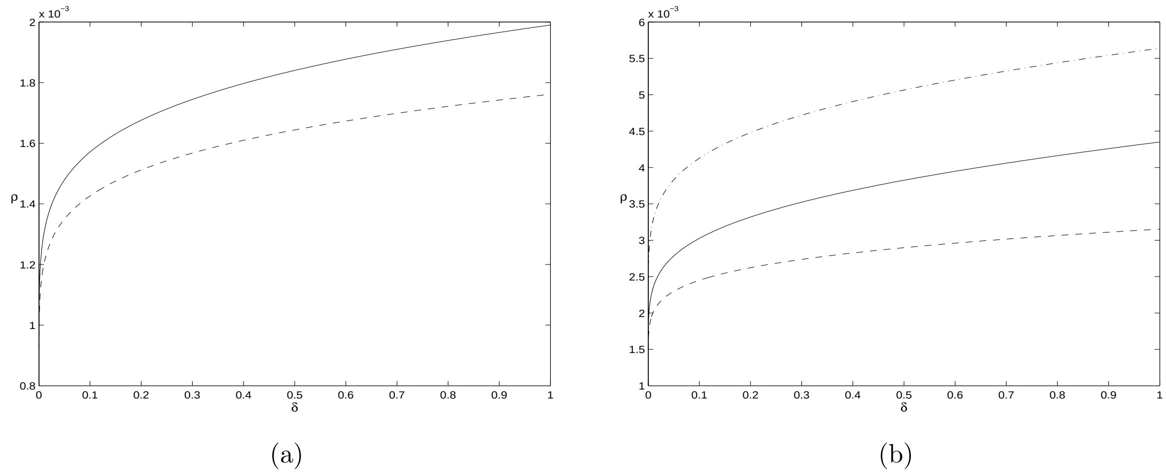


Figure 4: Lower bounds on the exact k-sparse recovery phase transition for Gaussian random matrices: (a) $\mathcal{R}(\delta, 2\rho) < .4531$ (solid), $\mathcal{R}(\delta, \rho) < .307$ (dash-dash). (b) $\rho_S^{cwx}(\delta)$ (dash-dot); $\rho_S^{bt}(\delta)$ (solid); $\rho_S^{fl}(\delta)$ (dash-dash).

shows that the smallest support size is not necessarily better than larger support sizes. This is certainly dependent on the method of proof, and it is plausible that improved proof techniques using the RIP may lead to weaker sufficient conditions with reduced support sizes in the future. Similarly, statements involving RIP constants are crucially dependent on the matrix ensemble chosen. Meanwhile, given that a quantitative method of comparison exists, namely the phase transition framework advocated by Donoho et al., such support size reductions in RIP constants can and should be examined for efficacy.

Epilogue

During the review process of this letter, the authors became aware of work by Cai et al. [6, 7] on sufficient RIP conditions for ℓ_1-regularization to exactly recover every $x \in \chi^N(k)$. Cai et al. made improvements on the upper bound for $R(2k, n, N)$ and announced what appears to be the first result involving a restricted isometry constant with support size k, namely $R(k, n, N) < .307$. As discussed in this letter, decreasing the support size of the RIP constants while decreasing the upper bound is not necessarily an improvement. For matrices from the Gaussian ensemble, $R(k, n, N) < .307$ is stronger than even the symmetric version of Thm. 2, $R(2k, n, N) < .4531$. See Fig. 4(a).

This reduction in support size was obtained by extending the work of Candès et al. [10, 11] and Foucart and Lai [21]. Two inequalities, the *Shifting Inequality* [7] and the *Square Root Lifting Inequality*[5] [6] were applied and a modified partitioning of the index set was performed to invoke these inequalities. The authors argue that their ability to decrease the support size of the RIP constants has lead to weaker conditions. The reduced support sizes highlighted in [6, 7] do not provide the weakest conditions from their analysis.

The techniques used by Cai et al. in both [6, 7] are quite valuable when applied to an ARIP analysis and without concern for the support size. Such an analysis leads to a family of sufficient conditions which further increase the lower bound on the phase transition for Gaussian matrices. For example, following [7] with an ARIP analysis produces the sufficient condition

$$\mu^{cwx}(k, n, N) := L(2k, n, N) + \frac{L(6k, n, N) + U(6k, n, N)}{4} < 1.$$

[5]The Square Root Lifting Inequality is essentially the same as the support set splitting of Blumenseth and Davies described in Sec. 4; as such it poses the same quantitative disadvantages.

In Fig. 4(b), the phase transition curve defined by $\rho_S^{cwx}(\delta)$ where $\mu^{cwx}\left(\delta, \rho_S^{cwx}(\delta)\right) \equiv 1$ is shown with the two curves from Fig. 2.

Acknowledgments

The authors would like to thank Edinburgh Compressed Sensing (E-CoS) and the referee for helpful suggestions which improved this letter.

References

[1] Jeffrey D. Blanchard, Coralia Cartis, and Jared Tanner. Compressed sensing: How sharp is the restricted isometry property? submitted, 2009.

[2] Jeffrey D. Blanchard, Coralia Cartis, and Jared Tanner. Phase transitions for restricted isometry properties. In *Proceedings of Signal Processing with Adaptive Sparse Structured Representations*, 2009.

[3] Jeffrey D. Blanchard, Coralia Cartis, Jared Tanner, and Andrew Thompson. Phase transitions for greedy sparse approximation algorithms. submitted, 2009.

[4] Thomas Blumensath and Mike E. Davies. Iterative hard thresholding for compressed sensing. *Appl. Comput. Harmon. Anal.*, 27(3):265–274, 2009.

[5] Thomas Blumensath and Mike E. Davies. Normalised iterative hard thresholding; guaranteed stability and performance. *IEEE J. Sel. Topics Sig. Proc.*, 2009. in press.

[6] T. Tony Cai, Lie Wang, and Guangwu Xu. New bounds for restricted isometry constants. submitted, online `arXiv:0911.1564v1`, 2009.

[7] T. Tony Cai, Lie Wang, and Guangwu Xu. Shifting inequality and recovery of sparse signals. *IEEE Trans. Sig. Proc.*, 2009. in press, `doi:10.1109/TSP.2009.2034936`.

[8] Emmanuel J. Candès. Compressive sampling. In *International Congress of Mathematicians. Vol. III*, pages 1433–1452. Eur. Math. Soc., Zürich, 2006.

[9] Emmanuel J. Candès. The restricted isometry property and its implications for compressed sensing. *C. R. Math. Acad. Sci. Paris*, 346(9-10):589–592, 2008.

[10] Emmanuel J. Candès, Justin K. Romberg, and Terence Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.*, 59(8):1207–1223, 2006.

[11] Emmanuel J. Candes and Terence Tao. Decoding by linear programming. *IEEE Trans. Inform. Theory*, 51(12):4203–4215, 2005.

[12] Emmanuel J. Candes and Terence Tao. Near-optimal signal recovery from random projections: universal encoding strategies? *IEEE Trans. Inform. Theory*, 52(12):5406–5425, 2006.

[13] Rick Chartrand. Nonconvex compressed sensing and error correction. In *32nd International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2007.

[14] David L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, 2006.

[15] David 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.

[16] David L. Donoho. High-dimensional centrally symmetric polytopes with neighborliness proportional to dimension. *Discrete Comput. Geom.*, 35(4):617–652, 2006.

[17] David L. Donoho and Victoria 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.

[18] David L. Donoho and Jared Tanner. Sparse nonnegative solutions of underdetermined linear equations by linear programming. *Proc. Natl. Acad. Sci. USA*, 102(27):9446–9451, 2005.

[19] David L. Donoho and Yaakov Tsaig. Fast solution of l1 minimization problems when the solution may be sparse. *IEEE Trans. Inform. Theory*, 54(11):4789–4812, 2008.

[20] Simon Foucart. A note on guaranteed sparse recovery via ℓ_1-minimization. *Appl. Comput. Harmon. Anal.*, 2010. in press, `doi:10.1016/j.acha.2009.10.004`.

[21] Simon Foucart and Ming-Jun 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.

[22] Deanna Needell and Joel Tropp. Cosamp: Iterative signal recovery from incomplete and inaccurate samples. *Appl. Comp. Harm. Anal.*, 26(3):301–321, 2009.

[23] Ryan Saab, Rick Chartrand, and Ozgur Yilmaz. Stable sparse approximations via nonconvex optimization. In *33rd International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2008.

[24] Ryan Saab and Ozgur Yilmaz. Sparse recovery by non-convex optimization - instance optimality. *Appl. Comput. Harmon. Anal.*, 2009. in press, `doi:10.1016/j.acha.2009.08.002`.