On Support Sizes of Restricted Isometry Constants

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

[a]*Department of Mathematics and Statistics, Grinnell College, Grinnell, IA 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. For qualitative comparison of sufficient conditions derived from an RIP analysis, the support size of the RIP constants is generally reduced as much as possible with the goal of achieving a support size of twice the sparsity of the target signal. Using a quantitative comparison via phase transitions for Gaussian measurement matrices, three examples from the literature of such support size reduction are investigated. In each case, utilizing a larger support size for the RIP constants results in a sufficient condition for exact sparse recovery satisfied, with high probability, 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, [6, 9, 13], 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 [9], namely the *restricted isometry property* (RIP). Recently, the RIP has been generalized to account for the asymmetry about the unit of the singular values of submatrices of the measurement matrix A, [1, 19]. In this article, we focus on Gaussian matrices, i.e. matrices whose entries are selected i.i.d. from the normal distribution $\mathcal{N}(0, 1/n)$, and denote the set of all k-sparse signals by $\chi^N(k) = \{x \in \mathbb{R}^N : \|x\|_0 \leq k\}$, where $\|x\|_0$ counts the number of nonzero entries of x.

Definition 1 (RIP [9] 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:

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

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

The standard (symmetric) *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)

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 $L(2k, n, N) < 1$. Second, the early sufficient conditions [8, 9] for successful k-sparse recovery via ℓ_1-regularization involved various support sizes. The results were eventually superseded by Candès's elegant sufficient RIP condition with support size of $2k$, i.e. $R(2k, n, N) < \sqrt{2} - 1$, [7]. 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 $2k$.

In this article, we employ the phase transition framework advocated by Donoho et. al. [12, 15, 16, 17, 18] 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 stricter sufficient conditions.

1.1. Organization and Notation

In the following, we present three instances in the literature where reducing the support sizes of the RIP constants results in a more stringent sufficient condition for sparse signal recovery. By using the quantitative comparisons available through the phase transition framework, outlined in Section 1.3, 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) [20], 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. (Section 1.2)

(ii) The current, generally accepted state of the art sufficient condition for ℓ_1-*regularization* obtained by Foucart and Lai [19] 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. (Section 2)

(iii) A clever 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. We apply the support set splitting to CoSaMP demonstrating that this method of support size reduction generally fails to provide a quantitative advantage. (Section 3)

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 $mk \leq n < N$.

1.2. 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 [20]. In this work, Needell and Tropp provide a sufficient

---

[2]An advantage of the asymmetric formulation of the RIP is that $L(2k, n, N) < 1$ is truly a necessary condition as opposed to the often stated, but not necessary requirement $R(2k, n, N) < 1$.
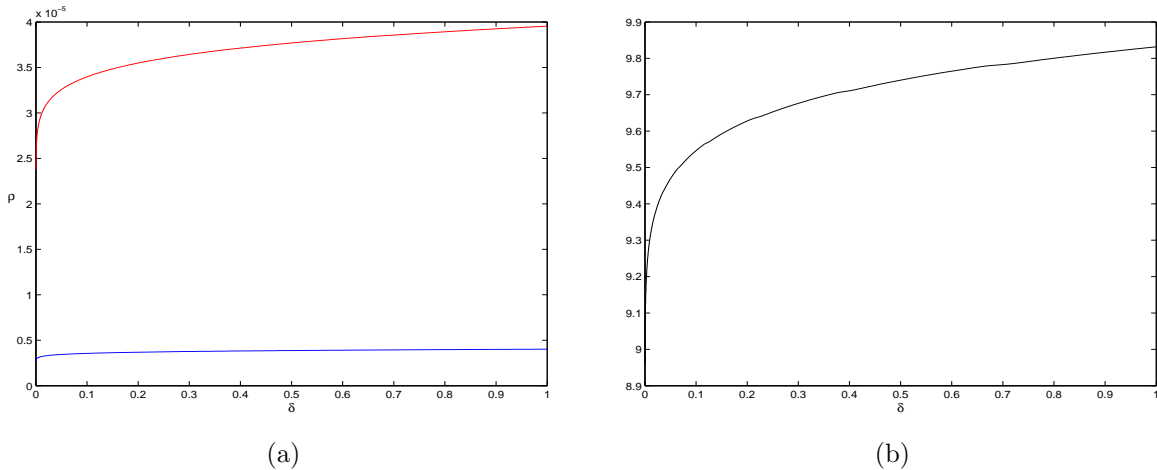
Figure 1: CoSaMP: (a) Lower bounds on the exact k-sparse recovery phase transition for Gaussian random matrices; $R(\delta, 4\rho) < 0.1$ (red), $R(\delta, 2\rho) < 0.025$ (blue). (b) The ratio of the two phase transition curves in (a).

condition for guaranteed recovery of k-sparse vectors, namely $R(4k, n, N) < 0.1$. The authors also provide a very nice bound on the growth rate of RIP constants, $R(ck, n, N) \leq c \cdot R(2k, n, N)$, [20, Cor. 3.4]. For qualitative comparison to current results for ℓ_1-regularization, such as Candès's $R(2k, n, N) < \sqrt{2} - 1$, [7], Needell and Tropp apply their bound on the growth of RIP constants to obtain the condition $R(2k, n, N) < 0.025$, [20, Rmk. 2.2], which is therefore also sufficient for the exact recovery of k-sparse signals. 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 $4k$ is considerably weaker than the condition with support size $2k$ when applied to Gaussian random matrices. By employing a bound (described in Section 1.3) $R(\delta, \rho)$ 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 Figure 1(a), lower bounds on the phase transition are displayed for the two conditions. With exponentially high 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 $R(4k, n, N) < 0.1$ with high probability has greater than 9.7 times the area of the phase space region determined by $R(2k, n, N) < 0.025$. Moreover, Figure 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.

1.3. 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, 9, 10, 14]. The current best known bounds for Gaussian matrices were derived in [1] and are denoted $L(\delta, \rho), U(\delta, \rho), R(\delta, \rho)$ where $\delta = \frac{n}{N}, \rho = \frac{k}{n}$. The following is an adaptation of [1, Thm. 1].

Theorem 1 (Blanchard, Cartis, Tanner [1]). *Let A be a matrix of size $n \times N$ whose entries are drawn i.i.d. from $\mathcal{N}(0, 1/n)$ and let $n \to \infty$ with $\frac{k}{n} \to \rho$ and $\frac{n}{N} \to \delta$. Let $L(\delta, \rho)$ and $U(\delta, \rho)$ be defined as in [1, Thm.*

3

1]. Define $R(\delta, \rho) = \max\{L(\delta, \rho), U(\delta, \rho)\}$. Then for any $\epsilon > 0$, as $n \to \infty$,

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

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

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

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 for 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 exponentially high 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 k-sparse signals.

2. 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 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,

$$\min \|x\|_1 \text{ subject to } y = Ax, \tag{7}$$

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

Currently, it is generally accepted that the state of the art sufficient condition obtained by RIP analysis for ℓ_1-regularization was proven by Foucart and Lai [19].

Theorem 2 (Foucart, Lai [19])**.** *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*

$$\mu^{fl}(k, n, N) := \frac{1 + \sqrt{2}}{4}\left(\frac{1 + U(2k, n, N)}{1 - L(2k, n, N)} - 1\right), \tag{8}$$

then ℓ_1-regularization (7) exactly recovers 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 (7), namely $3R(4k, n, N) + R(3k, n, N) < 2$, was obtained by Candès, Romberg, and Tao [8] from an RIP analysis. Chartrand [11] 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 [21, 22]. Here, the result is stated with an asymmetric RIP analysis following the proof in [11] which, in turn, is an adaptation of Candès, Romberg, and Tao's proof [8].

Theorem 3. *Suppose $k \in \mathbb{N}^+$ and $b > 2$ with $bk \in \mathbb{N}^+$. Let A be a matrix 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*

$$bL([b+1]k, n, N) + U(bk, n, N) < b - 1, \tag{9}$$

then ℓ_1-regularization (7) exactly recovers every $x \in \chi^N(k)$.

Proof. Let $x \in \chi^N(k)$ and $y = Ax$. Suppose z is a solution to (7). 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 $|T_J| \leq bk$. Let $T_{01} = T_0 \cup T_1$. Counting arguments and norm comparisons taken directly from [11] provide the following inequalities,

$$\|A_{T_{01}} h_{T_{01}}\|_2 \leq \sum_{j=2}^{J} \|A_{T_j} h_{T_j}\|_2, \tag{10}$$

$$\|h_{T_0^c}\|_1 \leq \|h_{T_0}\|_1, \tag{11}$$

$$\sum_{j=2}^{J} \|h_{T_j}\|_2 \leq \frac{1}{\sqrt{bk}} \|h_{T_0^c}\|_1. \tag{12}$$

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

$$\|A_{T_{01}} h_{T_{01}}\|_2 \geq \sqrt{1 - L([b+1]k, n, N)} \, \|h_{T_{01}}\|_2 \tag{13}$$

$$\text{and} \quad \|A_{T_j} h_{T_j}\|_2 \leq \sqrt{1 + U(bk, n, N)} \, \|h_{T_j}\|_2. \tag{14}$$

Therefore, inserting (13) and (14) into (10) yields

$$\|h_{T_{01}}\|_2 \leq \sqrt{\frac{1 + U(bk, n, N)}{1 - L([b+1]k, n, N)}} \sum_{j=2}^{J} \|h_{T_j}\|_2. \tag{15}$$

Combining (11), (12), (15), and the standard relationship between norms, we then have

$$\begin{aligned}
\|h_{T_{01}}\|_2 &\leq \sqrt{\frac{1 + U(bk, n, N)}{1 - L([b+1]k, n, N)}} \frac{1}{\sqrt{bk}} \|h_{T_0}\|_1 \\
&\leq \sqrt{\frac{1 + U(bk, n, N)}{1 - L([b+1]k, n, N)}} \sqrt{\frac{k}{bk}} \|h_{T_0}\|_2 \\
&\leq \sqrt{\frac{1 + U(bk, n, N)}{b - bL([b+1]k, n, N)}} \|h_{T_{01}}\|_2.
\end{aligned} \tag{16}$$

Squaring and rearranging (16),

$$\big([b-1] - [bL([b+1]k, n, N) + U(bk, n, N)] \big) \|h_{T_{01}}\|_2^2 \leq 0. \tag{17}$$

The hypothesis (9) ensures the left hand side of (17) is nonnegative and zero only when $\|h_{T_{01}}\|_2 = 0$. Therefore, $h = 0$ implying $z = x$. $\qquad\square$

Following the framework described in Section 1.3 and more formally developed in [1], we restate Theorems 2 and 3 in the language of phase transitions for Gaussian matrices. First, we state an equivalent
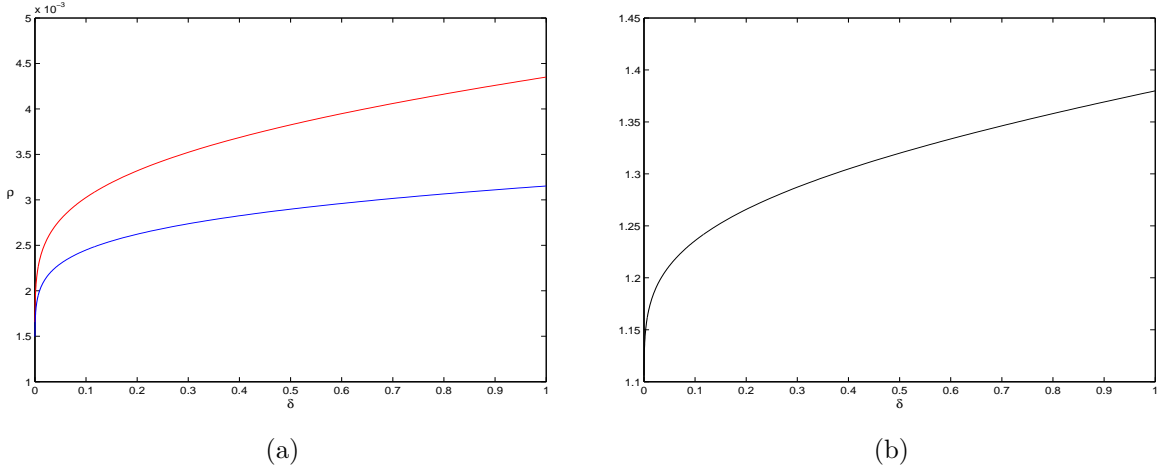
Figure 2: ℓ_1-regularization: (a) lower bounds on the exact k-sparse recovery phase transition for Gaussian random matrices; $\rho_S^{bt}(\delta; 11)$ (red) and $\rho_S^{fl}(\delta)$ (blue). (b) The improvement ratio $\frac{\rho_S^{bt}(\delta;11)}{\rho_S^{fl}(\delta)}$.

condition to (9), namely $\mu^{bt}(k, n, N; b) < 1$ where

$$\mu^{bt}(k, n, N; b) := \frac{bL([b+1]k, n, N) + U(bk, n, N)}{b - 1}. \tag{18}$$

Clearly, (18) defines a family of functions indexed by $b > 2$ with $bk \in \mathbb{N}^+$. Applying the bounds defined in Theorem 1 to (8) and (18), we obtain the functions

$$\mu^{fl}(\delta, \rho) := \frac{1 + \sqrt{2}}{4} \left(\frac{1 + U(\delta, 2\rho)}{1 - L(\delta, 2\rho)} - 1 \right) \quad \text{and} \quad \mu^{bt}(\delta, \rho; b) := \frac{bL(\delta, [b+1]\rho) + U(\delta, b\rho)}{b - 1}. \tag{19}$$

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 $\mu^{bt}(\delta, \rho; b) = 1$.

Theorem 4. *For any $\epsilon > 0$, with $\frac{n}{N} \to \delta$ and $\frac{k}{n} \to \rho$ as $n \to \infty$, 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 exactly recovered by ℓ_1-regularization (7), provided one of following conditions is satisfied:*

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

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

$\rho_S^{fl}(\delta)$ is displayed as the blue curve in Figure 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 red curve in Figure 2(a). 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 ranging from 1.11 to 1.38. By using a quantitative comparison, we see that the weakest RIP condition is not Theorem 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

$$11L(12k, n, N) + U(11k, n, N) < 10. \tag{20}$$

In a related direction, Saab, Chartrand, and Yilmaz [21] 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

6

setting. Saab and Yilmaz [22] further discuss a sufficient condition for ℓ_q-regularization, $q \in (0, 1]$, which is weaker than Theorem 2 as the support size of the RIP constants increases.

3. 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 $2k$. This support size reduction 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 $R(2k, n, N) < 1/4$. Thus the RIP support size is reduced from $3k$ to $2k$ at the expense of a factor of $\sqrt{2}$.

We compare the two conditions for Gaussian random matrices by means of the bounds defined in Theorem 1. The resulting lower bounds on the exact k-sparse recovery phase transition are displayed in Figure 3(a). 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 apply the support set splitting principle in the context of another algorithm and with an ARIP analysis, we can generalize the result as the following lemma.

Lemma 5. *Let S and T be disjoint sets of cardinality $m \cdot s$ and t respectively, where m, s and t are positive integers, and suppose that A has ARIP constants $L(s + t, n, N)$ and $U(s + t, n, N)$. Then, for any vector v,*

$$\|A_S^* A_T v_T\|_2 \leq \frac{\sqrt{m}}{2} \left\{ L(s + t, n, N) + U(s + t, n, N) \right\} \|v_T\|_2.$$

Proof. Split S into m equally-sized disjoint subsets S_i such that $S = \bigcup_{i=1}^m S_i$. Then we have, by the orthogonality of the S_i, and by [4, Lemma 2],

$$
\begin{aligned}
\|A_S^* A_T v_T\|_2^2 &= \|(\textstyle\sum_{i=1}^m A_{S_i})^* A_T v_T\|_2^2 \\
&= \textstyle\sum_{i=1}^m \|A_{S_i}^* A_T v_T\|_2^2 \\
&\leq \textstyle\sum_{i=1}^m \frac{1}{4} \left\{ L(s + t, n, N) + U(s + t, n, N) \right\}^2 \|v_T\|_2^2 \\
&= \frac{m}{4} \left\{ L(s + t, n, N) + U(s + t, n, N) \right\}^2 \|v_T\|_2^2,
\end{aligned}
$$

from which the result now follows by taking square roots. $\qquad\square$

As mentioned in Section 1.2, CoSaMP [20] is another greedy algorithm with a convergence guarantee involving restricted isometry constants. Largely following the method of proof in [20], an asymmetric condition guaranteeing convergence of CoSaMP was shown in [3] to be $\mu^{csp}(k, n, N) < 1$ for

$$\mu^{csp}(k, n, N) := \frac{1}{2} \left(2 + \frac{L(4k, n, N) + U(4k, n, N)}{1 - L(3k, n, N)} \right) \left(\frac{L(2k, n, N) + U(2k, n, N) + L(4k, n, N) + U(4k, n, N)}{1 - L(2k, n, N)} \right).$$
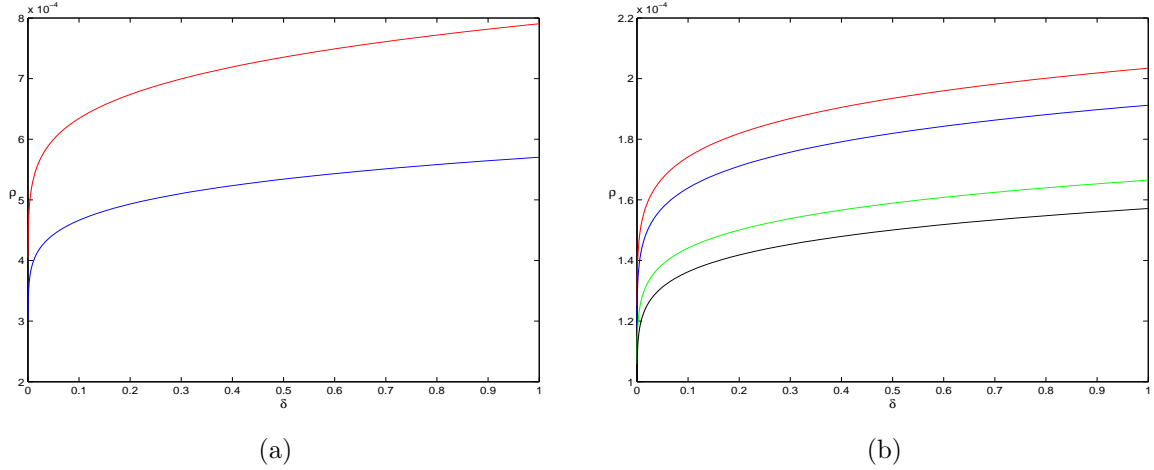
(21)

Figure 3: Lower bounds on the exact k-sparse recovery phase transition for Gaussian random matrices: (a) IHT: $R(\delta, 3\rho) < 1/\sqrt{8}$ (red), $R(\delta, 2\rho) < 1/4$ (blue). (b) CoSaMP: various cases with or without support set splitting, $\rho_S^{csp,1}(\delta)$ (red); $\rho_S^{csp,2}(\delta)$ (blue); $\rho_S^{csp,3}(\delta)$ (green); $\rho_S^{csp,4}(\delta)$ (black).

We apply Lemma 5 to the factors of $\mu^{csp}(k, n, N)$ involving $\{L(4k, n, N) + U(4k, n, N)\}$. To simplify the following discussion, we define $\phi(\cdot, n, N) = L(\cdot, n, N) + U(\cdot, n, N)$. A consideration of the argument in the original paper [20] shows that the first factor results from taking $|S| = 2k$ and $|T| = 2k$. We may therefore split S into two disjoint subsets of size k. By applying Lemma 5 with $m = 2$, we see that $\phi(4k, n, N)$ may be replaced by $\sqrt{2}\phi(3k, n, N)$. In the second case, we have $|S| = 3k$ and $|T| = k$. We therefore split this set S into three disjoint subsets of size k. By applying Lemma 5 with $m = 3$, we see that $\phi(4k, n, N)$ may be replaced by $\sqrt{3}\phi(2k, n, N)$. We consider now four alternative expressions for $\mu^{csp}(k, n, N)$: $\mu_1^{csp}(k, n, N)$ is the original expression (21), $\mu_2^{csp}(k, n, N)$ is obtained by substituting $\sqrt{2}\phi(3k, n, N)$ for $\phi(4k, n, N)$ in the first factor of (21), $\mu_3^{csp}(k, n, N)$ is obtained by substituting $\sqrt{3}\phi(2k, n, N)$ for $\phi(4k, n, N)$ in the second factor of (21), and $\mu_4^{csp}(k, n, N)$ is obtained by making both substitutions.

Applying the bounds from Theorem 1, we obtain asymptotic bounds on $\mu_i^{csp}(k, n, N)$ for $i = 1, \ldots, 4$:

$$\mu_1^{csp}(\delta, \rho) := \frac{1}{2}\left(2 + \frac{L(\delta, 4\rho) + U(\delta, 4\rho)}{1 - L(\delta, 3\rho)}\right)\left(\frac{L(\delta, 2\rho) + U(\delta, 2\rho) + L(\delta, 4\rho) + U(\delta, 4\rho)}{1 - L(\delta, 2\rho)}\right),$$

$$\mu_2^{csp}(\delta, \rho) := \frac{1}{2}\left(2 + \frac{\sqrt{2}(L(\delta, 3\rho) + U(\delta, 3\rho))}{1 - L(\delta, 3\rho)}\right)\left(\frac{L(\delta, 2\rho) + U(\delta, 2\rho) + L(\delta, 4\rho) + U(\delta, 4\rho)}{1 - L(\delta, 2\rho)}\right),$$

$$\mu_3^{csp}(\delta, \rho) := \frac{1}{2}\left(2 + \frac{L(\delta, 4\rho) + U(\delta, 4\rho)}{1 - L(\delta, 3\rho)}\right)\left(\frac{(1 + \sqrt{3})(L(\delta, 2\rho) + U(\delta, 2\rho))}{1 - L(\delta, 2\rho)}\right),$$

$$\mu_4^{csp}(\delta, \rho) := \frac{1}{2}\left(2 + \frac{\sqrt{2}(L(\delta, 3\rho) + U(\delta, 3\rho))}{1 - L(\delta, 3\rho)}\right)\left(\frac{(1 + \sqrt{3})(L(\delta, 2\rho) + U(\delta, 2\rho))}{1 - L(\delta, 2\rho)}\right).$$

Finally, define $\rho_S^{csp,i}(\delta)$ as the solution to $\mu_i^{csp}(\delta, \rho) = 1$ for $i = 1, \ldots, 4$.

Lower bounds on the exact k-sparse recovery phase transition for Gaussian random matrices are displayed in Figure 3(b): $\rho_S^{csp,1}(\delta)$ - red; $\rho_S^{csp,2}(\delta)$ - blue; $\rho_S^{csp,3}(\delta)$ - green; $\rho_S^{csp,4}(\delta)$ - black. In all three cases, splitting the support set results in a lower phase transition with the lowest phase transition occurring when both sets are split. In summary, an examination of the lower bounds on the phase transitions for Gaussian matrices derived from two well-known greedy algorithms shows that in each case support set splitting yields a stricter sufficient condition for exact k-sparse recovery.

4. Conclusion

The RIP is a versatile tool, although a strict condition, 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 (possibly with an ARIP analysis) and quantitatively comparing the resulting sufficient conditions with the phase transition framework shows that the smallest support size is not necessarily ideal. This is certainly dependent on the method of proof as the reduction in support size to $2k$ by Candès [7] clearly gave a weaker sufficient condition than alternative sufficient conditions known at the time. It is therefore plausible that improved proof techniques using the RIP can lead to weaker sufficient conditions in the future. 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.

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.*, 2009. in press, `doi:10.1016/j.acha.2009.04.002`.

[5] Thomas Blumensath and Mike E. Davies. Normalised iterative hard thresholding; guaranteed stability and performance. 2009. submitted.

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

[7] 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.

[8] 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.

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

[10] 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.

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

[12] David L. Donoho. Neighborly polytopes and sparse solution of underdetermined linear equations. Technical Report, Department of Statistics, Stanford University, 2004.

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

[14] 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.

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

[16] 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.

[17] 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.

[18] David L. Donoho and Yaakov Tsaig. Fast solution of l1 minimization problems when the solution may be sparse. submitted, 2006.

[19] 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.

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

[21] 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.

[22] Ryan Saab and Ozgur Yilmaz. Sparse recovery by non-convex optimization – instance optimality. submitted, online `arXiv:0809.0745v1`, 2008.