Recovery Analysis for Weighted ell_1-Minimization Using the Null Space Property

Mansour, H.; Saab, R.

TR2016-024 March 2016

Abstract

We study the recovery of sparse signals from underdetermined linear measurements when a potentially erroneous support estimate is available. Our results are twofold. First, we derive necessary and sufficient conditions for signal recovery from compressively sampled measurements using weighted ‘1-norm minimization. These conditions, which depend on the choice of weights as well as the size and accuracy of the support estimate, are on the null space of the measurement matrix. They can guarantee recovery even when standard ‘1 minimization fails. Second, we derive bounds on the number of Gaussian measurements for these conditions to be satisfied, i.e., for weighted ‘1 minimization to successfully recover all sparse signals whose support has been estimated sufficiently accurately. Our bounds show that weighted ‘1 minimization requires significantly fewer measurements than standard ‘1 minimization when the support estimate is relatively accurate.

Applied and Computational Harmonic Analysis

Recovery Analysis for Weighted ℓ_1-Minimization Using the Null Space Property

Hassan Mansour, Rayan Saab

Abstract

We study the recovery of sparse signals from underdetermined linear measurements when a potentially erroneous support estimate is available. Our results are twofold. First, we derive necessary and sufficient conditions for signal recovery from compressively sampled measurements using weighted ℓ_1-norm minimization. These conditions, which depend on the choice of weights as well as the size and accuracy of the support estimate, are on the null space of the measurement matrix. They can guarantee recovery even when standard ℓ_1 minimization fails. Second, we derive bounds on the number of Gaussian measurements for these conditions to be satisfied, i.e., for weighted ℓ_1 minimization to successfully recover all sparse signals whose support has been estimated sufficiently accurately. Our bounds show that weighted ℓ_1 minimization requires significantly fewer measurements than standard ℓ_1 minimization when the support estimate is relatively accurate.

I. Introduction

The application of ℓ_1 norm minimization for the recovery of sparse signals from incomplete measurements has become standard practice since the advent of compressed sensing [1]–[3]. Consider an arbitrary k-sparse signal $x \in \mathbb{R}^N$ and its corresponding linear measurements $y \in \mathbb{R}^m$ with $m < N$, where y results from the underdetermined system

$$y = Ax. \tag{1}$$

It is possible to exactly recover all such sparse x from y by solving the ℓ_1 minimization problem

$$\min_z \ \|z\|_1 \text{ subject to } y = Az \tag{2}$$

if A satisfies certain conditions [1]–[3]. In particular, these conditions are satisfied with high probability by many classes of random matrices, including those whose entries are i.i.d. Gaussian random variables, when* $m \gtrsim k \log(N/k)$.

One property of the measurement matrix A that characterizes sparse recovery from compressive measurements is the null space property (NSP) (see, e.g., [4]) defined below.

Hassan Mansour is with Mitsubishi Electric Research Laboratories, Cambridge, MA 02139 (mansour@merl.com).
Rayan Saab is with the Department of Mathematics, The University of California, San Diego (rsaab@ucsd.edu).

*We write $x \gtrsim y$ when $x \geq Cy$ for some constant independent of x and y.

Definition 1. [4] A matrix $A \in \mathbb{R}^{m \times N}$ is said to have the null space property of order k and constant C if for any vector $h : Ah = 0$, and for every index set $T \subset \{1 \ldots N\}$ with $|T| \leq k$, we have

$$\|h_T\|_1 \leq C\|h_{T^c}\|_1.$$

In this case, we say that A satisfies NSP(k, C).

NSP(k, C) with $C < 1$ is a necessary and sufficient condition on the matrix A for the recovery of all k-sparse vectors from their measurements using (2). For example, it was shown in [5, Section 9.4] using an *escape through the mesh argument* [6] based on, e.g., [7] (cf. [8], [9]) that Gaussian random matrices satisfy the null space property with probability greater than $1 - \epsilon$ when $m > ck \ln eN/k$. Here, c depends on C and ϵ, but the dependence is mild enough that $c \approx 8$ is a reasonable approximation when N is large and k/N is small.

While the ℓ_1 minimization problem in (2) is suitable for recovering signals with arbitrary support sets, it is often the case in practice that signals exhibit structured support sets, or that an estimate of the support can be identified. In such cases, one is interested in modifying (2) to weaken the exact recovery conditions. In this paper, we analyze a recovery method that incorporates support information by replacing (2) with weighted ℓ_1 minimization. In particular, given a support estimate $\widetilde{T} \subset \{1, ..., N\}$, we solve the optimization problem

$$\min_z \sum_{i=1}^{N} w_i|z_i| \text{ subject to } y = Az, \text{ where } w_i = \begin{cases} \mathrm{w} \in [0, 1], & i \in \widetilde{T} \\ 1, & i \in \widetilde{T}^c \end{cases}. \tag{3}$$

The idea behind such a modification is to choose the weight vector such that the entries of x that are "expected" to be large, i.e., those on the support estimate $\widetilde{T}$, are penalized less.

A. Prior work

The recovery of compressively sampled signals using prior support information has been studied in several works, e.g., [10]–[19]. Vaswani and Lu [11]–[13] proposed a modified compressed sensing approach that incorporates known support elements using a weighted ℓ_1 minimization approach, with zero weights on the known support. Their work derives sufficient recovery conditions that are weaker than the analogous ℓ_1 minimization conditions of [2] in the case where a large proportion of the support is known. This work was extended by Jacques in [14] to include compressible signals and account for noisy measurements.

Friedlander et al. [16] studied the case where non-zero weights are applied to the support estimate, further generalizing and refining the results of Vaswani and Lu; they derive tighter sufficient recovery conditions that depend on the accuracy and size of the support estimate. Mansour et al. [17] then extended these results to incorporate multiple support estimates with varying accuracies.

Khajehnejad et al. [15] also derive sufficient recovery conditions for compressively sampled signals with prior support information using weighted ℓ_1 minimization. They partition $\{1, \ldots, N\}$ to two sets such that the entries of x supported on each set have a fixed probability of being non-zero, albeit the probabilities differ between the sets. Thus, in this work the prior information is the knowledge of the partition and probabilities. More recently, Oymak et al. [18] adopt the same prior information setup as [15] and derive lower bounds on the the minimum number of Gaussian measurements required for successful recovery when the optimal weights are chosen for each set. Their results are asymptotic in nature and pertain to the non-uniform model where one fixes a signal and draws the matrix at random. In this model, every new instance of the problem requires a new draw of the random measurement matrix. In addition to differing in our model for prior information, our results are uniform in nature, i.e., they pertain to the model where the matrix is drawn once and successful recovery is guaranteed (with high probability) for all sparse signals with sufficiently accurate support information.

Recently, Rauhut and Ward [20] analyzed the effectiveness of weighted ℓ_1 minimization for the interpolation of smooth signals that also admit a sparse representation in an appropriate transform domain. Using a weighted robust null space property, they derive error bounds associated with recovering functions with coefficient sequences in weighted ℓ_p spaces, $0 < p < 1$. This differs from our work which focuses on the effect of the support estimate accuracy on the recovery guarantees, and the relationship between the weighted null space property and the standard null space property (e.g., [4]).

B. *Notation and preliminaries*

Throughout the paper, we adopt the following notation. As stated earlier, $x \in \mathbb{R}^N$ is the k-sparse signal to be recovered and $y \in \mathbb{R}^m$ denotes the vector of measurements, i.e., $y = Ax$. Thus, k, N, and m denote the number of non-zeros entries of x, its ambient dimension, and the number of measurements, respectively. $T \subset \{1, ..., N\}$ is the support of x, and $\widetilde{T}$ is the support estimate used in (3). The cardinality of $\widetilde{T}$ is $|\widetilde{T}| = \rho k$ for some $\rho > 0$ and the accuracy of $\widetilde{T}$ is $\alpha = \frac{|\widetilde{T} \cap T|}{|\widetilde{T}|}$. For an index set $V \subset \{1, ..., N\}$ we define

$$\Gamma_s(V) := \left\{ U \subset \{1, ..., N\} : \Big| (V \cap U^c) \cup (V^c \cap U) \Big| \leq s \right\}.$$

We use the notation x_T to denote the restriction of the vector x to the set T. We introduce a *weighted nonuniform* null space property that, as we prove in Section II, provides a necessary and sufficient condition for the recovery of sparse vectors supported on a fixed set using weighted ℓ_1 minimization (with constant weights applied to a support estimate).

Definition 2. Let $T \subset \{1 \dots N\}$ with $|T| \leq k$ and $\widetilde{T} \in \Gamma_s(T)$. A matrix $A \in \mathbb{R}^{m \times N}$ is said to have the weighted nonuniform null space property with parameters T and $\widetilde{T}$, and constant C if for any vector $h : Ah = 0$, we have

$$\mathrm{w}\|h_T\|_1 + (1 - \mathrm{w})\|h_S\|_1 \leq C\|h_{T^c}\|_1,$$

where $S = (\widetilde{T} \cap T^c) \cup (\widetilde{T}^c \cap T)$. In this case, we say A satisfies w-NSP$(T, \widetilde{T}, C)$.

Next we define a *weighted uniform* null space property that leads to necessary and sufficient conditions for the recovery of all k-sparse vectors from compressive measurements using weighted ℓ_1 minimization.

Definition 3. A matrix $A \in \mathbb{R}^{m\times N}$ is said to have the weighted null space property with parameters k and s, and constant C if for any vector $h : Ah = 0$, and for every index set $T \subset \{1 \dots N\}$ with $|T| \le k$ and $S \subset \{1 \dots N\}$ with $|S| \le s$, we have

$$\mathrm{w}\|h_T\|_1 + (1-\mathrm{w})\|h_S\|_1 \le C\|h_{T^c}\|_1.$$

In this case, we say A satisfies w-NSP(k, s, C).

Thus, the standard null space property of order k, i.e., NSP(k, C), is equivalent to 1-NSP(k, k, C).

Remark 3.1. There should be no confusion between the notation used for the weighted non-uniform and uniform null space properties, as one pertains to subsets and the other to sizes of subsets.

C. Main contributions

Necessary and sufficient conditions: Our first main result is Theorem 4, identifying necessary and sufficient conditions for weighted ℓ_1 minimization to recover all k-sparse vectors when the error in the support estimate is of size s or less.

Theorem 4. *Given a matrix $A \in \mathbb{R}^{m\times N}$, every k-sparse vector $x \in \mathbb{R}^N$ is the unique solution to all optimization problems* (3) *with $\widetilde{T} \in \Gamma_s\Big(supp(x)\Big)$ if and only if A satisfies* w-*NSP(k, s, C) for some $C < 1$.*

Essentially, the theorem relates the the size of the error in the support estimate $\widetilde{T}$ to the recoverability via weighted ℓ_1 minimization. In particular, the error in the support estimate includes both entries in $\widetilde{T}$ that are not in the true support T, as well as missed entries from T. We prove this theorem in Section II. There, we also compare ℓ_1 minimization to weighted ℓ_1 minimization. For example, we show that if the accuracy of the support estimate $\widetilde{T}$ is at least 50%, then weighted ℓ_1 minimization recovers x if ℓ_1 minimization recovers x (Corollary 12). Moreover, when the support accuracy exceeds 50% and the weights are sufficiently small, weighted ℓ_1 minimization can successfully recover x even when standard ℓ_1 minimization fails (Corollary 9).

Weaker conditions on the number of measurements: Our second main result deals with matrices $A \in \mathbb{R}^{m\times N}$ whose entries are i.i.d. Gaussian random variables. We establish a condition on the number of measurements, m, that yields the weighted null space property, and hence guarantees exact sparse recovery using weighted ℓ_1 minimization.

Theorem 5. *Let T and $\widetilde{T}$ be two subsets of $\{1, ..., N\}$ with $|T| \le k$ and $|(T \cap \widetilde{T}^c) \cup (T^c \cap \widetilde{T})| \le s$ and let A be a random matrix with independent zero-mean unit-variance Gaussian entries. Then A satisfies* w-*NSP$(T, \widetilde{T}, C)$ with probability exceeding $1 - \epsilon$ provided*

$$\frac{m}{\sqrt{m+1}} \ge \sqrt{k+s} + C^{-1}\sqrt{2(\mathrm{w}^2 k + s)\ln(eN/k)} + \left(\frac{1}{2\pi e^3}\right)^{1/4}\sqrt{\frac{k}{\ln(eN/k)}} + \sqrt{2\ln\epsilon^{-1}}.$$

We observe that in the limiting case of large m, N, k, and taking $\mathrm{w} = 0$, the condition in Theorem 5 simplifies to

$$m \gtrsim k + s\ln(eN/k),$$

which can be significantly smaller than the analogous condition $m \gtrsim k\ln(eN/k)$ of standard ℓ_1 minimization [5] especially when the support estimate is accurate, i.e., when s is small. In Section III, we prove a more general version of Theorem 5, namely Theorem 13. Theorem 13 suggests that the choice $\mathrm{w} = 1-\alpha$ gives the weakest condition on the number of measurements m. On the other hand, Proposition 10 shows that when $\alpha > 0.5$, and the weighted null space property holds for a weight w, it also holds for weights $\mathrm{v} < \mathrm{w}$. Taken together, these results indicate that while the bound on the number of measurements m is minimized for $\mathrm{w} = 1-\alpha$, recovery by weighted ℓ_1 minimization is also guaranteed for all weights $\mathrm{v} \in [0, 1-\alpha]$ (when $\alpha > 0.5$). We note that Theorem 13 also indicates that when $\alpha > 0.5$, then using any weight $\mathrm{w} < 1$ results in a weaker condition on the number of measurements m than standard ℓ_1 minimization. In Section III, we also develop the corresponding bounds that guarantee uniform recovery for arbitrary sets T and $\widetilde{T} \subset \{1 \dots N\}$. Finally, we present numerical simulations in Section IV that illustrate our theoretical results.

II. Weighted null space property

In what follows, we describe the relationship between the weighted and standard null space properties and their associated optimization problems. Specifically, Proposition 6 establishes $\mathrm{w}\text{-NSP}(k,s,C)$ with some $C < 1$ as a necessary condition for weighted ℓ_1-minimization to recover all k-sparse vectors x from their measurements Ax, given a support estimate with at most s errors. Proposition 7 establishes that the same weighted null space property is also sufficient. Together, Propositions 6 and 7 yield Theorem 4.

Proposition 8 relates the weighted null space property to the standard null space properties of size s, $k-s$, and k. As a consequence, Corollary 9 shows that weighted ℓ_1 minimization can succeed when ℓ_1 minimization fails provided the support estimate is accurate enough and the weights are small enough. Proposition 10 shows that if $s < k$, i.e. the support estimate is at least 50% accurate, then any matrix that satisfies $\mathrm{w}\text{-NSP}(k,s,C)$ also satisfies $\mathrm{v}\text{-NSP}(k,s,C)$ for all $\mathrm{v} < \mathrm{w}$. Corollary 11 shows that the standard null space property guarantees that weighted ℓ_1 minimization succeeds when the support estimate is at least 50% accurate, regardless of $\mathrm{w} \in (0,1]$. Corollary 12 establishes the equivalence of $\mathrm{w}\text{-NSP}(s,s,C_s)$ and $\mathrm{NSP}(s,C_s)$. This shows that weighted ℓ_1 minimization succeeds in recovering all s-sparse signals from a support estimate that is 50% accurate if and only if ℓ_1 minimization recovers all s-sparse signals.

Proposition 6. *Let A be an $m \times n$ matrix that does not satisfy* w-*NSP*(k,s,C) *for any $C < 1$. Then, there exists a k-sparse vector x satisfying $Ax = b$ and a set $\widetilde{T}$ with $|(\widetilde{T} \cap T^c) \cup (\widetilde{T}^c \cap T)| \leq s$ such that x is not the unique minimizer of the optimization problem* (3).

Proof: Since A does not satisfy $\mathrm{w}\text{-NSP}(k,s,C)$ for any $C < 1$, there exists a vector $h : Ah = 0$ and index sets T with $|T| \leq k$ and S with $|S| \leq s$ such that $Ah_T = -Ah_{T^c}$ and

$$\mathrm{w}\|h_T\|_1 + (1-\mathrm{w})\|h_S\|_1 \geq \|h_{T^c}\|_1.$$

Define $\widetilde{T} := (T^c \cap S) \cup (T \cap S^c)$, so that $S = (T \cap \widetilde{T}^c) \cup (T^c \cap \widetilde{T})$. Substituting for S, splitting the set T, and simplifying we obtain

$$\|h_{T\cap\widetilde{T}^c}\|_1 + \mathrm{w}\|h_{T\cap\widetilde{T}}\|_1 \geq \mathrm{w}\|h_{T^c\cap\widetilde{T}}\|_1 + \|h_{T^c\cap\widetilde{T}^c}\|_1.$$

In other words, the weighted ℓ_1-norm of the vector h_T equals or exceeds that of h_{T^c}. So h_T is not the unique minimizer of (3). This establishes the necessity of the w-NSP condition. ■

Proposition 7. *Let A be an $m \times n$ matrix that satisfies* w*-NSP*(k, s, C) *for some* $C < 1$. *Then, every k-sparse vector x is the unique minimizer of the optimization problem* (3) *provided $\widetilde{T}$ satisfies* $|(\widetilde{T} \cap T^c) \cup (\widetilde{T}^c \cap T)| \leq s$.

Proof: Let x^* be a minimizer of (3) and define $h := x^* - x$. Then by the optimality of x^*

$$\mathrm{w}\|x_{\widetilde{T}} + h_{\widetilde{T}}\|_1 + \|x_{\widetilde{T}^c} + h_{\widetilde{T}^c}\|_1 \leq \mathrm{w}\|x_{\widetilde{T}}\|_1 + \|x_{\widetilde{T}^c}\|_1.$$

Consequently,

$$\begin{aligned}\|x_{\widetilde{T}^c\cap T} + h_{\widetilde{T}^c\cap T}\|_1 + \|x_{\widetilde{T}^c\cap T^c} + h_{\widetilde{T}^c\cap T^c}\|_1 \quad &+ \quad \mathrm{w}\|x_{\widetilde{T}\cap T} + h_{\widetilde{T}\cap T}\|_1 + \mathrm{w}\|x_{\widetilde{T}\cap T^c} + h_{\widetilde{T}\cap T^c}\|_1 \\ &\leq \quad \|x_{\widetilde{T}^c\cap T}\|_1 + \|x_{\widetilde{T}^c\cap T^c}\|_1 + \mathrm{w}\|x_{\widetilde{T}\cap T}\|_1 + \mathrm{w}\|x_{\widetilde{T}\cap T^c}\|_1.\end{aligned}$$

Since x is strictly sparse and supported on the set T, we have $x_{T^c} = 0$. Thus, using the forward and reverse triangle inequalities, we have

$$\mathrm{w}\|h_{\widetilde{T}\cap T^c}\|_1 + \|h_{\widetilde{T}^c\cap T^c}\|_1 \leq \|h_{\widetilde{T}^c\cap T}\|_1 + \mathrm{w}\|h_{\widetilde{T}\cap T}\|_1.$$

Adding and subtracting $\mathrm{w}\|h_{\widetilde{T}^c\cap T^c}\|_1$ on the left hand side, and $\mathrm{w}\|h_{\widetilde{T}^c\cap T}\|_1$ on the right, we obtain

$$\begin{aligned}\mathrm{w}\|h_{\widetilde{T}\cap T^c}\|_1 + \mathrm{w}\|h_{\widetilde{T}^c\cap T^c}\|_1 \quad &+ \quad \|h_{\widetilde{T}^c\cap T^c}\|_1 - \mathrm{w}\|h_{\widetilde{T}^c\cap T^c}\|_1 \\ &\leq \mathrm{w}\|h_{\widetilde{T}\cap T}\|_1 + \mathrm{w}\|h_{\widetilde{T}^c\cap T}\|_1 + \|h_{\widetilde{T}^c\cap T}\|_1 - \mathrm{w}\|h_{\widetilde{T}^c\cap T}\|_1.\end{aligned}$$

Consequently

$$\mathrm{w}\|h_{T^c}\|_1 + (1-\mathrm{w})\|h_{\widetilde{T}^c\cap T^c}\|_1 \leq \mathrm{w}\|h_T\|_1 + (1-\mathrm{w})\|h_{\widetilde{T}^c\cap T}\|_1.$$

Finally, by adding $(1-\mathrm{w})\|h_{\widetilde{T}\cap T^c}\|_1$ to both sides we have

$$\|h_{T^c}\|_1 \leq \mathrm{w}\|h_T\|_1 + (1-\mathrm{w})\|h_{\widetilde{T}\cap T^c}\|_1 + (1-\mathrm{w})\|h_{\widetilde{T}^c\cap T}\|_1.$$

Setting $S = (\widetilde{T} \cap T^c) \cup (\widetilde{T}^c \cap T)$, we note that when $|S| \leq s$, the above inequality is in contradiction with w-NSP(k, s, C) for $C < 1$ unless $h = 0$. We thus conclude that $x^* = x$. ■

Proposition 8. *Let A be an $m \times n$ matrix that satisfies* 1*-NSP*(s, s, C_s) *for some* $C_s < 1$ *as well as* 1*-NSP*$(k-s, k-s, C_{k-s})$ *and* 1*-NSP*(k, k, C_k) *for some finite* C_{k-s}, C_k. *Then, A satisfies* w*-NSP*$(k, s, C(\mathrm{w}))$, *with* $C(\mathrm{w}) = \frac{(1+\mathrm{w})C_s C_{k-s} + C_s + \mathrm{w}C_{k-s}}{1 - C_s C_{k-s}}$.

Proof: Let h be any fixed vector in the null space of A and let T^* and S^* be the supports of its

k-largest and s-largest entries (in modulus), respectively. To check whether $\mathrm{w}\|h_T\|_1 + (1-\mathrm{w})\|h_S\|_1 \leq C\|h_{T^c}\|_1$ holds for all $T : |T| \leq k$ and $S : |S| \leq s$, it suffices to check whether $\mathrm{w}\|h_{T^*}\|_1 + (1-\mathrm{w})\|h_{S^*}\|_1 \leq C\|h_{(T^*)^c}\|_1$ holds. To see this, note that the left hand side is largest and the right hand side is smallest over all choices of $S : |S| \leq s$ and $T : |T| \leq k$ when $(T, S) = (T^*, S^*)$. Since A satisfies 1-NSP(s, s, C_s) and 1-NSP(k, k, C_k), h_{T^*} and h_{S^*} have no zero entries on T^* or S^*, respectively (otherwise we would have $h_{(T^*)^c} = 0$, contradicting the null space property). Thus, to prove w-NSP(k, s, C) we may now examine for any vector $h : Ah = 0$ only sets T, S with $S \subset T$ and $|T| = k$, $|S| = s$. Defining $\widetilde{T} := T \setminus S$ (which implies $S = T \setminus \widetilde{T}$, and $|\widetilde{T}| = k - s$) we have

$$\begin{aligned} \mathrm{w}\|h_T\|_1 + (1-\mathrm{w})\|h_S\|_1 &= \mathrm{w}\|h_{\widetilde{T}}\|_1 + \|h_{T\cap\widetilde{T}^c}\|_1 \\ &\leq \mathrm{w}\ C_{k-s}(\|h_{T\cap\widetilde{T}^c}\|_1 + \|h_{T^c}\|_1) + \|h_{T\cap\widetilde{T}^c}\|_1. \end{aligned} \tag{4}$$

Moreover, we have $\|h_{\widetilde{T}}\|_1 \leq C_{k-s}(\|h_{T\cap\widetilde{T}^c}\|_1 + \|h_{T^c}\|_1)$. Hence,

$$\begin{aligned} \|h_{T\cap\widetilde{T}^c}\|_1 &\leq C_s(\|h_{T^c}\|_1 + \|h_{\widetilde{T}}\|_1) \\ &\leq C_s\Big((1 + C_{k-s})\|h_{T^c}\|_1 + C_{k-s}\|h_{T\cap\widetilde{T}^c}\|_1\Big), \end{aligned}$$

and consequently $\|h_{T\cap\widetilde{T}^c}\|_1 \leq \frac{C_s(1+C_{k-s})}{1-C_sC_{k-s}}\|h_{T^c}\|_1$. Substituting in (4), we obtain

$$\mathrm{w}\|h_T\|_1 + (1-\mathrm{w})\|h_S\|_1 \leq \frac{(1+\mathrm{w})C_sC_{k-s} + C_s + \mathrm{w}C_{k-s}}{1 - C_sC_{k-s}}\|h_{T^c}\|_1, \tag{5}$$

which is the desired result. ■

Corollary 9. *Let k be a positive integer and suppose that $C_{k-s} > 1$ is the smallest constant so that the $m \times n$ matrix A satisfies 1-NSP$(k-s, k-s, C_{k-s})$. Suppose there exists an integer $s < k$ such that A satisfies 1-NSP(s, s, C_s) with constant $C_s < 1/(2C_{k-s} + 1)$. Let x be any k-sparse vector in $\mathbb{R}^n$, with supp$(x) = T$. If $\widetilde{T}$ satisfies $|\widetilde{T} \cap T^c| + |\widetilde{T}^c \cap T| \leq s$ and $\mathrm{w} \leq \frac{1-2C_sC_{k-s}-C_s}{C_{k-s}(C_s+1)}$, then $x = x^*(\mathrm{w}, \widetilde{T})$, the minimizer of (3) with $y = Ax$.*

Proof: Proposition 8 implies that A satisfies w-NSP$(k, s, C(\mathrm{w}))$ with $C(\mathrm{w}) = \frac{(1+\mathrm{w})C_sC_{k-s}+C_s+\mathrm{w}C_{k-s}}{1-C_sC_{k-s}}$. If $\mathrm{w} \leq \frac{1-2C_sC_{k-s}-C_s}{C_{k-s}(C_s+1)}$ and $C_s < \frac{1}{2C_{k-s}+1}$ then $0 \leq C(\mathrm{w}) < 1$, so Proposition 7 guarantees that $x = x^*(\mathrm{w}, \widetilde{T})$. ■

Proposition 10. *Let A be an $m \times n$ matrix that satisfies w-NSP(k, s, C). If $s \leq k$, then for every weight $\mathrm{v} \leq \mathrm{w}$, the matrix A satisfies v-NSP(k, s, C).*

Proof: Since A satisfies w-NSP(k, s, C), then any vector $h : Ah = 0$ satisfies

$$\mathrm{w}\|h_T\|_1 + (1-\mathrm{w})\|h_S\|_1 \leq C\|h_{T^c}\|_1,$$

for all index sets $T \subset \{1 \ldots N\}$ with $|T| \leq k$ and $S \subset \{1 \ldots N\}$ with $|S| \leq s$. In particular, consider the sets $T^* = T^*(h)$ and $S^* = S^*(h)$ indexing the largest k and s entries in magnitude of h. We have

$$\mathrm{w}\|h_{T^*}\|_1 + (1-\mathrm{w})\|h_{S^*}\|_1 \le C\|h_{T^{*c}}\|_1.$$

Let $\mathrm{v} < \mathrm{w}$ and write the above as

$$\mathrm{v}\|h_{T^*}\|_1 + (\mathrm{w}-\mathrm{v})\|h_{T^*}\|_1 + (1-\mathrm{w})\|h_{S^*}\|_1 \le C\|h_{T^{*c}}\|_1.$$

Since $s < k$, the set $S^* \subset T^*$, which implies

$$\mathrm{v}\|h_{T^*}\|_1 + (\mathrm{w}-\mathrm{v})\|h_{S^*}\|_1 + (1-\mathrm{w})\|h_{S^*}\|_1 \le C\|h_{T^{*c}}\|_1,$$

which is equivalent to

$$\mathrm{v}\|h_{T^*}\|_1 + (1-\mathrm{v})\|h_{S^*}\|_1 \le C\|h_{T^{*c}}\|_1.$$

Replacing T^* by an arbitrary T of the same size and S^* by an arbitrary S of the same size decreases the left hand side. Replacing T^{*c} by T^c increases the right hand side. So,

$$\mathrm{v}\|h_T\|_1 + (1-\mathrm{v})\|h_S\|_1 \le C\|h_{T^c}\|_1,$$

for all S, T with $|S| \le s$ and $|T| \le k$. This is v-NSP(k, s, C), and it holds for all $\mathrm{v} < \mathrm{w}$ once it holds for w. In particular, it holds for $\mathrm{v} = 0$. ■

Remark 10.1. The condition $s \le k$ is satisfied when a support estimate set $\widetilde{T}$ with $|(\widetilde{T} \cap T^c) \cup (\widetilde{T}^c \cap T)| \le s$ has an accuracy $\alpha \ge 0.5$. Therefore, Proposition 10 states that if the support estimate is at least 50% accurate, any matrix A that satisfies w-NSP(k, s, C) also satisfies v-NSP(k, s, C) for every weight $\mathrm{v} < \mathrm{w}$.

Corollary 11. *Let A be an $m \times n$ matrix that satisfies* 1*-NSP*(k, k, C_k) *with $C_k < 1$. Then, for every k-sparse vector x supported on some set T, and for every support estimate $\widetilde{T}$ with $\alpha := \frac{|T \cap \widetilde{T}|}{|\widetilde{T}|} \ge \frac{1}{2}$ it holds that $x = x^*(\mathrm{w}, \widetilde{T})$, the minimizer of* (3) *with $b = Ax$, and $0 \le \mathrm{w} < 1$.*

Proof: This follows from Proposition 10 by setting $\mathrm{w} = 1$ and $s = k$ (and applying Proposition 7). ■

Corollary 12. *The weighted null space property* w*-NSP*(s, s, C_s) *and the standard null space property* 1*-NSP*(s, s, C_s) *are equivalent.*

Proof: Proposition 8 with $k = s$, coupled with the observation $C_0 = 0$, yield one direction of the equivalence. The other direction, i.e., that w-NSP(s, s, C_s) implies 1-NSP(s, s, C_s) follows upon picking $S = T$ for any set $T : |T| \le s$ in the definition of the weighted null space property. ■

Remark 12.1. Corollary 12 in turn implies that weighted ℓ_1-minimization recovers all s-sparse signals x from noise-free measurements Ax given a support estimate that is 50% accurate if and only if ℓ_1 minimization recovers all s-sparse signals from their noise-free measurements.

III. GAUSSIAN MATRICES

It is known that Gaussian (and more generally, sub-Gaussian) matrices satisfy the standard null space property (with high probability) provided $m > Ck\log(n/k)$. It is also known, (see, e.g., [5, Theorem 10.11]) that if a matrix $A \in \mathbb{R}^{m\times n}$ guarantees recovery of all k-sparse vectors x via ℓ_1 minimization (2), then m must exceed $c_1 k\log(\frac{n}{c_2 k})$ for some appropriate constants c_1 and c_2. The purpose of this section is to show that weighted ℓ_1 minimization allows us to recover sparse vectors beyond this bound (i.e., with fewer measurements) given relatively accurate support estimates.

We begin with some simple observations to establish a rough lower bound on the number of measurements needed for weighted ℓ_1 minimization. We first observe that when $k \geq s$, w-NSP(k,s,C) implies 1-NSP(s,s,C), i.e., the standard null space property of size s. This can be seen by restricting T to be of size at most s and setting $S = T$ in the definition of the weighted null space property. Thus, w-NSP(k,s,C) guarantees recovery of all s sparse signals via ℓ_1 minimization. Consequently, it requires $m \geq c_1 s\log(\frac{N}{c_2 s})$ [5, Theorem 10.11]. Since in weighted ℓ_1 minimization s plays the role of the size of the error in the support estimate, then one may hope (in analogy with standard compressed sensing results) that $m \approx s\log\frac{N}{s}$ suffices for recovery, given an accurate support estimate. However, even if one had a perfect support estimate, k measurements are needed to directly measure the entries on the support. Combining these observations, we seek a bound on the number of measurements that scales (up to constants) like $k + s\log\frac{N}{s}$.

Indeed this can be deduced from Corollary 14, presented later in this section, which follows from our main technical result, Theorem 13 (which is a more general version of Theorem 5 in the Introduction). Corollary 14 entails that in the case of large m, N, k, for a fixed support estimate $\widetilde{T}$ all k-sparse vectors supported on any set $T \in \Gamma_s(\widetilde{T})$ can be recovered by solving (3) when $m \gtrsim k + s\log N/s$. We conclude the section with another corollary of Theorem 13, establishing a bound on the number of Gaussian measurements that guarantee w-NSP(k,s,C).

Theorem 13. *Let T and $\widetilde{T}$ be two subsets of $\{1,...,N\}$ with $|T| \leq k$ and $|(T\cap\widetilde{T}^c)\cup(T^c\cap\widetilde{T})| \leq s$ and let A be a random matrix with independent zero-mean unit-variance Gaussian entries. Then A satisfies* w-*NSP*$(T,\widetilde{T},C)$ *with probability exceeding $1-\epsilon$ provided*

$$\frac{m}{\sqrt{m+1}} \geq \sqrt{s+\alpha\rho k} + C^{-1}\sqrt{2((\mathrm{w}^2 - 2\mathrm{w}(1-\alpha))\rho k + s)\ln(eN/k)} + \left(\frac{1}{2\pi e^3}\right)^{1/4}\sqrt{\frac{k}{\ln(eN/k)}} + \sqrt{2\ln\epsilon^{-1}}. \tag{6}$$

Proof: Our proof will be a modified version of the analogous proof for the standard null space property for Gaussian matrices [7], cf., [5], [9]. Define the set

$$H_{T,\widetilde{T}} := \left\{h \in \mathbb{R}^N : \mathrm{w}\|h_T\|_1 + (1-\mathrm{w})\|h_{(T\cap\widetilde{T}^c)\cup(T^c\cap\widetilde{T})}\|_1 \geq C\|h_{T^c}\|_1\right\}.$$

Our aim is to show that for a random Gaussian matrix A with zero-mean and unit variance entries $\inf_{h\in H_{T,\widetilde{T}}}\|Ah\|_2 > 0$ with high probability. This will show that there are no vectors from $H_{T,\widetilde{T}}$ in the null space of A, i.e., that the weighted null space property holds over T and $\widetilde{T}$. To this end, we will utilize Gordon's escape through the mesh theorem [6], as in [7], cf., [5]. In short, the theorem states that for an $m \times N$ Gaussian matrix with zero-mean and unit-variance entries and for an arbitrary set $V \subset S^{N-1}$,

$$P\Big(\inf_{v\in V}\|Av\|_2 \le \frac{m}{\sqrt{m+1}} - \ell(V) - a\Big) \le e^{-a^2/2}, \tag{7}$$

where $\ell(V) := \mathbb{E}_{g\sim\mathcal{N}(0,I_N)} \sup_{v\in V}\langle g, v\rangle$ is the Gaussian width of V, cf. [9]. So we must estimate $\ell(H_{T,\widetilde{T}} \cap S^{N-1})$. Note that $H_{T,\widetilde{T}} \cap S^{N-1}$ is compact so the supremum in the definition of $\ell(H_{T,\widetilde{T}} \cap S^{N-1})$ can be replaced by a maximum. Moreover, note that

$$\max_{h\in H_{T,\widetilde{T}}\cap S^{N-1}} \langle g, h\rangle = \max_{h\in H_{T,\widetilde{T}}\cap S^{N-1}\cap\{h:\ h_i\ge 0\}} \sum_{i=1}^{N} |g_i| h_i.$$

Define the vector $\tilde{g}$ with entries $\tilde{g}_i = |g_i|$, $i \in \{1, ..., N\}$, the convex cone $\widetilde{H}_{T,\widetilde{T}} = H_{T,\widetilde{T}} \cap \{h \in \mathbb{R}^N : h_i \ge 0\}$, and its dual $\widetilde{H}^*_{T,\widetilde{T}} := \{z \in \mathbb{R}^N : \langle z, h\rangle \ge 0 \text{ for all } h \in \widetilde{H}_{T,\widetilde{T}}\}$. We may now use duality (see, e.g., [5][B.40]) to conclude that

$$\max_{h\in\widetilde{H}_{T,\widetilde{T}}} \langle \tilde{g}, h\rangle \le \min_{z\in\widetilde{H}^*_{T,\widetilde{T}}} \|\tilde{g} + z\|_2.$$

To bound the right hand side from above, we introduce for $t \ge 0$ (to be determined later) the set

$$Q^t_{T,\widetilde{T}} := \{z \in \mathbb{R}^N : \begin{cases} z_i = \mathrm{w}t \text{ for } i \in T\cap\widetilde{T}, \\ z_i = t \text{ for } i \in T\cap\widetilde{T}^c, \\ z_i = (1-\mathrm{w})t \text{ for } i \in T^c\cap\widetilde{T}, \\ z_i \ge -Ct \text{ for } i \in T^c\cap\widetilde{T}^c \end{cases} \}$$

and we observe that for any two vectors $z \in Q^t_{T,\widetilde{T}}$ and $h \in \widetilde{H}_{T,\widetilde{T}}$

$$\sum_{i=1}^{N} z_i h_i = t\mathrm{w}\sum_{i\in T\cap\widetilde{T}} h_i + t\sum_{i\in T\cap\widetilde{T}^c} h_i + t(1-\mathrm{w})\sum_{i\in T^c\cap\widetilde{T}} h_i + \sum_{i\in T^c\cap\widetilde{T}^c} z_i h_i \tag{8}$$

$$= t\mathrm{w}\|h_{T\cap\widetilde{T}}\|_1 + t(1-\mathrm{w})\|h_{(T\cap\widetilde{T}^c)\cup(T^c\cap\widetilde{T})}\|_1 + t\mathrm{w}\|h_{T\cap\widetilde{T}^c}\|_1 + \sum_{i\in T^c\cap\widetilde{T}^c} z_i h_i \tag{9}$$

$$= t(\mathrm{w}\|h_T\|_1 + (1-\mathrm{w})\|h_{(T\cap\widetilde{T}^c)\cup(T^c\cap\widetilde{T})}\|_1) + \sum_{i\in T^c\cap\widetilde{T}^c} z_i h_i \tag{10}$$

$$\ge t(\mathrm{w}\|h_T\|_1 + (1-\mathrm{w})\|h_{(T\cap\widetilde{T}^c)\cup(T^c\cap\widetilde{T})}\|_1) - Ct\|h_{T^c\cap\widetilde{T}^c}\|_1 \tag{11}$$

$$\ge t(\mathrm{w}\|h_T\|_1 + (1-\mathrm{w})\|h_{(T\cap\widetilde{T}^c)\cup(T^c\cap\widetilde{T})}\|_1 - C\|h_{T^c\cap\widetilde{T}^c}\|_1) \ge 0. \tag{12}$$

Hence $Q^t_{T,\widetilde{T}} \subset \widetilde{H}^*_{T,\widetilde{T}}$ and so for any $t \ge 0$

$$\max_{h\in H_{T,\widetilde{T}}\cap S^{N-1}} \langle g, h\rangle \le \min_{z\in\widetilde{H}^*_{T,\widetilde{T}}} \|\tilde{g} + z\|_2 \le \min_{z\in Q^t_{T,\widetilde{T}}} \|\tilde{g} + z\|_2.$$

Taking expectations and defining $S_{Ct} : \mathbb{R} \to \mathbb{R}$, the soft-thresholding operator with $S_\lambda(x) = \mathrm{sign}(x)(\max\{|x|-\lambda/2, 0\})$, we have

$$
\begin{aligned}
\ell(H_{T,\widetilde{T}} \cap S^{N-1}) \leq & \mathbb{E} \min_{z \in Q^t_{T,\widetilde{T}}} \|\tilde{g} + z\|_2 \\
\leq & \mathbb{E}\|\tilde{g}_{T\cup(T^c\cap\widetilde{T})} + z_{T\cup(T^c\cap\widetilde{T})}\|_2 + \mathbb{E} \min_{z \in Q^t_{T,\widetilde{T}}} \|\tilde{g}_{(T^c\cap\widetilde{T}^c)} + z_{(T^c\cap\widetilde{T}^c)}\|_2 \\
\leq & \mathbb{E}\|\tilde{g}_{T\cup(T^c\cap\widetilde{T})}\|_2 + \left(\sqrt{\mathrm{w}^2\alpha\rho k + (1-\mathrm{w})^2(1-\alpha)\rho k + (1-\alpha\rho)k}\right)t \\
& + \mathbb{E} \min_{z_i \geq -Ct} \left(\sum_{i \in T^c\cap\widetilde{T}^c} (\tilde{g}_i + z_i)^2\right)^{1/2} \\
\leq & \sqrt{(1+\rho-\alpha\rho)k} + \left(\sqrt{\mathrm{w}^2\alpha\rho k + (1-\mathrm{w})^2(1-\alpha)\rho k + (1-\alpha\rho)k}\right)t \\
& + \mathbb{E}\left(\sum_{i \in T^c\cap\widetilde{T}^c} S_{Ct}(\tilde{g}_i)^2\right)^{1/2} \\
= & \sqrt{s+\alpha\rho k} + \left(\sqrt{(\mathrm{w}^2 - 2\mathrm{w}(1-\alpha))\rho k + s}\right)t + \mathbb{E}\left(\sum_{i \in T^c\cap\widetilde{T}^c} S_{Ct}(\tilde{g}_i)^2\right)^{1/2} \\
\leq & \sqrt{s+\alpha\rho k} + \left(\sqrt{(\mathrm{w}^2 - 2\mathrm{w}(1-\alpha))\rho k + s}\right)t + \left((N-s-\alpha\rho k)\left(\frac{2}{\pi e}\right)^{1/2} \frac{e^{-(Ct)^2/2}}{(Ct)^2}\right)^{1/2}.
\end{aligned} \tag{13}
$$

Above, the third inequality utilizes the facts that $z_i = \mathrm{w}t$ when $i \in T \cap \widetilde{T}$ where $|T \cap \widetilde{T}| = \alpha\rho k$, $z_i = (1-\mathrm{w})t$ when $i \in (T^c \cap \widetilde{T})$ where $|T^c \cap \widetilde{T}| = (1-\alpha)\rho k$, and that $z_i = t$ when $i \in (T \cap \widetilde{T}^c)$ where $|(T \cap \widetilde{T}^c)| = (1-\alpha\rho)k$. Moreover, the fourth inequality follows from the well known bound on standard Gaussian random vectors $v \in \mathbb{R}^m$, namely $\mathbb{E}\|v\|_2 \leq \sqrt{m}$ and a straightforward computation for the soft-thresholding term. Similarly the fifth inequality results from a direct computation of $\mathbb{E}S_t(v)^2$ where v is a standard Gaussian vector (see [5, Section 9.2] for the details).

Picking $t = C^{-1}\sqrt{2\ln(eN/k)}$, we have

$$
\ell(H_{T,\widetilde{T}} \cap S^{N-1}) \leq \sqrt{s+\alpha\rho k} + C^{-1}\sqrt{2((\mathrm{w}^2 - 2\mathrm{w}(1-\alpha))\rho k + s)\ln(eN/k)} + \left(\frac{1}{2\pi e^3}\right)^{1/4}\sqrt{\frac{k}{\ln(eN/k)}}.
$$

We now apply Gordon's escape through the mesh theorem (7) to deduce

$$
P\left(\inf_{v \in V} \|Av\|_2 \leq \frac{m}{\sqrt{m+1}} - \ell(H_{T,\widetilde{T}} \cap S^{N-1}) - a\right) \leq e^{-a^2/2}.
$$

Choosing $a = \sqrt{2\ln\epsilon^{-1}}$, we obtain w-NSP$(T, \widetilde{T}, C)$, with probability exceeding $1-\epsilon$ provided

$$
\frac{m}{\sqrt{m+1}} \geq \sqrt{s+\alpha\rho k} + C^{-1}\sqrt{2((\mathrm{w}^2 - 2\mathrm{w}(1-\alpha))\rho k + s)\ln(eN/k)} + \left(\frac{1}{2\pi e^3}\right)^{1/4}\sqrt{\frac{k}{\ln(eN/k)}} + \sqrt{2\ln\epsilon^{-1}}.
$$

□

Remark 13.1. Notice that the term

$$(\mathrm{w}^2 - 2\mathrm{w}(1-\alpha))\rho k + s = \begin{cases} s \equiv (1+\rho-2\alpha\rho)k & \text{, when } \mathrm{w}=0 \\ k & \text{, when } \mathrm{w}=1 \end{cases}$$

Moreover, the inequality $\alpha\rho \leq 1$ holds.

Remark 13.2 (Choice of weights). The Gaussian width expression in (13) allows us to choose the weight w that minimizes the upper bound on $\ell(H_{T,\widetilde{T}} \cap S^{N-1})$, in turn leading to the smallest bound on the number of measurements – as can also be seen in (6). A simple calculation shows that this weight is $\mathrm{w} = 1-\alpha$. In fact, choosing $\mathrm{w} \in (1-2\alpha, 1)$ guarantees that the second term in (6) is smaller than the analogous term for standard ℓ_1-minimization, i.e., $(\mathrm{w}^2 - 2\mathrm{w}(1-\alpha))\rho k + s < k$. Applying Proposition 10 shows that when $\alpha > 0.5$, recovery can be performed with any $\mathrm{w} \in (0, 1-\alpha]$ with this number of measurements.

Remark 13.3. The condition in Theorem 5, i.e.,

$$\frac{m}{\sqrt{m+1}} \geq \sqrt{k+s} + C^{-1}\sqrt{2(\mathrm{w}^2 k + s)\ln(eN/k)} + \left(\frac{1}{2\pi e^3}\right)^{1/4}\sqrt{\frac{k}{\ln(eN/k)}} + \sqrt{2\ln\epsilon^{-1}},$$

is a stronger version of the analogous condition in Theorem 13, simplifying the dependence of m on s, the size of error in the support estimate. It is obtained by replacing $\alpha\rho$ by 1 and $(\mathrm{w}^2 - 2\mathrm{w}(1-\alpha))\rho k$ by $\mathrm{w}^2 k$ (as these are both upper bounds).

Corollary 14. *Let $\widetilde{T}$ be a subset of $\{1,...,N\}$ and let A be a random matrix with independent zero-mean unit-variance Gaussian entries. Then, with probability exceeding $1-\epsilon$, A satisfies w-NSP$(T, \widetilde{T}, C)$ for all sets $T \subset \{1,..,,N\}$ with $|T| \leq k \leq N/2$ and $|(T \cap \widetilde{T}^c) \cup (T^c \cap \widetilde{T})| \leq s \leq k$ provided*

$$\begin{aligned}\frac{m}{\sqrt{m+1}} \geq & \left(1 + \frac{1}{(2\pi e^3)^{1/4}\sqrt{\ln(eN/k)}}\right)\sqrt{k+s} + C^{-1}\sqrt{2(\mathrm{w}^2 k + s)\ln(eN/k)} \\ & + \sqrt{2\ln\epsilon^{-1} + (s+1)\ln(eN/s) + k}. \end{aligned} \tag{14}$$

Proof: The proof consists of bounding, from above, the number of sets T satisfying the hypotheses of the corollary, and applying Theorem 13 in conjunction with a union bound. Denote $p := |\widetilde{T}|$, then the number of sets T with $|(T \cap \widetilde{T}^c) \cup (T^c \cap \widetilde{T})| \leq s$ is bounded above by $\sum_{i=0}^{s} \binom{N-p}{i}\binom{p}{k-i}$. This bound is obtained by letting $|T \cap \widetilde{T}^c|$ range from 0 to s and some simple observations. First, there are $\binom{N-p}{i}$ ways to arrange the i entries of $T \cap \widetilde{T}^c$ in $\widetilde{T}^c$. Second, there remain $k-i$ entries of T, to be "placed" in $\widetilde{T}$ and there are $\binom{p}{k-i}$ ways to do this. In addition, $p \leq 2k$ as $|T \cap \widetilde{T}| \leq k$ and $|T^c \cap \widetilde{T}| \leq s \leq k$, so combining these observations we obtain

$$\sum_{i=0}^{s}\binom{N-p}{i}\binom{p}{k-i} \leq \binom{N-p}{s}\sum_{i=0}^{s}\binom{p}{k-i} \leq (s+1)\binom{N-p}{s}\binom{2k}{k}.$$

Specifically, the last inequality is due to the observation that $p \leq 2k$ and the bound $\binom{2k}{k-i} \leq \binom{2k}{k}$.

Thus, applying Theorem 13 with $\frac{\epsilon}{(s+1)\binom{N-p}{s}\binom{2k}{k}}$ in place of ϵ, and a union bound, we conclude that the probability that at least one set T with $|(T \cap \widetilde{T}^c) \cup (T^c \cap \widetilde{T})| \le s$ violates w-NSP$(T, \widetilde{T}, C)$ is bounded above by ϵ. Note the bound

$$\ln\left((s+1)\binom{N-p}{s}\binom{2k}{k}\right) \le (s+1)\ln(eN/s) + k,$$

which finishes the proof. ∎

Remark 14.1. In the limiting case of large m, N, k with small k/N the condition (14) simplifies to

$$m \ge (\sqrt{k+s} + C^{-1}\sqrt{2(\mathrm{w}^2 k + s)\ln(eN/k)} + \sqrt{2\ln\epsilon^{-1} + (s+1)\ln(eN/s) + k})^2,$$

which reveals the benefit of using weighted ℓ_1-minimization in reducing the number of measurements. In particular, taking $\mathrm{w} = 0$ leads to the bound

$$m \gtrsim k + s\ln(eN/s),$$

which is essentially as good as one can hope for; one needs k measurements to measure the non-zero entries even if the support was fully known, and about $s\ln(eN/s)$ measurements to recover the entries where the support estimate was erroneous.

Remark 14.2. In the case $s = 0$ (i.e., perfect support estimation), we have $\widetilde{T} = T$ hence the union bound in the proof above is only over one set, which leads to the better condition (when $\mathrm{w} = 0$, and in the limit of large k, N) $m > k$. This is to be expected as when the support is known and the weight on the support set T is zero, sparse recovery via weighted ℓ_1-minimization just requires that the matrix A_T is invertible. In the case of Gaussian matrices, this occurs with probability 1 if A_T is a square matrix.

To simplify the remaining discussion, we introduce one more definition and lemma, which will facilitate the proof of our final result, Corollary 17.

Definition 15. We say that an $m \times n$ matrix A satisfies w-NSP*(k, s, C) if for all sets $T : |T| = k$ and $S : |S| = s$ with $S \subset T$, and for all vectors $h : Ah = 0$ we have $\mathrm{w}\|h_T\|_1 + (1-\mathrm{w})\|h_S\|_1 \le C\|h_{T^c}\|_1$.

Lemma 16. *w-NSP(k, s, C) is equivalent to w-NSP*(k, s, C).*

Proof: The fact that w-NSP(k, s, C) implies w-NSP*(k, s, C) follows directly from the definition of the former. To see the reverse implication note that w-NSP*(k, s, C) implies that for any $h : Ah = 0$ and sets $T^* = T^*(h)$ and $S^* = S^*(h)$ indexing the largest k and s entries (in modulus) of h, we have $\mathrm{w}\|h_{T^*}\|_1 + (1-\mathrm{w})\|h_{S^*}\|_1 \le C\|h_{T^{*c}}\|_1$. Note that $|T^*| = k$ and $|S^*| = s$, otherwise $h_{T^{*c}} = 0$ and the property fails. Moreover for the vector h, T^* and S^* maximize the left-hand side and minimize the right hand side over all choices of $T : |T| \le k$ and $S : |S| \le s$. This argument works for all vectors in the null space so we are done. ∎

Corollary 17. *Let A be a random matrix with independent zero-mean unit-variance Gaussian entries. Then, with probability exceeding $1-\epsilon$, A satisfies w-NSP(k, s, C) for all sets $T, \widetilde{T} \subset \{1, .., , N\}$ with*

$|T| \le k \le N/2$ *and* $|(T \cap \widetilde{T}^c) \cup (T^c \cap \widetilde{T})| \le s \le k$ *provided*

$$\frac{m}{\sqrt{m+1}} \ge \left(1 + \frac{1}{(2\pi e^3)^{1/4}\sqrt{\ln(eN/k)}}\right)\sqrt{k+s} + C^{-1}\sqrt{2(\mathrm{w}^2 k + s)\ln(eN/k)} + \sqrt{2\ln\epsilon^{-1} + 2k\ln(eN/k)}. \tag{15}$$

Proof: The proof is similar to the proof of the previous corollary, albeit simpler. By Proposition 16 we only need to verify w-NSP*(k, s, C), i.e., the null space property pertaining to sets $S := (T \cap \widetilde{T}^c) \cup (T^c \cap \widetilde{T}) \subset T$. There are $\binom{N}{k}$ ways of selecting T and $\binom{k}{s}$ ways of selecting $S \subset T$. We thus apply Theorem 13 with ϵ replaced with $\frac{\epsilon}{\binom{N}{k}\binom{k}{s}}$ to complete the proof. ■

Remark 17.1. In the limiting case, we see that to obtain w-NSP(k, s, C), we still need a "baseline" of $\approx k\ln(eN/k)$ measurements. Nevertheless, this number increases like $C^{-1}(w^2k+s)\ln eN/k$, rather than as $C^{-1}k\ln eN/k$ as in the case of ℓ_1 minimization. Thus, robustness to noise (which improves as C decreases, see, e.g., [5]) costs fewer measurements in the case of weighted ℓ_1-minimization, provided the support estimate is relatively accurate. A detailed analysis of the noisy measurement case is outside the scope of this paper. However, numerical evaluations conducted in [16] suggest that weight values around $1 - \alpha$ achieve the best reconstruction.

Remark 17.2. For compressible signals, we note that numerical simulations from [16] suggest that weighted ℓ_1 minimization can outperform ℓ_1 minimization provided the weights are properly chosen and the prior support information (capturing the large coefficients) is accurate enough. However, a detailed theoretical analysis for compressible signals is outside the scope of this paper and will likely rely on generalizing the ℓ_2-robust null space property [5, Chapter 4] as well as the associated analysis.

IV. Numerical Examples

In this section, we present phase transition diagrams to illustrate the theoretical results presented above. We set $N = 500$ and we draw measurement matrices A with zero mean i.i.d. standard Gaussian random entries of dimensions $m \times N$ where we vary m between 50 and 250 with an increment of 25. We then generate k-sparse signals $x \in \mathbb{R}^N$, and vary k between $\lfloor \frac{m}{10} \rfloor$ and $\lfloor \frac{m}{2} \rfloor$. The nonzero values in x are drawn independently from a standard Gaussian distribution.

We generate 50 instances of A and x. For each instance, we compute the measurement vector $y = Ax$ and compare the recovery performance using both ℓ_1 and weighted ℓ_1 minimization. Support estimate sets $\widetilde{T}$ of size k with accuracies $\alpha \in \{0.1, 0.3, 0.7, 1\}$ are generated such that αk entries of $\widetilde{T}$ are chosen at random from the support of x. The remaining $(1-\alpha)k$ entries of $\widetilde{T}$ are chosen from outside the support of x. A weight $\mathrm{w} = 1 - \alpha$ is applied to the set $\widetilde{T}$ for every run of weighted ℓ_1 minimization.

Fig. 1 shows the phase transition diagram illustrating the exact recovery rate over 50 experiments using ℓ_1 minimization for the recovery of x from measurements $y = Ax$. The recovery rate transitions from a value of 1 (white region) indicating exact recovery for all 50 experiments to a value of zero (black region) indicating errors in the recovery. We highlighted the 0.85 empirical recovery rate threshold by a dashed red line, to compare the recovery performance with weighted ℓ_1 minimization.

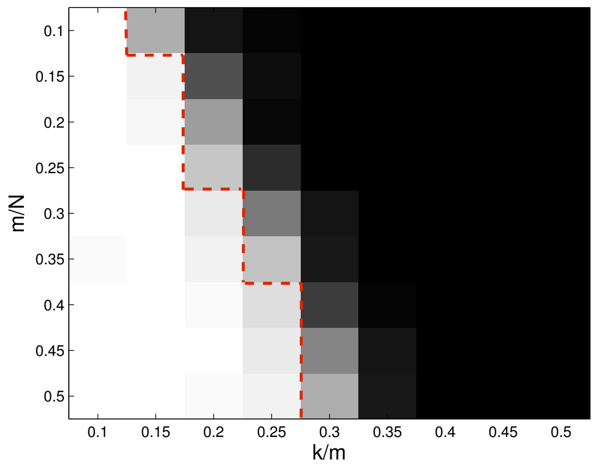


Fig. 1: Phase transition diagram showing the exact recovery rate using standard ℓ_1 minimization for 50 instances of the sparse vector x and measurement matrix A. The black region indicates a zero recovery rate, while the white region indicates full recovery with rate equal to 1. The dashed red line indicates the empirical recovery rate threshold that exceeds 0.85.

We illustrate the phase transitions of weighted ℓ_1 minimization in Figs. 2 (a)-(d), corresponding to support size $\rho = 1$ and accuracy $\alpha \in \{0.1, 0.3, 0.7, 1\}$, respectively. The solid red lines indicate the 0.85 empirical recovery rate thresholds for each of the weighted ℓ_1 problems. Notice that the recovery thresholds of the weighted ℓ_1 problems are consistently to the right of the standard ℓ_1 recovery threshold (dashed red line) when α is large enough. In some cases, the lines overlap exactly and only the solid red line is visible. For illustration purposes, we also include the threshold given by $m = k + s\log(N/s)$ (dashed green line) in the figure. This threshold is essentially the best that one can hope to achieve to guarantee exact recovery using weighted ℓ_1 minimization, as discussed in Section III.

We repeat the same experiment, albeit with different compressed sensing matrices and with $\rho = 0.7$ and $\rho = 1.3$ and present the results in Figs. 3 and 4, respectively. These experiments model the scenario where one does not have an exact estimate of the sparsity k. Here, again, we note that once α is large enough, weighted ℓ_1 minimization consistently outperforms the standard approach.

REFERENCES

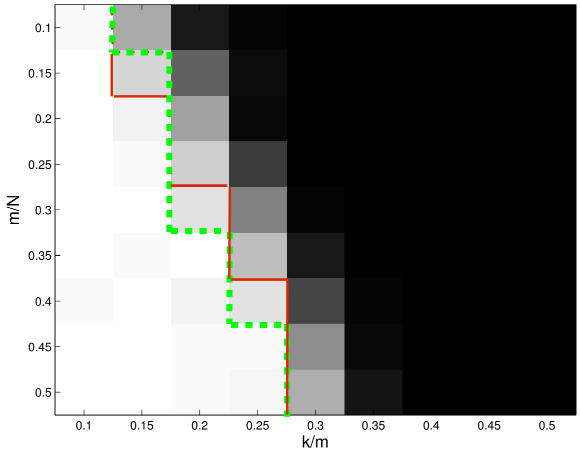
[1] D. Donoho, "Compressed sensing." *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[2] E. J. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, pp. 1207–1223, 2006.

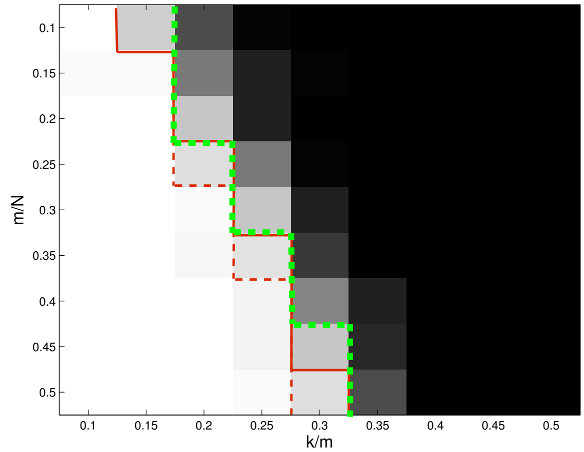
[3] ——, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, pp. 489–509, 2006.

[4] A. Cohen, W. Dahmen, and R. DeVore, "Compressed sensing and best k-term approximation," *Journal of the American Mathematical Society*, vol. 22, no. 1, pp. 211–231, 2009.

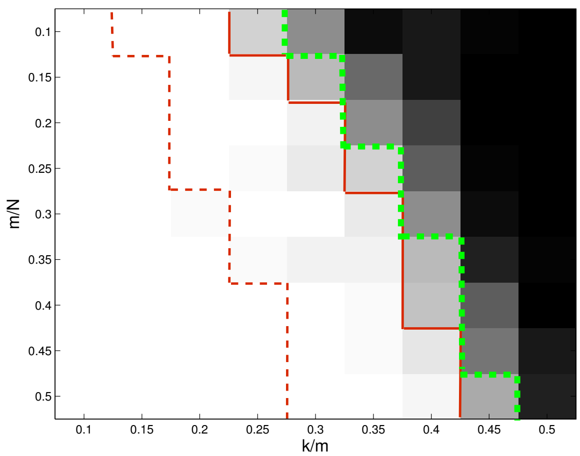
[5] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*. Birkhauser Basel, 2013.



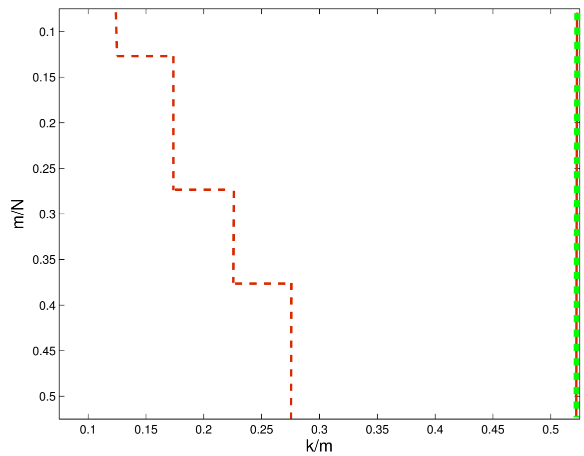
(a) $\alpha = 0.1$



(b) $\alpha = 0.3$



(c) $\alpha = 0.7$



(d) $\alpha = 1$

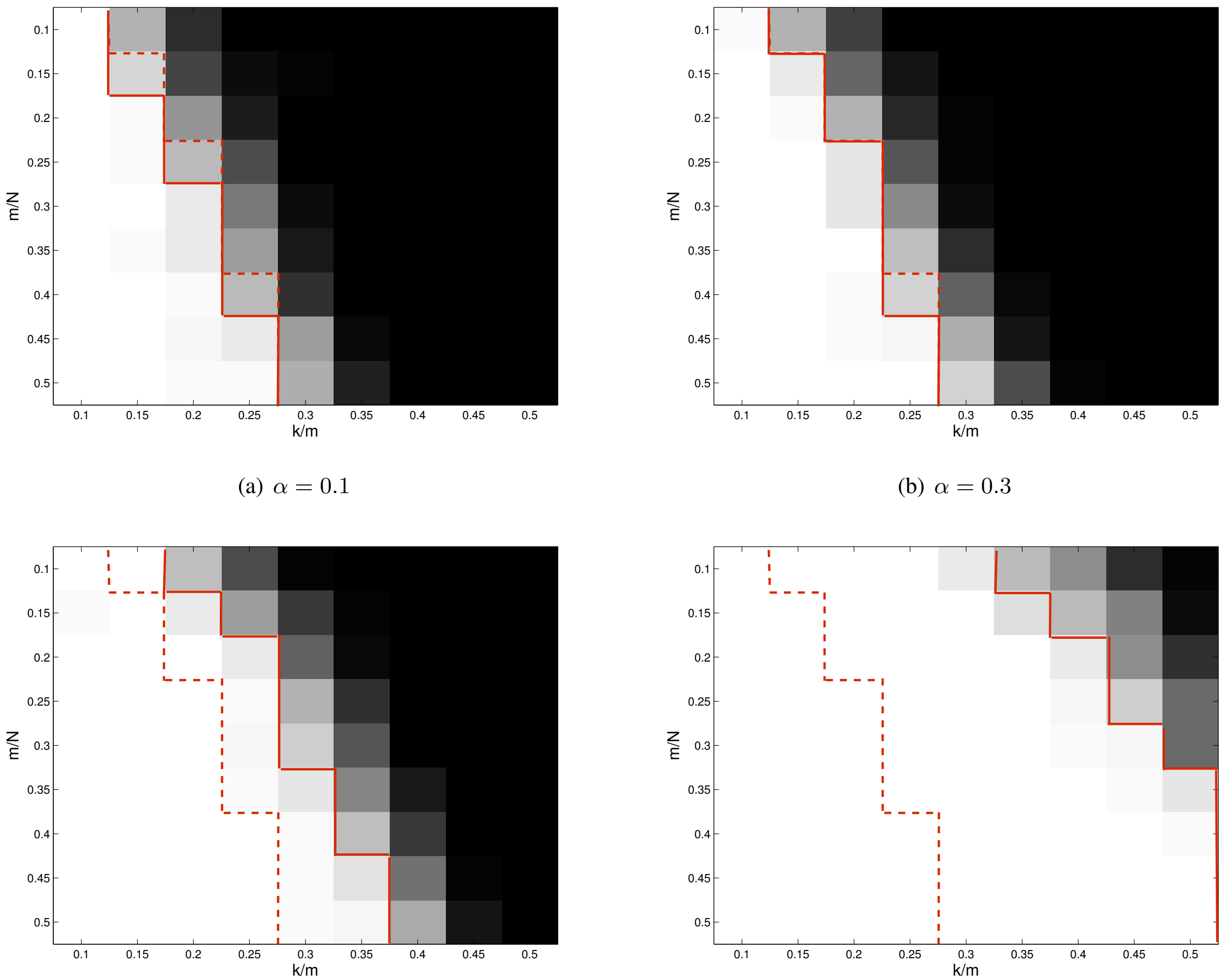
Fig. 2: Phase transition diagrams showing exact recovery rates using weighted ℓ_1 minimization with weights applied to support estimate sets $\widetilde{T}$ with $\rho = 1.3$ and varying accuracies $\alpha \in \{0.1, 0.3, 0.7, 1\}$. Here, $\mathrm{w} = 1 - \alpha$. The dashed red line corresponds to the empirical 0.85 rate threshold for standard ℓ_1 minimization. The solid red lines are the 0.85 rate thresholds for weighted ℓ_1 minimization. For comparison, the dashed green lines correspond to $m = k + s\log(N/s)$, where $s = (1 + \rho - 2\alpha\rho)k$.

[6] Y. Gordon, “On milman’s inequality and random subspaces which escape through a mesh in $\mathbb{R}^n$,” in *Geometric Aspects of Functional Analysis*, ser. Lecture Notes in Mathematics, J. Lindenstrauss and V. Milman, Eds. Springer Berlin Heidelberg, 1988, vol. 1317, pp. 84–106. [Online]. Available: http://dx.doi.org/10.1007/BFb0081737

[7] M. Rudelson and R.Vershynin, “On sparse reconstruction from fourier and gaussian measurements,” *Communications on Pure and Applied Mathematics, to appear. Conference version in: CISS 2006 (40th Annual Conference on Information Sciences and Systems).*, 2007.

[8] M. Stojnic, “Various threshold for ℓ_1 optimization in compressed sensing,” in *available at arXiv:0907.3666*, 2009.

[9] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky, “The convex geometry of linear inverse problems,”



(a) $\alpha = 0.1$

(b) $\alpha = 0.3$

(c) $\alpha = 0.7$

(d) $\alpha = 1$

Fig. 3: Phase transition diagrams showing exact recovery rates using weighted ℓ_1 minimization with weights applied to support estimate sets $\widetilde{T}$ with $\rho = 0.7$ and varying accuracies $\alpha \in \{0.1, 0.3, 0.7, 1\}$. Here, $\mathrm{w} = 1 - \alpha$. The dashed red line corresponds to the empirical 0.85 rate threshold for standard ℓ_1 minimization. The solid red lines are the 0.85 rate thresholds for weighted ℓ_1 minimization.

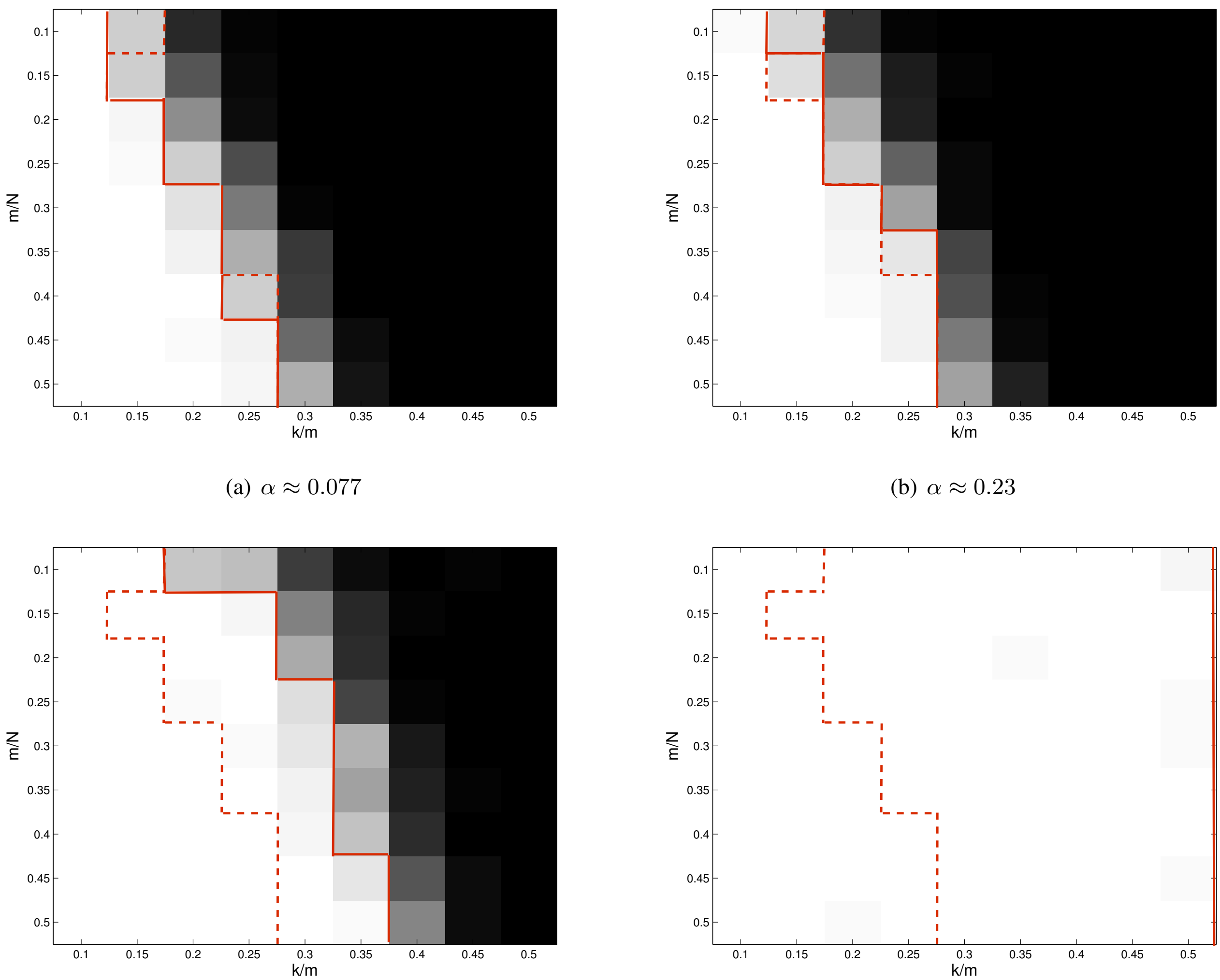
Foundations of Computational Mathematics, vol. 12, no. 6, pp. 805–849, 2012.

[10] R. von Borries, C. Miosso, and C. Potes, "Compressed sensing using prior information," in *2nd IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, CAMPSAP 2007.*, 12-14 2007, pp. 121 – 124.

[11] N. Vaswani and W. Lu, "Modified-CS: Modifying compressive sensing for problems with partially known support," *IEEE Trans. on Signal Processing*, vol. 58, no. 9, pp. 4595 – 4607, September 2010.

[12] ——, "Modified-CS: Modifying compressive sensing for problems with partially known support," in *IEEE International Symposium on Information Theory, ISIT 2009*, June 2009, pp. 488 – 492.

[13] W. Lu and N. Vaswani, "Exact reconstruction conditions and error bounds for regularized modified basis pursuit," in *Proc. of Asilomar Conf. on Signals, Systems and Computers*, July 2010.



(a) $\alpha \approx 0.077$ (b) $\alpha \approx 0.23$ (c) $\alpha \approx 0.54$ (d) $\alpha \approx 0.77$

Fig. 4: Phase transition diagrams showing exact recovery rates using weighted ℓ_1 minimization with weights applied to support estimate sets $\widetilde{T}$ with $\rho = 1.3$ and varying accuracies $\alpha \in \{0.1, 0.3, 0.7, 1\}/\rho$. Here, $\mathrm{w} = 1 - \alpha$. The dashed red line corresponds to the empirical 0.85 rate threshold for standard ℓ_1 minimization. The solid red lines are the 0.85 rate thresholds for weighted ℓ_1 minimization.

[14] L. Jacques, "A short note on compressed sensing with partially known signal support," *Signal Processing*, vol. 90, pp. 3308 – 3312, December 2010.

[15] M. Amin Khajehnejad, W. Xu, A. Salman Avestimehr, and B. Hassibi, "Weighted ℓ_1 minimization for sparse recovery with prior information," in *IEEE International Symposium on Information Theory, ISIT 2009*, June 2009, pp. 483 – 487.

[16] M. Friedlander, H. Mansour, R. Saab, and O. Yilmaz, "Recovering compressively sampled signals using partial support information," *IEEE Transactions on Information Theory*, vol. 58, no. 2, pp. 1122–1134, Feb 2012.

[17] H. Mansour and O. Yilmaz, "Weighted ℓ_1 minimization with multiple weighting sets," in *Proc. of the SPIE, Wavelets and Sparsity XIV*, August 2011, p. 13 pages, (invited).

[18] S. Oymak, M. Khajehnejad, and B. Hassibi, "Recovery threshold for optimal weight ℓ_1 minimization," in *IEEE International*

Symposium on Information Theory Proceedings (ISIT), July 2012, pp. 2032–2036.

[19] Y. Wang and W. Yin, “Sparse signal reconstruction via iterative support detection,” *SIAM Journal on Imaging Sciences*, pp. 462–491, 2010.

[20] H. Rauhut and R. Ward, “Interpolation via weighted ℓ_1 minimization,” *http://arxiv.org/abs/1308.0759*, 2014.