Weighted One-norm Minimization with Inaccurate Support Estimates:Sharp Analysis via the Null-space Property

Mansour, H.; Saab, R.

TR2015-026   April 2015

Abstract

We study the problem of recovering sparse vectors given possibly erroneous support estimates. First, we provide necessary and sufficient conditions for weighted l1 minimization to successfully recovery all sparse signals whose support estimate is sufficiently accurate. We relate these conditions to the analogous ones for l1 minimization, showing that they are equivalent when the support estimate is 50% accurate but that the weighted l1 conditions are easier to satisfy when the support is more than 50% accurate. Second, to quantify this improvement, we provide bounds on the number of Gaussian measurements that ensure, with high probability, that weighted l1 minimization succeeds. The resulting number of measurements can be significantly less than what is needed to ensure recovery via l1 minimization. Finally, we illustrate our results via numerical experiments.

IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)

WEIGHTED ONE-NORM MINIMIZATION WITH INACCURATE SUPPORT ESTIMATES: SHARP ANALYSIS VIA THE NULL-SPACE PROPERTY

Hassan Mansour

Mitsubishi Electric Research Laboratories
Cambridge, MA 02139
mansour@merl.com.

Rayan Saab

Department of Mathematics
The University of California, San Diego
La Jolla, CA 92093
rsaab@ucsd.edu.

ABSTRACT

We study the problem of recovering sparse vectors given possibly erroneous support estimates. First, we provide necessary and sufficient conditions for weighted ℓ_1 minimization to successfully recovery all sparse signals whose support estimate is sufficiently accurate. We relate these conditions to the analogous ones for ℓ_1 minimization, showing that they are equivalent when the support estimate is 50% accurate but that the weighted ℓ_1 conditions are easier to satisfy when the support is more than 50% accurate. Second, to quantify this improvement, we provide bounds on the number of Gaussian measurements that ensure, with high probability, that weighted ℓ_1 minimization succeeds. The resulting number of measurements can be significantly less than what is needed to ensure recovery via ℓ_1 minimization. Finally, we illustrate our results via numerical experiments.

Index Terms— Compressed sensing, compressive sampling, sparse approximation, ℓ_1 minimization, weighted ℓ_1 minimization, null space property

1. INTRODUCTION

Consider an arbitrary k-sparse signal in $\mathbb{R}^N$ (i.e., a vector $x \in \mathbb{R}^N$ with at most k non-zero entries) and consider its corresponding linear measurements $y \in \mathbb{R}^m$ with $m < N$, where

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

Compressed sensing [1, 2, 3] deals with recovering x from y, knowing A, using tractable algorithms. For example, in the compressed sensing context, the use of constrained ℓ_1 norm minimization for the recovery of sparse signals is now standard, (see, e.g., [1, 2, 3]). One estimates x via x^*, the minimizer of

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

This is primarily due to (2) lending itself to efficient convex optimization algorithms, and to the fact that when A is chosen appropriately, all sufficiently sparse x can be recovered

from y by solving the ℓ_1 minimization problem [1, 2, 3]. In particular, drawing the matrix at random from an appropriate distribution suffices, with high probability. For example, matrices whose entries are i.i.d. Gaussian random variables, work when[1] $m \gtrsim k \log(N/k)$.

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.

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 $T \subset \{1 \dots N\}$ with $|T| \leq k$, we have

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

In this case, we say that A satisfies $\text{NSP}(k, C)$.

A necessary and sufficient condition for the recovery of any k-sparse vector x from y using (2) is for the matrix A to satisfy $\text{NSP}(k, C)$ with $C < 1$. Moreover, it can be shown (see, e.g., [5, Section 9.4]) using tools developed in [6, 7] and [8] 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. Such a bound is essentially sharp. 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 (see, e.g., [5, Theorem 10.11]).

Nevertheless, one may hope that in the presence of useful information (such as a relatively accurate estimate of the support of the sparse signal) that one may be able to significantly reduce the number of measurements needed to guarantee recovery. In fact, it is often the case in practice that signals exhibit structured support sets, or that an estimate of the support can be identified (for example audio signals are low-frequency, the support of the transform domain coefficients of

---

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

videos does not change much from frame to frame). In this paper, we analyze a recovery method that incorporates support information by replacing (2) with weighted ℓ_1 minimization. Given a support estimate $\widetilde{T} \subset \{1, ..., N\}$, we estimate x via the minimizer of the optimization problem

$$\min_z \sum_{i=1}^N w_i |z_i| \text{ s.t. } y = Az, \text{ with } w_i = \begin{cases} w \in [0,1], & i \in \widetilde{T} \\ 1, & i \in \widetilde{T}^c \end{cases} .$$ (3)

The idea behind (3) is to choose $\text{w} := (w_i)_{i=1}^N$ such that the entries of x that are "expected" to be large, i.e., those on the support estimate $\widetilde{T}$, are penalized less.

1.1. Prior work

Incorporating support estimates into algorithms for recovering signals from compressed sensing measurements has been studied in, e.g., [9, 10, 11, 12, 13, 14, 15, 16, 17, 18]. Vaswani and Lu [10, 11, 12] used 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] when a large proportion of the support is known. Jacques [13] extended these results to deal with compressible signals and noisy measurements. Friedlander et al. [15] considered applying non-zero weights to the support estimate, further generalizing and refining the results of Vaswani and Lu. They provided tighter sufficient conditions for recovery, and their conditions depended on the accuracy and size of the support estimate. Subsequently, Mansour et al. [16] extended these results to the case when multiple support estimates with varying accuracies are available.

Using a different model for the support information, Khajehnejad et al. [14] derive sufficient recovery conditions for weighted ℓ_1 minimization to recover sparse signals from compressed sensing measurements. In [14], one partitions $\{1, \ldots, N\}$ to two sets and the entries of x supported on each set have a fixed probability of being non-zero, albeit the probabilities differ between the sets. Here the "support information" consists of knowing the partition and probabilities. Adopting the same model, Oymak et al. [17] derive 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. Here, every new instance of the problem requires a new draw of the random measurement matrix.

In this paper, we provide uniform recovery guarantees, i.e., the matrix is drawn once and successful recovery is guaranteed (with high probability) for all sparse signals when sufficiently accurate support information is available.

1.2. Notation

Throughout the paper, $\widetilde{T}$ is the support estimate used in (3). The size of $\widetilde{T}$ is ρk for some $\rho > 0$ and the accuracy of $\widetilde{T}$ is $\alpha = \frac{|\widetilde{T} \cap T|}{|\widetilde{T}|}$. For an index set $S \subset \{1, ..., N\}$ we define

$$\Gamma_s(S) := \left\{ T \subset \{1, ..., N\} : \left| (S \cap T^c) \cup (S^c \cap T) \right| \le s \right\}.$$

We denote the restriction of the vector x to the indices in $T \subset \{1, ..., N\}$ by x_T. Our results require that we introduce a *weighted non-uniform* null space property.

Definition 2. Consider the sets $T \subset \{1 \ldots N\}$ with $|T| \le k$ and $\widetilde{T} \in \Gamma_s(T)$, and define $S = (\widetilde{T} \cap T^c) \cup (\widetilde{T}^c \cap T)$. A matrix $A \in \mathbb{R}^{m \times N}$ is said to have the weighted nonuniform null space property over the sets T and $\widetilde{T}$ with constant C if for any vector $h : Ah = 0$, we have

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

In this case, we say A satisfies w-NSP$(T, \widetilde{T}, C)$.

We also require a uniform version of the above definition.

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 $T \subset \{1 \ldots N\}$ with $|T| \le k$ and $S \subset \{1 \ldots N\}$ with $|S| \le s$, we have

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

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

Thus, the standard null space property NSP(k, C) can be written as 1-NSP(k, k, C). Moreover, we note that 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.

2. RECOVERY GUARANTEES VIA THE WEIGHTED NULL-SPACE PROPERTY

Our first main result is Theorem 4. [2]

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 $\widetilde{T} \in \Gamma_s(\text{supp}(x))$ if and only if A satisfies w-NSP(k, s, C) for some $C < 1$.*

In particular, the theorem states that if A does not satisfy w-NSP(k, s, C) for any $C < 1$, then there exists a k-sparse vector x supported on some set T with $Ax = y$ and a set

---

[2]Due to space limitations, we omit the proofs of our results. They can be found in a technical report available at http://www.math.ucsd.edu/~rsaab/publications.htm and will be included in a journal version of this manuscript.

$\widetilde{T}$ with $|(\widetilde{T} \cap T^c) \cup (\widetilde{T}^c \cap T)| \leq s$ such that x is not the unique minimizer of (3). On the other hand, if A 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 $|(\widetilde{T} \cap T^c) \cup (\widetilde{T}^c \cap T)| \leq s$. Thus recovery is guaranteed via (3) provided the support estimate is accurate enough.

The following proposition shows that weighted ℓ_1 minimization can guarantee recovery even when ℓ_1 minimization cannot recover all k-sparse signals.

Proposition 5. *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*

$$C(\mathrm{w}) = \frac{(1 + \mathrm{w})C_s C_{k-s} + C_s + \mathrm{w}C_{k-s}}{1 - C_s C_{k-s}}.$$

To see that Proposition 5 shows that weighted ℓ_1 minimization can succeed where ℓ_1 minimization fails, first note that if $C_k > 1$ then there exists a k-sparse signal that cannot be recovered by ℓ_1 minimization. Moreover, if one chooses $\mathrm{w} \leq \frac{1 - 2C_s C_{k-s} - C_s}{C_{k-s}(C_s + 1)}$ and if $C_s < \frac{1}{2C_{k-s} + 1}$ then $0 \leq C(\mathrm{w}) < 1$, so Theorem 4 guarantees that x is the minimizer of (3) whenever $\widetilde{T} \in \Gamma_s(T)$, where T is the support of x. To further illustrate the relationship between ℓ_1 minimization and its weighted counterpart, we present a corollary of Theorem 4 and Proposition 5.

Corollary 6. *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.*

Corollary 6 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. To see this, note that w-NSP$(k, k, C_k), C_k < 1$ guarantees via Theorem 4 that every k-sparse vector x is the unique minimizer of (3) provided $\widetilde{T}$ satisfies $|(\widetilde{T} \cap T^c) \cup (\widetilde{T}^c \cap T)| \leq k$. Define $\rho := \frac{|\widetilde{T}|}{k}$; since the accuracy of the support estimate, $\alpha := \frac{|T \cap \widetilde{T}|}{|\widetilde{T}|} \geq 1/2$, then $|(T \cap \widetilde{T}^c) \cup (T^c \cap \widetilde{T})| = (1 + \rho - 2\alpha\rho)k \leq k$, so we are done.

3. GAUSSIAN MATRICES

Our second main result deals with matrices $A \in \mathbb{R}^{m \times N}$ whose entries are i.i.d. Gaussian random variables, which we refer to as Gaussian matrices. We establish a lower bound on m above which these matrices satisfy the weighted null space property (with high probability), and hence, guarantee exact sparse recovery using weighted ℓ_1 minimization. However, we begin with some simple observations to establish a rough lower bound on the number of measurements

necessary for weighted ℓ_1 minimization to guarantee exact recovery. Observe that w-NSP(k, s, C) (with $k > s$) implies 1-NSP(s, s, C), i.e., the standard null space property of size s (this can be seen from the definitions of the properties). Consequently, w-NSP(k, s, C) guarantees recovery of all s sparse signals via ℓ_1 minimization so it requires $m \geq c_1 s \log(\frac{N}{c_2 s})$ (see, e.g., [5, Theorem 10.11]). This may lead us to hope that $m \approx s \log \frac{N}{s}$ suffices for recovery given an accurate support estimate because in weighted ℓ_1 minimization s plays the role of the size of the support estimate's error. However, even if one had a perfect support estimate, k measurements would still be 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 $k + s \log \frac{N}{s}$. We begin with a non-uniform result (Theorem 7) whose proof consists of a suitably modified version of the analogous proof for the standard null space property for Gaussian matrices [19], cf., [5], [8]. Again, the details are omitted for space considerations.

Theorem 7. *Let T and $\widetilde{T}$ be two subsets of $\{1, ..., N\}$ with $|T| \leq k$ and $|(T \cap \widetilde{T}^c) \cup (T^c \cap \widetilde{T})| \leq s \leq k$ and let A be a random matrix with independent standard Gaussian entries. Then A satisfies w-NSP$(T, \widetilde{T}, C)$ with probability exceeding $1 - \epsilon$ provided*

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

Importantly, the theorem suggests that given a particular support estimate accuracy α (which in turn governs the relationship between s and k), one should optimize w to minimize the number of measurements in the theorem. Interestingly, the optimal choice satisfies $\mathrm{w} = 1 - \alpha$, which agrees with the intuition that a smaller weight should be applied when the confidence in the support estimate is high, and vice versa.

Moreover, one can use Theorem 7, combined with a counting argument and a union bound, to obtain a similar result which this time holds uniformly for all k-sparse signals supported on sets from $\Gamma_s(\widetilde{T})$.

Corollary 8. *Let $\widetilde{T}$ be a subset of $\{1, ..., N\}$ and let A be a random matrix with independent standard 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*

$$\frac{m}{\sqrt{m+1}} \geq \left(1 + \frac{1}{(2\pi e^3)^{1/4}\sqrt{\ln(eN/k)}}\right)\sqrt{k + s}$$
$$+ C^{-1}\sqrt{2(\mathrm{w}^2 k + s)\ln(eN/k)}$$
$$+ \sqrt{2 \ln \epsilon^{-1} + (s + 1)\ln(eN/s) + k}. \tag{5}$$

In the limiting case of large m, N, k with small k/N the condition (5) simplifies to

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

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
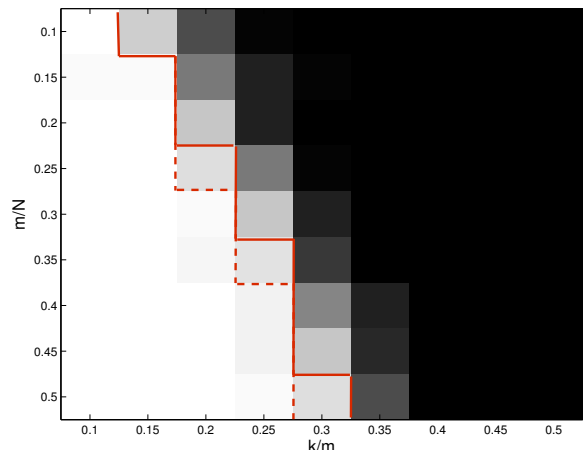
$$m \gtrsim k + (1 + C^{-2})s\ln(eN/s),$$

which is essentially as good as one can hope for; one needs k measurements to recover 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. In particular, this bound can be significantly smaller than the analogous $m \gtrsim (1 + C^{-1})^2 k \ln eN/k$ of standard ℓ_1 minimization [5] when s is significantly smaller than k, i.e., when the support estimate is very accurate.
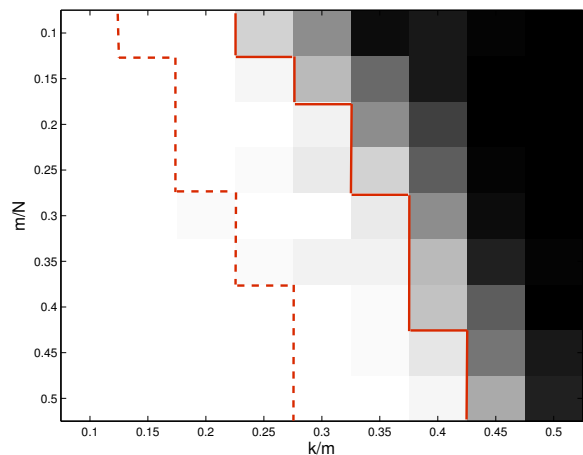
4. NUMERICAL EXPERIMENTS

We present numerical simulations to illustrate our theoretical results. Throughout, we set $N = 500$ and draw $m \times N$ matrices A with i.i.d. standard Gaussian random entries, with m ranging between 50 and 250 in increments of 25. We generate k-sparse signals $x \in \mathbb{R}^N$, and vary k between $\frac{m}{10}$ and $\frac{m}{2}$ in increments of $\frac{m}{20}$. 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 of ℓ_1 and weighted ℓ_1 minimization. Specifically, support estimate sets $\widetilde{T}$ of size k with accuracies $\alpha \in \{0.3, 0.7\}$ are generated such that αk entries of $\widetilde{T}$ are chosen at random from the support of x. The remaining 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 weighted ℓ_1 minimization.

We illustrate the performance of weighted ℓ_1 minimization in Figs. 1 (a)-(b), corresponding to support estimate accuracies of $\alpha = 0.3, 0.7$. 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 shifted to the right compared to the standard ℓ_1 recovery threshold (dashed red line) for values of α considered. Notice also that when $\alpha = 0.7$, weighted ℓ_1 minimization significantly outperforms standard ℓ_1.

(a) $\alpha = 0.3$

(b) $\alpha = 0.7$

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

5. 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] E. J. Candès, J. Romberg, and T. Tao, "Robust un-

certainty 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] Simon Foucart and Holger Rauhut, *A Mathematical Introduction to Compressive Sensing*, Birkhauser Basel, 2013.

[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*, Joram Lindenstrauss and VitaliD. Milman, Eds., vol. 1317 of *Lecture Notes in Mathematics*, pp. 84–106. Springer Berlin Heidelberg, 1988.

[7] M. Stojnic, "Various threshold for l1 optimization in compressed sensing," in *available at arXiv:0907.3666*, 2009.

[8] Venkat Chandrasekaran, Benjamin Recht, Pablo A. Parrilo, and Alan S. Willsky, "The convex geometry of linear inverse problems," *Foundations of Computational Mathematics*, vol. 12, no. 6, pp. 805–849, 2012.

[9] R. von Borries, C.J. 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.

[10] Namrata Vaswani and Wei 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.

[11] N. Vaswani and Wei Lu, "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.

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

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

[14] M. Amin Khajehnejad, Weiyu Xu, A. Salman Avestimehr, and B. Hassibi, "Weighted l1 minimization for sparse recovery with prior information," in *IEEE International Symposium on Information Theory, ISIT 2009*, June 2009, pp. 483 – 487.

[15] M.P. 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.

[16] 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).

[17] S. Oymak, M.A Khajehnejad, and B. Hassibi, "Recovery threshold for optimal weight l1 minimization," in *IEEE International Symposium on Information Theory Proceedings (ISIT)*, July 2012, pp. 2032–2036.

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

[19] 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.