SUPPORT DRIVEN REWEIGHTED ℓ_1 MINIMIZATION

Hassan Mansour and Özgür Yılmaz
University of British Columbia, Vancouver - BC, Canada

ABSTRACT

In this paper, we propose a support driven reweighted ℓ_1 minimization algorithm (SDRL1) that solves a sequence of weighted ℓ_1 problems and relies on the support estimate accuracy. Our SDRL1 algorithm is related to the IRL1 algorithm proposed by Candès, Wakin, and Boyd. We demonstrate that it is sufficient to find support estimates with *good* accuracy and apply constant weights instead of using the inverse coefficient magnitudes to achieve gains similar to those of IRL1. We then prove that given a support estimate with sufficient accuracy, if the signal decays according to a specific rate, the solution to the weighted ℓ_1 minimization problem results in a support estimate with higher accuracy than the initial estimate. We also show that under certain conditions, it is possible to achieve higher estimate accuracy when the intersection of support estimates is considered. We demonstrate the performance of SDRL1 through numerical simulations and compare it with that of IRL1 and standard ℓ_1 minimization.

Index Terms— Compressed sensing, iterative algorithms, weighted ℓ_1 minimization, partial support recovery

1. INTRODUCTION

Compressed sensing is a relatively new paradigm for the acquisition of signals that admit sparse or nearly sparse representations using fewer linear measurements than their ambient dimension [1, 2].

Consider an arbitrary signal $x \in \mathbb{R}^N$ and let $y \in \mathbb{R}^n$ be a set of measurements given by $y = Ax + e$, where A is a known $n \times N$ measurement matrix, and e denotes additive noise that satisfies $\|e\|_2 \leq \epsilon$ for some known $\epsilon \geq 0$. Compressed sensing theory states that it is possible to recover x from y (given A) even when $n \ll N$, that is, using very few measurements. When x is strictly sparse—i.e., when there are only $k < n$ nonzero entries in x—and when $e = 0$, one may recover an estimate $\hat{x}$ of the signal x by solving the constrained ℓ_0 minimization problem

$$\underset{u \in \mathbb{R}^N}{\text{minimize}} \ \|u\|_0 \text{ subject to } \ Au = y. \tag{1}$$

hassanm@cs.ubc.ca; oyilmaz@math.ubc.ca

Both authors were supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC) Collaborative Research and Development Grant DNOISE II (375142-08). Ö. Yılmaz was also supported in part by an NSERC Discovery Grant.

However, ℓ_0 minimization is a combinatorial problem and quickly becomes intractable as the dimensions increase. Instead, the convex relaxation

$$\underset{u \in \mathbb{R}^N}{\text{minimize}} \ \|u\|_1 \text{ subject to } \|Au - y\|_2 \leq \epsilon \qquad \text{(BPDN)}$$

also known as *basis pursuit denoise* (BPDN) can be used to recover an estimate $\hat{x}$. Candés, Romberg and Tao [2] and Donoho [1] show that (BPDN) can stably and robustly recover x from inaccurate and what appears to be "incomplete" measurements $y = Ax + e$ if A is an appropriate measurement matrix, e.g., a Gaussian random matrix such that $n \gtrsim k \log(N/k)$. Contrary to ℓ_0 minimization, (BPDN) is a convex program and can be solved efficiently. Consequently, it is possible to recover a stable and robust approximation of x by solving (BPDN) instead of (1) at the cost of increasing the number of measurements taken.

Several works in the literature have proposed alternate algorithms that attempt to bridge the gap between ℓ_0 and ℓ_1 minimization. These include using ℓ_p minimization with $0 < p < 1$ which has been shown to be stable and robust under weaker conditions than those of ℓ_1 minimization, see [3, 4, 5]. *Weighted ℓ_1 minimization* is another alternative if there is prior information regarding the support of the signal to-be-recovered as it incorporates such information into the recovery by weighted basis pursuit denoise (w-BPDN)

$$\underset{u}{\text{minimize}} \ \|u\|_{1,\mathrm{w}} \text{ subject to } \|Au - y\|_2 \leq \epsilon, \quad \text{(w-BPDN)}$$

where $\mathrm{w} \in [0, 1]^N$ and $\|u\|_{1,\mathrm{w}} := \sum_i \mathrm{w}_i |u_i|$ is the weighted ℓ_1 norm (see [6, 7, 8]). Yet another alternative, the *iterative reweighted ℓ_1 minimization* (IRL1) algorithm proposed by Candès, Wakin, and Boyd [9] and studied by Needell [10] solves a sequence of weighted ℓ_1 minimization problems with the weights $\mathrm{w}_i^{(t)} \approx 1/\left|x_i^{(t-1)}\right|$, where $x_i^{(t-1)}$ is the solution of the $(t-1)$th iteration and $\mathrm{w}_i^{(0)} = 1$ for all $i \in \{1 \dots N\}$.

In this paper, we propose a support driven iterative reweighted ℓ_1 (SDRL1) minimization algorithm that uses a small number of support estimates that are updated in every iteration and applies a constant weight on each estimate. The algorithm, presented in section 2, relies on the accuracy of each support estimate as opposed to the coefficient magnitude to improve the signal recovery. While we still lack a proof that SDRL1 outperforms ℓ_1 minimization, we present two results in section 3 that motivate SDRL1 and could lead towards such a proof. First, we prove that if x belongs to a class of signals that satisfy certain decay conditions and given a support estimate with accuracy larger than 50%, solving a weighted ℓ_1 minimization problem with constant weights is guaranteed to produce a support estimate with higher accuracy. Second, we show that under strict conditions related to the distribution of coefficients in a support estimate, it is possible to achieve higher estimate accuracy when the intersection of support estimates is considered. Finally, we demonstrate through numerical experiments in section 4 that the performance of our proposed algorithm is similar to that of IRL1.

Notation: For a vector $x \in \mathbb{R}^N$ and set $\Lambda \subset \{1 \dots N\}$, x_k and $x|_k$ refer to the largest k entries of x, $x(k)$ is the kth largest entry of x, x_Λ refers to the entries of x indexed by Λ, and $x^{(t)}$ is the vector x at interation t

2. ITERATIVE REWEIGHTED ℓ_1 MINIMIZATION

In this section, we give an overview of the IRL1 algorithm, proposed by Candès, Wakin, and Boyd [9] and present our proposed support driven reweighted ℓ_1 (SDRL1) algorithm.

2.1. The IRL1 algorithm

IRL1 algorithm solves a sequence of (w-BPDN) problems where the weights are chosen according to $\mathrm{w}_i = \frac{1}{|\tilde{x}_i|+a}$. Here $\tilde{x}_i$ is an estimate of the signal coefficient at index i (from the previous iteration) and a is a stability parameter. The choice of a affects the stability of the algorithm and different variations are proposed for the sparse, compressible, and noisy recovery cases. The algorithm is summarized in Algorithm 1.

Algorithm 1 IRL1 algorithm [9]

Input $y = Ax + e$
Output $x^{(t)}$
Initialize $\mathrm{w}_i^{(0)} = 1$ for all $i \in \{1 \dots N\}$, a
    $t = 0$, $x^{(0)} = 0$
while $\|x^{(t)} - x^{(t-1)}\|_2 \leq \mathrm{Tol}\|x^{(t-1)}\|_2$ **do**
    $t = t + 1$
    $x^{(t)} = \arg\min_u \|u\|_{1,W}$ s.t. $\|Au - y\|_2 \leq \epsilon$
    $\mathrm{w}_i = \frac{1}{|\tilde{x}_i|+a}$
end while

The rationale behind choosing the weights inversely proportional to the estimated coefficient magnitude comes from the fact that large weights encourage small coefficients and small weights encourage large coefficients. Therefore, if the true signal were known exactly, then the weights would be set equal to $\mathrm{w}_i = \frac{1}{|x_i|}$. Otherwise, weighting according to an approximation of the true signal and iterating was demonstrated to result in better recovery capabilities than standard ℓ_1 minimization. In [10], the error bounds for IRL1 were shown to be tighter than those of standard ℓ_1 minimization. However, aside from empirical studies, no provable results have yet been obtained to show that IRL1 outperforms standard ℓ_1.

2.2. Support driven reweighted ℓ_1 (SDRL1) algorithm

In [8], we showed that solving the weighted ℓ_1 problem with constant weights applied to a support estimate set $\widetilde{T}$ has better recovery guarantees than standard ℓ_1 minimization when the $\widetilde{T}$ is at least 50% accurate. Moreover, we showed in [11] that using multiple weighting sets improves on our previous result when additional information on the support estimate accuracy is available. Motivated by these works, we propose the SDRL1 algorithm (Algorithm 2), a support driven iterative reweighted ℓ_1 minimization algorithm, which identifies two support estimates that are updated in every iteration and applies constant weights on these estimates. The SDRL1 algorithm relies on the support estimate accuracy as opposed to the coefficient magnitude.

Algorithm 2 Support driven reweighted ℓ_1 (SDRL1) algorithm.

Input $y = Ax + e$
Output $x^{(t)}$
Initialize $\hat{p} = 0.99$, $\hat{k} = n\log(N/n)/2$,
    $\omega_1 = 0.5$, $\omega_2 = 0$, Tol, $T_1 = \emptyset$, $\Omega = \emptyset$,
    $t = 0$, $s^{(0)} = 0$, $x^{(0)} = 0$
while $\|x^{(t)} - x^{(t-1)}\|_2 \leq \mathrm{Tol}\|x^{(t-1)}\|_2$ **do**
    $t = t + 1$
    $\Omega = \mathrm{supp}(x^{(t-1)}|_{s^{(t-1)}}) \cap T_1$
    Set the weights equal to

$$\mathrm{w}_i = \begin{cases} 1, & i \in T_1^c \cap \Omega^c \\ \omega_1, & i \in T_1 \cap \Omega^c \\ \omega_2, & i \in \Omega \end{cases}$$

    $x^{(t)} = \arg\min_u \|u\|_{1,\mathbf{w}}$ s.t. $\|Au - y\|_2 \leq \epsilon$
    $l = \min_\Lambda |\Lambda|$ s.t. $\|x_\Lambda^{(t)}\|_2 \geq \hat{p}\|x^{(t)}\|_2$,
    $T_1 = \mathrm{supp}(x^{(t)}|_{s^{(t)}})$, where $s^{(t)} = \min\{l, \hat{k}\}$
end while

Note that we use two empirical parameters to control the size of the support estimate T_1. The first parameter $\hat{k}$ approximates the minimum sparsity level recoverable by (BPDN). The second parameter l is the number of largest coefficients of $x^{(t)}$ that contribute an ad hoc percentage $\hat{p}$ of the signal energy. The size of T_1 is set equal to the minimum of $\hat{k}$ and l.

3. MOTIVATING THEORETICAL RESULTS

The SDRL1 algorithm relies on two main premises. The first is the ability to improve signal recovery using a sufficiently accurate support estimate by solving a weighted ℓ_1 minimization problem with constants weights. The second is the intersection set of two support estimates has at least the higher accuracy of either set.

Let $x \in \mathbb{R}^N$ be an arbitrary signal and suppose we collect $n \ll N$ linear measurements $y = Ax$, $A \in \mathbb{R}^{n \times N}$ where n is small enough (or k is large enough) that it is not possible to recover x exactly by solving (BPDN) with $\epsilon = 0$. Denote by $\hat{x}$ the solution to (BPDN), and by $\hat{x}^\omega$ the solution to (w-BPDN) with weight ω applied to a support estimate set $\widetilde{T}$. Let x_k be the best k-term approximation of x and denote by $T_0 = \text{supp}(x_k)$ the support set of x_k.

Proposition 3.1. *Suppose that $\widetilde{T}$ is of size k with accuracy (with respect to T_0) $\alpha_0 = \frac{s_0}{k}$ for some integer $k/2 < s_0 < k$. If A has the restricted isometry property (RIP) with constant $\delta_{(a+1)k} < \frac{a-\gamma^2}{a+\gamma^2}$ for some $a > 1$ and $\gamma = \omega + (1-\omega)\sqrt{2-2\alpha_0}$, and if there exists a positive integer d_1 such that*

$$|x(s_0+d_1)| \geq (\omega\eta+1)\|x_{T_0^c}\|_1 + (1-\omega)\eta\|x_{T_0^c\cap\widetilde{T}^c}\|_1, \quad (2)$$

where $\eta = \eta_\omega(\alpha_0)$ is a well behaved constant, then the set $S = \text{supp}(x_{s_0+d_1})$ is contained in $T_\omega = \text{supp}(\hat{x}^\omega_k)$.

Remark 3.1.1. The constant $\eta_\omega(\alpha)$ is given explicitly by

$$\eta_\omega(\alpha) = \frac{2\left(\sqrt{1+\delta_{ak}} + \sqrt{a}\sqrt{1-\delta_{(a+1)k_0}}\right)}{\sqrt{a}\sqrt{1-\delta_{(a+1)k}} - (\omega+(1-\omega)\sqrt{2-2\alpha})\sqrt{1+\delta_{ak}}}.$$

Proof outline. The proof of Proposition 3.1 is a direct extension of our proof of Proposition 3.2 in [11]. In particular, we want to find the conditions on the signal x and the matrix A which guarantee that the set $S = \text{supp}(x_{s_0+d_1})$ is a subset of $T_\omega = \text{supp}(\hat{x}^\omega_k)$. This is achieved when $\hat{x}^\omega$ satisfies

$$\min_{j\in S} |\hat{x}^\omega(j)| \geq \max_{j\in T_\omega^c} |\hat{x}^\omega(j)|. \quad (3)$$

Since A has RIP with $\delta_{(a+1)k} < \frac{a-\gamma^2}{a+\gamma^2}$, it has the Null Space property (NSP) [12] of order k, i.e., for any $h \in \mathcal{N}(A)$, $Ah = 0$, then $\|h\|_1 \leq c_0\|h_{T_0^c}\|_1$, with $c_0 = 1 + \frac{\sqrt{1+\delta_{ak}}}{\sqrt{a}\sqrt{1-\delta_{(a+1)k}}}$. Define $h = \hat{x}^\omega - x$, then $h \in \mathcal{N}(A)$ and one can show that

$$\|h\|_1 \leq \eta\left(\omega\|x_{T_0^c}\|_1 + (1-\omega)\|x_{T_0^c\cap\widetilde{T}^c}\|_1\right) \quad (4)$$

In other words, (w-BPDN) is ℓ_1-ℓ_1 instance optimal with these error bounds. The proof of this fact is a direct extension of the ℓ_1-ℓ_1 instance optimality of (BPDN) as shown in [12] and we omit the details here. Next, we rewrite (4) as

$$\|h_{T_0}\|_1 \leq (\omega\eta+1)\|x_{T_0^c}\|_1 + (1-\omega)\eta\|x_{T_0^c\cap\widetilde{T}^c}\|_1 - \|\hat{x}^\omega_{T_0^c}\|_1.$$

To complete the proof, we make the following observations:
(i) $\min_{j\in S} |\hat{x}^\omega(j)| \geq \min_{j\in S} |x(j)| - \max_{j\in S} |x(j) - \hat{x}^\omega(j)|$,
(ii) $\|\hat{x}^\omega_{T_0^c}\|_1 \geq \max_{j\in T_\omega^c} |\hat{x}^\omega(j)|$
which after some manipulations –see [11], proof of Prop. 3.2 for details of a similar calculation in a different setting– imply

$$\begin{aligned}\min_{j\in S} |\hat{x}^\omega(j)| \geq \;& \max_{j\in T_\omega^c} |\hat{x}^\omega(j)| + \min_{j\in S} |x(j)| \\ & -(\omega\eta+1)\|x_{T_0^c}\|_1 + (1-\omega)\eta\|x_{T_0^c\cap\widetilde{T}^c}\|_1.\end{aligned} \quad (5)$$

Finally, we observe from (5) that (3) holds, i.e., $S \subseteq T_\omega$, if

$$|x(s_0+d_1)| \geq (\omega\eta+1)\|x_{T_0^c}\|_1 + (1-\omega)\eta\|x_{T_0^c\cap\widetilde{T}^c}\|_1.$$

□

Proposition 3.1 shows that if the signal x satisfies condition (2) and $\frac{s_0}{k} > 0.5$, then the support of the largest k coefficients of $\hat{x}^\omega$ contains at least the support of the largest $s_0 + d_1$ coefficients of x for some positive integer d_1.

Next we present a proposition where we focus on an idealized scenario: Suppose that the events $E_i := \{i \in T\}$, for $i \in \{1,\ldots,N\}$ and $T \subseteq \{1,\ldots,N\}$, are independent and have equal probability with respect to an appropriate discrete probability measure P. In this case, we show below, that the accuracy of $\Omega = \widetilde{T}\cap T_\omega$ is at least as high as the higher of the accuracies of $\widetilde{T}$ and T_ω. For simplicity, we use the notation $\mathrm{P}(T_0|\widetilde{T})$ to denote $\mathrm{P}(i \in T_0 | i \in \widetilde{T})$.

Proposition 3.2. *Let x be an arbitrary signal in $\mathbb{R}^N$ and denote by T_0 the support of the best k-term approximation of x. Let the sets $\widetilde{T}$ and T_ω be each of size k and suppose that $\widetilde{T}$ and T_ω contain the support of the largest s_0 and $s_1 > s_0$ coefficients of x, respectively. Define the set $\Omega = \widetilde{T}\cap T_\omega$. Given a discrete probability measure P, the events $E_i := \{i \in T\}$, for $i \in \{1,\ldots,N\}$ and $T \subseteq \{1,\ldots,N\}$, are independent and equiprobable. Then, for $\rho := \mathrm{P}(T_\omega|\widetilde{T}) \geq \frac{s_0}{k}$, the accuracy of the set Ω is given by*

$$\mathrm{P}(T_0|\Omega) = \frac{1}{\rho}\frac{s_0}{k}.$$

Proof outline. The proof follows directly using elementary tools in probability theory. In particular, we have $\mathrm{P}(T_0|\widetilde{T}) = \frac{s_0}{k}$, and $\mathrm{P}(T_0|T_\omega) = \frac{s_1}{k}$. Define $\rho = \mathrm{P}(T_\omega|\widetilde{T}) \geq \frac{s_0}{k}$, it is easy to see that $\mathrm{P}(T_0\cap\widetilde{T}|T_0\cap T_\omega) = \frac{s_0}{s_1}$ which leads to $\mathrm{P}(T_0\cap\Omega) = \mathrm{P}(T_0\cap T_\omega)\mathrm{P}(T_0\cap\widetilde{T}|T_0\cap T_\omega) = \frac{s_1}{N}\frac{s_0}{s_1} = \frac{s_0}{N}$. Consequently, $\mathrm{P}(T_0|\Omega) = \frac{\mathrm{P}(T_0\cap\Omega)}{\mathrm{P}(T_\omega|\widetilde{T})\mathrm{Pr}(\widetilde{T})} = \frac{s_0/N}{\rho(k/N)} = \frac{1}{\rho}\frac{s_0}{k}$. □

Proposition 3.2 indicates that as $\mathrm{Pr}(T_\omega|\widetilde{T}) \to \frac{s_0}{k}$, then $\mathrm{Pr}(T_0|\Omega) \to 1$. Therefore, when x satisfies (2) it could be beneficial to solve a weighted ℓ_1 problem where we can take advantage of the possible improvement in accuracy on the set $\widetilde{T}\cap T_\omega$. Finally, we note that there are more complex dependencies between the entries of $\widetilde{T}$ and T_ω of Algorithm 2 for which Proposition 3.2 does not account.

4. NUMERICAL RESULTS

We tested our SDRL1 algorithm by comparing its performance with IRL1 and standard ℓ_1 minimization in recovering synthetic signals x of dimension $N = 2000$. We first recover sparse signals from compressed measurements of x using matrices A with i.i.d. Gaussian random entries and dimensions $n \times N$ where $n \in \{N/10, N/4, N/2\}$. The sparsity of the signal is varied such that $k/n \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$. To quantify the reconstruction performance, we plot in Figure 1 the percentage of successful recovery averaged over 100 realizations of the same experimental conditions. The figure shows that both the proposed algorithm and IRL1 have a comparable performance which is far better than standard ℓ_1 minimization.

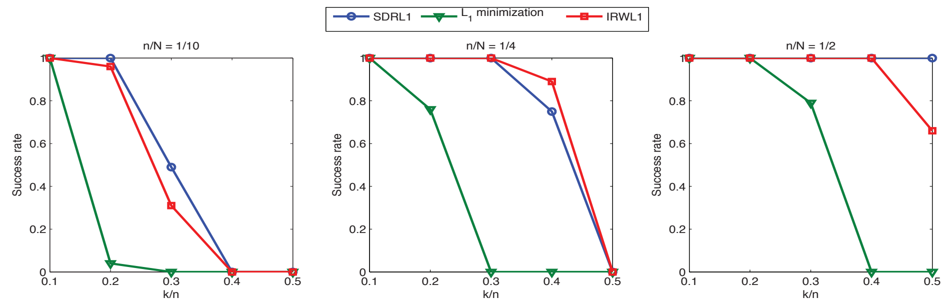

Fig. 1. Comparison of the percentage of exact recovery of *sparse signals* between the proposed SDRL1 algorithm, IRL1 [9], and standard ℓ_1 minimization. The signals have an ambient dimension $N = 2000$ and the sparsity and number of measurements are varied. The results are averaged over 100 experiments.

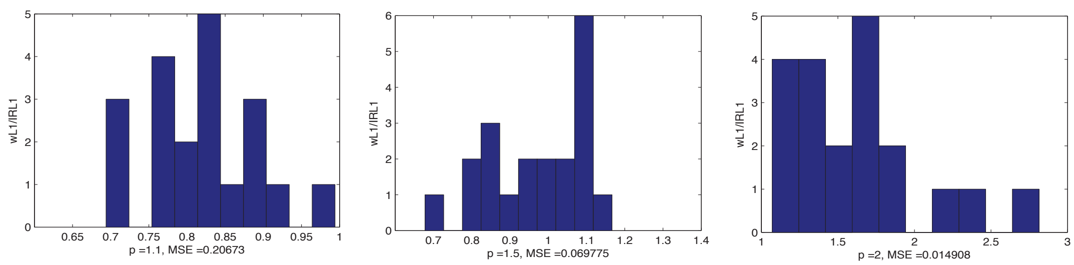

Fig. 2. Histogram of the ratio of the mean squared error (MSE) between the proposed SDRL1 and IRL1 [9] for the recovery of *compressible signals*. The signals x follow a power law decay such that $|x_i| = ci^{-p}$, for constant c and exponent p.

Next, we generate compressible signals with power law decay such that $x(i) = ci^{-p}$ for some constant c and decay power p. We consider the case where $n/N = 0.1$ and the decay power $p \in \{1.1, 1.5, 2\}$ and plot the ratio of the reconstruction error of SDRL1 over that of IRL1. Figure 2 shows the histograms of the ratio for 100 experiments each. Note that a ratio smaller than one means that our algorithm has a smaller reconstruction error than that of IRL1. The histograms indicate that both algorithms have a comparable performance for signals with different decay rates.

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] R. Gribonval and M. Nielsen, "Highly sparse representations from dictionaries are unique and independent of the sparseness measure," *Applied and Computational Harmonic Analysis*, vol. 22, no. 3, pp. 335–355, May 2007.

[4] R. Chartrand and V. Staneva, "Restricted isometry properties and nonconvex compressive sensing," *Inverse Problems*, vol. 24, no. 035020, 2008.

[5] R. Saab and O. Yilmaz, "Sparse recovery by non-convex optimization – instance optimality," *Applied and Computational Harmonic Analysis*, vol. 29, no. 1, pp. 30–48, July 2010.

[6] 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.

[7] N. Vaswani and Wei Lu, "Modified-CS: Modifying compressive sensing for problems with partially known support," *arXiv:0903.5066v4*, 2009.

[8] M. P. Friedlander, H. Mansour, R. Saab, and Ö. Yılmaz, "Recovering compressively sampled signals using partial support information," *to appear in the IEEE Trans. on Inf. Theory*.

[9] E. J. Candès, Michael B. Wakin, and Stephen P. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *The Journal of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 877–905, 2008.

[10] D. Needell, "Noisy signal recovery via iterative reweighted l1-minimization," in *Proceedings of the 43rd Asilomar conference on Signals, systems and computers*, 2009, Asilomar'09, pp. 113–117.

[11] H. Mansour and Ö. Yılmaz, "Weighted ℓ_1 minimization with multiple weighting sets," in *SPIE, Wavelets and Sparsity XIV*, August 2011.

[12] 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.