REFERENCE-BASED COMPRESSED SENSING: A SAMPLE COMPLEXITY APPROACH

João F. C. Mota† Lior Weizman‡ Nikos Deligiannis‡ Yonina C. Eldar† Miguel R. D. Rodrigues‡

† Electronic and Electrical Engineering Department, University College London, UK
‡ Department of Electrical Engineering, Technion — Israel Institute of Technology, Israel
¶ Department of Electrical and Informatics, Vrije Universiteit Brussel, Belgium

ABSTRACT

We address the problem of reference-based compressed sensing: reconstruct a sparse signal from few linear measurements using as prior information a reference signal, a signal similar to the signal we want to reconstruct. Access to reference signals arises in applications such as medical imaging, e.g., through prior images of the same patient, and compressive video, where previously reconstructed frames can be used as reference. Our goal is to use the reference signal to reduce the number of required measurements for reconstruction. We achieve this via a reweighted $\ell_1-\ell_1$ minimization scheme that updates its weights based on a sample complexity bound. The scheme is simple, intuitive and, as our experiments show, outperforms prior algorithms, including reweighted $\ell_1$ minimization, $\ell_1-\ell_1$ minimization, and modified CS.

Index Terms— Compressed sensing, reweighted $\ell_1$ minimization, prior information, sample complexity.

1. INTRODUCTION

Compressed Sensing allows acquiring signals at rates much lower than the Nyquist rate [1–3]. Applying it requires three elements: a basis in which the signals are sparse, an acquisition matrix with specific properties, and a nonlinear procedure to reconstruct signals from their measurements, e.g., $\ell_1$-norm minimization. After the initial work [1,2], much research focused on reducing acquisition rates even further, by leveraging more structured signal information [4–8], using prior information [9–13], or improving reconstruction algorithms, e.g., via reweighting schemes [17–21].

In this paper, we propose a reweighted scheme for a reconstruction problem that uses as prior knowledge a reference signal. Specifically, let $x^* \in \mathbb{R}^n$ be a sparse signal of which we have $m$ linear measurements $y = Ax^*$, where $A \in \mathbb{R}^{m \times n}$ is the measurement matrix (or its product with a sparsifying basis). Assume we know a reference signal $\mathbf{r} \in \mathbb{R}^n$, close to $x^*$ in the $\ell_1$-norm sense, i.e., $||x^* - \mathbf{r}||_1$ is assumed small. Using the measurements $y$ and reference $\mathbf{r}$, $x^*$ can be reconstructed via weighted $\ell_1-\ell_1$ minimization:

$$\min_{x} \left\| d \circ x \right\|_1 + \left\| w \circ (x - \mathbf{r}) \right\|_1 \quad \text{subject to} \quad Ax = y, \quad (1)$$

where $\circ$ denotes the entrywise product between two vectors, and $d, w \in \mathbb{R}_+^n$ have nonnegative entries. Problem (1) generalizes weighted $\ell_1$-norm minimization [10,22], in which $w$ is the zero vector, and also $\ell_1-\ell_1$ minimization [13,14], where both $d$ and $w$ are the vector of ones. Given that $d$ and $w$ are free parameters, they can be chosen in order to minimize the number of measurements required for reconstruction. In general, however, their optimal value depends on $x^*$ and is therefore unknown. To address this uncertainty, we consider a reweighting scheme: starting from arbitrary $d^0$ and $w^0$, we create a sequence $\{x^k\}_{k=1}^K$ such that, for $k = 1, \ldots, K$,

$$x^k \in \arg\min_{x} \left\| d^k \circ x \right\|_1 + \left\| w^k \circ (x - \mathbf{r}) \right\|_1 \quad \text{s.t.} \quad Ax = y, \quad (2)$$

where $d^k$ and $w^k$ are functions of $x^{k-1}$, the vector reconstructed at the previous iteration. If $d^k$ and $w^k$ are well-chosen, then the number of measurements to recover $x^*$ should decrease as we iterate (2). Our goal is to devise strategies to compute $d^k$ and $w^k$ at each iteration.

1.1. Overview and contributions

Our approach consists of two steps:

1) Obtaining a bound on the number of measurements $m$ above which (1) is guaranteed to reconstruct $x^*$; the bound depends on $x^*$ and is therefore uncomputable.

2) Computing $d^k$ and $w^k$ at iteration $k$ such that an approximation of the bound of step 1) is minimized; the approximation results from replacing the unknown signal $x^*$ by its current best estimate, $x^{k-1}$.

Our result establishing the bound in step 1) says that $O(\frac{1}{\zeta} \log n)$ measurements suffice to reconstruct $x^*$ via (1), where $\zeta$ and $\eta$ are functions of the weights $d$ and $w$. We show that if $d$ and $w$ are chosen properly, $\zeta/\eta^2$ can be made arbitrarily small, in which case the bound, and thus the number of required measurements, becomes a constant independent of $n$. This contrasts with known bounds for other problems, e.g., basis pursuit [7], weighted $\ell_1$ minimization [22], or simple $\ell_1-\ell_1$ minimization [13], which require $O(c \log n)$ measurements with $c$ having the same order of magnitude as the sparsity of $x^*$. We use the above property in the design of our reweighting scheme in step 2): at each iteration, $w$ and $d$ are computed so that $\zeta/\eta^2$ is minimized. To our knowledge, this approach to reweighting is the first one to use a sample complexity bound to update its weights. Although the bound looks complex [see (5)], the resulting scheme is simple and intuitive (see Algorithm 1). Furthermore, our experiments show that it outperforms prior reweighting schemes, including reweighted $\ell_1$ minimization [17], and static schemes that use prior information, such as $\ell_1-\ell_1$ minimization [13] and modified-CS [9].

1.2. Related Work

Reweighting has been applied in least squares problems as far back as [23,24]. For sparse reconstruction problems, [17] proposed a sim-
ple algorithm known as reweighted $\ell_1$ minimization: each weight $d_i$ is updated at iteration $k$ as $d_i^{k+1} = 1/(|x_i^k|^2 + \epsilon)$, where $\epsilon > 0$ and $x^* = \arg \min_{x \in S}$ is a solution of weighted $\ell_1$ minimization with weights $d_i$, i.e., (2) with $w^0 = 0$ (the zero vector). That algorithm and variations were analyzed in [25, 26]. Other reweighting schemes for sparse reconstruction include [18, 27], which solve simpler problems per iteration, namely least squares problems, and are therefore computationally more efficient. Regarding sparse reconstruction using prior information, [20, 21] proposed a reweighting algorithm for a slight variation of problem (1) in the context of MRI reconstruction. There, the weights are updated as $w^k = 1/(|x_i^k|^2 + 1)$ and $w_i^{k+1} = 1/(|x_i^k|^2 - |x_i^k|^2 + 1)$, and the resulting scheme is shown to significantly improve MRI reconstruction.

2. REWEIGHTED $\ell_1$-1 MINIMIZATION

2.1. Step 1: Bound on the number of measurements

The number of measurements that (1) requires to reconstruct $x^*$ depends on several problem parameters, namely on how the vectors $\tau$, $d$, and $w$ interact. To capture those interactions, we define the sets

$I = \{i : x_i^0 \neq 0\}, \quad J = \{i : x_i^0 \neq \tau_i\}, \quad K = \{i : d_i \neq w_i\}$

$I_+ = \{i : x_i^0 > 0\}, \quad J_+ = \{i : x_i^0 > \tau_i\}, \quad K_+ = \{i : d_i > w_i\}$

$I_- = \{i : x_i^0 < 0\}, \quad J_- = \{i : x_i^0 < \tau_i\}, \quad K_- = \{i : d_i < w_i\}.$

In words, $I, J, K$ are the supports of $x^*, x^* - \tau$, and $d - w$; the subscript + (resp. –) indicates that $x_i^0 > \tau_i$ (resp. $x_i^0 < \tau_i$) and also that there exist two (different) indices $i$ and $j$ such that $0 \neq x_i^0 \neq \tau_i$ and $x_j^0 = \tau_j$. If

$$m \geq 2 \frac{\zeta}{\eta} \log \left( \frac{n}{\eta + h} \right) + \frac{7}{5} (K + h) + \theta + 1,$$

then, with probability at least $1 - \exp \left( -\frac{1}{2} \left( m - \sqrt{mn} \right)^2 \right)$, $x^*$ is the unique solution of (1).

This theorem, whose proof uses the concept of Gaussian width [7, 28, 29], generalizes Theorem 1 in [13], which established a similar bound for the particular case $d = w = 1_n$.

We mentioned before that $\zeta/\eta$ can be made arbitrarily small. To see why, suppose $d$ and $w$ were selected so that $Q_+ = \emptyset$. Then, according to (4b)-(4c), the set over which $\zeta$ is defined, $I J$, does not intersect any of the sets over which $\eta$ is defined, i.e., $I J \cap Q_- = I J \cap (Q_+ \cup J' F') = \emptyset$. In other words, the set of components of $d$ and $w$ that contribute to $\zeta$ are independent from the components that contribute to $\eta$. Therefore, $\zeta/\eta$ can be arbitrarily small. As shown next, this is not the case for alternative reconstruction problems.

Comparison with other reconstruction problems. Table 1 compares our bound for Weighted $\ell_1$-$\ell_1$ minimization ($w_\ell_1-\ell_1$) with bounds obtained using similar tools for other methods: $\ell_1$-$\ell_1$ minimization [13], Modified-CS (Mod-CS) [9], and Basis Pursuit (BP) [30]. These problems have the same format as (1), but their objective functions are as shown in the table. In Mod-CS, $I$ is an estimate of the support $I$ of $x^*$ and is used as prior information. Prior information in $\ell_1$-$\ell_1$ is, as in our case, a reference signal $\tau$. Only BP uses no prior information. Table 1 also shows where the displayed bounds were computed. In the bound for Mod-CS, $0 < \xi < 1$, and $b$ is the sum of false negatives and false positives in the estimation of $I$, i.e., $b = |I \cap \tilde{I}| + |I^c \cap \tilde{I}|$.

Table 1. Sample complexity of alternative reconstruction schemes.

<table>
<thead>
<tr>
<th>Prob.</th>
<th>Objective Function</th>
<th>Bound</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$w_\ell_1-\ell_1$</td>
<td>$|d \circ x|_1 + |w \circ (x - \tau)|_1$</td>
<td>$O(2 n \log n)$</td>
<td>here</td>
</tr>
<tr>
<td>$\ell_1-\ell_1$</td>
<td>$|x|_1 + |x - \tau|_1$</td>
<td>$O(2 n \log n)$</td>
<td>[13]</td>
</tr>
<tr>
<td>Mod-CS</td>
<td>$\sum_{i \in I}</td>
<td>x_i</td>
<td>$</td>
</tr>
<tr>
<td>BP</td>
<td>$|x|_1$</td>
<td>$O(2 s \log n)$</td>
<td>[7]</td>
</tr>
</tbody>
</table>

Table 1 compares our bound for Weighted $\ell_1$-$\ell_1$ minimization ($w_\ell_1-\ell_1$) with bounds obtained using similar tools for other methods: $\ell_1$-$\ell_1$ minimization [13], Modified-CS (Mod-CS) [9], and Basis Pursuit (BP) [30]. These problems have the same format as (1), but their objective functions are as shown in the table. In Mod-CS, $I$ is an estimate of the support $I$ of $x^*$ and is used as prior information. Prior information in $\ell_1$-$\ell_1$ is, as in our case, a reference signal $\tau$. Only BP uses no prior information. Table 1 also shows where the displayed bounds were computed. In the bound for Mod-CS, $0 < \xi < 1$, and $b$ is the sum of false negatives and false positives in the estimation of $I$, i.e., $b = |I \cap \tilde{I}| + |I^c \cap \tilde{I}|$. Thus, for $\ell_1$-$\ell_1$ and Mod-CS, $\tilde{h}$ and $\tilde{b}$ measure the quality of the prior information: the better the quality, the smaller $\tilde{h}$ and $\tilde{b}$. This means the number of measurements required by $\ell_1$-$\ell_1$ and Mod-CS is determined by the quality of the prior information. For $w_\ell_1-\ell_1$ minimization, however, the ratio $\zeta/\eta$ can be arbitrarily small, independently of the quality of the prior information (of course, it has to have a “minimum quality” to satisfy the assumptions of Theorem 1; see footnote 1). Making $\zeta/\eta$ small, however, requires selecting the weights $d$ and $w$ properly. Our reweighting scheme, presented next, attempts to do exactly that.

2.2. Step 2: reweighting scheme

Algorithm 1 describes the method we propose. Its parameters are $r_{\min}$ and $r_{\max}$ which, as we will see, determine the amount by which the bound in (5) is minimized, $\epsilon_1$ and $\epsilon_2 > 0$, which are used in the definition of the sets $I$ and $J$, and the number of iterations $K$. At iteration $k$, the algorithm obtains an estimate $x^k$ of $x^*$ by solving weighted $\ell_1$-$\ell_1$ minimization with weights $d^k$ and $w^k$ (step 2). Note

1These assumptions can be stated equivalently as $I J \neq \emptyset$ and $I^c J^c \neq \emptyset$, and specify a minimum quality certificate for the prior information $\tau$. On the other hand, the assumptions $\zeta, \eta > 0$ are necessary to make (5) well-defined.

2http://www.ee.ucl.ac.uk/~jmota/reL1L1.pdf

3Note that the first term of (5) is dominant for sparse signals. In particular, it can be shown that $(7/5)(\tilde{h} + \tilde{b}) + \theta \leq (17/5)s + \pi$, where $s$ (resp. $\pi$) is the sparsity of $x^*$ (resp. $\tau$). This follows from $\tilde{h} + \tilde{b} = |I| \leq |I| = s$ and $\theta = |I| + |J| \leq 2s + \pi$. 

Algorithm 1  Reweighted $f_1$-$f_1$ minimization

**Input:** $A \in \mathbb{R}^{m \times n}$, $y \in \mathbb{R}^m$, $\pi \in \mathbb{R}^n$ (prior information)

**Parameters:** $0 < r_{\min} < r_{\max}$, $\epsilon_1, \epsilon_2 > 0$, $K$ (# iterations)

**Initialization:** $d^1 = r^1 = 1$, $k = 1$

1. for $k = 1, \ldots, K$ do
2. Obtain $x^k$ by solving
   
   $$\begin{align*}
   &\text{minimize } \|d^k \circ x\|_1 + \|w^k \circ (x - \pi)\|_1 \\
   &\text{subject to } Ax = y
   \end{align*}$$
3. Set $I^k = \{ i : |x^k_i| > \epsilon_1 \}$ and $J^k = \{ i : |x^k_i - \pi_i| > \epsilon_2 \}$
4. for $i = 1, \ldots, n$ do
5. if $i \in I^k \cap J^k$ then $d_{i}^{k+1} = w_{i}^{k+1} = r_{\min}$
6. else if $i \in I^k \setminus J^k$ then $d_{i}^{k+1} = r_{\min}$, $w_{i}^{k+1} = r_{\max}$
7. else if $i \in J^k \setminus I^k$ then $d_{i}^{k+1} = r_{\max}$, $w_{i}^{k+1} = r_{\min}$
8. else if $i \in J^k \cap J^k$ then $d_{i}^{k+1} + w_{i}^{k+1} = r_{\max}$
9. end
10. end for
11. end for

that because $d$ and $w$ are initialized as $1$, the first iteration is simply $f_1$-$f_1$ minimization [13]. Then, using $x^k$, the sets $I$ and $J$ are estimated via thresholding in step 3. Recall that $I$ and $J$ depend on the unknown vector $x^*$; so, we estimate them by using our current best guess: $x^k$. The weights $d$ and $w$ for the next iteration are then computed in steps 4-10. Note that they take only two values: $r_{\max}$ and $r_{\min}$. This is a consequence of the way we derive the algorithm, as explained later in the section. Although Algorithm 1 is derived with the goal of minimizing the bound in Theorem 1, the way it updates the weights is actually quite intuitive.

Intuition. Consider, for example, $i \in I^k \cap J^k$, i.e., it is estimated that $\pi_i \neq x^*_i \neq 0$ (step 5). The algorithm sets the corresponding weights $d_i$ and $w_i$ to a small value $r_{\min}$, which means that $x^*_i$ will be estimated solely from the measurements $y = Ax^*$. If, on the other hand, $i \in I^k \setminus J^k$, i.e., it is estimated that $\pi_i = x^*_i \neq 0$, the algorithm sets $d_i$, to a small value, to avoid penalizing large values for $x^*_i$, and sets $w_i$ to a large value, penalizing deviations from an apparently accurate component of $\pi$. Similarly, if $i \in J^k \setminus I^k$, i.e., it is estimated that $\pi_i \neq x^*_i = 0$, $w_i$ is set to a small value, since $\pi_i$ seems to be inaccurate, and $d_i$ is set to a large value, since $x^*_i$ is likely to be zero. Finally, if $i \in J^k \cap J^k$, i.e., it is estimated that $\pi_i = x^*_i = 0$, both $d_i$ and $w_i$ are set to large values since, very likely, $x^*_i$ is zero. These updates, beyond intuitive, lead to a reduction of the number of required measurements, as shown next.

Corollary 2. Let $x^*, \pi \in \mathbb{R}^n$ and $A \in \mathbb{R}^{m \times n}$ be as in Theorem 1. Consider Algorithm 1 and suppose the sets $I$ and $J$ are correctly estimated at iteration $K - 1$, i.e., $I^{K-1} = I$ and $J^{K-1} = J$. If the number of measurements satisfies

$$m \geq \left( \frac{r_{\min}}{r_{\max} - r_{\min}} \right)^2 \frac{n}{h} \log \left( \frac{n}{h + h} \right) + \left( 1 + 7 \right) + 1,$$

then, with probability at least $1 - \exp \left( -\frac{1}{2} (m - \sqrt{m})^2 \right)$, Algorithm 1 outputs $x^*$.

**Proof.** The weights used at iteration $K$ are computed at iteration $K - 1$. Hence, the last instance of $w \cdot f_1 - t_1$ in step 2 is solved with

$$d_i = r_{\min}, \quad w_i = r_{\min}, \quad \text{for all } i \in I^I$$

$$d_i = r_{\max}, \quad w_i = r_{\max}, \quad \text{for all } i \in I^J$$

$$d_i = r_{\max}, \quad w_i = r_{\min}, \quad \text{for all } i \in I^E$$

$$d_i = r_{\min}, \quad w_i = r_{\max}, \quad \text{for all } i \in I^W$$

Note that (7b) implies $I^J K_+ = \emptyset$ and (7c) implies $I^J K_- = \emptyset$, that is, $Q_+ = \emptyset$. This means the parameter $\xi$ in (4b) equals $r_{\min} \sum_{c \in J} (sg(x^*_c) + sg(x^*_c - \pi^c))^2 = r_{\min} h$, where we used (3). We also have $\theta = 0$ [cf. (4a)]. According to (4c) and (7b)-(7d), $\eta$ equals $r_{\max} - r_{\min}$ if $Q_- \neq \emptyset$, and 2$r_{\max}$ otherwise (note that, by assumption, $I^E J^I \neq \emptyset$; see footnote 1). As a consequence, the bound in (5) becomes

$$m \geq \left( \frac{r_{\min}}{2r_{\max}} \right)^2 \frac{n}{h} \log \left( \frac{n}{h + h} \right) + \left( 1 + 7 \right) + 1,$$

when $Q_- \neq \emptyset$, and becomes (6) otherwise. Note, however, that (6) implies (8). Therefore, whether or not $Q \neq \emptyset$, all the assumptions of Theorem 1 hold, and thus the statement of the corollary is true. $\square$

Although this result requires the strong assumption that $I$ and $J$ are correctly estimated at iteration $K - 1$, it shows that Algorithm 1 may reduce the number of required measurements significantly. If $r_{\max} \gg r_{\min}$, the dominant term of (6) becomes approximately $(r_{\min}/r_{\max})^2 \log n$. Thus, under the corollary’s assumptions, setting $r_{\max} = \sqrt{\log n} r_{\min}$ makes the number of required measurements required by Algorithm 1 a constant independent of $n$.

**Derivation of the scheme.** We now explain how to arrive at Algorithm 1. Given estimates of $I$ and $J$ at iteration $k$, we want to find $d$ and $w$ minimizing the ratio $\xi^2/\eta$, subject to $\eta > 0$ and $\eta > 0$ [cf. Theorem 1]. Such a problem is ill-posed, as it has no minimizer: the infimum is 0, but it can never be achieved because of the constraints. So, rather than minimizing $\xi^2/\eta^2$ formally i.e., with an optimization algorithm, we do it heuristically. In particular, we only allow two values for the weights: $r_{\min}$ and $r_{\max}$.

To aid our derivation, Table 2 shows the sets involved in the definitions of $\xi$ and $\eta$, and describes how the respective components of $d$ and $w$ should relate to minimize the $\xi^2/\eta^2$. Consider, for example, a component $i \in I^J$. It contributes $g_i(d_i, w_i) = (d_i s_g(x^*_i) + w_i s_g(x^*_i - \pi^c))^2$ to $\xi$ and has no influence on $\eta$. There are two scenarios: either $i \in I^J \cup J^I \cup J^E$, or $i \in I^J \cup J^E$ or $I^E$. In the former, we have $s_g(x^*_i) = s_g(x^*_i - \pi^c)$, and $g_i(d_i, w_i)$ has a unique minimizer at $d_i = w_i = 0$: $g_i(0,0) = 0$. However, we cannot set $d_i = w_i = 0$, since (5) is valid only for $d_i, w_i > 0$; rather, we set these components to a small value, $r_{\min}$, if $\geq 0$. When $i \in I^J \cup J^I \cup J^E$, $g_i(d_i, w_i)$ has an infinite set of minimizers, $\{ (d_i, w_i) : d_i = w_i \}$, from which we select $d_i = w_i = r_{\min}$ so that all the components in $I^J K_+$ are treated similarly; any other choice for a common value would also work. Consider now a component $i \in I^J K_+$: it contributes with $(d_i - w_i)^2$ to $\xi$ and the sum $d_i + w_i$, if small enough, may define $\eta$. To eliminate as many terms as possible from $\xi$, we make $I^J K_+ + 0$ by setting $d_i = r_{\max}$ and $w_i = r_{\min}$. The same reasoning applies to the components in $I^J K_-$. Making $I^J K_- = 0$ has a (positive) effect not mentioned in Table 2: 0 in (4a) is also minimized.

Regarding the components in $\eta$, consider $i \in I^J K_-$. Such a component has no influence on $\xi$. Hence, we simply want $|w_i - d_i|$ as large as possible. We achieve that by setting $d_i$ to a small value, $r_{\min}$, and $w_i$ to a large one, $r_{\max}$. Recall that $K_- = \{ i : d_i < w_i \}$.

---

4The threshold parameters $\epsilon_1$ and $\epsilon_2$ play a key role in the estimation of $I$ and $J$, and we recommend initializing them with large values and reduce them progressively at each iteration. The reason is to reduce the chance of misclassifying a component as belonging to one of these sets at an early stage.
where all the components in $\mathbf{x}$ are drawn from the standard Normal distribution $\mathcal{N}(0, 1)$, and the nonzero entries were generated uniformly at random. The values of the nonzero entries were drawn from $\mathcal{N}(0, 1)$. To illustrate the performance of Algorithm 1, we conducted experiments using synthetic data, described as follows.

**Experimental setup.** We generated a vector $\mathbf{x}^*$ of size $n = 1000$ with $s = 70$ nonzero entries, whose locations were selected uniformly at random. The values of the nonzero entries were drawn from the standard Normal distribution $\mathcal{N}(0, 1)$. The reference $\mathbf{x}$ was generated as $\mathbf{x} = \mathbf{x}^* + \mathbf{z}$, where $\mathbf{z}$ had sparsity 100 and a support that intersected the support of $\mathbf{x}^*$ in 60 locations and missed it in 40. The nonzero entries of $\mathbf{z}$ were drawn from $\mathcal{N}(0, 0.8)$. The number of measurements varied from $m = 1$ to 400 and, for each $m$, we generated 10 different matrices $A$ as in Theorem 1: $A_{ij} \sim \mathcal{N}(0, 1/m)$.

In Algorithm 1, we set $r_{\min} = 0.1$, $r_{\max} = 10$, $K = 15$ iterations, and $\ell_i$ and $\ell_j$ were initialized with 0 and 10, respectively, and decreased by 10% in each iteration. Each problem in step 2 of Algorithm 1 was solved by an ADMM solver for each subproblem of [20], we used SPGL1 [31] for each subproblem of [17]. For reference, we also compared with Mod-CS [9], which is a static algorithm, i.e., it uses no reweighting, but uses an estimate of the support of $\mathbf{x}$ as prior information. Naturally, we used the support of $\mathbf{x}$ as prior information.

**Results.** Fig. 1 shows the results of our experiments. The horizontal axis depicts the number of measurements $m$, the vertical axis the success rate over 10 different realizations of $A$. We consider that an algorithm reconstructed $\mathbf{x}^*$ successfully if the relative error of its output $\hat{x}$ was smaller than 0.1%, i.e., $\|\mathbf{x}^* - \hat{x}\|_2/\|\mathbf{x}^*\|_2 \leq 10^{-3}$.

The figure shows that Algorithm 1 had the best performance, requiring the least amount of measurements to reconstruct $\mathbf{x}^*$. The algorithm in [20] had the second best performance, followed by Mod-CS, $\ell_1$-minimization, and reweighted $\ell_1$ minimization. Note that $\ell_1$-minimization corresponds to one iteration of Algorithm 1. The plot then clearly shows that reweighting is an effective strategy to reduce the number of required measurements: in 15 iterations, the number of measurements required for reconstruction was reduced from 250 to 160, a reduction of 36%. Fig. 1 also shows a vertical line indicating the minimum theoretical value of the bound in (5), which shows that the margin for improvement is small.

**3. EXPERIMENTAL RESULTS**

To illustrate the performance of Algorithm 1, we conducted experiments using synthetic data, described as follows.

**Experimental setup.** We generated a vector $\mathbf{x}^*$ of size $n = 1000$ with $s = 70$ nonzero entries, whose locations were selected uniformly at random. The values of the nonzero entries were drawn from the standard Normal distribution $\mathcal{N}(0, 1)$. The reference $\mathbf{x}$ was generated as $\mathbf{x} = \mathbf{x}^* + \mathbf{z}$, where $\mathbf{z}$ had sparsity 100 and a support that intersected the support of $\mathbf{x}^*$ in 60 locations and missed it in 40. The nonzero entries of $\mathbf{z}$ were drawn from $\mathcal{N}(0, 0.8)$. The number of measurements varied from $m = 1$ to 400 and, for each $m$, we generated 10 different matrices $A$ as in Theorem 1: $\mathbf{A}_{ij} \sim \mathcal{N}(0, 1/m)$.

In Algorithm 1, we set $r_{\min} = 0.1$, $r_{\max} = 10$, $K = 15$ iterations, and $\ell_i$ and $\ell_j$ were initialized with 0 and 10, respectively, and decreased by 10% in each iteration. Each problem in step 2 of Algorithm 1 was solved by ADMM. We compared Algorithm 1 with the reweighted $\ell_1$-minimization scheme in [20, 21] and reweighted $\ell_1$ minimization [17]. Both algorithms ran for $K = 15$ iterations as well, and while we used the same ADMM solver for each subproblem of [20], we used SPGL1 [31] for each subproblem of [17]. For reference, we also compared with Mod-CS [9], which is a static algorithm, i.e., it uses no reweighting, but uses an estimate of the support of $\mathbf{x}^*$ as prior information. Naturally, we used the support of $\mathbf{x}$ as prior information.

**Results.** Fig. 1 shows the results of our experiments. The horizontal axis depicts the number of measurements $m$, the vertical axis the success rate over 10 different realizations of $A$. We consider that an algorithm reconstructed $\mathbf{x}^*$ successfully if the relative error of its output $\hat{x}$ was smaller than 0.1%, i.e., $\|\mathbf{x}^* - \hat{x}\|_2/\|\mathbf{x}^*\|_2 \leq 10^{-3}$.

The figure shows that Algorithm 1 had the best performance, requiring the least amount of measurements to reconstruct $\mathbf{x}^*$. The algorithm in [20] had the second best performance, followed by Mod-CS, $\ell_1$-minimization, and reweighted $\ell_1$ minimization. Note that $\ell_1$-minimization corresponds to one iteration of Algorithm 1. The plot then clearly shows that reweighting is an effective strategy to reduce the number of required measurements: in 15 iterations, the number of measurements required for reconstruction was reduced from 250 to 160, a reduction of 36%. Fig. 1 also shows a vertical line indicating the minimum theoretical value of the bound in (5), which shows that the margin for improvement is small.

**4. CONCLUSIONS**

We proposed a reweighted scheme for reference-based compressed sensing, in particular, weighted $\ell_1$-minimization. Our method differs from prior reweighting methods for either $\ell_1$-minimization or simple $\ell_1$ minimization by minimizing a sample complexity bound in each iteration. The resulting scheme is simple, intuitive, and shows excellent performance in practice. Possible research directions include understanding how the parameters of the algorithm affect its performance, and whether the sample complexity bound can be used to derive a stopping criterion.
5. REFERENCES


