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 Electronics and Informatics, Vrije Universiteit Brussel — iMinds, 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 ℓ_1-ℓ_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 ℓ_1 minimization, ℓ_1-ℓ_1 minimization, and modified CS.

Index Terms— Compressed sensing, reweighted ℓ_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., ℓ_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 [9–17], or improving reconstruction algorithms, e.g., via reweighting schemes [18–22].

In this paper, we propose a reweighted scheme for a reconstruction problem that uses as prior knowledge a reference signal. Specifically, let $x^\star \in \mathbb{R}^n$ be a sparse signal of which we have m linear measurements $y = Ax^\star$, 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 $\overline{x} \in \mathbb{R}^n$, close to $x^\star$ in the ℓ_1-norm sense, i.e., $\|x^\star - \overline{x}\|_1$ is assumed small. Using the measurements y and reference $\overline{x}$, $x^\star$ can be reconstructed via *weighted ℓ_1-ℓ_1 minimization*:

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & \left\| d \circ x \right\|_1 + \left\| w \circ (x - \overline{x}) \right\|_1 \\ \text{subject to} \quad & Ax = y \,, \end{aligned} \qquad (1)$$

where $\circ$ denotes the entrywise product between two vectors, and $d, w \in \mathbb{R}^n_+$ have nonnegative entries. Problem (1) generalizes weighted ℓ_1-norm minimization [10,23], in which w is the zero vector, and also ℓ_1-ℓ_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 $x^\star$ and is therefore unknown. To address this uncertainty, we consider a reweighting scheme: starting from arbitrary d^1 and w^1, we create a sequence $\{x^k\}_{k=1}^K$ such that, for $k = 1, \dots, K$,

$$\begin{aligned} x^k \in \quad \underset{x}{\arg\min} \quad & \left\| d^k \circ x \right\|_1 + \left\| w^k \circ (x - \overline{x}) \right\|_1 \\ \text{s.t.} \quad & Ax = y \,, \end{aligned} \qquad (2)$$

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 $x^\star$ 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 $x^\star$; the bound depends on $x^\star$ 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 $x^\star$ by its current best estimate, x^{k-1}.

Our result establishing the bound in step 1) says that $O(\frac{\zeta}{\eta^2} \log n)$ measurements suffice to reconstruct $x^\star$ via (1), where ζ and η are functions of the weights d and w. We show that if d and w are chosen properly, ζ/η^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 ℓ_1 minimization [23], or simple ℓ_1-ℓ_1 minimization [13], which require $O(c \log n)$ measurements with c having the same order of magnitude as the sparsity of $x^\star$. 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 ζ/η^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 ℓ_1-minimization [18], and static schemes that use prior information, such as ℓ_1-ℓ_1 minimization [13] and modified-CS [9].

1.2. Related Work

Reweighting has been applied in least squares problems as far back as [24,25]. For sparse reconstruction problems, [18] proposed a sim-

ple algorithm known as *reweighted ℓ_1 minimization*: each weight d_i is updated at iteration k as $d_i^{k+1} = 1/(|x_i^k| + \epsilon)$, where $\epsilon > 0$ and x^k is a solution of weighted ℓ_1 minimization with weights d^k, i.e., (2) with $w^k = 0_n$ (the zero vector). That algorithm and variations are analyzed in [26, 27]. Other reweighting schemes for sparse reconstruction include [19, 28], which solve simpler problems per iteration, namely least squares problems, and are therefore computationally more efficient. Regarding sparse reconstruction using prior information, [21, 22] proposed a reweighting algorithm for a slight variation of problem (1) in the context of MRI reconstruction. There, the weights are updated as $d_i^{k+1} = 1/(|x_i^k| + 1)$ and $w_i^{k+1} = 1/(|x_i^k - \overline{x}_i| + 1)$, and the resulting scheme is shown to significantly improve MRI reconstruction.

2. REWEIGHTED ℓ_1-ℓ_1 MINIMIZATION

2.1. Step 1: Bound on the number of measurements

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

$$I = \{i : x_i^\star \neq 0\} \quad J = \{i : x_i^\star \neq \overline{x}_i\} \quad K = \{i : d_i \neq w_i\}$$
$$I_+ = \{i : x_i^\star > 0\} \quad J_+ = \{i : x_i^\star > \overline{x}_i\} \quad K_+ = \{i : d_i > w_i\}$$
$$I_- = \{i : x_i^\star < 0\} \quad J_- = \{i : x_i^\star < \overline{x}_i\} \quad K_- = \{i : d_i < w_i\}.$$

In words, I, J, and K are the supports of $x^\star$, $x^\star - \overline{x}$, and $d - w$; the subscript $+$ (resp. $-$) restricts these supports to their positive (resp. negative) components. We represent set intersections as products: e.g., IJ denotes $I \cap J$. Using the above sets, we define [13]

$$\overline{h} := |I_+ J_+| + |I_- J_-| \qquad h := |I_+ J_-| + |I_- J_+|, \quad (3)$$

which are independent from d and w. As shown in [13], these parameters measure the quality of $\overline{x}$. In particular, ℓ_1-ℓ_1 minimization, i.e., (1) with $d = w = 1_n$, requires $O(\overline{h} \log n)$ measurements to reconstruct $x^\star$. To present our result, we need to define three additional parameters, all of which depend on d and w:

$$\theta := |IJ^c K_+| + |I^c J K_-| \tag{4a}$$

$$\zeta := \sum_{i \in IJ} \left[d_i \, \text{sg}(x_i^\star) + w_i \, \text{sg}(x_i^\star - \overline{x}_i) \right]^2 + \sum_{i \in Q_+} (d_i - w_i)^2 \tag{4b}$$

$$\eta := \min \left\{ \min_{i \in Q_-} |w_i - d_i| , \ \min_{i \in Q \cup I^c J^c} d_i + w_i \right\}, \tag{4c}$$

where $\text{sg}(\cdot)$ denotes the sign of a number, $Q_+ := IJ^c K_+ \cup I^c J K_-$, $Q_- := IJ^c K_- \cup I^c J K_+$, and $Q := Q_+ \cup Q_-$. The role played by $\overline{h}$ in ℓ_1-ℓ_1 minimization will now be played by the ratio ζ/η^2 in weighted ℓ_1-ℓ_1 minimization. In contrast with $\overline{h}$, however, ζ/η^2 can be manipulated because ζ and η depend on d and w.

Theorem 1. *Let $x^\star \in \mathbb{R}^n$ be the vector to reconstruct and $\overline{x} \in \mathbb{R}^n$ the prior information. Let $y = Ax^\star$, where the entries of $A \in \mathbb{R}^{m \times n}$ are drawn i.i.d. from the Gaussian distribution with zero mean and variance $1/m$. Assume d and w have positive entries, $\zeta > 0$, $\eta > 0$, and also that there exist two (different) indices i and j such that $0 \neq x_i^\star \neq \overline{x}_i$ and $x_j^\star = \overline{x}_j = 0$.[1] If*

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

---

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

Table 1. Sample complexity of alternative reconstruction schemes.

Prob.	Objective Function	Bound	Ref.
w-ℓ_1-ℓ_1	$\|d \circ x\|_1 + \|w \circ (x - \overline{x})\|_1$	$O\left(2\frac{\zeta}{\eta^2} \log n\right)$	here
ℓ_1-ℓ_1	$\|x\|_1 + \|x - \overline{x}\|_1$	$O\left(2\overline{h} \log n\right)$	[13]
Mod-CS	$\sum_{i \in \tilde{I}^c} |x_i|$	$O\left(2\frac{b}{C^2} \log n\right)$	[23]
BP	$\|x\|_1$	$O\left(2s \log n\right)$	[7]

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

This theorem, whose proof[2] uses the concept of Gaussian width [7, 29], generalizes Theorem 1 in [13], which established a similar bound for the particular case $d = w = 1_n$.

We mentioned before that ζ/η^2 can be made arbitrarily small.[3] To see why, suppose d and w were selected so that $Q_+ = \emptyset$. Then, according to (4b)-(4c), the set over which ζ is defined, IJ, does not intersect any of the sets over which η is defined, i.e., $IJ \cap Q_- = IJ \cap (Q_- \cup I^c J^c) = \emptyset$. In other words, the set of components of d and w that contribute to ζ are independent from the components that contribute to η. Therefore, ζ/η^2 can be arbitrarily small. As shown next, this is not the case of alternative reconstruction problems.

Comparison with other reconstruction problems. Table 1 compares our bound for Weighted ℓ_1-ℓ_1 minimization (w-ℓ_1-ℓ_1) with bounds obtained using similar tools for other methods: ℓ_1-ℓ_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, $\tilde{I}$ is an estimate of the support I of $x^\star$ and is used as prior information. Prior information in ℓ_1-ℓ_1 is, as in our case, a reference signal $\overline{x}$. Only BP uses no prior information. Table 1 also shows where the displayed bounds were computed. In the bound for Mod-CS, $0 < C < 1$, and b is the sum of false negatives and false positives in the estimation of I, i.e., $b := |I \cap \tilde{I}^c| + |I^c \cap \tilde{I}|$. Thus, for ℓ_1-ℓ_1 and Mod-CS, $\overline{h}$ and b measure the quality of the prior information: the better the quality, the smaller $\overline{h}$ and b. This means the number of measurements required by ℓ_1-ℓ_1 and Mod-CS is determined by the quality of the prior information. For w-ℓ_1-ℓ_1 minimization, however, the ratio ζ/η^2 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 ζ/η^2 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, $\epsilon_I, \epsilon_J > 0$, which are used in the estimation of the sets I and J, and the number of iterations K. At iteration k, the algorithm obtains an estimate x^k of $x^\star$ by solving weighted ℓ_1-ℓ_1 minimization with weights d^k and w^k (step 2). Note that because d and w are initialized as 1_n, the first iteration is simply

---

[2] http://www.ee.ucl.ac.uk/~jmota/reL1L1.pdf
[3] Note that the first term of (5) is dominant for sparse signals. In particular, $(7/5)(\overline{h}+h)+\theta \leq (17/5)s+\overline{s}$, where s (resp. $\overline{s}$) is the sparsity of $x^\star$ (resp. $\overline{x}$). This follows from $\overline{h}+h = |IJ| \leq |I| = s$ and $\theta \leq |I|+|J| \leq 2s+\overline{s}$.

Algorithm 1 Reweighted ℓ_1-ℓ_1 minimization

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

Parameters: $0 < r_{\min} < r_{\max}$, $\epsilon_I, \epsilon_J > 0$, K (# iterations)

Initialization: $d^1 = w^1 = 1_n$, $k = 1$

1: **for** $k = 1, \ldots, K$ **do**
2:     Obtain x^k by solving
$$\underset{x}{\text{minimize}} \quad \left\| d^k \circ x \right\|_1 + \left\| w^k \circ \left(x - \overline{x} \right) \right\|_1$$
$$\text{subject to} \quad Ax = y$$
3:     Set $I^k = \left\{ i : |x_i^k| > \epsilon_I \right\}$ and $J^k = \left\{ i : |x_i^k - \overline{x}_i| > \epsilon_J \right\}$
4:     **for** $i = 1, \ldots, n$ **do**
5:         **if** $i \in I^k J^k$ **then** $d_i^{k+1} = w_i^{k+1} = r_{\min}$
6:         **else if** $i \in I^k J^{c,k}$ **then** $d_i^{k+1} = r_{\min},\ w_i^{k+1} = r_{\max}$
7:         **else if** $i \in I^{c,k} J^k$ **then** $d_i^{k+1} = r_{\max},\ w_i^{k+1} = r_{\min}$
8:         **else if** $i \in I^{c,k} J^{c,k}$ **then** $d_i^{k+1} = w_i^{k+1} = r_{\max}$
9:         **end if**
10:     **end for**
11: **end for**

ℓ_1-ℓ_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 $x^\star$; so, we estimate them by using our current best guess: x^k.[4] 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, $i \in I^k J^k$, i.e., it is estimated that $\overline{x}_i \neq x_i^\star \neq 0$ (step 5). The algorithm sets the corresponding weights d_i and w_i to a small value $r_{\min}$, which means that $x_i^\star$ will be estimated solely from the measurements $y = Ax^\star$. If, on the other hand, $i \in I^k J^{c,k}$, i.e., it is estimated that $\overline{x}_i = x_i^\star \neq 0$, the algorithm sets d_i to a small value, to avoid penalizing large values for x_i^k, and sets w_i to a large value, penalizing deviations from an apparently accurate component of $\overline{x}$. Similarly, if $i \in I^{c,k} J^k$, i.e., it is estimated that $\overline{x}_i \neq x_i^\star = 0$, w_i is set to a small value, since $\overline{x}_i$ seems to be inaccurate, and d_i is set to a large value, since $x_i^\star$ is likely to be zero. Finally, if $i \in I^{c,k} J^{c,k}$, i.e., it is estimated that $\overline{x}_i = x_i^\star = 0$, both d_i and w_i are set to large values since, very likely, $x_i^\star$ is zero. These updates, beyond intuitive, lead to a reduction of the number of required measurements, as shown next.

Corollary 2. *Let $x^\star, \overline{x} \in \mathbb{R}^n$ 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*

$$m \geq \left(\frac{r_{\min}}{r_{\max} - r_{\min}} \right)^2 \overline{h} \log \left(\frac{n}{\overline{h} + h} \right) + \frac{7}{5} \left(\overline{h} + h \right) + 1, \quad (6)$$

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

---
[4]The threshold parameters ϵ_I and ϵ_J play a key role in the estimation of I and J, and we recommend initializing them with large values (w.r.t. the magnitudes of $x^\star$ and $\overline{x}$) 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.

Proof. The weights used at iteration K are computed at iteration $K - 1$. Hence, the last instance of w-ℓ_1-ℓ_1 in step 2 is solved with

$$d_i = r_{\min}, \ w_i = r_{\min}, \qquad \text{for all } i \in IJ \qquad (7a)$$
$$d_i = r_{\min}, \ w_i = r_{\max}, \qquad \text{for all } i \in IJ^c \qquad (7b)$$
$$d_i = r_{\max}, \ w_i = r_{\min}, \qquad \text{for all } i \in I^c J \qquad (7c)$$
$$d_i = r_{\max}, \ w_i = r_{\max}, \qquad \text{for all } i \in I^c J^c, \qquad (7d)$$

Note that (7b) implies $IJ^c K_+ = \emptyset$ and (7c) implies $I^c J K_- = \emptyset$, that is, $Q_+ = \emptyset$. This means the parameter ζ in (4b) equals $r_{\min}^2 \sum_{i \in IJ} (\text{sg}(x_i^\star) + \text{sg}(x_i^\star - \overline{x}_i))^2 = r_{\min}^2 \overline{h}$, where we used (3). We also have $\theta = 0$ [cf. (4a)]. According to (4c) and (7b)-(7d), η equals $r_{\max} - r_{\min}$ if $Q_- \neq \emptyset$, and $2r_{\max}$ otherwise (note that, by assumption, $I^c J^c \neq \emptyset$; see footnote 1). Then, (5) becomes

$$m \geq \left(\frac{r_{\min}}{2 r_{\max}} \right)^2 \overline{h} \log \left(\frac{n}{\overline{h} + h} \right) + \frac{7}{5} \left(\overline{h} + h \right) + 1 \qquad (8)$$

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 $(r_{\min}/r_{\max})^2 \overline{h} \log n$. Thus, under the corollary's assumptions, setting $r_{\max} \simeq \sqrt{\log n}\, r_{\min}$ makes the number of 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 ζ/η^2, subject to $\zeta > 0$ 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 ζ/η^2 formally, i.e., with an optimization algorithm, we do it heuristically. In particular, we allow only two values for the weights: $r_{\min}$ and $r_{\max}$.

To aid our derivation, Table 2 shows the sets involved in the definitions of ζ and η, and describes how the respective components of d and w should relate to minimize the ζ/η^2. Consider, for example, a component $i \in IJ$; it contributes $g_i(d_i, w_i) := (d_i \, \text{sg}(x_i^\star) + w_i \, \text{sg}(x_i^\star - \overline{x}_i))^2$ to ζ and has no influence on η. There are two scenarios: either $i \in I_+ J_+ \cup I_- J_-$ or $i \in I_+ J_- \cup I_- J_+$. In the former, we have $\text{sg}(x_i^\star) = \text{sg}(x_i^\star - \overline{x}_i)$, 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 $d, w > 0$; rather, we set these components to a small value, $r_{\min} > 0$. When $i \in I_+ J_- \cup I_- J_+$, $g_i(d_i, w_i)$ 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 IJ are treated similarly; any other choice for a common value would also work. Consider now a component $i \in IJ^c K_+$: it contributes with $(d_i - w_i)^2$ to ζ and the sum $d_i + w_i$, if small enough, may define η. To eliminate as many terms as possible from ζ, we make $IJ^c K_+$ empty by setting $d_i = r_{\min}$ and $w_i = r_{\max}$. The same reasoning applies to the components $i \in I^c J K_-$. Making $IJ^c K_+ = I^c J K_- = \emptyset$ has a (positive) side effect not mentioned in Table 2: θ in (4a) is also minimized.

Regarding the components in η, consider $i \in IJ^c K_-$. Such a component has no influence on ζ. 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 $K_- = \{i : d_i < w_i\}$; therefore, if we had switched the roles of d_i and w_i, we would have

Table 2. Derivation of the scheme. The third column shows the reasoning for minimizing ζ/η^2, the fourth the action we select.

Parameter	Set	Reasoning to minimize ζ/η^2	Action at iteration k
ζ	IJ	If $i \in I_+J_+ \cup I_-J_-$, set d_i and w_i as small as possible	$d_i^{k+1} = w_i^{k+1} = r_{\min}$
		If $i \in I_+J_- \cup I_-J_+$, set $d_i = w_i$	$d_i^{k+1} = w_i^{k+1} = r_{\min}$
	IJ^cK_+	Set $d_i \leq w_i$ to make $IJ^cK_+ = \emptyset$	$d_i^{k+1} = r_{\min},\ w_i^{k+1} = r_{\max}$
	I^cJK_-	Set $d_i \geq w_i$ to make $I^cJK_- = \emptyset$	$d_i^{k+1} = r_{\max},\ w_i^{k+1} = r_{\min}$
η	IJ^cK_-	Set d_i small and w_i large to make $|w_i - d_i|$ large	$d_i^{k+1} = r_{\min},\ w_i^{k+1} = r_{\max}$
	I^cJK_+	Set d_i large and w_i small to make $|w_i - d_i|$ large	$d_i^{k+1} = r_{\max},\ w_i^{k+1} = r_{\min}$
	IJ^cK	Set d_i large, w_i large, or both, to make $d_i + w_i$ large	$d_i^{k+1} = r_{\min},\ w_i^{k+1} = r_{\max}$
	I^cJK	Set d_i large, w_i large, or both, to make $d_i + w_i$ large	$d_i^{k+1} = r_{\max},\ w_i^{k+1} = r_{\min}$
	I^cJ^c	Set d_i large, w_i large, or both, to make $d_i + w_i$ large	$d_i^{k+1} = w_i^{k+1} = r_{\max}$

instead made IJ^cK_- empty. The same reasoning applies to the components in I^cJK_+. Now note that because $IJ^cK = IJ^cK_+ \cup IJ^cK_-$ and $I^cJK = I^cJK_- \cup I^cJK_+$, the action for the components in IJ^cK and I^cJK has already been determined. Namely, the 2nd and 4th lines of the table defined $d_i = r_{\min}$ and $w_i = r_{\max}$ for the components in IJ^cK_+ and IJ^cK_-, thus defining the action for all the components in IJ^cK. The same applies to the components in I^cJK (3rd and 5th lines). These actions do not conflict with our goal of making η as large as possible; rather, they reinforce it, as they align with the reasoning described in the table. Finally, the components $i \in I^cJ^c$ only influence η and, therefore, we set the respective d_i and w_i as large as possible: $d_i = w_i = r_{\max}$.

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 $x^\star$ 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 $\overline{x}$ was generated as $\overline{x} = x^\star + z$, where z had sparsity 100 and a support that intersected the support of $x^\star$ in 60 locations and missed it in 40. The nonzero entries of z were drawn from $\mathcal{N}(0, 0.8)$. The number of measurements varied m from 1 to 400 and, for each m, we generated 10 different matrices A as in Theorem 1: $A_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1/m)$.

In Algorithm 1, we set $r_{\min} = 0.1$, $r_{\max} = 10$, $K = 15$ iterations, and ϵ_I and ϵ_J were initialized with 0.5 and decreased by 10% in each iteration. Each problem in step 2 of Algorithm 1 was solved with ADMM [31]. We compared Algorithm 1 with the reweighted ℓ_1-ℓ_1 scheme in [21, 22] and reweighted ℓ_1 minimization [18]. Both algorithms ran for $K = 15$ iterations as well, and while we used the same ADMM solver for each subproblem of [21], we used SPGL1 [32] for each subproblem of [18]. All these algorithms have roughly the same computational complexity. For reference, we also compared with Mod-CS [9], a static algorithm (i.e., with no reweighting) that uses an estimate of the support of $x^\star$ as prior information. We used $\text{supp}(\overline{x})$ as such 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 $x^\star$ successfully if the relative error of its output $\widehat{x}$ was smaller than 0.1%, i.e., $\|\widehat{x} - x^\star\|_2 / \|x^\star\|_2 \leq 10^{-3}$.
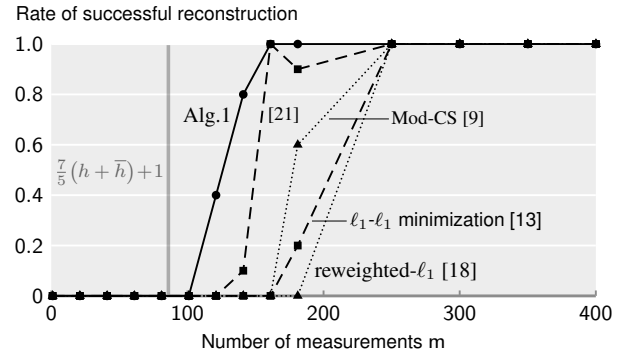
Fig. 1. Rate of reconstruction of Algorithm 1 and prior schemes. The vertical line shows the minimal theoretical value of (5).

The figure shows that Algorithm 1 had the best performance, requiring the least amount of measurements to reconstruct $x^\star$. The algorithm in [21] had the second best performance, followed by Mod-CS, ℓ_1-ℓ_1 minimization, and reweighted ℓ_1 minimization. Note that ℓ_1-ℓ_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), 85, obtained by ignoring the first term and considering $\theta = 0$. Since Algorithm 1 started reconstructing $x^\star$ using 120 measurements, this shows that the margin for improvement is small.

4. CONCLUSIONS

We proposed a reweighted scheme for reference-based compressed sensing, in particular, weighted ℓ_1-ℓ_1 minimization. Our method differs from prior reweighting methods for either ℓ_1-ℓ_1 minimization or simple ℓ_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

[1] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Info. Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[2] D. Donoho, "Compressed sensing," *IEEE Trans. Info. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[3] Y. C. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications*, Cambridge University Press, 2012.

[4] R. Tibshirani and M. Saunders, "Sparsity and smoothness via the fused lasso," *J. R. Statist. Soc. B*, vol. 67, no. 1, pp. 91–108, 2005.

[5] R. Baraniuk, V. Cevher, M. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Trans. Info. Theory*, vol. 56, no. 4, pp. 1982–2001, 2010.

[6] M. Duarte and Y. C. Eldar, "Structured compressed sensing: From theory to applications," *IEEE Trans. Sig. Proc.*, vol. 59, no. 9, pp. 4053–4085, 2011.

[7] V. Chandrasekaran, B. Recht, P. Parrilo, and A. Willsky, "The convex geometry of linear inverse problems," *Found. Computational Mathematics*, vol. 12, pp. 805–849, 2012.

[8] S. Oymak, A. Jalali, M. Fazel, Y. C. Eldar, and B. Hassibi, "Simultaneously structured models with application to sparse and low-rank matrices," *IEEE Trans. Info. Theory*, vol. 61, no. 5, pp. 2886–2908, 2015.

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

[10] M. Khajehnejad, W. Xu, A. Avestimehr, and B. Hassibi, "Analyzing weighted ℓ_1 minimization for sparse recovery with nonuniform sparse models," *IEEE Trans. Sig. Proc.*, vol. 59, no. 5, pp. 1985–2001, 2011.

[11] J. Scarlett, J. Evans, and S. Dey, "Compressed sensing with prior information: Information-theoretic limits and practical decoders," *IEEE Trans. Sig. Proc.*, vol. 61, no. 2, 2013.

[12] F. Renna, L. Wang, X. Yuan, J. Yang, G. Reeves, R. Calderbank, L. Carin, and M. R. D. Rodrigues, "Classification and reconstruction of high-dimensional signals from low-dimensional noisy features in the presence of side information," http://arxiv.org/abs/1412.0614, 2014.

[13] J. F. C. Mota, N. Deligiannis, and M. R. D. Rodrigues, "Compressed sensing with prior information: Optimal strategies, geometry, and bounds," http://arxiv.org/abs/1408.5250, 2014.

[14] J. F. C. Mota, N. Deligiannis, and M. R. D. Rodrigues, "Compressed sensing with side information: Geometrical interpretation and performance bounds," in *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 2014.

[15] J. F. C. Mota, N. Deligiannis, A. C. Sankaranarayanan, V. Cevher, and M. R. D. Rodrigues, "Adaptive-rate reconstruction of time-varying signals with application in compressive foreground extraction," accepted in IEEE Trans. Sig. Proc., http://arxiv.org/abs/1503.03231, 2016.

[16] J. F. C. Mota, N. Deligiannis, A. C. Sankaranarayanan, V. Cevher, and M. R. D. Rodrigues, "Dynamic sparse state estimation using ℓ_1-ℓ_1 minimization: Adaptive-rate measurement bounds, algorithms, and applications," in *IEEE Intern. Conf. Acoustics, Speech, and Sig. Processing (ICASSP)*, 2015.

[17] E. Zimos, J. F. C. Mota, M. R. D. Rodrigues, and N. Deligiannis, "Bayesian compressed sensing with heterogeneous side information," in *IEEE Data Compression Conf. (DCC)*, 2016.

[18] E. J. Candès, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *J Fourier Anal Appl*, vol. 14, pp. 877–905, 2008.

[19] R. Chartrand and W. Yin, "Iteratively reweighted algorithms for compressive sensing," in *IEEE Intern. Conf. Acoustics, Speech, and Sig. Processing (ICASSP)*, 2008, pp. 3869–3872.

[20] M. Asif and J. Romberg, "Dynamic updating for ℓ_1 minimization," *IEEE J. Selected Topics in Sig. Proc.*, vol. 4, no. 2, pp. 421–434, 2010.

[21] L. Weizman, Y. C. Eldar, and D. Ben-Bashat, "Compressed sensing for longitudinal MRI: An adaptive-weighted approach," *Medical Physics*, vol. 42, no. 9, pp. 5195–5207, 2015.

[22] L. Weizman, Y. C. Eldar, and D. Ben-Bashat, "Fast reference-based MRI," http://arxiv.org/abs/1508.02775, 2015.

[23] H. Mansour and R. Saab, "Recovery analysis for weighted ℓ_1-minimization using a null space property," Appl. Comput. Harmon. Anal., 2015.

[24] C. L. Lawson, *Contributions to the theory of linear least maximum approximations*, Ph.D. thesis, UCLA, 1961.

[25] A. E. Beaton and J. W. Tukey, "The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data," *Technometrics*, vol. 16, pp. 147–185, 1974.

[26] D. Needell, "Noisy signal recovery via iterative reweighted ℓ_1-minimization," in *Asilomar Conf. Signals, Systems, and Computers*, 2009, pp. 113–117.

[27] W. Xu, M. A. Khajehnejad, A. S. Avestimehr, and B. Hassibi, "Breaking through the thresholds: an analysis for iterative reweighted ℓ_1 minimization via the Grassmann angle framework," in *IEEE Intern. Conf. Acoustics, Speech, and Sig. Processing (ICASSP)*, 2010, pp. 5498–5501.

[28] I. Daubechies, R. DeVore, M. Fornasier, and C. S. Güntürk, "Iteratively reweighted least squares minimization for sparse recovery," *Communications on Pure and Applied Mathematics*, vol. LXIII, pp. 1–38, 2010.

[29] Y. Gordon, "On Milman's inequality and random subspaces which escape through a mesh in $\mathbb{R}^n$," in *Geometric Aspects of Functional Analysis, Israel Seminar 1986-1987. Lecture Notes in Mathematics*, 1988, pp. 84–106.

[30] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comp.*, vol. 20, no. 1, pp. 33–61, 1998.

[31] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[32] E. Berg and M. Friedlander, "Probing the Pareto frontier for basis pursuit solutions," *SIAM J. Sci. Comput.*, vol. 31, no. 2, pp. 890–912, 2008.