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

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 [9–17], 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 $x^* \in \mathbb{R}^n$ 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 $\tau \in \mathbb{R}^n$, close to $x^*$ in the $\ell_1$-norm sense, i.e., $\|x^* - \tau\|_1$ is assumed small. Using the measurements $y$ and reference $\tau$, $x^*$ can be reconstructed via weighted $\ell_1$-$\ell_1$ minimization:

$$
\begin{align*}
\text{minimize } \quad & \|d \circ x\|_1 + \|w \circ (x - \tau)\|_1 \\
\text{subject to } \quad & Ax = y,
\end{align*}
$$

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 [10,23], 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 $k = 1, \ldots, K$,

$$
x^k \in \arg \min \|d^k \circ x\|_1 + \|w^k \circ (x - \tau)\|_1 \\
\text{s.t. } Ax = y,
$$

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 $O(\frac{m}{\log n})$ 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 [23], 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 [18], 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 [24,25]. For sparse reconstruction problems, [18] proposed a sim-
ple algorithm known as reweighted $\ell_1$ minimization: each weight $d_i$ is updated at iteration $k$ as $d_i^{k+1} = 1/(|x_i^k| + \epsilon)$, where $\epsilon > 0$ and $x^k$ is a solution of weighted $\ell_1$ 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 $w_i^{k+1} = 1/(|x_i^k| - |x_i^k| + 1)$, and the resulting scheme is shown to significantly improve MRI reconstruction.

2. REWEIGHTED $\ell_1$-$\ell_1$ 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 $x$, $d$, and $w$ interact. To capture those sets, we define the sets

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

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

$I_- = \{i : x_i^* < 0\}, \quad J_- = \{i : x_i^* < |x_i^*|\}, \quad K_- = \{i : d_i < 0\}$

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

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

which are independent from $d$ and $w$. As shown in [13], these parameters measure the quality of $\mathcal{F}$. In particular, $\ell_1$-$\ell_1$ minimization, i.e., (1) with $d = w = 1_n$, requires $O(\mathcal{F} \log n)$ measurements to reconstruct $x^*$. To present our result, we need to define three additional parameters, all of which depend on $d$ and $w$:

$\theta := |I^c J K^c| + |I^c J K^c|$, $\zeta := \sum_{i \in J} d_i s_g(x_i^*) + w_i s_g(x_i^* - |x_i^*|)$ + $\sum_{i \in Q_\mathcal{F}} (d_i - w_i)^2$, $\eta := \min \{ \min_{i \in Q_\mathcal{F}} |w_i - d_i|, \min_{i \in Q_\mathcal{F} \cup J^c} d_i + w_i \}$.

Theorem 1. Let $x^* \in \mathbb{R}^n$ be the vector to reconstruct and $\mathcal{F} \in \mathbb{R}^n$ the prior information. Let $y = Ax^*$, 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 $0 \neq x_i^* \neq |x_i^*|, x_j^* = |x_j^*|$. If

$m \geq 2\frac{\zeta}{\eta^2} \left(\frac{\theta}{\zeta + \mathcal{F}}\right)^2 + \frac{7}{5} (\mathcal{F} + h) + \theta + 1$, (5)

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

This theorem, whose proof 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 $\zeta/\eta^2$ can be made arbitrarily small. To see why, suppose $d$ and $w$ were selected so that $Q_\mathcal{F} = \emptyset$. Then, according to (4b)-(4c), the set over which $\zeta$ is defined, $I^c J$, does not intersect any of the sets over which $\eta$ is defined, i.e., $I \cap Q_\mathcal{F} = I \cap (Q_\mathcal{F} \cup I^c J) = \emptyset$. In other words, the sets of components of $d$ and $w$ that contribute to $\zeta$ are independent from the components that contribute to $\eta$. Therefore, $\zeta/\eta^2$ 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 $\ell_1$-$\ell_1$ 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, $\tilde{I}$ is an estimate of the support of $x^*$ and is used as prior information. Prior information in $\ell_1$-$\ell_1$ is, as in our case, a reference signal $\mathcal{F}$. 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 $\ell_1$-$\ell_1$ and Mod-CS, $\mathcal{F}$ and $b$ measure the quality of the prior information: the better the quality, the smaller $\mathcal{F}$ and $b$. 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 $\zeta/\eta^2$ 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 $\zeta/\eta^2$ 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 $\tau_{\min}$ and $\tau_{\max}$ which, as we will see, determine the amount by which the bound in (5) is minimized, $\epsilon_1, \epsilon_2 > 0$, 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 $x^*$ by solving weighted $\ell_1$-$\ell_1$ 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

\[ \text{http://www.ee.ucl.ac.uk/~jmota/reL1L1.pdf} \]

\[ \text{Note that the first term of (5) is dominant for sparse signals. In particular, } (7/5)(\mathcal{F} + h) + \theta \leq (17/5)s^* + \mathcal{F}, \text{ where } s^* (\text{resp. } \mathcal{F}) \text{ is the sparsity of } x^* \text{ (resp. } \mathcal{F}). \]
Algorithm 1 Reweighted $\ell_1$-$\ell_1$ minimization

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

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

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

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

Proof. The weights used at iteration $k$ are computed at iteration $K - 1$. Hence, the last instance of $w^1 \cdot \ell_1 - \ell_1$ in step 2 is solved with
\begin{align*}
& d_i = r_{\min}, \quad w_i = r_{\min}, \quad \text{for all } i \in IJ \\
& d_i = r_{\max}, \quad w_i = r_{\max}, \quad \text{for all } i \in I^cJ \\
& d_i = r_{\max}, \quad w_i = r_{\min}, \quad \text{for all } i \in I^cI \\
& d_i = r_{\min}, \quad w_i = r_{\max}, \quad \text{for all } i \in IJ^c
\end{align*}

Note that (7b) implies $I^J \cup K_+ = \emptyset$ and (7c) implies $I^J \cup K_- = \emptyset$, that is, $Q_+ = \emptyset$. This means the parameter $\zeta$ in (4b) equals $r_{\min} - r_{\max}$ if $Q_+ \neq \emptyset$, and $2r_{\max}$ otherwise (note that, by assumption, $I^J \neq \emptyset$; see footnote 1). Then, (5) becomes
\begin{align*}
m \geq \left(\frac{r_{\min}}{2r_{\max}}\right)^2 2\pi \log \left(\frac{n}{\Lambda + h}\right) + \frac{7}{5}(\Lambda + h) + 1
\end{align*}

where $Q_+ \neq \emptyset$, 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.

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 $(r_{\min}/r_{\max})^2 2\pi \log n$. Thus, under the corollary’s assumptions, setting $r_{\max} = \sqrt{\log n} r_{\min}$ 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 $\zeta/\eta^2$, 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 $\zeta/\eta^2$ 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 $\zeta$ and $\eta$, and describes how the respective components of $d$ and $w$ should relate to minimizing the $\zeta/\eta^2$. Consider, for example, a component $i \in IJ$; it contributes $g_i(d_i, w_i) := (d_i \Delta g_i(x_i^*) + w_i \Delta g_i(x_i^* - \pi_i))^2$ to $\zeta$ and has no influence on $\eta$. There are two scenarios: either $i \in I_J \cup I_1 \cup J_1$, and $i \notin J_+ \cup J_-$. In the former, we have $\Delta g_i(x_i^*) = \Delta g_i(x_i^* - \pi_i)$, 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 $i \in J_+ \cup J_- \cup J_{1+} \cup J_{1-}$, we have 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 $i \in I^J K_+$; it contributes with $(d_i - w_i)^2$ to $\zeta$ and the sum $d_i + w_i$, if small enough, may define $\eta$. To eliminate as many terms as possible from $\zeta$, we make $I^J K_+$ empty by setting $d_i = r_{\max}$ and $w_i = r_{\min}$. The same reasoning applies to the components $i \in I^J K_-$. Making $I^J K_+ = I^J K_- = \emptyset$ has a (positive) side effect not mentioned in Table 2: $\theta$ in (4a) is also minimized.

Regarding the components in $\eta$, consider $i \in I^J K_-$. Such a component has no influence on $\zeta$. 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 discarded $i$.
instead made $I^*J^cK_-$ empty. The same reasoning applies to the components in $I^*J^cK_+$. Now note that because $I^*J^cK^c = I^*J^cK_+ \cup I^*J^cK_-$ and $I^*J^cK = I^*J^cK_+ \cup I^*J^cK_-$, the action for the components in $I^*J^cK$ and $I^*J^cK^c$ 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 $I^*J^cK_+$ and $I^*J^cK_-$, thus defining the action for all the components in $I^*J^cK$. The same applies to the components in $I^*J$ (3rd and 5th lines). These actions do not conflict with our goal of making $\eta$ as large as possible; rather, they reinforce it, as they align with the reasoning described in the table. Finally, the components $i \in I^*J^c$ only influence $\eta$ 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 $x^*$ 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 $\pi$ was generated as $\pi = x^* + z$, where $z$ had sparsity 100 and a support that intersected the support of $x^*$ 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 from $m = 1$ to 400 and, for each $m$, we generated 10 different matrices $A$ as in Theorem 1: $A_j \sim \mathcal{N}(0, 1/m)$.

In Algorithm 1, we set $r_{\min} = 0.1$, $r_{\max} = 10$, $K = 15$ iterations, and $\epsilon_2$ and $\epsilon_3$ 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 $\ell_1-\ell_1$ scheme in [21, 22] and reweighted $\ell_1$ 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 $x^*$ as prior information. We used supp$(\pi)$ 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 $x^*$ successfully if the relative error of its output $\hat{x}$ was smaller than 0.1%, i.e., $\|x^* - \hat{x}\|_2/\|x^*\|_2 \leq 10^{-3}$.

![Rate of successful reconstruction](image_url)

**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 $x^*$. The algorithm in [21] had the second best performance, followed by Mod-CS, $\ell_1-\ell_1$ 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), 85, obtained by ignoring the first term and considering $\theta = 0$. Since Algorithm 1 started reconstructing $x^*$ 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 $\ell_1-\ell_1$ minimization. Our method differs from prior reweighting methods for either $\ell_1-\ell_1$ 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


