A SIMPLER APPROACH TO WEIGHTED $\ell_1$ MINIMIZATION

Anilesh K. Krishnaswamy
Indian Institute of Technology Madras
Chennai, India – 600036

Samet Oymak and Babak Hassibi
California Institute of Technology
Pasadena, CA – 91125

ABSTRACT
In this paper, we analyze the performance of weighted $\ell_1$ minimization over a non-uniform sparse signal model by extending the “Gaussian width” analysis proposed in [1]. Our results are consistent with those of [7] which are currently the best known ones. However, our methods are less computationally intensive and can be easily extended to signals which have more than two sparsity classes. Finally, we also provide a heuristic for estimating the optimal weights, building on a more general model presented in [11]. Our results reinforce the fact that weighted $\ell_1$ minimization is substantially better than regular $\ell_1$ minimization and provide an easy way to calculate the optimal weights.

Index Terms—weighted $\ell_1$ minimization, compressed sensing, Gaussian measurements, recovery threshold, Gaussian width

1. INTRODUCTION
In recent years, sparse recovery problems have been the subject of great interest ([4,6,10]) due to their importance in various applications. It is now well known that there are algorithms which, under certain conditions, can provably recover the underlying sparse solution of a system. $\ell_1$ minimization is arguably the most popular among these as the $\ell_1$ norm is the tightest convex surrogate of the sparsity function (“$\ell_0$” norm). The most common form of the $\ell_1$ minimization problem is as follows:

$$
\min ||x||_1 \text{ subject to } Ax = y
$$

where $A \in \mathbb{R}^{m \times n}$ is the measurement matrix, $s \in \mathbb{R}^n$ is the $k$-sparse solution, $y = As \in \mathbb{R}^m$ is the corresponding measurement and we want the solution $x^*$ of (1) to be equal to $s$.

Clearly, this problem is interesting when the system is underdetermined and when certain properties of $A$, such as restricted isometry [4], can guarantee that $x^* = s$. An important problem is the characterization of the performance of the minimization (1). In particular, the relation between the sparsity of $s$, the number of measurements $m$ and the ambient dimension $n$ has been studied in great detail for Gaussian measurements ([1,3,5]).

In this work, we investigate a modified $\ell_1$ minimization algorithm, which significantly improves over regular $\ell_1$ minimization, by using prior information about the structure of the “non-uniformly sparse” solution as described in Definition 2.1.

Contributions: In [7], Khajehnejad et al. use a Grassman angle approach to calculate the optimal weights. We should emphasize that by optimal weights we mean the weights that minimize the number of Gaussian measurements required for weighted $\ell_1$ minimization to succeed in recovering the sparse signal. The approach of [7] requires extensive analytical work and the numerical procedure for optimal weight calculation is inefficient for more than two blocks. In this paper, we determine the optimal weights by extending the methods of [1]. In [1], Stojnic develops a new and relatively simpler method to obtain the minimum number of measurements $m$ to recover a sparsity of $k$ and finds the phase transition curve which matches with the results of [3,5], which are known to be exact. Consequently, without using complicated techniques, we are able to provide a method for obtaining optimal weights and because the analysis in [7] is based on [3,5], our results are consistent with [7]. Further, the resulting numerical procedure for optimal weight computation is cheaper and more clear. We also provide a heuristic method for estimating the optimal weight based on the more general “atomic decomposition” model presented in [11].

The structure of this paper is as follows. In the next section we define the “non-uniform sparse” model and weighted $\ell_1$ minimization and give other supplementary definitions. We then state the main theorems that will be used in this paper and proceed to the “Gaussian width” analysis of the weighted $\ell_1$ minimization.

2. SIGNAL MODEL
First, we introduce our model and algorithm.

Definition 2.1 (Non-uniformly Sparse Signal) Let $K = \{K_1, K_2, ..., K_u\}$ be a partition of $\{1, 2, ..., n\}$, i.e., $K_i \cap K_j = \emptyset$ for $i \neq j$, $\bigcup_{i=1}^{u} K_i = \{1, 2, ..., n\}$ and let $|K_i| = n_i = \gamma_i n$ for $1 \leq i \leq u$. Let $B = \{\beta_1, \beta_2, ..., \beta_u\}$ be a set of positive numbers in [0,1]. An $n \times 1$ vector $x = (x_1, x_2, ..., x_n)^T$ is said to be a non-uniformly sparse

978-1-4673-0046-9/12/$26.00 ©2012 IEEE 3621 ICASSP 2012
vector with sparsity fractions $B$ over $K$ if $k_i = \beta_i n_i$ entries
of $x$ in $K_i$ are non-zero. Finally, the overall sparsity of the
signal is denoted by $k = \sum_{i=1}^{n} k_i$ and $x \in \mathbb{R}^n$ is the vec-
tor that corresponds to the entries of $x$ in the $i$'th block for
$1 \leq i \leq u$.

Observe that, we don’t have any information about the entries inside a particular block hence it is natural to use the
same weight for all entries of the same block. We use the
following weighted $\ell_1$ minimization scheme:
\[
\min_{Ax = y} \|x\|_{1,w} = \min_{Ax = y} \sum_{i=1}^{n} w_i \|x^i\|_1.
\] (2)
where $w = \{w_1, w_2, ..., w_u\}$ is the weight vector we use. As
argued in [9], one would expect to associate larger weights
to the sparser blocks according to the so-called “punish the
sparser” scheme. By determining the optimum weight vec-
tor, we will show that this scheme can significantly improve
regular $\ell_1$ minimization.

Similar to relevant previous works, it is assumed that mea-
surement matrix is $m \times n$ Gaussian. Also, we assume the so-
called “linear regime” where $\{\gamma_i\}_{i=1}^{u}, \{\beta_i\}_{i=1}^{u}$ and $\alpha = \frac{m}{n}$
are constant. Our results will be true asymptotically with high
probability (a.w.h.p.), i.e. as $n \rightarrow \infty$.

Finally, the “recovery threshold” of a weighted algorithm
is the minimum $\alpha$ that can guarantee the successful recovery
of a signal obeying the non-uniform model a.w.h.p. Naturally,
the recovery threshold is a function of $\{\beta_i, \gamma_i, w_i\}_{i=1}^{u}$.

3. MAIN THEOREMS

We now state a seminal result from [8] which is the most im-
portant theorem used in this paper.

Theorem 3.1 (Escape through a mesh) Let $S$ be a subset of
the unit Euclidean sphere $S^{n-1}$ in $\mathbb{R}^n$. Let $\mathcal{Y}$ be a random
$(n - m)$-dimensional subspace of $\mathbb{R}^n$, distributed uniformly
in the Grassmanian with respect to the Haar measure. Let the
Gaussian width of $S$, denoted by $g(S)$, be defined as
\[
g(S) = E \sup_{v \in \mathcal{S}} \hat{h}^T v
\] (3)
where $\hat{h}$ is a random column vector in $\mathbb{R}^n$ with i.i.d. $\mathcal{N}(0, 1)$
components. Assume that $g(S) < (\sqrt{m} - \frac{1}{4\sqrt{m}})$. Then
\[
P(\mathcal{Y} \cap S = 0) > 1 - 3.5e^{-\frac{(\sqrt{m} - \frac{1}{4\sqrt{m}} - \Delta(S))^2}{18}}
\] (4)

Let $A$ be a $k \times n$ “Gaussian Matrix” and $\mathcal{N}(A)$ denote
the null space of $A$, i.e. $t \in \mathcal{N}(A) \iff At = 0$. As
mentioned in [10], it is a well known fact that $\mathcal{N}(A)$ is
distributed uniformly in the Grassmanian of $(n - k)$-dimen-
sional subspaces of $\mathbb{R}^n$, with respect to the Haar measure.

Since we use Gaussian measurements $A$, Theorem 3.1
helps one to analyze the null space of $A$. This idea was first
introduced by Stojnic, in [1], and was used to find various
thresholds for regular $\ell_1$ minimization [1], [12]. In this pa-
per, we show that same method can be extended to analyze
weighted $\ell_1$ minimization. To do this, we need to charac-
terize the undesired null space elements because whenever
$\mathcal{N}(A)$ does not contain them weighted $\ell_1$ minimization gives
us the desired sparse solution.

It is known that (see, e.g., [1], [3]), the particular loca-
tions and signs of the non-zero elements of the underlying
sparse solution $s$ are irrelevant to the analysis of the recovery
threshold. The following theorem gives a necessary and suf-
cient condition for the success of weighted $\ell_1$ minimization
(7, 9):

Theorem 3.2 (Null Space Condition) Assume $A \in \mathbb{R}^{m \times n}$
is Gaussian and let $s$ be a $k$-sparse vector as described in
Definition 2.1. Assume its non-zero components are negative.
Also, let the first $n_1$ components be in $K_1$, the next $n_2$ compo-
nents in $K_2$, and so on. Let the $j$'th component of the $i$'th
block $s^i$ be denoted by $s_k^i$ for $1 \leq i \leq u$ and $1 \leq j \leq n_i$.
Also let $s_1^i = s_2^i = ... = s_{n_i-k_i}^i = 0$. Then subject to the
observations $y = As$, minimization (2) will return $s$ as the
unique solution iff for any $t \in \mathcal{N}(A)$ we have:
\[
\sum_{i=1}^{u} \left( w_i \sum_{j=n_i-k_i+1}^{n_i} t^i_j \right) < \sum_{i=1}^{u} \left( w_i \sum_{j=1}^{n_i-k_i} |t^i_j| \right)
\] (5)

The next section is dedicated to the analysis of the
weighted recovery threshold using Theorems 3.1 and 3.2.

4. ANALYSIS OF THE RECOVERY THRESHOLD

For simplicity, we’ll focus on the case $u = 2$ and without loss of
generality assume $w_1 = 1$ and $w_2 = w$.

Let $S_{w}$ be the set of vectors which have unit $\ell_2$ norm
and which do not satisfy (2). For a given $w$, we will com-
pute a tight upper bound on $g(S_{w})$ and then equate that to
$\left(\sqrt{m} - \frac{1}{4\sqrt{m}}\right)$. We set $g(\hat{h}, S_{w}) = \max_{v \in S_{w}} (\hat{h}^T v)$ and, as
a first step, we’ll determine an upper bound $B_{w}$ on $g(\hat{h}, S_{w})$
for a fixed $\hat{h}$.

In the following analysis, due to insufficient space, we
will omit some of the intermediate steps and will try to em-
phasize the main points instead. For a complete discussion,
reader is referred to the longer version of this paper [14].

4.1. Upper bound $B_{w}$ on $g(\hat{h}, S_{w})$

Our analysis will mostly follow that of [1]. Let $\hat{h}_{i:1:(n_i-k_i)}^i =
(\hat{h}_{i:1}^1, \hat{h}_{i:1}^2, ..., \hat{h}_{i:1}^{n_i-k_i})^T$. Further, let $|\hat{h}_{i:1}^j|_{(j)}$ be the $j$'th smallest
magnitude of elements of $\hat{h}_{i:1:(n_i-k_i)}^i$. Let $h = (h^1, h^2)$ where
\[
h^i = (|\hat{h}_{i:1}^i|_{(1)}, |\hat{h}_{i:1}^i|_{(2)}, ..., |\hat{h}_{i:1}^i|_{(n_i-k_i)}, \hat{h}_{i:1}^{n_i-k_i+1}, ..., \hat{h}_{i:1}^{n_i})^T
\] (6)
By using elementary arguments, \( g(\tilde{h}, S_w) \) can be expressed as the outcome of the following optimization (11):

\[
\max_{y \in \mathbb{R}^n} (h^1)^T y^1 + (h^2)^T y^2
\]

subject to

i) \( y_j^i \geq 0 \) for all \( i = 1, 2 \) and \( 0 \leq j \leq (n_i - k_i) \)

ii) \( \sum_{i=1}^{2} \left( w_i \sum_{j=1}^{n_i - k_i} y_j^i \right) \geq \sum_{i=1}^{2} \left( w_i \sum_{j=1}^{n_i - k_i} |y_j^i| \right) \)

iii) \( \sum_{i=1}^{2} \left( \sum_{j=1}^{n_i} (y_j^i)^2 \right) \leq 1. \)

Let \( z = (z^1, w^2) \) where \( z^1 \in \mathbb{R}^{n_1} \) such that \( z^1_j = 1, 1 \leq j \leq n_1 - k_1 \) and \( z^1_j = -1, n_1 - k_1 + 1 \leq j \leq n_1 \). The next step is changing the above maximization problem to a minimization problem by writing out its dual which will provide us an upper bound \( g_{up}(\hat{h}, S_w) \):

\[
\min_{\gamma, \nu, \lambda} \max_y \ h^1 y^1 - \gamma \|y\|_2^2 + \gamma - \nu z^T y + \lambda^T y
\]

subject to \( \nu \geq 0, \lambda \geq 0 \)

\( \lambda^i_j \geq 0 \) for \( i = 1, 2 \) and \( 0 \leq j \leq n_i - k_i \).

After trivially solving the minimization over \( y \) and maximization over \( \gamma \) we end up with:

\[
g_{up}(\hat{h}, S_w) = \min_{\nu, \lambda} \|\lambda + \nu z\|_2
\]

subject to \( \nu \geq 0 \)

\( \lambda^i_j \geq 0 \) for \( i = 1, 2 \) and \( 0 \leq j \leq n_i - k_i \).

To tackle this problem, we deal with the square of the objective function in (6), namely \( \|\lambda + \nu z\|_2^2 \). Now let \( \lambda_1 = (\lambda_1^1, \lambda_1^2, \ldots, \lambda_1^{n_1}) \) where \( \lambda_1^i \leq (n_1 - k_1) \) and \( \lambda_2 = (\lambda_2^1, \lambda_2^2, \ldots, \lambda_2^{n_2}) \) where \( \lambda_2^i \leq (n_2 - k_2) \). Parameters \( c_1 \) and \( c_2 \) will be determined later. This makes the optimization over \( \nu \) simpler which can be done by differentiating the objective with respect to \( \nu \). Assuming \( \lambda \) is positive, \( \lambda^i_j \)'s can be determined by further differentiation and proceeding in a similar way to [1]. Eventually, we end up with the following lemma for fixed \( c_1, c_2 \).

**Lemma 4.1** Let \( c' = c_1 + w^2 c_2, n' = n_1 + w^2 n_2 \) and \( f(h, z, w, c_1, c_2) = \frac{(h^T z - \sum_{j=1}^{n'} h_j^1 - w \sum_{k=1}^{n'_1} h_k^2)}{n' - c'} \).

Assume

\[
f(h, z, w, c_1, c_2) \geq \max \{h_1^1, h_2^2\}
\]

Then \( g(\hat{h}, S_w) \) is upper bounded by \( g_{up}(\hat{h}, S_w, c_1, c_2) = \sqrt{\sum_{j=1}^{n'_1} (h_j^1)^2 + \sum_{k=c_2+1}^{n'_2} (h_k^2)^2 - (n' - c') f(h, z, w, c_1, c_2)^2} \)

Let \( F_\alpha(\cdot) \) be the magnitude distribution of a unit variance Gaussian RV. For any \( \epsilon > 0 \), if \( c_1, c_2 \) are chosen such that

\[
(1 - \epsilon)E[f(h, z, w, c_1, c_2)] = F_\alpha^{-1} \left( \frac{c_1 (1 + \epsilon)}{n_1 (1 - \beta_1)} \right)
\]

\[
(1 - \epsilon)E[f(h, z, w, c_1, c_2)] = F_\alpha^{-1} \left( \frac{c_2 (1 + \epsilon)}{n_2 (1 - \beta_2)} \right)
\]

it can be shown that as \( n \to \infty \), (8) will be satisfied except for an exponentially small probability. Consequently,

\[
E[g(\hat{h}, S_w)] \leq E[g_{up}(\hat{h}, S_w, c_1, c_2)] + o(1)
\]

\[
\leq \sqrt{E[(g_{up}(\hat{h}, S_w, c_1, c_2))^2]} + o(1)
\]

Hence, we have the following theorem:

**Theorem 4.1** Let \( \hat{h} \) be an i.i.d. Gaussian vector with entries having unit variance. \( \hat{h} \) is function of \( \tilde{h} \) as given in (5) and \( n', c' \) are same as in Lemma 4.1. Then, for a given weight \( w \)

\[
E \left[ \sum_{i=1}^{2} \sum_{j=c_1+1}^{n_i} (h_j^i)^2 \right] - (n' - c')E[f(h, z, w, c_1, c_2)]^2
\]

measurements are sufficient for recovering a signal that obeys Definition 2.1 w.h.p. Here \( c_1, c_2 \) are solutions of (9),(10).

Next, let \( \epsilon \to 0 \) in (9),(10) and using the fact that \( \hat{h} \) is Gaussian, we restate Theorem 4.2 for the more general case of \( u \geq 2 \) in a way that is accessible for numerical calculations.

**Theorem 4.2** Let \( \eta_i, \kappa, \mu \) be functions of \( \{\beta_i, \gamma_i, \theta_i\}_{i=1}^u \) defined as follows

\[
\eta_i = \text{erfinv} \left( \frac{1 - \theta_i}{1 - \beta_i} \right) \text{ for } 0 \leq i \leq u
\]

\[
\kappa = \left( \sum_{i=1}^u \mu_i \gamma_i^2 \right)^{-1}
\]

\[
\mu = \sum_{i=1}^u \gamma_i (1 - \beta_i) \sqrt{\frac{2}{\pi}} \exp(-\eta_i^2)
\]

where \( \theta_i = \left( 1 - \frac{\kappa}{\mu} \right) \).

**Step 1:** For \( 0 \leq i \leq u \) let \( \theta_i \) be the solution of:

\[
\sqrt{2} \eta_i = w_\kappa \mu
\]

**Step 2:** Then, an upper bound to the threshold of weighted \( \ell_1 \) minimization with weight \( w \) is called \( \alpha(w) \) and is given by

\[
\sum_{i=1}^u \gamma_i \left( 1 - \beta_i \right) \left( 1 + \frac{2 \eta_i}{\sqrt{\pi} \exp(\eta_i^2)} - \text{erf}(\eta_i) \right) + \beta_i - \kappa \mu
\]

Hence \( \alpha(w) \times n \) measurements are sufficient to recover a “nonuniformly sparse signal” given in Definition 2.1 w.h.p.

Observe that Steps 1 and 2 in Theorem 4.2 corresponds to (9),(10) and (??) respectively.

Finally, in order to determine the optimal weight, \( \alpha(w) \) is minimized over \( w \) which can be done numerically and will be the topic of the next section.
5. RESULTS

Comparison with previous work: Our results are consistent with [7] as simulations suggest we obtain the same optimal weighting given \( \{ \beta_i, \gamma_i \}_{i=1} \). Since [7] is based on [3] and exact, this shows our analysis yields the true optimal weight by using an alternative method.

On the other hand, our method is more accessible: Theorem 4.2 is concise and our numerical computations are much less intensive. For example, the weight curve in Figure 5 is smooth unlike those of [7]. Curves in [7] were not calculated with high density because of the high complexity of the optimal weight calculation [13].

A heuristic to estimate the optimal weight

In [11], a new framework is provided to determine the approximate thresholds using Gaussian width analysis using which it can be shown that regular \( \ell_1 \) requires \( \alpha \geq 2(1 + e^{-1}) \beta (\log((1 - \beta)/\beta) + 1.5) \) for success. Keeping this as the reference point, one can come up with the following closed form approximation for the optimal weight in the case of weighted \( \ell_1 \) minimization:

\[
w = \sqrt{\frac{1 + \log\left(\frac{1 - \beta_2}{\beta_2}\right)}{1 + \log\left(\frac{1 - \beta_1}{\beta_1}\right)}} \tag{16}
\]

It should be emphasized that (12) is independent of the block sizes \( \gamma_1, \gamma_2 \).

5.1. Simulation results and Conclusion

Figure 5 presents the threshold value \( \alpha(w) \) for different values of \( w = \frac{\alpha}{\gamma_i} \) given the values of \( \beta_1, \beta_2, \gamma_1, \gamma_2 \). The value of the recovery threshold corresponding to the weight given by the heuristic in (12) is highlighted.

In general, we see a substantial decrease in the recovery threshold from regular \( \ell_1 \) by the heuristic in (12) is highlighted. Although not very satisfactory, gives us approximately the optimal weight.

6. REFERENCES


