Projected ℓ_1 -Minimization for Compressed Sensing

Amin Khajehnejad, Matthew Thill and Babak Hassibi

Abstract—We propose a new algorithm to recover a sparse signal from a system of linear measurements. By projecting the measured signal onto a properly chosen subspace, we can use the projection to zero in on a low-sparsity portion of our original signal, which we can recover using ℓ_1 -minimization. We can then recover the remaining portion of our signal from an overdetermined system of linear equations. We prove that our scheme improves the threshold of ℓ_1 -minimization, and we derive an upper bound for this new threshold. We support our theoretical results with numerical simulations which demonstrate that certain classes of signals come close to achieving this upper bound.

Index terms— Compressed sensing, ℓ_1 -minimization, reweighted ℓ_1 -minimization, projected ℓ_1 -minimization.

I. INTRODUCTION

In the past few years, the field of compressed sensing has received substantial attention, addressing the problem of recovering a sparse signal from a relatively small set of linear measurements [1]. If \mathbf{x} is a real *n*-dimensional vector with *k* nonzero entries, and \mathbf{A} is an $m \times n$ real matrix, with k < m < n, we would like to efficiently determine \mathbf{x} knowing $\mathbf{y} = \mathbf{A}\mathbf{x}$. Much work has focused on determining sufficient conditions for variations of the ℓ_1 -minimization problem:

$$\min_{\mathbf{A}\hat{\mathbf{x}}=\mathbf{v}} ||\hat{\mathbf{x}}||_1,\tag{1}$$

to recover \mathbf{x} , either with certainty or with high probability [2]–[7]. A well established property of ℓ_1 -minimization is the phase transition property which holds in the regime of linear system dimensions. Proved originally by Donoho and Tanner [2], it is known that for the case of Gaussian measurement matrices, for any given ratio $\delta = \frac{m}{n}$, there exists a so-called weak threshold $\mu(\delta)$ for the sparsity of \mathbf{x} such that if $\frac{k}{n} \leq \mu(\delta)$, then ℓ_1 -minimization will recover \mathbf{x} with high probability. The practical performance of the ℓ_1 reconstruction algorithm exhibits a very sharp transition phase on the sides of the weak threshold, and as such the theoretical performance bounds of [2] are tight.

Certain extentions of the basis pursuit algorithm (a.k.a. ℓ_1 -minimization) have addressed the feasibility of finding reconstruction methods with better performance than regular ℓ_1 -minimization, and in particular with higher weak recovery thresholds. Examples of previous efforts along

those lines are [8], [9] which consider reconstruction of nonuniformly sparse signals given additional prior information, and [10]–[12] that study reweighted ℓ_1 -minimization schemes for improved recovery performance. In particular, the iterative reweighted algorithm introduced by Khajehnejad et al. [10] is based on identifying a set of indices S which contains a probabilistically large intersection with the support set of \mathbf{x} , possibly by choosing the klargest entries of the vector obtained from standard ℓ_1 minimization, and then solving a new biased program

$$\min_{\mathbf{A}\hat{\mathbf{x}}=\mathbf{y}} ||\hat{\mathbf{x}}_S||_1 + w||\hat{\mathbf{x}}_{\overline{S}}||_1,$$
(2)

where w is a weight parameter to be chosen greater than 1. Intuitively, this approach punishes indices in \overline{S} when attempting to identify the support set of \mathbf{x} .

As proved in [10], the above approach yields a recovery threshold that is strictly higher than that of regular ℓ_1 minimization, thus allowing for the recovery of signals with larger support sets. However, the very involved derivations of [10] make it difficult to compute explicit bounds for the threshold under this framework. The current paper presents a framework which covers certain cases of reweighted ℓ_1 -minimization, but is more conducive to the analysis and approximation of the threshold improvement over standard ℓ_1 -minimization.

We present an alternative extension of the basis pursuit algorithm, which lends itself to a formidable analysis of the performance bounds of the method. The algorithm is based on first identifying a candidate set S for the support set of \mathbf{x} , and considering the submatrix \mathbf{A}_S whose columns lie in the index set S. We then project \mathbf{y} onto the orthogonal complement to the column space of \mathbf{A}_{S} . The reduced projected system is then solved using another ℓ_1 -minimization, which is then exploited subsequently to recover \mathbf{x}_{S} through linear matrix inversions. In contrast to previously suggested iterative schemes, the proposed two-step projected ℓ_1 algorithm does not involve a set of weight parameters, which in other algorithms depends heavily on the signal distribution and the ideality of the sparsity assumption. In addition, we derive explicit numerical upper bounds on the performance improvement of the proposed method, and provide a numerical tool for a precise calculation of the new threshold. The obtained upper bounds are very close to the practical values of the threshold improvement verified by numerical simulations.

The organization of this paper is as follows: In Section II, we will establish some definitions and notations. In Section III, we introduce our algorithm. Section IV is dedicated to the analysis of the algorithm and how it improves the threshold of ℓ_1 -minimization. In Section V, we provide

The authors are with the Department of Electrical Engineering, California Institute of Technology, Pasadena, CA, USA. email:{amin,mthill, hassibi}@caltech.edu.

This work was supported in part by the National Science Foundation under grants CCF-0729203, CNS-0932428 and CCF-1018927, by the Office of Naval Research under the MURI grant N00014-08-1-0747, and by Caltech's Lee Center for Advanced Networking.

simulation results to compare the performance of our algorithm for various distributions of the support set, and analyze which distributions come closest to attaining the improved threshold.

II. DEFINITIONS AND MODEL

Let \mathbf{x} be a k-sparse, n-dimensional vector, with support set $supp(\mathbf{x})$. By this, we mean $|supp(\mathbf{x})| = k$. We will typically assume that \mathbf{x} is a Gaussian vector, that is, that the k entries $x_i, i \in supp(\mathbf{x})$ are i.i.d. $\mathcal{N}(0, 1)$ random variables. Much of our analysis, however, applies to other distributions as well. The goal is to recover \mathbf{x} from a system of m Gaussian linear measurements—that is, to determine \mathbf{x} from $\mathbf{y} = \mathbf{A}\mathbf{x}$, where, \mathbf{A} is an $m \times n$ matrix whose entries are i.i.d. Gaussian. Since recovery beyond the limits of ℓ_1 minimization is considered, the primary assumption here is that the number k of the nonzero entries of \mathbf{x} is larger than the weak recovery threshold of ℓ_1 -minimization, namely $k = (1 + \epsilon_0)\mu(\delta)n$ for some positive ϵ_0 .

III. PROJECTED ℓ_1 -MINIMIZATION (PJL1) ALGORITHM

We propose the following two step algorithm for reconstructing the sparse vector \mathbf{x} .

Algorithm 1 Two Step Projected ℓ_1 -minimization.

- 1: **Input:** Measurement matrix $\mathbf{A}^{m \times n}$, measurement vector $\mathbf{y}^{m \times 1}$.
- 2: Output: Sparse vector \mathbf{x} with $\mathbf{A}\mathbf{x} = \mathbf{y}$.
- 3: Perform a standard ℓ_1 -minimization to obtain an estimate $\hat{\mathbf{x}}$ for \mathbf{x} .
- 4: Let $S \subseteq [n]$ be the set of indices of the \tilde{k} largestmagnitude entries of $\hat{\mathbf{x}}$, where $\tilde{k} \leq \mu(\delta)n$, and \mathbf{A}_S the $m \times \tilde{k}$ submatrix of \mathbf{A} with columns indexed by S. Construct the $(m - \tilde{k}) \times m$ orthogonal transformation \mathbf{A}_S^{\perp} .
- 5: Let \overline{S} be the complement of S in [n], and let $\mathbf{A}_{\overline{S}}$ be the $m \times (n - \tilde{k})$ submatrix of \mathbf{A} with columns indexed by \overline{S} . Define the $(m - \tilde{k}) \times (n - \tilde{k})$ reduced measurement system $\mathbf{A}' := \mathbf{A}_{S}^{\perp} \mathbf{A}_{\overline{S}}$. Perform the reduced ℓ_{1} minimization,

$$\min_{\mathbf{A}'\hat{\mathbf{z}}=\mathbf{A}_{S}^{\perp}\mathbf{y}}||\hat{\mathbf{z}}||_{1}.$$

6: Finally, set $\hat{\mathbf{z}}' = \mathbf{A}_{S}^{\dagger}(\mathbf{y} - \mathbf{A}_{\overline{S}}\hat{\mathbf{z}})$, where \mathbf{A}_{S}^{\dagger} is the pseudoinverse of \mathbf{A}_{S} , defined as $\mathbf{A}_{S}^{\dagger} := (\mathbf{A}_{S}^{T}\mathbf{A}_{S})^{-1}\mathbf{A}_{S}^{T}$. Return $\hat{\mathbf{x}}' = \begin{bmatrix} \hat{\mathbf{z}}' \\ \hat{\mathbf{z}} \end{bmatrix}$.

In this algorithm, $\tilde{k} := |S|$ is a parameter which can be optimized depending on the signal. We require that \tilde{k} be less than $\mu(\delta)n$ because this is the region over which S is guaranteed to have significant overlap with the support set of \mathbf{x} with high probability. Intuitively, the reduced system $\mathbf{A}'\hat{\mathbf{z}} = \mathbf{A}_S^{\perp}\mathbf{y}$ in the algorithm allows us to search for a section of \mathbf{x} which we may expect to have relatively small sparsity. Once we have reconstructed this section of \mathbf{x} , we are left with an overdetermined system of equations, so we may perfectly recover the remaining section, corresponding to the smaller, high-sparsity portion of \mathbf{x} . In fact, Algorithm 1 is equivalent to the reweighted ℓ_1 -minimization problem:

$$\min_{\mathbf{x}_{\overline{S}}} w_{S} || \hat{\mathbf{x}}_{S} ||_{1} + w_{\overline{S}} || \hat{\mathbf{x}}_{\overline{S}} ||_{1},$$

where we set $w_S = 0$ and $w_{\overline{S}} = 1$. To see this we note that certainly the condition $\mathbf{A}\hat{\mathbf{x}} = \mathbf{y}$ implies that $\mathbf{A}_S^{\perp}\mathbf{A}_{\overline{S}}\hat{\mathbf{x}}_{\overline{S}} =$ $\mathbf{A}_S^{\perp}\mathbf{y}$. Conversely, we can write $\mathbf{y} = \mathbf{y}_S + \mathbf{y}_S^{\perp}$, where \mathbf{y}_S is in the column space of \mathbf{A}_S and \mathbf{y}_S^{\perp} is in the orthogonal complement to this column space. Then the condition $\mathbf{A}_S^{\perp}\mathbf{A}_{\overline{S}}\hat{\mathbf{x}}_{\overline{S}} = \mathbf{A}_S^{\perp}\mathbf{y}$ is equivalent to $\mathbf{A}_S^{\perp}\mathbf{A}_{\overline{S}}\hat{\mathbf{x}}_{\overline{S}} = \mathbf{y}_S^{\perp}$, and the vector difference $\mathbf{y} - \mathbf{A}_{\overline{S}}\hat{\mathbf{x}}_{\overline{S}}$ lies in the column space of \mathbf{A}_S , so there is some $\hat{\mathbf{x}}_S$ such that $\mathbf{A}_S\hat{\mathbf{x}}_S + \mathbf{A}_{\overline{S}}\hat{\mathbf{x}}_{\overline{S}} = \mathbf{y}$, and we see that the constraints for the two minimization problems are in fact equivalent.

IV. Analysis of the PJL1 Algorithm

To analyze the behavior of Algorithm 1, let S denote the \tilde{k} -support of $\hat{\mathbf{x}}$, *i.e.*, the \tilde{k} largest-magnitude entries of $\hat{\mathbf{x}}$, and suppose $|S \cap supp(\mathbf{x})| = (1 - \epsilon)k$, so that $|\overline{S} \cap supp(\mathbf{x})| = \epsilon k$. We first note that the modified measurement matrix $\mathbf{A}' = \mathbf{A}_S^{\perp} \mathbf{A}_{\overline{S}}$ will maintain the property of having a rotationally-invariant null space (that is, the distribution of the null space is invariant under unitary transformations). This is because the same is true of both \mathbf{A}_S^{\perp} and $\mathbf{A}_{\overline{S}}$ individually, and they are independent of each other. Consequently, the weak threshold μ applies to the reduced ℓ_1 -minimization in Algorithm 1.

We can now give the following upper bound on ϵ which will ensure that **x** will be recovered by Algorithm 1 with high probability:

Theorem 1: Let \mathbf{x} be a $k = (1 + \epsilon_0)\mu(\delta)n$ -sparse signal, and S a set of indices such that $|S \cap supp(\mathbf{x})| = (1 - \epsilon)k$. Then Algorithm 1 will recover \mathbf{x} with high probability provided that

$$\epsilon \leq \frac{1 - (1 + \epsilon_0)\mu(\delta)}{(1 + \epsilon_0)\mu(\delta)} \mu\left(\frac{\delta - (1 + \epsilon_0)\mu(\delta)}{1 - (1 + \epsilon_0)\mu(\delta)}\right).$$
(3)

Proof: Note that we can re-express the minimization in step 5 of the algorithm as

$$\min_{\mathbf{A}'\hat{\mathbf{z}}=\mathbf{A}'\mathbf{x}_{\overline{S}}} ||\hat{\mathbf{z}}||_1.$$

Since $\mathbf{x}_{\overline{S}}$ is ϵk -sparse, we see that $\hat{\mathbf{z}}$ will recover $\mathbf{x}_{\overline{S}}$ with high probability provided that $\frac{\epsilon k}{n-k} \leq \mu\left(\frac{m-k}{n-k}\right)$. Rewriting this in terms of ϵ_0 and δ , we obtain the desired result.

A. Upper Bound on Threshold Improvement

For a given δ , we denote the recovery threshold of the PJL1 algorithm by $\tilde{\mu}(\delta)$. Following Theorem 1, we note that in the above framework, the maximum size of the support set that can be recovered with high probability is

$$k = |S| + (n - |S|)\mu\left(\frac{m - |S|}{n - |S|}\right),\tag{4}$$

where $|S| \leq \mu(\delta)n$. Since $\mu(\cdot)$ is a sublinear function, it can be shown that for fixed $\delta = \frac{m}{n}$, the expression on the right hand side of (4) is an increasing function of |S|. Thus we obtain an upper bound by setting $|S| = \mu(\delta)n$. If we call our new threshold $\tilde{\mu}(\delta)$, then we can express this as follows:

$$\tilde{\mu}(\delta) \le \mu(\delta) + (1 - \mu(\delta))\mu\left(\frac{\delta - \mu(\delta)}{1 - \mu(\delta)}\right).$$
(5)

We stress that the above inequality clearly expresses the maximum possible improvement in the threshold of ℓ_1 -minimization as an additive term. In Figure 1, we plot both the standard ℓ_1 -minimization threshold (as computed using the results of [7]) as well as the upper bound for $\tilde{\mu}(\delta)$ from Equation (5).



Fig. 1. The improvement of the Projected ℓ_1 -minimization threshold (as estimated from the bound in Equation 5) over the threshold from standard ℓ_1 -minimization.

B. Exact Threshold Improvement

While (5) is a computable upper bound for the desired threshold, we could attempt to evaluate the threshold as follows: If we consider the fraction f of the set S which is actually contained in $supp(\mathbf{x})$, then the actual sparsity that can be recovered is

$$k = f|S| + (n - |S|)\mu\left(\frac{m - |S|}{n - |S|}\right).$$
 (6)

Thus, we could attempt to optimize over both |S| and f. We make the following definition:

Definition 1: Let \mathbf{x} be an *n*-dimensional signal, and let β and ν be given fractions between 0 and 1. Let \mathbf{A} be an $m \times n$ Gaussian measurement matrix, and let $\hat{\mathbf{x}}$ be the vector recovered by the ℓ_1 -minimization of (1). Then we define $\varphi(\beta, \nu)$ to be the largest fraction f such that if \mathbf{x} is νn -sparse, then with high probability, the largest βn entries of $\hat{\mathbf{x}}$ contain at least a fraction f of the support set of \mathbf{x} .

Essentially, $\varphi(\beta, \nu)$ is a probabilistic lower bound on the fraction f in Equation 6, and determines the support estimation capability of ℓ_1 -minimization. $\varphi(\beta, \nu)$ heavily depends on the distribution of the nonzero component of \mathbf{x} , and is generally larger for sparse signals with faster decaying dynamics. It follows from this definition that the real threshold of PJL1 algorithm, $\tilde{\mu}(\delta)$, is the solution to the following maximization program:

$$\tilde{\mu}(\delta) := \max_{\beta, f, \nu} \nu$$
s.t.
$$\begin{cases}
0 \le \beta \le \mu(\delta) \\
\nu \le \beta f + (1 - \beta)\mu\left(\frac{\delta - \beta}{1 - \beta}\right) \\
f \le \varphi(\beta, \nu)
\end{cases}$$
(7)

C. Tightness of Improvement Bound

In general, a tight bound for $\varphi(\beta,\nu)$ might not be available, and thus the computation of the exact numerical value for $\tilde{\mu}(\delta)$ is not necessarily tractable. In [10], it was shown that for certain distributions on the nonzero component of the signal (including Gaussian), $\varphi(\beta,\nu)$ is asymptotically computable as β and ν approach $\mu(\delta)$ and becomes arbitrarily close to 1. In general, building upon the derivations from [10], we obtain the following lemma, the proof of which is omitted due to lack of space:

Lemma 1: Let **x** be a k-sparse, n-dimensional real signal whose support set entries are i.i.d. with the pdf $p(\cdot)$. Let r be the minimum positive integer such that the r^{th} derivative of $p(\cdot)$ evaluated at 0, $p^{(r)}(0)$, is nonzero. Then if $\beta \leq \mu(\delta)$, we have

$$\varphi(\beta,\nu) = 1 - \mathcal{O}\left((\nu - \mu(\delta))^{1 + \frac{1}{r+1}}\right).$$
(8)

Intuitively, the value of $\varphi(\beta, \nu)$ in the maximization (7) which corresponds to the optimal value of ν should be close to 1, in which case the maximum is achieved when $\beta \approx \mu(\delta)$, giving us our bound. As a consequence of the above lemma, one can prove that for distributions $p(\cdot)$ with some finite order nonzero value at origin, the threshold $\tilde{\mu}(\delta)$ is strictly larger than $\mu(\delta)$. However, it does not provide a reasonable numerical estimation on how large this improvement is. Future work could focus on finding more exact approximations for $\varphi(\beta, \nu)$ and explicit computations of $\tilde{\mu}(\delta)$.

V. Simulations and Results

Figure 2 shows the results of our algorithm's attempts to reconstruct randomly-generated signals with various distributions for the support sets, and exhibits the improved recovery rates over those of standard ℓ_1 -minimization. Our algorithm demonstrates some improvement in the regime of Bernoulli signals, and significantly more for the classes of Gaussian and *exponentially-decaying* signals, which we define as follows:

Definition 2: Let **x** be a k-sparse, n-dimensional signal, with support set entries $x_1, x_2, ..., x_k$. Assume that the x_i

are ordered in decreasing order of magnitude, $|x_i| > |x_{i+1}|$. We say that **x** is an *exponentially decaying* signal if there is some positive constant c, independent of n and k, such that for each i = 1, ..., k - 1, we have $|x_{i+1}| < c|x_i|$.

In Figure 3, we show the results of the attempts of PJL1 to recover randomly-generated exponentially decaying signals for different values of $\delta = \frac{m}{n}$. We use a dimension of n = 100 in this case. Our results show that we observe nearly complete recovery of almost all signals provided that the sparsity fraction $\frac{k}{n}$ is less than the upper bound for $\tilde{\mu}(\delta)$ from Equation (5).



Fig. 2. Comparison of standard ℓ_1 -minimization with Projected ℓ_1 minimization for k-sparse signals with support sets having Bernoulli, Gaussian, and Exponentially-Decaying distributions. Here, we use values of m = 112, n = 200.



Fig. 3. Empirical recovery fraction of PJL1 algorithm for n = 100 (black and white cells); PJL1's improved threshold upper bound as in Equation (5) (red) and that of regular ℓ_1 -minimization (green). Black cells represent perfect recovery. The red threshold is nearly achieved by a signal with an exponentially-decaying distribution on its support set.

VI. CONCLUSION

We proved that the PJL1 algorithm has an increased threshold over that of standard ℓ_1 -minimization, and we derived an explicit upper bound for this threshold. We then exhibited numerical simulations which showed that this bound is nearly achieved by certain classes of supportset distributions, including Gaussians, and in the case of exponentially decaying signals for large n.

It would be interesting to derive tighter bounds on the threshold of PJL1 for various distributions, and to determine which distributions can achieve the upper bound for the threshold derived in this paper. Based on our algorithm's performance on exponentially decaying signals, it is reasonable to conjecture that distributions with sharply-decreasing tails might achieve the threshold bound obtained in this paper. It would also be worthwhile to analyze the robustness of the algorithm to noisy signals.

References

- Compressive sensing online resources at Rice university, http://www.dsp.ece.rice.edu/cs
- [2] D. Donoho and J. Tanner, "Thresholds for the Recovery of Sparse Solutions via L1 Minimization", Proceedings of the Conference on Information Sciences and Systems, March 2006.
- [3] D. Donoho," Compressed sensing", *IEEE Trans. on Information Theory*, 52(4), pp. 1289 1306, April 2006.
- [4] D. Donoho, "High-Dimensional Centrally Symmetric Polytopes with Neighborliness Proportional to Dimension", *Discrete and Computational Geometry*, 102(27), pp. 617-652, 2006, Springer.
- [5] E. Candés and T. Tao, "Decoding by linear programming", *IEEE Trans. on Information Theory*, 51(12), pp. 4203 4215, December 2005.
- [6] R. G. Baraniuk and M. B. Wakin "Random Projections of Smooth Manifolds", Journal of Foundations of Computational Mathematics, Volume 9, No.1, Feb. 2009.
- [7] M. Stojnic, "Various Thresholds for l₁-Optimization in Compressed Sensing," arXiv:0907.3666.
- [8] A. Khajehnejad, W. Xu, S. Avestimehr, B. Hassibi, "Analyzing Weighted L1 Minimization for Sparse Recovery With Nonuniform Sparse Models," IEEE Transactions on Signal Processing, Vol. 59(5), pp. 1985-2001, 2011.
- [9] N. Vaswani and W. Lu, "Modified-CS: Modifying Compressive Sensing for Problems with Partially Known Support", IEEE Trans. Signal Processing, September 2010.
- [10] A. Khajehnejad, W. Xu, S. Avestimehr and B. Hassibi, "Improved Sparse Recovery Thresholds with Two-Step Reweighted ℓ₁ Minimization", in Proceedings of ISIT Conference 2010.
- [11] D. Needell, "Noisy signal recovery via iterative reweighted L1minimization" Proc. Asilomar Conf. on Signals, Systems, and Computers, Pacific Grove, CA Nov. 2009.
- [12] E. J. Candés, M. B. Wakin, and S. Boyd, "Enhancing Sparsity by Reweighted 11 Minimization", *Journal of Fourier Analysis* and Applications, 14(5), pp. 877-905, special issue on sparsity, December 2008.
- [13] W. Xu and B. Hassibi, "On Sharp Performance Bounds for Robust Sparse Signal Recoveries," in Proceedings of ISIT Conference 2009.