Signal Recovery From Incomplete and Inaccurate Measurements via Regularized Orthogonal Matching Pursuit

Deanna Needell* and Roman Vershynin

***Abstract*—We demonstrate a simple greedy algorithm that can reliably recover a vector $v \in \mathbb{R}^d$ from incomplete and inaccurate measurements $x = \Phi v + e$. Here Φ is a $N \times d$ measurement matrix with $N \ll d$, and e is an error vector. Our algorithm, Regularized Orthogonal Matching Pursuit (ROMP), seeks to provide the benefits of the two major approaches to sparse recovery. It combines the speed and ease of implementation of the greedy methods with the strong guarantees of the convex programming methods.**

For any measurement matrix Φ that satisfies a quantitative restricted isometry principle, ROMP recovers a signal v with $O(n)$ nonzeros from its inaccurate measurements x in at most n iterations, where each iteration amounts to solving a Least Squares Problem. The noise level of the recovery is proportional to $\sqrt{\log n}\|e\|_2$. In particular, if the error term e vanishes the reconstruction is exact.

This stability result extends naturally to the very accurate recovery of approximately sparse signals.

***Index Terms*—Compressed Sensing, sparse approximation problem, Orthogonal Matching Pursuit, Uncertainty Principle.**

I. Introduction

A. Exact recovery by convex programming

The recent massive work in the area of Compressed Sensing, surveyed in [4], rigorously demonstrated that one can algorithmically recover sparse (and, more generally, compressible) signals from incomplete observations. The simplest model is a d-dimensional signal v with a small number of nonzeros:

$$v \in \mathbb{R}^d, \qquad |\mathrm{supp}(v)| \le n \ll d.$$

Such signals are called n-sparse. We collect $N \ll d$ nonadaptive linear measurements of v, given as $x = \Phi v$ where Φ is some N by d measurement matrix. The sparse recovery problem is to then efficiently recover the signal v from its measurements x.

A necessary and sufficient condition for exact recovery is that the map Φ be one-to-one on the set of n-sparse vectors. Much work has been done to show that under some circumstances, a convex optimization problem can be used to recover such signals (see e.g. [14], [7]). These results show that the sparse recovery problem is equivalent to the convex program

$$\min \|u\|_1 \qquad \text{subject to} \qquad \Phi u = x \tag{I.1}$$

and therefore is computationally tractable. Candès and Tao [7] provide a result showing that when the map Φ is an almost isometry on the set of $O(n)$-sparse vectors, the program (I.1) recovers sparse signals. This condition imposed on Φ is the restricted isometry condition:

Definition 1.1 (Restricted Isometry Condition): A measurement matrix Φ satisfies the *Restricted Isometry Condition* (RIC) with parameters (m, ε) for $\varepsilon \in (0,1)$ if we have

$$(1-\varepsilon)\|v\|_2 \le \|\Phi v\|_2 \le (1+\varepsilon)\|v\|_2 \qquad \text{for all } m\text{-sparse vectors.}$$

Under the Restricted Isometry Condition with parameters $(2n, \sqrt{2}-1)$, the convex program (I.1) exactly recovers an n-sparse signal v from its measurements x [7], [8].

The Restricted Isometry Condition can be viewed as an abstract form of the Uniform Uncertainty Principle of harmonic analysis ([9], see also [5] and [17]). Many natural ensembles of random matrices, such as partial Fourier, Bernoulli and Gaussian, satisfy the Restricted Isometry condition with parameters $n \ge 1$, $\varepsilon \in (0, 1/2)$ provided that

$$N = n\varepsilon^{-O(1)} \log^{O(1)} d;$$

see e.g. Section 2 of [20] and the references therein. Therefore, a computationally tractable exact recovery of sparse signals is possible with the number of measurements N roughly proportional to the sparsity level n, which is usually much smaller than the dimension d.

B. Exact recovery by greedy algorithms

An important alternative to convex programming is greedy algorithms, which have roots in Approximation Theory. A greedy algorithm computes the support of v iteratively, at each step finding one or more new elements (based on some "greedy" rule) and subtracting their contribution from the measurement vector x. The greedy rules vary. The simplest rule is to pick a coordinate of $\Phi^* x$ of the biggest magnitude; this defines the well known greedy algorithm called Orthogonal Matching Pursuit (OMP), known otherwise as Orthogonal Greedy Algorithm (OGA) [23].

Greedy methods are usually fast and easy to implement. For example, given $N \ge Cn\log(d/\delta)$ measurements with

D. Needell is with the Dept. of Mathematics, University of California, Davis, One Sheilds Ave., Davis CA 95616, USA e-mail: dneedell@math.ucdavis.edu.

R. Vershynin is with the Dept. of Mathematics, University of Michigan, 530 Church St., Ann Arbor MI 48105, USA email: romanv@umich.edu.

Partially supported by the Alfred P. Sloan Foundation and by NSF DMS grant 0652617

$\delta \in (0, 0.36)$, OMP succeeds in just n iterations except with probability δ [23]. Since each iteration amounts to solving one least-squares problem, its running time is always polynomial in n, N and d. Much promising work has been done on the complexity of linear programming techniques [1], and their applications to compressed sensing (see [15], [2], [24]). However, the greedy approach may still be more practical for many applications. For more discussion, see [23] and [20].

A variant of OMP was recently found in [20] that has guarantees similar to those of convex programming methods, with only an added theoretical logarithmic factor.[1] This greedy algorithm is called Regularized Orthogonal Matching Pursuit (ROMP); we state it in Section I-C below. Under the Restricted Isometry Condition with parameters $(2n, 0.03/\sqrt{\log n})$, ROMP exactly recovers an n-sparse signal v from its measurements x. Since this paper was written, other algorithms have been developed that also provide strong guarantees, and even without the logarithmic factor. See the remark at the end of Section IV for details.

Summarizing, *the restricted isometry principle is a guarantee for efficient sparse recovery; one can provably use either convex programming methods* (I.1) *or greedy algorithms (ROMP).*

C. Stable recovery by convex programming and greedy algorithms

A more realistic scenario is where the measurements are inaccurate (e.g. contaminated by noise) and the signals are not exactly sparse. In most situations that arise in practice, one cannot hope to know the measurement vector $x = \Phi v$ with arbitrary precision. Instead, it is perturbed by a small error vector: $x = \Phi v + e$. Here the vector e has unknown coordinates as well as unknown magnitude, and it needs not be sparse (as all coordinates may be affected by the noise). For a recovery algorithm to be stable, it should be able to approximately recover the original signal v from these perturbed measurements.

The stability of convex optimization algorithms for sparse recovery was studied in [12], [22], [13], [6]. Assuming that one knows a bound on the magnitude of the error, $\|e\| \le \delta$, and that the measurement matrix Φ has sufficiently small restricted isometry constants, it was shown in [6] that the solution $\hat{v}$ of the convex program

$$\min \|u\|_1 \qquad \text{subject to} \qquad \|\Phi u - x\|_2 \le \delta \qquad \text{(I.2)}$$

is a good approximation to the unknown signal: $\|v - \hat{v}\|_2 \le C\delta$.

In contrast, the stability of greedy algorithms for sparse recovery has not been well understood until recently. Numerical evidence [13] suggests that OMP should be less stable than the convex program (I.2), but no theoretical results have been known in either the positive or negative direction. The present paper seeks to remedy this situation.

We prove that the bound for the stability of ROMP has the same form as that of the convex program (I.2), up to a logarithmic factor. Although the logarithmic factor produces stronger requirements for the restricted isometry condition of the measurement matrix, we speculate that this factor is only an artifact of our proofs. This result essentially bridges a gap between convex programming and greedy approaches to sparse recovery.

[1] OMP itself does not have such strong guarantees, see [21].

REGULARIZED ORTHOGONAL MATCHING PURSUIT (ROMP)

INPUT: Measurement vector $x \in \mathbb{R}^N$ and sparsity level n
OUTPUT: Index set $I \subset \{1, \dots, d\}$, reconstructed vector $\hat{v} = y$

Initialize: Let the index set $I = \emptyset$ and the residual $r = x$. Repeat the following steps n times or until $|I| \ge 2n$:

Identify: Choose a set J of the n biggest nonzero coordinates in magnitude of the observation vector $u = \Phi^* r$, or all of its nonzero coordinates, whichever set is smaller.

Regularize: Among all subsets $J_0 \subset J$ with comparable coordinates:

$$|u(i)| \le 2|u(j)| \quad \text{for all } i, j \in J_0,$$

choose J_0 with the maximal energy $\|u|_{J_0}\|_2$.

Update: Add the set J_0 to the index set: $I \leftarrow I \cup J_0$, and update the residual:

$$y = \operatorname*{argmin}_{z \in \mathbb{R}^I} \|x - \Phi z\|_2; \qquad r = x - \Phi y.$$

Notation. Here and throughout we write $f|_T$ to denote the vector f restricted to the coordinates indexed by T.

Remark. The algorithm requires some knowledge about the sparsity level n, and there are several ways to estimate this parameter. One such way is to conduct empirical studies using various sparsity levels and select the level which minimizes $\|\Phi\hat{v} - x\|_2$ for the output $\hat{v}$. Testing sparsity levels from a geometric progression, for example, would not contribute significantly to the overall runtime.

Theorem 1.2 (Stability under measurement perturbations): Let Φ be a measurement matrix satisfying the Restricted Isometry Condition with parameters $(4n, \varepsilon)$ for $\varepsilon = 0.01/\sqrt{\log n}$. Let $v \in \mathbb{R}^d$ be an n-sparse vector. Suppose that the measurement vector Φv becomes corrupted, so that we consider $x = \Phi v + e$ where e is some error vector. Then ROMP produces an approximation to v that satisfies:

$$\|v - \hat{v}\|_2 \le 104\sqrt{\log n}\|e\|_2. \qquad \text{(I.3)}$$

Note that in the noiseless situation ($e = 0$) the reconstruction is exact: $\hat{v} = v$. This case of Theorem 1.2 was proved in [20].

Our stability result extends naturally to the even more realistic scenario where the signals are only approximately sparse. Here and henceforth, denote by f_m the vector of the m biggest coefficients in absolute value of f.

Corollary 1.3 (Stability of ROMP under signal perturbations): Let Φ be a measurement matrix satisfying the Restricted Isometry Condition with parameters $(8n, \varepsilon)$ for $\varepsilon = 0.01/\sqrt{\log n}$. Consider an arbitrary vector v in $\mathbb{R}^d$. Suppose that the measurement vector Φv becomes corrupted, so we consider $x = \Phi v + e$ where e is some error vector. Then ROMP produces an approximation to v_{2n} that satisfies:

$$\|\hat{v} - v_{2n}\|_2 \leq 159\sqrt{\log 2n}\Big(\|e\|_2 + \frac{\|v - v_n\|_1}{\sqrt{n}}\Big). \qquad \text{(I.4)}$$

Remarks. **1.** The term v_{2n} in the corollary can be replaced by $v_{(1+\delta)n}$ for any $\delta > 0$. This change will only affect the constant terms in the corollary.

2. We can apply Corollary 1.3 to the largest $2n$ coordinates of v and use Lemma 3.1 below to produce an error bound for the entire vector v. Along with the triangle inequality and the identity $v - v_{2n} = (v - v_n) - (v - v_n)_n$, these results yield:

$$\|\hat{v} - v\|_2 \leq 160\sqrt{\log 2n}\Big(\|e\|_2 + \frac{\|v - v_n\|_1}{\sqrt{n}}\Big). \qquad \text{(I.5)}$$

3. For the convex programming method (I.2), the stability bound (I.5) was proved in [6], and even without the logarithmic factor. We conjecture that this factor is also not needed in our results for ROMP.

4. Unlike the convex program (I.2), ROMP succeeds with absolutely no prior knowledge about the error e; its magnitude can be arbitrary. ROMP does however, require knowledge about the sparsity level n. Although often these parameters may be related, it may be more natural to impose sparsity awareness in some applications.

5. One can use ROMP to approximately compute a $2n$-sparse vector that is close to *the best* $2n$*-term approximation* v_{2n} of an arbitrary signal v. To this end, one just needs to retain the $2n$ biggest coordinates of $\hat{v}$. Indeed, Corollary 3.2 below shows that the best $2n$-term approximations of the original and the reconstructed signals satisfy:

$$\|v_{2n} - \hat{v}_{2n}\|_2 \leq 477\sqrt{\log 2n}\Big(\|e\|_2 + \frac{\|v - v_n\|_1}{\sqrt{n}}\Big).$$

6. An important special case of Corollary 1.3 is for the class of compressible vectors, which is a common model in signal processing, see [9], [11]. Suppose v is a compressible vector in the sense that its coefficients obey a power law: for some $p > 1$, the k-th largest coefficient in magnitude of v is bounded by $C_p k^{-p}$. Then (I.5) yields the following bound on the reconstructed signal:

$$\|v - \hat{v}\|_2 \leq C'_p \frac{\sqrt{\log n}}{n^{p-1/2}} + C''\sqrt{\log n}\|e\|_2. \qquad \text{(I.6)}$$

As observed in [6], without the logarithmic factor this bound would be optimal; no algorithm can perform fundamentally better.

The rest of the paper has the following organization. In Section II, we prove our main result, Theorem 1.2. In Section III, we deduce the extension for approximately sparse signals, Corollary 1.3, and a consequence for best n-term approximations, Corollary 3.2. In Section IV, we demonstrate some numerical experiments that illustrate the stability of ROMP.

Acknowledgment

We would like to thank the referees for a thorough reading of the manuscript and their comments and corrections. We would also like to thank Jared Tanner at the University of Edinburgh for his suggestions which helped improve the paper.

II. Proof of Theorem 1.2

We begin by showing that at every iteration of ROMP, either at least 50% of the selected coordinates from that iteration are from the support of the actual signal v, or the error bound already holds. This directly implies Theorem 1.2.

Theorem 2.1 (Stable Iteration Invariant of ROMP): Let Φ be a measurement matrix satisfying the Restricted Isometry Condition with parameters $(4n, \varepsilon)$ for $\varepsilon = 0.01/\sqrt{\log n}$. Let v be a non-zero n-sparse vector with measurements $x = \Phi v + e$. Then at any iteration of ROMP, after the regularization step where I is the current chosen index set, we have $J_0 \cap I = \emptyset$ and (at least) one of the following:

(i) $|J_0 \cap \text{supp}(v)| \geq \frac{1}{2}|J_0|$;
(ii) $\|v|_{\text{supp}(v)\setminus I}\|_2 \leq 100\sqrt{\log n}\|e\|_2$.

We show that the Iteration Invariant implies Theorem 1.2 by examining the three possible cases:

Case 1: (ii) occurs at some iteration. We first note that since $|I|$ is nondecreasing, if (ii) occurs at some iteration, then it holds for all subsequent iterations. To show that this would then imply Theorem 1.2, we observe that by the Restricted Isometry Condition and since $|\text{supp}(\hat{v})| \leq |I| \leq 3n$,

$$(1 - \varepsilon)\|\hat{v} - v\|_2 - \|e\|_2 \leq \|\Phi\hat{v} - \Phi v - e\|_2.$$

Then again by the Restricted Isometry Condition and definition of $\hat{v}$,

$$\|\Phi\hat{v} - \Phi v - e\|_2 \leq \|\Phi(v|_I) - \Phi v - e\|_2 \leq (1+\varepsilon)\|v|_{\text{supp}(v)\setminus I}\|_2 + \|e\|_2.$$

Thus we have that

$$\|\hat{v} - v\|_2 \leq \frac{1+\varepsilon}{1-\varepsilon}\|v|_{\text{supp}(v)\setminus I}\|_2 + \frac{2}{1-\varepsilon}\|e\|_2.$$

Thus (ii) of the Iteration Invariant would imply Theorem 1.2.

Case 2: (i) occurs at every iteration and J_0 is always non-empty. In this case, by (i) and the fact that J_0 is always non-empty, the algorithm identifies at least one element of the support in every iteration. Thus if the algorithm runs n iterations or until $|I| \geq 2n$, it must be that $\text{supp}(v) \subset I$, meaning that $v|_{\text{supp}(v)\setminus I} = 0$. Then by the argument above for Case 1, this implies Theorem 1.2.

Case 3: (i) occurs at each iteration and $J_0 = \emptyset$ for some iteration. By the definition of J_0, if $J_0 = \emptyset$ then $u = \Phi^* r = 0$ for that iteration. By definition of r, this must mean that

$$\Phi^*\Phi(v - y) + \Phi^* e = 0.$$

This combined with Part 1 of Proposition 2.2 below (and its proof, see [20]) applied with the set $I' = \text{supp}(v) \cup I$ yields

$$\|v - y + (\Phi^* e)|_{I'}\|_2 \leq 2.03\varepsilon\|v - y\|_2.$$

Then combinining this with Part 2 of the same Proposition, we have

$$\|v - y\|_2 \leq 1.1\|e\|_2.$$

Since $v|_{\text{supp}(v)\setminus I} = (v - y)|_{\text{supp}(v)\setminus I}$, this means that the error bound (ii) must hold, so by Case 1 this implies Theorem 1.2.

We now turn to the proof of the Iteration Invariant, Theorem 2.1. We will use the following proposition from [20].

Proposition 2.2 (Consequences of the RIC [20]): Assume a measurement matrix Φ satisfies the Restricted Isometry Condition with parameters $(2n, \varepsilon)$. Then the following holds.

1) *(Local approximation)* For every n-sparse vector $v \in \mathbb{R}^d$ and every set $I \subset \{1, \dots, d\}$, $|I| \le n$, the observation vector $u = \Phi^* \Phi v$ satisfies
$$\|u|_I - v|_I\|_2 \le 2.03\varepsilon \|v\|_2.$$
2) *(Spectral norm)* For any vector $z \in \mathbb{R}^N$ and every set $I \subset \{1, \dots, d\}$, $|I| \le 2n$, we have
$$\|(\Phi^* z)|_I\|_2 \le (1+\varepsilon)\|z\|_2.$$
3) *(Almost orthogonality of columns)* Consider two disjoint sets $I, J \subset \{1, \dots, d\}$, $|I \cup J| \le 2n$. Let P_I, P_J denote the orthogonal projections in $\mathbb{R}^N$ onto $\mathrm{range}(\Phi_I)$ and $\mathrm{range}(\Phi_J)$, respectively. Then
$$\|P_I P_J\|_{2\to 2} \le 2.2\varepsilon.$$

We prove Theorem 2.1 by inducting on each iteration of ROMP. We will show that at each iteration the set of chosen indices is disjoint from the current set I of indices, and that either (i) or (ii) holds. Clearly if (ii) held in a previous iteration, it would hold in all future iterations. Thus we may assume that (ii) has not yet held. Since (i) has held at each previous iteration, we must have
$$|I| \le 2n. \tag{II.1}$$

Consider an iteration of ROMP, and let $r \ne 0$ be the residual at the start of that iteration. Let J_0 and J be the sets found by ROMP in this iteration. As in [20], we consider the subspace
$$H := \mathrm{range}(\Phi_{\mathrm{supp}(v)\cup I})$$
and its complementary subspaces
$$F := \mathrm{range}(\Phi_I), \quad E_0 := \mathrm{range}(\Phi_{\mathrm{supp}(v)\setminus I}).$$
Part 3 of Proposition 2.2 states that the subspaces F and E_0 are nearly orthogonal. For this reason we consider the subspace:
$$E := F^\perp \cap H.$$

First we write the residual r in terms of projections onto these subspaces.

Lemma 2.3 (Residual): Here and onward, denote by P_L the orthogonal projection in $\mathbb{R}^N$ onto a linear subspace L. Then the residual r has the following form:
$$r = P_E \Phi v + P_{F^\perp} e.$$

Proof: By definition of the residual r in the ROMP algorithm, $r = P_{F^\perp} x = P_{F^\perp}(\Phi v + e)$. To complete the proof we need that $P_{F^\perp}\Phi v = P_E \Phi v$. This follows from the orthogonal decomposition $H = F + E$ and the fact that $\Phi v \in H$. ∎

Next we examine the missing portion of the signal as well as its measurements:
$$v_0 := v|_{\mathrm{supp}(v)\setminus I}, \quad x_0 := \Phi v_0 \in E_0. \tag{II.2}$$

In the next two lemmas we show that the subspaces E and E_0 are indeed close.

Lemma 2.4 (Approximation of the residual): Let r be the residual vector and x_0 as in (II.2). Then
$$\|x_0 - r\|_2 \le 2.2\varepsilon\|x_0\|_2 + \|e\|_2.$$

Proof: Since $v - v_0$ has support in I, we have $\Phi v - x_0 = \Phi(v - v_0) \in F$. Then by Lemma 2.3, $r = P_E \Phi v + P_{F^\perp} e = P_E x_0 + P_{F^\perp} e$. Therefore,
$$\|x_0 - r\|_2 = \|x_0 - P_E x_0 - P_{F^\perp} e\|_2 \le \|P_F x_0\|_2 + \|e\|_2.$$
Note that by (II.1), the union of the sets I and $\mathrm{supp}(v) \setminus I$ has cardinality no greater than $3n$. Thus by Part 3 of Proposition 2.2, we have
$$\|P_F x_0\|_2 + \|e\|_2 = \|P_F P_{E_0} x_0\|_2 + \|e\|_2 \le 2.2\varepsilon\|x_0\|_2 + \|e\|_2.$$
∎

Lemma 2.5 (Approximation of the observation): Let $u_0 = \Phi^* x_0$ and $u = \Phi^* r$. Then for any set $T \subset \{1, \dots, d\}$ with $|T| \le 3n$,
$$\|(u_0 - u)|_T\|_2 \le 2.4\varepsilon\|v_0\|_2 + (1+\varepsilon)\|e\|_2.$$

Proof: By Lemma 2.4 and the Restricted Isometry Condition we have
$$\begin{aligned}\|x_0 - r\|_2 &\le 2.2\varepsilon\|\Phi v_0\|_2 + \|e\|_2\\ &\le 2.2\varepsilon(1+\varepsilon)\|v_0\|_2 + \|e\|_2\\ &\le 2.3\varepsilon\|v_0\|_2 + \|e\|_2.\end{aligned}$$
Then by Part 2 of Proposition 2.2 we have the desired result,
$$\|(u_0 - u)|_T\|_2 \le (1+\varepsilon)\|x_0 - r\|_2.$$
∎

The result of the theorem requires us to show that we correctly gain a portion of the support of the signal v. To this end, we first show that ROMP correctly chooses a portion of the energy. The regularization step will then imply that the support is also selected correctly. We thus next show that the energy of u when restricted to the sets J and J_0 is sufficiently large.

Lemma 2.6 (Localizing the energy): Let u be the observation vector and v_0 be as in (II.2). Then $\|u|_J\|_2 \ge 0.8\|v_0\|_2 - (1+\varepsilon)\|e\|_2$.

Proof: Let $S = \mathrm{supp}(v) \setminus I$ be the missing support. Since $|S| \le n$, by definition of J in the algorithm, we have
$$\|u|_J\|_2 \ge \|u|_S\|_2.$$
By Lemma 2.5,
$$\|u|_S\|_2 \ge \|u_0|_S\|_2 - 2.4\varepsilon\|v_0\|_2 - (1+\varepsilon)\|e\|_2.$$
Since $v_0|_S = v_0$, Part 1 of Proposition 2.2 implies
$$\|u_0|_S\|_2 \ge (1 - 2.03\varepsilon)\|v_0\|_2.$$
These three inequalities yield
$$\begin{aligned}\|u|_J\|_2 &\ge (1 - 2.03\varepsilon)\|v_0\|_2 - 2.4\varepsilon\|v_0\|_2 - (1+\varepsilon)\|e\|_2\\ &\ge 0.8\|v_0\|_2 - (1+\varepsilon)\|e\|_2.\end{aligned}$$
This completes the proof. ∎

Lemma 2.7 (Regularizing the energy): Again let u be the observation vector and v_0 be as in (II.2). Then

$$\|u|_{J_0}\|_2 \geq \frac{1}{4\sqrt{\log n}}\|v_0\|_2 - \frac{\|e\|_2}{2\sqrt{\log n}}.$$

Proof: By Lemma 3.7 of [20] applied to the vector $u|_J$, we have

$$\|u|_{J_0}\|_2 \geq \frac{1}{2.5\sqrt{\log n}}\|u|_J\|_2.$$

Along with Lemma 2.6 this implies the claim. ■

We now conclude the proof of Theorem 2.1. The claim that $J_0 \cap I = \emptyset$ follows by the same arguments as in [20].

It remains to show its last claim, that either (i) or (ii) holds. Suppose (i) in the theorem fails. That is, suppose $|J_0 \cap \operatorname{supp}(v)| < \frac{1}{2}|J_0|$, which means

$$|J_0 \backslash \operatorname{supp}(v)| > \frac{1}{2}|J_0|.$$

Set $\Lambda = J_0 \backslash \operatorname{supp}(v)$. Since $|\Lambda| > \frac{1}{2}|J_0|$ and all coordinates of u in J_0 are within a factor of 2 of each other, we have

$$\|u|_{J_0 \cap \operatorname{supp}(v)}\|_2^2 < 4\|u|_\Lambda\|_2^2.$$

Since $\|u|_\Lambda\|_2^2 + \|u|_{J_0 \cap \operatorname{supp}(v)}\|_2^2 = \|u|_{J_0}\|_2^2$, this implies

$$\|u|_\Lambda\|_2 > \frac{1}{\sqrt{5}}\|u|_{J_0}\|_2.$$

Thus by Lemma 2.7,

$$\|u|_\Lambda\|_2 > \frac{1}{4\sqrt{5\log n}}\|v_0\|_2 - \frac{\|e\|_2}{2\sqrt{5\log n}}. \qquad \text{(II.3)}$$

Next, we also have

$$\|u|_\Lambda\|_2 \leq \|u|_\Lambda - u_0|_\Lambda\|_2 + \|u_0|_\Lambda\|_2. \qquad \text{(II.4)}$$

Since $\Lambda \subset J$ and $|J| \leq n$, by Lemma 2.5 we have

$$\|u|_\Lambda - u_0|_\Lambda\|_2 \leq 2.4\varepsilon\|v_0\|_2 + (1+\varepsilon)\|e\|_2.$$

By the definition of v_0 in (II.2), it must be that $v_0|_\Lambda = 0$. Thus by Part 1 of Proposition 2.2,

$$\|u_0|_\Lambda\|_2 \leq 2.03\varepsilon\|v_0\|_2.$$

Using the previous inequalities along with (II.4), we deduce that

$$\|u|_\Lambda\|_2 \leq 4.43\varepsilon\|v_0\|_2 + (1+\varepsilon)\|e\|_2.$$

This is a contradiction to (II.3) whenever

$$\varepsilon \leq \frac{0.02}{\sqrt{\log n}} - \frac{\|e\|_2}{\|v_0\|_2}.$$

If this is true, then indeed (i) in the theorem must hold. If it is not true, then by the choice of ε, this implies that

$$\|v_0\|_2 \leq 100\|e\|_2\sqrt{\log n}.$$

This proves Theorem 2.1. Next we turn to the proof of Corollary 1.3. ■

III. Approximately sparse vectors and best n-term approximations

A. Proof of Corollary 1.3

We first partition v so that $x = \Phi v_{2n} + \Phi(v - v_{2n}) + e$. Then since Φ satisfies the Restricted Isometry Condition with parameters $(8n, \varepsilon)$, by Theorem 1.2 and the triangle inequality,

$$\|v_{2n} - \hat{v}\|_2 \leq 104\sqrt{\log 2n}(\|\Phi(v - v_{2n})\|_2 + \|e\|_2), \quad \text{(III.1)}$$

The following lemma as in [16] relates the 2-norm of a vector's tail to its 1-norm. An application of this lemma combined with (III.1) will prove Corollary 1.3.

Lemma 3.1 (Comparing the norms): Let $w \in \mathbb{R}^d$, and let w_m be the vector of the m largest coordinates in absolute value from w. Then

$$\|w - w_m\|_2 \leq \frac{\|w\|_1}{2\sqrt{m}}.$$

Proof: By linearity, we may assume $\|w\|_1 = d$. Since w_m consists of the largest m coordinates of w in absolute value, we must have that $\|w - w_m\|_2 \leq \sqrt{d-m}$. (This is because the term $\|w - w_m\|_2$ is greatest when the vector w has constant entries.) Then by the arithmetic mean-geometric mean (AM-GM) inequality,

$$\|w - w_m\|_2\sqrt{m} \leq \sqrt{d-m}\sqrt{m} \leq (d-m+m)/2 = d/2 = \|w\|_1/2.$$

■

By Lemma 29 of [16], we have

$$\|\Phi(v - v_{2n})\|_2 \leq (1+\varepsilon)\Big(\|v - v_{2n}\|_2 + \frac{\|v - v_{2n}\|_1}{\sqrt{n}}\Big).$$

Applying Lemma 3.1 to the vector $w = v - v_n$ we then have

$$\|\Phi(v - v_{2n})\|_2 \leq 1.5(1+\varepsilon)\frac{\|v - v_n\|_1}{\sqrt{n}}.$$

Combined with (III.1), this proves the corollary.

B. Best n-term approximation

Often one wishes to find a *sparse* approximation to a signal. We now show that by simply truncating the reconstructed vector, a similar error bound still holds.

Corollary 3.2: Assume a measurement matrix Φ satisfies the Restricted Isometry Condition with parameters $(8n, \varepsilon)$ for $\varepsilon = 0.01/\sqrt{\log n}$. Let v be an arbitrary vector in $\mathbb{R}^d$, let $x = \Phi v + e$ be the measurement vector, and $\hat{v}$ the reconstructed vector output by the ROMP Algorithm. Then

$$\|v_{2n} - \hat{v}_{2n}\|_2 \leq 477\sqrt{\log 2n}\Big(\|e\|_2 + \frac{\|v - v_n\|_1}{\sqrt{n}}\Big),$$

where z_m denotes the best m-sparse approximation to z (i.e. the vector consisting of the largest m coordinates in absolute value).

Proof: Let $v_S := v_{2n}$ and $\hat{v}_T := \hat{v}_{2n}$, and let S and T denote the supports of v_S and $\hat{v_T}$ respectively. By Corollary 1.3, it suffices to show that $\|v_S - \hat{v}_T\|_2 \leq 3\|v_S - \hat{v}\|_2$.

Applying the triangle inequality, we have

$$\|v_S - \hat{v}_T\|_2 \leq \|(v_S - \hat{v}_T)|_T\|_2 + \|v_S|_{S\backslash T}\|_2 =: a + b.$$

We then have

$$a = \|(v_S - \hat{v}_T)|_T\|_2 = \|(v_S - \hat{v})|_T\|_2 \le \|v_S - \hat{v}\|_2$$

and

$$b \le \|\hat{v}|_{S\setminus T}\|_2 + \|(v_S - \hat{v})|_{S\setminus T}\|_2.$$

Since $|S| = |T|$, we have $|S\setminus T| = |T\setminus S|$. By the definition of T, every coordinate of $\hat{v}$ in T is greater than or equals to every coordinate of $\hat{v}$ in T^c in absolute value. Thus we have,

$$\|\hat{v}|_{S\setminus T}\|_2 \le \|\hat{v}|_{T\setminus S}\|_2 = \|(v_S - \hat{v})|_{T\setminus S}\|_2.$$

Thus $b \le 2\|v_S - \hat{v}\|_2$, and so

$$a + b \le 3\|v_S - \hat{v}\|_2.$$

This completes the proof. ∎

Remark. Corollary 3.2 combined with Corollary 1.3 and (I.5) implies that we can also estimate a bound on the whole signal v:

$$\|v - \hat{v}_{2n}\|_2 \le C\sqrt{\log 2n}\Big(\|e\|_2 + \frac{\|v - v_n\|_1}{\sqrt{n}}\Big).$$

IV. Numerical Examples

This section describes our numerical experiments that illustrate the stability of ROMP. We study the recovery error using ROMP for both perturbed measurements and signals. The empirical recovery error is actually much better than that given in the theorems.

First we describe the setup to our experimental studies. We run ROMP on various values of the ambient dimension d, the number of measurements N, and the sparsity level n, and attempt to reconstruct random signals. For each set of parameters, we perform 500 trials. Initially, we generate an $N \times d$ Gaussian measurement matrix Φ. For each trial, independent of the matrix, we generate an n-sparse signal v by choosing n components uniformly at random and setting them to one. In the case of perturbed signals, we add to the signal a d-dimensional error vector with Gaussian entries. In the case of perturbed measurements, we add an N-dimensional error vector with Gaussian entries to the measurement vector Φv. We then execute ROMP with the measurement vector $x = \Phi v$ or $x + e$ in the perturbed measurement case. After ROMP terminates, we output the reconstructed vector $\hat{v}$ obtained from the least squares calculation and calculate its distance from the original signal.

Figure 1 depicts the recovery error $\|v - \hat{v}\|_2$ when ROMP was run with perturbed measurements. This plot was generated with $d = 256$ for various levels of sparsity n. The horizontal axis represents the number of measurements N, and the vertical axis represents the average normalized recovery error. Figure 1 confirms the results of Theorem 1.2, while also suggesting that at least for typical signals the bound (I.3) given by the theorem appears to be satisfied without the $\sqrt{\log n}$ factor.

Figure 2 depicts the normalized recovery error when the signal was perturbed by a Gaussian vector. The figure confirms the results of Corollary 1.3 while also suggesting again that the logarithmic factor in the corollary is unnecessary.

Remark. Our work on ROMP has motivated the development of additional methods that indeed provide similar results but without the logarithmic factor. Compressive Sampling Matching Pursuit (CoSaMP) by Needell and Tropp and Subspace Pursuit (SP) by Dai and Milenkovic are greedy pursuits that incorporate ideas from ROMP, combinatorial algorithms, and convex optimization [19], [18], [10]. These improve upon the error bounds of ROMP by removing the logarithmic factor. In doing so, they lessen the requirements on the restricted isometry condition by this factor as well. The work on CoSaMP also analyzes the least squares step in the algorithm, showing how it can be done efficiently to the accuracy level needed to maintain the overall error bounds. With this analysis, the total runtime of CoSaMP is shown to be just $O(Nd)$. Recent work on thresholding algorithms such as Iterative Hard Thresholding (IHT) by Blumensath and Davies has also provided similar strong guarantees [3].

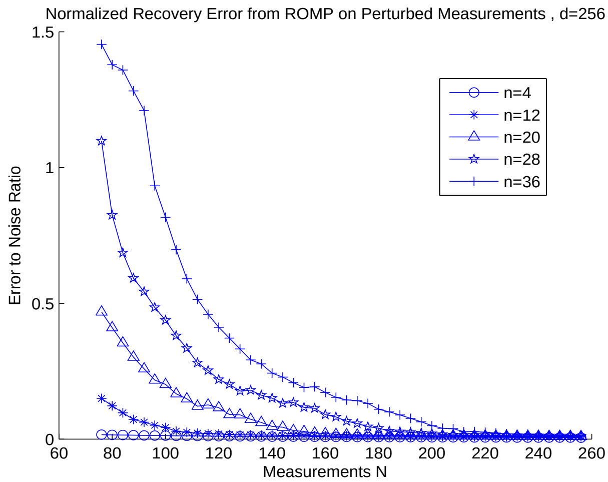

Fig. 1. The error to noise ratio $\frac{\|\hat{v}-v\|_2}{\|e\|_2}$ as a function of the number of measurements N in dimension $d = 256$ for various levels of sparsity n.

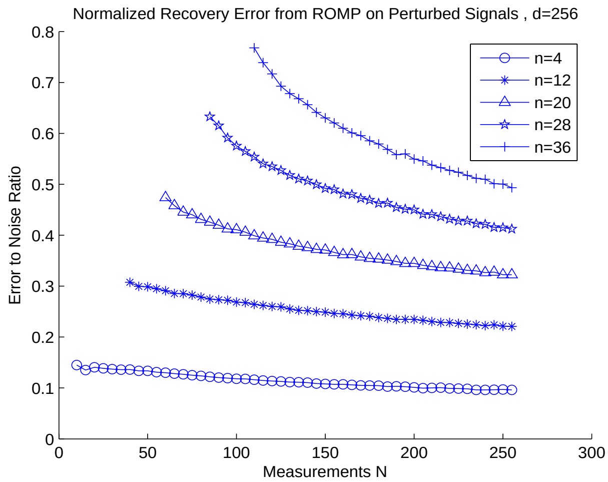

Fig. 2. The error to noise ratio $\frac{\|\hat{v}-v_{2n}\|_2}{\|v-v_n\|_1/\sqrt{n}}$ using a perturbed signal, as a function of the number of measurements N in dimension $d = 256$ for various levels of sparsity n.

REFERENCES

[1] K. M. Anstreicher. Linear programming in O$([n^3/\ln n]l)$ operations. *SIAM J. Optim.*, 9(4):803–812, 1999.

[2] E. van den Berg and M. P. Friedlander. In pursuit of a root. Tech. Rep. TR-2007-19, Department of Computer Science, University of British Columbia, Vancouver, June 2007.

[3] T. Blumensath and M. E. Davies. Iterative hard thresholding for compressed sensing. Preprint, Univ. of Edinburgh, 2009.

[4] E. Candès. Compressive sampling. In *Proc. International Congress of Mathematics*, volume 3, pages 1433–1452, Madrid, Spain, 2006.

[5] E. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete Fourier information. *IEEE Trans. Info. Theory*, 52(2):489–509, Feb. 2006.

[6] E. Candès, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, 59(8):1207–1223, 2006.

[7] E. Candès and T. Tao. Decoding by linear programming. *IEEE Trans. Inform. Theory*, 51:4203–4215, 2005.

[8] E. J. Candès. The restricted isometry property and its implications for compressed sensing. *C. R. Math. Acad. Sci. Paris, Serie I*, 346:589–592, 2008.

[9] E. J. Candès and T. Tao. Near optimal signal recovery from random projections: Universal encoding strategies? *IEEE Trans. Inform. Theory*, 52(12):5406–5425, Dec. 2006.

[10] W. Dai and O. Milenkovic. Subspace pursuit for compressive sensing: Closing the gap between performance and complexity. Submitted for publication, Mar. 2008.

[11] D. L. Donoho. Compressed sensing. *IEEE Trans. Info. Theory*, 52(4):1289–1306, Apr. 2006.

[12] D. L. Donoho. For most large underdetermined systems of linear equations the minimal l_1-norm solution is also the sparsest solution. *Comm. Pure Appl. Math.*, 59(6):797–829, 2006.

[13] D. L. Donoho, M. Elad, and V. N. Temlyakov. Stable recovery of sparse overcomplete representations in the presence of noise. Submitted to *IEEE Trans. Inform. Theory*, February 2004.

[14] David L. Donoho. Neighborly polytopes and sparse solutions of underdetermined linear equations. Technical report, 2005.

[15] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright. Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. *IEEE J. Selected Topics in Signal Processing: Special Issue on Convex Optimization Methods for Signal Processing*, 1(4):586–598, 2007.

[16] A. Gilbert, M. Strauss, J. Tropp, and R. Vershynin. One sketch for all: Fast algorithms for compressed sensing. In *Proc. 39th ACM Symp. Theory of Computing*, San Diego, June 2007.

[17] Yu. Lyubarskii and R. Vershynin. Uncertainty principles and vector quantization. Submitted for publication, 2008.

[18] D. Needell and J. A. Tropp. CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. ACM Technical Report 2008-01, California Institute of Technology, Pasadena, July 2008.

[19] D. Needell and J. A. Tropp. CoSaMP: Iterative signal recovery from noisy samples. *Appl. Comput. Harmon. Anal.*, 2008. DOI: 10.1016/j.acha.2008.07.002.

[20] D. Needell and R. Vershynin. Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit. *Found. Comput. Math.*, 2007. DOI: 10.1007/s10208-008-9031-3.

[21] H. Rauhut. On the impossibility of uniform sparse reconstruction using greedy methods. *Sampl. Theory Signal Image Process.*, 7(2):197–215, 2008.

[22] J. A. Tropp. Just relax: Convex programming methods for subset selection and sparse approximation. ICES Report 04-04, The University of Texas at Austin, 2004.

[23] J. A. Tropp and A. C. Gilbert. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Trans. Info. Theory*, 53(12):4655–4666, 2007.

[24] D. Goldfarb W. Yin, S. Osher and J. Darbon. Bregman iterative algorithms for compessed sensing and related problems. *SIAM Imaging Science*, 1(1):143–168, 2008.