4-1-2010

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

Deanna Needell
Claremont McKenna College

Roman Vershynin
University of California - Davis

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 close the gap between 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 Uniform Uncertainty 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.

1. INTRODUCTION

1.1. Exact recovery by convex programming.

The recent massive work in the area of Compressed Sensing, surveyed in [1], 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:

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

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. We then wish to 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. Candès and Tao [4] proved that under a stronger (quantitative) condition, the sparse recovery problem is equivalent to a convex program

$$(1.1) \qquad \min \|u\|_1 \qquad \text{subject to} \qquad \Phi u = x$$

and therefore is computationally tractable. This condition is that the map Φ is an almost isometry on the set of $O(n)$-sparse vectors:

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

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

Under the Restricted Isometry Condition with parameters $(3n, 0.2)$, the convex program (1.1) exactly recovers an n-sparse signal v from its measurements x [4].

The Restricted Isometry Condition can be viewed as an abstract form of the Uniform Uncertainty Principle of harmonic analysis ([5], see also [2] and [10]). Many natural ensembles of random matrices, such as partial Fourier, Bernoulli and Gaussian, satisfy the Restricted Isometry condition with parameters $n \geq 1$, $\varepsilon \in (0, 1/2)$ provided that

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

see e.g. Section 2 of [11] 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.

1.2. **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) [14].

Greedy methods are usually fast and easy to implement, which makes them popular with practitioners. For example, OMP needs just n iterations to find the support of an n-sparse signal v, and each iteration amounts to solving one least-squares problem; so its running time is always polynomial in n, N and d. In contrast, no known bounds are known on the running time of (1.1) as a linear program. Future work on customization of convex programming solvers for sparse recovery problems may change this picture, of course. For more discussion, see [14] and [11].

A variant of OMP was recently found in [11] that has guarantees essentially as strong as those of convex programming methods.[1] This greedy algorithm is called Regularized Orthogonal Matching Pursuit (ROMP); we state it in Section 1.3 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.

Summarizing, *the Uniform Uncertainty Principle is a guarantee for efficient sparse recovery; one can provably use either convex programming methods* (1.1) *or greedy algorithms (ROMP).*

---

[1]OMP itself does not have such strong guarantees, see [12].

1.3. **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 [7], [13], [8], [3]. Assuming that one knows a bound on the magnitude of the error, $\|e\| \leq \delta$, it was shown in [3] that the solution $\hat{v}$ of the convex program

$$(1.2) \qquad \min \|u\|_1 \qquad \text{subject to} \qquad \|\Phi u - x\|_2 \leq \delta$$

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

In contrast, the stability of greedy algorithms for sparse recovery has not been well understood. Numerical evidence [8] suggests that OMP should be less stable than the convex program (1.2), 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 *ROMP is as stable as the convex program* (1.2). This result essentially closes a gap between convex programming and greedy approaches to sparse recovery.

REGULARIZED ORTHOGONAL MATCHING PURSUIT (ROMP)

INPUT: Measurement vector $x \in \mathbb{R}^N$ and sparsity level n
OUTPUT: Index set $I \subset \{1, \ldots, d\}$, 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 $|I| \geq 2n$:
>> **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:
>>$$|u(i)| \leq 2|u(j)| \quad \text{for all } i, j \in J_0,$$
>> 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:
>>$$y = \operatorname*{argmin}_{z \in \mathbb{R}^I} \|x - \Phi z\|_2; \qquad r = x - \Phi y.$$

Theorem 1.2 (Stability under measurement perturbations)**.** *Assume a measurement matrix Φ satisfies the Restricted Isometry Condition with parameters $(4n, \varepsilon)$ for $\varepsilon = 0.01/\sqrt{\log n}$. Let v be an n-sparse vector 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 a good approximation to v:

$$\|v - \hat{v}\|_2 \le 104\sqrt{\log n}\|e\|_2.$$

Note that in the noiseless situation ($e = 0$) the reconstruction is exact: $\hat{v} = v$. This case of Theorem 1.2 was proved in [11].

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). *Assume a measurement matrix Φ satisfies 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 a good approximation to v_{2n}:*

$$(1.3) \qquad \|\hat{v} - v_{2n}\|_2 \le 159\sqrt{\log 2n}\Big(\|e\|_2 + \frac{\|v - v_n\|_1}{\sqrt{n}}\Big).$$

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.   By applying Corollary 1.3 to the largest $2n$ coordinates of v and using Lemma 3.1 below, we also have the error bound for the entire vector v:

$$(1.4) \qquad \|\hat{v} - v\|_2 \le 160\sqrt{\log 2n}\Big(\|e\|_2 + \frac{\|v - v_n\|_1}{\sqrt{n}}\Big).$$

3.   For the convex programming method (1.2), the stability bound (1.4) was proved in [3], 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 (1.2), ROMP succeeds with absolutely no prior knowledge about the error e; its magnitude can be arbitrary. In the terminology of [8], the convex programming approach needs to be "noise-aware" while ROMP needs not.

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 are close:

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

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 [5], [6]. 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 (1.4) yields the following bound on the reconstructed signal:

$$(1.5) \qquad \|v - \hat{v}\|_2 \le C_p' \frac{\sqrt{\log n}}{n^{p-1/2}} + C''\sqrt{\log n}\|e\|_2.$$

As observed in [3], this bound is optimal (within the logarithmic factor); no algorithm can perform fundamentally better.

The rest of the paper is organized as follows. In Section 2, we prove our main result, Theorem 1.2. In Section 3, we deduce the extension for approximately sparse signals, Corollary 1.3, and a consequence for best n-term approximations, Corollary 3.2. In Section 4, we demonstrate some numerical experiments that illustrate the stability of ROMP.

2. PROOF OF THEOREM 1.2

We shall prove a stronger version of Theorem 1.2, which states that *at every iteration* of ROMP, either at least 50% of the newly selected coordinates are from the support of the signal v, or the error bound already holds.

Theorem 2.1 (Iteration Invariant of ROMP). *Assume Φ satisfies the Restricted Isometry Condition with parameters $(4n, \varepsilon)$ for $\varepsilon = 0.01/\sqrt{\log n}$. Let $v \neq 0$ be an 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) $|J_0 \cap \mathrm{supp}(v)| \geq \frac{1}{2}|J_0|$;
(ii) $\||v|_{\mathrm{supp}(v) \setminus I}\|_2 \leq 100\sqrt{\log n}\|e\|_2$.

In other words, either at least 50% of the coordinates in the newly selected set J_0 belong to the support of v or the bound on the error already holds.

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 $|\mathrm{supp}(\hat{v})| \leq |I| \leq 3n$,

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

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

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

Thus we have that

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

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 $\mathrm{supp}(v) \subset I$, meaning that $v|_{\mathrm{supp}(v) \setminus I} = 0$. 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 [11]) applied with the set $I' = \operatorname{supp}(v) \cup I$ yields

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

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

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

Since $v|_{\operatorname{supp}(v)\setminus I} = (v - y)|_{\operatorname{supp}(v)\setminus I}$, 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 [11].

Proposition 2.2 (Consequences of Restricted Isometry Condition [11]). *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 $I \subset \{1, \ldots, d\}$, $|I| \le n$, 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 $I \subset \{1, \ldots, d\}$, $|I| \le 2n$, we have*

$$\|(\Phi^* z)|_I\|_2 \le (1 + \varepsilon)\|z\|_2.$$

(3) (Almost orthogonality of columns) *Consider two disjoint sets $I, J \subset \{1, \ldots, d\}$, $|I \cup J| \le 2n$. Let P_I, P_J denote the orthogonal projections in $\mathbb{R}^N$ onto $\operatorname{range}(\Phi_I)$ and $\operatorname{range}(\Phi_J)$, respectively. Then*

$$\|P_I P_J\|_{2 \to 2} \le 2.2\varepsilon.$$

The proof of Theorem 2.1 is by induction on the iteration of ROMP. The induction claim is that for all previous iterations, the set of newly chosen indices is disjoint from the set of previously chosen indices I, and 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

$$(2.1) \qquad\qquad\qquad\qquad |I| \le 2n.$$

Let $r \ne 0$ be the residual at the start of this iteration, and let J_0, J be the sets found by ROMP in this iteration. As in [11], we consider the subspace

$$H := \operatorname{range}(\Phi_{\operatorname{supp}(v)\cup I})$$

and its complementary subspaces

$$F := \operatorname{range}(\Phi_I), \quad E_0 := \operatorname{range}(\Phi_{\operatorname{supp}(v)\setminus I}).$$

The Restricted Isometry Condition in the form of Part 3 of Proposition 2.2 ensures that F and E_0 are almost orthogonal. Thus E_0 is close to the orthogonal complement of F in H,

$$E := F^\perp \cap H.$$

The residual r thus still has a simple description:

Lemma 2.3 (Residual). *Here and thereafter, let P_L denote the orthogonal projection in $\mathbb{R}^N$ onto a linear subspace L. Then*

$$r = P_E \Phi v + P_{F^\perp} e.$$

Proof. By definition of the residual in the 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 $\Phi v \in H$. $\qquad\square$

Now we consider the signal we seek to identify at the current iteration, and its measurements:

(2.2) $$v_0 := v|_{\mathrm{supp}(v) \setminus I}, \quad x_0 := \Phi v_0 \in E_0.$$

To guarantee a correct identification of v_0, we first state two approximation lemmas that reflect in two different ways the fact that subspaces E_0 and E are close to each other.

Lemma 2.4 (Approximation of the residual). *We have*

$$\|x_0 - r\|_2 \le 2.2\varepsilon \|x_0\|_2 + \|e\|_2.$$

Proof. By definition of F, we have $\Phi v - x_0 = \Phi(v - v_0) \in F$. Therefore, by Lemma 2.3, $r = P_E \Phi v + P_{F^\perp} e = P_E x_0 + P_{F^\perp} e$, and so

$$\|x_0 - r\|_2 = \|x_0 - P_E x_0 - P_{F^\perp} e\|_2 \le \|P_F x_0\|_2 + \|e\|_2.$$

Now we use Part 3 of Proposition 2.2 for the sets I and $\mathrm{supp}(v) \setminus I$ whose union has cardinality at most $3n$ by (2.1). It follows that

$$\|P_F x_0\|_2 + \|e\|_2 = \|P_F P_{E_0} x_0\|_2 + \|e\|_2 \le 2.2\varepsilon \|x_0\|_2 + \|e\|_2$$

as desired. $\qquad\square$

Lemma 2.5 (Approximation of the observation). *Consider the observation vectors $u_0 = \Phi^* x_0$ 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. Since $x_0 = \Phi v_0$, we have by Lemma 2.4 and the Restricted Isometry Condition that

$$\|x_0 - r\|_2 \le 2.2\varepsilon \|\Phi v_0\|_2 + \|e\|_2 \le 2.2\varepsilon(1 + \varepsilon)\|v_0\|_2 + \|e\|_2 \le 2.3\varepsilon \|v_0\|_2 + \|e\|_2.$$

To complete the proof, it remains to apply Part 2 of Proposition 2.2, which yields $\|(u_0 - u)|_T\|_2 \le (1 + \varepsilon)\|x_0 - r\|_2$. $\qquad\square$

We next show that the energy (norm) of u when restricted to J, and furthermore to J_0, is not too small. By the regularization step of ROMP, since all selected coefficients have comparable magnitudes, we will conclude that not only a portion of energy but also of the *support* is selected correctly, or else the error bound must already be attained. This will be the desired conclusion.

Lemma 2.6 (Localizing the energy). *We have* $\|u|_J\|_2 \geq 0.8\|v_0\|_2 - (1+\varepsilon)\|e\|_2$.

Proof. Let $S = \operatorname{supp}(v) \setminus I$. Since $|S| \leq n$, the maximality property of J in the algorithm implies that

$$\|u|_J\|_2 \geq \|u|_S\|_2.$$

By Lemma 2.5,

$$\|u|_S\|_2 \geq \|u_0|_S\|_2 - 2.4\varepsilon\|v_0\|_2 - (1+\varepsilon)\|e\|_2.$$

Furthermore, since $v_0|_S = v_0$, by Part 1 of Proposition 2.2 we have

$$\|u_0|_S\|_2 \geq (1 - 2.03\varepsilon)\|v_0\|_2.$$

Putting these three inequalities together, we conclude that

$$\|u|_J\|_2 \geq (1 - 2.03\varepsilon)\|v_0\|_2 - 2.4\varepsilon\|v_0\|_2 - (1+\varepsilon)\|e\|_2 \geq 0.8\|v_0\|_2 - (1+\varepsilon)\|e\|_2.$$

This proves the lemma.   □

We next bound the norm of u restricted to the smaller set J_0, again using the general property of regularization. In our context, Lemma 3.7 of [11] applied to the vector $u|_J$ yields

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

Along with Lemma 2.6 this directly implies:

Lemma 2.7 (Regularizing the energy). *We have*

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

We now finish the proof of Theorem 2.1. The claim that $J_0 \cap I = \emptyset$ follows by the same arguments as in [11].

The nontrivial part of the theorem is its last claim, that either (i) or (ii) holds. Suppose (i) in the theorem fails. Namely, suppose that $|J_0 \cap \operatorname{supp}(v)| < \frac{1}{2}|J_0|$, and thus

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

Set $\Lambda = J_0 \backslash \operatorname{supp}(v)$. By the comparability property of the coordinates in J_0 and since $|\Lambda| > \frac{1}{2}|J_0|$, there is a fraction of energy in Λ:

$$(2.3) \qquad \|u|_\Lambda\|_2 > \frac{1}{\sqrt{5}}\|u|_{J_0}\|_2 \geq \frac{1}{4\sqrt{5}\log n}\|v_0\|_2 - \frac{\|e\|_2}{2\sqrt{5}\log n},$$

where the last inequality holds by Lemma 2.7.

On the other hand, we can approximate u by u_0 as

(2.4) $$\|u|_\Lambda\|_2 \le \|u|_\Lambda - u_0|_\Lambda\|_2 + \|u_0|_\Lambda\|_2.$$

Since $\Lambda \subset J$, $|J| \le n$, and using Lemma 2.5, we have

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

Furthermore, by definition (2.2) of v_0, we have $v_0|_\Lambda = 0$. So, by Part 1 of Proposition 2.2,

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

Using the last two inequalities and (2.4), we conclude that

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

This is a contradiction to (2.3) so long as

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

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

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

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

3. Approximately sparse vectors and best n-term approximations

3.1. **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,

(3.1) $$\|v_{2n} - \hat{v}\|_2 \le 104\sqrt{\log 2n}(\|\Phi(v - v_{2n})\|_2 + \|e\|_2),$$

The following lemma as in [9] relates the 2-norm of a vector's tail to its 1-norm. An application of this lemma combined with (3.1) 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*

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

Proof. Let μ denote the $(n+1)th$ largest entry of w. If $\mu = 0$ then $w_m = 0$ so the claim holds. Thus we may assume this is not the case. Then we have

$$\frac{\|g - g_m\|_2}{\|g\|_1} \le \frac{\sqrt{\|g - g_m\|_1\|g - g_m\|_\infty}}{\|g_m\|_1 + \|g - g_m\|_1} \le \frac{\sqrt{m\mu^2}}{m\mu + m\mu}.$$

Simplifying gives the desired result. □

By Lemma 29 of [9], we have

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

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

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

Combined with (3.1), this proves the corollary.

3.2. **Best n-term approximation.** We now show that by truncating the reconstructed vector, we obtain a $2n$-sparse vector very close to the original signal.

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*

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

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 $\|v_S - \hat{v}_T\|_2 \le 3\|v_S - \hat{v}\|_2$.

Applying the triangle inequality, we have

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

We then have

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

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 equal 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. $\square$

Remark. Corollary 3.2 combined with Corollary 1.3 and (1.4) implies that we can also estimate a bound on the whole signal v:

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

4. Numerical Examples

This section describes our experiments that illustrate the stability of ROMP. We experimentally examine 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 of our experiments. For many values of the ambient dimension d, the number of measurements N, and the sparsity n, we reconstruct random signals using ROMP. For each set of values, 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 the bound may be improved by removing 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.
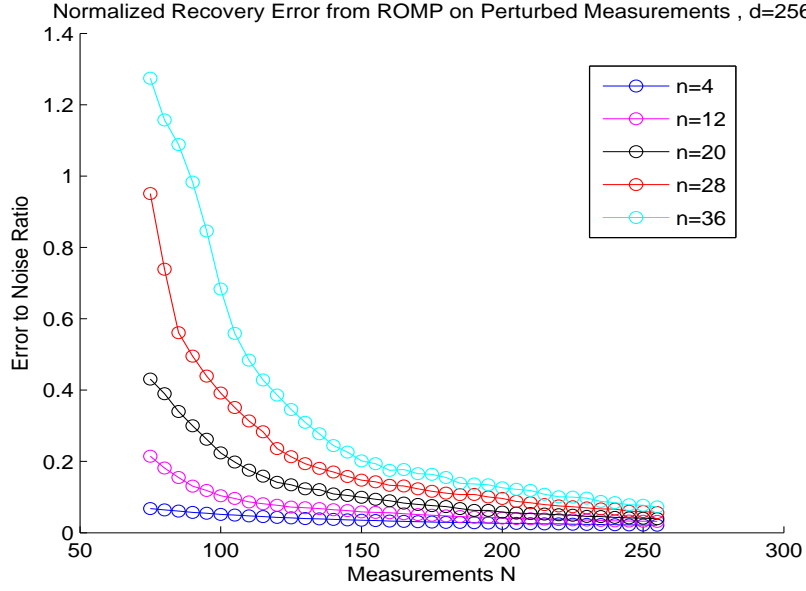
FIGURE 1. The error to noise ratio $\frac{\|\hat{v}-v\|_2}{\|e\|_2}$ as a function of the number of measurements N in dimension $d = 256$ for various levels of sparsity n.
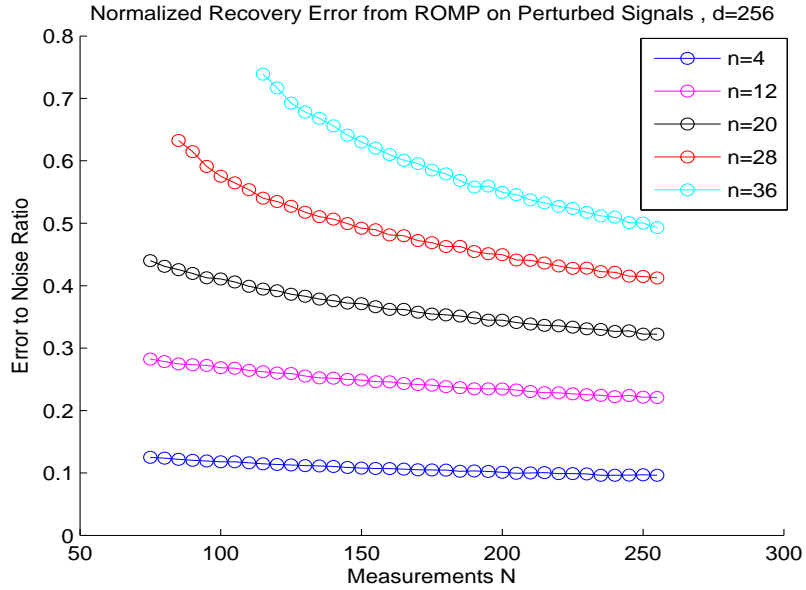

FIGURE 2. The error to noise ratio $\frac{\|\hat{v}-v_{2n}\|_2}{\|v-v_n\|_1/\sqrt{n}}$ using a perturbed signal, as a function of the number of measurements N in dimension $d = 256$ for various levels of sparsity n.

References

[1] E. Candès, *Compressive sampling*, Proc. International Congress of Mathematics, 3, pp. 1433-1452, Madrid, Spain, 2006.

[2] E. Candès, J. Romberg, T. Tao, *Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information*, IEEE Trans. Inform. Theory 52 (2006), 489-509.

[3] E. Candès, J. Romberg and T. Tao, *Stable signal recovery from incomplete and inaccurate measurements*, Comm. Pure Appl. Math. 59 (2006), 1207-1223.

[4] E. Candès, T. Tao, *Decoding by linear programming*, IEEE Trans. Inform. Theory 51 (2005), 4203–4215.

[5] E. Candès, T. Tao, *Near-optimal signal recovery from random projections: universal encoding strategies*, IEEE Trans. Inform. Theory 52 (2004), 5406–5425.

[6] D. Donoho, *Compressed sensing*, IEEE Trans. Inform. Theory 52 (2006), 1289–1306.

[7] D. Donoho, *For most large underdetermined systems of equations, the minimal l_1-norm near-solution approximates the sparsest near-solution*, Comm. Pure Appl. Math 59 (2006), 907–934

[8] D. Donoho, M. Elad, and V. Temlyakov, *Stable Recovery of Sparse Overcomplete Representations in the Presence of Noise*, IEEE Trans. Information Theory 52 (2006), 6-18.

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

[10] Yu. Lyubarskii, R. Vershynin, *Uncertainty principles and vector quantization*, submitted.

[11] D. Needell, R. Vershynin, *Uniform Uncertainty Principle and signal recovery via Regularized Orthogonal Matching Pursuit*, submitted.

[12] H. Rauhut, *On the impossibility of uniform recovery using greedy methods*, Sampling Theory in Signal and Image Processing, to appear

[13] J. Tropp, *Just relax: convex programming methods for identifying sparse signals in noise*, IEEE Trans. Inform. Theory 52 (2006), 1030–1051

[14] J. A. Tropp, A. C. Gilbert, *Signal recovery from random measurements via orthogonal matching pursuit*

Department of Mathematics, University of California, Davis, CA 95616, USA
E-mail address: {dneedell,vershynin}@math.ucdavis.edu