# Greedy Signal Recovery and Uncertainty Principles

Deanna Needell and Roman Vershynin

Department of Mathematics, University of California, Davis, CA 95616, USA

July 23, 2007

## ABSTRACT

This paper seeks to bridge the two major algorithmic approaches to sparse signal recovery from an incomplete set of linear measurements – $L_1$-minimization methods and iterative methods (Matching Pursuits). We find a simple regularized version of the Orthogonal Matching Pursuit (ROMP) which has advantages of both approaches: the speed and transparency of OMP and the strong uniform guarantees of the $L_1$-minimization. Our algorithm ROMP reconstructs a sparse signal in a number of iterations linear in the sparsity, and the reconstruction is exact provided the linear measurements satisfy the Uniform Uncertainty Principle. In the case of inaccurate measurements and approximately sparse signals, the noise level of the recovery is proportional to $\sqrt{\log n}\|e\|_2$ where $e$ is the error vector.

## 1. INTRODUCTION

Sparse recovery problems arise in many applications ranging from medical imaging to error correction. Suppose $v$ is an unknown $d$-dimensional signal with at most $n \ll d$ nonzero components:

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

We call such signals $n$-sparse. Suppose we are able to collect $N \ll d$ nonadaptive linear measurements of $v$, and wish to efficiently recover $v$ from these. The measurements are given as the vector $\Phi v \in \mathbb{R}^N$, where $\Phi$ is some $N \times d$ measurement matrix.

As discussed in,[2] exact recovery is possible with just $N = 2n$. However, recovery using only this property is not numerically feasible; the sparse recovery problem in general is known to be NP-hard. Nevertheless, a massive recent work in the emerging area of Compressed Sensing demonstrated that for several natural classes of measurement matrices $\Phi$, the signal $v$ can be exactly reconstructed from its measurements $\Phi v$ with

$$N = n \log^{O(1)}(d). \tag{1.1}$$

In other words, the number of measurements $N \ll d$ should be almost linear in the sparsity $n$. Survey[1] contains some of these results; the Compressed Sensing webpage[7] documents progress in this area.

The two major algorithmic approaches to sparse recovery are methods based on the $L_1$-minimization (see e.g.[3]) and iterative methods (such as Orthogonal Matching Pursuit, see[14]). We now briefly discuss the advantages and disadvantages of these methods. Then we propose a new iterative method that has advantages of both approaches.

Partially supported by the Alfred P. Sloan Foundation and by NSF DMS grants 0401032 and 0652617

Email: {dneedell,vershynin}@math.ucdavis.edu

### 1.1 Advantages and challenges of both approaches

The $L_1$-minimization has *strongest known guarantees* of sparse recovery. Once the measurement matrix $\Phi$ satisfies the Restricted Isometry Condition, this method works correctly for all sparse signals $v$.

Definition 1.1 (Restricted Isometry Condition). *A measurement matrix $\Phi$ satisfies the* Restricted Isometry Condition *(RIC) with parameters $(n, \varepsilon)$ for $\varepsilon \in (0, 1)$ if we have*

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

The Restricted Isometry Condition states that every set of $n$ columns of $\Phi$ forms approximately an orthonormal system. One can interpret the Restricted Isometry Condition as an abstract version of the Uniform Uncertainty Principle in harmonic analysis ([6] and also see discussions in[4] and[10]).

Also, the $L_1$-minimization is based on linear programming, and it is not very clear what this running time is, as there is no strongly polynomial time algorithm in linear programming yet.

Orthogonal Matching Pursuit is quite *fast*, both theoretically and experimentally. It makes $n$ iterations, where each iteration amounts to a multiplication by a $d \times N$ matrix $\Phi^*$ (computing the observation vector $u$), and solving a least squares problem in dimensions $N \times n$ (with matrix $\Phi_I$). This yields strongly polynomial running time. In practice, OMP is observed to perform faster and is easier to implement than $L_1$-minimization.[14] On the other hand, OMP has *weaker guarantees* of exact recovery. Unlike $L_1$-minimization, the guarantees of OMP are non-uniform: for each *fixed* sparse signal $v$ and not for *all* signals, the algorithm performs correctly with high probability. Moreover, OMP's condition on measurement matrices given in[14] is *more restrictive* than the Restricted Isometry Condition. In particular, it is not known whether OMP succeeds in the important class of partial Fourier measurement matrices.

### 1.2 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,[3],[15],[9].[5] Assuming that one knows a bound on the magnitude of the error, $\|e\| \le \delta$, it was shown in[5] that the solution $\hat{v}$ of the convex program

$$\min \|u\|_1 \qquad \text{subject to} \qquad \|\Phi u - x\|_2 \le \delta \tag{1.2}$$

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. Numerical evidence[9] 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.

### 1.3 Regularized OMP

This new algorithm for sparse recovery will perform correctly for all measurement matrices $\Phi$ satisfying the Restricted Isometry Condition, and for all sparse signals. It will also be stable, in the sense that small perturbations in the signal or measurements will not greatly affect recovery.

When we are trying to recover the signal $v$ from its measurements $x = \Phi v$, we can use the observation vector $u = \Phi^* x$ as a good *local approximation* to the signal $v$. Every $n$ coordinates of the signal $v$ look like correlations of the measurement vector $x$ with the almost orthonormal basis consisting of the columns of $\Phi$, and therefore

are close in the Euclidean sense to the coefficients of $v$ (see Proposition 2.2 below). To ensure that each of these coordinates gets an even share of information, we enforce a new regularization step. This leads to the following algorithm for stable sparse recovery:

REGULARIZED ORTHOGONAL MATCHING PURSUIT (ROMP)

INPUT: Measurement vector $x \in \mathbb{R}^N$ and sparsity level $n$
OUTPUT: Index set $I \subset \{1, \dots, d\}$, 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 $|I| \geq 2n$:

**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:

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

**Remark.** The identification and regularization steps of ROMP can be performed efficiently. In particular, the regularization step does *not* mean combinatorial complexity, but actually can be done in linear time. The running time of ROMP is thus comparable to that of OMP in theory, and is often better than OMP in practice. We discuss the runtime in detail in Section 3.

The main theorem of this paper states that ROMP yields exact sparse recovery in the noiseless case, and approximate recovery in general, provided that the measurement matrix satisfies the Restricted Isometry Condition.

THEOREM 1.2 (STABILITY UNDER MEASUREMENT PERTURBATIONS). *Assume a measurement matrix $\Phi$ 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 $\Phi v$ becomes corrupted, so we consider $x = \Phi v + e$ where $e$ is some error vector. Then ROMP produces a good approximation to $v$:*

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

**Remarks. 1.** Theorem 1.2 guarantees *exact sparse recovery* in the noiseless situation. Indeed, if $e = 0$ the reconstruction is exact: $\hat{v} = v$. In fact, for the noiseless case we actually only require $\varepsilon \leq 0.03/\sqrt{\log n}$ and the Restricted Isometry Condition with parameters $(2n, \varepsilon)$.[11]

**2.** Theorem 1.2 gives *uniform guarantees* of sparse recovery. Indeed, once the measurement matrix satisfies a deterministic condition (RIC), then our algorithm ROMP accurately recovers *every* sparse vector from its measurements. Uniform guarantees have been shown to be impossible for OMP,[13] and it has been an open problem to find a version of OMP with uniform guarantees (see[14]). Theorem 1.2 says that ROMP essentially settles this problem.

**3.** The logarithmic factor in $\varepsilon$ may be an artifact of the proof. At this moment, we do not know how to remove it.

**4.** Measurement matrices known to satisfy the Restricted Isometry Condition include random *Gaussian, Bernoulli and partial Fourier matrices*, with number of measurements $N$ almost linear in the sparsity $n$, i.e. as in (1.1). See[11] for detailed information. It has been unknown whether OMP gives sparse recovery for partial Fourier measurements (even with non-uniform guarantees). ROMP gives sparse recovery for these measurements, and even with uniform guarantees.

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 $f$. The following Corollary is proved in.[12]

COROLLARY 1.3 (STABILITY OF ROMP UNDER SIGNAL PERTURBATIONS). *Assume a measurement matrix $\Phi$ 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 $\Phi v$ becomes corrupted, so we consider $x = \Phi v + e$ where $e$ is some error vector. Then ROMP produces a good approximation to $v_{2n}$:*

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

**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 also have the error bound for the entire vector $v$ (see[12]):

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

**3.** For the convex programming method (1.2), the stability bound (1.4) was proved in,[5] 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,[9] 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,[12] 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,[6].[8] 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:

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

As observed in,[5] 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 discuss implementation, running time, and empirical performance of ROMP and also demonstrate some numerical experiments that illustrate the stability of ROMP.

## 2. PROOF OF THEOREM

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 $\Phi$ 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 $I' = \mathrm{supp}(v) \cup I$ 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 $v|_{\mathrm{supp}(v)\setminus I} = (v-y)|_{\mathrm{supp}(v)\setminus I}$, 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 $\Phi$ 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*

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

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

$$|I| \le 2n. \tag{2.1}$$

Let $r \neq 0$ 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

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

and its complementary subspaces

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

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 $\Phi v \in H$. □

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

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

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

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

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

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

as desired. □

LEMMA 2.5 (APPROXIMATION OF THE OBSERVATION). *Consider the observation vectors $u_0 = \Phi^* x_0$ and $u = \Phi^* r$. Then for any set $T \subset \{1, \ldots, d\}$ with $|T| \leq 3n$,*

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

*Proof.* Since $x_0 = \Phi v_0$, we have by Lemma 2.4 and the Restricted Isometry Condition that

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

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

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 $S = \mathrm{supp}(v) \setminus I$. 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 $|J_0 \cap \mathrm{supp}(v)| < \frac{1}{2}|J_0|$, and thus

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

Set $\Lambda = J_0 \backslash \mathrm{supp}(v)$. 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 $\Lambda$:

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

where the last inequality holds by Lemma 2.7.

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

$$\|u|_\Lambda\|_2 \leq \|u|_\Lambda - u_0|_\Lambda\|_2 + \|u_0|_\Lambda\|_2. \tag{2.4}$$

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

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

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

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

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

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

This is a contradiction to (2.3) so long as

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

If this is true, then indeed (i) in the theorem must hold. If it is not true, then by the choice of $\varepsilon$, 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.

## 3. IMPLEMENTATION AND EMPIRICAL PERFORMANCE OF ROMP

### 3.1 Running time

The Identification step of ROMP, i.e. selection of the subset $J$, can be done by *sorting* the coordinates of $u$ in the nonincreasing order and selecting $n$ biggest. Many sorting algorithms such as Mergesort or Heapsort provide running times of $O(d \log d)$.

The Regularization step of ROMP, i.e. selecting $J_0 \subset J$, can be done fast by observing that $J_0$ is an *interval* in the decreasing rearrangement of coefficients. Moreover, the analysis of the algorithm shows that instead of searching over all intervals $J_0$, it suffices to look for $J_0$ among $O(\log n)$ *consecutive intervals* with endpoints where the magnitude of coefficients decreases by a factor of 2. (these are the sets $A_k$ in the proof of Lemma 3.7 in[11]). Therefore, the Regularization step can be done in time $O(n)$.

In addition to these costs, the $k$-th iteration step of ROMP involves *multiplication* of the $d \times N$ matrix $\Phi^*$ by a vector, and solving the *least squares problem* with the $N \times |I|$ matrix $\Phi_I$, where $|I| \leq 2k \leq 2n$. For unstructured

matrices, these tasks can be done in time $dN$ and $O(n^2N)$ respectively. Since the submatrix of $\Phi$ when restricted to the index set $I$ is near an isometry, using an iterative method such as the Conjugate Gradient Method allows us to solve the least squares method in a constant number of iterations (up to a specific accuracy.) Using such a method then reduces the time of solving the least squares problem to just $O(nN)$. Thus in the cases where ROMP terminates after a fixed number of iterations, the total time to solve all required least squares problems would be just $O(nN)$. For structured matrices, such as partial Fourier, these times can be improved even more using fast multiply techniques.

In other cases, however, ROMP may need more than a constant number of iterations before terminating, say the full $O(n)$ iterations. In this case, it may be more efficient to maintain the QR factorization of $\Phi_I$ and use the Modified Gram-Schmidt algorithm. With this method, solving all the least squares problems takes total time just $O(n^2N)$. However, storing the QR factorization is quite costly, so in situations where storage is limited it may be best to use the iterative methods mentioned above.

ROMP terminates in at most $2n$ iterations. Therefore, for unstructured matrices using the methods mentioned above and in the interesting regime $N \geq \log d$, *the total running time of ROMP is O(dNn).* This is the same bound as for OMP.[14]

### 3.2 Experiments

This section describes our experiments that illustrate the signal recovery power of ROMP. We experimentally examine how many measurements $N$ are necessary to recover various kinds of $n$-sparse signals in $\mathbb{R}^d$ using ROMP. We also demonstrate that the number of iterations ROMP needs to recover a sparse signal is in practice at most linear the sparsity.

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 generate an $N \times d$ Gaussian measurement matrix $\Phi$ and then perform 500 independent trials. In a given trial, we generate an $n$-sparse signal $v$ in one of two ways. In either case, we first select the support of the signal by choosing $n$ components uniformly at random (independent from the measurement matrix $\Phi$). In the cases where we wish to generate flat signals, we then set these components to one. In the cases where we wish to generate sparse compressible signals, we set the $i^{th}$ component of the support to plus or minus $i^{-1/p}$ for a specified value of $0 < p < 1$. We then execute ROMP with the measurement vector $x = \Phi v$.

Figure 1 depicts the percentage (from the 500 trials) of sparse flat signals that were reconstructed exactly. 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 exact recovery percentage. Our results show that performance of ROMP is very similar to that of OMP which can be found in.[14]

Figure 2 demonstrates that the number of iterations needed for sparse compressible is higher than the number needed for sparse flat signals, as one would expect. The plot shows the average number of iterations (vertical axis) needed to recover a sparse compressible signal for various levels of sparsity $n$ (horizontal axis). The plot was generated for various values of the parameter $p$ (as discussed above) with $d = 10,000$ and $N = 200$. The plot suggests that for smaller values of $p$ (meaning signals that decay more rapidly) ROMP needs more iterations.

Figure 3 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 3 confirms the results of the Theorem, while also suggesting the bound may be improved by removing the $\sqrt{\log n}$ factor.

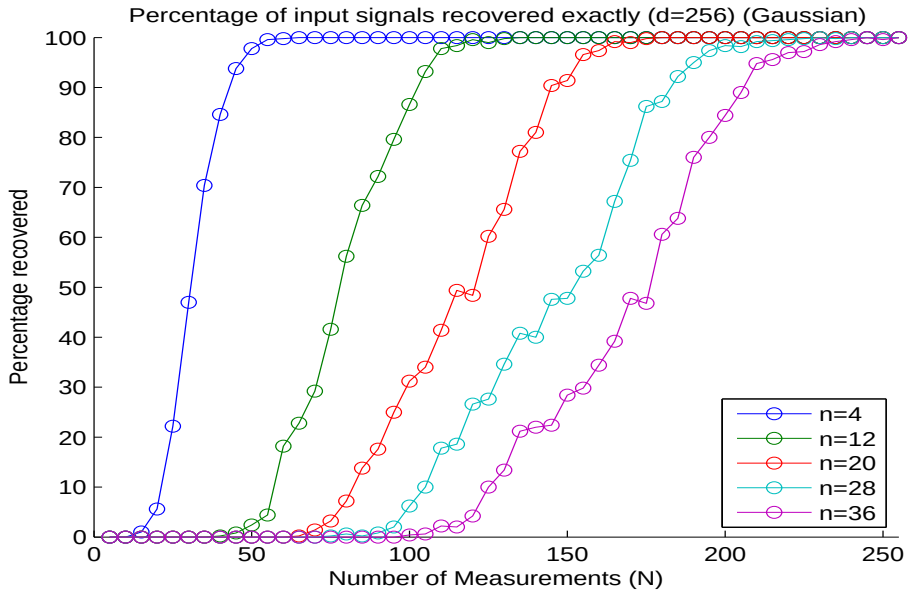


Figure 1. The percentage of sparse flat signals exactly recovered by ROMP as a function of the number of measurements $N$ in dimension $d = 256$ for various levels of sparsity $n$.

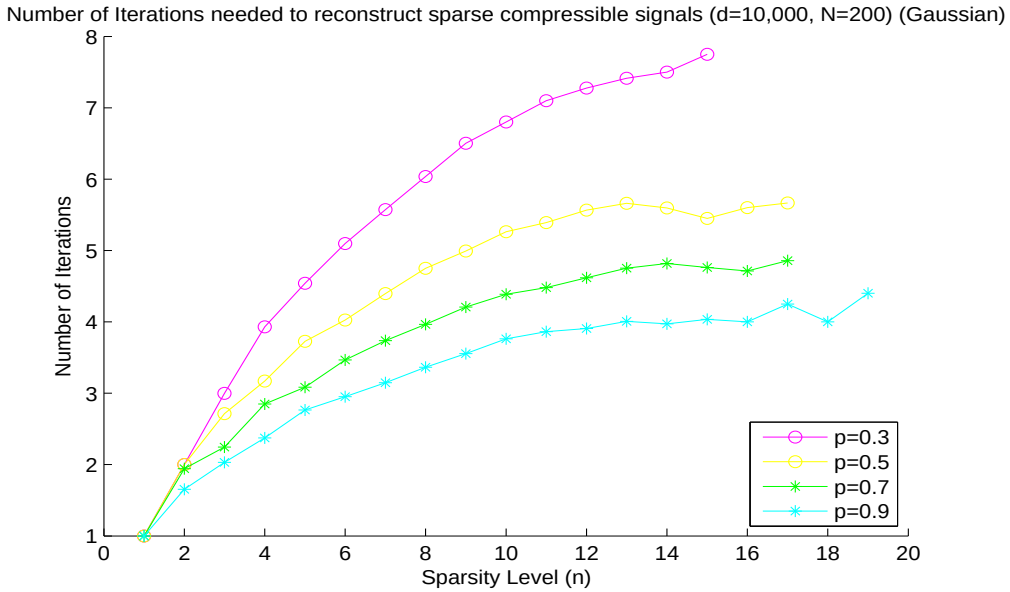


Figure 2. The number of iterations executed by ROMP as a function of the sparsity $n$ in dimension $d = 10,000$ with $N = 200$ for sparse compressible signals.

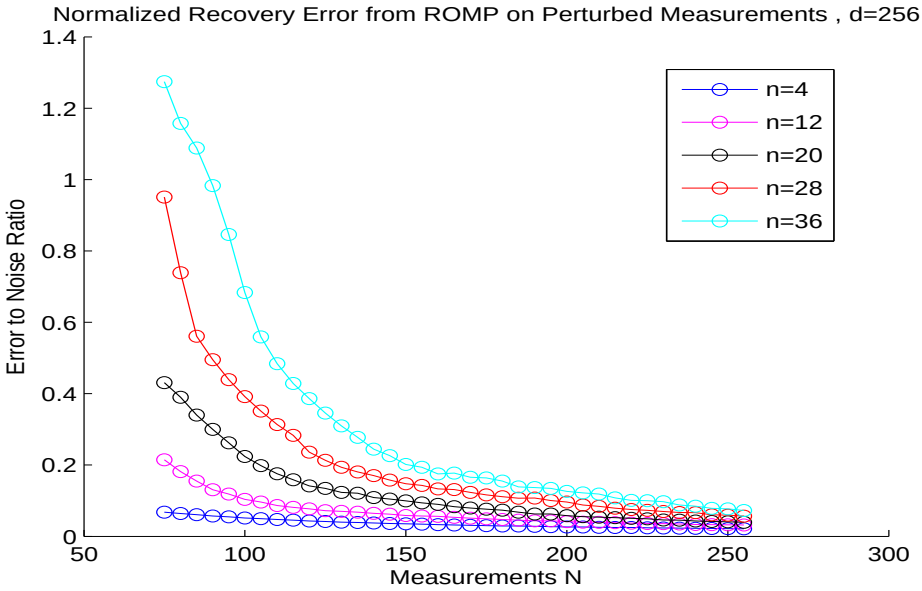


Figure 3. The error to noise ratio $\frac{\|\hat{v}-v\|_2}{\|e\|_2}$ as a function of the number of measurements $N$ in dimension $d = 256$ for various levels of sparsity $n$.

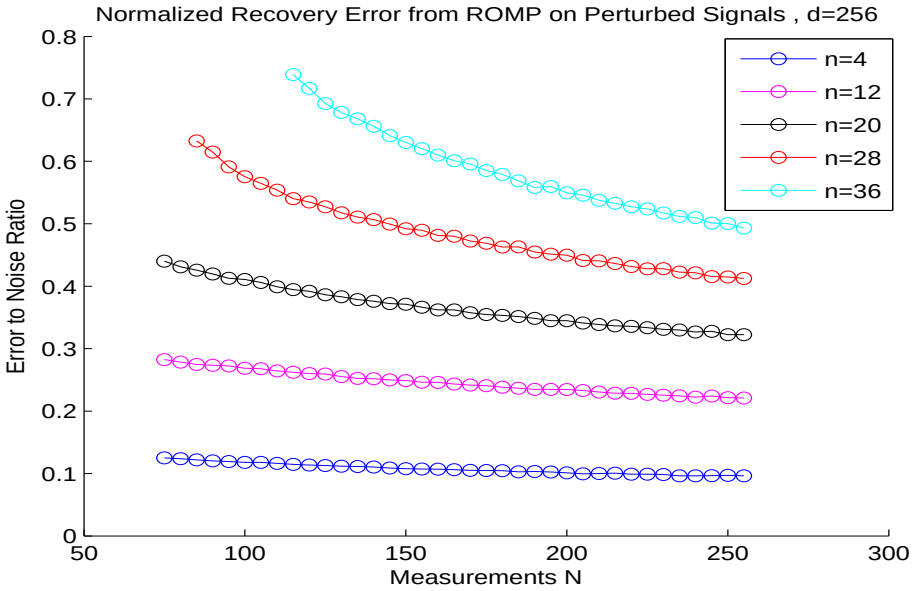


Figure 4. The error to noise ratio $\frac{\|\hat{v}-v_{2n}\|_2}{\|v-v_n\|_1/\sqrt{n}}$ using a perturbed signal, as a function of the number of measurements $N$ in dimension $d = 256$ for various levels of sparsity $n$.

## REFERENCES


1. E. Candès, *Compressive sampling*, Proc. International Congress of Mathematics, 3, pp. 1433-1452, Madrid, Spain, 2006.
2. A. Cohen, W. Dahmen, R. DeVore, *Compressed sensing and k-term approximation*, Manuscript (2007).
3. D. Donoho, Ph. Stark, *Uncertainty principles and signal recovery*, SIAM J. Appl. Math. 49 (1989), 906–931.
4. 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.
5. E. Candès, J. Romberg, T. Tao, *Stable signal recovery from incomplete and inaccurate measurements*, Comm. Pure Appl. Math. 59 (2006), 1207-1223
6. E. Candes, T. Tao, *Near-optimal signal recovery from random projections: universal encoding strategies*, IEEE Trans. Inform. Theory 52 (2004), 5406–5425.
7. Compressed Sensing webpage, `http://www.dsp.ece.rice.edu/cs/`
8. D. Donoho, *Compressed sensing*, IEEE Trans. Inform. Theory 52 (2006), 1289–1306.
9. 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.
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. D. Needell, R. Vershynin, *Signal Recovery from Inaccurate and Incomplete Measurements via Regularized Orthogonal Matching Pursuit*, submitted.
13. H. Rauhut, *On the impossibility of uniform recovery using greedy methods*, in preparation
14. J. A. Tropp, A. C. Gilbert, *Signal recovery from random measurements via orthogonal matching pursuit*, IEEE Trans. Info. Theory. To appear, 2007.
15. V. Temlyakov, *Nonlinear methods of approximation*, Found. Comput. Math. 3 (2003), 33-107.