Signal recovery from incomplete and inaccurate measurements via ROMP

Deanna Needell [(1)] and Roman Vershynin [(2)]

(1) Dept. of Mathematics, Univ. California, Davis, CA 95616, USA.
(2) Mathematics Dept., Univ. Michigan, Ann Arbor, MI 48109, USA.

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

dneedell@math.ucdavis.edu, romanv@umich.edu

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:

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

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

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

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 (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*

$$(1 - \varepsilon)\|v\|_2 \le \|\Phi v\|_2 \le (1 + \varepsilon)\|v\|_2,$$

for all m-sparse vectors.

Under the Restricted Isometry Condition with parameters $(2n, \sqrt{2} - 1)$, the convex program (1) exactly recovers an n-sparse signal v from its measurements x [4, 5].

The Restricted Isometry Condition can be viewed as an abstract form of the Uniform Uncertainty Principle of harmonic analysis ([6], see also [2] and [11]). 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 [13] 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) [16].

Greedy methods are usually fast and easy to implement. For example, except with probability d^{-1}, OMP needs just n iterations to find the support of an n-sparse signal v. Since each iteration amounts to solving one least-squares problem, its running time is always polynomial in n, N and d. In contrast, no known bounds are known on the running time of (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 [16] and [13].

A variant of OMP was recently found in [13] that has guarantees essentially as strong as those of convex programming methods (OMP itself does not have such strong guarantees, see [14]). 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) *or greedy algorithms (ROMP).*

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 [8], [15], [9], [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

$$\min \|u\|_1 \qquad \text{subject to} \qquad \|\Phi u - x\|_2 \leq \delta \tag{2}$$

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 [9] suggests that OMP should be less stable than the convex program (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* (2). This result essentially closes a gap between convex programming and greedy approaches to sparse recovery.

REGULARIZED ORTHOGONAL MATCHING PURSUIT

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 1 *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 2 (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 a good approximation to v:*

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

Note that in the noiseless situation ($e = 0$) the reconstruction is exact: $\hat{v} = v$. This case of Theorem 2 was proved in [13].
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 3 (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 a good approximation to v_{2n}:*

$$\|\hat{v} - v_{2n}\|_2 \leq 159\sqrt{\log 2n}\Big(\|e\|_2 + \frac{\|v - v_n\|_1}{\sqrt{n}}\Big). \tag{3}$$

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 can apply Corollary 3 to the largest $2n$ coordinates of v and use Lemma 5 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:

$$\|\hat{v} - v\|_2 \leq 160\sqrt{\log 2n}\Big(\|e\|_2 + \frac{\|v - v_n\|_1}{\sqrt{n}}\Big). \quad (4)$$

3. For the convex programming method (2), the stability bound (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 (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, it is straightforward to show (see [12]) that the best $2n$-term approximations of the original and the reconstructed signals are close:

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

6. An important special case of Corollary 3 is for the class of compressible vectors, which is a common model in signal processing, see [6], [7]. 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 (4) yields the following bound on the reconstructed signal:

$$\|v - \hat{v}\|_2 \leq C'_p \frac{\sqrt{\log n}}{n^{p-1/2}} + C''\sqrt{\log n}\|e\|_2. \quad (5)$$

As observed in [3], this bound is optimal (within the logarithmic factor); no algorithm can perform fundamentally better.
The rest of the paper has the following organization. In Section 2., we discuss our main result, Theorem 2. In Section 3., we deduce the extension for approximately sparse signals, Corollary 3, and a consequence for best n-term approximations, Corollary 6.

2. Theorem 2

The complete proof of Theorem 2 is given in [12], and parallels that given in [13]. The argument relies on the following iteration invariant, which shows that 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.

Theorem 4 (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) $|J_0 \cap \mathrm{supp}(v)| \geq \frac{1}{2}|J_0|$;

(ii) $\|v|_{\mathrm{supp}(v)\setminus I}\|_2 \leq 100\sqrt{\log n}\|e\|_2$.

This iteration invariant is proven in [12], and Theorem 2 follows immediately by examining the case where (ii) occurs at some iteration and the case where (i) occurs at every iteration. Theorem 2 then extends naturally to the case of noisy or compressible signals, as seen in the next section.

3. Approximately sparse vectors and best n-term approximations

3.1 Proof of Corollary 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 2 and the triangle inequality,

$$\|v_{2n} - \hat{v}\|_2 \leq 104\sqrt{\log 2n}(\|\Phi(v - v_{2n})\|_2 + \|e\|_2), \quad (6)$$

The following lemma as in [10] relates the 2-norm of a vector's tail to its 1-norm. An application of this lemma combined with (6) will prove Corollary 3.

Lemma 5 (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 AM-GM inequality,

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

This completes the proof. □
By Lemma 29 of [10], 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 5 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 (6), this proves the corollary.

3.2 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, we obtain a $2n$-sparse vector very close to the original signal.

Corollary 6 *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 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

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

We then have

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

and

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

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,

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

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

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

This completes the proof. □

Remark. Corollary 6 combined with Corollary 3 and (4) implies that we can also estimate a bound on the whole signal v:

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

References:

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

[2] 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.

[3] E. Candès, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math*, 59(8):1207–1223, 2006.

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

[5] E. J. Candès. The restricted isometry property and its implications for compressed sensing. Technical report, California Institute of Technology, 2008.

[6] E. J. Candès and T. Tao. Near optimal signal recovery from random projections: Universal encoding strategies? Submitted for publication, Nov. 2004.

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

[8] 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.

[9] 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.

[10] 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.

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

[12] D. Needell and R. Vershynin. Signal recovery from incomplete and inaccurate measurements via Regularized Orthogonal Matching Pursuit. Submitted for publication, October 2007.

[13] D. Needell and R. Vershynin. Uniform uncertainty principle and signal recovery via Regularized Orthogonal Matching Pursuit. Submitted for publication, July 2007.

[14] H. Rauhut. On the impossibility of uniform sparse reconstruction using greedy methods. To appear, *Sampl. Theory Signal Image Process.*, 2008.

[15] 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.

[16] 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.