

Uniform Uncertainty Principle and Signal Recovery via Regularized Orthogonal Matching Pursuit

Deanna Needell · Roman Vershynin

Received: 21 August 2007 / Revised: 6 February 2008 / Accepted: 12 March 2008 /
Published online: 5 June 2008
© SFOCM 2008

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 Orthogonal Matching Pursuit (ROMP) which has advantages of both approaches: the speed and transparency of OMP and the strong uniform guarantees of 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.

Keywords Signal recovery algorithms · Restricted isometry condition · Uncertainty principle · Basis pursuit · Compressed sensing · Orthogonal matching pursuit · Signal recovery · Sparse approximation

Mathematics Subject Classification (2000) 68W20 · 65T50 · 41A46

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

Communicated by Emmanuel Candès.

D. Needell · R. Vershynin (✉)

Department of Mathematics, University of California, One Shields Ave, Davis, CA 95616, USA
e-mail: vershynin@math.ucdavis.edu

D. Needell

e-mail: dneedell@math.ucdavis.edu

$n \ll d$ nonzero components:

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

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 Φ is some $N \times d$ measurement matrix.¹

As discussed in [5], 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, massive recent work in the emerging area of compressed sensing demonstrated that for several natural classes of measurement matrices Φ , the signal v can be exactly reconstructed from its measurements Φv with

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

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 [6] documents progress in this area.

The two major algorithmic approaches to sparse recovery are methods based on L_1 -minimization and iterative methods (matching pursuits). We now briefly describe these methods. Then we propose a new iterative method that has advantages of both approaches.

1.1 L_1 -Minimization

This approach to sparse recovery has been advocated over decades by Donoho and his collaborators (see, e.g., [10]). The sparse recovery problem can be stated as the problem of finding the sparsest signal v with the given measurements Φv :

$$\min \|u\|_0 \quad \text{subject to} \quad \Phi u = \Phi v \quad (L_0)$$

where $\|u\|_0 := |\text{supp}(u)|$. Donoho and his associates advocated the principle that for some measurement matrices Φ , the highly nonconvex combinatorial optimization problem (L_0) should be equivalent to its convex relaxation

$$\min \|u\|_1 \quad \text{subject to} \quad \Phi u = \Phi v \quad (L_1)$$

where $\|u\|_1 = \sum_i |u_i|$ denotes the ℓ_1 -norm of the vector $u = (u_1, \dots, u_d)$. The convex problem (L_1) can be solved using methods of convex and even linear programming.

The recent progress in the emerging area of compressed sensing pushed forward this program (see survey [1]). A necessary and sufficient condition of exact sparse recovery is that the map Φ be one-to-one on the set of n -sparse vectors. Candès and Tao [4] proved that a stronger quantitative version of this condition guarantees the equivalence of the problems (L_0) and (L_1) .

¹We chose to work with real numbers for simplicity of presentation; similar results hold over complex numbers.

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

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

The restricted isometry condition states that every set of m columns of Φ forms approximately an orthonormal system. One can interpret the restricted isometry condition as an abstract version of the uniform uncertainty principle in harmonic analysis ([3], see also discussions in [2, 14]).

Theorem 1.2 (Sparse recovery under RIC [4]) *Assume that the measurement matrix Φ satisfies the restricted isometry condition with parameters $(3n, 0.2)$. Then every n -sparse vector x can be exactly recovered from its measurements Φx as a unique solution to the convex optimization problem (L₁).*

In a lecture on compressive sampling, Candès sharpened this to work for the restricted isometry condition with parameters $(2n, \sqrt{2} - 1)$. Measurement matrices that satisfy the restricted isometry condition with number of measurements as in (1.1) include random Gaussian, Bernoulli, and partial Fourier matrices. Section 2 contains more detailed information.

1.2 Orthogonal Matching Pursuit (OMP)

An alternative approach to sparse recovery is via iterative algorithms, which find the support of the n -sparse signal v progressively. Once $S = \text{supp}(v)$ is found correctly, it is easy to compute the signal v from its measurements $x = \Phi v$ as $v = (\Phi_S)^{-1}x$, where Φ_S denotes the measurement matrix Φ restricted to columns indexed by S .

A basic iterative algorithm is Orthogonal Matching Pursuit (OMP), popularized and analyzed by Tropp and Gilbert in [20]; see [19] for a more general setting. OMP recovers the support of v , one index at a time, in n steps. Under a hypothetical assumption that Φ is an isometry, i.e., the columns of Φ are orthonormal; the signal v can be exactly recovered from its measurements $x = \Phi v$ as $v = \Phi^*x$.

The problem is that the $N \times d$ matrix Φ is never an isometry in the interesting range where the number of measurements N is smaller than the ambient dimension d . Even though the matrix is not an isometry, one can still use the notion of coherence in recovery of sparse signals. In that setting, greedy algorithms are used with incoherent dictionaries to recover such signals, see [8, 9, 11]. In our setting, for random matrices, one expects the columns to be approximately orthogonal and the *observation vector* $u = \Phi^*x$ to be a good approximation to the original signal v .

The biggest coordinate of the observation vector u in magnitude should thus be a nonzero coordinate of the signal v . We thus find one point of the support of v . Then OMP can be described as follows. First, we initialize the residual $r = x$. At each iteration, we compute the observation vector $u = \Phi^*r$. Denoting by I the coordinates selected so far, we solve a least squares problem and update the residual

$$y = \arg \min_{z \in \mathbb{R}^I} \|x - \Phi z\|_2; \quad r = x - \Phi y,$$

to remove any contribution of the coordinates in I . OMP then iterates this procedure n times, and outputs a set I of size n , which should equal the support of the signal v .

Tropp and Gilbert [20] analyzed the performance of OMP for Gaussian measurement matrices Φ ; a similar result holds for general sub-Gaussian matrices. They proved that for every fixed n -sparse d -dimensional signal v , and an $N \times d$ random Gaussian measurement matrix Φ , OMP recovers (the support of) v from the measurements $x = \Phi v$ correctly with high probability, provided the number of measurements is $N \sim n \log d$.

1.3 Advantages and Challenges of Both Approaches

The L_1 -minimization method has *strongest known guarantees* of sparse recovery. Once the measurement matrix Φ satisfies the restricted isometry condition, this method works correctly for all sparse signals v . No iterative methods have been known to feature such uniform guarantees with the exception of chaining pursuit [12] and the HHS algorithm [13], which, however, only work with specifically designed structured measurement matrices.

The restricted isometry condition is a natural abstract deterministic property of a matrix. Although establishing this property is often nontrivial, this task is *decoupled from the analysis* of the recovery algorithm.

L_1 -minimization is based on linear programming, which has its advantages and disadvantages. One thinks of linear programming as a black box, and any development of fast solvers will reduce the *running time* of the sparse recovery method. On the other hand, it is not very clear what this running time is, as there is no strongly polynomial time algorithm in linear programming yet. All known solvers take time polynomial not only in the dimension of the program d , but also on certain condition numbers of the program. While for some classes of random matrices the expected running time of linear programming solvers can be bounded (see the discussion in [18] and subsequent work in [21]), estimating condition numbers is hard for specific matrices. For example, there is no result yet showing that the restricted isometry condition implies that the condition numbers of the corresponding linear program is polynomial in d .

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 Φ^* (computing the observation vector u), and solving a least squares problem in dimensions at most $N \times n$ (with matrix Φ_I). This yields strongly polynomial running time. In practice, OMP is observed to perform faster and is easier to implement than L_1 -minimization [20]. For more details, see [20].

Orthogonal matching pursuit is quite *transparent*: at each iteration, it selects a new coordinate from the support of the signal v in a very specific and natural way. In contrast, the known L_1 -minimization solvers, such as the simplex method and interior point methods, compute a path toward the solution. However, the geometry of L_1 is clear, whereas the analysis of greedy algorithms can be difficult simply because they are iterative.

On the other hand, Orthogonal matching pursuit has *weaker guarantees* of exact recovery. Unlike L_1 -minimization, the guarantees of OMP are nonuniform: For each

fixed sparse signal v and not for *all* signals, the algorithm performs correctly with high probability. Rauhut has shown that uniform guarantees for OMP are impossible for natural random measurement matrices [16].

Moreover, OMP's condition on measurement matrices given in [20] 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.

These open problems about OMP, first stated in [20] and often reverberated in the compressed sensing community, motivated the present paper. We essentially settle them in positive by the following modification of orthogonal matching pursuit.

1.4 Regularized OMP

This new algorithm for sparse recovery will perform correctly for all measurement matrices Φ satisfying the restricted isometry condition, and for all sparse signals.

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 . Namely, the observation vector u encodes correlations of the measurement vector x with the columns of Φ . Note that Φ is a dictionary, and so since the signal v is sparse, x has a sparse representation with respect to the dictionary. By the restricted isometry condition, every n columns form approximately an orthonormal system. Therefore, every n coordinates of the observation vector u look like correlations of the measurement vector x with the orthonormal basis and, therefore, are close in the Euclidean norm to the corresponding n coefficients of v . This is documented in Proposition 3.2 below.

The local approximation property suggests to make use of the n biggest coordinates of the observation vector u , rather than one biggest coordinate as OMP did. We thus force the selected coordinates to be more regular (i.e., closer to uniform) by selecting only the coordinates with comparable sizes. To this end, a new *regularization* step will be needed to ensure that each of these coordinates gets an even share of information. This leads to the following algorithm for sparse recovery:

Regularized Orthogonal Matching Pursuit (ROMP)

Input: Measurement vector $x \in \mathbb{R}^N$ and sparsity level n

Output: Index set $I \subset \{1, \dots, d\}$

Initialize: Let the index set $I = \emptyset$ and the residual $r = x$.

Repeat the following steps until $r = 0$:

Identify: Choose a set J of the n biggest 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:

$$y = \arg \min_{z \in \mathbb{R}^I} \|x - \Phi z\|_2; \quad r = x - \Phi y.$$

Remark The identification and regularization steps of ROMP can be performed efficiently. In particular, the regularization step does *not* imply 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 Sect. 4.

The main theorem of this paper states that ROMP yields exact sparse recovery provided that the measurement matrix satisfies the restricted isometry condition.

Theorem 1.3 (Exact sparse recovery via ROMP) *Assume a measurement matrix Φ satisfies the restricted isometry condition with parameters $(2n, \varepsilon)$ for $\varepsilon = 0.03/\sqrt{\log n}$. Let v be an n -sparse vector in \mathbb{R}^d with measurements $x = \Phi v$. Then ROMP in at most n iterations outputs a set I such that*

$$\text{supp}(v) \subset I \quad \text{and} \quad |I| \leq 2n.$$

This theorem is proved in Sect. 3.

Remarks

1. Theorem 1.3 guarantees *exact sparse recovery*. Indeed, it is easy to compute the signal v from its measurements $x = \Phi v$ and the set I given by ROMP as $v = (\Phi_I)^{-1}x$, where Φ_I denotes the measurement matrix Φ restricted to columns indexed by I .
2. Theorem 1.3 gives *uniform guarantees* of sparse recovery. Indeed, once the measurement matrix satisfies a deterministic condition (RIC), then our algorithm ROMP correctly recovers *every* sparse vector from its measurements. Uniform guarantees have been shown to be impossible for OMP [16], and it has been an open problem to find a version of OMP with uniform guarantees (see [20]). Theorem 1.3 says that ROMP essentially settles this problem.
3. The logarithmic factor in ε 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). Section 2 contains detailed information. It has been unknown whether OMP gives sparse recovery for partial Fourier measurements (even with nonuniform guarantees). ROMP gives sparse recovery for these measurements, and even with uniform guarantees.

The rest of the paper is organized as follows. In Sect. 2, we describe known classes of measurement matrices satisfying the restricted isometry condition. In Sect. 3, we give the proof of Theorem 1.3. In Sect. 4, we discuss implementation, running time, and empirical performance of ROMP.

2 Measurement Matrices Satisfying the Restricted Isometry Condition

The only known measurement matrices known to satisfy the restricted isometry condition with number of measurements as in (1.1) are certain classes of random ma-

trices. The problem of deterministic constructions is still open. The known classes include: sub-Gaussian random matrices (in particular, Gaussian and Bernoulli), and random partial bounded orthogonal matrices (in particular, partial Fourier matrices).

Throughout the paper, $C, c, C_1, C_2, c_1, c_2, \dots$ denote positive absolute constants unless otherwise specified.

A *sub-Gaussian random matrix* Φ is a matrix whose entries are i.i.d. sub-Gaussian random variables with variance 1. A random variable X is sub-Gaussian if its tail distribution is dominated by that of the standard Gaussian random variable: there are constants $C_1, c_1 > 0$ such that $\mathbb{P}(|X| > t) \leq C_1 \exp(-c_1 t^2)$ for all $t > 0$. Examples of sub-Gaussian random variables are: standard Gaussian, Bernoulli (uniform ± 1), and any bounded random variables.

A *partial bounded orthogonal matrix* Φ is formed by N randomly uniformly chosen rows of an orthogonal $d \times d$ matrix Ψ , whose entries are bounded by C_2/\sqrt{d} , for some constant C_2 . An example of Ψ is the discrete Fourier transform matrix. Taking measurements Φv with a partial Fourier matrix thus amounts to observing N random frequencies of the signal v .

The following theorem documents known results on the restricted isometry condition for these classes of random matrices.

Theorem 2.1 (Measurement matrices satisfying RIC) *Consider an $N \times d$ measurement matrix Φ , and let $n \geq 1$, $\varepsilon \in (0, 1/2)$, and $\delta \in (0, 1)$.*

1. *If Φ is a sub-Gaussian matrix, then with probability $1 - \delta$, the matrix $(1/\sqrt{N})\Phi$ satisfies the restricted isometry condition with parameters (n, ε) provided that*

$$N \geq \frac{Cn}{\varepsilon^2} \log\left(\frac{d}{\varepsilon^2 n}\right).$$

2. *If Φ is a partial bounded orthogonal matrix, then with probability $1 - \delta$ the matrix $\sqrt{d}/N\Phi$ satisfies the Restricted Isometry Condition with parameters (n, ε) provided that*

$$N \geq C \left(\frac{n \log d}{\varepsilon^2} \right) \log\left(\frac{n \log d}{\varepsilon^2}\right) \log^2 d.$$

In both cases, the constant C depends only on the confidence level δ and the constants C_1, c_1, C_2 from the definition of the corresponding classes of matrices.

Remarks

1. The first part of this theorem is proved in [15]. The second part is from [17]; a similar estimate with somewhat worse exponents in the logarithms was proved in [3]. See these results for the exact dependence of C on the confidence level δ (although usually δ would be chosen to be some small constant itself).
2. In Theorem 1.3, we needed to use RIC for $\varepsilon = c_1/\sqrt{\log n}$. An immediate consequence of Theorem 2.1 is that sub-Gaussian matrices satisfy such RIC for the number of measurements

$$N \sim n \log^2 d$$

and partial bounded orthogonal matrices for

$$N \sim n \log^5 d.$$

These numbers of measurements guarantee exact sparse recovery using ROMP.

3 Proof of Theorem 1.3

We shall prove a stronger version of Theorem 1.3, which states that *at every iteration* of ROMP, at least 50% of the newly selected coordinates are from the support of the signal v .

Theorem 3.1 (Iteration invariant of ROMP) *Assume Φ satisfies the Restricted Isometry Condition with parameters $(2n, \varepsilon)$ for $\varepsilon = 0.03/\sqrt{\log n}$. Let $v \neq 0$ be an n -sparse vector with measurements $x = \Phi v$. Then at any iteration of ROMP, after the regularization step, we have $J_0 \neq \emptyset$, $J_0 \cap I = \emptyset$ and*

$$|J_0 \cap \text{supp}(v)| \geq \frac{1}{2}|J_0|. \quad (3.1)$$

In other words, at least 50% of the coordinates in the newly selected set J_0 belong to the support of v .

In particular, at every iteration ROMP finds at least one new coordinate in the support of the signal v . Coordinates outside the support can also be found, but (3.1) guarantees that the number of such “false” coordinates is always smaller than those in the support. This clearly implies Theorem 1.3.

Before proving Theorem 3.1, we explain how the restricted isometry condition will be used in our argument. RIC is necessarily a local principle, which concerns not the measurement matrix Φ as a whole, but its submatrices of n columns. All such submatrices Φ_I , $I \subset \{1, \dots, d\}$, $|I| \leq n$ are almost isometries. Therefore, for every n -sparse signal v , the observation vector $u = \Phi^* \Phi v$ approximates v locally, when restricted to a set of cardinality n . The following proposition formalizes these local properties of Φ on which our argument is based.

Proposition 3.2 (Consequences of Restricted Isometry Condition) *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 $I \subset \{1, \dots, d\}$, $|I| \leq n$, the observation vector $u = \Phi^* \Phi v$ satisfies*

$$\|u|_I - v|_I\|_2 \leq 2.03\varepsilon \|v\|_2.$$

- (2) (Spectral norm) *For any vector $z \in \mathbb{R}^N$ and every set $I \subset \{1, \dots, d\}$, $|I| \leq 2n$, we have*

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

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

$$\|P_I P_J\|_{2 \rightarrow 2} \leq 2.2\varepsilon.$$

Proof Part 1. Let $\Gamma = I \cup \text{supp}(v)$, so that $|\Gamma| \leq 2n$. Let Id_Γ denote the identity operator on \mathbb{R}^Γ . By the restricted isometry condition,

$$\|\Phi_\Gamma^* \Phi_\Gamma - \text{Id}_\Gamma\|_{2 \rightarrow 2} = \sup_{y \in \mathbb{R}^\Gamma, \|y\|_2=1} \left| \|\Phi_\Gamma y\|_2^2 - \|y\|_2^2 \right| \leq (1 + \varepsilon)^2 - 1 \leq 2.03\varepsilon.$$

Since $\text{supp}(v) \subset \Gamma$, we have

$$\|u|_\Gamma - v|_\Gamma\|_2 = \|\Phi_\Gamma^* \Phi_\Gamma v - \text{Id}_\Gamma v\|_2 \leq 2.03\varepsilon \|v\|_2.$$

The conclusion of Part 1 follows since $I \subset \Gamma$.

Part 2. Denote by Q_I the orthogonal projection in \mathbb{R}^d onto \mathbb{R}^I . Since $|I| \leq 2n$, the Restricted Isometry Condition yields

$$\|Q_I \Phi^*\|_{2 \rightarrow 2} = \|\Phi Q_I\|_{2 \rightarrow 2} \leq 1 + \varepsilon.$$

This yields the inequality in Part 2.

Part 3. The desired inequality is equivalent to

$$\frac{|\langle x, y \rangle|}{\|x\|_2 \|y\|_2} \leq 2.2\varepsilon \quad \text{for all } x \in \text{range}(\Phi_I), y \in \text{range}(\Phi_J).$$

Let $K = I \cup J$ so that $|K| \leq 2n$. For any $x \in \text{range}(\Phi_I), y \in \text{range}(\Phi_J)$, there are a, b so that

$$x = \Phi_K a, \quad y = \Phi_K b, \quad a \in \mathbb{R}^I, b \in \mathbb{R}^J.$$

By the restricted isometry condition,

$$\|x\|_2 \geq (1 - \varepsilon) \|a\|_2, \quad \|y\|_2 \geq (1 - \varepsilon) \|b\|_2.$$

By the proof of Part 2 above and since $\langle ab \rangle = 0$, we have

$$|\langle x, y \rangle| = |\langle (\Phi_K^* \Phi_K - \text{Id}_\Gamma) a, b \rangle| \leq 2.03\varepsilon \|a\|_2 \|b\|_2.$$

This yields

$$\frac{|\langle x, y \rangle|}{\|x\|_2 \|y\|_2} \leq \frac{2.03\varepsilon}{(1 - \varepsilon)^2} \leq 2.2\varepsilon,$$

which completes the proof. \square

We are now ready to prove Theorem 3.1.

Proof of Theorem 3.1 The proof is by induction on the iteration of ROMP. The induction claim is that for all previous iterations, the set of newly chosen indices J_0 is nonempty, disjoint from the set of previously chosen indices I , and (3.1) holds.

Let I be the set of previously chosen indices at the start of a given iteration. The induction claim easily implies that

$$|\text{supp}(v) \cup I| \leq 2n. \quad (3.2)$$

Let J_0, J , be the sets found by ROMP in the current iteration. By the definition of the set J_0 , it is nonempty.

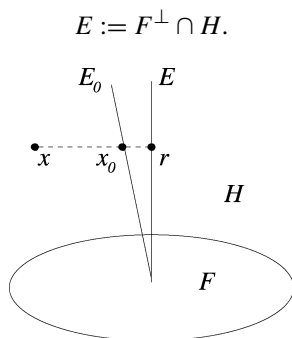
Let $r \neq 0$ be the residual at the start of this iteration. We shall approximate r by a vector in $\text{range}(\Phi_{\text{supp}(v) \setminus I})$. That is, we want to approximately realize the residual r as measurements of some signal which lives on the still unfound coordinates of the support of v . To that end, we consider the subspace

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

and its complementary subspaces

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

The restricted isometry condition in the form of Part 3 of Proposition 3.2 ensures that F and E_0 are almost orthogonal. Thus E_0 is close to the orthogonal complement of F in H ,



We will also consider the signal we seek to identify at the current iteration, its measurements, and its observation vector:

$$v_0 := v|_{\text{supp}(v) \setminus I}, \quad x_0 := \Phi v_0 \in E_0, \quad u_0 := \Phi^* x_0. \quad (3.3)$$

Lemma 3.5 will show that $\|(u - u_0)|_T\|_2$ for any small enough subset T is small, and Lemma 3.8 will show that $\|u|_{J_0}\|_2$ is not too small. First, we show that the residual r has a simple description:

Lemma 3.3 (Residual) *Here and thereafter, let P_L denote the orthogonal projection in \mathbb{R}^N onto a linear subspace L . Then*

$$r = P_E x.$$

Proof By definition of the residual in the algorithm, $r = P_{F^\perp}x$. Since $x \in H$, we conclude from the orthogonal decomposition $H = F + E$ that $x = P_Fx + P_Ex$. Thus $r = x - P_Fx = P_Ex$. \square

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. This will allow us to carry over information from E_0 to E .

Lemma 3.4 (Approximation of the residual) *We have*

$$\|x_0 - r\|_2 \leq 2.2\varepsilon \|x_0\|_2.$$

Proof By definition of F , we have $x - x_0 = \Phi(v - v_0) \in F$. Therefore, by Lemma 3.3, $r = P_Ex = P_Ex_0$, and so

$$x_0 - r = x_0 - P_Ex_0 = P_Fx_0 = P_FP_{E_0}x_0.$$

Now we use Part 3 of Proposition 3.2 for the sets I and $\text{supp}(v) \setminus I$ whose union has cardinality at most $2n$ by (3.2). It follows that $\|P_FP_{E_0}x_0\|_2 \leq 2.2\varepsilon \|x_0\|_2$ as desired. \square

Lemma 3.5 (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 $|T| \leq 2n$, we have*

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

Proof Since $x_0 = \Phi v_0$, we have by Lemma 3.4 and the restricted isometry condition that

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

To complete the proof, it remains to apply Part 2 of Proposition 3.2, which yields $\|(u_0 - u)|_T\|_2 \leq (1 + \varepsilon) \|x_0 - r\|_2$. \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 approximation lemmas, this will yield that ROMP selects at least a fixed percentage of energy of the still unidentified part of the signal. 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. This will be the desired conclusion.

Lemma 3.6 (Localizing the energy) *We have $\|u|_J\|_2 \geq 0.8 \|v_0\|_2$.*

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

$$\|u_0|_J\|_2 \geq \|u_0|_S\|_2.$$

Furthermore, since $v_0|_S = v_0$, by Part 1 of Proposition 3.2 we have

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

Putting these two inequalities together and using Lemma 3.5, we conclude that

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

This proves the lemma. \square

We next bound the norm of u restricted to the smaller set J_0 . We do this by first noticing a general property of regularization:

Lemma 3.7 (Regularization) *Let y be any vector in \mathbb{R}^m , $m > 1$. Then there exists a subset $A \subset \{1, \dots, m\}$ with comparable coordinates:*

$$|y(i)| \leq 2|y(j)| \quad \text{for all } i, j \in A, \quad (3.4)$$

and with big energy:

$$\|y|_A\|_2 \geq \frac{1}{2.5\sqrt{\log m}}\|y\|_2. \quad (3.5)$$

Proof We will construct at most $O(\log m)$ subsets A_k with comparable coordinates as in (3.4), and such that at least one of these sets will have large energy as in (3.5).

Let $y = (y_1, \dots, y_m)$, and consider a partition of $\{1, \dots, m\}$ using sets with comparable coordinates:

$$A_k := \{i : 2^{-k}\|y\|_2 < |y_i| \leq 2^{-k+1}\|y\|_2\}, \quad k = 1, 2, \dots$$

Let $k_0 = \lceil \log m \rceil + 1$, so that $|y_i| \leq \frac{1}{m}\|y\|_2$ for all $i \in A_k$, $k > k_0$. Then the set $U = \bigcup_{k \leq k_0} A_k$ contains most of the energy of y :

$$\|y|_{U^c}\|_2 \leq \left(m \left(\frac{1}{m}\|y\|_2\right)^2\right)^{1/2} = \frac{1}{\sqrt{m}}\|y\|_2 \leq \frac{1}{\sqrt{2}}\|y\|_2.$$

Thus,

$$\left(\sum_{k \leq k_0} \|y|_{A_k}\|_2^2\right)^{1/2} = \|y|_U\|_2 = (\|y\|_2^2 - \|y|_{U^c}\|_2^2)^{1/2} \geq \frac{1}{\sqrt{2}}\|y\|_2.$$

Therefore, there exists $k \leq k_0$ such that

$$\|y|_{A_k}\|_2 \geq \frac{1}{\sqrt{2k_0}}\|y\|_2 \geq \frac{1}{2.5\sqrt{\log m}}\|y\|_2,$$

which completes the proof. \square

In our context, Lemma 3.7 applied to the vector $u|_J$ along with Lemma 3.6 directly implies:

Lemma 3.8 (Regularizing the energy) *We have*

$$\|u|_{J_0}\|_2 \geq \frac{0.32}{\sqrt{\log n}} \|v_0\|_2.$$

We now finish the proof of Theorem 3.1.

To show the first claim that J_0 is nonempty, we note that $v_0 \neq 0$. Indeed, otherwise by (3.3), we have $I \subset \text{supp}(v)$, so by the definition of the residual in the algorithm, we would have $r = 0$ at the start of the current iteration, which is a contradiction. Then $J_0 \neq \emptyset$ by Lemma 3.8.

The second claim that $J_0 \cap I = \emptyset$, is also simple. Indeed, recall that by the definition of the algorithm, $r = P_{F^\perp} \in F^\perp = (\text{range}(\Phi_I))^\perp$. It follows that the observation vector $u = \Phi^* r$ satisfies $u|_I = 0$. Since by its definition the set J contains only nonzero coordinates of u , we have $J \cap I = \emptyset$. Since $J_0 \subset J$, the second claim $J_0 \cap I = \emptyset$ follows.

The nontrivial part of the theorem is its last claim, inequality (3.1). Suppose it fails. Namely, suppose that $|J_0 \cap \text{supp}(v)| < \frac{1}{2}|J_0|$, and thus

$$|J_0 \setminus \text{supp}(v)| > \frac{1}{2}|J_0|.$$

Set $\Lambda = J_0 \setminus \text{supp}(v)$. 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 :

$$\|u|_\Lambda\|_2 > \frac{1}{\sqrt{5}} \|u|_{J_0}\|_2 \geq \frac{1}{7\sqrt{\log n}} \|v_0\|_2, \quad (3.6)$$

where the last inequality holds by Lemma 3.8.

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

$$\|u|_\Lambda\|_2 \leq \|u|_\Lambda - u_0|_\Lambda\|_2 + \|u_0|_\Lambda\|_2. \quad (3.7)$$

Since $\Lambda \subset J$ and using Lemma 3.5, we have

$$\|u|_\Lambda - u_0|_\Lambda\|_2 \leq 2.4\varepsilon \|v_0\|_2.$$

Furthermore, by definition (3.3) of v_0 , we have $v_0|_\Lambda = 0$. So, by Part 1 of Proposition 3.2,

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

Using the last two inequalities in (3.7), we conclude that

$$\|u|_\Lambda\|_2 \leq 4.43\varepsilon \|v_0\|_2.$$

This is a contradiction to (3.6) so long as $\varepsilon \leq 0.03/\sqrt{\log n}$. This proves Theorem 3.1. \square

4 Implementation and Empirical Performance of ROMP

4.1 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). 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 Φ^* by a vector, and solving the *least squares problem* with the $N \times |I|$ matrix Φ_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 Φ 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 Φ_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 [20].

4.2 Nonsparse Signals

In many applications, one needs to recover a signal v which is not sparse, but close to being sparse in some way. Such are, for example, compressible signals, whose coefficients decay at a certain rate (see [3, 7]). To make ROMP work for such signals, one can replace the stopping criterion of exact recovery $r = 0$ by “repeat n times or until $r = 0$, whichever occurs first.” Note that we could amend the algorithm for sparse signals in this way as well, allowing for a specific level of accuracy to be attained before terminating.

We recently proved that ROMP is stable and guarantees approximate recovery of nonsparse signals with noisy measurements; this will be discussed in a forthcoming paper.

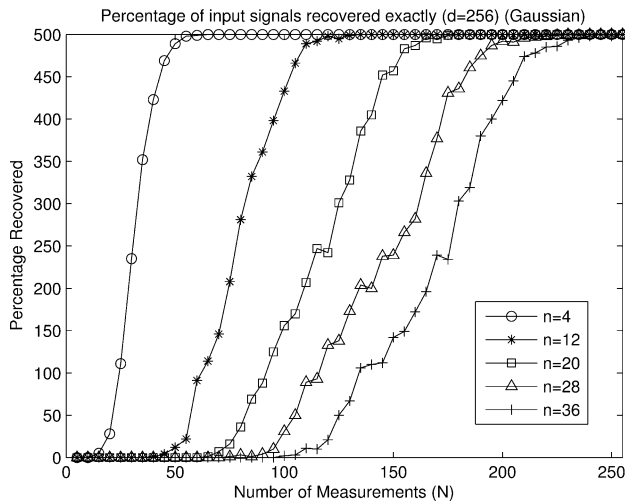


Fig. 1 The percentage of sparse flat signals exactly recovered by ROMP as a function of the number of measurements N in dimension $d = 256$ for various levels of sparsity n

4.3 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 Φ and then perform 500 independent trials. The results we obtained using Bernoulli measurement matrices were very similar. 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 Φ). In the cases where we wish to generate flat signals, we then set these components to one.² 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. We also performed this same test for sparse compressible signals and found the results very similar to those in

²Our work, as well as the analysis of Gilbert and Tropp [20], show that this is a challenging case for ROMP (and OMP).

Fig. 1. Our results show that performance of ROMP is very similar to that of OMP which can be found in [20].

Figure 2 depicts a plot of the values for N and n at which 99% of sparse flat signals are recovered exactly. This plot was generated with $d = 256$. The horizontal axis represents the number of measurements N , and the vertical axis the sparsity level n .

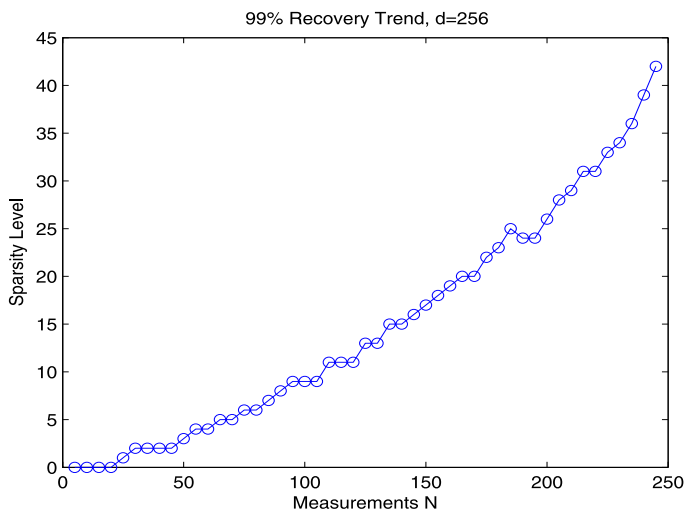


Fig. 2 The 99% recovery limit as a function of the sparsity n and the number of measurements N for sparse flat signals

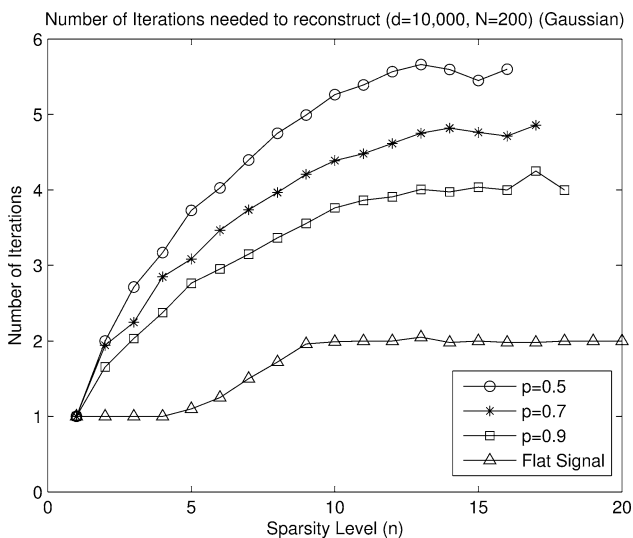


Fig. 3 The number of iterations executed by ROMP as a function of the sparsity n in dimension $d = 10,000$ with $N = 200$

Theorem 1.3 guarantees that ROMP runs with at most $O(n)$ iterations. Figure 3 depicts the number of iterations executed by ROMP for $d = 10,000$ and $N = 200$. ROMP was executed under the same setting as described above for sparse flat signals as well as sparse compressible signals for various values of p , and the number of iterations in each scenario was averaged over the 500 trials. These averages were plotted against the sparsity of the signal. As the plot illustrates, only 2 iterations were needed for flat signals even for sparsity n as high as 40. The plot also 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 suggests that for smaller values of p (meaning signals that decay more rapidly), ROMP needs more iterations. However, it shows that even in the case of $p = 0.5$, only 6 iterations are needed even for sparsity n as high as 20.

Acknowledgements The work was partially supported by the Alfred P. Sloan Foundation and by NSF DMS grants 0401032 and 0652617.

We would like to thank the referees for a thorough reading of the manuscript and making useful suggestions which greatly improved the paper.

References

1. E. Candès, Compressive sampling, in *Proceedings of International Congress of Mathematics*, vol. 3, Madrid, Spain, 2006, pp. 1433–1452.
2. E. Candès, J. Romberg, T. Tao, Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information, *IEEE Trans. Inf. Theory* **52**, 489–509 (2006).
3. E. Candès, T. Tao, Near-optimal signal recovery from random projections: universal encoding strategies, *IEEE Trans. Inf. Theory* **52**, 5406–5425 (2004).
4. E.J. Candès, T. Tao, Decoding by linear programming, *IEEE Trans. Inf. Theory* **51**, 4203–4215 (2005).
5. A. Cohen, W. Dahmen, R. DeVore, Compressed sensing and k -term approximation, Manuscript (2007).
6. Compressed sensing, webpage, <http://www.dsp.ece.rice.edu/cs/>.
7. D. Donoho, Compressed sensing, *IEEE Trans. Inf. Theory* **52**, 1289–1306 (2006).
8. D. Donoho, M. Elad, V. Temlyakov, Stable recovery of sparse overcomplete representations in the presence of noise, *IEEE Trans. Inf. Theory* **52**, 6–18 (2006).
9. D. Donoho, M. Elad, V. Temlyakov, On the Lebesgue type inequalities for greedy approximation, *J. Approx. Theory* **147**, 185–195 (2007).
10. D. Donoho, P. Stark, Uncertainty principles and signal recovery, *SIAM J. Appl. Math.* **49**, 906–931 (1989).
11. A. Gilbert, S. Muthukrishnan, M. Strauss, Approximation of functions over redundant dictionaries using coherence, in *The 14th Annual ACM–SIAM Symposium on Discrete Algorithms* (2003).
12. A. Gilbert, M. Strauss, J. Tropp, R. Vershynin, Algorithmic linear dimension reduction in the L_1 norm for sparse vectors, submitted. Conference version, in *Algorithmic Linear Dimension Reduction in the L_1 Norm for Sparse Vectors*, Allerton, 2006. 44th Annual Allerton Conference on Communication, Control, and Computing.
13. A. Gilbert, M. Strauss, J. Tropp, R. Vershynin, One sketch for all: fast algorithms for compressed sensing, in *STOC 2007. 39th ACM Symposium on Theory of Computing*, San Diego, 2007, to appear.
14. Y. Lyubarskii, R. Vershynin, Uncertainty principles and vector quantization, submitted.
15. S. Mendelson, A. Pajor, N. Tomczak-Jaegermann, Uniform uncertainty principle for Bernoulli and subgaussian ensembles, *Constr. Approx.*, submitted.
16. H. Rauhut, On the impossibility of uniform recovery using greedy methods, in *Sample Theory Signal Image Process.*, to appear.
17. M. Rudelson, R. Vershynin, On sparse reconstruction from Fourier and Gaussian measurements, *Commun. Pure Appl. Math.*, to appear. Conference version in *CISS 2006. 40th Annual Conference on Information Sciences and Systems*, Princeton.

18. D. Spielman, S.-H. Teng, Smoothed analysis of algorithms: why the simplex algorithm usually takes polynomial time, *J. ACM* **51**, 385–463 (2004).
19. V. Temlyakov, Nonlinear methods of approximation, *Found. Comput. Math.* **3**, 33–107 (2003).
20. J.A. Tropp, A.C. Gilbert, Signal recovery from random measurements via orthogonal matching pursuit, *IEEE Trans. Inf. Theory* **53**(12), 4655–4666 (2007).
21. R. Vershynin, Beyond Hirsch Conjecture: walks on random polytopes and smoothed complexity of the simplex method, submitted. Conference version in *FOCS 2006*. 47th Annual Symposium on Foundations of Computer Science, Berkeley, pp. 133–142.