Analysis of Orthogonal Matching Pursuit Using the Restricted Isometry Property

Mark A. Davenport, *Member, IEEE*, and Michael B. Wakin, *Member, IEEE*

Abstract—**Orthogonal matching pursuit (OMP) is the canonical greedy algorithm for sparse approximation. In this paper we demonstrate that the restricted isometry property (RIP) can be used for a very straightforward analysis of OMP. Our main conclusion is that the RIP of order $K + 1$ (with isometry constant $\delta < \frac{1}{3\sqrt{K}}$) is sufficient for OMP to exactly recover any K-sparse signal. The analysis relies on simple and intuitive observations about OMP and matrices which satisfy the RIP. For restricted classes of K-sparse signals (those that are highly compressible), a relaxed bound on the isometry constant is also established. A deeper understanding of OMP may benefit the analysis of greedy algorithms in general. To demonstrate this, we also briefly revisit the analysis of the regularized OMP (ROMP) algorithm.**

Index Terms—**Compressive sensing, greedy algorithms, orthogonal matching pursuit (OMP), redundant dictionaries, restricted isometry property (RIP), sparse approximation.**

I. INTRODUCTION

A. Orthogonal Matching Pursuit

ORTHOGONAL matching pursuit (OMP) is the canonical greedy algorithm for sparse approximation. Letting Φ denote a matrix of size $M \times N$ (where typically $M < N$) and y denote a vector in $\mathbb{R}^M$, the goal of OMP is to recover a coefficient vector $\hat{x} \in \mathbb{R}^N$ with roughly $K < M$ nonzero terms so that $\Phi\hat{x}$ equals y exactly or approximately. OMP is frequently used to find sparse representations for signals $y \in \mathbb{R}^M$ in settings where Φ represents an overcomplete dictionary for the signal space [1]–[3]. It is also commonly used in compressive sensing (CS), where $y = \Phi x$ represents compressive measurements of a sparse or nearly-sparse signal $x \in \mathbb{R}^N$ to be recovered [4]–[6].

One of the attractive features of OMP is its simplicity. The entire algorithm is specified in Algorithm 1, and it requires approximately the same number of lines of code to implement in a software package such as Matlab. Despite its simplicity, OMP is empirically competitive in terms of approximation performance [3], [7].

Theoretical analysis of OMP to date has concentrated primarily on two fronts. The first has involved the notion of a coherence parameter $\mu := \max_{i,j} |\langle \phi_i, \phi_j \rangle|$, where ϕ_i denotes column i of the matrix Φ. When the columns of Φ have unit norm and $\mu < \frac{1}{2K-1}$, it has been shown [3] that OMP will recover any K-sparse signal x from the measurements $y = \Phi x$. This guarantee is deterministic and applies to any matrix Φ having normalized columns and $\mu < \frac{1}{2K-1}$.

The second analytical front has involved the notion of probability. Suppose that $x \in \mathbb{R}^N$ with $\|x\|_0 := |\text{supp}(x)| \le K$ and that Φ is drawn from a suitable random distribution (independently of x) with $M = O(K \log(N))$ rows. Then with high probability, OMP will recover x exactly from the measurements $y = \Phi x$ [6]. It is not guaranteed, however, that any such fixed matrix will allow recovery of all sparse x simultaneously.

Algorithm 1 Orthogonal Matching Pursuit

---

input: Φ, y, stopping criterion

initialize: $r^0 = y$, $x^0 = 0$, $\Lambda^0 = \emptyset$, $\ell = 0$

while not converged **do**

match: $h^\ell = \Phi^T r^\ell$

identify: $\Lambda^{\ell+1} = \Lambda^\ell \cup \{\arg\max_j |h^\ell(j)|\}$
  (if multiple maxima exist, choose only one)

update: $x^{\ell+1} = \arg\min_{z:\text{supp}(z) \subseteq \Lambda^{\ell+1}} \|y - \Phi z\|_2$

$r^{\ell+1} = y - \Phi x^{\ell+1}$

$\ell = \ell + 1$

end while

output: $\hat{x} = x^\ell = \arg\min_{z:\text{supp}(z) \subseteq \Lambda^\ell} \|y - \Phi z\|_2$

B. Restricted Isometry Property

As an alternative to coherence and to probabilistic analysis, a large number of algorithms within the broader field of CS have been studied using the restricted isometry property (RIP) for the matrix Φ [8]. A matrix Φ satisfies the RIP of order K if there exists a constant $\delta \in (0, 1)$ such that

$$(1 - \delta)\|x\|_2^2 \le \|\Phi x\|_2^2 \le (1 + \delta)\|x\|_2^2 \qquad (1)$$

holds for all x such that $\|x\|_0 \le K$. In other words, Φ acts as an approximate isometry on the set of vectors that are K-sparse. Much is known about finding matrices that satisfy the RIP. For example, if we draw a random $M \times N$ matrix Φ whose entries

ϕ_{ij} are independent and identically distributed sub-Gaussian random variables, then provided that

$$M = O\left(\frac{K \log(N/K)}{\delta^2}\right) \qquad (2)$$

with high probability Φ will satisfy the RIP of order K [9], [10].

When it is satisfied, the RIP for a matrix Φ provides a sufficient condition to guarantee successful sparse recovery using a wide variety of algorithms [8], [11]–[19]. As an example, the RIP of order $2K$ (with isometry constant $\delta < \sqrt{2} - 1$) is a sufficient condition to permit ℓ_1-minimization (the canonical convex optimization problem for sparse approximation) to exactly recover any K-sparse signal and to approximately recover those that are nearly sparse [11]. The same RIP assumption is also a sufficient condition for robust recovery in noise using a modified ℓ_1-minimization [11].

Despite the considerable attention that has been paid to both OMP and the RIP, analysis of OMP using the RIP has been relatively elusive to date. However, several alternative greedy algorithms have been proposed—all essentially modifications of OMP—that are apparently much more amenable to RIP-based analysis. The regularized orthogonal matching pursuit (ROMP) [13], [14] and subspace pursuit (SP) [16] algorithms differ from OMP in the identification step, while the compressive sampling matching pursuit (CoSaMP) [15] and DThresh [17] algorithms differ from OMP in both the identification and the update steps. For each of these algorithms it has been shown that the RIP of order CK (where $C \geq 2$ is a constant depending on the algorithm) with δ adequately small is sufficient for exact recovery of K sparse signals.

C. Contributions

Our contributions in this paper are twofold. First, we begin in Section II with some very simple observations regarding OMP. Many of these facts are known to practitioners in the field but may not be obvious to a novice, and we feel that such readers may find value in a short exposition.

Critically, these observations also set the stage for our main results in Section III, in which we demonstrate that the RIP can be used for a very straightforward analysis of OMP. Our analysis revolves around three key facts: (1) that in each step of the algorithm, the residual vector r^ℓ can be written as a matrix times a sparse signal, (2) that this matrix satisfies the RIP, and (3) that consequently a sharp bound can be established for the vector h^ℓ of inner products. Our main conclusion, Theorem 3.1, states that the RIP of order $K + 1$ (with $\delta < \frac{1}{3\sqrt{K}}$) is sufficient for OMP to exactly recover any K-sparse signal in exactly K iterations. However, for restricted classes of K-sparse signals (those with sufficiently strong decay in the nonzero coefficients), a relaxed bound on the isometry constant can be used. We discuss such extensions of our results in Section IV. A deeper understanding of OMP may also benefit the analysis of greedy algorithms in general. To demonstrate this, we briefly revisit the analysis of the ROMP algorithm in Section IV.

D. Context

Let us place Theorem 3.1 in the context of the OMP literature. Using the RIP as a sufficient condition to guarantee OMP performance is apparently novel. Moreover, the fact that our bound requires only the RIP of order $K + 1$ is apparently unique among the published CS literature; much more common are results requiring the RIP of order $1.75K$ [12], $2K$ [11], [13], $3K$ [16], [18], $4K$ [15], and so on.[1] Of course, such results often permit the isometry constant to be much larger.[2]

If one wishes to use the RIP of order $K + 1$ as a sufficient condition for exact recovery of all K-sparse signals via OMP (as we have), then little improvement is possible in relaxing the isometry constant δ above $\frac{1}{3\sqrt{K}}$. In particular, there exists a matrix satisfying the RIP of order $K + 1$ with $\delta \leq \frac{1}{\sqrt{K}}$ for which there exists a K-sparse signal $x \in \mathbb{R}^N$ that cannot be recovered exactly via K iterations of OMP (this is conjectured in [16] with a suggestion for constructing such a matrix, and for the case $K = 2$ we have confirmed this via experimentation).

Unfortunately, from (2) we see that finding a matrix Φ satisfying the RIP of order $K + 1$ with an isometry constant $\delta < \frac{1}{3\sqrt{K}}$ will likely require $M = O(K^2 \log(N/K))$ random measurements. If one wishes to guarantee exact recovery of all K-sparse signals via OMP (as we have), then there is little room for further reducing M. In particular, it has been argued in a recent paper concerned with uniform guarantees for greedy algorithms [21] that when $M \lesssim K^{3/2}$, for most random $M \times N$ matrices Φ there will exist some K-sparse signal $x \in \mathbb{R}^N$ that cannot be recovered exactly via K iterations of OMP.

It is also worth comparing our RIP-based analysis with coherence-based analysis [3], as both techniques provide a sufficient condition for OMP to recover all K-sparse signals. It has been shown [6] that in a random $M \times N$ matrix, the coherence parameter μ is unlikely to be smaller than $\log(N)/\sqrt{M}$. Thus, to ensure $\mu < \frac{1}{2K-1}$, one requires $M = O(K^2 \log^2(N))$, which is roughly the same as what is required by our analysis. We note that neither result is strictly stronger than the other; we have confirmed experimentally that there exist matrices that satisfy our RIP condition but not the coherence condition, and vice versa.

Finally, we note that the aforementioned modifications of OMP (the ROMP, SP, CoSaMP, and DThresh algorithms) all have RIP-based guarantees of robust recovery in noise and stable recovery of nonsparse signals. Until recently, no such RIP-based or coherence-based guarantees had been established for OMP itself. However, there has been recent progress in using the RIP and similar conditions to analyze the performance of OMP on nonsparse signals [22]. The results of [22] can be adapted to provide a guarantee of exact recovery for sparse signals, but the assumptions required are stronger than the assumptions made in this work. Furthermore, a number of open questions remain concerning the performance of OMP on

---

[1]After the submission of this manuscript, it was shown in [20] that the RIP of order K with $\delta < 0.307$ is a sufficient condition for recovery via ℓ_1 minimization in the absence of noise.

[2]In general, it is important to note that a smaller order of the RIP is not necessarily a weaker requirement if the required constant is also significantly smaller. For example, Corollary 3.4 of [15] implies that if Φ satisfies the RIP of order $K + 1$ with constant δ, then Φ also satisfies the RIP of order $2K$ with constant 4δ.

nonsparse signals, and performance in the presence of noise has yet to be addressed. We speculate that our perspective may help to further the general understanding of OMP and perhaps provide a route to such guarantees. At present, however, this remains a topic of ongoing work.

E. Notation

Before proceeding, we set our notation. Suppose $\Lambda \subset \{1, 2, \ldots, N\}$. We let $\Lambda^c = \{1, 2, \ldots, N\}\backslash\Lambda$. By $x|_\Lambda$ we mean the length $|\Lambda|$ vector containing the entries of x indexed by Λ.

By Φ_Λ we mean the $M \times |\Lambda|$ matrix obtained by selecting the columns of Φ indexed by Λ, and by $\mathcal{R}(\Phi_\Lambda)$ we mean the range, or column space, of Φ_Λ. We will assume throughout that when $|\Lambda| \leq M$, Φ_Λ is full rank, in which case we let $\Phi_\Lambda^\dagger := (\Phi_\Lambda^T\Phi_\Lambda)^{-1}\Phi_\Lambda^T$ denote the Moore-Penrose pseudoinverse of Φ_Λ.

We denote the orthogonal projection operator onto $\mathcal{R}(\Phi_\Lambda)$ by $P_\Lambda := \Phi_\Lambda\Phi_\Lambda^\dagger$. Similarly, $P_\Lambda^\perp := (I - P_\Lambda)$ is the orthogonal projection operator onto the orthogonal complement of $\mathcal{R}(\Phi_\Lambda)$. We note that any orthogonal projection operator P obeys $P = P^T = P^2$.

Finally, we define $A_\Lambda := P_\Lambda^\perp\Phi$. This matrix is the result of orthogonalizing the columns of Φ against $\mathcal{R}(\Phi_\Lambda)$. It is, therefore, equal to zero on columns indexed by Λ.

II. OBSERVATIONS

Let us begin with some very simple observations regarding OMP as presented in Algorithm 1. The key idea is to try to iteratively estimate a set Λ that contains the locations of the nonzeros of x by starting with $\Lambda = \emptyset$ and then adding a new element to Λ in each iteration. In order to select which element to add, the algorithm also maintains a residual vector $r \notin \mathcal{R}(\Phi_\Lambda)$ that represents the component of the measurement vector y that cannot be explained by the columns of Φ_Λ. Specifically, at the beginning of the ℓth iteration, Λ^ℓ is our current estimate of $\text{supp}(x)$, and the residual r^ℓ is defined as $r^\ell = y - \Phi x^\ell$ where $\text{supp}(x^\ell) \subseteq \Lambda^\ell$. The element added to Λ^ℓ is the index of the column of Φ that has the largest inner product with r^ℓ.

Our first observation is that r^ℓ can be viewed as the orthogonalization of y against the previously chosen columns of Φ. To see this, note that the solution to the least squares problem in the update step is given by

$$x^\ell|_{\Lambda^\ell} = \Phi_{\Lambda^\ell}^\dagger y \quad \text{and} \quad x^\ell|_{(\Lambda^\ell)^c} = 0. \tag{3}$$

Thus, we observe that

$$r^\ell = y - \Phi x^\ell = y - \Phi_{\Lambda^\ell}\Phi_{\Lambda^\ell}^\dagger y = (I - P_{\Lambda^\ell})y = P_{\Lambda^\ell}^\perp y.$$

Note that it is not actually necessary to explicitly compute x^ℓ in order to calculate r^ℓ.

Our second observation is that, in the matching step, one may correlate r^ℓ either with the columns of Φ or with the columns of A_{Λ^ℓ}. To see this equivalence, observe that $r^\ell = P_{\Lambda^\ell}^\perp y = P_{\Lambda^\ell}^\perp P_{\Lambda^\ell}^\perp y = (P_{\Lambda^\ell}^\perp)^T P_{\Lambda^\ell}^\perp y$ and so

$$h^\ell = \Phi^T r^\ell = \Phi^T (P_{\Lambda^\ell}^\perp)^T P_{\Lambda^\ell}^\perp y = A_{\Lambda^\ell}^T r^\ell. \tag{4}$$

Incidentally, along these same lines, we observe that

$$h^\ell = \Phi^T r^\ell = \Phi^T P_{\Lambda^\ell}^\perp y = \Phi^T (P_{\Lambda^\ell}^\perp)^T y = A_{\Lambda^\ell}^T y.$$

From this, we note that it is not actually necessary to explicitly compute r^ℓ in order to calculate the inner products during the matching step; in fact, the original formulation of OMP was stated with instructions to orthogonalize the remaining columns of Φ against those previously chosen and merely correlate the resulting vectors against y [1], [2]. Additionally, we recall that, in A_{Λ^ℓ}, all columns indexed by Λ^ℓ will be zero. It follows that

$$h^\ell(j) = 0 \quad \forall j \in \Lambda^\ell \tag{5}$$

and so, since $\Lambda^\ell = \Lambda^{\ell-1} \cup \{j^*\}$ with $j^* \notin \Lambda^{\ell-1}$

$$|\Lambda^\ell| = \ell. \tag{6}$$

Our third observation is that, in the case of noise-free measurements $y = \Phi x$, we may write

$$r^\ell = P_{\Lambda^\ell}^\perp y = P_{\Lambda^\ell}^\perp \Phi x = A_{\Lambda^\ell} x.$$

Again recalling that all columns of A_{Λ^ℓ} indexed by Λ^ℓ are zero, we thus note that when $\text{supp}(x) \subseteq \Lambda^\ell$, $r^\ell = 0$, and from (3) we also know that $x^\ell = x$ exactly. It will also be useful to note that for the same reason, we can also write

$$r^\ell = A_{\Lambda^\ell}\widetilde{x}^\ell \tag{7}$$

where

$$\widetilde{x}^\ell|_{\Lambda^\ell} = 0 \quad \text{and} \quad \widetilde{x}^\ell|_{(\Lambda^\ell)^c} = x|_{(\Lambda^\ell)^c}. \tag{8}$$

III. ANALYSIS

Our analysis of OMP will center on the vector h^ℓ. In light of (4) and (7), we see that A_{Λ^ℓ} plays a role both in *constructing* and in *analyzing* the residual vector. In Lemma 3.2 below, we show that the matrix A_{Λ^ℓ} satisfies a modified version of the RIP. This allows us to very precisely bound the values of the inner products in the vector h^ℓ.

We begin with two elementary lemmas whose proofs are given in the Appendix and are also available in [23]. Our first result, which is a straightforward generalization of Lemma 2.1 of [11], states that RIP operators approximately preserve inner products between sparse vectors.

Lemma 3.1: Let $u, v \in \mathbb{R}^N$ be given, and suppose that a matrix Ψ satisfies the RIP of order $\max(\|u + v\|_0, \|u - v\|_0)$ with isometry constant δ. Then

$$|\langle \Psi u, \Psi v \rangle - \langle u, v \rangle| \leq \delta \|u\|_2 \|v\|_2. \tag{9}$$

One consequence of this result is that sparse vectors that are orthogonal in $\mathbb{R}^N$ remain nearly orthogonal after the application of Ψ. From this observation, it was demonstrated independently in [24] and [16] that if Φ has the RIP, then A_Λ satisfies a modified version of the RIP.

Lemma 3.2: Suppose that Φ satisfies the RIP of order K with isometry constant δ, and let $\Lambda \subset \{1, 2, \ldots, N\}$. If $|\Lambda| < K$ then

$$\left(1 - \frac{\delta}{1-\delta}\right) \|u\|_2^2 \leq \|A_\Lambda u\|_2^2 \leq (1+\delta)\|u\|_2^2 \qquad (10)$$

for all $u \in \mathbb{R}^N$ such that $\|u\|_0 \leq K - |\Lambda|$ and $\operatorname{supp}(u) \cap \Lambda = \emptyset$.

In other words, if Φ satisfies the RIP of order K, then A_Λ acts as an approximate isometry on every $(K - |\Lambda|)$-sparse vector supported on Λ^c. From (7), we recall that the residual vector in OMP is formed by applying A_{Λ^ℓ} to a sparse vector supported on $(\Lambda^\ell)^c$. Combining the above results, then, we may bound the inner products $h^\ell(j)$ as follows.

Lemma 3.3: Let $\Lambda \subset \{1, 2, \ldots, N\}$ and suppose $\widetilde{x} \in \mathbb{R}^N$ with $\operatorname{supp}(\widetilde{x}) \cap \Lambda = \emptyset$. Define

$$h = A_\Lambda^T A_\Lambda \widetilde{x}. \qquad (11)$$

Then if Φ satisfies the RIP of order $\|\widetilde{x}\|_0 + |\Lambda| + 1$ with isometry constant δ, we have

$$|h(j) - \widetilde{x}(j)| \leq \frac{\delta}{1-\delta} \|\widetilde{x}\|_2 \qquad (12)$$

for all $j \notin \Lambda$.

Proof: From Lemma 3.2, we have that the restriction of A_Λ to the columns indexed by Λ^c satisfies the RIP of order $(\|\widetilde{x}\|_0 + |\Lambda| + 1) - |\Lambda| = \|\widetilde{x}\|_0 + 1$ with isometry constant $\delta/(1-\delta)$. By the definition of h, we also know that

$$h(j) = \langle A_\Lambda \widetilde{x}, A_\Lambda e_j \rangle$$

where e_j denotes the jth vector from the cardinal basis. Now, suppose $j \notin \Lambda$. Then because $\|\widetilde{x} \pm e_j\|_0 \leq \|\widetilde{x}\|_0 + 1$ and $\operatorname{supp}(\widetilde{x} \pm e_j) \cap \Lambda = \emptyset$, we conclude from Lemma 3.1 that

$$|h(j) - \widetilde{x}(j)| = |\langle A_\Lambda \widetilde{x}, A_\Lambda e_j \rangle - \langle \widetilde{x}, e_j \rangle| \leq \frac{\delta}{1-\delta} \|\widetilde{x}\|_2 \|e_j\|_2.$$

Noting that $\|e_j\|_2 = 1$, we reach the desired conclusion. $\square$

With this bound on the inner products $h^\ell(j)$, we may derive a sufficient condition under which the identification step of OMP will succeed.

Corollary 3.1: Suppose that Λ, Φ, $\widetilde{x}$ meet the assumptions specified in Lemma 3.3, and let h be as defined in (11). If

$$\|\widetilde{x}\|_\infty > \frac{2\delta}{1-\delta} \|\widetilde{x}\|_2 \qquad (13)$$

we are guaranteed that $\arg\max_j |h(j)| \in \operatorname{supp}(\widetilde{x})$.

Proof: If (12) is satisfied, then for indices $j \notin \operatorname{supp}(\widetilde{x})$, we will have $|h(j)| \leq \frac{\delta}{1-\delta}\|\widetilde{x}\|_2$ [recall from (5) that $h(j) = 0$ for $j \in \Lambda$]. If (13) is satisfied, then there exists some $j \in \operatorname{supp}(\widetilde{x})$ with $|\widetilde{x}(j)| > \frac{2\delta}{1-\delta}\|\widetilde{x}\|_2$. From (12), and the triangle inequality, we conclude that for this index j, $|h(j)| > \frac{\delta}{1-\delta}\|\widetilde{x}\|_2$. $\square$

By choosing δ small enough, it is possible to guarantee that the condition (13) is satisfied. In particular, the lemma below follows from standard arguments.

Lemma 3.4: For any $u \in \mathbb{R}^N$, $\|u\|_\infty \geq \|u\|_2 / \sqrt{\|u\|_0}$.

Putting these results together, we can now establish our main theorem concerning OMP.

Theorem 3.1: Suppose that Φ satisfies the RIP of order $K+1$ with isometry constant $\delta < \frac{1}{3\sqrt{K}}$. Then for any $x \in \mathbb{R}^N$ with $\|x\|_0 \leq K$, OMP will recover x exactly from $y = \Phi x$ in K iterations.

Proof: The proof works by induction. We start with the first iteration where $h^0 = \Phi^T \Phi x$ and note that $\Phi = A_\emptyset$. Because $\|x\|_0 \leq K$, Lemma 3.4 states that $\|x\|_\infty \geq \frac{\|x\|_2}{\sqrt{K}}$. One can also check that $\delta < \frac{1}{3\sqrt{K}}$ implies that $\frac{2\delta}{1-\delta} < \frac{1}{\sqrt{K}}$. Therefore, we are guaranteed that (13) is satisfied, and so from Corollary 3.1 we conclude that $\arg\max_j |h^0(j)| \in \operatorname{supp}(x)$.

We now consider the general induction step. Suppose that we are at iteration ℓ and that all previous iterations have succeeded, by which we mean that $\Lambda^\ell \subseteq \operatorname{supp}(x)$. From (8), we know that $\operatorname{supp}(\widetilde{x}^\ell) \cap \Lambda^\ell = \emptyset$ and that $\|\widetilde{x}^\ell\|_0 \leq K - \ell$. From (6), we know that $|\Lambda^\ell| = \ell$. By assumption, Φ satisfies the RIP of order $K + 1 = (K - \ell) + \ell + 1 \geq \|\widetilde{x}^\ell\|_0 + |\Lambda^\ell| + 1$. Finally, using Lemma 3.4, we have that

$$\|\widetilde{x}^\ell\|_\infty \geq \frac{\|\widetilde{x}^\ell\|_2}{\sqrt{K - \ell}} \geq \frac{\|\widetilde{x}^\ell\|_2}{\sqrt{K}} > \frac{2\delta}{1-\delta}\|\widetilde{x}^\ell\|_2.$$

From Corollary 3.1, we conclude that $\arg\max_j |h^\ell(j)| \in \operatorname{supp}(\widetilde{x}^\ell)$ and, hence, $\Lambda^{\ell+1} \subseteq \operatorname{supp}(x)$. $\square$

IV. EXTENSIONS

A. Strongly-Decaying Sparse Signals

For even moderate values of the isometry constant δ there exist sparse signals that we can ensure are recovered exactly. For example, if the decay of coefficients is sufficiently strong in a sparse signal, we may use Lemma 3.3 to ensure that the signal entries are recovered in the order of their magnitude.

For any $x \in \mathbb{R}^N$ with $\|x\|_0 \leq K$ we denote by $x'(j)$ the entries of x ordered by magnitude, i.e.,

$$|x'(1)| \geq |x'(2)| \geq \cdots \geq |x'(K)| \geq 0$$

with $x'(K + 1) = x'(K + 2) = \cdots = x'(N) = 0$.

Theorem 4.1: Suppose that Φ satisfies the RIP of order $K+1$ with isometry constant $\delta < \frac{1}{3}$. Suppose $x \in \mathbb{R}^N$ with $\|x\|_0 \leq K$ and that for all $j \in \{1, 2, \ldots, K - 1\}$

$$\frac{|x'(j)|}{|x'(j+1)|} \geq \alpha.$$

If

$$\alpha > \frac{1 + 2\frac{\delta}{1-\delta}\sqrt{K - 1}}{1 - 2\frac{\delta}{1-\delta}} \qquad (14)$$

then OMP will recover x exactly from $y = \Phi x$ in K iterations.

Proof: The proof again proceeds by induction. At each stage, OMP will choose the largest entry of $\widetilde{x}^\ell$. To see this, note that by (12) we have $|h^\ell(j) - \widetilde{x}^\ell(j)| \leq \frac{\delta}{1-\delta}\|\widetilde{x}^\ell\|_2$. The

nonzero entries of $\tilde{x}^\ell$ will be comprised of $x'(\ell+1), x'(\ell+2), \ldots, x'(K)$. Thus

$$\begin{aligned}
\|\tilde{x}^\ell\|_2 &\leq \sqrt{|x'(\ell+1)|^2 + (K-1)\frac{|x'(\ell+1)|^2}{\alpha^2}} \\
&= \frac{|x'(\ell+1)|}{\alpha}\sqrt{\alpha^2 + (K-1)} \\
&\leq \frac{|x'(\ell+1)|}{\alpha}\left(\alpha + \sqrt{K-1}\right).
\end{aligned}$$

Now, for the specific index j at which $\tilde{x}^\ell$ has its largest entry, we have

$$\begin{aligned}
|h^\ell(j)| &\geq |x'(\ell+1)| - \frac{\delta}{1-\delta}\frac{|x'(\ell+1)|}{\alpha}\left(\alpha + \sqrt{K-1}\right) \\
&= \frac{|x'(\ell+1)|}{\alpha}\left(\alpha - \frac{\delta}{1-\delta}\left(\alpha + \sqrt{K-1}\right)\right) \quad (15)
\end{aligned}$$

while for all other values of j we have

$$\begin{aligned}
|h^\ell(j)| &\leq |x'(\ell+2)| + \frac{\delta}{1-\delta}\frac{|x'(\ell+1)|}{\alpha}\left(\alpha + \sqrt{K-1}\right) \\
&\leq \frac{|x'(\ell+1)|}{\alpha}\left(1 + \frac{\delta}{1-\delta}\left(\alpha + \sqrt{K-1}\right)\right). \quad (16)
\end{aligned}$$

From (14), it follows that (15) is greater than (16). $\qquad\square$

B. Analysis of Other Orthogonal Greedy Algorithms

We now demonstrate that the techniques used above can also be used to analyze other orthogonal greedy algorithms. We focus on ROMP [13], [14] for the purpose of illustration, but similar methods should be able to simplify the analysis of other orthogonal greedy algorithms such as SP [16].[3]

We first briefly describe the difference between ROMP and OMP, which lies only in the identification step: whereas OMP adds only one index to Λ^ℓ at each iteration, ROMP adds up to K indices to Λ^ℓ at each iteration. Specifically, ROMP first selects the indices corresponding to the K largest elements in magnitude of h^ℓ (or all nonzero elements of h^ℓ if h^ℓ has fewer than K nonzeros), and denotes this set as Ω^ℓ. The next step is to *regularize* this set so that the values are comparable in magnitude. To do this, define $R(\Omega^\ell) := \{\Omega \subseteq \Omega^\ell : |h^\ell(i)| \leq 2|h^\ell(j)| \,\forall i, j \in \Omega\}$, and set

$$\Omega_0^\ell := \mathrm{argmax}_{\Omega \in R(\Omega^\ell)}\|h^\ell|_\Omega\|_2$$

i.e., Ω_0^ℓ is the set with maximal energy among all regularized subsets of Ω^ℓ. Finally, setting $\Lambda^{\ell+1} = \Lambda^\ell \cup \Omega_0^\ell$, the remainder of the ROMP algorithm is identical to OMP.

In order to analyze ROMP, we will need only two preliminary lemmas from [13], which we state without proof. Note that Lemma 4.1, which is essentially a generalization of Lemma 3.3, is stated using slightly weaker assumptions than are used in

[13] and, to be consistent with the rest of our paper, uses the quadratic form of the RIP (whereas [13] uses the nonquadratic form). However, the present version can easily be obtained using the same proof techniques.

Lemma 4.1 [(1) in Prop. 3.2 of [13]]: Let $\Gamma \subset \{1, 2, \ldots, N\}$ and $x \in \mathbb{R}^N$ be given. Then if Ψ satisfies the RIP of order $|\mathrm{supp}(x) \cup \Gamma|$ with isometry constant δ, we have

$$\|(\Psi^T \Psi x)|_\Gamma - x|_\Gamma\|_2 \leq \delta\|x\|_2.$$

Lemma 4.2 (Lemma 3.7 of [13]): Let $u \in \mathbb{R}^K$, $K > 1$, be arbitrary. Then there exists a subset $\Gamma \subseteq \{1, \ldots, K\}$ such that $|u(i)| \leq 2|u(j)|$ for all $i, j \in \Gamma$ and

$$\|u|_\Gamma\|_2 \geq \frac{1}{2.5\sqrt{\log_2 K}}\|u\|_2.$$

Using these lemmas, we now provide a simplified proof of the main result of [13] concerning the recovery of sparse signals using ROMP.[4]

Theorem 4.2: Suppose that Φ satisfies the RIP of order $3K$ with isometry constant $\delta \leq 0.13/\sqrt{\log_2 K}$. Then for any $x \in \mathbb{R}^N$ with $\|x\|_0 \leq K$, ROMP will recover x exactly from $y = \Phi x$ in at most K iterations.

Proof: The proof works by showing that at each iteration

$$|\Omega_0^\ell \cap \mathrm{supp}(x)| \geq \frac{1}{2}|\Omega_0^\ell|. \quad (17)$$

If (17) is satisfied for $0, 1, \ldots, \ell-1$, then at iteration ℓ, we have that

$$|\Lambda^\ell \cap \mathrm{supp}(x)| \geq \frac{1}{2}|\Lambda^\ell|. \quad (18)$$

It follows that, before $|\Lambda^\ell|$ exceeds $2K$, we will have $\mathrm{supp}(x) \subseteq \Lambda^\ell$. Because Φ satisfies the RIP of order $3K > 2K$, at termination, Φ_{Λ^ℓ} will be full rank. From (3) we conclude that $x^\ell = x$ exactly.

To prove (17), we again proceed by induction. Hence, we assume that (17) holds for $0, 1, \ldots, \ell-1$, and thus (18) holds for iteration ℓ. We next assume for the sake of contradiction that (17) does not hold for iteration ℓ, i.e., that

$$|\Omega_0^\ell \setminus \mathrm{supp}(x)| > \frac{1}{2}|\Omega_0^\ell|. \quad (19)$$

Define the sets $T = \Omega_0^\ell \setminus \mathrm{supp}(x)$ and $S = \mathrm{supp}(x) \setminus \Lambda^\ell = \mathrm{supp}(\tilde{x}^\ell)$, where $\tilde{x}^\ell$ is defined as in (8). Recall that we can write $h^\ell = A_{\Lambda^\ell}^T A_{\Lambda^\ell} \tilde{x}^\ell$. Thus, using the assumption that $|T| > \frac{1}{2}|\Omega_0^\ell|$ and the facts that $T \subseteq \Omega_0^\ell$ and $\Omega_0^\ell \in R(\Omega^\ell)$, one can show that

$$\|h^\ell|_T\|_2 \geq \frac{1}{\sqrt{5}}\|h^\ell|_{\Omega_0^\ell}\|_2. \quad (20)$$

We now observe that

$$\|h^\ell|_{\Omega_0^\ell}\|_2 \geq \frac{1}{2.5\sqrt{\log_2 K}}\|h^\ell|_{\Omega^\ell}\|_2 \quad (21)$$

---

[3]Some of the greedy algorithms that have been proposed recently, such as CoSaMP [15] and DThresh [17], do not orthogonalize the residual against the previously chosen columns at each iteration, and so the techniques above cannot be directly applied to these algorithms. However, this orthogonalization step could easily be added (which in the case of CoSaMP yields an algorithm nearly identical to SP). Orthogonalized versions of these algorithms could then be studied using these techniques.

[4]Note that we assume that Φ satisfies the RIP of order $3K$ with constant $\delta \leq 0.13/\sqrt{\log_2 K}$. Using Corollary 3.4 of [15], we can replace this with the assumption that Φ satisfies the RIP of order $2K$ with constant $\delta \leq .043/\sqrt{\log_2 K}$.

which follows from Lemma 4.2 and the fact that Ω_0^ℓ is the maximal regularizing set. From the maximality of Ω^ℓ and the fact that $|S| \leq K$, we have that $\|h^\ell|_{\Omega^\ell}\|_2 \geq \|h^\ell|_S\|_2$, so that by combining (20) and (21), we obtain

$$\|h^\ell|_T\|_2 \geq \frac{1}{2.5\sqrt{5 \log_2 K}} \|h^\ell|_S\|_2. \qquad (22)$$

Note that $|S \cup \operatorname{supp}(\widetilde{x}^\ell)| = |S| \leq K$ and since $|\Lambda^\ell| \leq 2K$, from Lemma 3.2, we have that A_{Λ^ℓ} satisfies the RIP of order at least K with constant $\delta/(1-\delta)$, thus Lemma 4.1 implies that

$$\|h^\ell|_S - \widetilde{x}^\ell|_S\|_2 \leq \frac{\delta}{1-\delta} \|\widetilde{x}^\ell\|_2.$$

Since $\widetilde{x}^\ell|_{S^c} = 0$, $\|h^\ell|_S - \widetilde{x}^\ell|_S\|_2 \geq \|\widetilde{x}^\ell\|_2 - \|h^\ell|_S\|_2$, and thus

$$\|h^\ell|_S\|_2 \geq \frac{1-2\delta}{1-\delta} \|\widetilde{x}^\ell\|_2.$$

Hence

$$\|h^\ell|_T\|_2 \geq \frac{(1-2\delta)/(1-\delta)}{2.5\sqrt{5 \log_2 K}} \|\widetilde{x}^\ell\|_2. \qquad (23)$$

On the other hand, since $|\operatorname{supp}(\widetilde{x}^\ell)| + |\Lambda^\ell \cap \operatorname{supp}(x)| = K$, from (18), we obtain that $|\operatorname{supp}(\widetilde{x}^\ell)| \leq K - \frac{1}{2}|\Lambda^\ell|$. Thus, $|T \cup \operatorname{supp}(\widetilde{x}^\ell)| \leq |T| + |\operatorname{supp}(\widetilde{x}^\ell)| \leq 2K - \frac{1}{2}|\Lambda^\ell|$. Furthermore, A_{Λ^ℓ} satisfies the RIP of order $3K - |\Lambda^\ell| = 3K - \frac{1}{2}|\Lambda^\ell| - \frac{1}{2}|\Lambda^\ell|$. Since $|\Lambda^\ell| \leq 2K$, we have that A_{Λ^ℓ} satisfies the RIP of order at least $2K - \frac{1}{2}|\Lambda^\ell|$ with constant $\delta/(1-\delta)$. Thus, Lemma 4.1 also implies that

$$\|h^\ell|_T\|_2 = \|h^\ell|_T - \widetilde{x}^\ell|_T\|_2 \leq \frac{\delta}{1-\delta} \|\widetilde{x}^\ell\|_2. \qquad (24)$$

This is a contradiction whenever the right-hand-side of (23) is greater than the right-hand-side of (24), which occurs when $\delta < 1/(2 + 2.5\sqrt{5 \log_2 K})$. Since $\log_2 K \geq 1$, we can replace this with the slightly stricter condition $\delta < 1/((2 + 2.5\sqrt{5})\sqrt{\log_2 K}) \approx 0.1317/\sqrt{\log_2 K}$. $\qquad \square$

Observe that when $K = 1$, this proof (as well as the proofs in [13] and [14]) break down since Lemma 4.2 does not apply. However, when $K = 1$ the ROMP algorithm simply reduces to OMP. In this case we can apply Theorem 3.1 to verify that ROMP succeeds when $K = 1$ provided that Φ satisfies the RIP of order 2 with isometry constant $\delta < 1/3$.

APPENDIX

1) Proof of Lemma 3.1: We first assume that $\|u\|_2 = \|v\|_2 = 1$. From the fact that

$$\|u \pm v\|_2^2 = \|u\|_2^2 + \|v\|_2^2 \pm 2\langle u, v \rangle = 2 \pm 2\langle u, v \rangle$$

and since Ψ satisfies the RIP, we have that

$$(1-\delta)(2 \pm 2\langle u, v \rangle) \leq \|\Psi u \pm \Psi v\|_2^2 \leq (1+\delta)(2 \pm 2\langle u, v \rangle).$$

From the parallelogram identity, we obtain

$$\begin{aligned}
\langle \Psi u, \Psi v \rangle &= \frac{1}{4}\left(\|\Psi u + \Psi v\|_2^2 - \|\Psi u - \Psi v\|_2^2\right) \\
&\leq \frac{(1 + \langle u, v \rangle)(1 + \delta) - (1 - \langle u, v \rangle)(1 - \delta)}{2} \\
&= \langle u, v \rangle + \delta.
\end{aligned}$$

Similarly, one can show that $\langle \Psi u, \Psi v \rangle \geq \langle u, v \rangle - \delta$, and thus $|\langle \Psi u, \Psi v \rangle - \langle u, v \rangle| \leq \delta$. The result follows for u, v with arbitrary norm from the bilinearity of the inner product. $\qquad \square$

2) Proof of Lemma 3.2: From the definition of A_Λ, we may decompose $A_\Lambda u$ as $A_\Lambda u = \Phi u - P_\Lambda \Phi u$. Since P_Λ is an orthogonal projection, we can write

$$\|\Phi u\|_2^2 = \|P_\Lambda \Phi u\|_2^2 + \|A_\Lambda u\|_2^2. \qquad (25)$$

Our goal is to show that $\|\Phi u\|_2 \approx \|A_\Lambda u\|_2$, or equivalently, that $\|P_\Lambda \Phi u\|_2$ is small. Towards this end, we note that since $P_\Lambda \Phi u$ is orthogonal to $A_\Lambda u$

$$\begin{aligned}
\langle P_\Lambda \Phi u, \Phi u \rangle &= \langle P_\Lambda \Phi u, P_\Lambda \Phi u + A_\Lambda u \rangle \\
&= \langle P_\Lambda \Phi u, P_\Lambda \Phi u \rangle + \langle P_\Lambda \Phi u, A_\Lambda u \rangle \\
&= \|P_\Lambda \Phi u\|_2^2. \qquad (26)
\end{aligned}$$

Since P_Λ is a projection onto $\mathcal{R}(\Phi_\Lambda)$ there exists a $z \in \mathbb{R}^N$ with $\operatorname{supp}(z) \subseteq \Lambda$ such that $P_\Lambda \Phi u = \Phi z$. Furthermore, by assumption, $\operatorname{supp}(u) \cap \Lambda = \emptyset$. Hence $\langle u, z \rangle = 0$, and from the RIP and Lemma 3.1

$$\frac{|\langle P_\Lambda \Phi u, \Phi u \rangle|}{\|P_\Lambda \Phi u\|_2 \|\Phi u\|_2} = \frac{|\langle \Phi z, \Phi u \rangle|}{\|\Phi z\|_2 \|\Phi u\|_2} \leq \frac{|\langle \Phi z, \Phi u \rangle|}{(1-\delta)\|z\|_2 \|u\|_2} \leq \frac{\delta}{1-\delta}.$$

Combining this with (26), we obtain

$$\|P_\Lambda \Phi u\|_2 \leq \frac{\delta}{1-\delta} \|\Phi u\|_2.$$

Since we trivially have that $\|P_\Lambda \Phi u\|_2 \geq 0$, we can combine this with (25) to obtain

$$\left(1 - \left(\frac{\delta}{1-\delta}\right)^2\right) \|\Phi u\|_2^2 \leq \|A_\Lambda u\|_2^2 \leq \|\Phi u\|_2^2.$$

Since $\|u\|_0 \leq K$, we can use the RIP to obtain

$$\left(1 - \left(\frac{\delta}{1-\delta}\right)^2\right)(1-\delta)\|u\|_2^2 \leq \|A_\Lambda u\|_2^2 \leq (1+\delta)\|u\|_2^2,$$

which simplifies to (10). $\qquad \square$

REFERENCES

[1] Y. Pati, R. Rezaifar, and P. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," presented at the 27th Asilomar Conf. Signals, Systems and Comput., Nov. 1993.

[2] G. Davis, S. Mallat, and Z. Zhang, "Adaptive time-frequency decompositions," *SPIE J. Opt. Eng.*, vol. 33, no. 7, pp. 2183–2191, 1994.

[3] J. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.

[4] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[5] E. Candès, "Compressive sampling," in *Proc. Int. Congr. Mathematics*, Madrid, Spain, 2006, vol. 3, pp. 1433–1452.

[6] J. Tropp and A. Gilbert, "Signal recovery from partial information via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[7] S. Kunis and H. Rauhut, "Random sampling of sparse trigonometric polynomials II – Orthogonal matching pursuit versus basis pursuit," *Found. Comput. Math.*, vol. 8, no. 6, pp. 737–763, 2008.

[8] E. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[9] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Const. Approx.*, vol. 28, no. 3, pp. 253–263, 2008.

[10] R. DeVore, G. Petrova, and P. Wojtaszczyk, "Instance-optimality in probability with an ℓ_1-minimization decoder," *Appl. Comput. Harmon. Anal.*, vol. 27, no. 3, pp. 275–288, 2009.

[11] E. Candès, "The restricted isometry property and its implications for compressed sensing," *Compte Rendus de l'Academie des Sciences*. Paris, France, 2008, vol. 346, pp. 589–592, ser. I.

[12] T. Cai, G. Xu, and J. Zhang, "On recovery of sparse signals via ℓ_1 minimization," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3388–3397, Jul. 2009.

[13] D. Needell and R. Vershynin, "Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit," *Found. Comput. Math.*, vol. 9, no. 3, pp. 317–334, 2009.

[14] D. Needell and R. Vershynin, "Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 310–316, Apr. 2010.

[15] D. Needell and J. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmon. Anal.*, vol. 26, no. 3, pp. 291–432, 2008.

[16] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2230–2249, May 2009.

[17] A. Cohen, W. Dahmen, and R. DeVore, "Instance optimal decoding by thresholding in compressed sensing," in *Proc. 8th Int. Int. Conf., Jun. 16–20, 2010, El Escorial, Madrid, Spain*, vol. 505, p. 1, 2010, Amer. Mathematical Society.

[18] T. Blumensath and M. Davies, "Iterative hard thresholding for compressive sensing," *Appl. Comput. Harmon. Anal.*, vol. 27, no. 3, pp. 265–274, 2009.

[19] R. Chartrand and V. Staneva, "Restricted isometry properties and nonconvex compressive sensing," *Inv. Probl.*, vol. 24, no. 3, pp. 1–14, 2008.

[20] T. Cai, L. Wang, and G. Xu, *New Bounds for Restricted Isometry Constants* [Online]. Available: http://arxiv.org/abs/0911.1564

[21] H. Rauhut, "On the impossibility of uniform sparse reconstruction using greedy methods," *Sampl. Theory Signal Image Process.*, vol. 7, no. 2, pp. 197–215, 2008.

[22] P. Bechler and P. Wojtaszczyk, *Error Estimates for Orthogonal Matching Pursuit and Random Dictionaries*, preprint.

[23] M. Davenport, P. Boufounos, M. Wakin, and R. Baraniuk, "Signal processing with compressive measurements," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 445–460, Apr. 2010.

[24] M. Davenport, P. Boufounos, and R. Baraniuk, "Compressive domain interference cancellation," presented at the Workshop on Signal Processing with Adaptive Sparse Structured Representations (SPARS), Saint-Malo, France, Apr. 2009.

Mark A. Davenport (S'04–M'10) received the B.S.E.E., M.S., and Ph.D. degrees in electrical and computer engineering in 2004, 2007, and 2010, and the B.A. degree in managerial studies in 2004, all from Rice University, Houston, TX.

He is currently an NSF Mathematical Sciences Postdoctoral Research Fellow in the Department of Statistics at Stanford University, Stanford, CA. His research interests include compressive sensing, nonlinear approximation, and the application of low-dimensional signal models to a variety of problems in signal processing and machine learning. He is also cofounder and an editor of *Rejecta Mathematica*.

Dr. Davenport shared the Hershel M. Rich Invention Award from Rice in 2007 for his work on the single-pixel camera and compressive sensing.

Michael B. Wakin (S'01–M'07) received the B.S. degree in electrical engineering and the B.A. degree in mathematics in 2000 (*summa cum laude*), the M.S. degree in electrical engineering in 2002, and the Ph.D. degree in electrical engineering in 2007 from Rice University, Houston, TX.

He was an NSF Mathematical Sciences Postdoctoral Research Fellow at the California Institute of Technology, Pasadena, from 2006 to 2007 and an Assistant Professor at the University of Michigan, Ann Arbor, from 2007 to 2008. He is now an Assistant Professor in the Division of Engineering at the Colorado School of Mines, Golden. His research interests include sparse, geometric, and manifold-based models for signal and image processing, approximation, compression, compressive sensing, and dimensionality reduction.

Dr. Wakin shared the Hershel M. Rich Invention Award from Rice University in 2007 for the design of a single-pixel camera based on compressive sensing, and in 2008, he received the DARPA Young Faculty Award for his research in compressive multisignal processing for environments such as sensor and camera networks.