

Stability and Robustness of Weak Orthogonal Matching Pursuits*

Simon Foucart[†], Drexel University

Abstract

A recent result establishing, under restricted isometry conditions, the success of sparse recovery via Orthogonal Matching Pursuit using a number of iterations proportional to the sparsity level is extended to Weak Orthogonal Matching Pursuits. The new result also applies to a Pure Orthogonal Matching Pursuit, where the index appended to the support at each iteration is chosen to maximize the subsequent decrease of the squared-norm of the residual vector.

1 Introduction and Main Results

This note considers the basic Compressive Sensing problem, which aims at reconstructing s -sparse vectors $\mathbf{x} \in \mathbb{C}^N$ (i.e., vectors with at most s nonzero entries) from the knowledge of measurement vectors $\mathbf{y} = A\mathbf{x} \in \mathbb{C}^m$ with $m \ll N$. By now, it is well established — see [1, 2] and references therein — that this task can be carried out efficiently using the realization of a random matrix as a measurement matrix $A \in \mathbb{C}^{m \times N}$ and by using ℓ_1 -minimization as a reconstruction map $\Delta : \mathbb{C}^m \rightarrow \mathbb{C}^N$. There exist alternatives to the ℓ_1 -minimization, though, including as a precursor the Orthogonal Matching Pursuit introduced by Mallat and Zhang in [7] in the context of sparse approximation. The basic idea of the algorithm is to construct a target support by adding one index at a time, and to find the vector with this target support that best fits the measurement. It had become usual, when trying to reconstruct s -sparse vectors, to analyze the convergence of this algorithm after s iterations, since the s th iterate is itself an s -sparse vector. For instance, the exact reconstruction of all s -sparse vectors $\mathbf{x} \in \mathbb{C}^N$ from $\mathbf{y} = A\mathbf{x} \in \mathbb{C}^m$ via s iterations of Orthogonal Matching Pursuit can be established under a condition on the coherence of the matrix A , or even on its cumulative coherence, see [10] for details. Lately, there has also been some work [3, 5, 6, 8] establishing sparse recovery under some (stronger than desired) conditions on the restricted isometry constants of the measurement matrix. We recall briefly that the s th restricted isometry constant δ_s of A is

*This work was triggered by the participation of the author to the NSF supported Workshop in Linear Analysis and Probability at Texas A&M University

[†]The author is supported by the NSF grant DMS-1120622

defined as the smallest constant $\delta \geq 0$ such that

$$(1 - \delta)\|\mathbf{z}\|_2^2 \leq \|A\mathbf{z}\|_2^2 \leq (1 + \delta)\|\mathbf{z}\|_2^2 \quad \text{for all } s\text{-sparse } \mathbf{z} \in \mathbb{C}^N.$$

However, it was observed by Rauhut [9] and later by Mo and Shen [8] that s iterations of Orthogonal Matching Pursuit are not enough to guarantee s -sparse recovery under (more natural) restricted isometry conditions. Nonetheless, it was recently proved by Zhang [11] that $30s$ iterations are enough to guarantee s -sparse recovery provided $\delta_{31s} < 1/3$. Zhang's result also covers the important case of vectors $\mathbf{x} \in \mathbb{C}^N$ which are not exactly sparse and which are measured with some error via $\mathbf{y} = A\mathbf{x} + \mathbf{e} \in \mathbb{C}^m$ — in fact, it covers the case of measurement maps $A : \mathbb{C}^N \rightarrow \mathbb{C}^m$ that are not necessarily linear, too. This note imitates the original arguments of Zhang and extends the results to a wider class of algorithms. Namely, we consider algorithms of the following type:

Generic Orthogonal Matching Pursuit

Input: measurement matrix A , measurement vector \mathbf{y} , index set S^0 .

Initialization: $\mathbf{x}^0 = \operatorname{argmin} \{\|\mathbf{y} - A\mathbf{z}\|_2, \operatorname{supp}(\mathbf{z}) \subseteq S^0\}$.

Iteration: repeat the following steps until a stopping criterion is met at $n = \bar{n}$

$$\text{(GOMP}_1) \quad S^{n+1} = S^n \cup \{j^{n+1}\},$$

$$\text{(GOMP}_2) \quad \mathbf{x}^{n+1} = \operatorname{argmin} \{\|\mathbf{y} - A\mathbf{z}\|_2, \operatorname{supp}(\mathbf{z}) \subseteq S^{n+1}\}.$$

Output: \bar{n} -sparse vector $\mathbf{x}^* = \mathbf{x}^{\bar{n}}$.

The recipe for choosing the index j^{n+1} distinguishes between different algorithms, which are customarily initiated with $S^0 = \emptyset$ and $\mathbf{x}^0 = 0$. The classical Orthogonal Matching Pursuit algorithm corresponds to a choice where

$$(1) \quad |(A^*(\mathbf{y} - A\mathbf{x}^n))_{j^{n+1}}| = \max_{1 \leq j \leq N} |(A^*(\mathbf{y} - A\mathbf{x}^n))_j|.$$

The Weak Orthogonal Matching Pursuit algorithm with parameter $0 < \rho \leq 1$ corresponds to a choice where

$$(2) \quad |(A^*(\mathbf{y} - A\mathbf{x}^n))_{j^{n+1}}| \geq \rho \max_{1 \leq j \leq N} |(A^*(\mathbf{y} - A\mathbf{x}^n))_j|.$$

We also single out a choice of $j^{n+1} \notin S^n$ where

$$(3) \quad \frac{|(A^*(\mathbf{y} - A\mathbf{x}^n))_{j^{n+1}}|}{\operatorname{dist}(\mathbf{a}_{j^{n+1}}, \operatorname{span}\{\mathbf{a}_i, i \in S^n\})} = \max_{j \notin S^n} \frac{|(A^*(\mathbf{y} - A\mathbf{x}^n))_j|}{\operatorname{dist}(\mathbf{a}_j, \operatorname{span}\{\mathbf{a}_i, i \in S^n\})},$$

where $\mathbf{a}_1, \dots, \mathbf{a}_N \in \mathbb{C}^m$ denote the columns of A . We call this (rather unpractical) algorithm Pure Orthogonal Matching Pursuit, because it abides by a pure greedy strategy of reducing as much as possible the squared ℓ_2 -norm of the residual at each iteration. Indeed, the following observation can be made when $\mathbf{a}_1, \dots, \mathbf{a}_N$ all have norm one.

Theorem 1. For any Generic Orthogonal Matching Pursuit algorithm applied with a matrix $A \in \mathbb{C}^{m \times N}$ whose columns $\mathbf{a}_1, \dots, \mathbf{a}_N$ are ℓ_2 -normalized, the squared ℓ_2 -norm of the residual decreases at each iteration according to

$$\|\mathbf{y} - A\mathbf{x}^{n+1}\|_2^2 = \|\mathbf{y} - A\mathbf{x}^n\|_2^2 - \Delta_n,$$

where the quantity Δ_n satisfies

$$\begin{aligned} \Delta_n &= \|A(\mathbf{x}^{n+1} - \mathbf{x}^n)\|_2^2 = x_{j^{n+1}}^{n+1} \overline{(A^*(\mathbf{y} - A\mathbf{x}^n))_{j^{n+1}}} = \frac{|(A^*(\mathbf{y} - A\mathbf{x}^n))_{j^{n+1}}|^2}{\text{dist}(\mathbf{a}_{j^{n+1}}, \text{span}\{\mathbf{a}_i, i \in S^n\})^2} \\ &\geq |(A^*(\mathbf{y} - A\mathbf{x}^n))_{j^{n+1}}|^2. \end{aligned}$$

This result reveals that the quantity Δ_n is computable at iteration n . This fact is apparent from its third expression but it is hidden in its first two expressions. Maximizing Δ_n leads to the Pure Orthogonal Matching Pursuit, as hinted above. We informally remark at this point that Δ_n almost equals its lower bound when A has small restricted isometry constants, i.e., when its columns are almost orthonormal. Hence, under the restricted isometry property, the Pure Orthogonal Matching Pursuit algorithm almost coincides with the classical Orthogonal Matching Pursuit algorithm. This is made precise below.

Theorem 2. Given an integer $\bar{n} \geq 1$ for which $\delta_{\bar{n}} < (\sqrt{5} - 1)/2$, the Pure Orthogonal Matching Pursuit algorithm iterated at most \bar{n} times is a Weak Orthogonal Matching Pursuit algorithm with parameter $\rho := \sqrt{1 - \delta_{\bar{n}}^2}/(1 - \delta_{\bar{n}})$.

The proofs of Theorems 1 and 2 are postponed until Section 3. We now state the main result of this note, which reduces to Zhang's result for the classical Orthogonal Matching Pursuit algorithm when $\rho = 1$ (with slightly improved constants). Below, the notation $\mathbf{x}_{\bar{S}}$ stands for the vector equal to \mathbf{x} on the complementary set \bar{S} of S and to zero on the set S , and the notation $\sigma_s(\mathbf{x})_1$ stands for the ℓ_1 -error of best s -term approximation to \mathbf{x} .

Theorem 3. Let $A \in \mathbb{C}^{m \times N}$ be a matrix with ℓ_2 -normalized columns. For all $\mathbf{x} \in \mathbb{C}^N$ and all $\mathbf{e} \in \mathbb{C}^m$, let (\mathbf{x}^n) be the sequence produced from $\mathbf{y} = A\mathbf{x} + \mathbf{e}$ and $S^0 = \emptyset$ by the Weak Orthogonal Matching Pursuit algorithm with parameter $0 < \rho \leq 1$. If $\delta_{(1+3r)s} \leq 1/6$, $r := \lceil 3/\rho^2 \rceil$, then there is a constant $C > 0$ depending only on $\delta_{(1+3r)s}$ and on ρ such that, for any $S \subseteq \{1, \dots, N\}$ with $\text{card}(S) = s$,

$$(4) \quad \|\mathbf{y} - A\mathbf{x}^{2rs}\|_2 \leq C\|A\mathbf{x}_{\bar{S}} + \mathbf{e}\|_2.$$

Furthermore, if $\delta_{(2+6r)s} \leq 1/6$, then, for all $1 \leq p \leq 2$,

$$(5) \quad \|\mathbf{x} - \mathbf{x}^{4rs}\|_p \leq \frac{C}{s^{1-1/p}} \sigma_s(\mathbf{x})_1 + Ds^{1/p-1/2} \|\mathbf{e}\|_2$$

for some constants $C, D > 0$ depending only on $\delta_{(2+6r)s}$ and on ρ .

2 Sparse Recovery via Weak Orthogonal Matching Pursuits

This section is dedicated to the proof of Theorem 3. We stress once again the similitude of the arguments presented here with the ones given by Zhang in [11]. Throughout this section and the next one, we will make use of the simple observation that

$$(6) \quad (A^*(\mathbf{y} - A\mathbf{x}^k))_{S^k} = 0 \quad \text{for all } k \geq 1.$$

Indeed, with $n := k - 1 \geq 0$, the definition of \mathbf{x}^{n+1} via (GOMP₂) implies that the residual $\mathbf{y} - A\mathbf{x}^{n+1}$ is orthogonal to the space $\{A\mathbf{z}, \text{supp}(\mathbf{z}) \subseteq S^{n+1}\}$. This means that, for all $\mathbf{z} \in \mathbb{C}^N$ supported on S^{n+1} , we have $0 = \langle \mathbf{y} - A\mathbf{x}^{n+1}, A\mathbf{z} \rangle = \langle A^*(\mathbf{y} - A\mathbf{x}^{n+1}), \mathbf{z} \rangle$, hence the claim. It will also be convenient to isolate the following statement from the flow of the argument.

Lemma 4. Let $A \in \mathbb{C}^{m \times N}$ be a matrix with ℓ_2 -normalized columns. For $\mathbf{x} \in \mathbb{C}^N$ and $\mathbf{e} \in \mathbb{C}^m$, let (\mathbf{x}^n) be the sequence produced from $\mathbf{y} = A\mathbf{x} + \mathbf{e}$ by the Weak Orthogonal Matching Pursuit algorithm with parameter $0 < \rho \leq 1$. For any $n \geq 0$, any index set U not included in S^n , and any vector $\mathbf{u} \in \mathbb{C}^N$ supported on U ,

$$(7) \quad \begin{aligned} \|\mathbf{y} - A\mathbf{x}^{n+1}\|_2^2 &\leq \|\mathbf{y} - A\mathbf{x}^n\|_2^2 - \rho^2 \frac{\|A(\mathbf{u} - \mathbf{x}^n)\|_2^2}{\|\mathbf{u}_{S^n}\|_1^2} \max\{0, \|\mathbf{y} - A\mathbf{x}^n\|_2^2 - \|\mathbf{y} - A\mathbf{u}\|_2^2\} \\ &\leq \|\mathbf{y} - A\mathbf{x}^n\|_2^2 - \frac{\rho^2(1-\delta)}{\text{card}(U \setminus S^n)} \max\{0, \|\mathbf{y} - A\mathbf{x}^n\|_2^2 - \|\mathbf{y} - A\mathbf{u}\|_2^2\}, \end{aligned}$$

where $\delta := \delta_{\text{card}(U \cup S^n)}$.

Proof. The second inequality follows from the first one by noticing that

$$\begin{aligned} \|A(\mathbf{u} - \mathbf{x}^n)\|_2^2 &\geq (1-\delta)\|\mathbf{u} - \mathbf{x}^n\|_2^2 \geq (1-\delta)\|(\mathbf{u} - \mathbf{x}^n)_{\overline{S^n}}\|_2^2, \\ \|\mathbf{u}_{\overline{S^n}}\|_1^2 &\leq \text{card}(U \setminus S^n)\|(\mathbf{u} - \mathbf{x}^n)_{\overline{S^n}}\|_2^2 = \text{card}(U \setminus S^n)\|(\mathbf{u} - \mathbf{x}^n)_{\overline{S^n}}\|_2^2. \end{aligned}$$

Now, according to Theorem 1, it is enough to prove that

$$(8) \quad |(A^*(\mathbf{y} - A\mathbf{x}^n))_{j^{n+1}}|^2 \geq \rho^2 \frac{\|A(\mathbf{u} - \mathbf{x}^n)\|_2^2}{\|\mathbf{u}_{\overline{S^n}}\|_1^2} \{\|\mathbf{y} - A\mathbf{x}^n\|_2^2 - \|\mathbf{y} - A\mathbf{u}\|_2^2\}$$

when $\|\mathbf{y} - A\mathbf{x}^n\|_2^2 \geq \|\mathbf{y} - A\mathbf{u}\|_2^2$. Making use of (6), we observe on the one hand that

$$(9) \quad \begin{aligned} \Re\langle A(\mathbf{u} - \mathbf{x}^n), \mathbf{y} - A\mathbf{x}^n \rangle &= \Re\langle \mathbf{u} - \mathbf{x}^n, A^*(\mathbf{y} - A\mathbf{x}^n) \rangle = \Re\langle (\mathbf{u} - \mathbf{x}^n)_{\overline{S^n}}, (A^*(\mathbf{y} - A\mathbf{x}^n))_{\overline{S^n}} \rangle \\ &\leq \|(\mathbf{u} - \mathbf{x}^n)_{\overline{S^n}}\|_1 \| (A^*(\mathbf{y} - A\mathbf{x}^n))_{\overline{S^n}} \|_\infty \leq \|\mathbf{u}_{\overline{S^n}}\|_1 | (A^*(\mathbf{y} - A\mathbf{x}^n))_{j^{n+1}} | / \rho. \end{aligned}$$

We observe on the other hand that

$$(10) \quad \begin{aligned} 2\Re\langle A(\mathbf{u} - \mathbf{x}^n), \mathbf{y} - A\mathbf{x}^n \rangle &= \|A(\mathbf{u} - \mathbf{x}^n)\|_2^2 + \|\mathbf{y} - A\mathbf{x}^n\|_2^2 - \|A(\mathbf{u} - \mathbf{x}^n) - (\mathbf{y} - A\mathbf{x}^n)\|_2^2 \\ &= \|A(\mathbf{u} - \mathbf{x}^n)\|_2^2 + \{\|\mathbf{y} - A\mathbf{x}^n\|_2^2 - \|\mathbf{y} - A\mathbf{u}\|_2^2\} \\ &\geq 2\|A(\mathbf{u} - \mathbf{x}^n)\|_2 \sqrt{\|\mathbf{y} - A\mathbf{x}^n\|_2^2 - \|\mathbf{y} - A\mathbf{u}\|_2^2}. \end{aligned}$$

Combining the squared versions of (9) and (10), we arrive at

$$\|A(\mathbf{u} - \mathbf{x}^n)\|_2^2 \{ \|\mathbf{y} - A\mathbf{x}^n\|_2^2 - \|\mathbf{y} - A\mathbf{u}\|_2^2 \} \leq \|\mathbf{u}_{\overline{S^n}}\|_1^2 |(A^*(\mathbf{y} - A\mathbf{x}^n))_{j^{n+1}}|^2 / \rho^2.$$

Rearranging the terms leads to the desired inequality (8). \square

Theorem 3 would actually holds for any sequence obeying the conclusion of Lemma 4, by virtue of the following proposition.

Proposition 5. Let $A \in \mathbb{C}^{m \times N}$ be a matrix with ℓ_2 -normalized columns. For an s -sparse vector $\mathbf{x} \in \mathbb{C}^N$ and for $\mathbf{e} \in \mathbb{C}^m$, let $\mathbf{y} = A\mathbf{x} + \mathbf{e}$ and let (\mathbf{x}^n) be a sequence satisfying (7). If $\delta_{(1+3\lceil 3/\rho^2 \rceil)_s} \leq 1/6$, then

$$\|\mathbf{y} - A\mathbf{x}^{\bar{n}}\|_2 \leq C\|\mathbf{e}\|_2, \quad \bar{n} := 2\lceil 3/\rho^2 \rceil \text{card}(S \setminus S^0),$$

where $C > 0$ is a constant depending only on $\delta_{(1+3\lceil 3/\rho^2 \rceil)_s}$ and on ρ .

Proof. The proof proceeds by induction on $\text{card}(S \setminus S^0)$, where $S := \text{supp}(\mathbf{x})$. If it is zero, i.e., if $S \subseteq S^0$, then the definition of \mathbf{x}^0 implies

$$\|\mathbf{y} - A\mathbf{x}^0\|_2 \leq \|\mathbf{y} - A\mathbf{x}\|_2 = \|\mathbf{e}\|_2,$$

and the result holds with $C = 1$. Let us now assume that the result holds up to an integer $s' - 1$, $s' \geq 1$, and let us show that it holds when $\text{card}(S \setminus S^0) = s'$. We consider subsets of $S \setminus S^0$ defined by

$$U^0 = \emptyset \text{ and } U^\ell = \{\text{indices of } 2^{\ell-1} \text{ largest entries of } \mathbf{x}_{\overline{S^0}} \text{ in modulus}\} \text{ for } \ell \geq 1,$$

to which we associate the vectors

$$\tilde{\mathbf{x}}^\ell := \mathbf{x}_{\overline{S^0 \cup U^\ell}}, \quad \ell \geq 0.$$

Note that the last U^ℓ , namely $U^{\lceil \log_2(s') \rceil + 1}$, is taken to be the whole set $S \setminus S^0$ (it may have less than $2^{\ell-1}$ elements), in which case $0 = \|\tilde{\mathbf{x}}^\ell\|_2^2 \leq \|\tilde{\mathbf{x}}^{\ell-1}\|_2^2 / \mu$ for a constant μ to be chosen later. We may then consider the smallest integer $1 \leq L \leq \lceil \log_2(s') \rceil + 1$ such that

$$\|\tilde{\mathbf{x}}^{L-1}\|_2^2 \geq \mu \|\tilde{\mathbf{x}}^L\|_2^2.$$

Its definition implies the (possibly empty) list of inequalities

$$\|\tilde{\mathbf{x}}^0\|_2^2 < \mu \|\tilde{\mathbf{x}}^1\|_2^2, \dots, \|\tilde{\mathbf{x}}^{L-2}\|_2^2 < \mu \|\tilde{\mathbf{x}}^{L-1}\|_2^2.$$

For each $\ell \in [L]$, we apply inequality (7) to the vector $\mathbf{u} = \mathbf{x} - \tilde{\mathbf{x}}^\ell$ supported on $S^0 \cup U^\ell$ while noticing that $(S^0 \cup U^\ell) \cup S^n \subseteq S \cup S^n$ and $(S^0 \cup U^\ell) \setminus S^n \subseteq (S^0 \cup U^\ell) \setminus S^0 = U^\ell$, and we subtract

$\|\mathbf{y} - A\mathbf{u}\|_2^2 = \|A\tilde{\mathbf{x}}^\ell + \mathbf{e}\|_2^2$ from both sides to obtain

$$\begin{aligned} & \max\{0, \|\mathbf{y} - A\mathbf{x}^{n+1}\|_2^2 - \|A\tilde{\mathbf{x}}^\ell + \mathbf{e}\|_2^2\} \\ & \leq \left(1 - \frac{\rho^2(1 - \delta_{s+n})}{\text{card}(U^\ell)}\right) \max\{0, \|\mathbf{y} - A\mathbf{x}^n\|_2^2 - \|A\tilde{\mathbf{x}}^\ell + \mathbf{e}\|_2^2\} \\ & \leq \exp\left(-\frac{\rho^2(1 - \delta_{s+n})}{\text{card}(U^\ell)}\right) \max\{0, \|\mathbf{y} - A\mathbf{x}^n\|_2^2 - \|A\tilde{\mathbf{x}}^\ell + \mathbf{e}\|_2^2\}. \end{aligned}$$

For any $K \geq 0$ and any $n, k \geq 0$ satisfying $n + k \leq K$, we derive by immediate induction that

$$\begin{aligned} & \max\{0, \|\mathbf{y} - A\mathbf{x}^{n+k}\|_2^2 - \|A\tilde{\mathbf{x}}^\ell + \mathbf{e}\|_2^2\} \\ & \leq \exp\left(-\frac{k\rho^2(1 - \delta_{s+K})}{\text{card}(U^\ell)}\right) \max\{0, \|\mathbf{y} - A\mathbf{x}^n\|_2^2 - \|A\tilde{\mathbf{x}}^\ell + \mathbf{e}\|_2^2\}. \end{aligned}$$

By separating cases in the rightmost maximum, we easily deduce that

$$\|\mathbf{y} - A\mathbf{x}^{n+k}\|_2^2 \leq \exp\left(-\frac{k\rho^2(1 - \delta_{s+K})}{\text{card}(U^\ell)}\right) \|\mathbf{y} - A\mathbf{x}^n\|_2^2 + \|A\tilde{\mathbf{x}}^\ell + \mathbf{e}\|_2^2.$$

For some integer κ to be chosen later, applying this successively with

$$k_1 := \kappa \text{card}(U^1), \dots, k_L := \kappa \text{card}(U^L), \quad \text{and} \quad K := k_1 + \dots + k_L,$$

yields, with $\nu := \exp(\kappa\rho^2(1 - \delta_{s+K}))$,

$$\begin{aligned} \|\mathbf{y} - A\mathbf{x}^{k_1}\|_2^2 & \leq \frac{1}{\nu} \|\mathbf{y} - A\mathbf{x}^0\|_2^2 + \|A\tilde{\mathbf{x}}^1 + \mathbf{e}\|_2^2, \\ \|\mathbf{y} - A\mathbf{x}^{k_1+k_2}\|_2^2 & \leq \frac{1}{\nu} \|\mathbf{y} - A\mathbf{x}^{k_1}\|_2^2 + \|A\tilde{\mathbf{x}}^2 + \mathbf{e}\|_2^2, \\ & \vdots \\ \|\mathbf{y} - A\mathbf{x}^{k_1+\dots+k_{L-1}+k_L}\|_2^2 & \leq \frac{1}{\nu} \|\mathbf{y} - A\mathbf{x}^{k_1+\dots+k_{L-1}}\|_2^2 + \|A\tilde{\mathbf{x}}^L + \mathbf{e}\|_2^2. \end{aligned}$$

Dividing the first row by ν^{L-1} , the second row by ν^{L-2} , and so on, then summing everything, we obtain

$$\|\mathbf{y} - A\mathbf{x}^K\|_2^2 \leq \frac{\|\mathbf{y} - A\mathbf{x}^0\|_2^2}{\nu^L} + \frac{\|A\tilde{\mathbf{x}}^1 + \mathbf{e}\|_2^2}{\nu^{L-1}} + \dots + \frac{\|A\tilde{\mathbf{x}}^{L-1} + \mathbf{e}\|_2^2}{\nu} + \|A\tilde{\mathbf{x}}^L + \mathbf{e}\|_2^2.$$

Taking into account that $\mathbf{x} - \tilde{\mathbf{x}}^0$ is supported on $S^0 \cup U^0 = S^0$, the definition of \mathbf{x}^0 implies that $\|\mathbf{y} - A\mathbf{x}^0\|_2^2 \leq \|\mathbf{y} - A(\mathbf{x} - \tilde{\mathbf{x}}^0)\|_2^2 = \|A\tilde{\mathbf{x}}^0 + \mathbf{e}\|_2^2$, hence

$$\|\mathbf{y} - A\mathbf{x}^K\|_2^2 \leq \sum_{\ell=0}^L \frac{\|A\tilde{\mathbf{x}}^\ell + \mathbf{e}\|_2^2}{\nu^{L-\ell}} \leq \sum_{\ell=0}^L \frac{2(\|A\tilde{\mathbf{x}}^\ell\|_2^2 + \|\mathbf{e}\|_2^2)}{\nu^{L-\ell}}.$$

Let us remark that, for $\ell \leq L - 1$ and also for $\ell = L$,

$$\|A\tilde{\mathbf{x}}^\ell\|_2^2 \leq (1 + \delta_s) \|\tilde{\mathbf{x}}^\ell\|_2^2 \leq (1 + \delta_s) \mu^{L-1-\ell} \|\tilde{\mathbf{x}}^{L-1}\|_2^2.$$

As a result, we have

$$\begin{aligned}\|\mathbf{y} - A\mathbf{x}^K\|_2^2 &\leq \frac{2(1 + \delta_s)\|\tilde{\mathbf{x}}^{L-1}\|_2^2}{\mu} \sum_{\ell=0}^L \left(\frac{\mu}{\nu}\right)^{L-\ell} + 2\|\mathbf{e}\|_2^2 \sum_{\ell=0}^L \frac{1}{\nu^{L-\ell}} \\ &\leq \frac{2(1 + \delta_s)\|\tilde{\mathbf{x}}^{L-1}\|_2^2}{\mu(1 - \mu/\nu)} + \frac{2\|\mathbf{e}\|_2^2}{1 - \nu}.\end{aligned}$$

We choose $\mu = \nu/2$ so that $\mu(1 - \mu/\nu)$ takes its maximal value $\nu/4$. It follows that, with $\alpha := \sqrt{8(1 + \delta_s)/\nu}$ and $\beta := \sqrt{2/(1 - \nu)}$,

$$(11) \quad \|\mathbf{y} - A\mathbf{x}^K\|_2 \leq \alpha \|\tilde{\mathbf{x}}^{L-1}\|_2 + \beta \|\mathbf{e}\|_2.$$

On the other hand, with $\gamma := \sqrt{1 - \delta_{s+K}}$, we have

$$\begin{aligned}\|\mathbf{y} - A\mathbf{x}^K\|_2 &= \|A(\mathbf{x} - \mathbf{x}^K) + \mathbf{e}\|_2 \geq \|A(\mathbf{x} - \mathbf{x}^K)\|_2 - \|\mathbf{e}\|_2 \\ &\geq \gamma \|\mathbf{x} - \mathbf{x}^K\|_2 - \|\mathbf{e}\|_2 \geq \gamma \|\mathbf{x}_{\overline{S^K}}\|_2 - \|\mathbf{e}\|_2.\end{aligned}$$

We deduce that

$$(12) \quad \|\mathbf{x}_{\overline{S^K}}\|_2 \leq \frac{\alpha}{\gamma} \|\tilde{\mathbf{x}}^{L-1}\|_2 + \frac{\beta + 1}{\gamma} \|\mathbf{e}\|_2.$$

Let us now choose $\kappa = \lceil 3/\rho^2 \rceil$, which guarantees that

$$(13) \quad \frac{\alpha}{\gamma} = \sqrt{\frac{8(1 + \delta_s)}{(1 - \delta_{s+K}) \exp(\kappa\rho^2(1 - \delta_{s+K}))}} < 1, \quad \text{since } \delta_{s+K} \leq \delta_{(1+3\lceil 3/\rho^2 \rceil)s} \leq 1/6,$$

where we have used the fact that

$$K = \kappa(1 + \dots + 2^{L-2} + \text{card}(U^L)) < \kappa(2^{L-1} + s') \leq 3\kappa s' \leq 3\lceil 3/\rho^2 \rceil s.$$

Thus, in the case $((\beta + 1)/\gamma)\|\mathbf{e}\|_2 < (1 - \alpha/\gamma)\|\tilde{\mathbf{x}}^{L-1}\|_2$, we derive from (12) that

$$\|\mathbf{x}_{\overline{S^K}}\|_2 < \|\tilde{\mathbf{x}}^{L-1}\|_2, \quad \text{i.e.,} \quad \|(\mathbf{x}_{\overline{S^0}})_{S \setminus S^K}\|_2 < \|(\mathbf{x}_{\overline{S^0}})_{(S \setminus S^0) \setminus U^{L-1}}\|_2.$$

But according to the definition of U^{L-1} , this yields

$$\text{card}(S \setminus S^K) < \text{card}((S \setminus S^0) \setminus U^{L-1}) = s' - 2^{L-1}.$$

Continuing the algorithm from iteration K now amounts to starting it from iteration 0 with \mathbf{x}^0 replaced by \mathbf{x}^K , therefore the induction hypothesis implies that

$$\|\mathbf{y} - A\mathbf{x}^{K+\bar{n}}\|_2 \leq C\|\mathbf{e}\|_2, \quad \bar{n} := 2\lceil 3/\rho^2 \rceil (s' - 2^{L-1}).$$

Thus, since we also have the bound $K \leq \kappa(1 + \dots + 2^{L-2} + 2^{L-1}) < \lceil 3/\rho^2 \rceil \cdot 2^L$, the number of required iterations satisfies $K + \bar{n} \leq 2\lceil 3/\rho^2 \rceil s'$, as expected. In the alternative case where $((\beta + 1)/\gamma)\|\mathbf{e}\|_2 \geq (1 - \alpha/\gamma)\|\tilde{\mathbf{x}}^{L-1}\|_2$, the situation is easier, since (11) yields

$$\|\mathbf{y} - A\mathbf{x}^K\|_2 \leq \frac{\alpha(\beta + 1)}{\gamma - \alpha} \|\mathbf{e}\|_2 + \beta \|\mathbf{e}\|_2 =: C\|\mathbf{e}\|_2,$$

where the constant $C \geq 1$ depends only on $\delta_{(1+3\lceil 3/\rho^2 \rceil)s}$ and on ρ . This shows that the induction hypothesis holds when $\text{card}(S \setminus S^0) = s'$. The proof is now complete. \square

With Proposition 5 at hand, proving the main theorem is almost immediate.

Proof of Theorem 3. Given $S \subseteq [N]$ for which $\text{card}(S) = s$, we can write $\mathbf{y} = A\mathbf{x}_S + \mathbf{e}'$ where $\mathbf{e}' := A\mathbf{x}_{\bar{S}} + \mathbf{e}$. Applying Proposition 5 to \mathbf{x}_S and \mathbf{e}' with $S^0 = \emptyset$ then gives the desired inequality

$$\|\mathbf{y} - A\mathbf{x}^{2rs}\|_2 \leq C\|\mathbf{e}'\|_2 = C\|A\mathbf{x}_{\bar{S}} + \mathbf{e}\|_2$$

for some constant $C > 0$ depending only on $\delta_{(1+3r)s}$ and on ρ . Deducing the second part of the theorem from the first part is rather classical (see [4], for instance), and we therefore omit the justification of this part. \square

Remark. For the Pure Orthogonal Matching Pursuit algorithm, the parameter ρ depends on the restricted isometry constants as $\rho^2 = 1 - \delta_n^2/(1 - \delta_n)$. In this case, to guarantee that $\alpha/\gamma < 1$ in (13), we may choose $\kappa = 3$ when $\delta_{s+K} \leq \delta_n \leq 1/6$. This would yield the estimate (4) in $6s$ iterations provided $\delta_{10s} \leq 1/6$, and estimate (5) in $12s$ iterations provided $\delta_{20s} \leq 1/6$.

3 Pure OMP as a Weak OMP under RIP

This section is dedicated to the proofs of Theorems 1 and 2. The latter justifies that the Pure Orthogonal Matching Pursuit algorithm is interpreted as a Weak Pure Orthogonal Matching Pursuit algorithm with parameter depending on the restricted isometry constants.

Proof of Theorem 1. We start by noticing that the residual $\mathbf{y} - A\mathbf{x}^{n+1}$ is orthogonal to the space $\{A\mathbf{z}, \text{supp}(\mathbf{z}) \subseteq S^{n+1}\}$, and in particular to $A(\mathbf{x}^{n+1} - \mathbf{x}^n)$, so that

$$\|\mathbf{y} - A\mathbf{x}^n\|_2^2 = \|\mathbf{y} - A\mathbf{x}^{n+1} + A(\mathbf{x}^{n+1} - \mathbf{x}^n)\|_2^2 = \|\mathbf{y} - A\mathbf{x}^{n+1}\|_2^2 + \|A(\mathbf{x}^{n+1} - \mathbf{x}^n)\|_2^2.$$

This establishes the first expression for Δ_n . As for the other statements about Δ_n , we separate first the case $j^{n+1} \in S^n$. Here, the ratio $|(A^*(\mathbf{y} - A\mathbf{x}^n))_{j^{n+1}}|/\text{dist}(\mathbf{a}_{j^{n+1}}, \text{span}\{\mathbf{a}_i, i \in S^n\})$ is not given any meaning, but we have $\Delta_n = 0$ in view of $S^{n+1} = S^n$ and $\mathbf{x}^{n+1} = \mathbf{x}^n$, and the awaited equality $(A^*(\mathbf{y} - A\mathbf{x}^n))_{j^{n+1}} = 0$ follows from (6). We now place ourselves in the case $j^{n+1} \notin S^n$. For the second expression of Δ_n , we keep in mind that $\mathbf{x}^{n+1} - \mathbf{x}^n$ is supported on S^{n+1} , that $(A^*\mathbf{y})_{S^{n+1}} = (A^*A\mathbf{x}^{n+1})_{S^{n+1}}$, and that $(A^*(\mathbf{y} - A\mathbf{x}^n))_{S^n} = 0$ to write

$$\begin{aligned} \|A(\mathbf{x}^{n+1} - \mathbf{x}^n)\|_2^2 &= \langle \mathbf{x}^{n+1} - \mathbf{x}^n, A^*A(\mathbf{x}^{n+1} - \mathbf{x}^n) \rangle = \langle \mathbf{x}^{n+1} - \mathbf{x}^n, (A^*A(\mathbf{x}^{n+1} - \mathbf{x}^n))_{S^{n+1}} \rangle \\ &= \langle \mathbf{x}^{n+1} - \mathbf{x}^n, (A^*(\mathbf{y} - A\mathbf{x}^n))_{S^{n+1}} \rangle = \langle (\mathbf{x}^{n+1} - \mathbf{x}^n)_{S^{n+1} \setminus S^n}, (A^*(\mathbf{y} - A\mathbf{x}^n))_{S^{n+1} \setminus S^n} \rangle \\ &= x_{j^{n+1}}^{n+1} \overline{(A^*(\mathbf{y} - A\mathbf{x}^n))_{j^{n+1}}}. \end{aligned}$$

For the third expression, we first notice that the step (GOMP₂) is equivalent to

$$\mathbf{x}^{n+1} = A_{S^{n+1}}^\dagger \mathbf{y}, \quad A_{S^{n+1}}^\dagger := (A_{S^{n+1}}^* A_{S^{n+1}})^{-1} A_{S^{n+1}}^*.$$

We decompose $A_{S^{n+1}}$ as $A_{S^{n+1}} = \begin{bmatrix} A_{S^n} & | & \mathbf{a}_{j^{n+1}} \end{bmatrix}$. We then derive that

$$A_{S^{n+1}}^* \mathbf{y} = \begin{bmatrix} A_{S^n}^* \mathbf{y} \\ \mathbf{a}_{j^{n+1}}^* \mathbf{y} \end{bmatrix}, \quad A_{S^{n+1}}^* A_{S^{n+1}} = \begin{bmatrix} A_{S^n}^* A_{S^n} & | & A_{S^n}^* \mathbf{a}_{j^{n+1}} \\ \mathbf{a}_{j^{n+1}}^* A_{S^n} & | & 1 \end{bmatrix}.$$

Among the several ways to express the inverse of a block-matrix, we select

$$(A_{S^{n+1}}^* A_{S^{n+1}})^{-1} = \begin{bmatrix} M_1 & | & M_2 \\ M_3 & | & M_4 \end{bmatrix},$$

where M_3 and M_4 (note that M_4 is a scalar in this case) are given by

$$\begin{aligned} M_4 &= (1 - \mathbf{a}_{j^{n+1}}^* A_{S^n} (A_{S^n}^* A_{S^n})^{-1} A_{S^n}^* \mathbf{a}_{j^{n+1}})^{-1} = (1 - \mathbf{a}_{j^{n+1}}^* A_{S^n} A_{S^n}^\dagger \mathbf{a}_{j^{n+1}})^{-1}, \\ M_3 &= -M_4 \mathbf{a}_{j^{n+1}}^* A_{S^n} (A_{S^n}^* A_{S^n})^{-1}. \end{aligned}$$

It follows from the block-decomposition of $\mathbf{x}^{n+1} = (A_{S^{n+1}}^* A_{S^{n+1}})^{-1} A_{S^{n+1}}^* \mathbf{y}$ that

$$x_j^{n+1} = M_3 A_{S^n}^* \mathbf{y} + M_4 \mathbf{a}_{j^{n+1}}^* \mathbf{y} = M_4 \mathbf{a}_{j^{n+1}}^* (-A_{S^n} A_{S^n}^\dagger \mathbf{y} + \mathbf{y}) = M_4 \mathbf{a}_{j^{n+1}}^* (\mathbf{y} - A \mathbf{x}^n).$$

We note that $\mathbf{a}_{j^{n+1}}^* (\mathbf{y} - A \mathbf{x}^n)$ is simply $(A^* (\mathbf{y} - A \mathbf{x}^n))_{j^{n+1}}$. As for M_4 , we note that $A_{S^n} A_{S^n}^\dagger \mathbf{a}_{j^{n+1}}$ is the orthogonal projection $P_n \mathbf{a}_{j^{n+1}}$ of $\mathbf{a}_{j^{n+1}}$ onto the space $\{A \mathbf{z}, \text{supp}(\mathbf{z}) \subseteq S^n\} = \text{span}\{\mathbf{a}_i, i \in S^n\}$, so that

$$M_4 = (1 - \langle P_n \mathbf{a}_{j^{n+1}}, \mathbf{a}_{j^{n+1}} \rangle)^{-1} = (1 - \|P_n \mathbf{a}_{j^{n+1}}\|_2^2)^{-1} = \text{dist}(\mathbf{a}_{j^{n+1}}, \text{span}\{\mathbf{a}_i, i \in S^n\})^{-2}.$$

Note that $\|P_n \mathbf{a}_{j^{n+1}}\|_2 < \|\mathbf{a}_{j^{n+1}}\|_2 = 1$ justifies the existence of M_4 and that $\|P_n \mathbf{a}_{j^{n+1}}\|_2 \geq 0$ shows that $M_4 \geq 1$, which establishes the lower bound on Δ_n . \square

Proof of Theorem 2. According to the previous arguments and to the definition of the Pure Orthogonal Matching Pursuit algorithm, for any $j \notin S^n$, we have

$$\begin{aligned} \frac{|(A^* (\mathbf{y} - A \mathbf{x}^n))_{j^{n+1}}|^2}{1 - \|P_n \mathbf{a}_{j^{n+1}}\|_2^2} &= \frac{|(A^* (\mathbf{y} - A \mathbf{x}^n))_{j^{n+1}}|^2}{\text{dist}(\mathbf{a}_{j^{n+1}}, \text{span}\{\mathbf{a}_i, i \in S^n\})^2} \geq \frac{|(A^* (\mathbf{y} - A \mathbf{x}^n))_j|^2}{\text{dist}(\mathbf{a}_j, \text{span}\{\mathbf{a}_i, i \in S^n\})^2} \\ &\geq |(A^* (\mathbf{y} - A \mathbf{x}^n))_j|^2. \end{aligned}$$

But, in view of $|\langle A \mathbf{u}, A \mathbf{v} \rangle| \leq \delta_k \|\mathbf{u}\|_2 \|\mathbf{v}\|_2$ for all disjointly supported $\mathbf{u}, \mathbf{v} \in \mathbb{C}^N$ satisfying $\text{card}(\text{supp}(\mathbf{u}) \cup \text{supp}(\mathbf{v})) \leq k$, we have

$$\begin{aligned} \|P_n \mathbf{a}_{j^{n+1}}\|_2^2 &= \langle P_n \mathbf{a}_{j^{n+1}}, \mathbf{a}_{j^{n+1}} \rangle = \langle A_{S^n} A_{S^n}^\dagger \mathbf{a}_{j^{n+1}}, \mathbf{a}_{j^{n+1}} \rangle = \langle A A_{S^n}^\dagger \mathbf{a}_{j^{n+1}}, A \mathbf{e}_{j^{n+1}} \rangle \\ &\leq \delta_{n+1} \|A_{S^n}^\dagger \mathbf{a}_{j^{n+1}}\|_2 \leq \frac{\delta_{n+1}}{\sqrt{1 - \delta_n}} \|A A_{S^n}^\dagger \mathbf{a}_{j^{n+1}}\|_2 = \frac{\delta_{n+1}}{\sqrt{1 - \delta_n}} \|P_n \mathbf{a}_{j^{n+1}}\|_2. \end{aligned}$$

After simplification, we obtain $\|P_n \mathbf{a}_{j^{n+1}}\|_2 \leq \delta_{n+1} / (\sqrt{1 - \delta_n})$. The condition $\delta_{\bar{n}} \leq (\sqrt{5} - 1)/2$ ensures that this bound does not exceed one. It then follows that, for $j \notin S^n$,

$$|(A^* (\mathbf{y} - A \mathbf{x}^n))_{j^{n+1}}|^2 \geq \left(1 - \frac{\delta_{n+1}^2}{1 - \delta_n}\right) |(A^* (\mathbf{y} - A \mathbf{x}^n))_j|^2.$$

The latter also holds for $j \in S^n$, in view of (6), hence (2) is established when $n + 1 \leq \bar{n}$ with $\rho := \sqrt{1 - \delta_{\bar{n}}^2} / (1 - \delta_{\bar{n}})$. \square

References

- [1] E. J. CANDÈS, *Compressive sampling*, Proceedings of the International Congress of Mathematicians, Madrid, Spain, 2006.
- [2] D. L. DONOHO, *Compressed sensing*, IEEE Transactions on Information Theory, 52(4), 1289–1306, 2006.
- [3] M. DAVENPORT AND M. WAKIN, *Analysis of Orthogonal Matching Pursuit using the Restricted Isometry Property*, IEEE Transactions on Information Theory 56(9), 4395–4401, 2010.
- [4] S. FOUCART, *Hard Thresholding Pursuit: an Algorithm for Compressive Sensing*, SIAM Journal on Numerical Analysis 49(6), 2543–2563, 2011.
- [5] E. LIU AND V. TEMLYAKOV, *Orthogonal Super Greedy Algorithm and Applications in Compressed Sensing*, preprint.
- [6] R. MALEH, *Improved RIP Analysis of Orthogonal Matching Pursuit*, preprint.
- [7] S.G MALLAT AND Z. ZHANG, *Matching pursuits with time-frequency dictionaries*, IEEE Transactions on Signal Processing, 41(12), 3397–3415, 1993.
- [8] Q. MO AND Y. SHEN, *Remarks on the Restricted Isometry Property in Orthogonal Matching Pursuit algorithm*, preprint.
- [9] H. RAUHUT, *On the impossibility of uniform sparse reconstruction using greedy methods*, Sampling Theory in Signal and Image Processing, 7(2),197–215, 2008.
- [10] J. TROPP, *Greedy is Good: Algorithmic Results for Sparse Approximation*, IEEE Transactions on Information Theory 50(10), 2231–2242, 2004.
- [11] T. ZHANG, *Sparse Recovery with Orthogonal Matching Pursuit under RIP*, IEEE Transactions on Information Theory 57(9), 6215–6221, 2011.