# GREEDY TYPE ALGORITHMS FOR RIP MATRICES. A STUDY OF TWO SELECTION RULES.

EUGENIO HERNÁNDEZ AND DANIEL VERA

ABSTRACT. On [24] some consequences of the Restricted Isometry Property (RIP) of matrices have been applied to develop a greedy algorithm called "ROMP" (Regularized Orthogonal Matching Pursuit) to recover sparse signals and to approximate non-sparse ones. These consequences were subsequently applied to other greedy and thresholding algorithms like "SThresh", "CoSaMP", "StOMP" and "SWCGP". In this paper, we find another consequence of the RIP property and use it to analyze the approximation to $k$-sparse signals with *Stagewise Weak* versions of Gradient Pursuit (SWGP), Matching Pursuit (SWMP) and Orthogonal Matching Pursuit (SWOMP) algorithms described in in [5]. We combine the above mentioned algorithms with another selection rule similar to the ones that appeared in [8] and [15] showing that results are obtained with less restrictions in the RIP constant, but we need a smaller threshold parameter for the coefficients. The results of some experiments are shown.

## 1. INTRODUCTION

One problem in Compressed Sensing (CS) is to reconstruct a $k-$sparse vector (all except $k$ elements are zero) $\mathbf{x} \in \mathbb{R}^N$ from a lower dimension vector $\mathbf{y} = \mathbf{\Phi x}$, where $\mathbf{\Phi} \in \mathbb{R}^{m \times N}$ is called a CS matrix or measurements ensemble. The aim of CS is to compress the signal while taking samples at the same time in such a way that a "good" reconstruction is possible. The next definition (see [6]) is a sufficient condition for the so called CS matrices to yield exact reconstruction of sparse signals with the Basis Pursuit (BP) algorithm or $\ell^1$ minimization with equality constraints (as firstly proposed by Donoho and collaborators for dictionaries in the signal processing community). This property will play a central role in the results developed here.

**Definition 1.1.** *Given $k \in \mathbb{N}$, a matrix $\mathbf{\Phi} \in \mathbb{C}^{m \times N}$ ($m > k$) is said to satisfy the Restricted Isometry Property with parameter $\delta_k$, $0 < \delta_k < 1$ (called the Restricted Isometry Constant), if*

$$(1 - \delta_k) \|\mathbf{x}\|^2_{\ell^2(\mathbb{R}^N)} \leq \|\mathbf{\Phi x}\|^2_{\ell^2(\mathbb{R}^m)} \leq (1 + \delta_k) \|\mathbf{x}\|^2_{\ell^2(\mathbb{R}^N)} \qquad (1.1)$$

*for **all** $k-$sparse vectors $\mathbf{x} \in \mathbb{R}^N$.*

It is known that those matrices that satisfy RIP and allow the least number of measurements $m$ for reconstruction of all sparse signals are some random matrices (therefore reconstruction is in probability). For example, in [6] and [14], it is shown that it suffices to take $m$ linearly with the sparsity $k$ and polylogarithmic with the

ambient dimension $N$, *i.e.* $m \geq Ck \log(N/k)$ for Gaussian and Bernoulli matrices, and $m \geq Ck \log^6(N)$ for Fourier matrices (for better bounds see [26]) to reconstruct, with high probability, a $k$-sparse vector with the BP algorithm.

It has become standard to use greedy algorithms to iteratively identify the support $\Gamma^\sharp := \mathrm{supp}\,(\mathbf{x})$ of a sparse signal. This is done by computing the inner product of the residue $\mathbf{r}^{n-1}$ of the approximation at step $n-1$ and the columns of the matrix $\mathbf{\Phi}$, *i.e.* $\mathbf{g}^n = \mathbf{\Phi}^* \mathbf{r}^{n-1}$ ($\mathbf{\Phi}^*$ denotes de transpose of $\mathbf{\Phi}$) , and then select the largest element(s) (in absolute value) in $\mathbf{g}^n = (g_1^n, \ldots, g_N^n)^t$, where each $g_i^n = \langle \phi_i, \mathbf{r}^{n-1} \rangle$ and $\phi_i$ is the $i$-th column of $\mathbf{\Phi}$. Since $\mathbf{\Phi}$ verifies the RIP property, then such an inner product gives an idea on where the support may be because the square of the energies of the $k$-sparse vector signal $\mathbf{x}$ and the observation vector $\mathbf{y} = \mathbf{\Phi}\mathbf{x}$ should not differ more than $\delta_k$. To see this more clearly, suppose we know the true support of $\mathbf{x}$, $\Gamma^\sharp = \mathrm{supp}\,(\mathbf{x})$, and $\mathbf{y} = \mathbf{\Phi}\mathbf{x}$ is the observation. Writing $\mathbf{\Phi}_\Gamma$ (resp. $\mathbf{x}_\Gamma$) to denote the matrix $\mathbf{\Phi}$ (resp. the vector $\mathbf{x}$) restricted to the columns (resp. the elements) indexed by $\Gamma \subset \{1, 2, \ldots, N\}$, and $(\mathbf{\Phi}_{\Gamma^\sharp})^\dagger = (\mathbf{\Phi}_{\Gamma^\sharp}^* \mathbf{\Phi}_{\Gamma^\sharp})^{-1} \mathbf{\Phi}_{\Gamma^\sharp}^*$ for the pseudo inverse of $\mathbf{\Phi}_{\Gamma^\sharp}$ (which exists by (1.1)), we can recover $\mathbf{x}$ from $\mathbf{y}$ using $(\mathbf{\Phi}_{\Gamma^\sharp})^\dagger$ since $\mathbf{y} = \mathbf{\Phi}\mathbf{x} = \mathbf{\Phi}_{\Gamma^\sharp} \mathbf{x}_{\Gamma^\sharp}$ and

$$(\mathbf{\Phi}_{\Gamma^\sharp})^\dagger \mathbf{y} = (\mathbf{\Phi}_{\Gamma^\sharp}^* \mathbf{\Phi}_{\Gamma^\sharp})^{-1} \mathbf{\Phi}_{\Gamma^\sharp}^* \mathbf{\Phi}_{\Gamma^\sharp} \mathbf{x}_{\Gamma^\sharp} = \mathbf{x}_{\Gamma^\sharp}. \tag{1.2}$$

We shall use the notation $\mathbb{R}^\Gamma$ to denote the subspace of the ambient space $\mathbb{R}^N$ with significant coordinates in $\Gamma \subset \{1, \ldots, N\}$. Notice the prominent role of $\mathbf{\Phi}_{\Gamma^\sharp}^* : \mathbb{R}^m \to \mathbb{R}^{\Gamma^\sharp}$ in the above argument. The matrix $\mathbf{\Phi}^*$ (the transpose of $\mathbf{\Phi}$) will be used in the algorithms below. Also, $\mathbf{r} \in \mathrm{span}\,(\mathbf{\Phi}_\Gamma)$ means that $\mathbf{r}$ is a linear combination of the columns of $\mathbf{\Phi}$ indexed by $\Gamma$.

Section 2 studies conditions to identify the support of a sparse signal sensed with RIP matrices using different greedy type algorithms. A review of *Stagewise Weak* versions of the Gradient Pursuit (SWGP), Matching Pursuit (SWMP), and Orthogonal Matching Pursuit (SWOMP) algorithms, as first proposed in [5], is done in Subsection 2.1. The main novelty, as pointed out in [5], is that not only one but several elements are allowed to be selected in each iteration. This is a feature also present in [15], a paper published in 2012 but circulated since 2006 as a preprint. Weak algorithms were used in non-linear approximation theory before the development of Compressed Sensing (see [17], [2] and more recently [12], [27], [28], [29] and the references therein), but in all of these works only one element was selected at each iteration.

Known properties of RIP matrices are stated in Subsection 2.2, while a new property of these matrices is proved in Subsection 2.3 (see Lemma 2.2). We give conditions on $\mathbf{\Phi}$ and on the weakness parameter $\alpha$ of the selection rule to identify the support of a sparse signal in Subsection 2.4 (see Theorem 2.3) for all of the *Stagewise Weak* algorithms mentioned above.

Another selection rule, called **relaxed** in this paper, is introduced in Section 3 given rise to new algorithms that we name Relaxed Weak Gradient Pursuit (RWGP), Relaxed Weak Matching Pursuit (RWMP) and Relaxed Weak Orthogonal Matching Pursuit (RWOMP). The strategy of selecting several elements in each iteration is also

used in these algorithms. In the Relaxed selection rule, elements are chosen if their magnitude is larger than a fraction of the energy of the residue at an iteration. This procedure has also been used in [8] and [15] (see the second paragraph in Section 3). The name Weak Relaxed has appeared in the non-linear approximation theory associated to greedy algorithms (see [12], [27]), but with a different meaning that in this paper. In Theorem 3.1 we give conditions on a matrix $\mathbf{\Phi}$ satisfying RIP and on the weakness parameter $\alpha$ of the Relaxed selection rule to identify the support of a sparse signal.

In Section 4 the convergence of all the above algorithms is studied. The energy of the residual of the observation at iteration $n$ is compared with the energy at iteration $n-1$, i.e, $\|\mathbf{r}^n\|_{\ell^2} \leq C_k \|\mathbf{r}^{n-1}\|_{\ell^2}$. For the GP, SWGP and RWGP algorithms we establish in Theorem 4.1 the above inequality with $C_k = (1 - \frac{1-\delta_k}{k(1+\delta_k)})^{1/2} < 1$; this is a more explicit version than the ones already known for GP and SWGP (see details in Section 4). For the SWMP, SWOMP, RWMP, and RWOMP the result is stated in Theorem 4.7 giving $C'_k = (1 - \frac{(1-\delta_k)^2}{k})^{1/2} < 1$. Notice that Gradient Pursuit algorithms seem to have stronger rate of convergence than Matching Pursuit ones, a fact present in the experiments shown for images in Section 6.

In Section 5 we study the behavior of the selection rules with particular Gaussian and Bernoulli random matrices not necessarily satisfying RIP. Here we prove that with high probability the algorithms allow to recover the position of the $k$ entries of a given $k$-sparse signal. Finally, some experiments are shown in Section 6 for the algorithms described in the above sections, and compare results with already existing algorithms to recover sparse signals and approximate compressible images with a sparse representation.

## 2. Support Identification with RIP.

We will review Stagewise Weak versions of the Gradient Pursuit (SWGP), Matching Pursuit (SWMP) and Orthogonal Matching Pursuit (SWOMP) algorithms and some consequences of the RIP property. Then, we will develop a new consequence of RIP and find some conditions so the algorithms select elements on the support of a $k-$sparse signal $\mathbf{x}$ on each iteration.

### 2.1. **SWGP, SWMP and SWOMP algorithms.** 
The Stagewise Weak Gradient Pursuit (SWGP), Stagewise Weak Matching Pursuit (SWMP) and Stagewise Weak Orthogonal Matching Pursuit (SWOMP) algorithms select a set of elements (possibly not new) in each iteration by comparing the maximum of the inner products between the columns of the measurement ensemble with the residue at the previous iteration. This stage in the algorithms is the selection rule, where a weakness parameter $\alpha$ is introduced. The main differences between these algorithms are the direction search and the way to update the approximation. For some history on MP and OMP see [30] and references therein. The *stagewise weak selection rule* is defined in [5].

## STAGEWISE WEAK MATCHING PURSUIT (SWMP)

The MP algorithm was the first of the greedy algorithms to be used in signal processing (see [23]); as far as we know, the weakness parameter appeared for the first time in [17] and has been extensively used in non-linear approximation (see [12], [27], [28], [29] and the references therein).

In the initialisation we set: at iteration $n = 0$ the residue is $\mathbf{r}^0 = \mathbf{y}$, the approximation to the observation is $\mathbf{y}^0 = 0$ and the estimation of the signal is $\mathbf{x}^0 = 0$. Recall that $\phi_i$ denotes the $i$-th column of $\mathbf{\Phi}$. The loop until some criteria are met follows the next steps:

- $\mathbf{g}^n = \mathbf{\Phi}^* \mathbf{r}^{n-1}$, the "proxy" of the signal.
- $\mathcal{I}_n := \mathcal{I}_n(\alpha) := \left\{ i : |g_i^n| \geq \alpha \, \|\mathbf{\Phi}^* \mathbf{r}^{n-1}\|_{\ell^\infty(\mathbb{R}^N)} \right\}$, the stagewise weak selection rule with $0 < \alpha \leq 1$.
- $\mathbf{y}^n = \mathbf{y}^{n-1} + \sum_{i \in \mathcal{I}_n} g_i^n \phi_i = \mathbf{y}^{n-1} + \sum_{i \in \mathcal{I}_n} \langle \phi_i, \mathbf{r}^{n-1} \rangle \phi_i$, approximation to the observation.
- $x_i^n = x_i^{n-1} + g_i^n = x_i^{n-1} + \langle \phi_i, \mathbf{r}^{n-1} \rangle$, $i \in \mathcal{I}_n$ and $x_j^n = x_j^{n-1}$ if $j \notin \mathcal{I}_n$, estimation of the signal.
- $\mathbf{r}^n = \mathbf{r}^{n-1} - \sum_{i \in \mathcal{I}_n} g_i^n \phi_i = \mathbf{r}^{n-1} - \sum_{i \in \mathcal{I}_n} \langle \phi_i, \mathbf{r}^{n-1} \rangle \phi_i$, the residual.

By recursion one can see from the definition of $\mathbf{y}^n$ and $\mathbf{r}^n$ that $\mathbf{r}^n = \mathbf{r}^{n-1} - (\mathbf{y}^n - \mathbf{y}^{n-1}) = \cdots = \mathbf{y} - \mathbf{y}^n$.

## STAGEWISE WEAK ORTHOGONAL MATCHING PURSUIT (SWOMP)

SWOMP is similar to SWMP. Once $\mathcal{I}_n$ has been selected in SWMP, the updating of the approximation with $\sum_{i \in \mathcal{I}_n} g_i^n \phi_i$ might not be the best approximation from the subspace spanned by the columns $\{\phi_i\}_{i \in \mathcal{I}_n}$. For SWOMP instead, the update of the approximation is $\mathbf{y}^n = P_{\Gamma^n} \mathbf{y}$ where $P_{\Gamma^n}$ is the orthogonal projection into the span of the columns $\Gamma^n$ of $\mathbf{\Phi}$: $\mathbf{y}^n = \mathbf{\Phi}_{\Gamma^n} \mathbf{\Phi}_{\Gamma^n}^\dagger \mathbf{y} = \mathbf{\Phi}_{\Gamma^n} (\mathbf{\Phi}_{\Gamma^n}^* \mathbf{\Phi}_{\Gamma^n})^{-1} \mathbf{\Phi}_{\Gamma^n}^* \mathbf{y}$ on the indices $\Gamma^n = \bigcup_{k=1}^n \mathcal{I}_k$ and therefore the residue becomes $\mathbf{r}^n = \mathbf{y} - P_{\Gamma^n} \mathbf{y}$. The residue is then orthogonal to all elements previously selected as can be seen from

$$\begin{aligned} \mathbf{\Phi}_{\Gamma^n}^* \mathbf{r}^n &= \mathbf{\Phi}_{\Gamma^n}^* (\mathbf{y} - \mathbf{\Phi}_{\Gamma^n} \mathbf{\Phi}_{\Gamma^n}^\dagger \mathbf{y}) \\ &= \mathbf{\Phi}_{\Gamma^n}^* \mathbf{y} - \mathbf{\Phi}_{\Gamma^n}^* \mathbf{\Phi}_{\Gamma^n} (\mathbf{\Phi}_{\Gamma^n}^* \mathbf{\Phi}_{\Gamma^n})^{-1} \mathbf{\Phi}_{\Gamma^n}^* \mathbf{y} = 0. \end{aligned}$$

Thus, at every iteration new elements are selected. The initialisation is as in SWMP and the recursion loop is:

- $\mathbf{g}^n = \mathbf{\Phi}^* \mathbf{r}^{n-1}$, the "proxy" of the signal.
- $\mathcal{I}_n := \mathcal{I}_n(\alpha) := \left\{ i : |g_i^n| \geq \alpha \, \|\mathbf{\Phi}^* \mathbf{r}^{n-1}\|_{\ell^\infty(\mathbb{R}^N)} \right\}$, the stagewise weak selection rule with $0 < \alpha \leq 1$.
- $\Gamma^n = \Gamma^{n-1} \bigcup \mathcal{I}_n$, update of selected elements.
- $\mathbf{x}^n := \mathbf{x}^n|_{\Gamma^n} = \mathbf{\Phi}_{\Gamma^n}^\dagger \mathbf{y}$, estimation of the signal.
- $\mathbf{y}^n = \mathbf{\Phi}_{\Gamma^n} \mathbf{\Phi}_{\Gamma^n}^\dagger \mathbf{y}$, approximation to the observation.
- $\mathbf{r}^n = \mathbf{y} - \mathbf{y}^n$, the residual.

## STAGEWISE WEAK GRADIENT PURSUIT (SWGP)

The Gradient Pursuit (GP) is described in [4] and the Stagewise Weak Gradient Pursuit (SWGP) was developed in [5] (together with some other variants). We describe SWGP since GP is obtained from SWGP by setting the weakness parameter $\alpha = 1$ in the selection rule.

At iteration $n = 0$ we have: $\mathbf{x}^0 = 0$, the estimation of the signal; $\mathbf{r}^0 = \mathbf{y}$, the residue; $\Gamma^0 = \emptyset$, the support. Then, the recursion at step $n$ until some criteria are met is the following:

- $\mathbf{g}^n = \mathbf{\Phi}^* \mathbf{r}^{n-1}$, the "proxy" of the signal.
- $\mathcal{I}_n := \mathcal{I}_n(\alpha) := \{i : |g_i^n| \geq \alpha \|\mathbf{r}^{n-1}\|_{\ell^\infty(\mathbb{R}^N)}\}$, the stagewise weak selection rule with $0 < \alpha \leq 1$.
- $\Gamma^n = \Gamma^{n-1} \bigcup \mathcal{I}_n$, the updated support.
- $\mathbf{d}_{\Gamma^n}^n = \mathbf{\Phi}_{\Gamma^n}^* \mathbf{r}^{n-1}$, the updated direction.
- $a^n = \frac{\langle \mathbf{r}^{n-1}, \mathbf{\Phi}_{\Gamma^n} \mathbf{d}_{\Gamma^n}^n \rangle}{\|\mathbf{\Phi}_{\Gamma^n} \mathbf{d}_{\Gamma^n}^n\|_{\ell^2(\mathbb{R}^m)}^2}$, the optimised step.
- $\mathbf{x}^n := \mathbf{x}_{\Gamma^n}^n = \mathbf{x}^{n-1} + a^n \mathbf{d}_{\Gamma^n}^n$, the estimation to the signal.
- $\mathbf{y}^n = \mathbf{y}^{n-1} + a^n \mathbf{\Phi}_{\Gamma^n} \mathbf{d}_{\Gamma^n}^n$, the approximation to the observation.
- $\mathbf{r}^n = \mathbf{r}^{n-1} - a^n \mathbf{\Phi}_{\Gamma^n} \mathbf{d}_{\Gamma^n}^n$.

Again, one can prove that $\mathbf{r}^n = \mathbf{y} - \mathbf{y}^n$.

In [5] it is proved that SWGP converges to the solution at least as good as the simpler version GP.

## 2.2. Some Consequences of RIP.

This section is based on some implicit results in [6], see especially Lemma 2.1, that were made explicit in [24]; for the proofs the reader can also see [8].

**Lemma 2.1.** *Assume that $\mathbf{\Phi} \in \mathbb{R}^{m \times N}$ satisfies RIP with $\delta_k$, $|\Gamma| \leq k$, $supp(\mathbf{u}) = \Gamma$. Then*

$$\|\mathbf{\Phi}_\Gamma\|_{\ell^2(\mathbb{R}^\Gamma) \to \ell^2(\mathbb{R}^m)} = \|\mathbf{\Phi}_\Gamma^*\|_{\ell^2(\mathbb{R}^m) \to \ell^2(\mathbb{R}^\Gamma)} \leq (1 + \delta_k)^{1/2}, \tag{2.1}$$

$$(1 - \delta_k) \|\mathbf{u}\|_{\ell^2(\mathbb{R}^\Gamma)} \leq \|\mathbf{\Phi}_\Gamma^* \mathbf{\Phi}_\Gamma \mathbf{u}\|_{\ell^2(\mathbb{R}^\Gamma)} \leq (1 + \delta_k) \|\mathbf{u}\|_{\ell^2(\mathbb{R}^\Gamma)}, \tag{2.2}$$

$$(1 + \delta_k)^{-1} \|\mathbf{u}\|_{\ell^2(\mathbb{R}^\Gamma)} \leq \|(\mathbf{\Phi}_\Gamma^* \mathbf{\Phi}_\Gamma)^{-1} \mathbf{u}\|_{\ell^2(\mathbb{R}^\Gamma)} \leq (1 - \delta_k)^{-1} \|\mathbf{u}\|_{\ell^2(\mathbb{R}^\Gamma)}, \tag{2.3}$$

*For disjoint sets $\Gamma', \Gamma$ such that $|\Gamma' \bigcup \Gamma| \leq k$, we have*

$$\|\mathbf{\Phi}_{\Gamma'}^* \mathbf{\Phi}_\Gamma \mathbf{u}\|_{\ell^2(\mathbb{R}^{\Gamma'})} \leq \delta_k \|\mathbf{u}\|_{\ell^2(\mathbb{R}^\Gamma)}. \tag{2.4}$$

All of these results have become standard for the greedy algorithms in compressed sensing.

## 2.3. One more consequence.

We derive here another bound as a consequence of the RIP property.

**Lemma 2.2.** *Let $\mathbf{\Phi}$ verify RIP with parameter $\delta_k$ and let $|\Gamma| \leq k$. For all $\mathbf{r} \in span(\mathbf{\Phi}_\Gamma)$*

$$\|\mathbf{\Phi}_\Gamma^* \mathbf{r}\|_{\ell^2(\mathbb{R}^\Gamma)} \geq (1 - \delta_k)^{\frac{1}{2}} \|\mathbf{r}\|_{\ell^2(\mathbb{R}^m)} \tag{2.5}$$

**Proof.** For the pseudo-inverse we have

$$\|\mathbf{\Phi}_\Gamma^\dagger \mathbf{r}\|_{\ell^2(\mathbb{R}^\Gamma)} = \|(\mathbf{\Phi}_\Gamma^* \mathbf{\Phi}_\Gamma)^{-1} \mathbf{\Phi}_\Gamma^* \mathbf{r}\|_{\ell^2(\mathbb{R}^\Gamma)}$$
$$\leq (1 - \delta_k)^{-1} \|\mathbf{\Phi}_\Gamma^* \mathbf{r}\|_{\ell^2(\mathbb{R}^\Gamma)},$$

where we have used (2.3) in the last inequality. We next use this result and the Schwarz inequality to get

$$
\begin{aligned}
\|\mathbf{\Phi}_\Gamma^* \mathbf{r}\|_{\ell^2(\mathbb{R}^\Gamma)}^2 \ &\geq\ (1-\delta_k)\left\|\mathbf{\Phi}_\Gamma^\dagger \mathbf{r}\right\|_{\ell^2(\mathbb{R}^\Gamma)} \|\mathbf{\Phi}_\Gamma^* \mathbf{r}\|_{\ell^2(\mathbb{R}^\Gamma)} \\
&\geq\ (1-\delta_k)\langle \mathbf{\Phi}_\Gamma^\dagger \mathbf{r}, \mathbf{\Phi}_\Gamma^* \mathbf{r}\rangle = (1-\delta_k)(\mathbf{\Phi}_\Gamma^\dagger \mathbf{r})^* \mathbf{\Phi}_\Gamma^* \mathbf{r} \\
&=\ (1-\delta_k)\mathbf{r}^* \mathbf{\Phi}_\Gamma (\mathbf{\Phi}_\Gamma^* \mathbf{\Phi}_\Gamma)^{-1} \mathbf{\Phi}_\Gamma^* \mathbf{r} \\
&=\ (1-\delta_k)\mathbf{y}_\Gamma^* \mathbf{\Phi}_\Gamma^* \mathbf{\Phi}_\Gamma (\mathbf{\Phi}_\Gamma^* \mathbf{\Phi}_\Gamma)^{-1} \mathbf{\Phi}_\Gamma^* \mathbf{\Phi}_\Gamma \mathbf{y}_\Gamma \\
&=\ (1-\delta_k)\mathbf{y}_\Gamma^* \mathbf{\Phi}_\Gamma^* \mathbf{\Phi}_\Gamma \mathbf{y}_\Gamma = (1-\delta_k)\|\mathbf{r}\|_{\ell^2(\mathbb{R}^m)}^2,
\end{aligned}
$$

using the fact that, since $\mathbf{r}$ is in the span of $\mathbf{\Phi}_\Gamma$, we may write $\mathbf{r} = \mathbf{\Phi}_\Gamma \mathbf{y}_\Gamma$ for some $\mathbf{y}_\Gamma$, and conclude the proof. ∎

2.4. **Support Identification with SWGP, SWMP and SWOMP.** We now give sufficient conditions on the matrix $\mathbf{\Phi}$ so that the SWGP, SWMP and SWOMP algorithms select elements on the support $\Gamma^\sharp$ of the sparse signal $\mathbf{x}$. The SWOMP algorithm will select new atoms in each iteration due to the orthogonality with previous residues, therefore convergence to exact reconstruction of $k-$sparse vectors is guaranteed in at most $k$ iterations when the condition is met. As previously mentioned, the stagewise weak selection rule is

$$
\mathcal{I}_n := \mathcal{I}_n(\alpha) := \left\{ i : |g_i^n| \geq \alpha \left\|\mathbf{\Phi}^* \mathbf{r}^{n-1}\right\|_{\ell^\infty(\mathbb{R}^N)} \right\}, \tag{2.6}
$$

for some $\alpha \in (0,1]$. For the next result we follow the line of reasoning of [30], (see also [16] and [5]), where the results are given for OMP on *quasi-incoherent dictionaries*. Here we use Lemma 2.2.

**Theorem 2.3.** *Let $\mathbf{\Phi}$ satisfies RIP with $\delta_{k+1}$. A sufficient condition for the SWGP, SWMP and SWOMP algorithms with selection rule $\mathcal{I}_n(\alpha)$ given by (2.6), $0 < \alpha \leq 1$, to identify elements in the support $\Gamma^\sharp$ of the $k-$sparse signal $\mathbf{x}$ is*

$$
\alpha > \frac{\sqrt{k}\delta_{k+1}}{1-\delta_k}. \tag{2.7}
$$

**Proof.** Since $\operatorname{supp}(\mathbf{x}) = \Gamma^\sharp$, then $\mathbf{r}^0 \in \operatorname{span}(\mathbf{\Phi}_{\Gamma^\sharp})$. The algorithms update the approximations and residuals precisely in the indices contained in $\mathcal{I}_n$, so that to proceed by induction we can assume that after $n-1$ iterations we have $\mathbf{r}^{n-1} \in \operatorname{span}(\mathbf{\Phi}_{\Gamma^\sharp})$. We drop the superindex of $\mathbf{r}$ for the proof. The condition $\mathcal{I}_n \subset \Gamma^\sharp$ is implied by

$$
\left\|\mathbf{\Phi}_{\Gamma^{\sharp c}}^* \mathbf{r}\right\|_{\ell^\infty(\mathbb{R}^{\Gamma^{\sharp c}})} < \alpha \left\|\mathbf{\Phi}_{\Gamma^\sharp}^* \mathbf{r}\right\|_{\ell^\infty(\mathbb{R}^{\Gamma^\sharp})}.
$$

Using the assumption on $\mathbf{r}$, we can write $\mathbf{r} = \mathbf{\Phi}_{\Gamma^\sharp} \mathbf{y}_{\Gamma^\sharp}$ for some $\mathbf{y}_{\Gamma^\sharp}$. Rearranging to express the condition as a quotient (called the *greedy selection ratio*), squaring and choosing $\lambda$ as one index in $\Gamma^{\sharp c}$ with the largest value, it yields

$$
\begin{aligned}
\frac{\left\|\mathbf{\Phi}_{\Gamma^{\sharp c}}^* \mathbf{r}\right\|_{\ell^\infty(\mathbb{R}^{\Gamma^{\sharp c}})}^2}{\left\|\mathbf{\Phi}_{\Gamma^\sharp}^* \mathbf{r}\right\|_{\ell^\infty(\mathbb{R}^{\Gamma^\sharp})}^2} \ &=\ \frac{\left\|\mathbf{\Phi}_\lambda^* \mathbf{r}\right\|_{\ell^\infty(\mathbb{R}^\lambda)}^2}{\left\|\mathbf{\Phi}_{\Gamma^\sharp}^* \mathbf{r}\right\|_{\ell^\infty(\mathbb{R}^{\Gamma^\sharp})}^2} \leq \frac{\left\|\mathbf{\Phi}_\lambda^* \mathbf{r}\right\|_{\ell^2(\mathbb{R}^\lambda)}^2}{\frac{1}{k}\left\|\mathbf{\Phi}_{\Gamma^\sharp}^* \mathbf{r}\right\|_{\ell^2(\mathbb{R}^{\Gamma^\sharp})}^2} \\
&=\ \frac{k\left\|\mathbf{\Phi}_\lambda^* \mathbf{\Phi}_{\Gamma^\sharp} \mathbf{y}_{\Gamma^\sharp}\right\|_{\ell^2(\mathbb{R}^\lambda)}^2}{\left\|\mathbf{\Phi}_{\Gamma^\sharp}^* \mathbf{\Phi}_{\Gamma^\sharp} \mathbf{y}_{\Gamma^\sharp}\right\|_{\ell^2(\mathbb{R}^{\Gamma^\sharp})}^2} \leq \frac{k\delta_{k+1}^2}{(1-\delta_k)^2}
\end{aligned}
$$

where the first inequality is due to usual norm inequalities and the second by (2.4) in the numerator and consecutive application of (2.5) and the left-hand side of (1.1) in the denominator. Comparing the right side of the last inequality with $\alpha^2$ ends the proof. ∎

**Corollary 2.4.** *With the conditions of Theorem 2.3, if any of the algorithms has selected $k$ elements, then it has found the whole support of the $k$-sparse vector $\mathbf{x}$.*

Since SWOMP, as well as OMP, select new elements at each iteration, in at most $k$ steps they recover the $k$-sparse signal by (1.2). Thus we have

**Corollary 2.5.** *With the conditions of Theorem 2.3, the SWOMP and OMP algorithms recover every $k$-sparse vector $\mathbf{x}$ in at most $k$ iterations.*

**Remark 2.6.** *Observe that $\delta_k \leq \delta_{k+1}$ since the set of all $k$-sparse vectors is contained in the set of all $(k+1)$-sparse vectors. Moreover, to achieve (2.7) the RIP constants $\delta_{k+1}$ and $\delta_k$ must satisfy $\delta_{k+1} < \alpha(1 - \delta_k)/\sqrt{k}$. This gives a restriction on $k$, that is $k < (\alpha(1 - \delta_k)/\delta_{k+1})^2$.*

**Remark 2.7.** *The proof of Theorem 2.3 follow the ideas of the proof in [30], regarding deterministic quasi-incoherent dictionaries, which uses the fact that $\mathbf{r} = P_{\Gamma^\sharp}\mathbf{r} = \mathbf{\Phi}_{\Gamma^\sharp}\mathbf{\Phi}_{\Gamma^\sharp}^\dagger\mathbf{r}$ (that is OMP type algorithms) and then bounds with usual norm inequalities. The condition obtained is then called* Exact Reconstruction Condition $ERC_\alpha$ *for* quasi-incoherent dictionaries *since the result is on OMP and therefore exact reconstruction is guaranteed in at most $k$ iterations; a similar result is called* Stability Condition *in [16] since it is applied to MP and no exact reconstruction of sparse signals is guaranteed in $k$ iterations in this case. In our case, a straight use of the consequences of the RIP property lead us to the result.*

**Remark 2.8.** *Suppose that $\delta_{k+1} < \frac{1}{1+\sqrt{k}}$. Since $\delta_k \leq \delta_{k+1}$ we have*

$$\delta_{k+1} < \frac{1}{1+\sqrt{k}} = \frac{1 - \frac{1}{1+\sqrt{k}}}{\sqrt{k}} \leq \frac{1 - \delta_k}{\sqrt{k}}.$$

*Thus, condition (2.7) is satisfied with $\alpha = 1$. By Corollary 2.5, if $\delta_{k+1} < \frac{1}{1+\sqrt{k}}$ the OMP algorithm recovers any $k$-sparse vector $\mathbf{x}$ in at most $k$ iterations. This result has recently appeared in [20] and [22] (see also [11] and [18] for previous smaller bounds on $\delta_{k+1}$). Moreover, it is proved in [22] that there exits a $k$-sparse vector $\mathbf{x} \in \mathbb{R}^{k+1}$ and a $(k+1) \times (k+1)$ matrix $\Phi$ satisfying RIP with $\delta_{k+1} = \frac{1}{\sqrt{k}}$ such that OMP does not recover $\mathbf{x}$ in $k$ iterations (proving a conjecture estated in [9]). Thus the bound $\delta_{k+1} < \frac{1}{1+\sqrt{k}}$ is nearly optimal.*

## 3. Support Identification with a Relaxed Weak Selection Rule

Next, we consider another decision rule to select indices in the true support of a sparse signal. With the same notation as in section 2, let

$$\tilde{\mathcal{I}}_n := \tilde{\mathcal{I}}_n(\tilde{\alpha}) := \left\{ i : |g_i^n| \geq \tilde{\alpha} \left\|\mathbf{r}^{n-1}\right\|_{\ell^2(\mathbb{R}^m)} \right\}, \tag{3.1}$$

be the **relaxed weak** selection rule. Rule (3.1) compares the absolute value of $g_i^n = \langle \phi_i, \mathbf{r}^{n-1} \rangle$ with $\|\mathbf{r}^{n-1}\|_{\ell^2(\mathbb{R}^m)}$. It can be proved (see the proof of Theorem 9.10 in [21], p. 422) that there exists $\beta_0 > 0$ such that for any $\mathbf{x} \in \mathbb{R}^m$, $\sup_{i=1,\dots,N} |\langle \phi_i, \mathbf{x} \rangle| \geq \beta_0 \|\mathbf{x}\|_{\ell^2(\mathbb{R}^m)}$. Thus, $\tilde{\mathcal{I}}_n(\alpha)$ is a non empty set if we choose $\alpha \leq \beta_0$.

Selection rule (3.1) is similar to the ones considered in [15] and [8]. In [15] the rule is

$$\mathcal{J}_n(t) := \left\{ i : |g_i^n| \geq \frac{t}{\sqrt{m}} \|\mathbf{r}^{n-1}\|_{\ell^2(\mathbb{R}^m)} \right\},$$

and the algorithm developed is called StOMP. Thus, $\tilde{\mathcal{I}}_n(\tilde{\alpha})$ is $\mathcal{J}_n(t)$ with $\tilde{\alpha} = t/\sqrt{m}$. In [8] the rule is

$$\mathcal{D}_n(\delta) := \left\{ i : |g_i^n| \geq \frac{t}{\sqrt{k}} \|\mathbf{r}^{n-1}\|_{\ell^2(\mathbb{R}^m)} \right\},$$

and the algorithm developed is called DTresh (and its cousin STresh). Thus, $\tilde{\mathcal{I}}_n(\tilde{\alpha})$ is $\mathcal{D}_n(t)$ with $\tilde{\alpha} = t/\sqrt{k}$. In both of these algorithms the update of the approximation is done using the same updating as the OMP algorithm.

In rule (3.1) we compare with the energy of the residual $\|\mathbf{r}^{n-1}\|_{\ell^2(\mathbb{R}^m)}$ instead of comparing with the maximum of the correlations of the residue with the "atoms" of $\mathbf{\Phi}$, i.e. $\|\mathbf{\Phi}^* \mathbf{r}^{n-1}\|_{\ell^\infty(\mathbb{R}^N)}$. We choose the term "**relaxed**" since, as we will see in Theorem 3.1, the condition on $\delta_{k+1}$ and $\delta_k$ is weaker than the one required in Theorem 2.3. Therefore, we decided to call the greedy algorithms with this selection rule **Relaxed** WGP, WMP or WOMP algorithms, writing RWGP, RWMP and RWOMP respetively.

We still consider $\mathbf{x}$ a $k-$sparse vector with support on $\Gamma^\sharp$.

**Theorem 3.1.** *Let $\mathbf{\Phi}$ satisfies RIP with $\delta_{k+1}$. A sufficient condition for the RWGP, RWMP and RWOMP algorithms to identify elements on the support $\Gamma^\sharp$ of $\mathbf{x}$ with the **relaxed** weak selection rule $\tilde{\mathcal{I}}_n(\tilde{\alpha})$ given by (3.1) is that $\tilde{\mathcal{I}}_n(\tilde{\alpha}) \neq \emptyset$ and*

$$\tilde{\alpha} > \frac{\delta_{k+1}}{(1 - \delta_k)^{1/2}}. \tag{3.2}$$

**Proof.** It is basically the same proof as for Theorem 2.3. Again, we drop the superindex of $\mathbf{r}$ and make the assumption $\mathbf{r} \in \text{span}(\mathbf{\Phi}_{\Gamma^\sharp})$ to proceed by induction. The condition $\tilde{\mathcal{I}}_n \subset \Gamma^\sharp$ is implied by

$$\|\mathbf{\Phi}_{\Gamma^{\sharp c}}^* \mathbf{r}\|_{\ell^\infty(\Gamma^{\sharp c})} < \tilde{\alpha} \|\mathbf{r}\|_{\ell^2(\mathbb{R}^m)}.$$

As before, rearranging to express it as a quotient (called *relaxed weak greedy selection ratio*), squaring and choosing $\lambda$ as one index in $\Gamma^{\sharp c}$ with the largest value of $\mathbf{\Phi}_{\Gamma^{\sharp c}}^* \mathbf{r}$, it

gives

$$
\begin{aligned}
\frac{\left\|\mathbf{\Phi}_{\Gamma^{\sharp c}}^{*}\mathbf{r}\right\|_{\ell^{\infty}(\mathbb{R}^{\Gamma^{\sharp c}})}^{2}}{\|\mathbf{r}\|_{\ell^{2}(\mathbb{R}^{m})}^{2}}
\ &= \ \frac{\left\|\mathbf{\Phi}_{\lambda}^{*}\mathbf{r}\right\|_{\ell^{\infty}(\mathbb{R}^{\lambda})}^{2}}{\left\|\mathbf{\Phi}_{\Gamma^{\sharp}}\mathbf{y}_{\Gamma^{\sharp}}\right\|_{\ell^{2}(\mathbb{R}^{m})}^{2}}
= \frac{\left\|\mathbf{\Phi}_{\lambda}^{*}\mathbf{\Phi}_{\Gamma^{\sharp}}\mathbf{y}_{\Gamma^{\sharp}}\right\|_{\ell^{2}(\mathbb{R}^{\lambda})}^{2}}{\left\|\mathbf{\Phi}_{\Gamma^{\sharp}}\mathbf{y}_{\Gamma^{\sharp}}\right\|_{\ell^{2}(\mathbb{R}^{m})}^{2}} \\[2mm]
&\leq \ \frac{\left\|\mathbf{\Phi}_{\lambda}^{*}\mathbf{\Phi}_{\Gamma^{\sharp}}\right\|_{\ell^{2}(\mathbb{R}^{\Gamma^{\sharp}})\to\ell^{2}(\mathbb{R}^{\lambda})}^{2}\left\|\mathbf{y}_{\Gamma^{\sharp}}\right\|_{\ell^{2}(\mathbb{R}^{\Gamma^{\sharp}})}^{2}}{\left\|\mathbf{\Phi}_{\Gamma^{\sharp}}\mathbf{y}_{\Gamma^{\sharp}}\right\|_{\ell^{2}(\mathbb{R}^{m})}^{2}} \\[2mm]
&\leq \ \frac{\delta_{k+1}^{2}}{(1-\delta_{k})},
\end{aligned}
$$

where the first and second inequalities are due to usual norm inequalities and the third by (2.4) in the numerator and the left side of (1.1) in the denominator. Comparing the left hand side of the last inequality with $\tilde{\alpha}^{2}$ ends the proof. ∎

**Remark 3.2.** *Suppose* $\mathbf{r} \in span\,(\mathbf{\Phi}_{\Gamma^{\sharp}})$ *where* $\mathbf{\Phi}$ *is a RIP matrix; then*

$$
\left\|\mathbf{\Phi}_{\Gamma^{\sharp}}^{*}\mathbf{r}\right\|_{\ell^{\infty}(\mathbb{R}^{\Gamma^{\sharp}})} \geq \frac{1}{\sqrt{k}}\left\|\mathbf{\Phi}_{\Gamma^{\sharp}}^{*}\mathbf{r}\right\|_{\ell^{2}(\mathbb{R}^{\Gamma^{\sharp}})}^{2} \geq \frac{(1-\delta_{k})^{1/2}}{\sqrt{k}}\left\|\mathbf{r}\right\|_{\ell^{2}(\mathbb{R}^{\Gamma^{\sharp}})}^{2},
$$

*by the usual norm inequality and Lemma 2.2. Thus taking*

$$
\tilde{\alpha} \leq \frac{(1-\delta_{k})^{1/2}}{\sqrt{k}} \tag{3.3}
$$

*we always have* $\tilde{\mathcal{I}}_{n}(\tilde{\alpha}) \neq \emptyset$ *at each iteration of the algorithms with the Relaxed selection rule. Notice that (3.2) and (3.3) could hold at the same time for some value of* $\tilde{\alpha}$ *only if*

$$
\frac{\delta_{k+1}}{(1-\delta_{k})^{1/2}} \leq \frac{(1-\delta_{k})^{1/2}}{\sqrt{k}},
$$

*which gives the following restriction:* $\sqrt{k} \leq \frac{1-\delta_{k}}{\delta_{k+1}}$.

**Remark 3.3.** *Theorem 3.1 could be considered as an Exact Reconstruction Condition for the* **Relaxed** *Weak OMP (RWOMP) algorithm with RIP matrices and with parameter* $\delta_{k+1}$ *(see Corollary 2.5) as long as (3.3) is satisfied. The probability of success depends exclusively on the probability that a random ensemble verifies RIP. It has been proved that Gaussian, Bernoulli and partial Fourier matrices verify RIP with very high probability (exponential concentration) as long as the number of measurements* $m \geq Ck\log(N/k)$ *for the first two ensembles (see [6],[7]) and* $m \geq Ck\log^{5}(N)$ *for the random Fourier (see [6], [26]).*

## 4. CONVERGENCE. THE SPARSE CASE.

In this section we obtain convergence rates for the SWGP, SWMP and SWOMP algorithms and their **relaxed** counterparts for matrices satisfying RIP. The results here are given in terms of the reduccion of the energy of the residuals $\mathbf{r}^{n} = \mathbf{y} - \mathbf{y}^{\mathbf{n}}$ of the observation $\mathbf{y} = \mathbf{\Phi}\mathbf{x}$ rather in the energy of the residuals of the approximation, $\mathbf{x} - \mathbf{x}^{n}$, as is done in [25] for CoSaMP and in [8] for DThresh.

4.1. **Convergence of GP, SWGP and RWGP.** In [4] the analysis of convergence of the GP is based on an existence theorem (see Theorem 9.10 in [21]) when $\mathbf{\Phi} \in \mathbb{C}^{N \times N_1}$ is a dictionary (this means that $\mathbf{\Phi}$ contains at least a base for $\mathbb{R}^N$ and thus $N_1 \geq N$) that verifies the *Exact Reconstruction Condition* $\mathrm{ERC}_\alpha(\Gamma)$ for *quasi-incoherent* dictionaries (see [30]). The analysis in [31] is done for random *admissible* matrices. The result in the theorem below is obtained for matrices satisfying RIP. We still consider $\mathbf{x}$ a $k-$sparse vector with support on $\Gamma^\sharp$.

**Theorem 4.1.** *Consider the algorithms* GP, SWGP *and* RWGP *and suppose that at iteration $n$ we have $\Gamma^s \subset \Gamma^\sharp$, $s = 1, 2, \ldots, n$. Let $\mathbf{\Phi}$ verifies RIP with $\delta_k$. Then, for all $k$-sparse vectors $\mathbf{x} \in \mathbb{R}^N$ (supp $(\mathbf{x}) = \Gamma^\sharp$),*

$$\|\mathbf{r}^n\|_{\ell^2(\mathbb{R}^m)}^2 \leq C_k \|\mathbf{r}^{n-1}\|_{\ell^2(\mathbb{R}^m)}^2, \tag{4.1}$$

*with $C_k = (1 - \frac{1-\delta_k}{k(1+\delta_k)})^{1/2} < 1$. In the case RWGP suppose $\tilde{\alpha} \leq \frac{(1-\delta_k)^{1/2}}{\sqrt{k}}$ so that $\tilde{\mathcal{I}}_n(\tilde{\alpha}) \neq \emptyset$.*

**Proof.** To shorten notation we will write $\mathbf{d}^n = \mathbf{d}_{\Gamma^n}^n$. We have

$$\begin{aligned}
&\|\mathbf{r}^n\|_{\ell^2(\mathbb{R}^m)}^2 \\
&= \langle \mathbf{r}^{n-1} - a^n \mathbf{\Phi}_{\Gamma^n} \mathbf{d}^n, \mathbf{r}^{n-1} - a^n \mathbf{\Phi}_{\Gamma^n} \mathbf{d}^n \rangle \\
&= \|\mathbf{r}^{n-1}\|_{\ell^2(\mathbb{R}^m)}^2 - a^n \langle \mathbf{r}^{n-1}, \mathbf{\Phi}_{\Gamma^n} \mathbf{d}^n \rangle - a^n \langle \mathbf{\Phi}_{\Gamma^n} \mathbf{d}^n, \mathbf{r}^{n-1} \rangle + \langle a^n \mathbf{\Phi}_{\Gamma^n} \mathbf{d}^n, a^n \mathbf{\Phi}_{\Gamma^n} \mathbf{d}^n \rangle \\
&= \|\mathbf{r}^{n-1}\|_{\ell^2(\mathbb{R}^m)}^2 - 2 \frac{|\langle \mathbf{\Phi}_{\Gamma^n} \mathbf{d}_{\Gamma^n}^n, \mathbf{r}^{n-1} \rangle|^2}{\|\mathbf{\Phi}_{\Gamma^n} \mathbf{d}^n\|_{\ell^2(\mathbb{R}^m)}^2} + \frac{|\langle \mathbf{\Phi}_{\Gamma^n} \mathbf{d}^n, \mathbf{r}^{n-1} \rangle|^2}{\|\mathbf{\Phi}_{\Gamma^n} \mathbf{d}^n\|_{\ell^2(\mathbb{R}^m)}^4} \|\mathbf{\Phi}_{\Gamma^n} \mathbf{d}^n\|_{\ell^2(\mathbb{R}^m)}^2 \\
&= \|\mathbf{r}^{n-1}\|_{\ell^2(\mathbb{R}^m)}^2 - \frac{|\langle \mathbf{r}^{n-1}, \mathbf{\Phi}_{\Gamma^n} \mathbf{d}^n \rangle|^2}{\|\mathbf{\Phi}_{\Gamma^n} \mathbf{d}^n\|_{\ell^2(\mathbb{R}^m)}^2}. \tag{4.2}
\end{aligned}$$

Since $\mathbf{d}^n = \mathbf{\Phi}_{\Gamma^n}^* \mathbf{r}^{n-1}$, the second term above can be bounded below by

$$\begin{aligned}
\frac{|\langle \mathbf{r}^{n-1}, \mathbf{\Phi}_{\Gamma^n} \mathbf{d}^n \rangle|^2}{\|\mathbf{\Phi}_{\Gamma^n} \mathbf{d}^n\|_{\ell^2(\mathbb{R}^m)}^2} &= \frac{|\langle \mathbf{\Phi}_{\Gamma^n}^* \mathbf{r}^{n-1}, \mathbf{d}^n \rangle|^2}{\|\mathbf{\Phi}_{\Gamma^n} \mathbf{d}^n\|_{\ell^2(\mathbb{R}^m)}^2} \\
&\geq \frac{\|\mathbf{\Phi}_{\Gamma^n}^* \mathbf{r}^{n-1}\|_{\ell^2(\mathbb{R}^{\Gamma^n})}^4}{\|\mathbf{\Phi}_{\Gamma^n}\|_{\ell^2(\mathbb{R}^{\Gamma^n}) \to \ell^2(\mathbb{R}^m)}^2 \|\mathbf{\Phi}_{\Gamma^n}^* \mathbf{r}^{n-1}\|_{\ell^2(\mathbb{R}^{\Gamma^n})}^2} \\
&\geq \frac{\|\mathbf{\Phi}_{\Gamma^n}^* \mathbf{r}^{n-1}\|_{\ell^2(\mathbb{R}^{\Gamma^n})}^2}{1 + \delta_k}. \tag{4.3}
\end{aligned}$$

We have

$$\|\mathbf{\Phi}_{\Gamma^n}^* \mathbf{r}^{n-1}\|_{\ell^\infty(\mathbb{R}^{\Gamma^n})} \geq \|\mathbf{\Phi}_{\mathcal{I}_n}^* \mathbf{r}^{n-1}\|_{\ell^\infty(\mathbb{R}^{\mathcal{I}_n})} = \|\mathbf{\Phi}^* \mathbf{r}^{n-1}\|_{\ell^\infty(\mathbb{R}^N)}, \tag{4.4}$$

because $\Gamma^n \supset \mathcal{I}_n$ in the inequality and because the definition of $\mathcal{I}_n$ (in GP and SWGP) in the equality. Since we are supposing that $\Gamma^n \subset \Gamma^\sharp$, using (4.4) yields

$$\|\mathbf{\Phi}_{\Gamma^n}^* \mathbf{r}^{n-1}\|_{\ell^2(\mathbb{R}^{\Gamma^n})} \geq \|\mathbf{\Phi}_{\Gamma^n}^* \mathbf{r}^{n-1}\|_{\ell^\infty(\mathbb{R}^{\Gamma^n})} \geq \|\mathbf{\Phi}^* \mathbf{r}^{n-1}\|_{\ell^\infty(\mathbb{R}^N)} \geq \|\mathbf{\Phi}_{\Gamma^\sharp}^* \mathbf{r}^{n-1}\|_{\ell^\infty(\mathbb{R}^{\Gamma^\sharp})}. \tag{4.5}$$

By (4.5) and usual norm inequality

$$\|\mathbf{\Phi}_{\Gamma^n}^* \mathbf{r}^{n-1}\|_{\ell^2(\mathbb{R}^{\Gamma^n})} \geq \|\mathbf{\Phi}_{\Gamma^\sharp}^* \mathbf{r}^{n-1}\|_{\ell^\infty(\mathbb{R}^{\Gamma^\sharp})} \geq \frac{1}{k} \|\mathbf{\Phi}_{\Gamma^\sharp}^* \mathbf{r}^{n-1}\|_{\ell^2(\mathbb{R}^{\Gamma^\sharp})}.$$

Since $\mathbf{r}^n \in \mathrm{span}\,(\mathbf{\Phi}_{\Gamma^\sharp})$ (assuming conditions in Theorem 2.3 are met and because $\mathbf{r}^n = \mathbf{y} - \mathbf{y}^n$ in GP and SWGP), by Lemma 2.2 we can write

$$\left\|\mathbf{\Phi}_{\Gamma^n}^* \mathbf{r}^{n-1}\right\|_{\ell^2(\mathbb{R}^{\Gamma^n})}^2 \geq \frac{1-\delta_k}{k} \left\|\mathbf{r}^{n-1}\right\|_{\ell^2(\mathbb{R}^m)}^2. \tag{4.6}$$

Substituting (4.6) in (4.3) and (4.2) one gets

$$\left\|\mathbf{r}^n\right\|_{\ell^2(\mathbb{R}^m)}^2 \leq \left(1 - \frac{1-\delta_k}{k(1+\delta_k)}\right) \left\|\mathbf{r}^{n-1}\right\|_{\ell^2(\mathbb{R}^m)}^2,$$

which shows the result for GP and SWGP algorithms.

For RWGP the selection rule

$$\tilde{\mathcal{I}}_n(\tilde{\alpha}) = \left\{i : \left|\langle \phi_i, \mathbf{r}^{n-1}\rangle\right| \geq \tilde{\alpha} \left\|\mathbf{r}^{n-1}\right\|_{\ell^2(\mathbb{R}^m)}\right\}$$

do not allow us to write (4.4). But in this case using Remark 3.2 we can write

$$\begin{aligned}
\left\|\mathbf{\Phi}_{\Gamma^n}^* \mathbf{r}^{n-1}\right\|_{\ell^2(\mathbb{R}^{\Gamma^n})}^2 &\geq \left\|\mathbf{\Phi}_{\Gamma^n}^* \mathbf{r}^{n-1}\right\|_{\ell^\infty(\mathbb{R}^{\Gamma^n})}^2 \geq \left\|\mathbf{\Phi}_{\tilde{\mathcal{I}}_n}^* \mathbf{r}^{n-1}\right\|_{\ell^\infty(\mathbb{R}^{\tilde{\mathcal{I}}_n})}^2 \\
&\geq \frac{1-\delta_k}{k} \left\|\mathbf{r}^{n-1}\right\|_{\ell^2(\mathbb{R}^m)}^2,
\end{aligned}$$

which shows (4.6) for RWGP and therefore the result.  ∎

**Remark 4.2.** *In* [4, *Theorem* 3], *it is proved that* $\|\mathbf{r}^n\|_{\ell^2(\mathbb{R}^m)}^2 \leq c \|\mathbf{r}^{n-1}\|_{\ell^2(\mathbb{R}^m)}^2$ *with* $c = (1 - \frac{\omega}{\|\mathbf{\Phi}\|_{\ell^2}^2})$ *and* $\omega$ *is a positive real number such that* $\|\mathbf{\Phi}\mathbf{x}\|_{\ell^\infty(\mathbb{R}^N)}^2 > \omega \|\mathbf{x}\|_{\ell^2(\mathbb{R}^m)}^2$ *for all* $\mathbf{x} \in \mathbb{R}^m$. *Our Theorem 4.1 gives a value of* $C_k$ *depending on the restricted isometry constant* $\delta_k$ *and the sparseness* $k$. *This value of* $C_k$ *is less than 1, but very close to 1 when* $k$ *increases. It is easy to show that it can never hold* $C_k \leq 1/2$ *when* $k \geq 2$.

Using Theorems 2.3, 3.1 and 4.1 we deduce the following:

**Corollary 4.3.** *Suppose that* $\mathbf{\Phi}$ *satisfies RIP with constants* $\delta_k$ *and* $\delta_{k+1}$.

*i) Suppose* $\frac{\sqrt{k}\delta_{k+1}}{1-\delta_k} < 1$; *then, the GP algorithm satisfies* (4.1).

*ii) Let* $0 < \alpha \leq 1$ *and suppose* $\frac{\sqrt{k}\delta_{k+1}}{1-\delta_k} < \alpha$; *then, the SWGP algorithm with selection rule* $\mathcal{I}_n(\alpha)$ *satisfies* (4.1).

*iii) Suppose that* $\frac{\delta_{k+1}}{1-\delta_k} < \frac{1}{\sqrt{k}}$ *and that* $\tilde{\alpha}$ *is chosen such that* $\frac{\delta_{k+1}}{(1-\delta_k)^{1/2}} < \tilde{\alpha} < \frac{(1-\delta_k)^{1/2}}{\sqrt{k}}$; *then, the RWGP algorithm with selection rule* $\tilde{\mathcal{I}}_n(\tilde{\alpha})$ *satisfies* (4.1).

**Remark 4.4.** *Convergence in algorithms of type Gradient Pursuit (GP, WGP y RWGP) is given in Theorem 4.1 in terms of the convergence of the energy of residual* $\mathbf{r}^n = \mathbf{y} - \mathbf{y}^n$. *If* $\mathbf{\Phi}$ *satisfies RIP, convergence of the residual in terms of estimation* $\mathbf{x} - \mathbf{x}^n$ *can be obtained from the last result. For algorithms of type GP it is not hard to show that* $\mathbf{y}^n = \mathbf{\Phi}\mathbf{x}^n$. *Then,*

$$\begin{aligned}
\|\mathbf{y} - \mathbf{y}^n\|_{\ell^2(\mathbb{R}^m)} &= \|\mathbf{\Phi}\mathbf{x} - \mathbf{\Phi}\mathbf{x}^n\|_{\ell^2(\mathbb{R}^m)} = \|\mathbf{\Phi}_{\Gamma^\sharp}(\mathbf{x} - \mathbf{x}^n)\|_{\ell^2(\mathbb{R}^m)} \\
&\geq (1-\delta_k)^{1/2} \|\mathbf{x} - \mathbf{x}^n\|_{\ell^2(\mathbb{R}^N)}
\end{aligned} \tag{4.7}$$

*because the left-hand side of RIP as long as* $\Gamma^n \subset \Gamma^\sharp$ *(which is verified under the conditions of Theorems 2.3 and 3.1). Analogously, if we also assume* $\Gamma^{n-1} \subset \Gamma^\sharp$ *we*

*have*

$$\begin{aligned}
\left\|\mathbf{y} - \mathbf{y}^{n-1}\right\|_{\ell^2(\mathbb{R}^m)} &= \left\|\mathbf{\Phi}\mathbf{x} - \mathbf{\Phi}\mathbf{x}^{n-1}\right\|_{\ell^2(\mathbb{R}^m)} = \left\|\mathbf{\Phi}_{\Gamma^\sharp}(\mathbf{x} - \mathbf{x}^{n-1})\right\|_{\ell^2(\mathbb{R}^m)} \\
&\leq (1 + \delta_k)^{1/2} \left\|\mathbf{x} - \mathbf{x}^{n-1}\right\|_{\ell^2(\mathbb{R}^N)}
\end{aligned} \tag{4.8}$$

*because the right-hand side of RIP. From (4.7) and (4.8) we deduce*

$$\left\|\mathbf{x} - \mathbf{x}^n\right\|_{\ell^2(\mathbb{R}^N)} \leq C_k \left(\frac{1 + \delta_k}{1 - \delta_k}\right)^{1/2} \left\|\mathbf{x} - \mathbf{x}^{n-1}\right\|_{\ell^2(\mathbb{R}^N)}, \tag{4.9}$$

*where $C_k$ is the constant in Theorem 4.1. Observe that, besides conditions of Corollary 4.3, if we want to assure convergence in $\ell^2(\mathbb{R}^N)$ of $\mathbf{x}^n$ to $\mathbf{x}$ we need*

$$\left(\frac{1 + \delta_k}{1 - \delta_k}\right)^{1/2}\left(1 - \frac{1 - \delta_k}{k(1 + \delta_k)}\right)^{1/2} < 1,$$

*which requires $\delta_k < \frac{1}{2k+1}$.*

If from some iteration we had $\Gamma^n = \Gamma^\sharp = \text{supp}(\mathbf{x})$ reduction of energy of residuals would be faster than the given in (4.1), as the next result shows.

**Theorem 4.5.** *Consider GP, WGP and RWGP algorithms and suppose that at iteration $n_0$ we have $\Gamma^{n_0} = \Gamma^\sharp = sop(\mathbf{x})$. Suppose that $\alpha, \tilde{\alpha} > \frac{\delta_{k+1}}{(1-\delta_k)^{1/2}}$ and $\mathbf{\Phi}$ verifies RIP with parameter $\delta_{k+1}$. Then, for all $n \geq n_0$,*

$$\left\|\mathbf{r}^n\right\|_{\ell^2(\mathbb{R}^m)} \leq D_k \left\|\mathbf{r}^{n-1}\right\|_{\ell^2(\mathbb{R}^m)} \tag{4.10}$$

*with $D_k = \left(1 - \frac{1 - \delta_k}{1 + \delta_k}\right)^{1/2} = \left(\frac{2\delta_k}{1 + \delta_k}\right)^{1/2} < 1$.*

**Proof.** Equality (4.2) and inequality (4.3) in the proof of Theorem 4.1 are still valid in our context. Since $\Gamma^{n_0} = \Gamma^\sharp$ we have $\Gamma^n = \Gamma^\sharp$ for $n \geq n_0$. The fact that $\alpha, \tilde{\alpha} > \frac{\delta_{k+1}}{(1-\delta_k)^{1/2}}$ allow us to conclude $\tilde{\mathcal{I}}_{n_0}, \tilde{\mathcal{I}}_n \subset \Gamma^\sharp$.

Hence, we can replace $\Gamma^n$ by $\Gamma^\sharp$ in (4.3) and since $\mathbf{r}^{n-1} \in \text{span}(\mathbf{\Phi}_{\Gamma^\sharp})$, by Lemma 2.2 we can write

$$\begin{aligned}
\left\|\mathbf{\Phi}_{\Gamma^n}^* \mathbf{r}^{n-1}\right\|_{\ell^2(\mathbb{R}^m)}^2 &= \left\|\mathbf{\Phi}_{\Gamma^\sharp}^* \mathbf{r}^{n-1}\right\|_{\ell^2(\mathbb{R}^m)}^2 \\
&\geq (1 - \delta_k) \left\|\mathbf{r}^{n-1}\right\|_{\ell^2(\mathbb{R}^m)}^2.
\end{aligned}$$

Substituting this inequality in (4.2) and (4.3) we obtain the result for the three algorithms.                                                                                                   ∎

**Remark 4.6.** *Constant $D_k$ of Theorem 4.5 can be made as close to $0$ as desired taking $\delta_k$ small enough. In particular, it is enough to take $\delta_k \leq 1/7 \approx 0.143$ to reduce the energy of the residuals by half in just one iteration, if conditions of Theorem 4.1 are satisfied.*

*Reasoning as in Remark 4.4 we have*

$$\left\|\mathbf{x} - \mathbf{x}^n\right\|_{\ell^2(\mathbb{R}^N)} \leq \left(\frac{1 + \delta_k}{1 - \delta_k}\right)^{1/2}\left(1 - \frac{1 - \delta_k}{1 + \delta_k}\right)^{1/2} \left\|\mathbf{x} - \mathbf{x}^{n-1}\right\|_{\ell^2(\mathbb{R}^N)}$$

*for which it is enough to take $\delta_k < 1/3$ to assure convergence of $\mathbf{x}^n$ to $\mathbf{x}$ in $\ell^2(\mathbb{R}^N)$ once we have $\Gamma^{n_0} = \Gamma^\sharp$.*

4.2. **Convergence of SWMP and RWMP.** The results on convergence in this section are similar to those obtained in [16] for quasi-incoherent dictionaries.

**Theorem 4.7.** *Consider the algorithms SWMP and RWMP. Suppose that conditions of Theorems 2.3 (for SWMP) and 3.1 are verified so that $\Gamma^s \subset \Gamma^\sharp$, $s = 1, 2, \ldots$. Then, for every $k$-sparse vector $\mathbf{x} \in \mathbb{R}^N$ with supp $(\mathbf{x}) = \Gamma^\sharp$,*

$$\|\mathbf{r}^n\|^2_{\ell^2(\mathbb{R}^m)} \leq C'_k \|\mathbf{r}^{n-1}\|^2_{\ell^2(\mathbb{R}^m)}, \tag{4.11}$$

*with $C'_k = (1 - \frac{(1-\delta_k)^2}{k})^{1/2} < 1$.*

**Proof.** We have for SWMP as well as for RWMP algorithms that

$$
\begin{aligned}
\|\mathbf{r}^n\|^2_{\ell^2(\mathbb{R}^m)} &= \langle \mathbf{r}^{n-1} - \boldsymbol{\Phi}_{\mathcal{I}_n} \boldsymbol{\Phi}^*_{\mathcal{I}_n} \mathbf{r}^{n-1}, \mathbf{r}^{n-1} - \boldsymbol{\Phi}_{\mathcal{I}_n} \boldsymbol{\Phi}^*_{\mathcal{I}_n} \mathbf{r}^{n-1} \rangle \\
&= \|\mathbf{r}^{n-1}\|^2_{\ell^2(\mathbb{R}^m)} - \langle \mathbf{r}^{n-1}, \boldsymbol{\Phi}_{\mathcal{I}_n} \boldsymbol{\Phi}^*_{\mathcal{I}_n} \mathbf{r}^{n-1} \rangle \\
&\quad - \langle \boldsymbol{\Phi}_{\mathcal{I}_n} \boldsymbol{\Phi}^*_{\mathcal{I}_n} \mathbf{r}^{n-1}, \mathbf{r}^{n-1} \rangle + \|\boldsymbol{\Phi}_{\mathcal{I}_n} \boldsymbol{\Phi}^*_{\mathcal{I}_n} \mathbf{r}^{n-1}\|^2_{\ell^2(\mathbb{R}^m)} \\
&= \|\mathbf{r}^{n-1}\|^2_{\ell^2(\mathbb{R}^m)} - 2\|\boldsymbol{\Phi}^*_{\mathcal{I}_n} \mathbf{r}^{n-1}\|^2_{\ell^2(\mathbb{R}^{\mathcal{I}_n})} + \|\boldsymbol{\Phi}_{\mathcal{I}_n} \boldsymbol{\Phi}^*_{\mathcal{I}_n} \mathbf{r}^{n-1}\|^2_{\ell^2(\mathbb{R}^m)} .(4.12)
\end{aligned}
$$

By (2.1) we have $\|\boldsymbol{\Phi}_{\mathcal{I}_n} \boldsymbol{\Phi}^*_{\mathcal{I}_n} \mathbf{r}^{n-1}\|^2_{\ell^2(\mathbb{R}^m)} \leq (1 + \delta_k) \|\boldsymbol{\Phi}^*_{\mathcal{I}_n} \mathbf{r}^{n-1}\|^2_{\ell^2(\mathbb{R}^{\mathcal{I}_n})}$ since $\mathcal{I}_n \subset \Gamma^\sharp$ and $|\Gamma^\sharp| = k$. Hence, from (4.12) we have

$$
\begin{aligned}
\|\mathbf{r}^n\|^2_{\ell^2(\mathbb{R}^m)} &\leq \|\mathbf{r}^{n-1}\|^2_{\ell^2(\mathbb{R}^m)} - 2\|\boldsymbol{\Phi}^*_{\mathcal{I}_n} \mathbf{r}^{n-1}\|^2_{\ell^2(\mathbb{R}^{\mathcal{I}_n})} + (1 + \delta_k)\|\boldsymbol{\Phi}^*_{\mathcal{I}_n} \mathbf{r}^{n-1}\|^2_{\ell^2(\mathbb{R}^{\mathcal{I}_n})} \\
&= \|\mathbf{r}^{n-1}\|^2_{\ell^2(\mathbb{R}^m)} - (1 - \delta_k)\|\boldsymbol{\Phi}^*_{\mathcal{I}_n} \mathbf{r}^{n-1}\|^2_{\ell^2(\mathbb{R}^{\mathcal{I}_n})} . \tag{4.13}
\end{aligned}
$$

For SWMP as well as RWMP we have

$$\|\boldsymbol{\Phi}^*_{\mathcal{I}_n} \mathbf{r}^{n-1}\|^2_{\ell^2(\mathbb{R}^{\mathcal{I}_n})} \geq \frac{1 - \delta_k}{k} \|\mathbf{r}^{n-1}\|^2_{\ell^2(\mathbb{R}^m)} . \tag{4.14}$$

For SWMP the proof is as in (4.6) from Theorem 4.1 and it is not hard to prove it also for RWMP.

Substituting (4.14) in (4.13) we have

$$\|\mathbf{r}^n\|^2_{\ell^2(\mathbb{R}^m)} \leq (1 - \frac{(1 - \delta_k)^2}{k}) \|\mathbf{r}^{n-1}\|^2_{\ell^2(\mathbb{R}^m)}$$

which is the desired result. ■

**Corollary 4.8.** *Suppose $\boldsymbol{\Phi}$ satisfies RIP with constants $\delta_k$ and $\delta_{k+1}$.*

   a) *Let $0 < \alpha \leq 1$ and suppose that $\frac{\sqrt{k}\delta_{k+1}}{1-\delta_k} < \alpha$; then, SWMP with selection rule $\mathcal{I}_n(\alpha)$ satisfies (4.11).*
   b) *Suppose that $\frac{\delta_{k+1}}{1-\delta_k} < \frac{1}{\sqrt{k}}$ and $\tilde{\alpha}$ is chosen so that*

$$\frac{\delta_{k+1}}{(1-\delta_k)^{1/2}} < \tilde{\alpha} < \frac{(1-\delta_k)^2}{\sqrt{k}};$$

   *then, RWMP with selection rule $\tilde{\mathcal{I}}_n(\tilde{\alpha})$ satisfies (4.11).*

**Remark 4.9.** *Constant $C'_k$ of Theorem 4.7 is a numer less than 1, so there is always a decreasing in the residual energy. However, it is close to 1 when $k$ increases. It is easy to show that we can never have $C'_k \leq 1/2$ when $k \geq 2$.*

**Remark 4.10.** *As in Remark 4.4 the decreasing of the residual energy given in (4.11) for SWMP y RWMP can be translated to convergence of the estimation. Thus,*

$$\|\mathbf{x} - \mathbf{x}^n\|_{\ell^2(\mathbb{R}^N)} \leq (\frac{1 + \delta_k}{1 - \delta_k})^{1/2} C_k' \|\mathbf{x} - \mathbf{x}^{n-1}\|_{\ell^2(\mathbb{R}^N)}.$$

*For $(\frac{1+\delta_k}{1-\delta_k})^{1/2}(1 - \frac{(1-\delta_k)^2}{k})^{1/2}$ to be less than 1 we must have*

$$1 - \frac{(1 - \delta_k)^2}{k} < \frac{1 - \delta_k}{1 + \delta_k} \Leftrightarrow \frac{2\delta_k}{1 + \delta_k} < \frac{(1 - \delta_k)^2}{k}.$$

*Since $(1 + \delta_k)(1 - \delta_k)^2 = (1 - \delta_k)(1 - \delta_k^2) = 1 - \delta_k - \delta_k^2 + \delta_k^3$ last inequality is equivalent to*

$$2k\delta_k < 1 - \delta_k - \delta_k^2 + \delta_k^3 \Leftrightarrow \delta_k^3 - \delta_k^2 - (1 + 2k)\delta_k + 1 > 0.$$

*If $\delta_k < \frac{1}{2k+2}$, we have*

$$1 > \delta_k + (2k + 1)\delta_k > \delta_k^2 + (2k + 1)\delta_k > \delta_k^2 + (2k + 1)\delta_k - \delta_k^3,$$

*from which*

$$\delta_k < \frac{1}{2k + 2}$$

*is enough to have convergence of approximations in WMP y RWMP.*

### 4.3. **Convergencia de WOMP y RWOMP.** 
In WOMP as well as in RWOMP the residual $\mathbf{r}^n$ is the vector that carries out the distance from $\mathbf{y}$ to the subspace $V_n = \{\mathbf{\Phi}_{\Gamma^n}\mathbf{x} : \mathbf{x} \in \mathbb{R}^n \text{ with } \text{supp}(\mathbf{x}) \subset \Gamma^n\}$. Since $\Gamma^{n-1} \subset \Gamma^n$, it is clear that

$$\|\mathbf{r}^n\|_{\ell^2(\mathbb{R}^m)} \leq \|\mathbf{r}^{n-1}\|_{\ell^2(\mathbb{R}^m)}$$

is always accomplished. A strict inequality can be obtained observing that the residual in WOMP or RWOMP, temporarily denoted $\mathbf{r}^n_{OMP}$, has an energy no larger than that for WGP, RWGP, WMP or RWMP, temporarily denoted $\mathbf{r}^n_{GP}$ and $\mathbf{r}^n_{MP}$, since $\mathbf{y}^n_{GP}$ as well as $\mathbf{y}^n_{MP}$ are elements from $V_n$.

Therefore, if conditions of Theorems 2.3 are verified (for WGP) or 3.1 (for RWGP) we can apply Theorem 4.1 to get

$$\begin{aligned}
\|\mathbf{r}^n_{OMP}\|_{\ell^2(\mathbb{R}^m)} &\leq \|\mathbf{r}^n_{GP}\|_{\ell^2(\mathbb{R}^m)} \leq C_k \|\mathbf{r}^{n-1}_{GP}\|_{\ell^2(\mathbb{R}^m)} \\
&\leq C_k^n \|\mathbf{r}^0_{GP}\|_{\ell^2(\mathbb{R}^m)} \leq C_k^n \|\mathbf{y}\|_{\ell^2(\mathbb{R}^m)},
\end{aligned} \tag{4.15}$$

with $C_k < 1$, the constant of Theorem 4.1.

Analogously, but using Theorem 4.7 we have

$$\|\mathbf{r}^n_{OMP}\|_{\ell^2(\mathbb{R}^m)} \leq C_k'^n \|\mathbf{y}\|_{\ell^2(\mathbb{R}^m)}, \tag{4.16}$$

with $C_k' < 1$ the constant of Theorem 4.7. Among the constants $C_k$ and $C_k'$ the relation is $C_k < C_k'$ since

$$\begin{aligned}
C_k < C_k' &\Leftrightarrow \frac{1 - \delta_k}{k(1 + \delta_k)} > \frac{(1 - \delta_k)^2}{k} \Leftrightarrow \frac{1}{1 + \delta_k} > 1 - \delta_k \\
&\Leftrightarrow 1 > 1 - \delta_k^2,
\end{aligned}$$

and the last inequality is true because $0 < \delta_k < 1$. Therefore, (4.15) gives faster convergence than (4.16) and it proves that GP algorithms converge faster than MP algorithms.

Since $\mathbf{r}_{OMP}^n$ is a vector perpendicular to $V_n$, the constant in (4.15) could be improved, at least in principle. Since

$$\mathbf{r}_{OMP}^n = \mathbf{y} - \mathbf{y}^n = \mathbf{y} - \mathbf{\Phi}_{\Gamma^n} \mathbf{\Phi}_{\Gamma^n}^\dagger \mathbf{y},$$

where $\mathbf{\Phi}_{\Gamma^n}^\dagger$ is the pseudo-inverse of $\mathbf{\Phi}_{\Gamma^n}$, we have to study

$$
\begin{aligned}
\|\mathbf{r}_{OMP}^n\|_{\ell^2(\mathbb{R}^m)}^2 \;=\;& \|\mathbf{y}\|_{\ell^2(\mathbb{R}^m)}^2 - \langle \mathbf{y}, \mathbf{\Phi}_{\Gamma^n} \mathbf{\Phi}_{\Gamma^n}^\dagger \mathbf{y} \rangle - \langle \mathbf{\Phi}_{\Gamma^n} \mathbf{\Phi}_{\Gamma^n}^\dagger \mathbf{y}, \mathbf{y} \rangle \\
&+ \; \left\| \mathbf{\Phi}_{\Gamma^n} \mathbf{\Phi}_{\Gamma^n}^\dagger \mathbf{y} \right\|_{\ell^2(\mathbb{R}^m)}^2 .
\end{aligned}
\tag{4.17}
$$

We are not yet able to find a bound for (4.17) of the form

$$\|\mathbf{r}_{OMP}^n\|_{\ell^2(\mathbb{R}^m)} \leq B_k^n \|\mathbf{y}\|_{\ell^2(\mathbb{R}^m)} \tag{4.18}$$

with $B_K < C_k$. In the case that conditions of Theorems 2.3 and 3.1 are satisfied (for example, if $\delta_{k+1} < \alpha \frac{1-\delta_k}{\sqrt{k}}$ for WOMP and $\frac{\delta_{k+1}}{(1-\delta_k)^{1/2}} < \tilde{\alpha} < \frac{(1-\delta_k)^{1/2}}{\sqrt{k}}$ for RWOMP) inequalities (4.15) and (4.16) are trivial if $n \geq k$ since algorithms of type OMP identify the support of a $k$-sparse signal $\mathbf{x}$ in as much $k$ iterations, and then $\mathbf{r}^n = 0$.

Therefore, it is only interesting to find bounds of the form (4.18) from expression (4.17) if it is satisfied with values of $\delta_{k+1}$ and $\delta_k$ less restrictive than those in Theorems 2.3 and 3.1, for which algorithms do not identify the support of $\mathbf{x}$.

## 5. Behavior of the selection rules for some random matrices

The reader can find in [13] a way to construct matrices that satisfy RIP deterministically. These matrices are of order $m \times N$ with $m = p^2$ ($p$ a prime number) and $N = p^{r+1}$, $0 < r < p$, and satisfy RIP with $k < \frac{p}{r} + 1$ and $\delta_k = (k-1)r/p$. Therefore, $m = p^2 > (k-1)^2 r^2 > (k-1)^2$; which gives a value of $m$ much larger than the necessary to recover a $k$-sparse signal with the $\ell^1$ minimization which is $m \geq Ck \log(N/k)$.

It is known (see [3]) that there exist random matrices that satisfy RIP with parameter $\delta_k$ for any $m \geq Ck \log(N/k)$ with probability greater than $1 - 2e^{-c'm}$. Among those are the matrices that satisfy an inequality known as **concentration of measure**, namely,

$$\mathbb{P}\left\{ \left| \|\mathbf{\Phi}(\omega)\mathbf{x}\|_{\ell^2(\mathbb{R}^m)} - \|\mathbf{x}\|_{\ell^2(\mathbb{R}^N)} \right| \geq \varepsilon \|\mathbf{x}\|_{\ell^2(\mathbb{R}^N)}^2 \right\} \leq 2e^{mc_0(\varepsilon)}, \quad 0 < \varepsilon < 1, \tag{5.1}$$

where the probability is taken over all random matrices $\mathbf{\Phi}(\omega)$ of order $m \times N$ and $c_0(\varepsilon) > 0$ is a constant that depends only on $\varepsilon$.

An example of such matrices are those $\mathbf{\Phi} = (\phi_{i,j})$ such that $\phi_{i,j}$ is an independent Gaussian random variable $N(0, 1/\sqrt{m})$, that is, with mean 0 and standard deviation $1/\sqrt{m}$. In this case we have $c_0(\varepsilon) = \frac{\varepsilon^2}{4} - \frac{\varepsilon^3}{6}$ (see [10]).

Another example are the matrices whose entries are independent Bernoulli random variables with values $\{-1/\sqrt{m}, 1/\sqrt{m}\}$, with probability $1/2$ each. In this case we also have $c_0(\varepsilon) = \frac{\varepsilon^2}{4} - \frac{\varepsilon^3}{6}$ (see [1]).

In this section we will study the behavior of the selection rules and algorithms given in 2.1 and 3 with respect to the random matrices $\mathbf{\Phi}$ as just described above. The aim is to prove directly that this kind of matrices select elements of the support of a $k$-sparse signal with high probability, following the reasoning given in [31] for *Orhogonal Matching Pursuit* (OMP).

We start with a result on random processes (see [31] and the references cited in this article), whose proof is given for completeness:

**Lemma 5.1.**     a) *Let $\mathbf{z}$ be a vector with dimension $m$ whose components are Gaussian r.v. $N(0, 1/\sqrt{m})$ i.i.d. Independently a vector $\mathbf{u}$ unitary in $\ell^2(\mathbb{R}^m)$ is chosen. We have, for $0 < \varepsilon \le 1$,*

$$\mathbb{P}\left\{|\langle \mathbf{u}, \mathbf{z}\rangle| \ge \varepsilon\right\} \le e^{-\frac{\varepsilon^2}{2}m}. \tag{5.2}$$

b) *Let $\mathbf{w}$ a vector of dimension $m$ whose entries are symmetric Bernoulli r.v. $\{-1/\sqrt{m}, 1/\sqrt{m}\}$ i.i.d. Independently a vector $\mathbf{u}$ unitary in $\ell^2(\mathbb{R}^m)$ is chosen. We have, for $0 < \varepsilon \le 1$,*

$$\tag{5.3}$$

$$\mathbb{P}\left\{|\langle \mathbf{u}, \mathbf{w}\rangle| \ge \varepsilon\right\} \le 2e^{-\frac{\varepsilon^2}{2}m}.$$

**Proof**. The inner product $\langle \mathbf{u}, \mathbf{z}\rangle = \sum_{i=1}^m u_i z_i$ is a Gaussian r.v. (the sum of Gaussian r.v. is a Gaussian r.v.) with mean $\mathbb{E}\{\langle \mathbf{u}, \mathbf{z}\rangle\} = \sum_{i=1}^m u_i \mathbb{E}\{z_i\} = 0$ and standard deviation

$$(\mathbb{E}\{|\langle \mathbf{u}, \mathbf{z}\rangle|^2\})^{1/2} = (\sum_{i=1}^m u_i^2 \mathbb{E}\{z_i^2\} + \sum_{i=1}^m \sum_{j \neq i} u_i u_j \mathbb{E}\{z_i z_j\})^{1/2} = \frac{1}{\sqrt{m}},$$

since the r.v. $z_i$ are independents and that $\mathbf{u}$, unitary, is independent from $\mathbf{z}$. Hence, since $|\langle \mathbf{u}, \mathbf{z}\rangle|$ is symmetric

$$\mathbb{P}\left\{|\langle \mathbf{u}, \mathbf{z}\rangle| \ge \varepsilon\right\} = 2\mathbb{P}\left\{\langle \mathbf{u}, \mathbf{z}\rangle \ge \varepsilon\right\} = 2\frac{\sqrt{m}}{\sqrt{2\pi}} \int_\varepsilon^\infty e^{-\frac{1}{2}mx^2} dx = \sqrt{\frac{2}{\pi}} \int_{\varepsilon\sqrt{m}}^\infty e^{-\frac{y^2}{2}} dy \tag{5.4}$$

making the change of variable $\sqrt{m}x = y$. Let

$$I = \int_{\varepsilon\sqrt{m}}^\infty e^{-\frac{y^2}{2}} dy.$$

We have

$$
\begin{aligned}
I^2 &= \left(\int_{\varepsilon\sqrt{m}}^\infty e^{-\frac{y^2}{2}} dy\right)\left(\int_{\varepsilon\sqrt{m}}^\infty e^{-\frac{x^2}{2}} dx\right) = \int_{\varepsilon\sqrt{m}}^\infty \left(\int_{\varepsilon\sqrt{m}}^\infty e^{-\frac{y^2+x^2}{2}} dy\right) dx \\
&\le \int_{R_{\varepsilon,m}} \int e^{-\frac{y^2+x^2}{2}} dx dy,
\end{aligned}
$$

where

$$R_{\varepsilon,m} = \{(x,y) = w \in \mathbb{R}^2 : x, y \ge 0, \|w\|_2 \ge \varepsilon\sqrt{2m}\}.$$

Passing to polar coordinates

$$I^2 \le \int_0^{\pi/2} \int_{\varepsilon\sqrt{2m}}^\infty e^{-\frac{r^2}{2}} r\, dr\, d\theta = \frac{\pi}{2}\left[-e^{-\frac{r^2}{2}}\right]_{\varepsilon\sqrt{2m}}^\infty = \frac{\pi}{2}e^{-\frac{\varepsilon^2 m}{2}}.$$

Therefore, $I \le \sqrt{\frac{\pi}{2}}e^{-\frac{\varepsilon^2 m}{2}}$, which substituting in (5.4) yields (5.2).

b) In this case the Hoeffding inequality can be applied (see Theorem 4 in [19]) to get

$$
\mathbb{P}\left\{|\langle \mathbf{u}, \mathbf{w}\rangle \geq \varepsilon|\right\} = 2\mathbb{P}\left\{\sum_{i=1}^{m} u_i w_i \geq \varepsilon\right\} = 2\mathbb{P}\left\{\sum_{i=1}^{m} u_i w_i - \mathbb{E}\left\{\sum_{i=1}^{m} u_i w_i\right\} \geq \varepsilon\right\}
$$

$$
= 2e^{-2\varepsilon^2 / \sum_{i=1}^{m}(u_i/\sqrt{m}-(-u_i/\sqrt{m}))^2} = 2e^{-\frac{\varepsilon^2 m}{2}},
$$

since $\mathbb{E}\left\{\sum_{i=1}^{m} u_i w_i\right\} = \sum_{i=1}^{m} u_i \mathbb{E}\left\{w_i\right\} = 0$ because $\mathbf{u}$ is independent from the r.v. $w_i$ and $\|\mathbf{u}\|_2 = 1$. ∎

In this section we will consider random matrices $\boldsymbol{\Phi}(\omega) \in \mathbb{R}^{m \times N}$ that satisfy next conditions, similar to those conditions for admissible matrices in [31]:

(M1) The columns of $\boldsymbol{\Phi}(\omega)$ are statistically independents.

(M2) For each column $\phi_j(\omega)$, $j = 1, \ldots, N$, of $\boldsymbol{\Phi}(\omega)$ we have $\mathbb{E}\left\{\|\phi_j(\omega)\|_{\ell^2(\mathbb{R}^m)}^2\right\} = 1$.

(M3) Let $\mathbf{u} \in \mathbb{R}^m$ a vector with $\|\mathbf{u}\|_{\ell^2(\mathbb{R}^m)} \leq 1$. If $\phi(\omega)$ is a column of $\boldsymbol{\Phi}(\omega)$ independent from $\mathbf{u}$,

$$
\mathbb{P}\left\{|\langle \phi(\omega), \mathbf{u}\rangle| \geq \varepsilon\right\} \leq q_1 e^{-c_1 \varepsilon^2 m},
$$

with $q_1, c_1$ constants, $q_1 \geq 1$.

(M4) For every set $\Gamma \subset \{1, \ldots, N\}$ with $|\Gamma| \leq k < N$ and for every $\mathbf{r} \in \text{span}\,(\boldsymbol{\Phi}_\Gamma(\omega))$ we have

$$
\mathbb{P}\left\{\|\boldsymbol{\Phi}_\Gamma^*(\omega)\mathbf{r}\|_{\ell^2(\mathbb{R}^\Gamma)} \geq \frac{1}{2}\|r\|_{\ell^2(\mathbb{R}^m)}\right\} \geq 1 - q_2 D^k e^{-c_2 m},
$$

with $q_2, D$ and $c_2$ constants, $q_2, D > 1$.

These properties are satisfied by the Gaussian and Bernoulli random matrices as described at the beginning of this section. Property (M3) es the content of Lemma 5.1 (observe that if $\|\mathbf{u}\|_2 \leq 1$ Lemma 5.1 is also verified since if we substitute $\mathbf{u}$ for a unitary vector in its direction then the probability increases) and condition (M4) is proved in Lemma 5.1 in [3], where a proof is given from a concentration of measure inequality as that in (5.1).

5.1. **Probabilistic support identification for relaxed algorithms.** In this section we show that matrices satisfying (M1), (M2), (M3) and (M4) allow to identify indices in the support of a $k$-sparse signal with high probability. The result follows the arguments in [31], with necessary modifications to fit the relaxed selection rule $\tilde{I}(\tilde{\alpha})$ given in (3.1).

**Theorem 5.2.** *Let $\boldsymbol{\Phi}(\omega) \in \mathbb{R}^{m \times N}$ be a random matrix satisfyiing (M1), (M2), (M3) and (M4). Let $\mathbf{x} \in \mathbb{R}^N$ with $\text{supp}\,(\mathbf{x}) = \Gamma^\sharp$ y $|\Gamma^\sharp| \leq k < N$. Sea $\mathbf{y} = \boldsymbol{\Phi}\mathbf{x}$. Suppose that $\tilde{\alpha} \leq 1/2\sqrt{k}$ and given $l$, $1 \leq l < N$,*

$$
m \geq \max\{\frac{1}{c_1 \tilde{\alpha}^2} \ln q_3 l(N-k), \frac{2k}{c_2} \ln D\}, \tag{5.5}
$$

*with $q_3 = q_1 + q_2$. Algorithms RWMP, RWOMP y RWGP, with selection rule $\tilde{I}_n(\tilde{\alpha})$ given by(3.1), identify elements from $\Gamma^\sharp$ in the first $l$ iterations with probability greater or equal to*

$$
1 - q_3 l(N-k)e^{-c_1 \tilde{\alpha}^2 m}.
$$

**Proof**. For any of the given algorithms with selection rule $\tilde{I}_n(\tilde{\alpha})$, let $E_n$ be the set of matrices that identify elements from $\Gamma^\sharp = \text{sop}\,(\mathbf{x})$ at iteration $n$, $n = 1, 2, 3, \ldots$.

We start bounding $\mathbb{P}\{E_1\}$, that is, the probability that a matrix identify elements from $\Gamma^\sharp$ in the first iteration. For a random matrix $\mathbf{\Phi}(\omega)$ to belong to $E_1$ it is enough that the next two conditions are satisfied:

($A_1$) That $\tilde{I}_1(\tilde{\alpha}) \neq \emptyset$.
($B_1$) For $\mathbf{u}_0 = \mathbf{r}^0 / \|\mathbf{r}^0\|_{\ell^2(\mathbb{R}^m)}$ (we remind that $\mathbf{r}^0 = \mathbf{y}$) it should be true that

$$\frac{\left\|\mathbf{\Phi}^*_{\Gamma^{\sharp c}}(\omega)\mathbf{r}^0\right\|_{\ell^\infty(\mathbb{R}^{\Gamma^{\sharp c}})}}{\|\mathbf{r}^0\|_{\ell^2(\mathbb{R}^m)}} = \left\|\mathbf{\Phi}^*_{\Gamma^{\sharp c}}(\omega)\mathbf{u}_0\right\|_{\ell^\infty(\mathbb{R}^{\Gamma^{\sharp c}})} = \max_{j \in \Gamma^{\sharp c}} |\langle \mathbf{u}_0, \phi_j(\omega)\rangle| < \tilde{\alpha}.$$

With abuse of notation, we call $A_1$ and $B_1$ the sets of matrices that satisfy ($A_1$) and ($B_1$) respectively. We have

$$\mathbb{P}\{E_1\} \geq \mathbb{P}\{A_1 \cap B_1\} = \mathbb{P}\{A_1|B_1\}\,\mathbb{P}\{B_1\}. \tag{5.6}$$

To estimate $\mathbb{P}\{B_1\}$ we write

$$\begin{aligned}
\mathbb{P}\{B_1\} &= \mathbb{P}\left\{\max_{j \in \Gamma^{\sharp c}} |\langle \mathbf{u}_0, \phi_j(\omega)\rangle| < \tilde{\alpha}\right\} = \mathbb{P}\left\{\cap_{j \in \Gamma^{\sharp c}}\{|\langle \mathbf{u}_0, \phi_j(\omega)\rangle| < \tilde{\alpha}\}\right\} \\
&= \prod_{j \in \Gamma^{\sharp c}} \mathbb{P}\{|\langle \mathbf{u}_0, \phi_j(\omega)\rangle| < \tilde{\alpha}\}
\end{aligned}$$

due to the independence of the columns $\phi_j$ from $\mathbf{\Phi}(\omega)$ expressed in (M1). Since $\mathbf{u}_0 = \mathbf{y}/\|\mathbf{y}\|_2$ is unitary and $\mathbf{y} = \mathbf{\Phi}\mathbf{x} = \mathbf{\Phi}_{\Gamma^\sharp}\mathbf{x}_{\Gamma^\sharp}$ only depends on the columns $\phi_j$ with $j \in \Gamma^\sharp$, $\mathbf{u}_0$ is independent from the columns $\phi_j$ with $j \in \Gamma^{\sharp c}$, and we can use (M3) to get

$$\mathbb{P}\{B_1\} \geq \prod_{j \in \Gamma^{\sharp c}} (1 - q_1 e^{-c_1\tilde{\alpha}^2 m}) = (1 - q_1 e^{-c_1\tilde{\alpha}^2 m})^{N-k} \geq 1 - q_1(N-k)e^{-c_1\tilde{\alpha}^2 m}, \tag{5.7}$$

since $(1-x)^n \geq 1 - nx$ if $n \geq 1$ and $x \leq 1$.

On the other hand, conditioned to $B_1$ we have that $\mathbf{\Phi} \in A_1 \Leftrightarrow \|\mathbf{\Phi}_{\Gamma^\sharp}\mathbf{y}\|_\infty \geq \tilde{\alpha}\|\mathbf{y}\|_2$. Since $\tilde{\alpha} \leq \frac{1}{2\sqrt{k}}$,

$$\begin{aligned}
\mathbb{P}\{A_1|B_1\} &= \mathbb{P}\left\{\|\mathbf{\Phi}^*_{\Gamma^\sharp}(\omega)\mathbf{y}\|_{\ell^\infty(\mathbb{R}^{\Gamma^\sharp})} \geq \tilde{\alpha}\|\mathbf{y}\|_{\ell^2(\mathbb{R}^m)}\right\} \\
&\geq \mathbb{P}\left\{\|\mathbf{\Phi}^*_{\Gamma^\sharp}(\omega)\mathbf{y}\|_{\ell^2(\mathbb{R}^{\Gamma^\sharp})} \geq \tilde{\alpha}\sqrt{k}\|\mathbf{y}\|_{\ell^2(\mathbb{R}^m)}\right\} \\
&\geq \mathbb{P}\left\{\|\mathbf{\Phi}^*_{\Gamma^\sharp}(\omega)\mathbf{y}\|_{\ell^2(\mathbb{R}^{\Gamma^\sharp})} \geq \frac{1}{2}\|\mathbf{y}\|_{\ell^2(\mathbb{R}^m)}\right\}.
\end{aligned}$$

Using now (M4) we deduce (observe that $\mathbf{y} = \mathbf{\Phi}\mathbf{x}$, therefore $\mathbf{y} \in \text{span}\,(\mathbf{\Phi}_{\Gamma^\sharp})$):

$$\mathbb{P}\{A_1|B_1\} \geq 1 - q_2 D^k e^{-c_2 m} \geq 1 - q_2 e^{-c_2 m/2} \tag{5.8}$$

taking $-c_2 m + k\ln D \leq -c_2 m/2$, that is, $m \geq \frac{2k}{c_2}\ln D$.

Substituting (5.7) and (5.8) in (5.6) we get

$$\begin{aligned}
\mathbb{P}\{E_1\} &\geq (1 - q_1(N_k)e^{-c_1\tilde{\alpha}^2 m})(1 - q_2 e^{-c_2 m/2}) \\
&\geq 1 - q_1(N-k)e^{-c_1\tilde{\alpha}^2 m} - q_2 e^{-c_2 m/2} \\
&\geq 1 - q_3(N-k)e^{-c_1\tilde{\alpha}^2 m}, \tag{5.9}
\end{aligned}$$

since $\tilde{\alpha} < 1/2$, with $q_3 = q_1 + q_2$.

We have proved the result for $l = 1$. Suppose the result is true until iteration $l - 1$, this is,

$$\mathbb{P}\left\{\cap_{n=1}^{l-1} E_n\right\} \geq 1 - q_3(l-1)(N-k)e^{-c_1\tilde{\alpha}^2 m}. \tag{5.10}$$

Until iteration $l$ we have

$$\mathbb{P}\left\{\cap_{n=1}^{l} E_n\right\} = \mathbb{P}\left\{E_l \mid \cap_{n=1}^{l-1} E_n\right\} \mathbb{P}\left\{\cap_{n=1}^{l-1} E_n\right\}. \tag{5.11}$$

For a random matrix $\mathbf{\Phi}(\omega)$ to belong to $E_l$ it is enough that next two conditions are satisfied:

($A_l$) That $\tilde{I}_l(\tilde{\alpha}) \neq \emptyset$.
($B_l$) For $\mathbf{u}_{l-1} = \mathbf{r}^{l-1} / \left\|\mathbf{r}^{l-1}\right\|_2$ it should be true that

$$\frac{\left\|\mathbf{\Phi}_{\Gamma^{\sharp c}}^*(\omega)\mathbf{r}^{l-1}\right\|_{\ell^\infty(\mathbb{R}^{\Gamma^{\sharp c}})}}{\left\|\mathbf{r}^{l-1}\right\|_{\ell^2(\mathbb{R}^m)}} = \left\|\mathbf{\Phi}_{\Gamma^{\sharp c}}^*(\omega)\mathbf{u}_{l-1}\right\|_{\ell^\infty(\mathbb{R}^{\Gamma^{\sharp c}})} = \max_{j \in \Gamma^{\sharp c}} |\langle \mathbf{u}_{l-1}, \phi_j(\omega)\rangle| < \tilde{\alpha}.$$

With abuse of notation, we call $A_l$ and $B_l$ the sets of matrices that satisfy ($A_l$) and ($B_l$) respectively. We have

$$\mathbb{P}\left\{E_l \mid \cap_{n=1}^{l-1} E_n\right\} \geq \mathbb{P}\left\{A_l \mid B_l \cap (\cap_{n=1}^{l-1} E_n)\right\} \mathbb{P}\left\{B_l \mid \cap_{n=1}^{l-1} E_n\right\}. \tag{5.12}$$

With the same reasoning that leads to (5.7) we get

$$\mathbb{P}\left\{B_l \mid \cap_{n=1}^{l-1} E_n\right\} \geq 1 - q_1(N-k)e^{-c_1\tilde{\alpha}^2 m} \tag{5.13}$$

since we are conditioned to the fact that the algorithm has selected indices from $\Gamma^{\sharp}$ until iteration $l - 1$, we have that $\mathbf{r}^{l-1}$ only depends on the columns $\phi_j$ with $j \in \Gamma^{\sharp}$, therefore $\mathbf{u}_{l-1}$ is a unitary vector independent from the columns $\phi_j$ with $j \in \Gamma^{\sharp c}$.

Conditioned to $B_l$, $\mathbf{\Phi} \in A_l \Leftrightarrow \left\|\mathbf{\Phi}_{\Gamma^{\sharp}}(\omega)\mathbf{r}^{l-1}\right\|_\infty \geq \tilde{\alpha}\left\|\mathbf{r}^{l-1}\right\|_2$. Since $\tilde{\alpha} \leq \frac{1}{2\sqrt{k}}$, we have

$$\begin{aligned}
\mathbb{P}\left\{A_l \mid B_l \cap (\cap_{n=1}^{l-1} E_n)\right\} &= \mathbb{P}\left\{\{\left\|\mathbf{\Phi}_{\Gamma^{\sharp}}^*(\omega)\mathbf{r}^{l-1}\right\|_{\ell^\infty(\mathbb{R}^{\Gamma^{\sharp}})} \geq \tilde{\alpha}\left\|\mathbf{r}^{l-1}\right\|_{\ell^2(\mathbb{R}^m)}\} \mid \cap_{n=1}^{l-1} E_n\right\} \\
&\geq \mathbb{P}\left\{\{\left\|\mathbf{\Phi}_{\Gamma^{\sharp}}^*(\omega)\mathbf{y}\right\|_{\ell^2(\mathbb{R}^{\Gamma^{\sharp}})} \geq \tilde{\alpha}\sqrt{k}\left\|\mathbf{r}^{l-1}\right\|_{\ell^2(\mathbb{R}^m)}\} \mid \cap_{n=1}^{l-1} E_n\right\} \\
&\geq \mathbb{P}\left\{\left\|\mathbf{\Phi}_{\Gamma^{\sharp}}^*(\omega)\mathbf{r}^{l-1}\right\|_{\ell^2(\mathbb{R}^{\Gamma^{\sharp}})} \geq \frac{1}{2}\left\|\mathbf{r}^{l-1}\right\|_{\ell^2(\mathbb{R}^m)} \mid \cap_{n=1}^{l-1} E_n\right\}.
\end{aligned}$$

It is now possible to use (M4) since $\mathbf{r}^{l-1} \in \text{span}\left(\mathbf{\Phi}_{\Gamma^{\sharp}}\right)$ (which is not hard to prove) and since we are computing the probability conditioned to the fact that the algorithm has selected indices from $\Gamma^{\sharp}$. Hence, with the same reasoning that leads to (5.8) we get

$$\mathbb{P}\left\{A_l \mid B_l \cap (\cap_{n=1}^{l-1} E_n)\right\} \geq 1 - q_2 e^{-c_2 m/2}. \tag{5.14}$$

From (5.13) and (5.14) we deduce

$$\mathbb{P}\left\{E_l \mid \cap_{n=1}^{l-1} E_n\right\} \geq 1 - q_3(N-k)e^{-c_1\tilde{\alpha}^2 m} \tag{5.15}$$

reasoning as in the chain of inequalities that leads to (5.9). Substitute (5.15) and (5.10) in (5.11) to get

$$\begin{aligned}
\mathbb{P}\left\{\cap_{n=1}^{l} E_n\right\} &\geq (1 - q_3(N-k)e^{-c_1\tilde{\alpha}^2 m/2})(1 - q_3(l-1)(N-k)e^{-c_1\tilde{\alpha}^2 m}) \\
&\geq 1 - q_3(N-k)e^{-c_1\tilde{\alpha}^2 m/2} - q_3(l-1)(N-k)e^{-c_1\tilde{\alpha}^2 m} \\
&= 1 - q_3 l(N-k)e^{-c_1\tilde{\alpha}^2 m},
\end{aligned} \tag{5.16}$$

which is the desired estimation. Condition (5.5) assures that this probability is greater than zero.  ∎

**Remark 5.3.** *Taking $\tilde{\alpha} = \frac{1}{2\sqrt{k}}$, condition (5.5) is written*

$$m \geq \max\{\frac{4k}{c_1} \ln q_3 l(N-k), \frac{2k}{c_2} \ln D\}.$$

*Hence, if we take $m \geq Ck \ln l(N-k)$, with $C$ large enough, inequality (5.5) is guaranteed.*

For RWOMP with parameter $\tilde{\alpha}$ each iteration adds one element at least, as long as $\tilde{\mathcal{I}}_n(\tilde{\alpha}) \neq \emptyset$. In this situation, RWOMP identifies every index within $\Gamma^\sharp$ in as much $k$ iteration. Besides, the OMP type algorithms recover any $k$-sparse vector $\mathbf{x}$ once the support is known. Since $k(N-k) \leq N^2/4$ if $k < N$, we have the following corollary.

**Corollary 5.4.** *Suppose the same hypothesis that in Theorem 5.2 substituting (5.5) by*

$$m \geq \max\{\frac{1}{c_1\tilde{\alpha}^2} \ln q_3 \frac{N^2}{4}, \frac{2k}{c_2} \ln D\}. \tag{5.17}$$

*Then, RWOMP algorithm recovers the $k$-sparse vector $\mathbf{x}$ in the first $k$ iterations with probability greater or equal to*

$$1 - q_3 \frac{N^2}{4} e^{-c_1\tilde{\alpha}^2 m}.$$

**Remark 5.5.** *Suppose we want to obtain the result in Theorem 5.2 with a probability greater than or equal to $1 - \beta$ for a number $\beta \in (0, 1)$. It would be enough to take*

$$1 - q_3 l(N-k) e^{-c_1\tilde{\alpha}^2 m} \geq 1 - \beta.$$

*This is accomplished if*

$$\beta e^{c_1\tilde{\alpha}^2 m} \geq q_3 l(N-k),$$

*for which it is enough to take*

$$m \geq \max\{\frac{2}{c_1\tilde{\alpha}^2} \ln \frac{q_3 l(N-k)}{\beta}, \frac{2k}{c_2} \ln D\}. \tag{5.18}$$

*To be sure that*

$$\ln \frac{q_3 l(N-k)}{\beta} > 1$$

*for all $q_3 \geq 1$, $l < N$ and $k < N$, it suffices to take $\beta \in (0, 1/e)$.*

*An analogous comment can be done for Corollary 5.4, and in this case (5.17) should be substituted by*

$$m \geq \max\{\frac{1}{c_1\tilde{\alpha}^2} \ln \frac{q_3 N^2}{4\beta}, \frac{2k}{c_2} \ln D\} \tag{5.19}$$

*for RWOMP to recover a $k$-sparse signal $\mathbf{x}$ within the first $k$ iterations with probability greater than or equal to $1 - \beta$.*

5.2. **Probabilistic support identification with selection rule $\mathcal{I}(\alpha)$.** In this section we will study the behavior of matrices satisfying (M1), (M2), (M3) and (M4) with respect to the algorithms with selection rule

$$\mathcal{I}_n(\alpha) = \{i : \left|\langle \phi_i, \mathbf{r}^{n-1}\rangle\right| \geq \alpha \left\|\mathbf{\Phi}^*\mathbf{r}^{n-1}\right\|_{\ell^\infty(\mathbb{R}^N)}\}, \tag{5.20}$$

$0 < \alpha \leq 1$. An advantage of $\mathcal{I}_n(\alpha)$ with respect to the selection rule $\tilde{\mathcal{I}}_n(\tilde{\alpha})$ given in (3.1) is that $\mathcal{I}_n(\alpha) \neq \emptyset$ for all $\alpha \in (0,1)$ and for all $n = 1, 2, \ldots$.

There is a disadvantage, however. If we want to prove that $\mathcal{I}_n(\alpha) \subset \Gamma^\sharp = \text{supp}(\mathbf{x})$ we need

$$\frac{\left\|\mathbf{\Phi}^*_{\Gamma^{\sharp c}}\mathbf{r}^{n-1}\right\|_\infty}{\left\|\mathbf{\Phi}^*\mathbf{r}^{n-1}\right\|_\infty} < \alpha$$

to be satisfied. If we wrote $\mathbf{u}_{n-1} = \mathbf{r}^{n-1}/\left\|\mathbf{\Phi}^*\mathbf{r}^{n-1}\right\|_\infty$ as in the proof of Theorem 5.2 we would not have $\|\mathbf{u}_{n-1}\|_2 \leq 1$, and could not use (M3). Property (M4) saves this situation as shown in the next result.

**Theorem 5.6.** *Choose $\mathbf{\Phi}(\omega) \in \mathbb{R}^{m \times N}$ a random matrix satisfying (M1), (M2), (M3) y (M4). Let $\mathbf{x} \in \mathbb{R}^N$ with $\text{supp}(\mathbf{x}) = \Gamma^\sharp$ and $\left|\Gamma^\sharp\right| \leq k < N$. Let $\mathbf{y} = \mathbf{\Phi}\mathbf{x}$. Suppose that*

$$m \geq \max\{\frac{4k}{c_1\alpha^2}\ln q_3 l(N-k), \frac{2k}{c_2}\ln D\}, \tag{5.21}$$

*with $q_3 = q_1 + q_2$. Algorithms WMP, WOMP and WGP, with selection rule $\mathcal{I}_n(\alpha)$ given in 5.20, identify elements from $\Gamma^\sharp$ in the first $l$ iterations with probability greater than or equal to*

$$1 - q_3 l(N-k)e^{-c_1\frac{\alpha^2}{4k}m}.$$

**Proof.** For any of the given algorithms with selection rule $\mathcal{I}_n(\alpha)$, let $E_n$ be the set of matrices that identify elements from $\Gamma^\sharp = \text{supp}(\mathbf{x})$ at iteration $n$, $n = 1, 2, \ldots$.

We start bounding $\mathbb{P}\{E_1\}$, that is, the probability that a matrix identify elements from $\Gamma^\sharp$ in the first iteration. For a random matrix $\mathbf{\Phi}(\omega)$ belong to $E_1$ it is enough that

$$\frac{\left\|\mathbf{\Phi}^*_{\Gamma^{\sharp c}}(\omega)\mathbf{r}^0\right\|_{\ell^\infty(\mathbb{R}^{\Gamma^{\sharp c}})}}{\left\|\mathbf{\Phi}^*_{\Gamma^\sharp}(\omega)\mathbf{r}^0\right\|_{\ell^\infty(\mathbb{R}^N)}} < \alpha, \quad (\mathbf{r}^0 = \mathbf{y}), \tag{5.22}$$

is verified. Let $\mathbf{u}_0 = \mathbf{y}/2\left\|\mathbf{\Phi}^*_{\Gamma^\sharp}\mathbf{y}\right\|$; we have

$$\frac{\left\|\mathbf{\Phi}^*_{\Gamma^{\sharp c}}(\omega)\mathbf{y}\right\|_\infty}{\left\|\mathbf{\Phi}^*_{\Gamma^\sharp}(\omega)\mathbf{y}\right\|_\infty} \leq \frac{\sqrt{k}\left\|\mathbf{\Phi}^*_{\Gamma^{\sharp c}}(\omega)\mathbf{y}\right\|_\infty}{\left\|\mathbf{\Phi}^*_{\Gamma^\sharp}(\omega)\mathbf{y}\right\|_2} = 2\sqrt{k}\left\|\mathbf{\Phi}^*_{\Gamma^{\sharp c}}(\omega)\mathbf{u}_0\right\|_\infty$$

$$= 2\sqrt{k}\sup_{j \in \Gamma^{\sharp c}}\left|\langle\phi_j(\omega), \mathbf{u}_0\rangle\right|. \tag{5.23}$$

Let $A_1$ be the set of matrices satisfying (M1), (M2), (M3) and (M4) such that

$$\sup_{j \in \Gamma^{\sharp c}}\left|\langle\phi_j(\omega), \mathbf{u}_0\rangle\right| < \frac{\alpha}{2\sqrt{k}}. \tag{5.24}$$

From (5.23) we deduce that if $\mathbf{\Phi}(\omega) \in A_1$, then $\mathbf{\Phi}(\omega)$ satisfy (5.22) and we have $\mathbf{\Phi}(\omega) \in E_1$. Therefore,

$$\mathbb{P}\{E_1\} \geq \mathbb{P}\{A_1\}. \tag{5.25}$$

Let $B_1$ be the set of matrices $\mathbf{\Phi}(\omega)$ satisfying (M1), (M2), (M3) and (M4) such that

$$\left\|\mathbf{\Phi}_{\Gamma^\sharp}^*(\omega)\mathbf{y}\right\|_2 \geq \frac{1}{2}\left\|\mathbf{y}\right\|_2. \tag{5.26}$$

From (5.25) we deduce

$$\mathbb{P}\left\{E_1\right\} \geq \mathbb{P}\left\{A_1\right\} \geq \mathbb{P}\left\{A_1 \cap B_1\right\} = \mathbb{P}\left\{A_1|B_1\right\}\mathbb{P}\left\{B_1\right\}. \tag{5.27}$$

Since $\mathbf{y} = \mathbf{\Phi}\mathbf{x} = \mathbf{\Phi}_{\Gamma^\sharp}\mathbf{x}_{\Gamma^\sharp} \in \mathrm{span}\,(\mathbf{\Phi}_{\Gamma^\sharp})$ from (M4) we deduce

$$\mathbb{P}\left\{B_1\right\} \geq 1 - q_2 D^k e^{-c_2 m} \geq 1 - q_2 e^{-c_2 m/2} \tag{5.28}$$

taking $-c_2 m + k \ln D \leq -c_2 m/2$, this is, $m \geq \frac{2k}{c_2}\ln D$. Conditioned to $B_1$ the vector $\mathbf{u}_0 = \mathbf{y}/2\left\|\mathbf{\Phi}_{\Gamma^\sharp}^*\mathbf{y}\right\|_2$ satisfy

$$\left\|\mathbf{u}_0\right\|_2 = \frac{1}{2}\frac{\left\|\mathbf{y}\right\|_2}{\left\|\mathbf{\Phi}_{\Gamma^\sharp}^*\mathbf{y}\right\|_2} \leq 1.$$

Therefore, to bound

$$\mathbb{P}\left\{A_1|B_1\right\} = \mathbb{P}\left\{\sup_{j \in \Gamma^{\sharp c}} |\langle \phi_j(\omega), \mathbf{u}_0\rangle| < \frac{\alpha}{2\sqrt{k}}|B_1\right\}$$

we use (M3) to get

$$\begin{aligned}
\mathbb{P}\left\{A_1|B_1\right\} &= \mathbb{P}\left\{\cap_{j \in \Gamma^{\sharp c}}\{|\langle \phi_j(\omega), \mathbf{u}_0\rangle| < \frac{\alpha}{2\sqrt{k}}\}|B_1\right\} \\
&= \prod_{j \in \Gamma^{\sharp c}} \mathbb{P}\left\{|\langle \phi_j(\omega), \mathbf{u}_0\rangle| < \frac{\alpha}{2\sqrt{k}}|B_1\right\} \\
&\geq \left(1 - q_1 e^{-c_1 \frac{\alpha^2}{4k} m}\right)^{N-k}
\end{aligned}$$

due to the independence of the columns $\phi_j$ of $\mathbf{\Phi}(\omega)$ expressed in (M1) and because $\mathbf{u}_0 = \mathbf{y}/2\left\|\mathbf{\Phi}_{\Gamma^\sharp}^*\mathbf{y}\right\|_2$ only depends on the columns $\phi_j$ with $j \in \Gamma^\sharp$, and therefore independent of the columns $\phi_j$ with $j \in \Gamma^{\sharp c}$.

Since $(1-x)^n \geq 1 - nx$ if $n \geq 1$, $x \leq 1$, we can write

$$\mathbb{P}\left\{A_1|B_1\right\} \geq 1 - q_1(N-k)e^{-c_1 \frac{\alpha^2}{4k} m}. \tag{5.29}$$

Substituting (5.29) and (5.28) in (5.27) we get

$$\begin{aligned}
\mathbb{P}\left\{E_1\right\} &\geq (1 - q_1(N-k)e^{-c_1 \frac{\alpha^2}{4k} m})(1 - q_2 e^{-c_1 \frac{m}{2}}) \\
&\geq 1 - q_1(N-k)e^{-c_1 \frac{\alpha^2}{4k} m} - q_2 e^{-c_1 \frac{m}{2}} \\
&\geq 1 - q_3(N-k)e^{-c_1 \frac{\alpha^2}{4k} m} \tag{5.30}
\end{aligned}$$

with $q_3 = q_1 + q_2$, since $\frac{\alpha^2}{4k} \leq 1/4 < 1/2$. This proves the result for $l = 1$.

Suppose that the result is true until iteration $l-1$, this is

$$\mathbb{P}\left\{\cap_{n=1}^{l-1} E_n\right\} \geq 1 - q_3(l-1)(N-k)e^{-c_1 \frac{\alpha^2}{4k} m}. \tag{5.31}$$

Until iteration $l$ we have

$$\mathbb{P}\left\{\cap_{n=1}^l E_n\right\} = \mathbb{P}\left\{E_l| \cap_{n=1}^{l-1} E_n\right\}\mathbb{P}\left\{\cap_{n=1}^{l-1} E_n\right\}. \tag{5.32}$$

Write $\mathbf{u}_{l-1} = \mathbf{r}^{l-1}/2 \left\|\mathbf{\Phi}^*_{\Gamma^\sharp}\mathbf{r}^{l-1}\right\|_2$. Let $A_l$ be the set of matrices satisfying (M1), (M2), (M3) and (M4) such that

$$\sup_{j\in\Gamma^{\sharp c}} |\langle \phi_j(\omega), \mathbf{u}_{l-1}\rangle| < \frac{\alpha}{2\sqrt{k}}. \tag{5.33}$$

If $\mathbf{\Phi} \in A_l$ we have

$$
\begin{aligned}
\frac{\left\|\mathbf{\Phi}^*_{\Gamma^{\sharp c}}(\omega)\mathbf{r}^{l-1}\right\|_\infty}{\left\|\mathbf{\Phi}^*_{\Gamma^\sharp}(\omega)\mathbf{r}^{l-1}\right\|_\infty} &\leq \frac{\sqrt{k}\left\|\mathbf{\Phi}^*_{\Gamma^{\sharp c}}(\omega)\mathbf{r}^{l-1}\right\|_\infty}{\left\|\mathbf{\Phi}^*_{\Gamma^\sharp}(\omega)\mathbf{r}^{l-1}\right\|_2} = 2\sqrt{k}\left\|\mathbf{\Phi}^*_{\Gamma^{\sharp c}}(\omega)\mathbf{u}_{l-1}\right\|_\infty \\
&= 2\sqrt{k}\sup_{j\in\Gamma^{\sharp c}} |\phi_j(\omega), \mathbf{u}_{l-1}| < \alpha,
\end{aligned}
$$

and this is enough to assure that $\mathcal{I}_n(\alpha) \subset \Gamma^\sharp$, this is, $\mathbf{\Phi} \in E_l$. Therefore,

$$\mathbb{P}\left\{E_l| \cap_{n=1}^{l-1} E_n\right\} \geq \mathbb{P}\left\{A_l| \cap_{n=1}^{l-1} E_n\right\}. \tag{5.34}$$

Let $B_l$ be the set of matrices $\mathbf{\Phi}(\omega)$ satisfying (M1), (M2), (M3) and (M4) such that

$$\left\|\mathbf{\Phi}^*_{\Gamma^\sharp}(\omega)\mathbf{r}^{n-1}\right\|_2 \geq \frac{1}{2}\left\|\mathbf{r}^{n-1}\right\|_2. \tag{5.35}$$

From (5.34) we deduce

$$\mathbb{P}\left\{E_l| \cap_{n=1}^{l-1} E_n\right\} \geq \mathbb{P}\left\{A_l \cap B_l| \cap_{n=1}^{l-1} E_n\right\} = \mathbb{P}\left\{A_l|B_l \cap (\cap_{n=1}^{l-1} E_n)\right\}\mathbb{P}\left\{B_l| \cap_{n=1}^{l-1} E_n\right\}. \tag{5.36}$$

Conditioned to $B_l$,

$$\|\mathbf{u}_{l-1}\|_2 = \frac{1}{2}\frac{\left\|\mathbf{r}^{l-1}\right\|_2}{\left\|\mathbf{\Phi}^*_{\Gamma^\sharp}\mathbf{r}^{l-1}\right\|_2} \leq 1$$

due to (5.35). By (M1) the columns $\phi_j$ of $\mathbf{\Phi}(\omega)$ are independents among them. Besides, conditioned to $\cap_{n=1}^{l-1}E_n$, the vector $\mathbf{r}^{n-1}$ only depends on the columns $\phi_j$ with $j \in \Gamma^\sharp$ since $\mathcal{I}_n(\alpha) \subset \Gamma^\sharp$, $n = 1, 2, \ldots, l-1$. Hence, $\mathbf{u}_{l-1}$ is independent from the columns $\phi_j$ with $j \in \Gamma^{\sharp c}$ and we can use (M3). Then,

$$
\begin{aligned}
\mathbb{P}\left\{A_l|B_l \cap (\cap_{n=1}^{l-1}E_n)\right\} &= \mathbb{P}\left\{\sup_{j\in\Gamma^{\sharp c}} |\langle \phi_j(\omega), \mathbf{u}_{l-1}\rangle| < \frac{\alpha}{2\sqrt{k}}|B_l \cap (\cap_{n=1}^{l-1}E_n)\right\} \\
&= \prod_{j\in\Gamma^{\sharp c}} \mathbb{P}\left\{|\langle \phi_j(\omega), \mathbf{u}_{l-1}\rangle| < \frac{\alpha}{2\sqrt{k}}|B_l \cap (\cap_{n=1}^{l-1}E_n)\right\} \\
&\geq (1 - q_1 e^{-c_1\frac{\alpha^2}{4k}m})^{N-k} \geq 1 - q_1(N-k)e^{-c_1\frac{\alpha^2}{4k}m}. \tag{5.37}
\end{aligned}
$$

Conditioned to $\cap_{n=1}^{l-1}E_n$, it is not hard to prove that the vector $\mathbf{r}^{n-1} \in \operatorname{span}(\mathbf{\Phi}_{\Gamma^\sharp})$. We can use (M4) to get

$$\mathbb{P}\left\{B_l| \cap_{n=1}^{l-1} E_n\right\} \geq 1 - q_1 D^k e^{-c_2 m} \geq 1 - q_1 e^{-c_2 m/2} \tag{5.38}$$

taking $-c_2 m + k\ln D \leq -c_2\frac{m}{2}$, this is, $m \geq \frac{2k}{c_2}\ln D$.

Substituting (5.36) y (5.38) en (5.35) and proceed as in the calculations that lead to (5.30) to get

$$\mathbb{P}\left\{E_l| \cap_{n=1}^{l-1} E_n\right\} \geq 1 - q_3(N-k)e^{-c_1\frac{\alpha^2}{4k}m}, \tag{5.39}$$

with $q_3 = q_1 + q_2$. Substitute (5.39) and (5.31) in (5.32) to get

$$
\begin{aligned}
\mathbb{P}\left\{\cap_{n=1}^{l} E_n\right\} &\geq (1 - q_3(N-k)e^{-c_1\frac{\alpha^2}{4k}m})(1 - q_3(l-1)(N-k)e^{-c_1\frac{\alpha^2}{4k}m}) \\
&\geq 1 - q_3(N-k)e^{-c_1\frac{\alpha^2}{4k}m} - q_3(l-1)(N-k)e^{-c_1\frac{\alpha^2}{4k}m} \\
&= 1 - q_3 l(N-k)e^{-c_1\frac{\alpha^2}{4k}m},
\end{aligned}
$$

which is what we wanted to prove. For the last probability be greater than 0 we must take

$$
m \geq \frac{4k}{c_1\alpha^2}\ln q_3 l(N-k).
$$

$\blacksquare$

We can now state similar comments to those at the end of the proof of Theorem 5.2. We emphasize next corollary, that in the case $\alpha = 1$ gives Theorem 6 of [31].

**Corollary 5.7.** *Choose the same conditions as in Theorem 5.6, substituting (5.21) by*

$$
m \geq C\frac{k}{\alpha^2}\ln(q_3 N^2/4) \tag{5.40}
$$

*with $C$ large enough. Then, WOMP algorithm recovers the $k$-sparse vector $\mathbf{x}$ in the first $k$ iterations with probability greater than or equal to*

$$
1 - q_3\frac{N^2}{4}e^{-c_1\frac{\alpha^2}{4k}m}.
$$

## 6. Some Experiments.

In this section we present experiments on the sparse recovery and non sparse approximation problems for the orthogonal and gradient algorithms with the selection rule $\tilde{\mathcal{I}} = \tilde{\mathcal{I}}(\alpha)$ given by (3.1) and compare them against the results obtained with the more classical selection rule $\mathcal{I} = \mathcal{I}(\alpha)$ given by (2.6). In every set of experiments one Gaussian matrix was created of order $m \times N$. For some signals the results are shown as percentage of elements recovered and for others we use the **signal-to-noise ratio** of the energy of the original signal $\mathbf{x} \in \mathbb{R}^N$ and the energy of the difference between the signal and the approximation $\mathbf{a}$, given by

$$
SNR = 10\log_{10}(\frac{\|\mathbf{x}\|_2}{\|\mathbf{x} - \mathbf{a}\|_2}).
$$

Figure 1 shows the percentage of elements recovered with the RWOMP algorithm with parameter $\alpha = 0.125$. The Gaussian matrices generated have $N = 256$ and $m = 10\ell$, $\ell = 1, 2, \ldots, 25$. The sparsity levels have been chosen to be $k = 4, 12, 20, 28, 36$, (each graph corresponds to one of them). For each pair $(m, k)$, 200 experiments were run for different signals. The results can be compared with those obtained on Figure 1 of [31] for OMP ($\alpha = 1$). The parameters $N, k$ and $m$ take the same value, but 1000 experiments for each set were performed in [31]. The results for RWOMP are better that those for OMP in [31] for $k = 20, 28, 36$.

Next, for computational purposes, we introduce a minor modification of the RWOMP algorithm, called $k$-RWOMP algorithm. At each iteration in RWOMP we keep the $k$-largest elements of the orthogonal approximation $\mathbf{x}^n$. Figure 2 shows the results of applying the $k$-RWOMP with the same parameters of the experiments described for
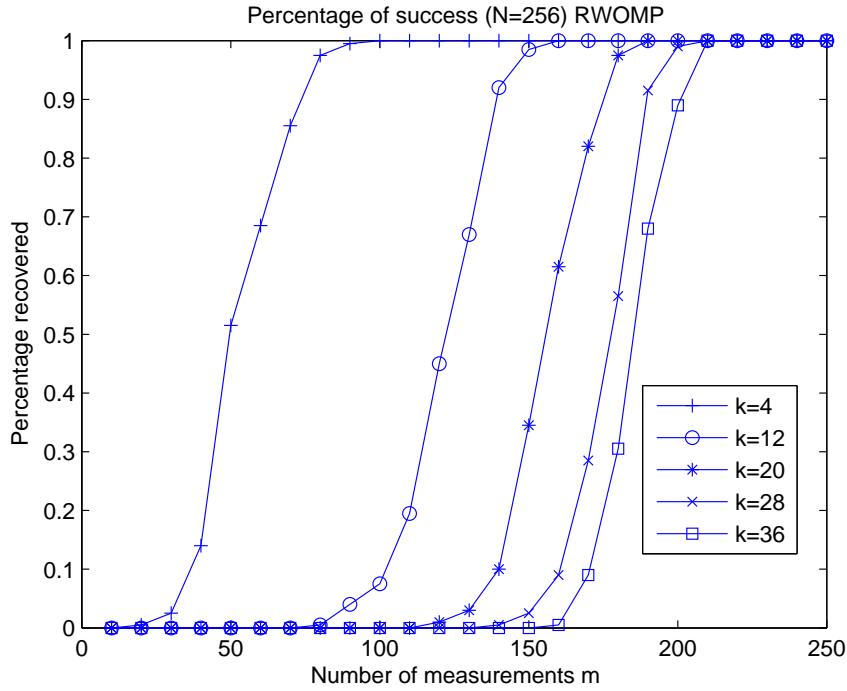
FIGURE 1. Recovery of sparse signals with RWOMP, $\alpha = 0.125$. For each set $(m, k)$, 200 experiments were generated with one Gaussian matrix of order $m \times 256$. Up to $k$ iterations are allowed.

Figure 1. As can be seen, in Figure 2 exact recovery is achieved with smaller values of $m$.

The next set of experiments is done on images. We take an image of $64 \times 64$ pixels, which is part of Lena. Two decompositions of the wavelet transform were performed using *Daubechies 5* wavelets. The compressed sensing was done only on the detail coefficients. The sparsity $k$ is calculated as the integer part of 5% of $L$, where $L$ is the number of vertical, horizontal or diagonal coefficients at each decomposition level. The number of measurements at each level is the integer part of $k \log_2(L/k)$ for vertical, horizontal and diagonal coefficients. The k-RWOMP and k-RWGP algorithms (those with selection rule $\tilde{\mathcal{I}}(\alpha)$ as defined in (3.1)) were run with $\tilde{\alpha} = 0.125$ and $\tilde{\alpha} = 0.15$, respectively. The WGP algorithm (with selection rule $\mathcal{I}(\alpha)$ as defined in (2.6)) was run with $\alpha = 0.8$. The relevant data is given in Figure 3. When the sparsity goes to 10% of $L$ it takes several minutes for the CoSaMP algorithm to stop, whereas for the rest of algorithms it takes almost the same time. For each algorithm, the running times and the $SNR$'s differ for different $\alpha$'s and, moreover, they may even differ with the same $\alpha$'s due to the randomness of $\mathbf{\Phi}$. Also visually there is an improvement for the k-Relaxed Weak algorithms for which there are less artifacts in the smoother regions of the image. Judging by the SNR, k-RWGP gives the best approximation.
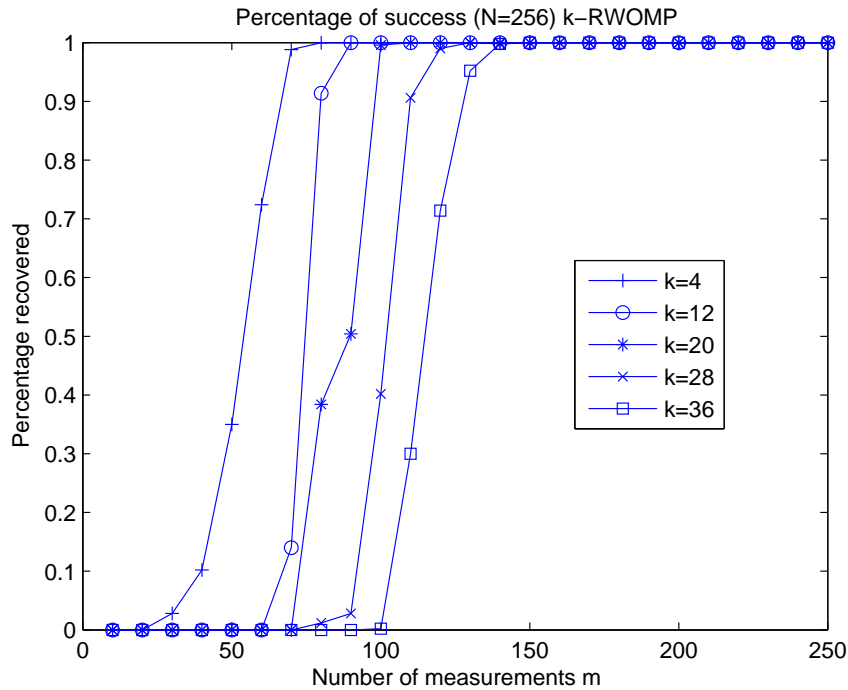
FIGURE 2. Recovery of sparse signals with $k-$RWOMP, with the same parameters as in Figure 1. The results are better than in Figure 1 for RWOMP. Up to $k$ iterations are allowed.

## REFERENCES

[1] D. Achlioptas, *Database-friendly Random Projections: Johnson-Lindenstrauss with Binary Coins*, Journal of Comp. & Sys. Sci., 66 (4), (2003), p.671-687, special issue of invited papers from PODS'01.

[2] A.R. Barron, "Universal approximation bounds for superposition of $n$ sigmoidal functions", *IEEE Trans. Inform. Thoery*, vol. 39, pp. 930-945, 1993.

[3] R. Baraniuk, M. Davenport, R. DeVore and M. Wakin, *A Simple Proof of the Restricted Isometry Property for Random Matrices*, Constructive Approximation, vol. 28, no. 3, pp. 253-263, 2008.

[4] T. Blumensath, M. E. Davies; "Gradient Pursuits", *IEEE Transactions on Signal Processing,* vol. 56, num. 6, pp. 2370-2382, 2008.

[5] T. Blumensath, M. E. Davies, "Stagewise Weak Gradient Pursuits", *IEEE Transactions on Signal Processing* vol. 57, num. 11, pp. 4333-4346, 2009.

[6] E. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. on Information Theory,* vol. 51, num. 12, pp. 4203-4215, 2005.

[7] E. Candès and T. Tao, "Near optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. on Information Theory,* vol. 52, num. 12, pp. 5406-5425, 2006.

[8] A. Cohen, W. Dahmen, R.A. DeVore, "Instance Optimal Decoding by Thersholding in Compressed Sensing," *Proceedings of the 8th International Conference in Harmonic Analysis and Partial Differential Equations, El Escorial 2008*, Contemporary Mathematics, AMS, 2009.

[9] W. Dai, O. Milenkovic, "Subspace pursuit for compressive sensing: closing the gap between performance and complexity", *IEEE Trans, Inform, Theory*, vol. 55, no 5, pp. 2230-2249, 2009.

[10] S. Dasgupta and A. Gupta, *An elementary proof of the Johnson-Lindenstrauss lemma*, ICSI Technical Report TR-99-006, U.C. Berkeley, 1999.
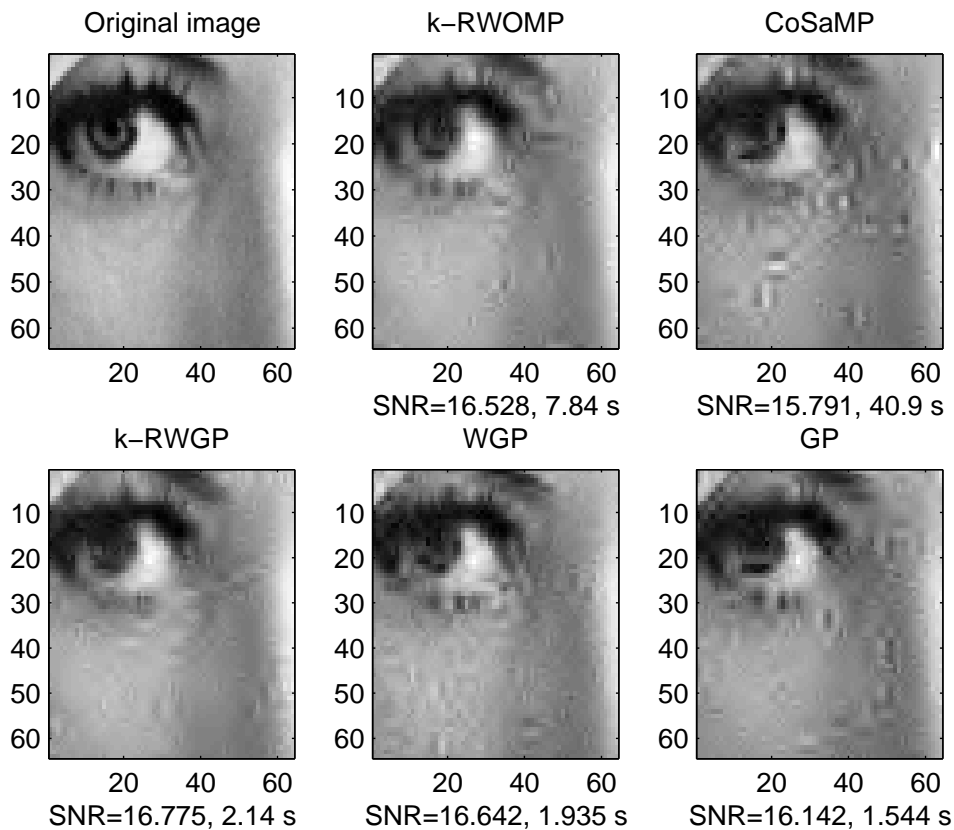
FIGURE 3. k-RWOMP and k-RWGP vs CoSaMP, SWGP and GP.

[11] M. Davenport, M. Watkin, "Analysis of Orthogonal Matching Pursuit using the Restricted Isometry Property", *IEEE Trans, Inform, Theory*, vol. 56, no 9, pp. 4395-4401, 2010.

[12] R.A. DeVore, V.N, Temlyakov, "Some remarks on greedy algoritms", *Adv. in Comput. Math.*, vol. 5, pp. 173-187, 1996.

[13] R. A. DeVore, *Deterministic construction of compressed sensing matrices*, Journal of Complexity, 23:918-925, 2007.

[14] D.L. Donoho, "Compressed sensing", *IEEE Trans. on Information Theory,* vol. 52, num. 4, pp. 1289-1306, 2006.

[15] D.L. Donoho, Y. Tsaig, I. Drori, J-L. Starck, "Sparse Solution of Underdetermined Linear Equations by Stagewise Orthogonal Matching Pursuit", *IEEE Trans. on Information Theory,* vol. 58, num. 2, pp. 1094-1121, 2012.

[16] R. Gribonval, P. Vandergheynst, "On the exponential convergence of Matching Pursuits in Quasi-Incoherent Dictionaries," *IEEE Trans. Information Theory,* vol. 52, num. 1, pp. 255-261, 2006.

[17] L. Jones, "On a conjecture of Huber concerning the convergence of projection pursuit regression", *The Annals of Statistics*, vol. 15, pp. 880-882, 1987.

[18] E. Liu, V. N. Temlyakov, "Orthogonal supergreedy algorithm and applications in Compressed Sensing", Preprint.

[19] G. Lugosi. "Concentration-of-measure inequalities. Lecture Notes, december 2004," [On-line]. Available on *http://www.econ.upf.edu/ lugosi/surveys.html*

[20] R. Maleh, "Improved RIP Analysis of Orthogonal Matching Pursuit", Preprint.

[21] S. Mallat. *A Wavelet tour on signal processing*, Academic Press, 2nd. ed., 1999.

[22] Q. Mo, Y. Shen, "Remarks on the Restricted Isometry Property in Orthogonal Matching Pursuit algorithm", Preprint, 2012.

[23] S. Mallat, Z. Zhang, "Matching Pursusit with time-frequency dictionaries", *IEEE Trans. Signal Processing*, vol. 41, num. 12, pp. 3397-3415, 1993.

[24] D. Needell, R. Vershynin, "Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit," *Found. Comput. Math.*, vol. 9, pp. 317-334, 2009.

[25] D. Needell, J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmon. Anal.*, vol. 26, num. 3, pp. 301-321, 2009.

[26] M. Rudelson, R. Vershynin, "Sparse reconstruction by convex relaxation: Fourier and Gaussian measurements," *CISS 2006 (40th Annual Conference on Information Sciences and Systems)*.

[27] V.N. Temlyakov, "Weak greedy algorithms", *Adv. in Comput. Math.*, vol. 12, pp. 213-227, 2000.

[28] V.N. Temlyakov, "Nonlinear methods of approximation", *Found. Comput. Math.*, vol. 17, pp. 269-280, 2003.

[29] V.N. Temlyakov, "Greedy approximation", *Cambridge Monographs on Applied and Computational Mathematics*, Cambtidge University Press, 2011.

[30] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Info. Theory*, vol. 50, num. 10, pp. 2231-2242, 2004.

[31] J. A. Tropp, A. Gilbert, "Signal recovery from random measurements via Orthogonal Matching Pursuit," *IEEE Trans. Info. Theory*, vol. 53, num. 12, pp. 4655-4666, 2007.

EUGENIO HERNÁNDEZ, DEPARTAMENTO DE MATEMÁTICAS, UNIVERSIDAD AUTÓNOMA DE MADRID, 28049 MADRID, SPAIN
  *E-mail address*: eugenio.hernandez@uam.es

DANIEL VERA, DEPARTAMENTO DE MATEMÁTICAS, UNIVERSIDAD AUTÓNOMA DE MADRID, 28049 MADRID, SPAIN
  *E-mail address*: daniel.vera@uam.es