



# Challenging Restricted Isometry Constants with Greedy Pursuit

Charles H Dossal, Gabriel Peyré, Jalal M. Fadili

## ► To cite this version:

Charles H Dossal, Gabriel Peyré, Jalal M. Fadili. Challenging Restricted Isometry Constants with Greedy Pursuit. ITW'09, Oct 2009, Taormina, Italy. pp.475-479. hal-00373450v2

**HAL Id: hal-00373450**

**<https://hal.science/hal-00373450v2>**

Submitted on 13 Jul 2009

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Challenging Restricted Isometry Constants with Greedy Pursuit

Charles Dossal  
IMB  
Université Bordeaux 1  
charles.dossal@math.u-bordeaux1.fr

Gabriel Peyré  
Ceremade  
CNRS-Université Paris-Dauphine  
gabriel.peyre@ceremade.dauphine.fr

Jalal Fadili  
GREYC  
CNRS-ENSICAEN-Université Caen  
jalal.fadili@greyc.ensicaen.fr

**Abstract**—This paper proposes greedy numerical schemes to compute lower bounds of the restricted isometry constants that are central in compressed sensing theory. Matrices with small restricted isometry constants enable stable recovery from a small set of random linear measurements. We challenge this compressed sampling recovery using greedy pursuit algorithms that detect ill-conditioned sub-matrices. It turns out that these sub-matrices have large isometry constants and hinder the performance of compressed sensing recovery.

## I. COMPRESSED SAMPLING RECOVERY

Compressed sampling paradigm consists in acquiring a small number of linear measurements  $y = Ax$ , where  $x \in \mathbb{R}^N$  is the high resolution signal to recover, and  $y \in \mathbb{R}^P$  is the vector of measurements with  $P \ll N$ . Using the underdetermined sensing matrix  $A$  that is drawn from a proper random ensemble, and having knowledge of the measurements  $y$ , the sparsest vector giving rise to these measurements is sought. Throughout this paper, we consider the case where the entries of  $A$  are independent and identically distributed  $\mathcal{N}(0, 1/P)$ .

For noiseless measurements  $y = Ax$ , the recovery of a sparse vector  $x$  is achieved by solving the convex program

$$\min_{\tilde{x} \in \mathbb{R}^N} \|\tilde{x}\|_1 \quad \text{subj. to} \quad A\tilde{x} = y, \quad (1)$$

where  $\|\tilde{x}\|_1 = \sum_i |\tilde{x}_i|$ .

The vector  $x$  is said to be identifiable if the solution  $x^*$  to (1) is unique and coincides exactly with  $x$ . To ensure identifiability, several sufficient conditions on  $x$  were considered in the literature. Of particular interest are those relying on the sparsity (or cardinality of the support)  $\|x\|_0 = |I(x)|$ , where the support of  $x$  is

$$I(x) = \{i \mid x_i \neq 0\}.$$

With high probability on the sampling matrix  $A$ , compressed sampling theory [1], [2] asserts that any vector satisfying

$$\|x\|_0 \leq \rho(P/N)P \quad (2)$$

is identifiable for  $\rho(\eta) > 0$ .

**Restricted isometry based criteria.** The seminal work of Donoho [1], Candès, Romberg and Tao [2] has focused on the stability of the compressed sampling decoder. Towards this goal, these authors introduced the restricted isometry property (RIP), with the RIP constants  $0 < \delta_s^{\min} \leq \delta_s^{\max} < 1$ . These

constants are the smallest numbers such that for every vector  $x \in \mathbb{R}^N$  with  $\|x\|_0 \leq s$ ,

$$(1 - \delta_s^{\min})\|x\|^2 \leq \|Ax\|^2 \leq (1 + \delta_s^{\max})\|x\|^2. \quad (3)$$

Let  $A_I = (a_i)_{i \in I}$  be a sub-matrix of  $s = |I|$  columns extracted from  $A$ . Then (3) is equivalent to saying that for all  $I$  such that  $|I| = s$ , the smallest and largest eigenvalues  $\lambda^{\min}(A_I)$  and  $\lambda^{\max}(A_I)$  of the Gram matrix  $A_I^* A_I$  are respectively lower-bounded and upper-bounded by  $1 - \delta_s^{\min}$  and  $1 + \delta_s^{\max}$ . Thus, the RIP constants are equivalently defined as

$$\delta_s^{\min} = \max_{|I|=s} \delta^{\min}(A_I) \quad \text{and} \quad \delta_s^{\max} = \max_{|I|=s} \delta^{\max}(A_I)$$

$$\text{where} \quad \begin{cases} \delta^{\min}(A_I) = 1 - \lambda^{\min}(A_I), \\ \delta^{\max}(A_I) = \lambda^{\max}(A_I) - 1 \end{cases}.$$

The original work of Candès et al. [2] considers a symmetric RIP constant  $\delta_s^{\min} = \max(\delta_s^{\max}, \delta_s)$ . These authors proved that a small enough value of  $\delta_{2s}$  ensures identifiability of all  $s$ -sparse vectors. For instance, it is proved in [3] that  $\delta_{2s} \leq \sqrt{2} - 1$  guarantees identifiability of all  $s$ -sparse vectors. This is achieved with high probability on  $A$  if  $s \leq CP/\log(N/P)$ , which corresponds to condition (2) with  $\rho(\eta) \leq C/\log(\eta^{-1})$  with  $\eta^{-1} = N/P$  the redundancy of the matrix  $A$ .

It turns out that the largest and smallest eigenvalues  $\lambda^{\min}(A_I)$  and  $\lambda^{\max}(A_I)$  do not deviate from 1 at the same rate. Using asymmetric RIP constants, Foucart and Lai [4] show that

$$(4\sqrt{2} - 3)\delta_{2s}^{\min} + \delta_{2s}^{\max} < 4(\sqrt{2} - 1) \quad (4)$$

implies identifiability of all  $s$ -sparse vectors. Blanchard et al. [5] determine  $\rho_0$  such that with high probability on  $A$

$$\|x\|_0 \leq \rho_0(P/N)P \quad (5)$$

ensures that condition (4) is in force. Condition (5) guarantees not only identifiability, but also robustness to noisy measurements. This however causes the function  $\rho_0(\eta)$  to be quite small, and for instance  $\rho_0(1/2) = 0.003$  and  $\rho_0(1/4) = 0.0027$ .

**Topological noiseless criteria.** More precise identifiability criteria exist that do not rely solely on the sparsity level  $\|x\|_0$  of  $x$ . Donoho [6] devises a topological recovery criterion that allows him to derive an explicit sharp value for  $\rho(\eta)$  in (2).

For instance  $\rho(1/2) \approx 0.089$  and  $\rho(1/4) \approx 0.065$ . This is better than the RIP-based bound  $\rho_0$  given above, since the RIP condition imposes a stronger metric constraint on the sensing matrix. These bounds based on topological criteria are however not robust to noise, and we restrict in this paper our analysis to the RIP condition.

This topological criterion is local and ensures recovery of a given  $x$  alone. This allows Donoho to devise a bound that ensures the recovery of most sparse vectors. This bound is better than (5) since it does not constrain all  $s$ -sparse vectors but only most of these vectors. In contrast, recovery conditions based on restricted isometry constants are global. The recovery of some  $x \in \mathbb{R}^N$  does not only rely on the conditioning of the matrix  $A_I$  for  $I = I(x)$ . In particular, a single matrix  $A_I$ , with  $|I| \leq 2s$  may be a problem for the identifiability of many  $s$ -sparse vectors  $x$  if  $\delta^{\min}(A_I)$  or  $\delta^{\max}(A_I)$  is too large.

## II. GREEDY SINGULAR VALUE PURSUIT

The exact computation of the RIP constants  $\delta_s^{\min}$  and  $\delta_s^{\max}$  is combinatorial since it requires an exhaustive enumeration of all sub-matrices  $A_I$  for  $|I| = s$ , which might take an exponential time in  $N$ . We here propose to compute good approximate lower-bounds  $\tilde{\delta}_s^{\min}$  and  $\tilde{\delta}_s^{\max}$  by considering only a small sub-set of the whole set of supports.

The set of all supports is a lattice ordered by inclusion, that is visualized using a graph structure, as depicted in Figure 1. A small sub-set of this lattice is computed by several traversals, starting from the singleton supports  $I$  of size  $|I| = 1$ .

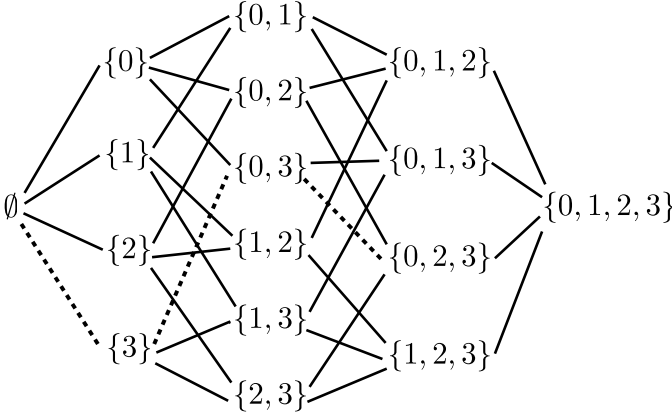


Figure 1. Lattice of the set of supports, for  $N = 4$ . Dashed: a possible path followed by the algorithm to select a matrix  $A_I$  with  $I = \{0, 2, 3\}$ .

A step of the traversal, that follows an edge in the lattice, corresponds to a greedy extension  $I \leftarrow I \cup \{i\}$  computed by adding a properly selected index  $i \notin I$  to increase the size of a support  $I$ . This new index is added so as to maximize the value of  $\delta^{\min}(A_{I \cup \{i\}})$  or  $\delta^{\max}(A_{I \cup \{i\}})$ . The corresponding brute force greedy search is described in Algorithm 1 for the computation of  $\tilde{\delta}_s^{\max}$ . A similar algorithm computes  $\tilde{\delta}_s^{\min}$  by adding at each step the index  $i$  that maximizes  $\delta^{\min}(A_{I \cup \{i\}})$ .

Figure 2, dashed curve, shows examples of the lower bounds  $\tilde{\delta}_s^{\min}$  and  $\tilde{\delta}_s^{\max}$  computed with this brute-force greedy algorithm for several values of  $s$ .

## Algorithm 1: Brute force greedy pursuit.

---

**Initialization:** set  $\mathcal{I}^{(1)} = \{\{0\}, \{1\}, \dots, \{N-1\}\}$  ;  
**for**  $k = 2, \dots, s$  **do**  
    **Initialization:**  $\mathcal{I}^{(k)} = \emptyset$  ;  
    **for**  $I \in \mathcal{I}^{(k-1)}$  **do**  
        Compute  $i^* = \operatorname{argmax}_{i \notin I} \delta^{\max}(A_{I \cup \{i\}})$  ;  
        Set  $\mathcal{I}^{(k)} \leftarrow \mathcal{I}^{(k)} \cup \{I \cup \{i^*\}\}$  ;  
    Set  $k \leftarrow k + 1$ .  
**Return:**  $\tilde{\delta}_s^{\max} = \max_{I \in \mathcal{I}^{(s)}} \delta^{\max}(A_I)$ .

---

## III. BOUNDING THE SINGULAR VALUES

The brute force greedy scheme is still too computationally intensive to be applicable to large scale sampling matrices. This is because the evaluation of the isometry constant  $\delta^{\min}(A_{I \cup \{i\}})$  is required for all candidate extensions  $i \notin I$ . This necessitates the computation of a large amount of minimum and maximum singular values of  $P \times s$  matrices for an increasing value of  $s$ .

To alleviate this computational difficulty, we propose a family of criteria that bounds the minimum and maximum singular values of each sub-matrix indexed by  $I$  by using the "radius" of clustering of its columns, see Proposition 1. A specific instantiation of these bounds leads to one that is easy to compute, and that can be approximately optimized by computing simple inner products with a properly chosen vector. This is made formal in Proposition 2.

**Bounding singular values by clustering.** We define two regions of  $\mathbb{R}^P$  associated to a non-zero vector  $d \in \mathbb{R}^P$

$$\begin{cases} C_d = \{v \in \mathbb{R}^P \mid |\langle d, v \rangle| \geq 1\}, \\ C_d^c = \{v \in \mathbb{R}^P \mid |\langle d, v \rangle| \leq 1\}. \end{cases}$$

**Remark 1.** In the normalized case—columns of  $A$  have unit  $\ell_2$ -norm— $C_d$  and its complement have a nice geometrical interpretation. In such a case,  $C_d$  corresponds to a double-spherical cap, whose radius  $r$  satisfies  $r^2 = 1 - 1/\|d\|^2$ . Similarly,  $C_d^c$  defines a band on the unit sphere in  $\mathbb{R}^P$ .

The following proposition shows that clustering the vectors  $\{a_i\}_{i \in I}$  within these regions allows one to lower-bound the maximum or minimum RIP constants of  $A_I$ .

**Proposition 1.** If  $\{a_i\}_{i \in I} \subset C_d$ , then

$$\delta_s^{\max} \geq s/\|d\|^2 - 1. \quad (6)$$

If  $\{a_i\}_{i \in I} \subset C_d^c$  and if  $d \in \operatorname{Span}(a_i)_{i \in I}$  then

$$\delta_s^{\min} \geq 1 - s/\|d\|^2. \quad (7)$$

*Proof:* We prove (6) and (7) can be proved similarly. The orthogonal projection  $\tilde{d}$  of  $d$  onto  $\operatorname{Span}(a_i)_{i \in I}$  reads  $\tilde{d} = A_I A_I^+ d = (A_I^+)^* A_I^* d$  where  $A_I^+ = (A_I^* A_I)^{-1} A_I^*$  is the pseudo-inverse of  $A_I$ .

Since  $\{a_i\}_{i \in I} \subset C_d$ , we have

$$\forall i \in I, \langle \tilde{d}, a_i \rangle = \langle d, a_i \rangle \geq 1.$$

This shows that

$$\|A_I^* \tilde{d}\|^2 = \sum_{i \in I} |\langle \tilde{d}, a_i \rangle|^2 = \|A_I^* d\|^2 \geq s,$$

and hence

$$\|d\|^2 \geq \|\tilde{d}\|^2 \geq \lambda^{\min}((A_I^+)^*) \|A_I^* d\|^2 \geq \frac{s}{\lambda^{\max}(A_I)}.$$

Given a sub-matrix  $A_I$ , a precise estimate of  $\delta^{\max}(A_I)$  is obtained by maximizing the right hand side of (6). This is achieved by identifying the region  $\mathcal{C}_d$  that encloses the columns of  $A_I$ , and corresponding to the smallest  $\|d\|$ .

**Clustering with appropriate  $d$ .** Finding such an optimal cluster of points is however difficult in large dimension. It is thus desirable to compute an approximate clustering using a well chosen vector  $d$  for the region  $\mathcal{C}_d$ .

In the following, we consider a vector  $d = d(\sigma)$  defined by the vectors  $\{\sigma_i a_i\}_I$ , where

$$\{\sigma_i\}_{i \in I} \in \{+1, -1\}^{|I|}$$

is an appropriate set of signs. The following proposition ensures that this vector can be easily computed.

**Proposition 2.** *For a given set of signs  $\{\sigma_i\}_{i \in I}$  such that  $A_I$  has full rank, the vector*

$$d(\sigma) = A_I(A_I^* A_I)^{-1} \sigma, \quad (8)$$

is such that

$$\forall i \in I, \quad \langle \sigma_i a_i, d(\sigma) \rangle = 1. \quad (9)$$

Any other region  $\mathcal{C}_d$  with another vector  $d$  that satisfies this property leads to a worse lower bound on  $\delta_s^{\max}$ .

*Proof:* By definition  $\langle a_i, d(\sigma) \rangle = \sigma_i$ , and any other vector  $d$  with this property satisfies  $\|d\| > \|d(\sigma)\|$ . Indeed,  $\langle a_i, d - d(\sigma) \rangle = 0$ , and since  $d(\sigma) \in \text{Span}(a_i)_{i \in I}$ , we have  $\langle d(\sigma), d - d(\sigma) \rangle = 0$  implying that  $\|d\|^2 = \|d - d(\sigma)\|^2 + \|d(\sigma)\|^2 > \|d(\sigma)\|^2$ . In view of the right hand side of (6), the conclusion follows. ■

This region  $\mathcal{C}_{d(\sigma)}$  is an optimal choice to estimate  $\delta^{\max}(A_I)$  using (6) if we restrict the choice to regions whose boundary contains the vectors  $\{\sigma_i a_i\}_{i \in I}$ , which corresponds to condition (9). Better estimates might be obtained using another region  $\mathcal{C}_d$  that pass only through a subset of these vectors, or that is defined using different signs, but it is not obvious how to compute them efficiently. Once a set of signs  $\sigma$  is fixed, the vector  $d(\sigma)$  (and corresponding region) is fast to compute as it only requires inverting an overdetermined linear system.

**Remark 2.** *When the columns of  $A$  have unit norm, it is worth noting the following geometrical facts:*

- Any other choice of  $d$  in Proposition 2 leads to a larger spherical cap  $\mathcal{C}_d$ .
- The cap  $\mathcal{C}_{d(\sigma)}$  draws a circle on the unit sphere which is a circumcircle cap since it passes through all the points  $\{\sigma_i a_i\}_{i \in I}$ . The vector  $d(\sigma)$  intersects the circumcircle at its circumcenter.

#### IV. GREEDY PURSUIT USING $d(\sigma)$

Proposition 1, applied to the region  $\mathcal{C}_{d(\sigma)}$  defined from Proposition 2 leads to the following lower bounds on the RIP constants of  $A_I$

$$\begin{cases} \delta_s^{\max} \geq s/\|d(\sigma)\|^2 - 1, \\ \delta_s^{\min} \geq 1 - s/\|d(\sigma)\|^2, \end{cases} \quad (10)$$

and these bounds are expected to be reasonably tight since  $d(\sigma)$  pass through the points  $\{\sigma_i a_i\}_{i \in I}$ .

**Minimal and maximal extensions.** To extract a sub-matrix  $A_I$  with a large isometry constant  $\delta^{\max}(A_I)$ , we thus propose a greedy scheme that iteratively extends both the support  $I$  and the set of signs.

An elementary step of the scheme extends the sign vector  $\sigma$  into

$$\tilde{\sigma} = \sigma + \zeta \Delta_i \quad \text{with} \quad \zeta \in \{+1, -1\},$$

for  $i \notin I$ , where  $\Delta_i$  is the Dirac vector at location  $i$ . The support is thus extended from  $I = I(\sigma)$  to  $\tilde{I} = I(\tilde{\sigma}) = I \cup \{i\}$ . In view of (10), the choice of  $i$  and  $\zeta$  should be made in order to minimize or maximize  $\|d(\tilde{\sigma})\|$ . The following proposition gives essential guidelines to reformulate and solve this optimization problem.

**Proposition 3.** *Let  $\tilde{a}_i \in \text{Span}(a_j, j \in \tilde{I})$  be the dual vector such that*

$$\forall j \in I, \quad \langle \tilde{a}_i, a_j \rangle = 0 \quad \text{and} \quad \langle \tilde{a}_i, a_i \rangle = 1.$$

Then

$$\|d(\tilde{\sigma})\|^2 = C + \|\tilde{a}_i\| |\langle d(\sigma), a_i \rangle - \zeta|,$$

where  $C$  does not depend on  $i$ .

*Proof:* Since  $d(\sigma) \in \text{Span}(a_j, j \in I)$  and  $I \subset \tilde{I}$ , we have

$$\langle d(\sigma) - d(\tilde{\sigma}), d(\sigma) \rangle = 0.$$

Consequently

$$\|d(\tilde{\sigma})\|^2 = \|d(\sigma)\|^2 + \|d(\sigma) - d(\tilde{\sigma})\|^2.$$

We then have that

$$d(\tilde{\sigma}) - d(\sigma) = -\tilde{a}_i (\langle d(\sigma), a_i \rangle - \zeta), \quad \forall j \in \tilde{I},$$

which implies that

$$\|d(\sigma) - d(\tilde{\sigma})\| = \|\tilde{a}_i\| |\langle d(\sigma), a_i \rangle - \zeta|.$$

Finding an extension that maximizes (resp. minimizes)  $\|d(\tilde{\sigma})\|$  is thus equivalent to maximizing (resp. minimizing)  $\|\tilde{a}_i\| |\langle d(\sigma), a_i \rangle - \zeta|$  over both  $i$  and  $\zeta$ . Calculating  $\|\tilde{a}_i\|$  for all possible  $i \notin I$  is computationally demanding since it requires to solve an over-determined system of linear equations for each  $i$ .

We thus select an approximately optimal extension by maximizing or minimizing  $|\langle d(\sigma), a_i \rangle - \zeta|$  instead of  $\|\tilde{a}_i\| |\langle d(\sigma), a_i \rangle - \zeta|$ . This optimization can be solved in closed

form, and defines the extension  $\tilde{\sigma} = \sigma + \zeta \Delta_i$  for the approximation of  $\delta_s^{\max}$  using  $s/\|d(\tilde{\sigma})\|^2 - 1$  where

$$\begin{cases} i = \operatorname{argmin}_{j \notin I(\sigma)} |1 - |\langle d(\sigma), a_j \rangle||, \\ \zeta = \operatorname{sign}(\langle d(\sigma), a_i \rangle). \end{cases} \quad (11)$$

Similarly, the approximation of  $\delta_s^{\min}$  is obtained using

$$\begin{cases} i = \operatorname{argmax}_{j \notin I(\sigma)} |\langle d(\sigma), a_j \rangle|, \\ \zeta = -\operatorname{sign}(\langle d(\sigma), a_i \rangle). \end{cases} \quad (12)$$

**Greedy pursuit with  $d(\sigma)$ .** By exploiting Proposition 3, our iterative scheme suggests to compute a lower bound  $\tilde{\delta}_s^{\max}$  by iteratively extending a support according to (11). This greedy pursuit is summarised in Algorithm 2. A similar algorithm allows one to estimate  $\tilde{\delta}_s^{\min}$  by using the extension rule (12).

---

**Algorithm 2:** (Pruned-Weak)-Greedy pursuit with parameters  $Q$  and  $R$ .

---

**Initialization:** set  $\Sigma^{(1)} = \{\Delta_0, \Delta_1, \dots, \Delta_{N-1}\}$  ;  
**for**  $k = 2, \dots, s$  **do**  
  **Initialization:**  $\Sigma^{(k)} = \emptyset$  ;  
  **for**  $\sigma \in \Sigma^{(k-1)}$  **do**  
    Compute for each  
     $i \in \operatorname{argmin}_{j \notin I(\sigma)}^{[R]} |1 - |\langle d(\sigma), a_j \rangle||$  ;  
    Compute  $\zeta = \operatorname{sign}(\langle d(\sigma), a_i \rangle)$  ;  
    Set  $\Sigma^{(k)} \leftarrow \Sigma^{(k)} \cup \{\sigma + \zeta \Delta_i\}$  ;  
  **Pruning:** set  $\Sigma^{(k)} = \operatorname{argmin}_{\sigma \in \Sigma^{(k)}}^{[Q]} \|d(\sigma)\|$  ;  
  Set  $k \leftarrow k + 1$ .  
**Return:**  $\tilde{\delta}_s^{\max} = \max_{\sigma \in \Sigma^{(s)}} \delta_s^{\max}(A_{I(\sigma)})$ .

---

The algorithm is accelerated by pruning the candidate set  $\Sigma^{(k)}$  at each iteration. This pruning corresponds to the extraction of the  $Q$  signs  $\sigma \in \Sigma^{(k)}$  corresponding to the smallest values of  $\|d(\sigma)\|$ , which is formally written as

$$\operatorname{argmin}_{\sigma \in \Sigma^{(k)}}^{[Q]} \|d(\sigma)\|.$$

This keeps only those signs that generate the largest maximum singular value, as approximately measured by  $s/\|d(\sigma)\|^2$ .

As our algorithm is greedy by nature, the selection rules (11)-(12) may be too stringent, and important candidate extensions can be missed while the algorithm evolves. We then consider the so-called weak selection rule which keeps the  $R$  smallest values of  $|1 - |\langle d(\sigma), a_j \rangle||$  for  $\tilde{\delta}_s^{\max}$ , and largest values of  $|\langle d(\sigma), a_j \rangle|$  for  $\tilde{\delta}_s^{\min}$ . For  $\tilde{\delta}_s^{\max}$ , this is written

$$\operatorname{argmin}_{j \notin I(\sigma)}^{[R]} |1 - |\langle d(\sigma), a_j \rangle||,$$

and similarly for  $\tilde{\delta}_s^{\min}$ . Corresponding variants of greedy pursuit are called weak greedy pursuit with parameter  $R$ .

**Empirical evaluation of our greedy pursuit bound.** To assess the performance of our greedy pursuit, Algorithm 2

(with  $R = Q = 1$ ), we compare it with the brute force pursuit, Algorithm 1, that is expected to perform better since at each of its step, it maximizes the RIP constants. Figure 2 shows that this is indeed the case, but the gap between the two estimates of the RIP constants provided by the two methods is rather small. This proves numerically that the heuristic (10) is accurate in practice.

We also compared the performance of the different variants of Algorithm 2: greedy pursuit with no pruning ( $Q = N$ ,  $R = 1$ ), and pruned weak greedy pursuit with  $Q = N/4$  and  $R = 4$ . The increase in performance brought by the pruned weak greedy variant is slightly more salient as the sparsity level increases.

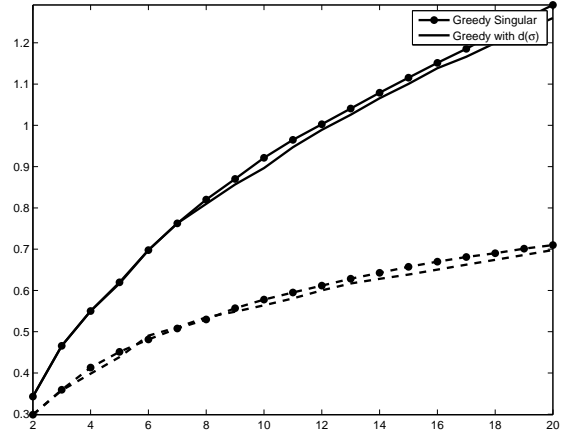


Figure 2. Comparison of the brute force pursuit and our greedy pursuit with  $Q = N$ ,  $R = 1$  for  $(N, P) = (2000, 500)$ . Solid and dashed lines correspond respectively to  $\tilde{\delta}_s^{\max}$  and  $\tilde{\delta}_s^{\min}$ , as a function of  $s$  on the  $x$  axis.

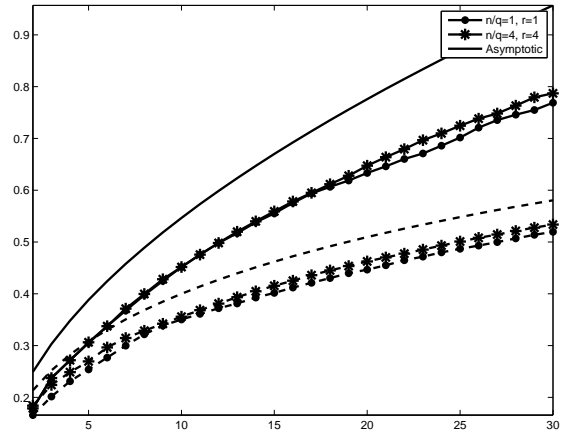


Figure 3. Comparison of the different variants of Algorithm 2 for  $(N, P) = (8000, 2000)$ . Circles: greedy pursuit  $Q = R = 1$ . Asterisks: pruned weak greedy pursuit  $Q = N/4$ ,  $R = 4$ . The curve without markers corresponds to the asymptotic upper bound of [5]. Solid and dashed lines correspond respectively to  $\tilde{\delta}_s^{\max}$  and  $\tilde{\delta}_s^{\min}$ , as a function of  $s$  on the  $x$  axis.

Figure 4 shows that for a fixed value of  $\eta = P/N$  and  $s/P = 10^{-2}$ , the estimates of the RIP constants provided by Algorithm 2 are close to being constant when the size ( $P = 100s, N = P/\eta$ ) of the sensing matrix varies. This

$P$	250	500	1000	2000
$s_0^*(1/4, P)$	2	3	5	8
$\lceil \rho_0(1/4)P \rceil$	1	2	3	6

Table I

OUR NUMERICALLY COMPUTED CRITICAL SPARSITY LEVELS  $s_0^*(\eta, P)$  VERSUS THE THEORETICAL UPPER-BOUND OF [5]  $\rho_0(1/4)P \sim 0.0027P$ .

is consistent with the asymptotic upper bound of restricted isometry constants provided by Blanchard et al. [5]. A slight drift is however observed as the sparsity increases, which might be a consequence of the greedy nature of the algorithm. This numerical result tends to prove that the existence of ill-conditioned sub-matrices at such small sparsity levels is not restricted to small dimensions.

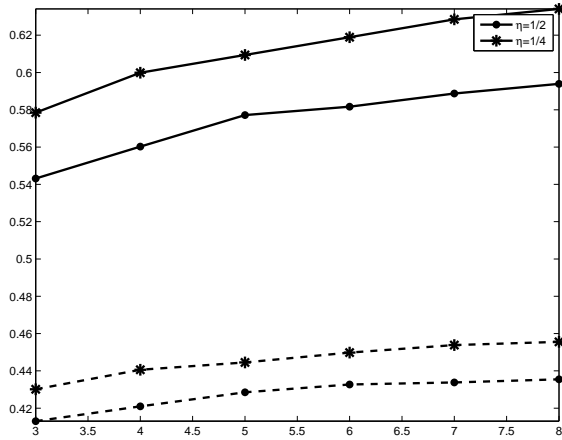


Figure 4. Plot of  $\tilde{\delta}_s^{\max}$  (solid curves) and  $\tilde{\delta}_s^{\min}$  (dashed curves) as a function of  $s$  on the  $x$  axis, for two values of  $\eta = P/N$  and for  $P = 100s$ . The curves are obtained by averaging the value of  $\delta_s^{\max}$  and  $\delta_s^{\min}$  for 5 realizations of the random matrix  $A$ .

**Empirical sparsity bounds for RIP condition.** For each underdeterminacy value  $\eta = P/N \leq 1$ , we computed  $s_0^*(\eta, P)$ , the minimum sparsity  $s$  for which our empirical estimates invalidate condition (4), hence  $\ell^1$ -identifiability,

$$(4\sqrt{2} - 3)\tilde{\delta}_{2s}^{\min} + \tilde{\delta}_{2s}^{\max} \geq 4(\sqrt{2} - 1). \quad (13)$$

Figure 5 depicts our numerical estimate of the bound (4) for varying  $s$ . Table I reports our numerically computed critical sparsity levels  $s_0^*(\eta, P)$  for  $\eta = 1/4$ , and compares this numerical evidence with the theoretical bound of Blanchard et al. [5]  $\rho_0(1/4) \sim 0.0027$ .

**Computational speed.** A chief advantage of our (weak) greedy algorithm over brute force pursuit is that it has a much lowest computational load while leading to comparable estimates of the RIP constants. This is clearly testified by the execution times reported in Table II for two typical problem sizes. Note also that the pruned weak greedy variant is faster than the greedy version owing to the pruning step.

## V. CONCLUSION

We have proposed in this paper a new greedy algorithm to find sub-matrices with a small number of columns that are

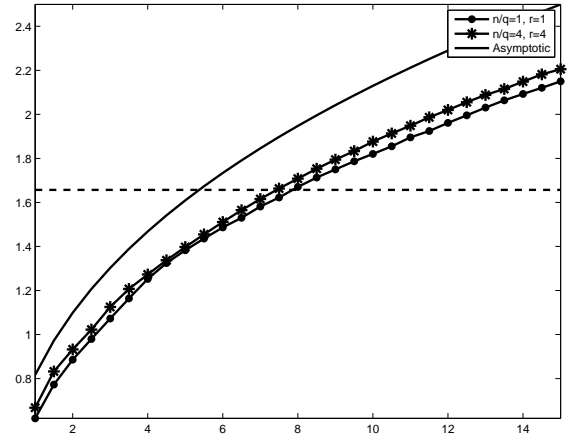


Figure 5. Display of  $(4\sqrt{2} - 3)\tilde{\delta}_{2s}^{\min} + \tilde{\delta}_{2s}^{\max}$  for  $(N, P) = (8000, 2000)$ , as a function of  $s$  on the  $x$  axis, computed using our greedy pursuit with no pruning (circles) and pruning (asterisks,  $Q = N/4$ ). The solid curve corresponds to the asymptotic upper bound of [5]. The dashed line corresponds to the limit  $y = 4(\sqrt{2} - 1)$  below which condition (13) is in force.

$N = 800, P = 200, s = 10$

Algorithm	Time (s)	$\tilde{\delta}_s^{\max}$
Singular value pursuit	2145	0.596
Greedy pursuit ( $Q = N, R = 1$ )	11.1	0.595
Pruned weak greedy pursuit ( $N/Q = 10, R = 4$ )	6.0	0.597
Pruned weak greedy pursuit ( $N/Q = 100, R = 4$ )	5.1	0.597

$N = 4000, P = 1000, s = 30$

Algorithm	Time (s)	$\tilde{\delta}_s^{\max}$
Singular value pursuit	—	—
Greedy pursuit ( $Q = N, R = 1$ )	356	0.627
Pruned weak greedy pursuit ( $N/Q = 10, R = 4$ )	113	0.634
Pruned weak greedy pursuit ( $N/Q = 100, R = 4$ )	66.3	0.629

Table II

COMPARISON OF COMPUTATION TIMES FOR THE EVALUATION OF  $\tilde{\delta}_s^{\max}$ .

ill-conditioned. This allows us to check numerically sparsity-based criteria for compressed sampling recovery based on the restricted isometry property.

## REFERENCES

- [1] D. Donoho, "Compressed sensing," *IEEE Trans. Info. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [2] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Info. Theory*, vol. 52, no. 2, pp. 489–509, 2006.
- [3] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *Compte Rendus de l'Academie des Sciences, Paris, Série I*, no. 346, pp. 589–592, 2008.
- [4] S. Foucart and M.-J. Lai, "Sparsest solutions of underdetermined linear systems via  $\ell_q$ -minimization for  $0 < q \leq 1$ ," to appear in *Applied and Computational Harmonic Analysis*, 2009.
- [5] J. Blanchard, C. Cartis, and J. Tanner, "The restricted isometry property and  $\ell^q$ -regularization: Phase transitions for sparse approximation," *Preprint*, 2009.
- [6] D. L. Donoho, "High-dimensional centrally symmetric polytopes with neighborliness proportional to dimension," *Discrete & Computational Geometry*, vol. 35, no. 4, pp. 617–652, 2006.