A Numerical Exploration of Compressed Sampling Recovery

Charles Dossal, Gabriel Peyré, Jalal M. Fadili

A Numerical Exploration of Compressed Sampling Recovery

Charles Dossal[a], Gabriel Peyré[b], Jalal Fadili[c]

[a]*IMB Université Bordeaux 1,*
351, cours de la Libération F-33405 Talence cedex, France
[b]*CNRS and CEREMADE, Université Paris-Dauphine,*
Place du Maréchal De Lattre De Tassigny, 75775 Paris Cedex 16, France
[c]*GREYC, CNRS-ENSICAEN-Université Caen,*
6 Bd du Maréchal Juin 14050 Caen Cedex, France

Abstract

This paper explores numerically the efficiency of ℓ^1 minimization for the recovery of sparse signals from compressed sampling measurements in the noiseless case. This numerical exploration is driven by a new greedy pursuit algorithm that computes sparse vectors that are difficult to recover by ℓ^1 minimization. The supports of these pathological vectors are also used to select sub-matrices that are ill-conditionned. This allows us to challenge theoretical identifiability criteria based on polytopes analysis and on restricted isometry conditions. We evaluate numerically the theoretical analysis without resorting to Monte-Carlo sampling, which tends to avoid worst case scenarios. [1]

Key words: Compressed sensing, ℓ^1 minimization, restricted isometry constant, polytopes.

1. Introduction

1.1. Compressed Sampling

Compressed sampling acquisition. Compressed sampling is a new sampling theory that uses a fixed set of linear measurements together with a non-linear reconstruction. For the recovery of a signal from a small number of measurements to be efficient, compressed sampling makes use of a sampling operator that is drawn from certain random distributions.

The idea of performing randomized compressed acquisition of signals was introduced independently by Candès et al. [1] and Donoho [2]. This emerging sampling paradigm could have far reaching applications in several fields where data acquisition is slow and costly, such as medical imaging [3] or astronomical imaging [4]. It is thus important to better understand the theoretical guarantees of perfect recovery that randomized acquisition can offer.

[1]This work is supported by ANR grant NatImages ANR-08-EMER-009.

Email addresses: charles.dossal@math.u-bordeaux1.fr (Charles Dossal), gabriel.peyre@ceremade.dauphine.fr (Gabriel Peyré), jalal.fadili@greyc.ensicaen.fr (Jalal Fadili)

ℓ^1 *recovery.* The sampling operator computes the projection of the data $x \in \mathbb{R}^N$ on a set of $P \ll N$ sampling vectors, which can be written as a matrix-vector multiplication $y = Ax$ where $A \in \mathbb{R}^{P \times N}$ is the sampling matrix.

The signal x can be recovered either exactly or accurately from these measurements y by exploiting its exact or approximate sparsity, which means that all entries x but a few are zero or small. With a proper change of basis, this is extended to signals which are sparse in an orthogonal basis, such as a wavelet basis for natural images.

Throughout this paper, we consider the case where the entries of $A = (a_i)_{i=0}^{N-1} \in \mathbb{R}^{P \times N}$ are independent and identically distributed (iid) $\mathcal{N}(0, 1/P)$. We note that many theoretical recovery results extends to more general random distributions such as Bernoulli matrices, random projectors [5] or partial Fourier measurements [6]. The greedy algorithm described in this paper is expected to work also well for these distributions. In particular, the heuristics developed in this paper are expected to be accurate for distributions that are exactly or approximately invariant under rotation, which holds true when all entries of A are drawn independently from the gaussian distribution.

1.2. Identifiability

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

$$\min_{\tilde{x} \in \mathbb{R}^N} \|\tilde{x}\|_1 \quad \text{subj. to} \quad A\tilde{x} = y, \qquad \text{where} \quad \|\tilde{x}\|_1 = \sum_i |\tilde{x}_i|. \tag{1}$$

Such an ℓ^1 recovery program has been introduced by Chen, Donoho and Saunders under the name of Basis Pursuit (BP) for sparse coding [7]. The vector x is said to be identifiable if the solution $x^\star$ to (1) is unique and coincides exactly with x.

The optimization problem (1) can be recast as a linear program, which can be solved using iterative methods such as interior points algorithms [7] or the Douglas Rachford algorithm [8, 9].

Other recovery methods with theoretical recovery guarantees exist. Basis pursuit denoising [7] is able to cope with noisy measurements, and can be solved for instance with proximal iterative thresholding algorithms [10, 11, 12, 13], accelerated iterative thresholding [14, 15] and the Nesterov algorithm [16]. Greedy algorithms such as CoSaMP [17] have been proved to work under conditions similar to ℓ^1 minimization. Non-convex regularization using ℓ^p functionals for $p < 1$ can be solved approximately using re-weighting schemes [18, 19], and can improve numerically the recovery.

Remark 1. *Since we are interested in the recovery with ℓ^1 minimization from noiseless measurements, the identifiability of x depends only on* $\operatorname{sign}(x)$*. In the following, and without loss of generality, we consider vectors $x \in \mathbb{R}^N$ such that the entries $x_i \in \{-1, 0, 1\}, \forall\, i$.*

2. Criteria for Identifiability

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

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

With high probability on the sampling matrix A, a sparsity-based recovery criterion asserts that any vector satisfying

$$\|x\|_0 = \#\{i \setminus x_i \neq 0\} \leqslant \rho(P/N)P \tag{2}$$

is identifiable for $\rho(\eta) > 0$, where $\eta = P/N < 1$ corresponds to the undersampling rate, so that $\eta^{-1} > 1$ is the redundancy of A.

For a given undersampling ratio $\eta < 1$, a critical question is to know as precisely as possible the value of $\rho(\eta)$, since this value gives a threshold (possibly in a worst case setting) below which exact or accurate recovery by ℓ^1-minimization is theoretically guaranteed. Many recovery criteria have been proposed in the literature. They lead to different estimates of $\rho(\eta)$, possibly with robustness to noise or imperfect sparsity.

2.1. Deterministic Necessary Conditions

Generic deterministic necessary conditions based on the mutual coherence of the matrix A where introduced by several authors, see for instance [20, 21, 22, 23, 24, 25]. They usually lead to overly pessimistic sparsity bounds, especially for random matrices.

These necessary recovery conditions are refined by considering not only the sparsity $\|x\|_0$ but also the support and the sign pattern of the non-zero entries of x indexed by the support $I(x)$ of x. Such criteria use the interactions between the columns of $A_I = (a_i)_{i\in I}$ and the other ones $(a_i)_{i\notin I}$, where the sub-matrix A_I is the restriction of A to the columns indexed by $I(x)$. Fuchs [26] proved that a sufficient condition for x to be identifiable is

$$F(x) = \max_{i\notin I} |\langle a_i,\, d(x)\rangle| < 1 \tag{3}$$

$$\text{where} \quad d(x) = A_I(A_I^* A_I)^{-1}\,\mathrm{sign}(x_I), \tag{4}$$

see also Tropp [27] for a similar result.

Wainwright [28] considers a condition of the form (3) to ensure sparsity pattern recovery from noisy measurements by solving the penalized ℓ^1 optimization (the so-called Basis Pursuit DeNoising, BPDN [7]). He also established that violation of (3) is sufficient for failure of the BPDN in recovering the support set. This analysis was specialized to the case of standard gaussian sensing matrices to derive sharp sufficient conditions of identifiability of the optimal decoder, as well as necessary conditions that any recovery procedure should satisfy to guarantee identifiability of sparse enough vectors[29].

2.2. Restricted Isometry Based Criteria

The seminal work of Donoho [2] and Candès et al. [1, 30] has focused on the stability of the compressed sampling decoder. This analysis leads to an estimation of $\rho(P/N)$ which is nearly constant up to a logarithmic term.

Candès et al. [1, 30] introduced the restricted isometry property (RIP), with the RIP constants $0 < \delta_s^{\min} \leqslant \delta_s^{\max} < 1$. These constants are the smallest numbers such that for every vector $x \in \mathbb{R}^N$ with $\|x\|_0 \leqslant s$,

$$(1 - \delta_s^{\min})\|x\|^2 \leqslant \|Ax\|^2 \leqslant (1 + \delta_s^{\max})\|x\|^2. \tag{5}$$

Condition (5) is equivalent to saying that for all I such that $|I| = s$, the smallest and largest eigenvalues $\lambda^{\min}(A_I^* A_I)$ and $\lambda^{\max}(A_I^* A_I)$ of the Gram matrix $A_I^* A_I$ are respectively bounded below and above by $1 - \delta_s^{\min}$ and $1 + \delta_s^{\max}$. Thus, the RIP constants are equivalently defined as

$$\delta_s^{\min} = \max_{|I|=s} \delta^{\min}(A_I) \quad \text{and} \quad \delta_s^{\max} = \max_{|I|=s} \delta^{\max}(A_I) \tag{6}$$

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

The original work of Candès et al. [1] considers a symmetric RIP constant $\delta_s = \max(\delta_s^{\max}, \delta_s^{\min})$. These authors proved that a small enough value of δ_{2s} ensures identifiability of all s-sparse vectors. For instance, it is proved in [31] that $\delta_{2s} \leqslant \sqrt{2} - 1$ guarantees identifiability of all s-sparse vectors. This is achieved with high probability on A if $s \leqslant CP/\log(N/P)$, which corresponds to condition (2) with $\rho(\eta) \leqslant C/\log(\eta^{-1})$ with $\eta^{-1} = N/P$ the redundancy of the matrix A.

It turns out that the largest and smallest eigenvalues $\lambda^{\min}(A_I^* A_I)$ and $\lambda^{\max}(A_I^* A_I)$ do not deviate from 1 at the same rate. Using asymmetric RIP constants, Foucart and Lai [32] have shown that

$$(4\sqrt{2} - 3)\delta_{2s}^{\min} + \delta_{2s}^{\max} < 4(\sqrt{2} - 1) \tag{7}$$

implies identifiability of all s-sparse vectors. Blanchard et al. [33] determine ρ_0 such that with high probability on A

$$\|x\|_0 \leqslant \rho_0(P/N)P \tag{8}$$

ensures that condition (7) is in force. Condition (8) 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 $\rho_0(1/4) = 0.0027$.

2.3. Topologically-based Criteria

Donoho [34, 35] gave a topological necessary and sufficient condition for the identifiability of a vector by considering the lower-dimensional projection $A(B_1)$ of the ℓ^1 ball

$$B_1 = \{\tilde{x} \setminus \|\tilde{x}\|_1 \leqslant 1\} .$$

The centro-symmetric polytope $A(B_1)$ is the image of the ℓ^1 ball, and is also the convex hull of $\{\pm a_i\}_i$. The $\|x\|_0$-dimensional face $f_x \subset A(B_1)$ selected by x is the convex hull of $\{\text{sign}(x_i)a_i\}_{i \in I}$. Donoho [34] showed that

$$x \text{ is identifiable} \quad \Longleftrightarrow \quad f_x \in \partial A(B_1) , \tag{9}$$

where $\partial A(B_1)$ is the boundary of the polytope $A(B_1)$. Dossal [36] proved that this topological condition is equivalent to having x as the limit of x_n where $F(x_n) < 1$, where F is defined in (3). This can be interpreted as x being in the closure of the set of all vectors satisfying (3).

Using (9), Donoho and Tanner [34, 37] determine, in the noiseless case $y = Ax$, a precise asymptotic value for $\rho(\eta)$ in (2) when P and N tend to infinity. For instance $\rho(1/2) \approx 0.089$ and $\rho(1/4) \approx 0.065$.

The function $\rho(\eta)$ induced by the topological condition (9) is sharper than $\rho_0(\eta)$ defined in (8)obtained from the RIP-based condition (7). This can be interpreted in the light of several arguments. First, $\rho(\eta)$ originates from an asymptotic analysis performed under a union bound, while the topological analysis gives a sharp asymptotic bound. Furthermore, condition (9) exploits fully the exact geometry of ℓ^1 minimization, while the RIP condition is more flexible and applies to other sparse recovery schemes including non-convex ℓ_p-minimization and greedy pursuit, see for instance [17]. However, the RIP condition ensures robustness to compressible signals and stability to noise if BP is replaced with a noise-aware variant [1], while condition (9) does not imply such a stability property.

2.4. Numerical Evaluation of Recovery Criteria

The numerical evaluation of compressed sensing is usually performed by Monte-Carlo sampling over the set of s-sparse vectors. Numerous simulations suggest that recovery of most s-sparse vectors is obtained for values of s/P that can be as large as $1/4$ for reasonable problem sizes (N, P), see for instance [38]. We intend to perform a worse case analysis which Monte-Carlo sampling fails to capture.

The asymptotic evaluation of $\rho(P/N)$ performed by Donoho and Tanner [34, 37] suggests that these Monte-Carlo simulations are far from capturing the true value of $\rho(P/N)$. This is because pathological sparse vectors that are not identifiable are difficult to find by random sampling. In the noiseless case, an asymptotic explanation for these over-optimistic Monte-Carlo simulations is given by Tanner and Donoho [37], that analyze the recovery of *almost all* sparse vectors, see also [39] for simulations with matrices drawn from several random ensembles.

The situation for recovery conditions based on restricted isometry constants is even worse, since they are intractable to compute exactly. Although no exact asymptotic for these conditions is known, the careful asymptotic analysis of Blanchard et al. [33] suggests that the RIP leads to small estimates for $\rho(P/N)$.

The work of Juditsky and Nemirovski studies a necessary and sufficient condition that ensures recovery of all s-sparse vectors [40]. They also show that their analysis lead to verifiable sufficient conditions for identifiability by solving BP. A numerical exploration is carried out for small scale problems and with values of $\eta = P/N$ close to one. Instead of testing the RIP condition (5) which is combinatorial, d'Aspremont and El Ghaoui [41] propose a semidefinite convex relaxation to derive a bound on the nullspace property of a matrix A. This allowed them to compute numerical bounds of critical sparsity levels ensuring exact recovery. Our work is complementary to these approaches, because we study heuristics that lead to fast greedy algorithms that enable an exploration in high dimension and for high or low redundancies.

2.5. Contributions

This paper studies both topological and RIP conditions, and find non-asymptotic upper bounds on the sparsity conditions obtained by both approaches. We show numerically that the bounds provided by both approaches are quite sharp in a non-asymptotic regime.

Our main contribution is a new greedy pursuit algorithm that can be used to challenge both kinds of conditions. For example, for $(N, P) = (4000, 1000)$, the algorithm computes a non-identifiable vector x such that $\|x\|_0 = 79$. The algorithm also reveals that $\delta_{10}^{\max} \geqslant 0.58$ and $\delta_{10}^{\min} \geqslant 0.42$ so that condition (7) is not fulfilled even for $s = 5$.

This is the first time a numerical scheme leads to such a conclusion, mainly because previous experimental studies where based on Monte-Carlo sampling of sparse vectors. Such randomized numerical experiments tend to avoid pathological cases, and are thus far from reaching the theoretical bounds. For instance, using the distribution of the eigenvalues of a Wishart matrix, it can be shown that the probability that a random sub-matrix A_I of $|I| = 10$ columns satisfies $\lambda^{\max}(A_I^* A_I) \geqslant 1.58$ is less than 4×10^{-6}, whereas our algorithm is able to find such a sub-matrix.

3. Greedy Singular Value Pursuit

The problem of computing lower bounds on s-restricted isometry constants corresponds to selecting support I of size $s = |I|$ so that the matrix A_I is ill-conditionned. Before detailing in Section 5 fast algorithms that select this support I as the support $I = I(x)$ of an appropriate signed

vector x, we show in this section a brute force greedy scheme that directly computes the support I.

3.1. Support Extensions

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. Here we 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 for $N = 4$. A small sub-set of this lattice is computed by several traversals, starting from the singleton supports I of size $|I| = 1$. Figure 1 shows in dashed line an example of such a traversal. The idea behind our numerical scheme is to select carefully this traversal.

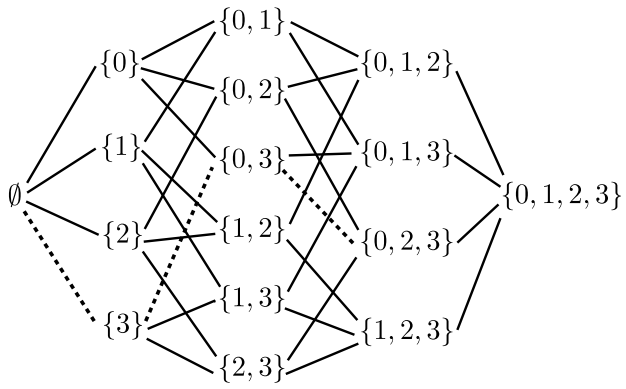

Figure 1: *Lattice of the set of supports, for $N = 4$. Dashed: a possible path followed by the algorithm to select a matrix A_I with $I = \{0, 2, 3\}$.*

3.2. Greedy Pursuit Algorithm

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. According to (6), 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 singular value pursuit 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 4 shows examples of the lower bounds $\tilde{\delta}_s^{\min}$ and $\tilde{\delta}_s^{\max}$ computed with this brute-force greedy singular value pursuit for several values of s.

4. Interior Facets and Non-Identifiable Vectors

The brute force greedy scheme is computationally too intensive to be applicable to large scale sampling matrices. This is because the evaluation of the isometry constants $\delta^{\min}(A_{I\cup\{i\}})$ or $\delta^{\max}(A_{I\cup\{i\}})$ is required for all candidate extensions $i \notin I$. This necessitates the computation of

Algorithm 1: Greedy singular value pursuit.

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

a large amount of minimum and maximum singular values of $P \times s$ matrices for an increasing value of s.

Furthermore, the direct extension of supports is suitable to detect ill-conditioned sub-matrices, but does not make sense to compute sparse non-identifiable vectors $x \in \mathbb{R}^N$. Indeed, this requires not only the computation of a support I, but also the optimization of a sign $x_i \in \{+1, -1\}$ for each $i \in I = I(x)$ to make the vector as difficult as possible to identify.

To address both issues, we derive in this section two heuristics that indicate whether x is a good candidate for both non-identifiability and to select an ill-posed sub-matrix $A_{I(x)}$.

4.1. An Heuristic for Identifiability

From (9) we deduce that a non identifiable vector x corresponds to a face f_x belonging to the interior of the polytope $A(B_1)$. In other words, the distance of the face f_x to the center of the polytope is a good indicator of identifiability. This distance is stated in the following result.

Proposition 1. *For any vector x such that $\operatorname{rank}(A_I) = |I|$ for $I = I(x)$, the distance from the face f_x to 0 is $\frac{1}{\|d(x)\|_2}$, where $d(x) = A_I(A_I^* A_I)^{-1} \operatorname{sign}(x_I)$.*

Proof. Distance of f_x to the 0 is the maximum of the distance between any hyperplane $\mathcal{H}$ containing f_x and 0. The definition of $d(x)$ implies that $A_I^* d(x) = \operatorname{sign}(x_I)$, which in turn yields $\langle d(x), (\operatorname{sign} x_i) a_i \rangle = 1$ for all $i \in I$. The hyperplane

$$\mathcal{H}_x = \{u \setminus \langle d(x), u \rangle = 1\}$$

is such that for all $i \in I$, $(\operatorname{sign} x_i) a_i \in \mathcal{H}_x$ and thus $f_x \subset \mathcal{H}_x$. The distance between $\mathcal{H}_x$ and 0 is $1/\|d(x)\|_2$.
Let $\mathcal{H}_1 = \{u \setminus \langle c, u \rangle = 1\}$ be another hyperplane such that $(\operatorname{sign} x_i) a_i \in \mathcal{H}_1$, for all $i \in I$. The distance between $\mathcal{H}_1$ and 0 is $\frac{1}{\|c\|}$. For all $i \in I$, we have $\langle c, a_i \rangle = \langle d(x), a_i \rangle$ and thus $\langle c - d(x), a_i \rangle = 0$. Since $d(x) \in \operatorname{Span}(a_i)_{i \in I}$, $\langle c - d(x), d(x) \rangle = 0$ and then $\|c\|^2 = \|c - d(x)\|^2 + \|d(x)\|_2^2 > \|d(x)\|_2^2$, which completes the proof. □

Figure 2 illustrates this proposition for $P = 2$ and $N = 3$. This property, together with condition (9), suggests that a vector x having a small value of $1/\|d(x)\|_2$ is more likely to be non-identifiable. This heuristic is particularly relevant when the matrix is random and invariant under rotation, which is true when all entries of A are iid gaussian.

Figure 3 estimates with a Monte-Carlo sampling the ratio of vectors that are identifiable, according to the sparsity $\|x\|_0$ and to a quantized value of $\|d(x)\|_2$. The curve parameterized by

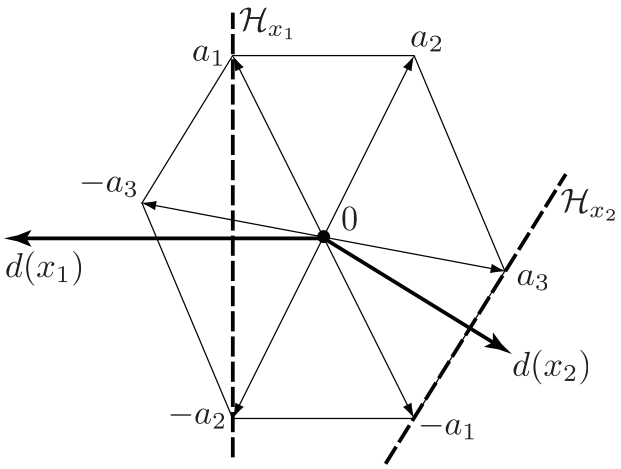

Figure 2: *Geometry of ℓ^1 recovery, for $N = 3$ and $P = 2$. The vector $x_1 = (2, -3, 0)$ is not identifiable because f_{x_1} is inside the polytope $A(B_1)$, and has a large $\|d(x_1)\|$. On the contrary, $x_2 = (-5, 0, 3)$ is identifiable because $f_{x_1} \in \partial A(B_1)$, and it has a small $\|d(x_1)\|$.*

$\|d(x)\|_2$ exhibits a phase transition that is even sharper than the curve parameterized by sparsity (each dot on the curves accounts for 1000 random realizations of the signal).

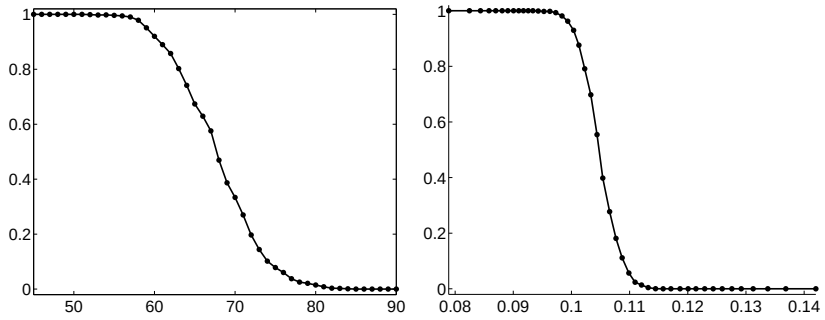

Figure 3: *Left: ratio of identifiable vectors as a function of $\|x\|_0$, for $(P, N) = (250, 1000)$. Right: ratio of identifiable vectors as a function of $\|d(x)\|_2$.*

The numerical evidence provided by Figure 3 suggests that non-identifiable vectors might be found not just by increasing the sparsity of a given vector, but also by decreasing the value of $1/\|d(x)\|_2$.

4.2. An Heuristic for Matrix Conditioning

Bounding Singular Values by Clustering. We define two regions of $\mathbb{R}^P$ associated to a non-zero vector $d \in \mathbb{R}^P$

$$\begin{cases} C_d = \left\{ v \in \mathbb{R}^P \setminus |\langle d, v \rangle| \geqslant 1 \right\}, \\ C_d^c = \left\{ v \in \mathbb{R}^P \setminus |\langle d, v \rangle| \leqslant 1 \right\}. \end{cases}$$

Remark 2. *In the normalized case—columns of A have unit ℓ^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 bound below the maximum or minimum RIP constants.

Proposition 2. *If $\{a_i\}_{i\in I} \subset C_d$, then*

$$\delta_s^{\max} \geqslant s/\|d\|^2 - 1. \tag{10}$$

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

$$\delta_s^{\min} \geqslant 1 - s/\|d\|^2. \tag{11}$$

Proof. We prove (10) and (11) can be proved similarly. The orthogonal projection $\tilde{d}$ of d onto $\mathrm{Span}(a_i)_{i\in I}$ 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

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

This shows that

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

and hence

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

□

Remark 3. *Given a sub-matrix A_I, a precise estimate of $\delta^{\max}(A_I)$ is obtained by maximizing the right hand side of* (10)*. This is achieved by identifying the region 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 high dimension. It is thus desirable to compute an approximate clustering based on a well chosen vector d for the region C_d.

In the following, we use the set of signs

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

so that $I = I(x)$, and we guide our choice of d based on the signed vectors $\{x_i a_i\}_I$. In particular, we require that the boundary of the cap C_d passes through all these signed vectors, which corresponds to

$$\forall\, i \in I, \quad \langle x_i a_i,\, d(x)\rangle = 1.$$

The following proposition shows that the vector $d(x)$, already introduced in (4) is a reasonable choice that matches this constraint.

Proposition 3. *For a given set of signs $\{x_i\}_{i\in I}$ such that A_I has full rank, the vector*

$$d(x) = A_I(A_I^* A_I)^{-1} x, \tag{12}$$

satisfies

$$\forall\, i \in I, \quad \langle x_i a_i,\, d(x)\rangle = 1. \tag{13}$$

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

Proof. By definition $\langle a_i,\, d(x)\rangle = x_i$, and any other vector d with this property satisfies $\|d\| > \|d(x)\|_2$. Indeed, $\langle a_i,\, d - d(x)\rangle = 0$, and since $d(x) \in \mathrm{Span}(a_i)_{i\in I}$, we have $\langle d(x),\, d - d(x)\rangle = 0$ implying that $\|d\|^2 = \|d - d(x)\|^2 + \|d(x)\|_2^2 > \|d(x)\|_2^2$. In view of the right hand side of (10), the conclusion follows. □

This region $C_{d(x)}$ is an optimal choice to estimate $\delta^{\max}(A_I)$ using (10) if we restrict the choice to regions whose boundary contains the vectors $\{x_i a_i\}_{i\in I}$, which corresponds to condition (13). Better estimates might be obtained using another region 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 in high dimension. Once a set of signs x is fixed, the vector $d(x)$ (and corresponding region) is fast to compute as it only requires inverting an overdetermined linear system.

Remark 4. *When the columns of A have unit norm, it is worth noting the following geometrical facts:*

- *Any other choice of d in Proposition 3 leads to a larger spherical cap C_d.*
- *The cap $C_{d(x)}$ draws a circle on the unit sphere which is a circumcircle cap since it passes through all the points $\{x_i a_i\}_{i\in I}$. The vector $d(x)$ intersects the circumcircle at its circumcenter.*

5. Greedy Pursuits Using $d(x)$

Proposition 2 together with Proposition 3 suggest that efficient bounds on the restricted isometry constants are obtained by finding a sparse vector x that maximizes or minimizes $1/\|d(x)\|_2$. Similarly, Proposition 1 suggests that minimizing $1/\|d(x)\|_2$ is a good strategy to search for sparse non-identifiable vectors. This section shows how the minimization or maximization of $1/\|d(x)\|_2$ can be performed approximately using greedy algorithms.

5.1. 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 x into

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

for $i \notin I$, where Δ_i is the Dirac vector at location i. The support is thus extended from $I = I(x)$ to $\tilde{I} = I(\tilde{x}) = I \cup \{i\}$. In view of Propositions 2 and 3 (see (16)), the choice of i and ζ should be made in order to minimize or maximize $\|d(\tilde{x})\|$. The following proposition gives essential guidelines to reformulate and solve this optimization problem.

Proposition 4. *Let $\tilde{a}_i \in \mathrm{Span}(a_j,\, j \in \tilde{I})$ be the dual vector such that*

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

Then

$$\|d(\tilde{x})\|^2 = \|d(x)\|_2^2 + \|\tilde{a}_i\|^2 |\langle d(x),\, a_i \rangle - \zeta|^2\ .$$

Proof. Since $d(x) \in \mathrm{Span}(a_j,\, j \in I)$ and $I \subset \tilde{I}$, we have

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

Consequently

$$\|d(\tilde{x})\|_2^2 = \|d(x)\|_2^2 + \|d(x) - d(\tilde{x})\|_2^2.$$

Moreover, we have

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

which implies that

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

□

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

We thus select an approximately optimal extension by maximizing or minimizing $|\langle d(x),\, a_i \rangle - \zeta|$ instead of $\|\tilde{a}_i\| |\langle d(x),\, a_i \rangle - \zeta|$. This optimization can be solved in closed form, and defines the extension $\tilde{x} = x + \zeta^+ \Delta_{i^+}$ that maximizes $1/\|d(x)\|_2$

$$\begin{cases} i^+ = \underset{j \notin I(x)}{\mathrm{argmin}}\ |1 - |\langle d(x),\, a_j \rangle||, \\ \zeta^+ = \mathrm{sign}(\langle d(x),\, a_{i^+} \rangle). \end{cases} \tag{14}$$

Similarly, the extension $\tilde{x} = x + \zeta^- \Delta_{i^-}$ that minimizes $1/\|d(x)\|_2$

$$\begin{cases} i^- = \underset{j \notin I(x)}{\mathrm{argmax}}\ |\langle d(x),\, a_j \rangle|, \\ \zeta^- = -\,\mathrm{sign}(\langle d(x),\, a_{i^-} \rangle). \end{cases} \tag{15}$$

5.2. Greedy Pursuit Algorithms

Starting from an initial candidate set of 1-sparse vectors $\Sigma_{\max}^{(1)} = \{\Delta_0, \Delta_1, \ldots, \Delta_{N-1}\}$, the greedy pursuit algorithm for maximizing $1/\|d(x)\|_2$ builds a candidate set $\Sigma_{\max}^{(k)}$ for each sparsity $k \leqslant s$. The algorithm iteratively applies the extension (14) to each $x \in \Sigma_{\max}^{(k-1)}$ to obtain $\Sigma_{\max}^{(k)}$.

As our algorithm is greedy by nature, the selection rule (14) may be too stringent, and important candidate extensions can be missed while the algorithm evolves. We thus use a weak greedy selection rule which keeps the $R \geqslant 1$ best candidate extensions. This selection of the best R extensions of a given $x \in \Sigma_{\max}^{(k-1)}$ is written

$$\underset{j \notin I(x)}{\mathrm{argmin}^{[R]}}\ |1 - |\langle d(x),\, a_j \rangle|,$$

where the notation $\text{argmin}^{[R]}$ indicates that we select the list of the R elements of $I(x)^c$ corresponding to the smallest values of $|1 - |\langle d(x),\, a_j\rangle||$.

The algorithm is accelerated by pruning the candidate set $\Sigma_{\max}^{(k)}$ at each iteration. This pruning corresponds to the extraction of the Q vectors $x \in \Sigma_{\max}^{(k)}$ corresponding to the largest values of $\|d(x)\|_2$, which is formally written as

$$\underset{x \in \Sigma_{\max}^{(k)}}{\text{argmin}^{[Q]}} \|d(x)\|_2.$$

Algorithm 2 details the steps of the resulting maximum pursuit algorithm.

Algorithm 2: Maximum greedy pursuit algorithm.

Parameter: pruning rate Q, extension rate R, sparsity s ;
Initialization: set $\Sigma_{\max}^{(1)} = \{\Delta_0, \Delta_1, \ldots, \Delta_{N-1}\}$;
for $k = 2, \ldots, s$ **do**
  Initialization: $\Sigma_{\max}^{(k)} = \emptyset$;
  for *each* $x \in \Sigma_{\max}^{(k-1)}$ **do**
    Compute $\tilde{I} = \underset{j \notin I(x)}{\text{argmin}^{[R]}} |1 - |\langle d(x),\, a_j\rangle||$;
    for *each* $i \in \tilde{I}$ **do**
      Compute $\zeta^+ = \text{sign}(\langle d(x),\, a_i\rangle)$;
      Set $\Sigma_{\max}^{(k)} \leftarrow \Sigma_{\max}^{(k)} \cup \{x + \zeta^+ \Delta_i\}$;
  Pruning: Set $\Sigma_{\max}^{(k)} = \underset{x \in \Sigma_{\max}^{(k)}}{\text{argmin}^{[Q]}} \|d(x)\|_2$;
  Set $k \leftarrow k + 1$.

Similarly, a minimum greedy pursuit algorithm builds a set of k-sparse vectors $\Sigma_{\min}^{(k)}$ for minimizing $1/\|d(x)\|_2$ by iteratively applying the extension (15) to each $x \in \Sigma_{\min}^{(k-1)}$ to obtain $\Sigma_{\min}^{(k)}$. Algorithm 3 below summarizes the steps of the algorithm.

Algorithm 3: Minimum greedy pursuit algorithm.

Parameter: pruning rate Q, extension rate R, sparsity s. ;
Initialization: set $\Sigma_{\min}^{(1)} = \{\Delta_0, \Delta_1, \ldots, \Delta_{N-1}\}$;
for $k = 2, \ldots, s$ **do**
  Initialization: $\Sigma_{\min}^{(k)} = \emptyset$;
  for *each* $x \in \Sigma_{\min}^{(k-1)}$ **do**
    Compute $\tilde{I} = \underset{j \notin I(x)}{\text{argmax}^{[R]}} |\langle d(x),\, a_j\rangle|$;
    for *each* $i \in \tilde{I}$ **do**
      Compute $\zeta^- = -\text{sign}(\langle d(x),\, a_i\rangle)$;
      Set $\Sigma_{\min}^{(k)} \leftarrow \Sigma_{\min}^{(k)} \cup \{x + \zeta^- \Delta_i\}$;
  Pruning: Set $\Sigma_{\min}^{(k)} = \underset{x \in \Sigma_{\min}^{(k)}}{\text{argmax}^{[Q]}} \|d(x)\|_2$;
  Set $k \leftarrow k + 1$.

P	125	250	500	1000
$s^\star(1/4, P)$	10	20	42	79
$\lceil \rho(1/4)P \rceil$	9	17	33	65

Table 1: Our numerically computed critical sparsity levels $s^\star(1/4, P)$ versus the theoretical upper-bound of [34] $\rho(1/4)P \sim 0.065P$.

6. Numerical Results

We apply our greedy pursuit Algorithm 3 to obtain sparse non identifiable vectors, and Algorithms 2 and 3 to get lower bounds on the restricted isometry constants.

6.1. Non Identifiable Vectors

Proposition 1 suggests that a sparse vector with a small value of $1/\|d(x)\|_2$ is likely to be difficult to identify. We thus use the candidate set $\Sigma^{(s)}_{\min}$ computed with the minimum greedy pursuit Algorithm 3 as a pool of s-sparse vectors to challenge the compressed sensing recovery.

Given $\eta = P/N \leqslant 1$, we compute

$$s^\star(\eta, P) = \min\left\{ s \setminus \exists\, x \text{ not identifiable in } \Sigma^{(s)}_{\min} \right\}.$$

The computation of $s^\star$ is carried out using a dichotomy search on the sparsity s to find some candidate set $\Sigma^{(s)}_{\min}$ that contains a non-identifiable vector. Each time, the identifiability is tested by solving the BP problem (1). We use a Douglas-Rachford iterative splitting scheme which is computationally efficient to solve (1) for large scale data, see for instance [8, 9], even though any other solver can be used.

Table 1 reports our numerical findings for $\eta = 1/4$, and compares this numerical evidence with the sharp theoretical bound of Donoho [34] $\rho(1/4) \sim 0.065$. For instance, with $N = 1000$ and $P = 250$, we are able to find a 20-sparse vector that is non-identifiable. In contrast, Monte Carlo sampling with 1000 random vectors for each sparsity s does not reveal any non-identifiable vector for $s < 54$, as shown in Figure 3.

6.2. Sub-Matrix Conditioning

Proposition 2, applied to the region $C_{d(x)}$ defined from Proposition 3 leads to the following lower bounds on the RIP constants

$$\left\{ \begin{array}{l} \delta^{\max}_s \geqslant s/\|d(x)\|_2^2 - 1, \\ \delta^{\min}_s \geqslant 1 - s/\|d(x)\|_2^2, \end{array} \right. \tag{16}$$

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

Empirical sparsity bounds for RIP condition. We thus use the candidate set $\Sigma^{(s)}_{\max}$ which is the outcome of Algorithm 2 to compute a numerical lower bound on the upper restricted isometry constant

$$\tilde{\delta}^{\max}_s = \max_{x \in \Sigma^{(s)}_{\max}} \delta^{\max}(A_{I(x)}).$$

Similarly, a numerical lower bound on the lower restricted isometry constant is obtained through Algorithm 3

$$\tilde{\delta}^{\min}_s = \max_{x \in \Sigma^{(s)}_{\min}} \delta^{\min}(A_{I(x)}).$$

Empirical evaluation of our greedy pursuit bound. To assess the performance of our greedy pursuits, Algorithms 3 and 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 4 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 provides numerical evidence that the heuristic (16) is remarkably accurate in practice.

We also compared the performance of Algorithms 3 and 2 for several values of the parameters Q and R: greedy pursuit with no pruning ($Q = N$, $R = 1$), and pruned weak greedy pursuit with $Q = N/4$ and $R = 4$. The results are depicted in Figure 5. The increase in performance brought by the pruned weak greedy variant becomes slightly more salient as the sparsity level increases.

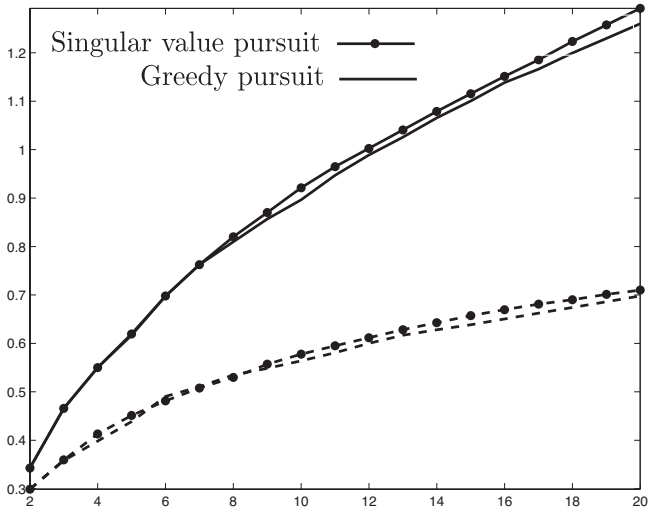

Figure 4: *Comparison of the brute force greedy singular value pursuit and our greedy pursuit with $Q = N$, $R = 1$ for $(N, P) = (2000, 500)$. Solid and dashed lines correspond respectively to $\tilde{\delta}_s^{\max}$ and $\tilde{\delta}_s^{\min}$, as a function of s on the x axis.*

Figure 6 shows that for a fixed value of $\eta = P/N$ and $s/P = 10^{-2}$, the estimates of the RIP constants provided by Algorithms 3 and 2 are close to being constant when the size ($P = 100s, N = P/\eta$) of the sensing matrix varies. This is consistent with the asymptotic upper bound of the restricted isometry constants provided by Blanchard et al. [33]. 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.

Empirical sparsity bounds for RIP condition. For each undersampling rate value $\eta = P/N \leqslant 1$, we compute $s_0^\star(\eta, P)$, the minimum sparsity s for which our empirical estimates invalidate condition (7), hence ℓ^1-identifiability, i.e.

$$(4\sqrt{2} - 3)\tilde{\delta}_{2s}^{\min} + \tilde{\delta}_{2s}^{\max} \geqslant 4(\sqrt{2} - 1)\,. \tag{17}$$

Figure 7 depicts our numerical estimate of the bound (7) for varying s. Table 2 reports our numerically computed critical sparsity levels $s_0^\star(\eta, P)$ for $\eta = 1/4$, and compares this numerical evidence with the theoretical bound of Blanchard et al. [33] $\rho_0(1/4) \sim 0.0027$.

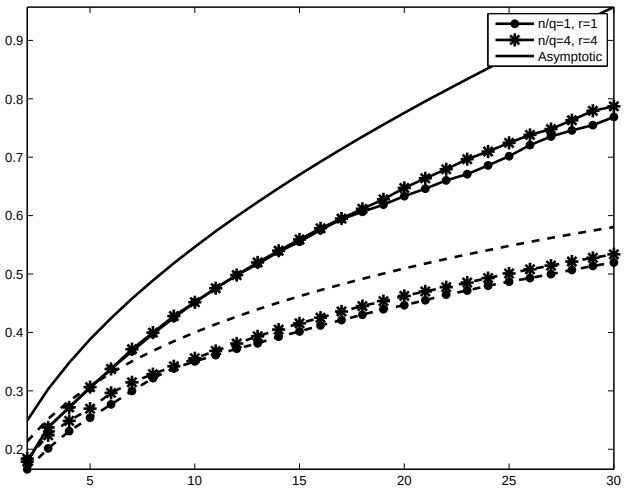

Figure 5: *Comparison of the different variants of Algorithm 3 and 2 for* $(N, P) = (8000, 2000)$. *Solid and dashed lines correspond respectively to* $\tilde{\delta}_s^{\max}$ *and* $\tilde{\delta}_s^{\min}$, *as a function of* s *on the x axis. Circles: greedy pursuit* $Q = R = 1$. *Asterisks: pruned weak greedy pursuit* $Q = N/4$, $R = 4$. *The curves without circles or asterisks corresponds to the asymptotic upper bounds of [33].*

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

Table 2: Our numerically computed critical sparsity levels $s_0^\star(\eta, P)$ versus the theoretical upper-bound of [33] $\rho_0(1/4)P \sim 0.0027P$.

Importance of the sign information. Computing $\delta_s^{\max}$ or $\delta_s^{\min}$ requires the selection of a poorly conditioned sub-matrix A_I, and thus a careful selection of a support I. Our estimations $\tilde{\delta}_s^{\max}$ or $\tilde{\delta}_s^{\min}$ make use of Algorithms 2 and 3, which not only build supports I but also signs $\{x_i\}_{i\in I}$. One can thus wonder how much these signs are useful for the computation of a support I. We thus compare our estimation $\tilde{\delta}_s^{\max}$, with an estimation obtained with algorithm 2 without taking into account sign information. This corresponds to remove the sign computation $\zeta^+ = \mathrm{sign}(\langle d(x), a_{i^+}\rangle)$ obtained from (14) and impose $\zeta^+ = 1$. Figure 8 shows that the sign information is crucial to obtain efficient restricted isometry bounds.

Computational speed. A chief advantage of our (weak) greedy algorithm over the greedy singular value pursuit is that it has a much lower computational load while leading to comparable estimates of the RIP constants. This is clearly testified by the execution times reported in Table 3 for two typical problem sizes. Note also that the pruned weak greedy variant is faster than the greedy version owing to the pruning step.

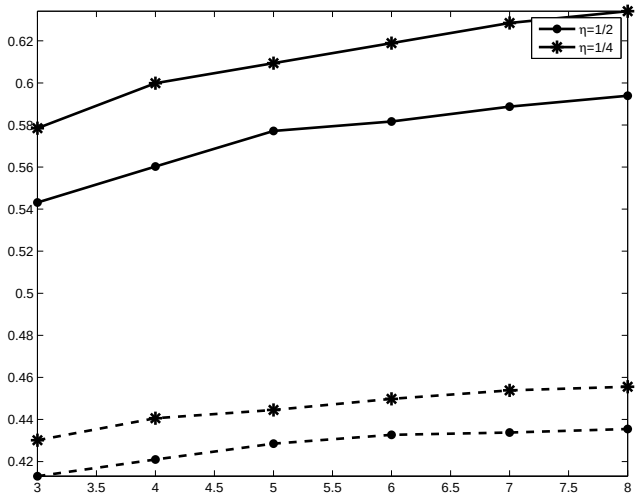

Figure 6: *Plot of $\tilde{\delta}_s^{\max}$ (solid curves) and $\tilde{\delta}_s^{\min}$ (dashed curves) as a function of s on the x axis, for two values of $\eta = P/N$ and for $P = 100s$. The curves are obtained by averaging the value of $\tilde{\delta}_s^{\max}$ and $\tilde{\delta}_s^{\min}$ for 5 realizations of the random matrix A.*

7. Conclusion

We have proposed in this paper a new greedy algorithm to find sparse vectors that are not identifiable and sub-matrices with a small number of columns that are ill-conditioned. This allows us to check numerically sparsity-based criteria for compressed sampling recovery based either on polytope projection or on the RIP. Our numerical findings shows that even in a non-asymptotic setting, the worse case theoretical bounds are quite sharp, which contrast with conclusions that are usually drawn from Monte Carlo sampling of sparse vectors.

Acknowledgments

We would like to thank the referees for their remarks that helped to improve the quality of the manuscript.

References

[1] E. Candès, J. Romberg, T. Tao, Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information, IEEE Trans. Info. Theory 52 (2) (2006) 489–509.
[2] D. Donoho, Compressed sensing, IEEE Trans. Info. Theory 52 (4) (2006) 1289–1306.
[3] D. D. M. Lustig, J. Pauly, Sparse MRI: The application of compressed sensing for rapid MR imaging, Magnetic Resonance in Medicine 58 (6) (2007) 1182–1195.
[4] J. Bobin, J.-L. Starck, R. Ottensamer, Compressed sensing in astronomy, IEEE Journal of Selected Topics in Signal Processing 2 (5) (2008) 718–726.
[5] E. Candès, Compressive sampling, Proceedings of the International Congress of Mathematicians, Madrid, Spain 3 (2006) 1433–1452.
[6] M. Rudelson, R.Vershynin, On sparse reconstruction from fourier and gaussian measurements, Commun. on Pure and Appl. Math. 61 (8) (2008) 1025–1045.

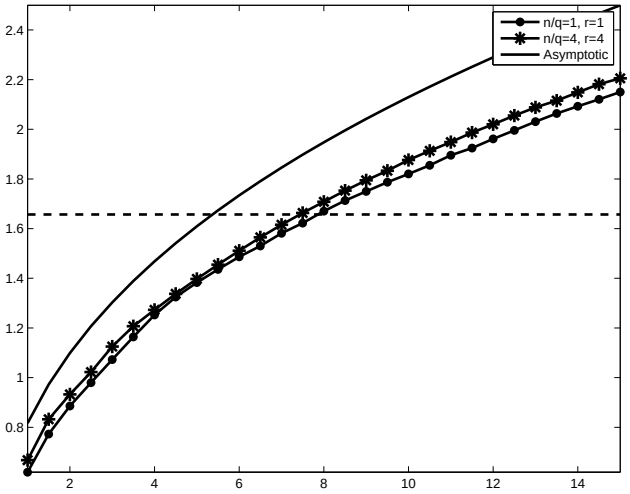

Figure 7: *Display of* $(4\sqrt{2}-3)\tilde{\delta}_{2s}^{\min}+\tilde{\delta}_{2s}^{\max}$ *for* $(N, P) = (8000, 2000)$*, as a function of* s *on the* x *axis, computed using our greedy pursuit with no pruning (circles) and pruning (asterisks,* $Q = N/4$*). The solid curve corresponds to the asymptotic upper bound of [33]. The dashed line corresponds to the limit* $y = 4(\sqrt{2}-1)$ *below which condition* (17) *is fulfilled.*

[7] S. S. Chen, D. Donoho, M. Saunders, Atomic decomposition by basis pursuit, SIAM Journal on Scientific Computing 20 (1) (1998) 33–61.

[8] P. L. Combettes, J.-C. Pesquet, A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery, IEEE Journal of Selected Topics in Signal Processing 1 (4) (2007) 564–574.

[9] M. Fadili, J.-L. Starck, Monotone operator splitting for fast sparse solutions of inverse problems, in: Proc. of IEEE ICIP, Cairo, Egypt, 2009.

[10] M. Figueiredo, R. Nowak, An EM Algorithm for Wavelet-Based Image Restoration, IEEE Trans. Image Proc. 12 (8) (2003) 906–916.

[11] I. Daubechies, M. Defrise, C. D. Mol, An iterative thresholding algorithm for linear inverse problems with a sparsity constraint, Commun. on Pure and Appl. Math. 57 (2004) 1413–1541.

[12] J. Bect, L. Blanc Féraud, G. Aubert, A. Chambolle, A ℓ_1-unified variational framework for image restoration, in: Proc. of ECCV04, Springer-Verlag, 2004, pp. Vol IV: 1–13.

[13] P. L. Combettes, V. R. Wajs, Signal recovery by proximal forward-backward splitting, SIAM Journal on Multiscale Modeling and Simulation 4 (4) (2005) 1168–1200.

[14] E. v. Berg, M. P. Friedlander, Probing the pareto frontier for basis pursuit solutions, SIAM J. Sci. Comp. 31 (2) (2008) 890–912.

[15] J. M. B. Dias, M. A. T. Figueiredo, A new twIST: Two-step iterative shrinkage/thresholding algorithms for image restoration, IEEE Trans. Image Proc. 16 (12) (2007) 2992–3004.

[16] Y. Nesterov, Smooth minimization of non-smooth functions, Math. Program. 103 (1, Ser. A) (2005) 127–152.

[17] D. Needell, J. A. Tropp, Cosamp: Iterative signal recovery from incomplete and inaccurate samples, J. of Appl. and Comput. Harmonic Analysis 26 (2009) 301–321.

[18] I. F. Gorodnitsky, B. D. Rao, Sparse signal reconstruction from limited data using FOCUSS: a re-weighted minimum norm algorithm, IEEE Trans. Image Proc. 45 (3) (1997) 600–616.

[19] E. J. Candes, M. B. Wakin, S. P. Boyd, Enhancing sparsity by reweighted L1 minimization, J. Fourier Anal. Appl. 14 (5) (2008) 877–905.

[20] D. Donoho, X. Huo, Uncertainty principles and ideal atomic decomposition, IEEE Trans. on Information Theory 47 (7) (2001) 2845–2862.

[21] A. Bruckstein, M. Elad, A generalized uncertainty principle and sparse representation in pairs of r^n bases, IEEE Trans. on Information Theory 48 (2002) 2558–2567.

[22] D. Donoho, M. Elad, Optimally sparse representation in general (non-orthogonal) dictionaries via ℓ^1 minimization, Proc. Nat. Aca. Sci. 100 (2003) 2197–2202.

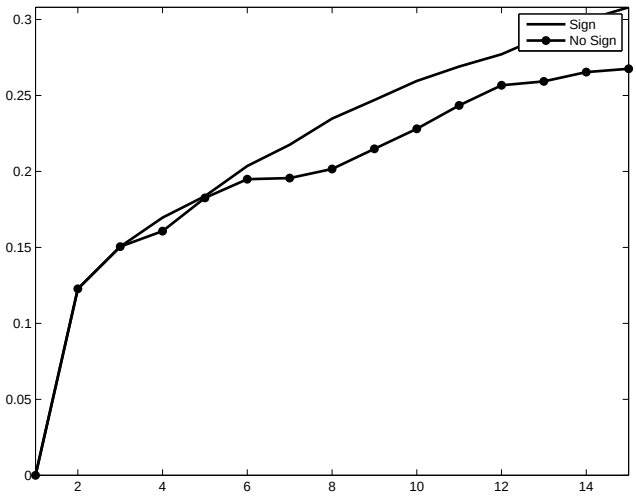

Figure 8: *Display of* $\tilde{\delta}_s^{\max}$ *for* $(N, P) = (4000, 1000)$*, as a function of* s *on the* x *axis, computed using our greedy pursuit (plain curve) and without using the sign information (points).*

[23] R. Gribonval, M. Nielsen, Sparse representations in unions of bases, IEEE Trans. Info. Theory 49 (12) (2003) 3320–3325.

[24] A. Feuer, A. Nemirovski, On sparse representation in pairs of bases, IEEE Trans. Info. Theory 49 (6) (2003) 1579–1581.

[25] D. L. Donoho, M. Elad, V. N. Temlyakov, Stable recovery of sparse overcomplete representations in the presence of noise, IEEE Trans. Info. Theory 52 (1) (2006) 6–18.

[26] J.-J. Fuchs, On sparse representations in arbitrary redundant bases, IEEE Trans. Info. Theory 50 (6) (2004) 1341–1344.

[27] J. A. Tropp, Just relax: convex programming methods for identifying sparse signals in noise, IEEE Trans. Info. Theory 52 (3) (2006) 1030–1051.

[28] M. Wainwright, Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ^1-constrained quadratic programs, IEEE Trans. Info. Theory 55 (5) (2009) .

[29] M. Wainwright, Information-theoretic limits on sparsity recovery in the high-dimensional and noisy setting, Preprint (2009) .

[30] E. Candès, T. Tao, Decoding by linear programming, IEEE Trans. Info. Theory 51 (12) (2005) 4203–4215.

[31] E. J. Candès, The restricted isometry property and its implications for compressed sensing, Compte Rendus de l'Academie des Sciences, Paris, Série I (346) (2008) 589–592.

[32] S. Foucart, M.-J. Lai, Sparsest solutions of underdetermined linear systems via ℓ_q-minimization for $0 < q \leqslant 1$, J. of Appl. and Comput. Harmonic Analysis 26 (3) (2009) 395–407.

[33] J. Blanchard, C. Cartis, J. Tanner, Compressed sensing: How sharp is the RIP?, Preprint (2009) .

[34] D. L. Donoho, High-dimensional centrally symmetric polytopes with neighborliness proportional to dimension, Discrete & Computational Geometry 35 (4) (2006) 617–652.

[35] D. L. Donoho, Neighborly polytopes and sparse solutions of underdetermined linear equations, Technical report Stanford University, Department of Statistics (2006) .

[36] C. Dossal, A necessary and sufficient condition for exact recovery by ℓ_1 minimization, Preprint Hal-00164738 (2007) .

[37] D. Donoho, J. Tanner, Neighborliness of randomly-projected simplices in high dimensions, Proc. National Academy of Sciences 27 (102) (2005) 9452–9457.

[38] D. Donoho, Y. Tsaig, Extensions of compressed sensing, Signal Processing 86 (3) (2006) 549–571.

[39] D. Donoho, J. Tanner, Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing, Philosophical Transactions of The Royal Society A 1906 (367) (2009) 4273–4293.

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

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 3: Comparison of computation times for the evaluation of $\tilde{\delta}_s^{\max}$.

[40] A. Juditsky, A. Nemirovski, On verifiable sufficient conditions for sparse signal recovery via L1 minimization, Preprint (2009) .

[41] A. d'Aspremont, L. E. Ghaoui, Testing the nullspace property using semidefinite programming, To appear in Mathematical Programming (2009) .