



HAL
open science

A Numerical Exploration of Compressed Sampling Recovery

Charles H Dossal, Gabriel Peyré, Jalal M. Fadili

► **To cite this version:**

Charles H Dossal, Gabriel Peyré, Jalal M. Fadili. A Numerical Exploration of Compressed Sampling Recovery. Linear Algebra and its Applications, 2009. hal-00402455v1

HAL Id: hal-00402455

<https://hal.science/hal-00402455v1>

Submitted on 7 Jul 2009 (v1), last revised 27 Nov 2009 (v2)

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.

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-conditioned. 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 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 $P \ll N$. 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.

The resolution of the ill-posed inverse problem $y = Ax$ is stabilized by considering a matrix $A = (a_i)_{i=0}^{N-1} \in \mathbb{R}^{P \times N}$ that is drawn from a proper random ensemble. 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.

Throughout this paper, we consider the case where the entries of A 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

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

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

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

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

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^* to (1) is unique and coincides exactly with x .

Other recovery algorithms with theoretical recovery guarantees exist, such as for instance greedy algorithms or variants [8]. The numerical study of the recovery using these algorithms is out of the scope of this paper.

Remark 1. *Since we are interested in the recovery with ℓ^1 minimization from noiseless measurements, we note, using for example (9), that the identifiability of x depends only on $\text{sign}(x)$. 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. Previous Works

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 , a sparsity-based recovery criterion asserts that any vector satisfying

$$\|x\|_0 = \#\{i \mid x_i \neq 0\} \leq \rho(P/N)P \quad (2)$$

is identifiable for $\rho(\eta) > 0$, where $\eta = P/N < 1$ is the underdeterminacy of the sampling system (or equivalently $\eta^{-1} > 1$ is the redundancy of A).

For a given underdeterminacy 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) beyond 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 [9, 10, 11, 12, 13, 14]. They usually lead to over-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 [15] proved that a sufficient condition for x to be identifiable is

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

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

see also Tropp [16] for a similar result.

Condition of the form (3) has been considered by Wainwright [17] 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[18].

2.2. Restricted Isometry Based Criteria

The seminal work of Donoho [2] and Candès et al. [1, 19] 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 logarithmic term. For the analysis, 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 $\|x\|_0 \leq s$,

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

Condition (5) 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

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

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

The original work of Candès et al. [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 [20] 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 [21] have shown that

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

implies identifiability of all s -sparse vectors. Blanchard et al. [22] determine ρ_0 such that with high probability on A

$$\|x\|_0 \leq \rho_0(P/N)P \quad (8)$$

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 [23, 24] 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

$$B_1 = \{\tilde{x} \mid \|\tilde{x}\|_1 \leq 1\}.$$

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 [23] showed that

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

where $\partial A(B_1)$ is the boundary of the polytope $A(B_1)$. Dossal [25] proved that this topological condition is equivalent to having x as the limit of x_n where $F(x_n) < 1$, i.e. that x is in the closure of the set of all vectors satisfying (3).

Using (9), Donoho and Tanner [23, 26] 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 $\rho(1/4) \approx 0.065$.

This bound is better than (8) because it is a local criterion that only ensures recovery of a single vector, and is thus unable to ensure robustness to noise. In contrast, recovery conditions based on the RIP are global. The recovery of some $x \in \mathbb{R}^N$ does not rely solely on the conditioning of its associated sub-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 either $\delta^{\min}(A_I)$ or $\delta^{\max}(A_I)$ is too large.

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 all tested 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 [27].

The asymptotic evaluation of $\rho(P/N)$ performed by Donoho and Tanner [23, 26] 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 [26], that analyze the recovery of *almost all* sparse vectors, see also [28] 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. [22] 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 [29]. 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 [30] 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 $\delta_{10}^{\max} \geq 0.58$ and $\delta_{10}^{\min} \geq 0.42$ 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 were 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 $\lambda^{\max}(A_I) \geq 1.58$ 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-conditioned. Before detailing in Section 5 fast algorithms that select this 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 δ_s^{\min} and δ_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 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.

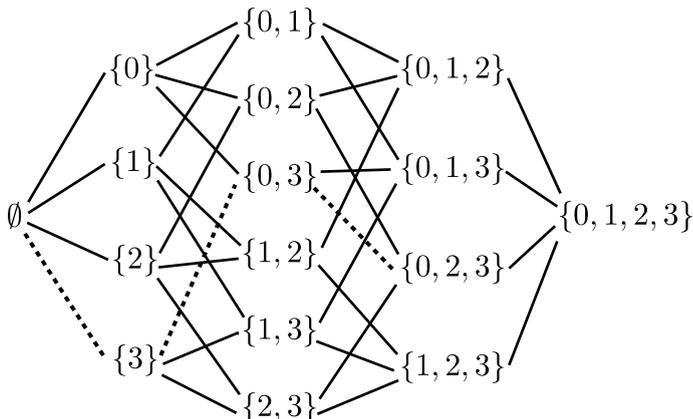


Figure 1: Lattice of the set of supports, for $N = 4$. Dashed: a possible path followed by the algorithm to select a matrix A_I with $I = \{0, 2, 3\}$.

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 $\delta^{\min}(A_{I \cup \{i\}})$.

Algorithm 1: Greedy singular value 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: $\tilde{\delta}_s^{\max} = \max_{I \in \mathcal{I}^{(s)}} \delta^{\max}(A_I)$.

Figure 4, lines with filled circles, 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 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 $\operatorname{rank}(A_I) = |I|$, the distance from the face f_x to 0 is $\frac{1}{\|d(x)\|_2}$, where $d(x)$ is defined in (4).*

Proof. The 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 $A_1^*d(x) = \text{sign}(x_I)$, which in turn yields $\langle d(x), (\text{sign } x_i)a_i \rangle = 1$ for all $i \in I$. The hyperplane

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

is such that for all $i \in I$, $(\text{sign } x_i)a_i \in \mathcal{H}_x$ and thus $f_x \subset \mathcal{H}_x$. The distance between \mathcal{H}_x and 0 is $1/\|d(x)\|_2$.

Let $\mathcal{H}_1 = \{u \mid \langle c, u \rangle = 1\}$ be another hyperplane such that $(\text{sign } x_i)a_i \in \mathcal{H}_1$, for all $i \in I$. The distance between \mathcal{H}_1 and 0 is $1/\|c\|$. 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 $d(x) \in \text{Span}(a_i)_{i \in I}$, $\langle c - d(x), d(x) \rangle = 0$ and then $\|c\|^2 = \|c - d(x)\|^2 + \|d(x)\|_2^2 > \|d(x)\|_2^2$, which completes the proof. \square

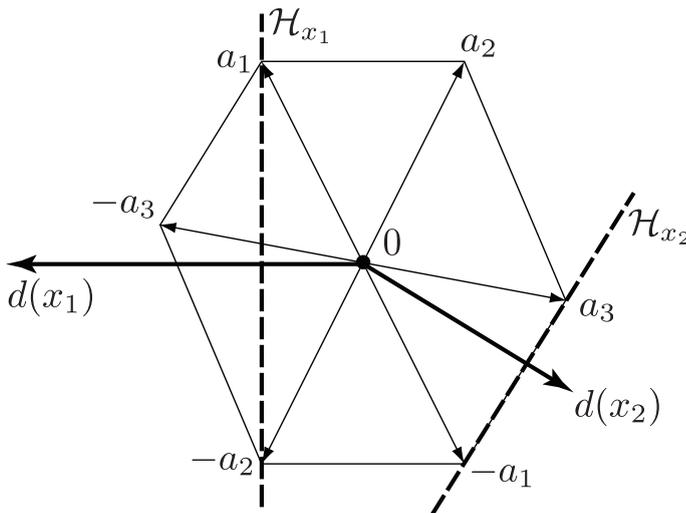


Figure 2: Geometry of ℓ^1 recovery, for $N = 3$ and $P = 2$. The vector $x_1 = (2, -3, 0)$ is not identifiable because f_{x_1} is inside the polytope $A(B_1)$, and has a large $\|d(x_1)\|$. On the contrary, $x_2 = (-5, 0, 3)$ is identifiable because $f_{x_2} \in \partial A(B_1)$, and it has a small $\|d(x_2)\|$.

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 $\|d(x)\|_2$ 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).

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 $d \in \mathbb{R}^P$

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

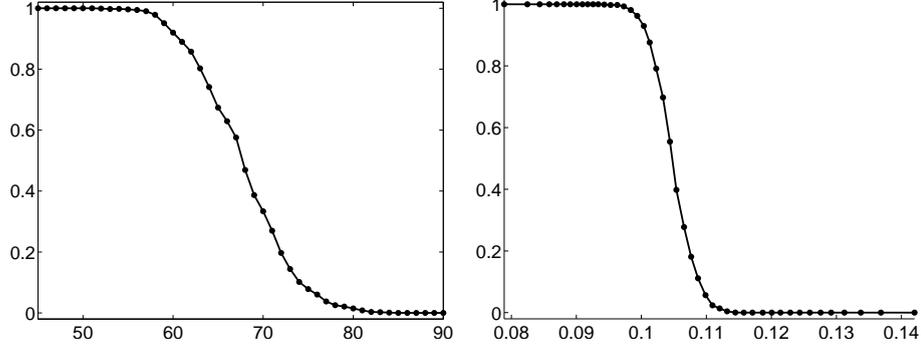


Figure 3: *Left: ratio of identifiable vectors as a function of $\|x\|_0$, for $(P, N) = (250, 1000)$. Right: ratio of identifiable vectors as a function of $\|d(x)\|_2$.*

Remark 2. *In the normalized case—columns of A have unit ℓ^2 -norm— \mathcal{C}_d and its complement have a nice geometrical interpretation. In such a case, \mathcal{C}_d corresponds to a double-spherical cap, whose radius r satisfies $r^2 = 1 - 1/\|d\|^2$. Similarly, \mathcal{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.

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

$$\delta_s^{\max} \geq s/\|d\|^2 - 1. \quad (10)$$

If $\{a_i\}_{i \in I} \subset \mathcal{C}_d^c$ and if $d \in \text{Span}(a_i)_{i \in I}$ then

$$\delta_s^{\min} \geq 1 - s/\|d\|^2. \quad (11)$$

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

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

□

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 \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 high dimension. It is thus desirable to compute an approximate clustering based on a well chosen vector d for the region \mathcal{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 \mathcal{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*

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

is such that $\langle x_i a_i, d(x) \rangle = 1$. Any other region \mathcal{C}_d with another vector d that satisfies this property leads to a worse lower bound on δ_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 $d(x) \in \text{Span}(a_i)_{i \in I}$, 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. \square

This region $\mathcal{C}_{d(x)}$ is an optimal choice to estimate $\delta^{\max}(A_I)$ using (10) if we restrict the choice to regions that pass through the vectors $\{x_i a_i\}_{i \in I}$. 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 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 \mathcal{C}_d .*
- *The cap $\mathcal{C}_{d(x)}$ 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 Proposition 2-3 (see (15)), 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 $\tilde{a}_i \in \text{Span}(a_j, j \in \tilde{I})$ be the dual vector such that*

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

Then

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

Proof. Since $d(x) \in \text{Span}(a_j, j \in I)$ 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

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

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 to solve an over-determined system of linear equations for each i .

We thus select an approximately optimal extension by maximizing or minimizing $|\langle d(x), a_i \rangle - \zeta|$ 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 $1/\|d(x)\|_2$

$$\begin{cases} i^+ = \underset{j \notin I(x)}{\operatorname{argmin}} |1 - |\langle d(x), a_j \rangle||, \\ \zeta^+ = \operatorname{sign}(\langle d(x), a_{i^+} \rangle). \end{cases} \quad (13)$$

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

$$\begin{cases} i^- = \underset{j \notin I(x)}{\operatorname{argmax}} |\langle d(x), a_j \rangle|, \\ \zeta^- = -\operatorname{sign}(\langle d(x), a_{i^-} \rangle). \end{cases} \quad (14)$$

5.2. Greedy Pursuit Algorithms

Starting from an initial candidate set of 1-sparse vectors $\Sigma_{\max}^{(1)} = \{\delta_0, \delta_1, \dots, \delta_{N-1}\}$, the greedy pursuit algorithm for maximizing $1/\|d(x)\|_2$ builds a candidate set $\Sigma_{\max}^{(k)}$ for each sparsity $k \leq s$. The algorithm iteratively applies the extension (13) to each $x \in \Sigma_{\max}^{(k-1)}$ to obtain $\Sigma_{\max}^{(k)}$.

As our algorithm is greedy by nature, the selection rule (13) 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 $R \geq 1$ best candidate extensions. This selection of the best R extensions of a given $x \in \Sigma_{\max}^{(k-1)}$ is written

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

where the notation $\operatorname{argmin}^{[R]}$ 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

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

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 $\Sigma_{\max}^{(1)} = \{\delta_0, \delta_1, \dots, \delta_{N-1}\}$;

for $k = 2, \dots, s$ **do**

Initialization: $\Sigma_{\max}^{(k)} = \emptyset$;

for each $x \in \Sigma_{\max}^{(k-1)}$ **do**

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

for each $i \in \tilde{I}$ **do**

Compute $\zeta^+ = \operatorname{sign}(\langle d(x), a_i \rangle)$;

Set $\Sigma_{\max}^{(k)} \leftarrow \Sigma_{\max}^{(k)} \cup \{x + \zeta^+ \delta_i\}$;

Pruning: Set $\Sigma_{\max}^{(k)} = \operatorname{argmin}_{x \in \Sigma_{\max}^{(k)}}^{[Q]} \|d(x)\|_2$;

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 (14) to each $x \in \Sigma_{\min}^{(k-1)}$ to obtain $\Sigma_{\min}^{(k)}$. Algorithm 3 below summarizes the steps of the algorithm.

6. Numerical Results

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

Algorithm 3: Minimum greedy pursuit algorithm.

Parameter: pruning rate Q , extension rate R , sparsity s . ;

Initialization: set $\Sigma_{\min}^{(1)} = \{\delta_0, \delta_1, \dots, \delta_{N-1}\}$;

for $k = 2, \dots, s$ **do**

Initialization: $\Sigma_{\min}^{(k)} = \emptyset$;

for each $x \in \Sigma_{\min}^{(k-1)}$ **do**

Compute $\tilde{I} = \operatorname{argmax}_{j \notin I(x)}^{[R]} |\langle d(x), a_j \rangle|$;

for each $i \in \tilde{I}$ **do**

Compute $\zeta^- = -\operatorname{sign}(\langle d(x), a_i \rangle)$;

Set $\Sigma_{\min}^{(k)} \leftarrow \Sigma_{\min}^{(k)} \cup \{x + \zeta^- \delta_i\}$;

Pruning: Set $\Sigma_{\min}^{(k)} = \operatorname{argmax}_{x \in \Sigma_{\min}^{(k)}}^{[Q]} \|d(x)\|_2$;

Set $k \leftarrow k + 1$.

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 $\Sigma_{\min}^{(s)}$ computed with the minimum greedy pursuit Algorithm 3 as a pool of s -sparse vectors to challenge the compressed sensing recovery.

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

$$s^*(\eta, P) = \min \left\{ s \mid \exists x \text{ not identifiable in } \Sigma_{\min}^{(s)} \right\}.$$

The computation of s^* is carried out using a dichotomy search on the sparsity s to find some candidate set $\Sigma_{\min}^{(s)}$ 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 [31, 32], even though any other solver can be used.

The following table reports our numerical findings for $\eta = 1/4$, and compares this numerical evidence with the sharp theoretical bound of Donoho [23] $\rho(1/4) \sim 0.065$.

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

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 $\mathcal{C}_{d(x)}$ defined from Proposition 3 leads to the following lower bounds on the RIP constants

$$\begin{cases} \delta_s^{\max} \geq s/\|d(x)\|_2^2 - 1, \\ \delta_s^{\min} \geq 1 - s/\|d(x)\|_2^2, \end{cases} \quad (15)$$

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 $\Sigma_{\max}^{(s)}$ which is the outcome of Algorithm 2 to compute a numerical lower bound on the upper restricted isometry constant

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

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

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

Empirical evaluation of our greedy pursuit bound. To assess the performance of our greedy pursuits, Algorithms 3-2 (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 proves numerically that the heuristic (15) is remarkably accurate in practice.

We also compared the performance of Algorithms 3-2 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.

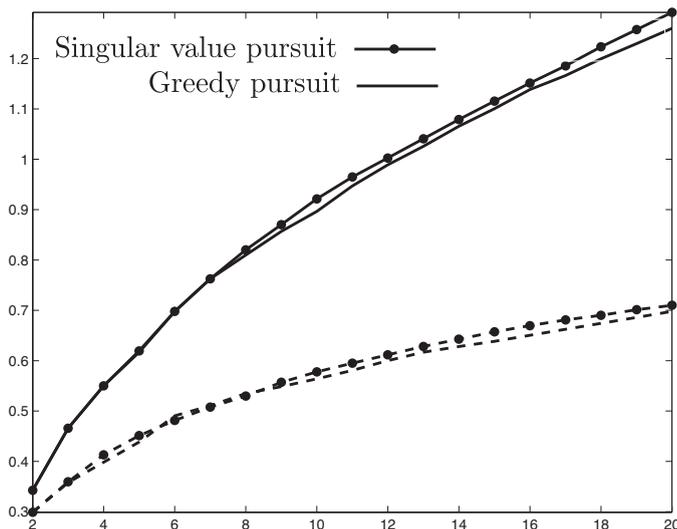


Figure 4: Comparison of the brute force greedy singular value pursuit and our greedy pursuit with $Q = N, R = 1$ for $(N, P) = (2000, 500)$. Solid and dashed lines correspond respectively to $\tilde{\delta}_s^{\max}$ and $\tilde{\delta}_s^{\min}$.

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 3-2 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. [22]. 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 underdeterminacy value $\eta = P/N \leq 1$, we computed $s_0^*(\eta, P)$, the minimum sparsity s for which our empirical estimates inval-

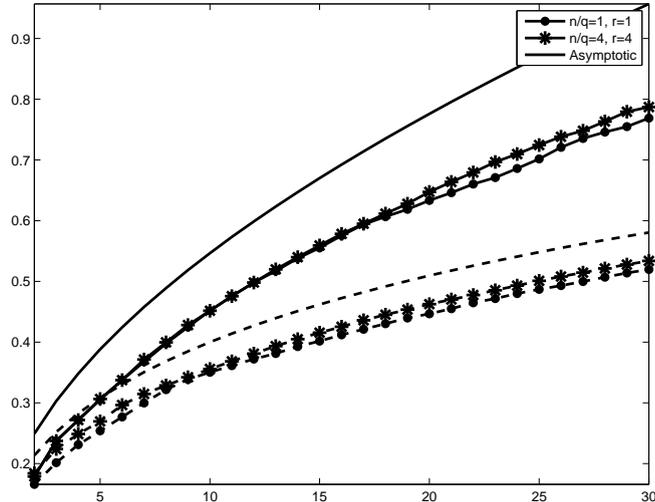


Figure 5: Comparison of the different variants of Algorithm 3-2 for $(N, P) = (8000, 2000)$. Solid and dashed lines correspond respectively to $\tilde{\delta}_s^{\max}$ and $\tilde{\delta}_s^{\min}$. Circles: greedy pursuit $Q = R = 1$. Asterisks: pruned weak greedy pursuit $Q = N/4, R = 4$. The curves without circles or asterisks corresponds to the asymptotic upper bounds of [22].

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 1: Our numerically computed critical sparsity levels $s_0^*(\eta, P)$ versus the theoretical upper-bound of [22] $\rho_0(1/4)P \sim 0.0027P$.

idate condition (7), hence ℓ^1 -identifiability, i.e.

$$(4\sqrt{2} - 3)\tilde{\delta}_{2s}^{\min} + \tilde{\delta}_{2s}^{\max} \geq 4(\sqrt{2} - 1). \quad (16)$$

Figure 7 depicts our numerical estimate of the bound (7) for varying s . Table 1 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. [22] $\rho_0(1/4) \sim 0.0027$.

Computational speed. A chief advantage of our (weak) greedy algorithm over the greedy singular value 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 2 for two typical problem sizes. Note also that the pruned weak greedy variant is faster than the greedy version owing to the pruning step.

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.

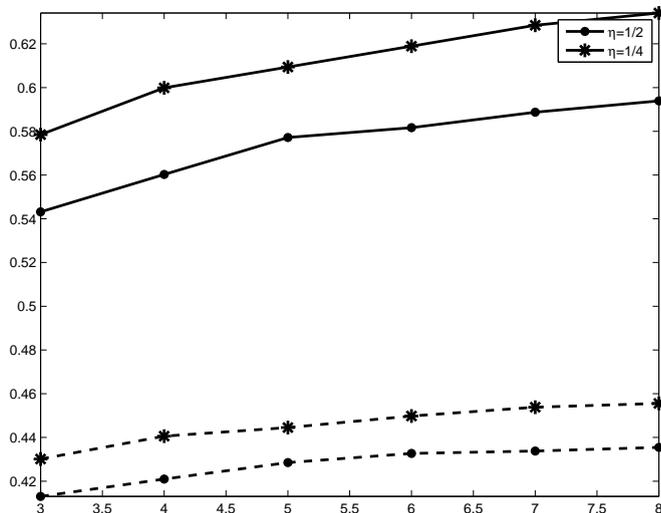


Figure 6: Plot of $\tilde{\delta}_s^{\max}$ (solid curves) and $\tilde{\delta}_s^{\min}$ (dashed curves) as a function of s , for two values of $\eta = P/N$ and for $P = 100s$. The curves are obtained by averaging the value of $\tilde{\delta}_s^{\max}$ and $\tilde{\delta}_s^{\min}$ for 5 realizations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

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

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

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.
- [6] M. Rudelson, R. Vershynin, On sparse reconstruction from fourier and gaussian measurements, *Commun. on Pure and Appl. Math.* 61 (8) (2008) 1025–1045.

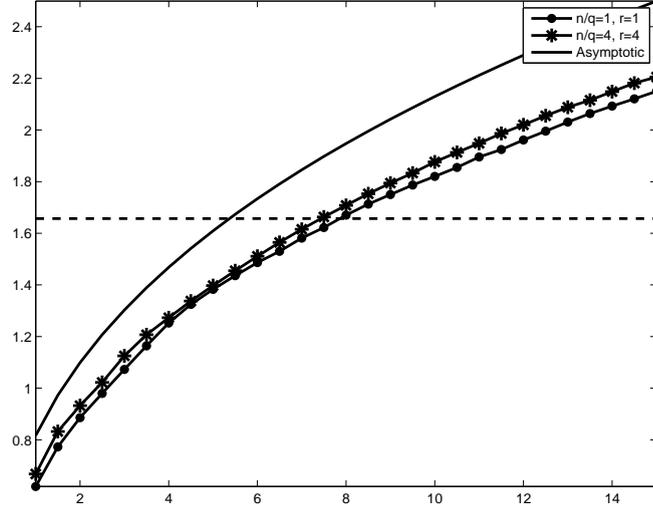


Figure 7: Display of $(4\sqrt{2} - 3)\tilde{\delta}_{2s}^{\min} + \tilde{\delta}_{2s}^{\max}$ for $(N, P) = (8000, 2000)$, computed using our greedy pursuit with no pruning (circles) and pruning (asterisks, $Q = N/4$). The solid curve corresponds to the asymptotic upper bound of [22]. The dashed line corresponds to the limit $y = 4(\sqrt{2} - 1)$ below which condition (16) 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] 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.
- [9] D. Donoho, X. Huo, Uncertainty principles and ideal atomic decomposition, IEEE Trans. on Information Theory 47 (7) (2001) 2845–2862.
- [10] A. Bruckstein, M. Elad, A generalized uncertainty principle and sparse representation in pairs of \mathbf{r}^n bases, IEEE Trans. on Information Theory 48 (2002) 2558–2567.
- [11] D. Donoho, M. Elad, Optimally sparse representation in general (non-orthogonal) dictionaries via ℓ^1 minimization, Proc. Nat. Aca. Sci. 100 (2003) 2197–2202.
- [12] R. Gribonval, M. Nielsen, Sparse representations in unions of bases, IEEE Trans. Info. Theory 49 (12) (2003) 3320–3325.
- [13] A. Feuer, A. Nemirovski, On sparse representation in pairs of bases, IEEE Trans. Info. Theory 49 (6) (2003) 1579–1581.
- [14] 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.
- [15] J.-J. Fuchs, On sparse representations in arbitrary redundant bases, IEEE Trans. Info. Theory 50 (6) (2004) 1341–1344.
- [16] J. A. Tropp, Just relax: convex programming methods for identifying sparse signals in noise, IEEE Trans. Info. Theory 52 (3) (2006) 1030–1051.
- [17] M. Wainwright, Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ^1 -constrained quadratic programs, IEEE Trans. Info. Theory 55 (5).

- [18] M. Wainwright, Information-theoretic limits on sparsity recovery in the high-dimensional and noisy setting, Preprint.
- [19] E. Candès, T. Tao, Decoding by linear programming, *IEEE Trans. Info. Theory* 51 (12) (2005) 4203–4215.
- [20] 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.
- [21] S. Foucart, M.-J. Lai, Sparsest solutions of underdetermined linear systems via ℓ_q -minimization for $0 < q \leq 1$, to appear in *Applied and Computational Harmonic Analysis*.
- [22] J. Blanchard, C. Cartis, J. Tanner, The restricted isometry property and ℓ^q -regularization: Phase transitions for sparse approximation, Preprint.
- [23] D. L. Donoho, High-dimensional centrally symmetric polytopes with neighborliness proportional to dimension, *Discrete & Computational Geometry* 35 (4) (2006) 617–652.
- [24] D. L. Donoho, Neighborly polytopes and sparse solutions of underdetermined linear equations, Technical report Stanford University, Department of Statistics.
- [25] C. Dossal, A necessary and sufficient condition for exact recovery by ℓ_1 minimization, Preprint.
- [26] D. Donoho, J. Tanner, Neighborliness of randomly-projected simplices in high dimensions, *Proc. National Academy of Sciences* 27 (102) (2005) 9452–9457.
- [27] D. Donoho, Y. Tsaig, Extensions of compressed sensing, *Signal Processing* 86 (3) (2006) 549–571.
- [28] D. Donoho, J. Tanner, Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing, to appear in *Philosophical Transactions of The Royal Society A*.
- [29] A. Juditsky, A. Nemirovski, On verifiable sufficient conditions for sparse signal recovery via ℓ_1 minimization, Preprint.
- [30] A. d’Aspremont, L. E. Ghaoui, Testing the nullspace property using semidefinite programming, Preprint.
- [31] 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.
- [32] M. Fadili, J.-L. Starck, Monotone operator splitting for fast sparse solutions of inverse problems, in: *Proc. of IEEE ICIP, Cairo, Egypt, 2009*.