A greedy algorithm to extract sparsity degree for l1/l0-equivalence in a deterministic context

Nelly Pustelnik, Charles H Dossal, Flavius Turcu, Yannick Berthoumieu, Philippe Ricoux

A greedy algorithm to extract sparsity degree for ℓ_1/ℓ_0-equivalence in a deterministic context [1]

N. Pustelnik[2], Ch. Dossal[3], F. Turcu[4], Y. Berthoumieu[4], Ph. Ricoux[5]

May 28, 2013

Abstract

This paper investigates the problem of designing a deterministic system matrix, that is measurement matrix, for sparse recovery. An efficient greedy algorithm is proposed in order to extract the class of sparse signal/image which cannot be reconstructed by ℓ_1-minimization for a fixed system matrix. Based on the polytope theory, the algorithm provides a geometric interpretation of the recovery condition considering the seminal work by Donoho. The paper presents an additional condition, extending the Fuchs/Tropp results, in order to deal with noisy measurements. Simulations are conducted for tomography-like imaging system in which the design of the system matrix is a difficult task consisting of the selection of the number of views according to the sparsity degree.

[1]This work was supported by grant from TOTAL SA.

[2]Nelly Pustelnik is with the Laboratoire de Physique de l'ENS Lyon, CNRS UMR 5672, F69007 Lyon, France. E-mail: `nelly.pustelnik@ens-lyon.fr`

[3]Charles Dossal is with the IMB, Université de Bordeaux, CNRS UMR 5584, 33405 Talence cedex, France. E-mail: `charles.dossal@math.u-bordeaux1.fr`

[4]Flavius Turcu and Yannick Berthoumieu are with the IMS, Université de Bordeaux, UMR CNRS 5218, 33405 Talence cedex, France. E-mail: `flavius.turcu@u-bordeaux1.fr; yannick.berthoumieu@ims-bordeaux.fr`

[5]Philippe Ricoux is with TOTAL SA, Direction Scientifique, Paris La Défense, France. E-mail: `philippe.ricoux@total.com`.

1 Introduction

The main goal of compressed sensing is to design a system matrix $A \in \mathbb{R}^{M\times N}$ with $M < N$ for which every s-sparse signals $\overline{x} \in \mathbb{R}^N$ can be recovered from the observations $y = A\overline{x}$. The sparsity degree s denotes the number of nonzero components in the signal. The considered problem may include an additive perturbation that leads to an observation vector $y = A\overline{x} + n$ where $n \in \mathbb{R}^M$. The objective of designing a system matrix involves to specify the smallest number M of required observations we need as well as the way to acquire them (e.g. random sampling or regular sampling). Moreover, we have to recall that this design will be obviously dependent on the sparsity degree s and of the signal size N.

The classical approach to look for some sufficiently sparse solution consists to solve:

$$\widehat{x} \in \operatorname*{Argmin}_{x\in\mathbb{R}^N} \|x\|_1 \quad \text{subject to} \quad \|y - Ax\|_2 \leq \epsilon, \tag{1}$$

where $\epsilon > 0$ and the ℓ_1-norm is formally defined as, for every $x = (x_i)_{1\leq i\leq N} \in \mathbb{R}^N$, $\|x\|_1 = \sum_{i=1}^N |x_i|$. Numerous algorithms have been proposed to solve problem (1) or its Lagrangian formulation [1, 2, 3, 4]. By making use of the ℓ_1/ℓ_0-equivalence guarantees, the latest non-smooth convex optimization techniques propose a specific framework to exactly recover sparse signals by ℓ_1-minimization.

The main theoretical results about sparse recovery by ℓ_1-minimization are briefly recalled below. On the one hand, sufficient conditions were proposed by:

- Donoho and Huo [5] with the concept of coherence for a matrix A. This allows us to characterize ℓ_1/ℓ_0-equivalence and thus leads to $\overline{x} = \widehat{x}$,
- Candès et al. [6] through the restricted isometry property (RIP) onto A to establish that $\overline{x} = \widehat{x}$,
- Fuchs [7] and Tropp [8] using first order necessary condition and then the subdifferential of the ℓ_1 norm in order to prove that $\overline{x} = \widehat{x}$.

Note that these sufficient conditions can include robustness to noise. On the other hand, Donoho gives a necessary and sufficient condition based on polytope theory [9] to prove that $\overline{x} = \widehat{x}$. Its dual interpretation is known as the null space property.

Apart from the coherence property proposed in [5], these theoretical results require NP-hard computations to test their validity. Normalized random system matrices have been largely studied in the compressed sensing literature in order to simplify the recovery conditions. Indeed, such an assumption onto the system matrix enables to control the associated eigenvalue distribution and thus to obtain an explicit relation between the observation number M, the signal/image size N, and the sparsity degree s. However, such a relation does not exist for deterministic matrices such those encountered in tomography applications. Until now, in order to design the system matrix for a specific sparsity degree, it seems that most of the existing works dealing with a deterministic context have introduced some randomness in their formulation in order to justify the good results obtained from ℓ_1-minimization [10, 11, 12].

In this work we propose a greedy algorithm based on Donoho results [9] in order to extract the class of sparse signals which cannot be reconstructed by ℓ_1-minimization for a fixed system matrix. A consequence will be to extract an approximation of the largest sparsity degree s allowing us to recover every s-sparse vectors by ℓ_1-minimization based on a given deterministic matrix A.

The paper is organized as follows. Section 2 details a greedy algorithm designed by Dossal et al. [13] which initially considers neighborly condition of a polytope associated with a random matrix. Section 3 presents the adaptation of this greedy algorithm in order to deal with a deterministic matrix and details how to extract the approximation of the sparsity degree s that allows us to recover every s-sparse signal by ℓ_1-minimization. The way to integrate robustness to noise in the sparsity extraction is also presented. Finally, Section 4 illustrates the performance of the algorithm in a context of tomography.

2 Polytope theory for sparse recovery

2.1 Theoretical results

In [9], Donoho describes the ℓ_1/ℓ_0-equivalence by considering ideas from the convex polytope theory. In this work, Donoho introduced a necessary and sufficient condition based on the neighborly property of a polytope.

Definition 2.1 For every $i \in \{1, \dots, N\}$, let a_i denote the i-th column of A. The quotient polytope associate to A is formed by taking the convex hull of the $2N$ points $(\pm a_i)$ in $\mathbb{R}^M$. A polytope P is called (s-1)-neighborly if every subset of s elements $(\pm_{i_\ell} a_{i_\ell})_{\ell=1}^{s}$ are the vertices of a face of P.

An illustration of a polytope P is provided in Figure 1(a) for $N = 3$ and $M = 2$. In this example, it appears that P has $2N = 6$ vertices and is 0-neighborly but not 1-neighborly (e.g. (a_1, a_3) does not span a face of P).

Theorem 2.2 *[9, Theorem 1] Let A be a $M \times N$ matrix with $M < N$. These two properties of A are equivalent:*
(i) The quotient polytope P has $2N$ vertices and is (s-1)-neighborly;
(ii) Whenever $y = Ax$ has a solution $\overline{x}$ having at most s nonzeros, $\overline{x}$ is the unique optimal solution of the ℓ_1-minimization problem.

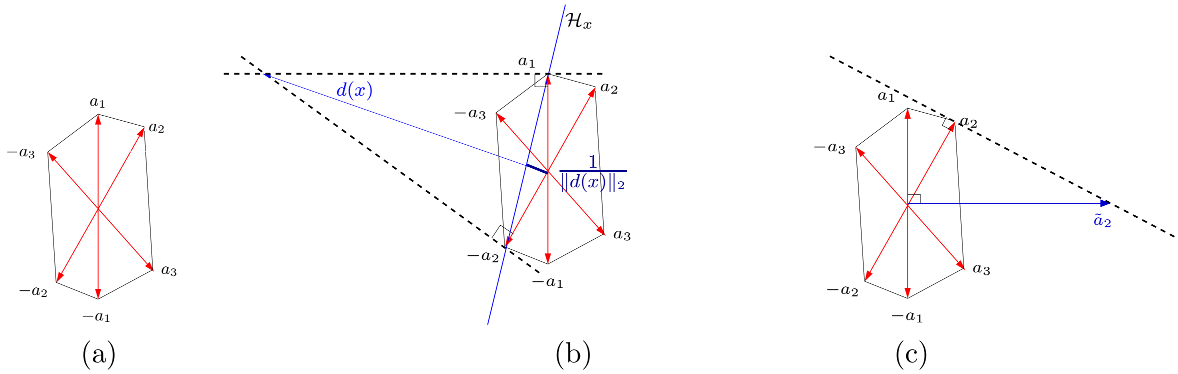

Figure 1: Illustration relative to the polytope formed by taking the convex hull of the $2N$ points $(\pm a_i)$ in $\mathbb{R}^M$.

2.2 Greedy algorithm to extract vectors inside the polytope for random matrices

The geometric interpretation of Donoho was considered by Dossal et al. [13], in a context of normalized random matrices, in order to extract non-ℓ_1-identifiable vectors.

Regarding Theorem 2.2, a non-ℓ_1-identifiable vector denotes a vector $x \in \mathbb{R}^N$ with a support I for which the image of the ℓ_1-ball associated to the support I is inside the polytope. In other words, non-ℓ_1-identifiable vectors have a small distance from the center of the polytope to the hyperplane $\mathcal{H}_x$ (hyperplane going through $\{\text{sign}(x_i)a_i\}_{i\in I}$ where $I \subset \{1,\dots,N\}$). This distance [13, Proposition 1], illustrated in Figure 1(b), is $1/\mathcal{D}_x$ where

$$\mathcal{D}_x = \|d(x)\|_2 \quad \text{with} \quad \begin{cases} d(x) = A_I(A_I^* A_I)^{-1}\,\text{sign}(x_I), \\ A_I = (a_i)_{i\in I}. \end{cases}$$

In Figure 1(b), we can notice that $(\boldsymbol{a}_1, -\boldsymbol{a}_2)$ does not span a face of P and has a large $\mathcal{D}_{\boldsymbol{x}}$ while $(\boldsymbol{a}_1, \boldsymbol{a}_2)$ spans a face of P and has a smaller value of $\mathcal{D}_{\boldsymbol{x}}$.

It results that looking for non-ℓ_1-identifiable vectors leads to search vectors x with the largest measure $\mathcal{D}_x$. Consequently, Dossal et al. [13] have proposed an algorithm allowing to extract sparse vectors with the largest $\mathcal{D}_x$.

The greedy algorithm proposed by Dossal et al. [13] is recalled in Algorithm 1 and the associated complexity is evaluated in Proposition 2.3.

Algorithm 1 constructs a set of s-sparse vectors with the largest $\mathcal{D}_x$ values. At each iteration, the new set of non-identifiable k-sparse vectors $\Sigma_{\max}^{(k)}$ is built from the previous vector set $\Sigma_{\max}^{(k-1)}$ (e.g. set of vectors with a sparsity degree $k-1$). It results that each step looks for the k-sparse vectors $\widetilde{x}$ such that $\widetilde{x} = x + o\Delta_i$ where x denotes a (k-1)-sparse vector from $\Sigma_{\max}^{(k-1)}$, $o \in \{-1,+1\}$ and Δ_i is a Dirac vector at the location i. In Algorithm 1, the notation $\arg\max^{[R]}$ (resp. $\arg\max^{[Q]}$) involves to keep the R indexes and signum which lead to the maximum $\|d(x+o\Delta_i)\|_2$ (resp. the Q vectors which lead to the maximum $\|d(x)\|_2$).

Algorithm 1 [13] - Extract sparse vectors with the large $\mathcal{D}_x$.

Set the pruning rate Q and the extension rate R,
Set the sparsity degree S,
Set $\Sigma_{\max}^{(1)} = \{\Delta_1, \ldots, \Delta_N\}$,
For $k = 2, \ldots, S$
    $\overline{\Sigma}_{\max} = \emptyset$,
    For every $x \in \Sigma_{\max}^{(k-1)}$
        $(\hat{I}, \hat{O}) = \underset{i \notin I(x); o \in \{-1,+1\}}{\arg\max^{[R]}} \|d(x + o\Delta_i)\|_2$
        For $j \in \{1, \ldots, R\}$
            $\overline{\Sigma}_{\max} = \overline{\Sigma}_{\max} \cup \{x + \hat{O}_j \Delta_{\hat{I}_j}\}$
    Set $\Sigma_{\max}^{(k)} = \underset{x \in \overline{\Sigma}_{\max}}{\arg\max^{[Q]}} \|d(x)\|_2$

Proposition 2.3 *The iteration complexity of Algorithm 1 is*

$$O\big(2Q(N-k+1)(N(k+1)+k^3)\big) \ll O(\mathrm{C}_k^N).$$

Regarding Proposition 2.3, it appears that the computational cost of each iteration is too high to be used in real experiments. However, it is possible to write [13, Proposition 4]:

$$\|d(\widetilde{x})\|_2^2 = \|d(x)\|_2^2 + \|\widetilde{a}_i\|_2^2 |\langle d(x), a_i \rangle - o|^2 \tag{2}$$

where x denotes a k-sparse vector with the support I, $o \in \{-1, +1\}$, $\widetilde{x}$ denotes a $(k+1)$-sparse vector with the support $I \cup \{i\}$, and $\widetilde{a}_i \in \mathrm{Span}\,(a_j, j \in I \cup \{i\})$ such that $\langle \widetilde{a}_i, a_i \rangle = 1$, and, for every $j \in I$, $\langle \widetilde{a}_i, a_j \rangle = 0$. Note that $\langle \cdot, \cdot \rangle$ denotes the scalar product. An illustration of $\widetilde{a}_i$ is given in Figure 1(c). An accelerated version of Algorithm 1 for random matrices was proposed in [13] by making the assumption that $\|\widetilde{a}_i\|_2$ is close to 1. In the next section, we refer to this accelerated version by Algorithm 1bis.

3 Evaluate sparsity in a deterministic context

In some real applications such as tomographic imaging, the inversion problem issue does not involve a random system matrix. Thus, an interesting question is how to get such an efficient algorithm considering a deterministic matrix. Moreover, it can be noticed that the polytope theory proposed by Donoho is not adapted to the noisy case. Another natural question is how to introduce robustness to noise. In this section, an adaptation of Algorithm 1 in a deterministic context is proposed and the noisy case is handled via the derivation of a new criterion based on Fuchs/Tropp theorems [7, 8] .

3.1 Adaptation of Algorithm 1 to deterministic matrices

For deterministic matrices, the accelerated version of Algorithm 1 can no longer be used due to the fact that $\|\widetilde{a}_i\|_2$ cannot be discarded. However, in order to reduce the computational cost of

Algorithm 1 (stated in Proposition 2.3), we consider Equation (2) and give the closed form of $\|\widetilde{a}_i\|_2$.

Proposition 3.1 *Let* $\widetilde{a}_i \in \mathrm{Span}\,(a_j, j \in I \cup \{i\})$ *and such that* $\langle \widetilde{a}_i, a_i \rangle = 1$ *and* $\langle \widetilde{a}_i, a_j \rangle = 0$*, for every* $j \in I$*. It results that*

$$\widetilde{a}_i = \frac{a_i - A_I(A_I^* A_I)^{-1} A_I^* a_i}{\big\langle a_i, a_i - A_I(A_I^* A_I)^{-1} A_I^* a_i \big\rangle}. \tag{3}$$

The computation of $d(\widetilde{x})$ can thus be expressed as a function of $d(x)$ and $\widetilde{a}_i$. For each sparsity degree k in Algorithm 1, this expression leads to the computation of Q matrix inversions of size $(k-1) \times (k-1)$ rather than $Q \times (N-k)$ matrix inversions of size $k \times k$. The proposed algorithm is detailled in Algorithm 2 and the associated computational cost is specified in Proposition 3.2.

Algorithm 2 Accelerated version of Algorithm 1 for deterministic matrices.

Set the pruning rate Q and the extension rate R,
Set the sparsity degree s,
Set $\Sigma_{\max}^{(1)} = \{\Delta_1, \ldots, \Delta_N\}$,
For $k = 2, \ldots, s$
  $\overline{\Sigma}_{\max} = \emptyset$,
  For every $x \in \Sigma_{\max}^{(k-1)}$
    Compute the matrix inversion involved in (3)
    $(\hat{I}, \hat{O}) = \underset{\substack{i \notin I(x);\\ o \in \{-1,+1\}}}{\arg\max^{[R]}} \; \|d(x)\|_2^2 + \|\widetilde{a}_i\|_2^2 |\langle d(x), a_i \rangle - o|^2$
    For $j \in \{1, \ldots, R\}$
      $\overline{\Sigma}_{\max} = \overline{\Sigma}_{\max} \cup \{x + \hat{O}_j \Delta_{\hat{I}_j}\}$
  Set $\Sigma_{\max}^{(k)} = \underset{x \in \overline{\Sigma}_{\max}}{\arg\max^{[Q]}} \|d(x)\|_2$

Proposition 3.2 *The iteration complexity of Algorithm 2 is*

$$O\big(Q(N(k-1) + (k-1)^3) + 2Q(N-(k-1))(N(k+4))\big) \ll O\big(2Q(N-k+1)(N(k+1)+k^3)\big)$$

In Figure 2, we compare the original algorithm (Algorithm 1), the accelerated version of this algorithm designed for random matrices (Algorithm 1bis), and the proposed accelerated version devoted to the deterministic matrices (Algorithm 2). The evaluation of the proposed algorithm is presented both in a context of random matrix and of tomography, i.e. A denotes either a Radon transform (this matrix is obtained with the MATLAB implementation of the Radon transform) where $N = 20 \times 20$ and $M = 198$ (that corresponds to 4 angles) or a random matrix of the same size. We compare these algorithms in terms of computation time and of maximum extracted $\mathcal{D}_x$ values. The pruning rate Q and the extension rate R are fixed to $Q = N$ and $R = 1$. It appears that in a deterministic context (bottom figures), the extraction performances (i.e. find sparse vectors with large $\mathcal{D}_x$) of Algorithm 2 are similar to those of Algorithm 1 with a much better convergence rate while the extraction performance are better than the accelerated version considering Algorithm 1bis. However, note that in a random context (top figures), the proposed approach leads to smaller improvements. To sum up, these results illustrate the relevance of the proposed algorithm in order to easily handle deterministic matrices with higher dimensionality.

3.2 Extract sparsity with Algorithm 2

Considering the tomography-like experiment detailed above, we present in Figure 3 the reconstruction results obtained by ℓ_1-minimization for vectors with large $\mathcal{D}_x$ considering different sparsity degrees (i.e. vectors in $\Sigma_{\max}^{(k)}$). We also present the obtained reconstruction vectors with small $\mathcal{D}_x$ that require to compute the "minimum version of Algorithm 2" (i.e. vectors choose in $\Sigma_{\min}^{(k)}$). In these experiments the ℓ_1-minimization algorithm is FISTA [3] and the stopping criterion takes in consideration the evolution of the relative error between $\overline{x}$ and $\widehat{x}$ ($< 10^{-10}$) as well as the iteration number ($< 10^6$).

Algorithm 2 allows us to extract k-sparse vectors with the largest value of $\mathcal{D}_x$. It results that if the vector $x \in \Sigma_{\max}^{(k)}$ having the largest $\mathcal{D}_x$ value cannot be recovered by ℓ_1-minimization, **a good approximation of the sparsity degree s for which every s-sparse vectors can be reconstructed by ℓ_1-minimization from $z = A\overline{x}$ is the largest $s < k$.**

3.3 Noisy case

We have mentionned in the introduction that Fuchs [7] and also Tropp [8] proposed a sufficient condition in order to recover sparse vectors by ℓ_1-minimization. Contrary to Theorem 2.2, this condition is not a necessary condition but it has the nice property that it can be extended in order to take into account the robustness w.r.t noise. Here, we propose a new result inspired by Fuchs/Tropp result which allows us to easily control the reconstruction error in the noisy case.

Proposition 3.3 *Let $I \subset \{1, \ldots, N\}$ denote a set of index such that $|I| = s$ and let $J = \{1, \ldots, N\} \setminus I$. Let*

$$\mathrm{ERC}(I) = \max_{j \in J} \|(A_I^* A_I)^{-1} A_I^* a_j\|_1. \tag{4}$$

We assume that:
1) $\mathrm{ERC}(I) < 1$,
2) $\gamma > \frac{\max_{j \in J} \|a_j\|_2 \|n\|_2}{1 - \mathrm{ERC}(I)}$.

Then, it results that the support of the solution $\widehat{x}$ of

$$\arg\min_{x \in \mathbb{R}^N} \frac{1}{2}\|y - Ax\|_2^2 + \gamma\|x\|_1, \tag{5}$$

is included in the support of $\overline{x}$ and

$$\|\hat{x} - \overline{x}\|_2 \leq \big(\lambda_{\min}(A_I^* A_I)\big)^{-1} \left(\|A_I^* n\|_2 + \frac{\sqrt{|I|}\max_{j \in J} \|a_j\|_2 \|n\|_2}{1 - \mathrm{ERC}(I)} \right). \tag{6}$$

4 Experimental results

We consider a problem of few angle tomography for sparse data. It appears that some industrial materials which requires to be studied through a tomographic process exhibit sparsity properties. The goal of this experiment is to design the system matrix (i.e. find the adapted number of views) according to a given sparsity degree.

The system matrix A is associated to a Radon tranform. The MATLAB implementation makes it possible to select the number and the location of the polar angles (between 0^o and 180^o). In this experiment we fix the angle between two views. The experiments have been held for images of size $N = 32 \times 32$. Algorithm 2 is successively employed with a system matrix associated to 6 view angles ($M = 294$), 9 view angles ($M = 441$), and 12 view angles ($M = 588$). Moreover, due to the positivity of the data, $o = +1$ in Algorithm 2.

In Table 1, we evaluate the sparsity degree allowing us to recover every sparse vectors. The first row details the state-of-the-art results related to the coherence [7] such that:

$$s < \frac{1}{2}\Big(1 + \frac{1}{\mu(A)}\Big) \quad \text{where} \quad \mu(A) = \max \frac{< a_i, a_j >}{\|a_i\|_2^2 \|a_j\|_2^2}.$$

The second row presents the approximated value of the sparsity degree extracted by considering the proposed approach described in Section 3.2.

In Table 2, we evaluate the sparsity degree allowing us to recover every sparse vectors in the noisy case. We have filled in the table in considering Proposition 3.3 where I denotes the support of the vector $x \in \Sigma_{\max}^{(s)}$, extracted with Algorithm 2, with the largest $\mathcal{D}_x$. The values of $\|A_I^* n\|_2$ and $\|n\|_2$ are obtained by a Monte-Carlo process with 100 realizations of a vector $n \sim \mathcal{N}(0, \sigma^2)$. The robustness to noise expressed by Proposition 3.3 requires to insure the convergence inside the support. This is a strong condition and it explains why the extracted sparsity is small compare to the results presented in Table 1.

5 Conclusion

We propose an efficient method to upper bound the sparsity degree s for which every s-sparse vectors can be reconstructed by ℓ_1-minimization according to a specific system matrix. Such a value is important to know in a context where we want to be sure that ℓ_1-minimization leads to the exact true sparse solution or, in presence of noise, to a solution for which we control the error.

Note that the proposed method does not directly give a relation between the sparsity and the size of the matrix but one might construct it by considering system matrices A with different sizes.

In a future work, a parallel implementation will be done in order to extract the sparsity degree for real tomography matrices. Moreover, we should notice that this approach can be considered for various contexts in inverse problems such as restoration or inpainting, and also in the case where A models the product of the system matrix with a frame transform.

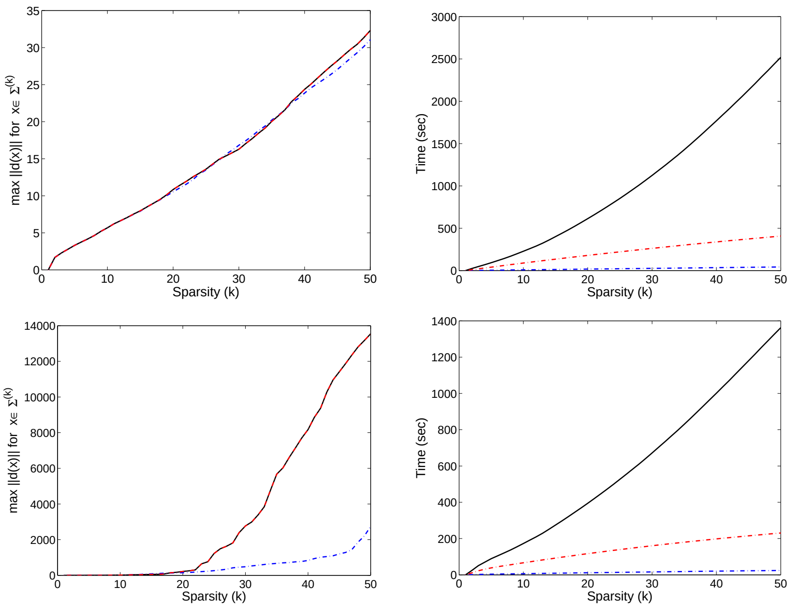

Figure 2: Algorihm 1 (solid black), Algorithm 1bis (dash-dotted blue), Algorithm 2 (dash-dotted red). The bottom figures present the results obtained with a tomography-like matrix while the top figures illustrate the results for a normalized random matrix.

	6 views ($M = 294$)	9 views ($M = 441$)	12 views ($M = 588$)
Sparsity (Coherence)	2	3	5
Sparsity (Proposed method)	**39**	**120**	**144**

Table 1: Sparsity s allowing us to recover every s-sparse vectors by ℓ_1-minimization in the absence of noise consideration. The second row presents the approximation of the sparsity obtained with the proposed approach. Results for three different configurations of the tomography-like matrix A.

References

[1] P. L. Combettes and V. R. Wajs, "Signal recovery by proximal forward-backward splitting," *Multiscale Model. and Simul.*, vol. 4, no. 4, pp. 1168–1200, Nov. 2005.

[2] J. M. Bioucas-Dias and M. A. T. Figueiredo, "A new TwIST: two-step iterative shrinkage/thresholding algorithms for image restoration," *IEEE Trans. Image Process.*, vol. 16, no. 12, pp. 2992–3004, Dec 2007.

[3] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imaging Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

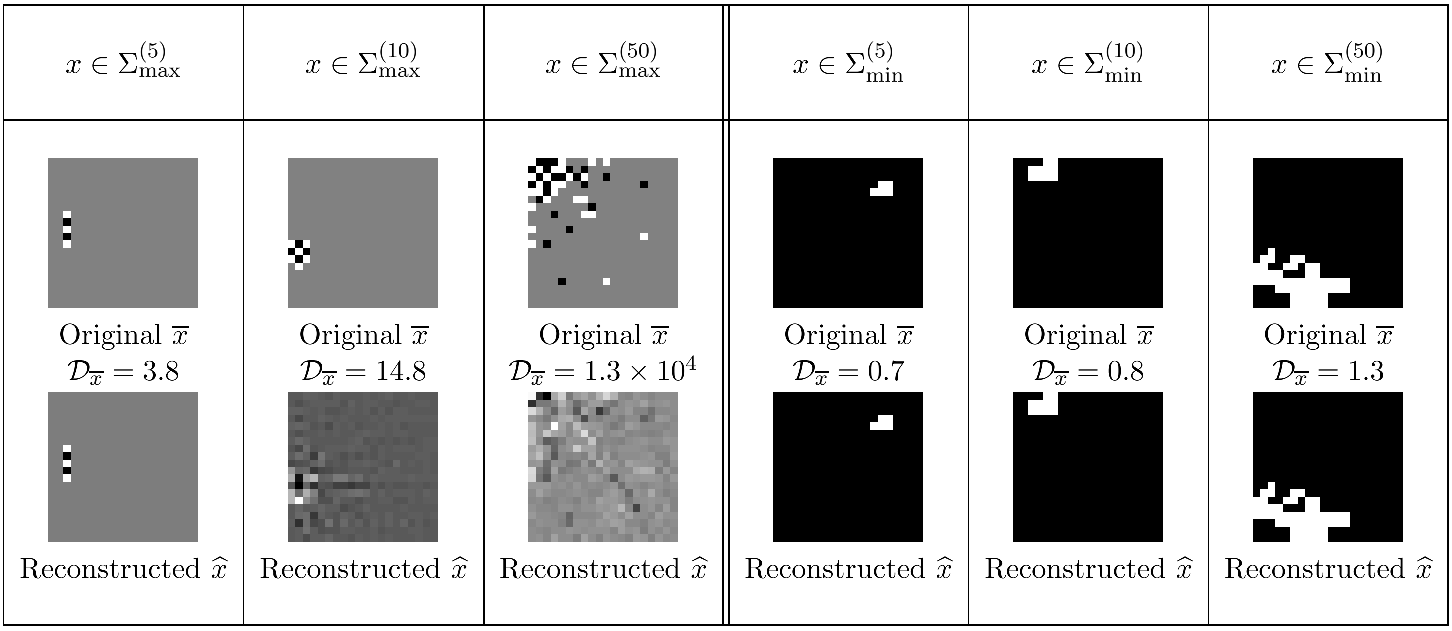

Figure 3: Reconstruction results from $z = A\overline{x}$ for sparse vectors with $s = 5$, $s = 10$, and $s = 50$ extracted with Algorithm 2 (1-3 columns) or with the "minimum version of Algorithm 2" (4-6 columns).

	6 views ($M = 294$)	9 views ($M = 441$)	12 views ($M = 588$)
Error $=0$ $\sigma^2 = 0$	11	12	12
Error $\leq 10^{-5}$ $\sigma^2 = 10^{-4}$	8	10	10
Error $\leq 10^{-1}$ $\sigma^2 = 10^{-2}$	8	10	10

Table 2: Sparsity s allowing us to recover every s-sparse vectors by ℓ_1-minimization with noise consideration and thus Fuchs/Tropp criterion. Results for three different configurations of the tomography-like matrix A.

[4] Y. E. Nesterov, "Primal-dual subgradient methods for convex problems," *Math. Program.*, vol. 120, no. 1, pp. 221–259, 2009.

[5] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Inform. Theory*, vol. 47, no. 7, pp. 2845–2862, Nov. 2001.

[6] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[7] J.-J. Fuchs, "On sparse representations in arbitrary redundant bases," *IEEE Trans. Inform. Theory*, vol. 50, no. 6, pp. 1341–1344, Jun. 2004.

[8] J. A. Tropp, "Just relax: Convex programming methods for identifying sparse signals in noise," *IEEE Trans. Inform. Theory*, vol. 52, pp. 1030–1051, Mar. 2006.

[9] David L. Donoho, "Neighborly polytopes and sparse solutions of underdetermined linear equations," Tech. Rep., Department of Statistics, Stanford University, 2004.

[10] E. J. Candès and J. Romberg, "Practical signal recovery from random projections," in *Wavelet Applications in Signal and Image Processing XI, Proc. SPIE*, 2004, vol. 5914.

[11] D. Liang, H. F. Zhang, and L. Ying, "Compressed-sensing photoacoustic imaging based on random optical illumination," *Int. J. Functional Informatics and Personalised Medicine*, vol. 2, no. 4, pp. 394–406, 2009.

[12] Y. Lu, X. Zhang, A. Douraghy, D. Stout, J. Tian, T. F. Chan, and A . F. Chatziioannou, "Source reconstruction for spectrally-resolved bioluminescence tomography with sparse a priori information," *Optics Express*, vol. 17, no. 10, pp. 8062–8080, 2009.

[13] Ch. Dossal, G. Peyré, and J. Fadili, "A numerical exploration of compressed sampling recovery," *Linear Algebra and Appl.*, vol. 432, no. 7, pp. 1663–1679, 2010.