# Spectrum Insensitive Sparse Recovery with Iterative Affine Projections

Nicolae Cleju, Iulian B. Ciocoiu
Faculty of Electronics, Telecommunications and Information Technology
"Gheorghe Asachi" Technical University of Iasi
700506, Iasi, Romania
Email: {nikcleju, iciocoiu}@etti.tuiasi.ro

***Abstract*—We propose a novel greedy algorithm for $\ell_0$-based sparse signal recovery, inspired by Iterative Hard Thresholding, which alternates a gradient descent step towards minimizing the sparsity error with a projection step on the affine solution space $\mathbf{y} = A\mathbf{x}$. We provide a theoretical guarantee based on Restricted Isometry Property for successful recovery of exact sparse signals, in the noiseless case, which does not depend on the singular values spectrum of the dictionary. This improves signal recovery by providing robustness in case of ill-conditioned dictionaries, as learned and coherent dictionaries tend to be. Simulation results on noiseless exact-sparse recovery indicate improvements compared to similar algorithms, especially in the case of ill-conditioned dictionaries.**

## I. Introduction

Sparse signal recovery [1], [2] is a well known problem that has been studied extensively in the last two decades. Given a signal $\mathbf{x} \in \mathbb{R}^n$ that is sparse, $\|x\|_0 = s$, the goal is to recover it from a reduced set of $m$ linear measurements arranged as rows in a acquisition matrix $A \in \mathbb{R}^{m \times n}$, possibly affected by additive noise $e$:

$$\mathbf{y} = A\mathbf{x} + \mathbf{e}$$

The signal $\mathbf{x}$ is recovered from the smaller dimensional measurement vector $\mathbf{y}$, in general, via an optimization problem seeking to find the sparsest solution to the equation system:

$$\mathbf{x} = \arg\min_x \|x\|_p \text{ subject to } \|\mathbf{y} - A\mathbf{x}\|_2 \leq \epsilon.$$

The recovery problem can be formulated in a different number of ways, all expressing the fundamental goal of finding the sparsest representation of $\mathbf{y}$ within the dictionary $A$. Usually, the $\ell_0$ or the $\ell_1$ norm are used as a sparsity measure. In this paper we focus on the noiseless case, which means the quadratic constraint becomes an exact constraint $\mathbf{y} = A\mathbf{x}$, and we use the $\ell_0$ norm. In this case the optimization problem becomes:

$$\mathbf{x} = \arg\min_x \|x\|_0 \text{ subject to } \mathbf{y} = A\mathbf{x}. \tag{1}$$

The recovery problem carries over when signal $\mathbf{x}$ is not itself sparse in the time domain, but has a sparse representation $\gamma$ in some basis or overcomplete dictionary $D$, i.e. $\mathbf{x} \approx D\gamma$, and thus we have $\mathbf{y} \approx AD\gamma$. In this case, first the sparse decomposition $\gamma$ is obtained as the solution of the core optimization problem, where the product $A_{ef} = A \cdot D$ acts as an effective dictionary or acquisition matrix for the sparse representation vector. The signal $\mathbf{x}$ itself is then found from its representation $\gamma$. As the core optimization is the same, in this paper we take (1) to cover both cases, with $\mathbf{x}$ denoting the sparse signal and $A$ the dictionary or the effective dictionary, depending on the problem at hand.

In general, recovery is successful only if the columns of $A$ satisfy certain incoherence properties [3]. The Restricted Isometry Property (RIP) [4] is a well known way of capturing this behavior. The matrix $A$ is said to satisfy RIP of order $s$ with constant $\delta_s$ if, for all $s$-sparse vectors $\mathbf{x}$, it holds that:

$$(1 - \delta_s)\|x\|_2^2 \leq \|A\mathbf{x}\|_2^2 \leq (1 + \delta_s)\|x\|_2^2$$

Alternatively, this can be expressed as a bound on the spectral norm of all submatrices $A_s$ obtained from $s$ columns of $A$ [5]:

$$\|I - A_s^T A_s\|_{2\to 2} \leq \delta_s,$$

where $\|\cdot\|_{2\to 2}$ designates the spectral norm of a matrix. For a matrix with small RIP constant, all submatrices of $s$ columns are close to orthonormal, which keeps the norm $\|A\mathbf{x}\|_2$ close to that of $\mathbf{x}$. A dictionary having a small enough RIP constant is a sufficient condition for the success of most of the sparse recovery algorithms in literature, see e.g.[5].

There are a large number of approaches for solving (1). Of particular interest for this paper is the well known Iterative Hard Thresholding (IHT) algorithm [6]. IHT repeatedly performs the following operation:

$$\mathbf{x}^{k+1} = H_s(\mathbf{x}^k + \mu \cdot A^T(x - A\mathbf{x}^k)), \tag{2}$$

where $H_s(\cdot)$ is the hard thresholding operator, that keeps only the $s$ absolute largest entries of a vector and zeroes all the others. The iteration (2) consists of an update step followed by a thresholding step. The update equation

$$\mathbf{v}^k = \mathbf{x}^k + \mu \cdot A^T(x - A\mathbf{x}^k)$$

represents a gradient descent step of size $\mu$, from the current $\mathbf{x}^k$ towards minimizing the error term $\|y - A\mathbf{x}^k\|_2^2$. The resulting vector $\mathbf{v}^k$ is then hard-thresholded, producing an improved $s$-sparse candidate, and the process is repeated. An adequate choice of a step-size is important for ensuring convergence, and thus adaptive ways of choosing $\mu$ have been proposed [7].

978-9-0827-9703-9/19/$31.00 ©2019 IEEE

## II. THE ITERATIVE AFFINE PROJECTION ALGORITHM

Our approach starts from observing that in IHT the hard thresholding operator $H_s(\cdot)$ actually accomplishes two things simultaneously:

(i) it determines the candidate support set at the current iteration, $T^k$, consisting of the locations of the largest absolute values, and

(ii) it projects orthogonally the input vector on the subspace $I_{T^k}$ spanned by the canonical basis vectors corresponding to $T^k$.

In general, having a candidate support set $T^k$ obtained after step (i), the quality of any candidate solution $\mathbf{x}^k$ (not necessarily $s$-sparse itself) is determined by two error terms:

$$E(\mathbf{x}^k) = \|\mathbf{y} - A\mathbf{x}^k\|_2^2 + \lambda \|\mathbf{x}^k_{T_c^k}\|_2^2 \tag{3}$$

The first term measures how accurate is the input signal $\mathbf{y}$ represented, whereas the second term quantifies how much of the candidate solution extends outside the presumed support. Here, the set $T_c^k$ denotes the cosupport, i.e. set of atoms outside the candidate support $T^k$, and the notation $\mathbf{x}_{T_c^k}$ designates the restriction of $\mathbf{x}$ to the entries from the cosupport. Throughout the paper we refer to the two terms as the *representation error* and the *sparsity error* of the candidate $\mathbf{x}^k$. The weighting factor $\lambda$ reflects the relative importance attributed to these errors. Note that an $s$-sparse exact solution $\mathbf{x}^*$, if it exists, satisfies both terms exactly so that the overall error is zero.

In IHT, the hard thresholding operation ensures that the sparsity error of $\mathbf{x}^k$ will be rigorously 0, via orthogonal projection on the canonical subspace $T^k$, resulting in an exact $s$-sparse signal. The non-zero representation error is tolerated, and is used for the gradient descent step that follows.

In this paper we propose taking the alternative route. Instead of enforcing the sparsity term via orthogonal projection, we choose to enforce the representation term $\|\mathbf{y} - A\mathbf{x}^k\|_2^2 = 0$, via orthogonal projection on the affine solution space of $\mathbf{y} = A\mathbf{x}$. The result is a vector $\mathbf{x}^k$ having a non-zero sparsity error $\|\mathbf{x}^k_{T_c^k}\|_2^2$, which is then reduced by gradient descent. This mirrors the behavior of IHT, but with the sparsity and representation error terms exchanged.

We refer to this algorithm as Iterative Affine Projection (IAP), summarized as Algorithm 1. An IAP iteration starts with a current candidate solution $\mathbf{x}^k$ from the affine solution space of $y = A\mathbf{x}$, and ends with the next candidate solution $\mathbf{x}^{k+1}$ from the same affine space. All candidate solutions live in this affine space, i.e. they satisfy the equation $y = A\mathbf{x}$.

1. **Gradient descent**. Since the current solution $\mathbf{x}^k$ satisfies the representation term exactly, it must be non $s$-sparse (otherwise it is already the final solution). A gradient descent step of size $\mu$ is taken to reduce its sparsity error $\|\mathbf{x}^k_{T_c^k}\|_2^2$:

$$\mathbf{v}^k \leftarrow \mathbf{x}^k - \mu \cdot \mathbf{x}^k_{T_c^k}$$

The smallest entries of $\mathbf{x}^k$ are shrinked with a factor of $\mu$, while the $s$ largest absolute values are preserved. Note that if $\mu = 1$ this amounts to hard-thresholding the vector $\mathbf{x}^k$, but this value is not mandatory.

2. **Orthogonal projection**. Bring the result $\mathbf{v}^k$ back to the affine solution space via orthogonal projection:

$$\mathbf{x}^{k+1} \leftarrow \mathbf{x}^0 + N^T N \mathbf{v}^k,$$

Here $\mathbf{x}^0 = A^\dagger \mathbf{y}$ is the least-squares solution of the system, and $N$ denotes a matrix whose rows form a basis of the null space of $A$. For simplicity, we take the rows of $N$ to be orthonormal, such that $N^T N$ represents the projection operator on the null space of $A$. The affine solution space of $\mathbf{y} = A\mathbf{x}$ consists of the least-squares solution $\mathbf{x}^0$ plus all the vectors from the null space.

**Algorithm 1** Iterative Affine Projection (IAP)

**Input:** $A =$ system matrix
**Input:** $N =$ orthonormal basis for the null space of $A$
**Input:** $y =$ input signal

$k \leftarrow 0$
$\mathbf{x}^0 \leftarrow A^\dagger \mathbf{y}$
**while** not finished **do**
  Gradient descent step (shrink) to reduce sparsity error

$$\mathbf{v}^k \leftarrow \mathbf{x}^k - \mu \cdot \mathbf{x}^k_{T_c^k}$$

  Project back on affine solution space

$$\mathbf{x}^{k+1} \leftarrow \mathbf{x}^0 + N^T N \mathbf{v}^k$$

**end while**

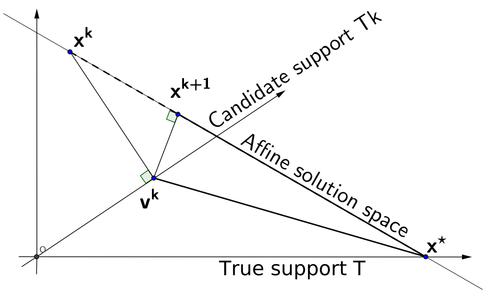



Fig. 1: Graphical depiction of the IAP iteration with $\mu = 1$, in a 3D space. Step 1: $\mathbf{x}^k$ is hard thresholded to $\mathbf{v}^k$. Step 2: $\mathbf{x}^k$ is projected on the affine solution space, resulting in the next candidate

Replacing $\mathbf{v}^k$ with its definition leads to the convenient single-line update equation covering both steps:

$$\begin{aligned} \mathbf{x}^{k+1} &\leftarrow \mathbf{x}^0 + N^T N (\mathbf{x}^k - \mu \cdot \mathbf{x}^k_{T_c^k}) \\ &\leftarrow \mathbf{x}^k - \mu \cdot N^T N \mathbf{x}^k_{T_c^k}. \end{aligned} \tag{4}$$

Here, we have used the fact that $\mathbf{x}^k$ is itself in the affine solution space, so it holds that $\mathbf{x}^k = \mathbf{x}^0 + N^T N \mathbf{x}^k$.

A conceptual graphical representation of a IAP iteration is depicted in Fig. 1, for a step size $\mu = 1$. Candidate $\mathbf{x}^k$ lives in the affine solution space of $\mathbf{y} = A\mathbf{x}$. It is shrinked towards the candidate support $T^k$, which may be different than the true support $T$. Here, $\mu = 1$ so the result $\mathbf{v}^k$ is actually $s$-sparse, but for different sizes of $\mu$ the vector $\mathbf{v}^k$ can be outside the subspace spanned by $T^k$. Then $\mathbf{x}^{k+1}$ is the projection of $\mathbf{v}^k$ on the affine solution space, advancing towards the true solution $\mathbf{x}^*$. The true solution $\mathbf{x}^*$ lives in the same affine space, and it is $s$-sparse itself.

## III. Theoretical guarantee for exact recovery

Consider a signal $\mathbf{y}$ having an $s$-sparse representation $\mathbf{y} = A\mathbf{x}^\star, \|\mathbf{x}^\star\|_0 = s$. We derive a theoretical guarantee for successful recovery of $\mathbf{x}^\star$ in the noiseless case, based on the RIP constant of order $2s$, not of the dictionary itself but of its right singular matrix. In this proof, we use a step size $\mu = 1$. The proof is inspired by the analysis in [5].

**Theorem III.1.** *Let $\mathbf{x}^*$ be an unknown $n$-dimensional $s$-sparse signal, $\|\mathbf{x}^*\|_0 = s$. Let $A$ be a $m \times n$ matrix, $m \leq n$, and $\mathbf{y} = A\mathbf{x}$. Consider an SVD decomposition $A = USV^T$, with $V^T$ in reduced form (of size $m \times n$).*

*If the RIP constant of order $2s$ of $V^T$ satisfies*

$$\delta_{2s}^{V^T} < \frac{1}{2} \tag{5}$$

*then the IAP algorithm converges linearly to the true solution $\mathbf{x}^*$.*

*Proof.* Consider the candidate solution at step $k$, $\mathbf{x}^k$, the step towards the $s$-sparse solution, $\mathbf{v}^k$, the next solution at step $(k+1)$, $\mathbf{x}^{k+1}$, and the true solution $\mathbf{x}^*$. We refer to the graphical representation in Fig. 1 for an illustration.

The proof is based on comparing the lengths of the following vectors: (i) $\mathbf{x}^{k+1} - \mathbf{x}^*$, (ii) $\mathbf{v}^k - \mathbf{x}^*$, and (iii) $\mathbf{x}^k - \mathbf{x}^*$. In particular we focus on showing that $\|\mathbf{x}^{k+1} - \mathbf{x}^*\|_2 < \|\mathbf{v}^k - \mathbf{x}^*\|_2 < \|\mathbf{x}^k - \mathbf{x}^*\|_2$.

For proving the first inequality, observe that the difference $\mathbf{x}^* - \mathbf{x}^{k+1}$ is the projection on the null space of the vector difference $\mathbf{x}^* - \mathbf{v}^k$, since $\mathbf{x}^{k+1}$ is the orthogonal projection of $\mathbf{v}^k$ on the affine solution space, and $\mathbf{x}^*$ is a point belonging to the same affine space. We have thus

$$\begin{aligned} \|\mathbf{x}^{k+1} - \mathbf{x}^*\|_2 &= \|N^T N(\mathbf{v}^k - \mathbf{x}^*)\|_2 \\ &= \|(I - A^\dagger A)(\mathbf{v}^k - \mathbf{x}^*)\|_2, \end{aligned}$$

where we have rewritten the null space projection operator $N^T N$ in a more convenient form $(I - A^\dagger A)$.

Considering the SVD decomposition $A = USV^T$ , we can rewrite $A^\dagger A = VV^T$. Now, observe that both vectors on the right side, $\mathbf{v}^k$ and $\mathbf{x}^*$, are $s$-sparse. With their common support defined as $L = supp(\mathbf{x}^k) \cup supp(\mathbf{x}^*)$, $|L| \leq 2s$, in the right side term of the equation we can restrict $V$ to the support set $L$ , as follows:

$$\|\mathbf{x}^{k+1} - \mathbf{x}^*\|_2 = \|(I - V_L V_L^T)(\mathbf{v}^k - \mathbf{x}^*)\|_2,$$

We use the equivalence $\delta_{|L|}^{V^T} = \|(I - V_L V_L^T)\|_{2\to 2}$ to further rewrite:

$$\begin{aligned} \|\mathbf{x}^{k+1} - \mathbf{x}^*\|_2 &= \|(I - V_L V_L^T)(\mathbf{v}^k - \mathbf{x}^*)\|_2 \\ &\leq \|(I - V_L V_L^T)\|_{2\to 2} \|(\mathbf{v}^k - \mathbf{x}^*)\|_2 \\ &\leq \delta_{2s}^{V^T} \|(\mathbf{v}^k - \mathbf{x}^*)\|_2 \end{aligned}$$

The proof of the second inequality is as follows: since $\mathbf{v}^k$ is the hard thresholding of $\mathbf{x}^k$, it is closer to $\mathbf{x}^k$ than any other $s$-sparse vector, including $\mathbf{x}^*$. Then $\|\mathbf{x}^k - \mathbf{v}^k\|_2 \leq \|\mathbf{x}^k - \mathbf{x}^*\|_2$. Considering now the triangle inequality in the triangle $(\mathbf{x}^k, \mathbf{v}^k, \mathbf{x}^*)$, the length of side $\mathbf{v}^k - \mathbf{x}^*$ is smaller than the sum of the other two, and since $\|\mathbf{x}^k - \mathbf{x}^*\|_2$ is larger, we have:

$$\|\mathbf{v}^k - \mathbf{x}^*\|_2 \leq 2\|(\mathbf{x}^k - \mathbf{x}^*)\|_2$$

Putting both inequalities together we have:

$$\begin{aligned} \|\mathbf{x}^{k+1} - \mathbf{x}^*\|_2 &\leq \delta_{2s}^{V^T} \|\mathbf{v}^k - \mathbf{x}^*)\|_2 \\ &\leq 2\delta_{2s}^{V^T} \|\mathbf{x}^k - \mathbf{x}^*\|_2 \end{aligned}$$

Hence, if $V^T$ has a RIP constant

$$\delta_{2s}^{V^T} < \frac{1}{2},$$

the error between the new candidate $\mathbf{x}^{k+1}$ and the true solution $\mathbf{x}^*$ decreases at every step with at least a constant factor, and the algorithm converges linearly to the true solution $\mathbf{x}^*$. □

The theoretical guarantee (5) shows that the IAP algorithm depends on the RIP properties of the tight frame $V^T$ formed by the most significant right singular vectors of the dictionary matrix $A$. As such, given an SVD decomposition $A = USV^T$, the algorithm is insensitive to the singular values spectrum $S$. This "spectral insensitivity" is beneficial for learned and coherent dictionaries which are usually significantly ill-conditioned, having wide-ranging singular values. This feature is specifically tested in Tests 2 and 3 from Section IV.

## IV. Simulation results

We test the exact recovery of $s$-sparse signals with IAP against well known $\ell_0$-based iterative algorithms: Iterativa Hard Thresholding (IHT), Orthogonal Matching Pursuit (OMP) [8], [9], and Approximate Message Passing (AMP) [10].

The first test is a synthetic test with i.i.d. random generated Gaussian dictionaries of size $m \times n$, thus very close to being tight frames. We plot the phase transition image depicting the percentage of exactly recovered sparse signals as a function of the compression ratio $\delta = \frac{m}{n}$ and relative sparsity $\rho = \frac{s}{m}$. For each ($\delta$, $\rho$) pair we generate a random Gaussian dictionary, 20 exact $s$-sparse signals $\mathbf{x}$ of size $n = 200$, and we attempt recovery of the sparse signals $\mathbf{x}$ from the measurements vector $\mathbf{y} = A\mathbf{x}$ with the algorithms under test. Exact recovery is defined as a having a relative error less than $10^{-6}$ of the norm of the true $\mathbf{x}$. The percentage of exactly recovered signals is depicted graphically, with pure white indicating $100\%$, pure black $0\%$, and shades of gray — intermediate values.

Fig.3 depicts the phase transitions plots obtained with IAP, IHT, OMP and AMP. For IAP we use step size $\mu = 1$, for IHT we use a variant with adaptive step size [7], since it consistently provides better results. The results show IAP

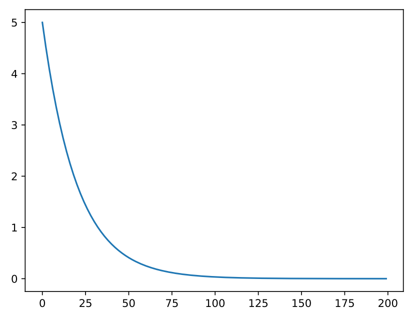



Fig. 2: Exponentially decaying spectrum of singular values used for the random dictionaries in Test 2

to have above average results, but all the algorithms perform rather well.

The second tests verifies the spectral insensitivity of IAP, in a compressed sensing scenario. For each ($\delta$, $\rho$) pair we generate a random Gaussian dictionary $D$ of size $n \times n = 200 \times 200$, but we replace the singular values with an exponential decaying function as in Fig.2. The atoms are then normalized. We generate random $s$-sparse data as exact linear combinations of the atoms of $D$, $\mathbf{x} = D\gamma$, $\|\gamma\|_0 = s$, which are then multiplied with a random Gaussian acquisition matrix $P$ of size $m \times n$. We then attempt recovery of the sparse decomposition vector $\gamma$, from which we reconstruct the signals $\mathbf{x}$ themselves. Exact recovery is again defined as a having a relative error less than $10^{-6}$ of the norm of the true $\mathbf{x}$.

The results obtained in Fig.4 clearly prove that IAP is essentially immune to the ill-conditioned nature of a dictionary, whereas the other algorithms are significantly affected.

Finally, in a third test we replace the random dictionaries with a dictionary of $8 \times 8$ pixel image patches learned with the K-SVD [11] algorithm from the USC-SIPI Miscellaneous dataset, which better illustrates a real-life dictionary with coherence and possible dependencies between the atoms. The dictionary size is $64 \times 80$. As before, from the dictionary we generate exact-sparse signals as random combinations of the atoms, and we test the algorithms' recovery performance of the signals from random projections.

The results in Fig.5 show that exact recovery of sparse signals with the learned dictionary is much more problematic. However, when it comes to comparing the algorithms, IAP clearly surpasses both IHT and OMP in this test, with a smaller margin over AMP as well.

## V. Conclusions

This paper presents Iterative Affine Projection, a proposed sparse recovery algorithm similar to Iterative Hard Thresholding, but with a different internal handling of the sparsity and representation error terms. The algorithm is based on iterative projections on the affine solution space of the system, alternating with gradient descent steps for improving sparsity.

We prove a theoretical guarantee for the success of noiseless recovery of sparse signals, which is based on a RIP condition on the right singular matrix of the dictionary, and thus is independent of the dictionary's singular value spectrum. This feature provides increased robustness to ill conditioned dictionaries, which are often encountered with learned dictionaries.

Noiseless recovery tests against some of the well-known greedy algorithms in literature show notable gains for various types of dictionaries. Significant improvements are observed in the case of ill-conditioned dictionaries, but also smaller improvements even in the case of well-conditioned ones. The results suggest that the proposed algorithm may be a viable solution for sparse signal recovery in general.

## Acknowledgment

This work was supported by a grant of Ministery of Research and Innovation, CNCS - UEFISCDI, project number PN-III-P1-1.1-PD-2016-0127, within PNCDI III.

## References

[1] E. J. Candès, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, pp. 1207–1223, 2006.

[2] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[3] E. J. Candes and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse Problems*, vol. 23, no. 3, pp. 969–985, Jun. 2007.

[4] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathematique*, vol. 346, pp. 589–592, 2008.

[5] S. Foucart, "Sparse recovery algorithms: Sufficient conditions in terms of restricted isometry constants," in *Approximation Theory XIII: San Antonio 2010*, 2012, pp. 65–77.

[6] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Applied and Computational Harmonic Analysis*, vol. 27, no. 3, pp. 265–274, 2009.

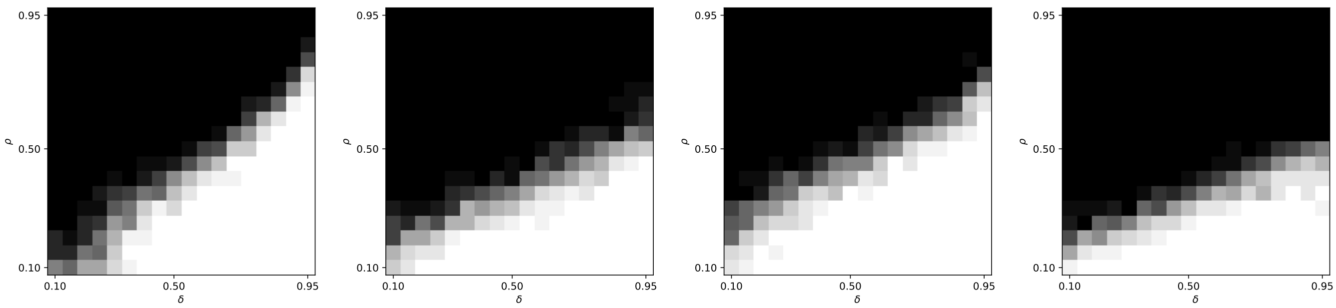
[7] T. Blumensath, "Accelerated iterative hard thresholding," *Signal Processing*, vol. 92, no. 3, pp. 752–756, 2012.

[8] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. 27th Annual Asilomar Conf. on Signals, Systems, and Computers*, 1993, pp. 40–44.

[9] G. Davis, S. Mallat, and M. Avellaneda, "Adaptive greedy approximations," *Constructive approximation*, pp. 57–98, 1997.

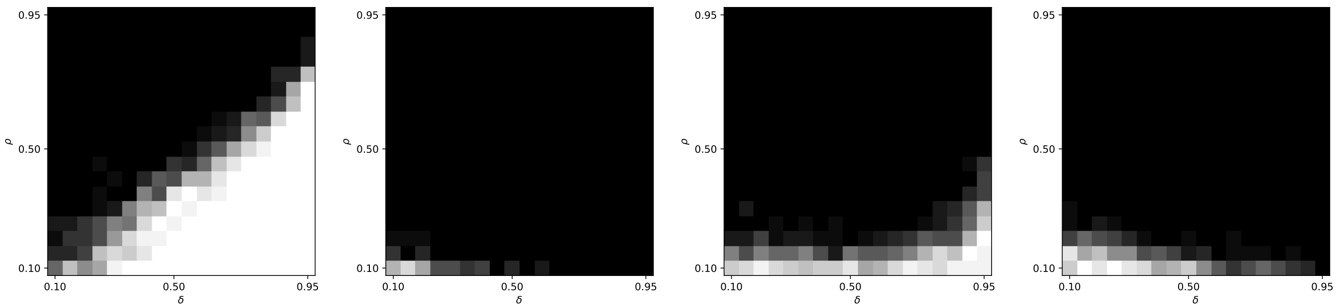
[10] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proceedings of the National Academy of Sciences*, vol. 106, no. 45, pp. 18914–18919, 2009.

[11] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, p. 4311, 2006.



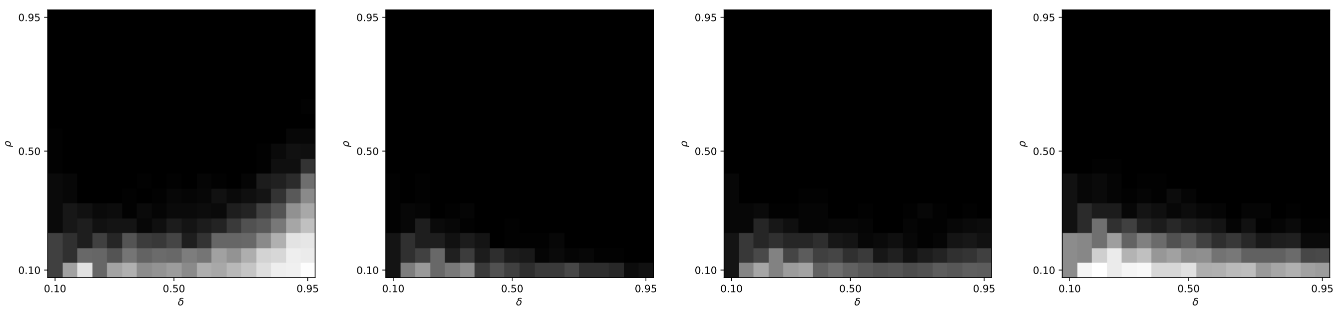

(a) IAP (b) IHT (adaptive $\mu$) (c) OMP (d) AMP

Fig. 3: **Test 1.** Percentage of exact-sparse signals reconstructed perfectly with different recovery algorithms, using random Gaussian dictionaries: (a) the proposed Iterative Affine Projection, (b) Iterative Hard Thresholding, (c) Orthogonal Matching Pursuit and (d) Approximate Message Passing. White indicates 100% recovered signals and black 0%.




(a) IAP (b) IHT (adaptive $\mu$) (c) OMP (d) AMP

Fig. 4: **Test 2.** Percentage of exact-sparse signals reconstructed perfectly with different recovery algorithms, using Gaussian dictionaries with exponentially decaying singular values: (a) the proposed Iterative Affine Projection, (b) Iterative Hard Thresholding, (c) Orthogonal Matching Pursuit and (d) Approximate Message Passing. White indicates 100% recovered signals and black 0%.




(a) IAP (b) IHT (adaptive $\mu$) (c) OMP (d) AMP

Fig. 5: **Test 3.** Percentage of exact-sparse signals reconstructed perfectly with different recovery algorithms, using a image patch dictionary learned with K-SVD: (a) the proposed Iterative Affine Projection, (b) Iterative Hard Thresholding, (c) Orthogonal Matching Pursuit and (d) Approximate Message Passing. White indicates 100% recovered signals and black 0%.