

RECONSTRUCTION IN COMPRESSIVE SENSING USING AFFINE SCALING TRANSFORMATIONS WITH VARIABLE- p DIVERSITY MEASURE

Rufino J. Domínguez^a, Sergio D. Cabrera^b, J. Gerardo Rosiles^b, and Javier Vega-Pineda^a

^aDiv. de Posgrado en Electrónica, Inst. Tec. de Chihuahua, 31310 Chihuahua, Chih. MÉXICO

E-mail: rjdominguez@itchihuahua.edu.mx

^bDept. of Electrical and Comp. Eng., University of Texas at El Paso, El Paso, TX 79968-0523 USA

ABSTRACT

The *Affine Scaling Transformation* (AST) family of algorithms can solve the minimization of the $\ell_{(p \leq 1)}$, p -norm-like diversity measure for an underdetermined linear inverse problem. The AST algorithms can therefore be used to solve the sparse signal recovery problem that arises in *Compressive Sensing*. In this paper, we continue to investigate the application of the iterative AST family of algorithms with a dynamical adjustment of the p parameter to improve convergence speed and signal recovery accuracy. In our previous work we experimentally determined that any p in $[0, 1]$ can give the sparse solution when exact recovery is possible, however, the behavior of the algorithm is highly dependent on this parameter. In addition, the *best-approximation* error, for those cases where exact recovery is not possible, is also highly dependent on p . Using various criteria, we propose and evaluate some strategies to vary the values of p as a function of the iteration in the AST algorithm. The goal in these strategies for a variable- p AST algorithm is to capture the benefits of the $p=0$ and the $p=1$ fixed- p approaches simultaneously.

Index Terms— Affine Scaling Transformation, diversity measure, compressive sensing, sparse signal recovery.

1. INTRODUCTION AND MOTIVATION

The problem addressed in this paper is the recovery of a discrete signal from a few linear measurements as encountered in the Compressive Sensing (CS) framework. We use the CS notation as presented in [1] and focus on the Affine Scaling Transformation (AST) family of algorithms. Let M be the number of linear measurements used in recovery of a signal with N total samples, with K non-zero components in the transform domain, such that $K < M \ll N$. We call the measurements matrix Φ , the resulting measurements vector \underline{y} , which is taken from a signal \underline{x} which is sparse in a specific transform domain. Thus, the desired solution \underline{x} is the sparse transform so that, $\underline{x} = \Psi \underline{z}$ and Ψ is the

unitary $N \times N$ transform matrix. The problem to be solved is thus to find \underline{x} from \underline{y} from the underdetermined linear system:

$$\Phi \Psi^H \underline{x} = \underline{y} \quad \text{or} \quad \Theta \underline{x} = \underline{y} \quad (1)$$

Since there are an infinite number of solutions for \underline{x} , looking for sparse, or the sparsest solution, reduces the set of solutions of interest and the problem is more restricted because of this additional goal. The common approach in CS is to find sparse solutions by minimizing the ℓ_1 norm of the signal \underline{x} , problem known as *Basis Pursuit* [9]:

$$\min \|\underline{x}\|_1 \quad \text{subject to} \quad \Theta \underline{x} = \underline{y}$$

The greedy algorithms are another strategy used frequently [12]. Our approach is to use the AST family of algorithms to minimize the $\ell_{(p \leq 1)}$ diversity measure as described in [11], and therefore, it includes the ℓ_1 norm minimization as a special case. More importantly, the parameter p can be used to improve the properties of the algorithm and to evaluate the alternatives in the best-approximation problem.

In AST, the problem to solve is:

$$\min E^{(p)}(\underline{x}) \quad \text{subject to} \quad \Theta \underline{x} = \underline{y} \quad (2)$$

where $E^{(p)}(\underline{x})$ is the diversity measure defined as:

$$E^{(p)}(x) = \text{sgn}(p) \sum_{i=1}^N |x_i|^p \quad (3)$$

The strategy in AST algorithm is performing a change of variable:

$$\underline{q} = \mathbf{W}^{-1} \underline{x} \quad \text{or} \quad \underline{x} = \mathbf{W} \underline{q} \quad (4)$$

where \mathbf{W} is a matrix defined by:

$$\mathbf{W} = \text{diag}(|x_i|^{1-p/2}) \quad (5)$$

Then, we have a transformed problem in terms of the variable \underline{q} :

$$\min E^{(p)}(\mathbf{W} \underline{q}) \quad \text{subject to} \quad \Theta \mathbf{W} \underline{q} = \bar{\Theta} \underline{q} = \underline{y} \quad (6)$$

To perform this minimization, at each iteration of the AST a *pseudoinverse* is computed using $\bar{\Theta}$, a modified Θ matrix whose columns are weighted by the magnitude of the previous solution. Thus, columns of Θ that contributed significantly in the previous solution are favored to be contributors in the next solution. Successive weighting produces a successive de-emphasis of components that are not present in most previous solutions and in the limit, the number of surviving components is a minimum (at most $M=\text{rank}(\Theta)$). The procedure is equivalent to solving a series of underdetermined problems $\bar{\Theta}^{(k-1)} \underline{q}^{(k-1)} = \underline{y}$ by minimizing a weighted ℓ_2 -norm per iteration to obtain $\underline{x}^{(k)} = \mathbf{W}^{(k-1)} \underline{q}^{(k-1)}$ [3]. The convergence analysis of the AST algorithm can be found in [10] and [11].

In [2], the authors presented an extensive empirical evaluation of AST for CS. They measured the mean and standard deviation of the number of iterations for convergence of the algorithm, for several realizations of the measurement matrix Φ . The experiment was repeated for different values of p in the $[0, 1]$. The authors found experimentally that the AST algorithm, for $p=1$, requires 3-5 times more iterations to converge to its solution than AST for $p=0$. On a separate experiment they observed the behavior of the M value. The minimal M needed for perfect recovery is approximately the same on the average for all values of $p < 1$, however, as p decreases from $p=1$ to $p=0$, a larger standard deviation is obtained over many repetitions of the experiment (each corresponds to a different measurement matrix Φ). In each solution for a sparse signal, AST is allowed to converge using a fixed value of p .

These observations were the motivation to modify the AST formulation to achieve the reduction in the number of the algorithm iterations, modification that we present in this paper. The goal was to lead the reduction to a similar level than the obtained in the $p=0$ case, and with less variation in the minimal M value for perfect recovery. The strategy consists in making p a changing parameter as the AST algorithm execution progresses in a manner that will gradually traverse the range from $p=1$ to $p=0$.

2. AST ALGORITHM WITH VARIABLE p .

Figure 1 shows the flow diagram of the AST with p a variable parameter within the algorithm. Basically, it is the same AST algorithm described in [2], but with an additional step, the adjustment of the p parameter at each iteration. Again, the desired approach is to have a scheme that will start in the first iteration with $p=1$ and then gradually decrement it toward zero in subsequent iterations.

We introduce three approaches for p variation during AST execution: (a) a linear variation, (b) a variation as a function of the convergence error, and (c) a variation as a function of

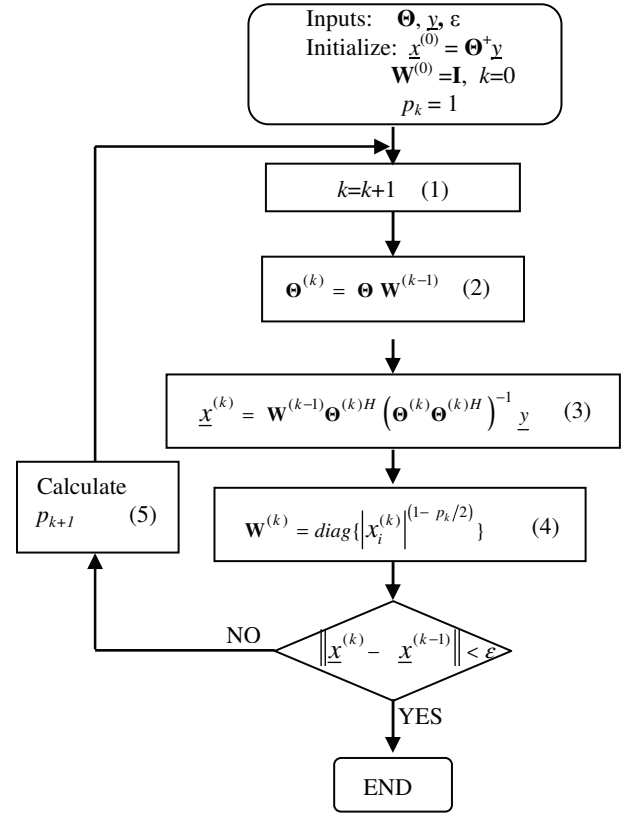


Figure 1. Flow diagram of the Affine Scaling Transformation (AST) with variable p .

the gradient of the convergence error. The first two strategies are empirical approaches, being the first attempts at adjusting the p parameter. They served as models to develop the third approach, a more mathematically founded strategy.

2.1. Linear variation of p .

This approach initializes p to the value 1 and then p is decremented by a fixed amount γ . If p reaches zero before the convergence is achieved, the remaining iterations use $p=0$. On the other hand, convergence may be achieved before p becomes zero. The p parameter adjustment is defined by:

$$p_{k+1} = \max\{p_k - \gamma, 0\} \quad (7)$$

In the case where $\gamma=0.1$, the p parameter becomes zero after 11 iterations.

2.2. Variation of p as a function of the convergence error.

The linear variation of p is independent of the behavior of the algorithm through the iterations. In order to make the variation of p sensitive to the general behavior of the

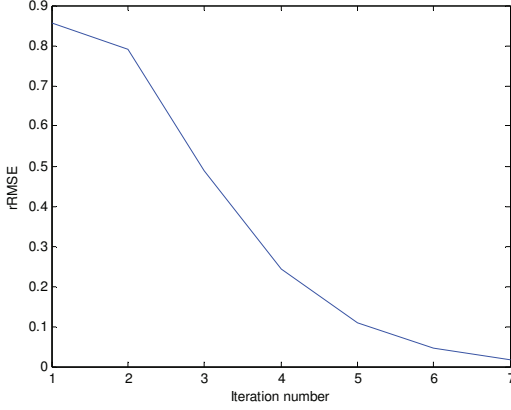


Figure 2.- Typical behavior of the $rRMSE_k$ measure for $p=1$.

algorithm and its convergence trend, we associate p with the convergence measure used to stop the algorithm. We use the norm of the difference between the solution calculated in the current iteration ($\underline{x}^{(k)}$ or simply \underline{x}_k) and the solution calculated in the previous iteration (\underline{x}_{k-1}). We call this measure $RMSE_k$, that is:

$$RMSE_k = \|\underline{x}_k - \underline{x}_{k-1}\| \quad (8)$$

If this norm is lower than a certain value (ε), we stop the iterations (see figure 1). In order to adjust p and manage values in the range $[0,1]$, we define the relative error:

$$rRMSE_k = \frac{\|\underline{x}_k - \underline{x}_{k-1}\|}{\|\underline{x}_k\|} \quad (9)$$

In our experiments it has been noted that during the first few iterations of AST, the $rRMSE_k$ value is largest compared with its value in the subsequent iterations. As the algorithm iterates, the $rRMSE_k$ value diminishes, becoming nearly zero at the end of the sequence of iterations, leading to convergence. This behavior is showed in the figure 2 which was generated for the fixed p AST using $p=1$. The second strategy to modify the p parameter at each iteration is to assign to p the value:

$$p_{k+1} = \min(rRMSE_k, 1) \quad (10)$$

2.3. Variation of p as a function of the gradient of the convergence error.

The proposed third strategy involves the gradient of the $rRMSE_k$ measure. That is, we can adjust the p parameter based on the rate/change in the $rRMSE_k$ measure. We use a scaled approximation of this gradient, computing the simple difference between the $rRMSE_k$ value in the current iteration, and the corresponding value in the previous iteration:

$$\nabla rRMSE_k = rRMSE_k - rRMSE_{k-1} \quad (11)$$

We propose the adjustment of the p parameter to proceed as follows:

$$p_{k+1} = p_k + \alpha \frac{\nabla rRMSE_k}{rRMSE_k} \quad (12)$$

where α is a positive constant which regulates the use of the gradient term.

To understand the motivation for this approach, we must keep in mind the way that AST works [11]. It has four main steps: 1- Perform a change of variable by an affine scaling transformation: $\underline{q}_k = \mathbf{W}^{-1}\underline{x}_k$, where $\mathbf{W} = \text{diag}(|x_k|^{1-p/2})$. 2- calculate a search direction using the steepest-descend method, 3- take a step in such direction (in the ' \underline{q} domain') to obtain \underline{q}_{k+1} , and 4- transform back to the ' \underline{x} domain' using $\underline{x}_{k+1} = \mathbf{W}\underline{q}_{k+1}$ to find the new solution point \underline{x}_{k+1} . In the step two the search direction is calculated as the negative of the gradient of the objective function (2) and projected to the null space of the matrix $\mathbf{\Theta}$, in order to maintain the feasibility imposed by the constraint in (2). In the step three, to move from \underline{q}_k to \underline{q}_{k+1} , is necessary to settle down a step-length. In [11] it is shown that the step-length implicit in the algorithm formulation shown in the figure 1 is $1/p_k$.

Now, we observe that (12) works like a feedback system. Note that p_{k+1} is a function of $rRMSE_k$, but $rRMSE_k$ is a function of p_k , due to the use of p_k in step (4) of figure 1. Let us consider the iteration k where the current parameter value is p_k , and the step length used to calculate \underline{x}_{k+1} is $1/p_k$. Under ideal conditions this causes the gradient term in (12) to be negative and p_{k+1} will be lower than p_k . The trend of the p value is diminishing, as is desired.

On the other hand, if the new value p_{k+1} become too low, then in the iteration $k+1$ the step length will be too long and the gradient term in (12) will be positive. This will cause that p_{k+2} to be bigger than p_{k+1} . As a consequence, the step length in iteration $k+2$ will be lower, the gradient term in (12) will be negative and p_{k+3} will be lower than p_{k+2} . Thus, the trend to decreasing p still holds. This feedback-like mechanism stabilizes the global diminishing trend of the p parameter.

3. EXPERIMENTS.

We are working on the problem of recovery of a harmonic signal with a sparse Discrete Fourier Transform (DFT) spectrum from measurements which are random projections. For the simulation, we generate a 256 sample signal \underline{s} from $K=4$ non-zero, randomly positioned DFT values of different amplitudes and phases. Thus, the transform matrix Ψ is unitary with entries given from:

$$\Psi_{r,c}^H = \frac{1}{\sqrt{N}} e^{j\frac{2\pi}{N}(r-1)(c-1)} \quad (13)$$

The M by N measurements matrix Φ is selected to have entries which are zero-mean Gaussian random values. Additionally, the rows of the matrix Φ are first orthonormalized, in order to fulfill the requirement of incoherence with the matrix Ψ . The coherence μ between two matrixes \mathbf{A} and \mathbf{B} of the same size is defined as [4, 5, 7]

$$\mu(\mathbf{A}, \mathbf{B}) = \sqrt{N} \cdot \max_{1 \leq k, j \leq N} \left| \langle a_k, b_j \rangle \right| \quad (14)$$

where \mathbf{A} and \mathbf{B} are both orthonormal matrixes. The vectors a_k and b_j are row and column vectors of \mathbf{A} and \mathbf{B} respectively. Using the definition (14), the coherence values are in the range $[1, \sqrt{N}]$. In [7] it is shown that the minimum number of linear measures M required for exact recovery of the signal is proportional to $\mu^2(\Phi_N, \Psi)$, where Φ_N is a complete measurements matrix which includes Φ as a submatrix of M rows. Thus, in compressive sensing, it is desirable for $\mu(\Phi, \Psi)$ to be as small as possible, in order for the lower bound on M to be as small as possible. That is, Φ_N and Ψ must be both incoherent matrixes.

3.1. Minimal M for perfect recovery and number of iterations.

We perform an experiment in order to determine the minimal value of M to obtain perfect recovery, that is, the minimal number of random projections required for perfect recovery of the sparse signal of length N with K non-zero transform domain components. We consider perfect recovery to occur when the norm of the difference between the original signal \underline{s} and the recovered signal $\hat{\underline{s}}$, divided by the norm of the original signal is less or equal to 0.005 or 0.5%. On the other hand, we also monitor and assess changes in the number of iterations required for the AST algorithm to converge. In this case, the $rRMSE$ must be below the value 0.001 or 0.1% to achieve convergence.

In order to compare our results with the results reported in [2], we worked on one of the same test signals and have included the results for the three approaches of AST with variable p described in section 2. To determine the minimal value of M for perfect recovery, we perform the recovery of a signal with $K=4$ components starting with $M=5$, and then incrementing this value one by one until the perfect recovery is achieved and register the last M value used. We repeat the entire process 70 times using the same harmonic signal \underline{s} but generating different measurement matrices Φ (always orthonormal, zero-mean Gaussian random values). This matrix is generated once with a maximum number of rows. As M increases, more rows are taken from this matrix. At the end we calculate the average of the resulting minimal M values, and the standard deviation as a measure of variability

Smallest M for Perfect Recovery, Gradient-RMSE Approach, Average = 20.3857 std is 5.9668

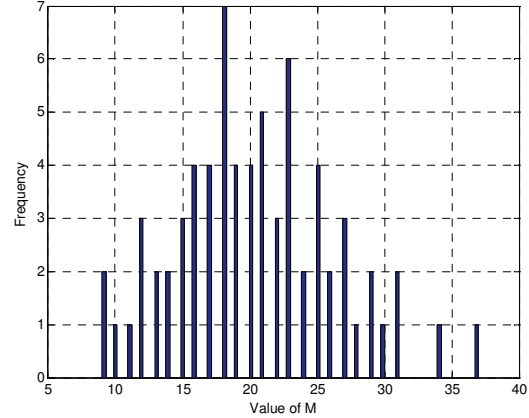


Figure 3. Frequency histogram of the minimal M for perfect recovery, using p variation as a function of the gradient of the error.

Number of iterations of AST, Gradient-RMSE Approach, mean is 14.3514 std is 4.1029

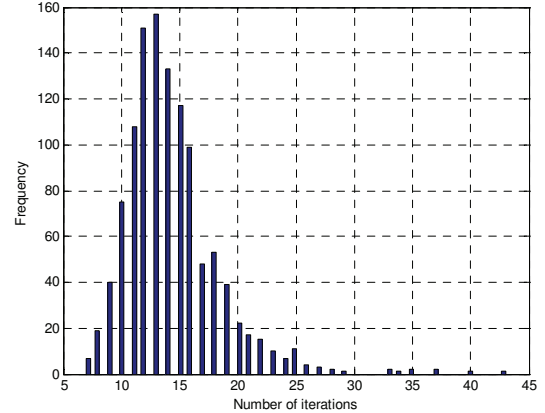


Figure 4. Frequency histogram of the number of iterations required for convergence of the AST with p variable algorithm as a function of the gradient of the error.

in this quantity. Furthermore, we register the number of iterations required for convergence each time the AST algorithm is executed within each cycle and their averages and standard deviations are computed.

Figure 3 shows for the gradient approach, the resulting histogram of minimal M for perfect recovery for all the 70 times that the experiment was repeated. The minimal M average for perfect recovery is 20, being the worst case equal to 37. The standard deviation is 5.97. Is remarkable that two times (the best cases), with just nine linear measures it was possible the perfect recovery of the 256 length signal. Figure 4 shows also for the gradient approach, the resulting histogram for the number of iterations required for convergence. The average number of iterations is 14, with a standard deviation of 4.1. The histograms for the other approaches are very similar. Table 1 resumes the results for all the approaches.

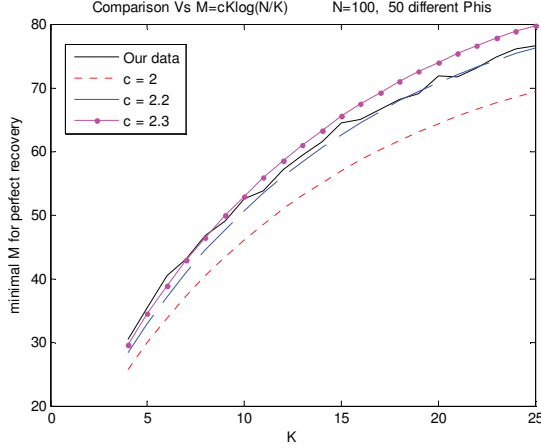


Figure 5. Comparison of the behavior of minimal M for perfect recovery as a function of K , with fixed N using AST algorithm, against the $M=cK\log(N/K)$ function.

3.2. Minimal M for perfect recovery as function of K and N .

In [6] it is showed that the number of linear measurements M required to have a successful recovery in a general Compressive Sensing framework is related with N (the signal length) and K (the signal diversity) as follows:

$$M \geq cK \log(N/K) \quad (15)$$

for some constant c , where the natural log function is used. We were interested in verify the consistency of AST with this inequality. The following experiment was performed. For N fixed, obtain the minimal M average for perfect recovery for different K values, repeating each K value several times for the same signal but with different measurement matrix Φ each time.

Figure 5 shows a comparison of the behavior of minimal M for perfect recovery as function of K , with fixed N (continuous black trace), against the $M=cK\log(N/K)$ function for three different values of the constant c . We used $N = 100$, K values from 4 to 25, (that is, the maximal diversity tested was the 25% of the signal length), and 50 repetitions of each K value with different Φ each time.

We observe in figure 5 that the behavior of the minimal M using the AST with p -variable algorithm (for the gradient of the error approach) follows very close to the relation $M=cK\log(N/K)$ for $c = 2.22$ approximately. This result confirms the consistency of the AST p -variable algorithm. In the conclusion section we comment more about this result.

3.3. Best approximation behavior.

In this experiment we observe the behavior of the different AST approaches when the M value is lower than the minimal M for perfect recovery, to see how close to the original signal their respective approximations are.

We define the $\%RMSE$ measure of the approximation of the recovered signal \underline{s}^* to the original signal \underline{s} as:

$$\%RMSE = \frac{\|\underline{s} - \underline{s}^*\|}{\|\underline{s}\|} \quad (16)$$

In the experiment we observe the behavior of the $\%RMSE$ value as function of M , for a wide enough range of the M values, due the variability of the minimal M for perfect recovery of the different AST approaches, including fixed AST and p -variable AST. For each value of M , we repeated the recovery 70 times, and compute the average of the $\%RMSE$ and its standard deviation. We repeat this for all approaches of AST.

Figure 6 shows the plot of the $\%RMSE$ average as function of M . Figure 7 shows the plot of the $\%RMSE$ standard deviation as function of M . From this perspective, the linear approach and the fixed p AST with $p = 1$ have the better behavior, in $\%RMSE$ average as well as standard deviation.

The AST p -variable as function of the convergence error ($rRMSE$) and Gradient approaches have the second place respect the $\%RMSE$ average, but with poor standard deviation behavior. The worst case in both, $\%RMSE$ average and standard deviation is for Fixed AST with $p=0$ approach.

4. RESULTS AND CONCLUSIONS.

Table 1 shows the summary of all the methods including the fixed p AST for $p=0$ and $p=1$, in order to compare with the variable p AST versions. We indicate the best and worst approach for each quality metric. The test signal used is the same as in [2] of length $N=256$ with $K=4$ non-zero DFT components. The improvement respect the fixed- p approaches is seen in terms of the simultaneous reduction in the number of iterations to a level close to the best ($p=0$ case) and the reduction in mean and std for the minimum M for perfect recovery to nearly the same as the best ($p = 1$ case). For the linear AST, the parameter p is reduced from $p=1$ to $p=0$ after 11 iterations. For the gradient-based AST, the parameter α is chose to have the value 0.4.

In summary, the results tell us that the variable- p approaches are good compromises that combine some of the advantages of the two fixed- p approaches for $p=0$ and $p=1$. Among the variable p approaches, there is no single winner. The best in terms of both M metrics is the linear while the best in terms of average iterations is the gradient-based method.

We also showed that the AST algorithm follows closely the relation between M and N and K given by (15), as is stated in [6]. In that work the authors found experimentally that (15) holds for $3 \leq c \leq 5$ using the Basis Pursuit approach. In our experiments with the AST algorithm, we found that it

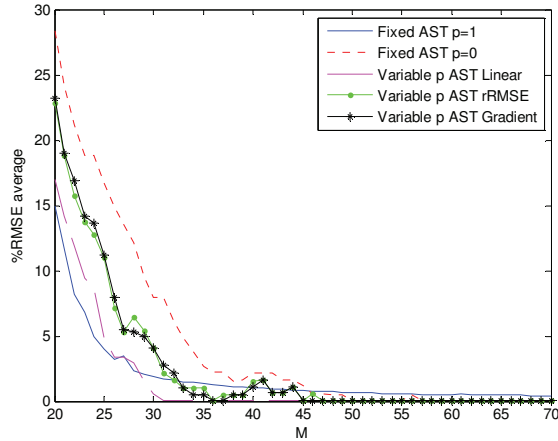


Figure 6. %RMSE average as function of M .

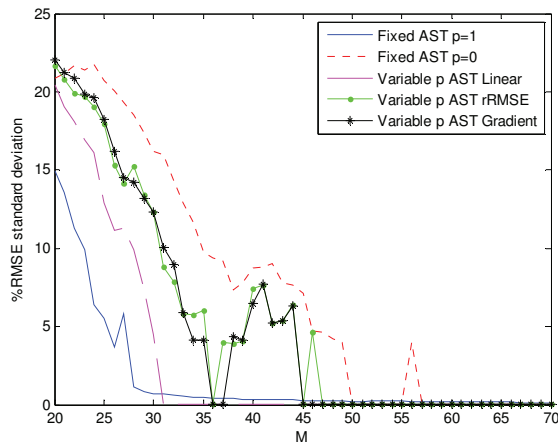


Figure 7. %RMSE standard deviation as function of M .

TABLE 1- Comparison of various versions of AST. The best choice is highlighted, and the worst value is underlined.

Version	Min. M		Iterations	
	Mean	Std	Mean	Std
$p=0$	23.8	6.93	13.2	3.83
LINEAR	19.7	5.1	16.3	3.46
rRMSE	20.9	5.5	14.8	3.77
Gradient	20.4	5.96	14.3	4.1
$p=1$	19.7	3.25	60.3	25.7

holds for $c \approx 2.22$. The fact that the c value found in our experiments is lower than the c value reported in [6], and therefore the number of linear measurements M is also lower, was expected. In [8] is demonstrated that fewer measurements are required for exact recovery minimizing $\ell_{(p<1)}$ than minimizing $\ell_{(p=1)}$ (as in [6] using Basis Pursuit).

In the approaches proposed here, $p=1$ is used initially but in all subsequent iterations p remains below 1. Thus, it is expected that the variable- p AST approaches need less linear

measurements than Basis Pursuit [6] which is what we have found experimentally.

5. ACKNOWLEDGEMENTS

This work was partially supported by the SEP-DGEST grant 833.08-P (México) and by the SEP-DGEST PROMEP scholarship 022005015P (México).

6. REFERENCES

- [1] R. Baraniuk, "Compressive Sensing", *IEEE Signal Processing Magazine*, vol. 24, No. 4, pp. 118, 119, 120, 124, July 2007.
- [2] S.D. Cabrera, J.G. Rosiles, and A.E. Brito, "Affine Scaling Transformation Algorithms for Harmonic Retrieval in a Compressive Sampling Framework", *Proc. of Wavelets XII*, SPIE vol. 6701, San Diego, CA, August 2007.
- [3] S.D. Cabrera and T.W. Parks, "Extrapolation and Spectral Estimation with Iterative Weighted Norm Modification", *IEEE Trans. on Signal Processing*, vol. 39-4, pp. 842-851, April 1991.
- [4] E.J. Candès, "Compressive Sampling", *Proc. of Int. Congress of Mathematicians*, pp. 1433-1452, Madrid, 2006.
- [5] E.J. Candès and J. Romberg, "Sparsity and Incoherence in Compressive Sampling", Preprint. October 2006.
- [6] E.J. Candès and T. Tao, "Near Optimal Signal Recovery From Random Projections: Universal Encoding Strategies?", *IEEE Trans. on Information Theory*, vol. 52, pp. 5406-5425, 2006.
- [7] E.J. Candès and M.B. Wakin, "An Introduction to Compressive Sensing", *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21-30, March 2008.
- [8] R. Chartrand, "Exact reconstructions of sparse signals via nonconvex minimization". *IEEE Signal Proc. Lett.*, 14(10), pp. 707-710, 2007
- [9] D.L. Donoho, "Compressed Sensing", *IEEE Trans. on Information Theory*, vol. 52, No.4, pp. 1289-1306, 2006.
- [10] I.F. Gorodnitsky and B.D. Rao, "Convergence analysis of a class of adaptive weighted norm extrapolation algorithms". *Proc. Twenty-seventh Asilomar Conf. Signals, Syst., Comput.*, Monterey, CA, vol. I, pp. 339-343, Nov. 1993.
- [11] B.D. Rao and K. Kreutz-Delgado, "An affine scaling methodology for best basis selection", *IEEE Trans. on Signal Processing*, vol. 47, No. 1, pp. 187-200, Jan. 1999.
- [12] J.A. Tropp, "Greed Is Good: Algorithmic Results for Sparse Approximation", *IEEE Transactions on information Theory*, vol. 50, no. 10, pp. 2231 – 2242, 2004.