Advances in Pure Mathematics, 2017, 7, 66-74
http://www.scirp.org/journal/apm
ISSN Online: 2160-0384
ISSN Print: 2160-0368

A Generalized Elastic Net Regularization with Smoothed ℓ_0 Penalty

Sisu Li, Wanzhou Ye

Department of Mathematics, College of Science, Shanghai University, Shanghai, China
Email: wzhy@shu.edu.cn

How to cite this paper: Li, S.S. and Ye, W.Z. (2017) A Generalized Elastic Net Regularization with Smoothed ℓ_0 Penalty. *Advances in Pure Mathematics*, 7, 66-74.
http://dx.doi.org/10.4236/apm.2017.71006

Received: December 6, 2016
Accepted: January 21, 2017
Published: January 24, 2017

Abstract

This paper presents an accurate and efficient algorithm for solving the generalized elastic net regularization problem with smoothed ℓ_0 penalty for recovering sparse vector. Finding the optimal solution to the unconstrained ℓ_0 minimization problem in the recovery of compressive sensed signals is an NP-hard problem. We proposed an iterative algorithm to solve this problem. We then prove that the algorithm is convergent based on algebraic methods. The numerical result shows the efficiency and the accuracy of the algorithm.

Keywords

Sparse Vector, Compressed Sense, Elastic Net Regularization, ℓ_0 Minimization

1. Introduction

Compressive sensing (CS) has been emerging as a very active research field and brought about great changes in the field of signal processing during recent years with broad applications such as compressed imaging, analog-to-information conversion, biosensors, and so on [1] [2] [3]. Meanwhile, the ℓ_0 norm based signal recovery is attractive in compressed sensing as it can facilitate exact recovery of sparse signal with very high probability [4] [5]. Mathematically, the problem can be presented as

$$\min_{x\in R^N} \|x\|_0 \text{, subject to } Ax = y, \tag{1}$$

where $y \in R^m, A \in R^{m\times N}$ is a measurement matrix, $\|\cdot\|_2$ denotes the Euclidean norm and $\|x\|_0$, formally called the quasi-norm, denotes the number of the nonzero components of $x = (x_1, x_2, \cdots, x_n)^T \in R^N$, and the $\lambda > 0$ is a regularization parameter.

We can then solve the unconstrained ℓ_0 regularization problem

$$\min_{x\in R^N}\left\{\frac{1}{2}\|Ax-y\|_2^2+\lambda\|x\|_0\right\}, \tag{2}$$

A natural approach to this problem is to solve a convex relaxation ℓ_1 regularization problem [6] [7] as following

$$\min_{x\in R^N}\left\{\frac{1}{2}\|Ax-y\|_2^2+\lambda\|x\|_1\right\}, \tag{3}$$

where the $\|x\|_1=\sum_{i=1}^N\|x_i\|$ is the ℓ_1 norm. Undoubtly, the ℓ_1 regularization has many applications [8] [9] and can be solved by many classic algorithms such as the iterative soft thresholding algorithm [7], the LARs [10], etc. An effective regression method, Lasso [11], has a very close relationship with the ℓ_1 regularization as well. In 2005, Zou *et al.* proposed the following algorithm, which called the elastic net regularization [12]

$$\min_{x\in R^N}\left\{\frac{1}{2}\|Ax-y\|_2^2+\lambda_1\|x\|_1+\lambda_2\|x\|_2\right\}, \tag{4}$$

where the $\lambda_1,\lambda_2>0$ are two regularization parameters. It is proved in many papers that the elastic net regularization outperforms the Lasso in prediction accuracy. Cands proved that as long as A satisfies the RIP condition with a constant parameter, the ℓ_1 minimization can yield an equivalent solution as that of ℓ_0 minimization [13]. So in general, the ℓ_1 regularization problem can be regard as an aproach to the ℓ_0 regularization. Therefore, we shall consider a generalized elastic net regularization problem with ℓ_0 penalty:

$$\min_{x\in R^N}\left\{\frac{1}{2}\|Ax-y\|_2^2+\lambda_1\|x\|_0+\lambda_2\|x\|_2^2\right\}, \tag{5}$$

Unfortunately, the ℓ_0 norm minimization problem is NP-hard [14]. And due to the sparsity of the solution x, we could turn out to calculate the following generalized elastic net regularization with smoothed ℓ_0 penalty:

$$\min_{x\in R^N}\left\{\frac{1}{2}\|Ax-y\|_2^2+\lambda_1\|x\|_{0,\delta}+\lambda_2\|x\|_2^2\right\}, \tag{6}$$

where $\|x\|_{0,\delta}=\sum_{i=1}^N\frac{x_i^2}{x_i^2+\delta}$, the $\delta>0$ is a parameter which approaches zero in order to approximate $\|x\|_0$.

In this paper,we propose an iterative algorithm for recovering sparse vectors which substitute the ℓ_0 penalty with a function [15]. And by adding an ℓ_2 term, we can prove that the algorithm is convergent based on the algebraic methods. In the experiment part, we compare the algorithm with the ℓ_1 soft thresholding algorithm (ITH) [16]. And the output results show an outstanding success of the new method.

The rest of this paper is organized as follows. We develop the new algorithm in Section 2 and prove its convergence in Section 3. Experiments on accuracy and efficiency are reported in Section 4. Finally, we conclude this paper in

Section 5.

2. Problem Reformulation

The reconstruction method discussed in this paper is for directly approaching the l_0 norm and obtaining its minimal solution with suitably designed objective functions. We denote by $C_\delta\left(x,\lambda_1,\lambda_2\right)$ the objective function of the minimization problem (6).

$$C_\delta\left(x,\lambda_1,\lambda_2\right)=\frac{1}{2}\|Ax-y\|_2^2+\lambda_1\sum_{i=1}^{N}\frac{x_i^2}{x_i^2+\delta}+\lambda_2\|x\|_2^2. \tag{7}$$

Our goal is to minimize the objective function. For any $\delta>0$ and $\lambda_1,\lambda_2>0$, the minimization problem is convex coercie, thus it has a solution. So the optimal solution of (7) can be given according to the optimal condition.

$$A^{\mathrm{T}}\left(A\hat{x}-y\right)+\left[\frac{2\lambda_1\hat{x}_i\delta}{\left(\hat{x}_i^2+\delta\right)^2}\right]_{1\le i\le N}+2\lambda_2\hat{x}=0. \tag{8}$$

Then we can present the following iterative algorithm to solve the above minimization problem.

Algorithm 1 Iterative Algorithm for Generalized Elastic Net Regularization with Smoothed ℓ_0 Penalty (IAGENR-L0)

0: Given vector y, matrix A, choose parameters $\delta>0,\lambda_1,\lambda_2>0$ and initialize $x_0\in R^N$

1: **for** $k=1$ **do**

2: Compute the following system for $x^{(k+1)}$

$$\left[\frac{2\lambda_1\delta x_j^{k+1}}{\left(\delta+\left(x_j^k\right)^2\right)^2}\right]_{1\le j\le N}+A^{\mathrm{T}}\left(Ax^{k+1}-y\right)+2\lambda_2x^{k+1}=0 \tag{9}$$

3: Or compute the equivalent equation

$$\left(A^{\mathrm{T}}A+\mathrm{Diag}\left[\frac{2\lambda_1\delta}{\left(\delta+\left(x_i^k\right)^2\right)^2},i=1,2,\cdots,N\right]\right)x^{k+1}=A^{\mathrm{T}}y \tag{10}$$

4: Stop when $\left\|x^{k+1}-x^k\right\|_2<\delta$

5: **end for**

6: Output the vector $x_0\in R^N$

3. Convergence of the Algorithm

In this section, we prove that the algorithm is convergent. Firstly, we start from the lemma 1 [17] which we can deduce the inquality directly by using the mean value theorem.

Lemma 1. *Given* $\delta>0$ *, then the inequality*

$$\frac{x^2}{x^2+\delta}-\frac{y^2}{y^2+\delta}-\frac{2\delta\left(x-y\right)y}{\left(x^2+\delta\right)^2}\ge\frac{\delta\left(x-y\right)^2}{\left(x^2+\delta\right)^2} \tag{11}$$

holds for any $x,y\in R$.

Proof. We first denote $f\left(x^2\right)=\frac{x^2}{x^2+\delta}$, then by the mean value theorem, we have

$$f\left(x^2\right)-f\left(y^2\right)=f'(\xi)\left(x^2-y^2\right) \text{ where } \xi \text{ between } x^2 \text{ and } y^2. \tag{12}$$

So we have

$$\frac{x^2}{x^2+\delta}-\frac{y^2}{y^2+\delta}=\frac{\delta\left(x^2-y^2\right)}{(\xi+\delta)^2}=\frac{2\delta(x-y)y+\delta(x-y)^2}{\left(x^2+\delta\right)^2}. \tag{13}$$

Thus, we can simplify the inequality as follow:

$$\frac{x^2}{x^2+\delta}-\frac{y^2}{y^2+\delta}-\frac{2\delta\left(x^2-y^2\right)}{\left(x^2+\delta\right)^2}\geq\frac{\delta(x-y)^2}{\left(x^2+\delta\right)^2}.$$

This inequality of (11) holds no matter $x^2>y^2$, $x^2<y^2$ or $x^2=y^2$. And the next Lemma proves that the sequence $x^{(k)}$ drives the function $C_\delta\left(x,\lambda_1,\lambda_2\right)$ downhill. □

Lemma 2. *For any* $\delta>0$ *and* $\lambda_1,\lambda_2>0$, *let* $x^{(k+1)}$ *be the solution of* (9) *for* $k=1,2,3,\cdots$ *Then we can have*

$$\left\|Ax^k-Ax^{k+1}\right\|_2^2\leq 2\left(C_\delta\left(x^k,\lambda_1,\lambda_2\right)-C_\delta\left(x^{k+1},\lambda_1,\lambda_2\right)\right). \tag{14}$$

Furthermore,

$$\left\|x^k-x^{k+1}\right\|_2^2\leq c\left(C_\delta\left(x^k,\lambda_1,\lambda_2\right)-C_\delta\left(x^{k+1},\lambda_1,\lambda_2\right)\right). \tag{15}$$

where c is a positive constant that depends on λ_2.

Proof.

$$\begin{aligned}
C_\delta\left(x^k,\lambda_1,\lambda_2\right)-C_\delta\left(x^{k+1},\lambda_1,\lambda_2\right)&=\lambda_1\sum_{i=1}^N\left(\frac{\left(x_i^k\right)^2}{\left(x_i^k\right)^2+\delta}-\frac{\left(x_i^{k+1}\right)^2}{\left(x_i^{k+1}\right)^2+\delta}\right)\\
&\quad+\lambda_2\left(\left\|x^k\right\|_2^2-\left\|x^{k+1}\right\|_2^2\right)+\frac{1}{2}\left(\left\|Ax^k-y\right\|_2^2-\left\|Ax^{k+1}-y\right\|_2^2\right)\\
&=\lambda_1\sum_{i=1}^N\left(\frac{\left(x_i^k\right)^2}{\left(x_i^k\right)^2+\delta}-\frac{\left(x_i^{k+1}\right)^2}{\left(x_i^{k+1}\right)^2+\delta}\right)\\
&\quad+\frac{1}{2}\left\|Ax^k-Ax^{k+1}\right\|_2^2+\lambda_2\left\|x^k-x^{k+1}\right\|_2^2\\
&\quad+2\lambda_2\left(x^k-x^{k+1}\right)^{\mathrm{T}}x^{k+1}+\left(Ax^k-Ax^{k+1}\right)^{\mathrm{T}}\left(Ax^{k+1}-y\right).
\end{aligned} \tag{16}$$

Using (9). The last term in (16) can be simplified to be

$$\begin{aligned}
\left(Ax^k-Ax^{k+1}\right)^{\mathrm{T}}\left(Ax^{k+1}-y\right)&=\left(Ax^k-Ax^{k+1}\right)^{\mathrm{T}}\left[-\left(A^T\right)^{-1}\left(2\lambda_2x^{k+1}+\frac{2\lambda_1\delta x^{k+1}}{\left(\left(x^k\right)^2+\delta\right)^2}\right)\right]\\
&=-\sum_{i=1}^N\frac{2\lambda_1\delta x_i^{k+1}+\left(x_i^k-x_i^{k+1}\right)}{\left(\left(x^k\right)^2+\delta\right)^2}-2\lambda_2\left(x^k-x^{k+1}\right)^Tx^{k+1}.
\end{aligned} \tag{17}$$

Substituting (15) into (16) and using (11),

$$\begin{aligned}
&C_\delta\left(x^k,\lambda_1,\lambda_2\right)-C_\delta\left(x^{k+1},\lambda_1,\lambda_2\right)\\
&=\lambda_1\sum_{i=1}^{N}\left(\frac{\left(x_i^k\right)^2}{\left(x_i^k\right)^2+\delta}-\frac{\left(x_i^{k+1}\right)^2}{\left(x_i^{k+1}\right)^2+\delta}-\frac{2\delta x_i^{k+1}\left(x_i^k-x_i^{k+1}\right)}{\left(\left(x^k\right)^2+\delta\right)^2}\right)\\
&\quad+\frac{1}{2}\left\|Ax^k-Ax^{k+1}\right\|_2^2+\lambda_2\left\|x^k-x^{k+1}\right\|_2^2\\
&\geq\sum_{i=1}^{N}\frac{\delta\lambda_1\left(x^k-x_i^{k+1}\right)^2}{\left(\left(x^k\right)^2+\delta\right)^2}+\frac{1}{2}\left\|Ax^k-Ax^{k+1}\right\|_2^2+\lambda_2\left\|x^k-x^{k+1}\right\|_2^2.
\end{aligned}\tag{18}$$

□

Since $\sum_{i=1}^{N}\frac{\delta\lambda_1\left(x^k-x_i^{k+1}\right)^2}{\left(\left(x^k\right)^2+\delta\right)^2}\geq 0$ for any x^k and x^{k+1}. From (18) we can obtain the results of (14) and (15) with $C=\frac{1}{\lambda_2}$.

Lemma 3. ([18], Theorem 3.1) Let $P\left(z,\overline{w}\right)=0$ to be given, and let $Q\left(z,\left(\overline{a}\right),\left(\overline{c}\right)\right)=0$ be its corresponding highest ordered system of equations. If $Q\left(z,\left(\overline{a}\right),\left(\overline{c}\right)\right)=0$ has only the trivial solution $z=0$, then $P\left(z,\overline{w}\right)=0$ has $\beta=\prod_{i=1}^{m}q_i$ solutions, where q_i is the degree of P_i.

Theorem 1. *For any* $\delta>0$ *and* $\lambda_1,\lambda_2>0$. *Then the iterative solutions* x^k *in* (9) *converge to* x^*, *that is* $\lim_{k\to\infty}x^k=x^*$ *and* x^* *is a critical point of* (6).

Proof. Here, we need to prove that the sequence x^k is bounded. We assume that x^{k_i} is one convergent subsequence of x^k and its limit point is x^*. By (15) we know that the sequence x^{k_i+1} also converges to x^*. If we replace x^k, x^{k+1} with x^{k_i}, x^{k_i+1} in (10) and letting $i\to\infty$ yields.

□

$$\frac{2\lambda_1\delta x_j^*}{\left(\left(x_j^*\right)^2+\delta\right)^2}+A^{\mathrm{T}}\left(Ax^*-y\right)+2\lambda_2x^*=0.\tag{19}$$

And this implies that the limit point which converges to any convergent subsequence of x^k is the critical point of (8). In order to prove the convergence of sequence x^k, we need to prove that the limit point set M, which contains all the limit points of convergent subsequence of x^k is a finite set. So we have to prove that the following equation has finite solutions.

$$\left[\frac{2\lambda_1\delta u_j}{\left(\left(u_i\right)^2+\delta\right)^2}\right]_{1\leq i\leq N}+A^{\mathrm{T}}\left(Au-y\right)+2\lambda_2u=0.\tag{20}$$

where $u=\left(u_1,u_2,\cdots,u_N\right)^{\mathrm{T}}\in R^N$. We can rewrite (20) as follow:

$$\left[\frac{2\lambda_1\delta u_j}{\left(\left(u_i\right)^2+\delta\right)^2}\right]_{1\leq i\leq N}+\left(A^{\mathrm{T}}A+2\lambda_2I_N\right)u-A^{\mathrm{T}}y=0.\tag{21}$$

It is obvious that $A^{\mathrm{T}}A+2\lambda_2 I_N$ is a positive definite matrix, $A^{\mathrm{T}}y\in R^N$ is the $N\times N$ identity matrix. Then the (21) can be rewritten as the following equation:

$$2\lambda_1\delta u+B\left(\left(A^{\mathrm{T}}A+2\lambda_2 I_N\right)u-A^{\mathrm{T}}y\right)=0. \tag{22}$$

where B is an $N\times N$ diagonal matrix with diagonal entries $B_{ii}=\left(\left(u_i\right)^2+\delta\right)^2, i=1,2,3,\cdots,N$. We denote $A^{\mathrm{T}}A+2\lambda_2 I_N=\left(a_{ii}\right)_{N\times N}$ and $A^{\mathrm{T}}y=\left(q_1,q_2,\cdots,q_N\right)^{\mathrm{T}}$. Then

$$\begin{cases} 2\lambda_1\delta u_1+\left(a_{11}u_1+a_{12}u_2+\cdots+a_{1N}u_N-q_1\right)\left(u_1^2+\delta\right)^2=0, \\ 2\lambda_1\delta u_2+\left(a_{21}u_1+a_{22}u_2+\cdots+a_{2N}u_N-q_2\right)\left(u_2^2+\delta\right)^2=0, \\ \quad\cdots\cdots\cdots\cdots\cdots\cdots \\ 2\lambda_1\delta u_N+\left(a_{N1}u_1+a_{N2}u_2+\cdots+a_{NN}u_N-q_N\right)\left(u_N^2+\delta\right)^2=0. \end{cases} \tag{23}$$

If we want to prove that (23) has finite solutions, then we can prove the (22) system has finite solutions. According to lemma 3, if we prove that the highest ordered system of (23) has only trivial solution, then it's easy to conclude that the Equation (23) has finite solutions.

$$\begin{cases} \left(a_{11}u_1+a_{12}u_2+\cdots+a_{1N}u_N\right)u_1^4=0, \\ \left(a_{21}u_1+a_{22}u_2+\cdots+a_{2N}u_N\right)u_2^4=0, \\ \quad\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \left(a_{N1}u_1+a_{N2}u_2+\cdots+a_{NN}u_N\right)u_N^4=0. \end{cases} \tag{24}$$

We prove the system (24) has only trivial solution. We assume that $u=\left(u_1,u_2,\cdots,u_s,0,\cdots,0\right)^{\mathrm{T}}\in R^N$ is a nonzero solution of (24), $u_i\neq 0$ for $i=1,2,\cdots,s$, $1\leq s\leq N$. Then we have

$$Cu^s=0. \tag{25}$$

where $C=\left(a_{ii}\right)_{s\times s}$ is the $s\times s$ leading principle submatrix of matrix $A^{\mathrm{T}}A+2\lambda_2 I_N$ is the positive definite, therefore the matrix C is positive definite as well. So we have $u_i=0$ for $i=1,2,\cdots,s$. This contradicts the assumption of $u_i\neq 0$, $i=1,2,\cdots,s$, $1\leq i\leq s$.

Therefore, the system (24) has only trivial solution. So the Equation (20) has finite solutions. Since all the limit points of convergent subsequence of $x^{(k)}$ satisfies the Equation (20) and we have proved that (20) has finite solutions. So the limit point set M is a finite set. Combining with $\left\|x^{(k+1)}-x^{(k)}\right\|_2\to 0$ as $k\to\infty$, we thus obtain that the sequence $x^{(k)}$ is convergent and limit x^* is a critical point of problem (7).

4. Numerical Experiments

In this section, we present some numerical experiments to show the efficiency and the accuracy of the Algorithm 1 for sparse vector recovery. We compare the performance of Algorithm 1 with ℓ_1 IST [3]. In the test, the matrix A had the size of 100×250, which is $m=100$ and $N=250$. All the experiments were performed in Matlab and all the experimental results were averaged over 100 independent trials for various sparsity s.

The experiment results contain two parts: the first one focuses on the comparison of the two algorithms in accuracy; the other one focuses on the efficiency of the two algorithms. In the experiments, the mean squared error of the original vector and the result is recorded as

$$\text{MSE} = \left\| x^k - x^0 \right\|_2^2 / N \tag{26}$$

4.1. Comparison on the Accuracy

The matrix $A \in R^{100\times 250}$ and the original sparse vector $x^0 \in R^{250}$ was generated randomly according to the standard Gaussian distribution with N-length and s-sparse, which varies as 2, 4, 6, 8, ⋯, 48. The location of the nonzero elements were randomly generated. The regularization parameters were set as $\delta = 10^{-6}$ and $\lambda_1 = 10^{-3}, \lambda_2 = 10^{-5}$. All the other parameters of the two algorithms were set to be the same.The results are shown in Figure 1.

The Figure 1 shows that the convergence error MSE for the two algorithms tends to be stable at last for different sparsity *s*. We can also observe that the MSE of the LAGENR-L0 is lower than the IST which demonstrates that our algorithm is not only convergent,but also outperforms the IST in accuracy.

4.2. Comparison on the Efficiency

In this subsection, we focus on the speed of the two algorithms. We conduct various experiments to test the effectiveness of the proposed algorithm. Table 1 report the numerical results of the two algorithms for recovering vectors for different sparsity level. From the results, we can see that the IAGENR-L0 performs much better than IST in efficiency and the accuracy.

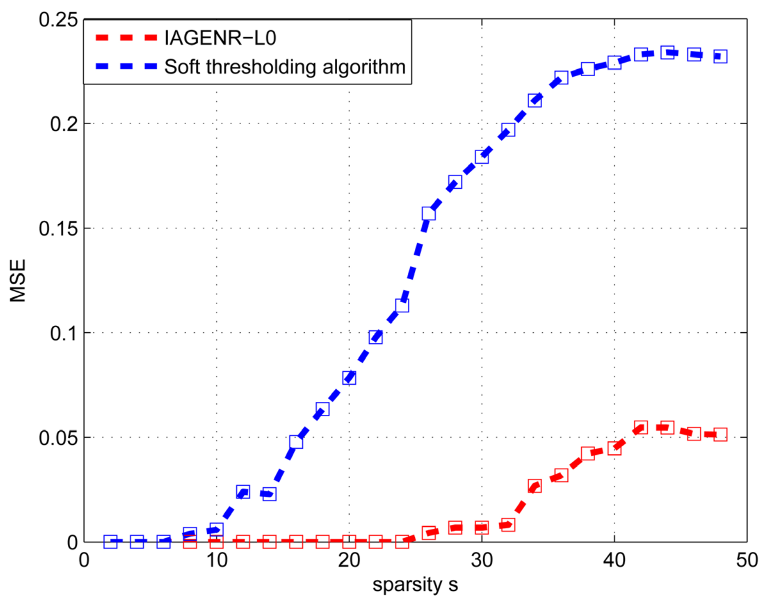

Figure 1. Comparison of the convergence error $\text{MSE} = \left\| x^k - x^0 \right\|_2^2 / N$ for both IAGENR-L0 and IST.

Table 1. The iteration time of the IAGENR-L0 and the IST for different sparsity level.

Sparsity	Algorithm	Time	MSE
2	IAGENR-L0	0.038234 s	7.35e−007
	IST	0.721920 s	1.68e−006
4	IAGENR-L0	0.088917 s	1.03e−006
	IST	0.943177 s	2.21e−006
8	IAGENR-L0	0.031317 s	1.83e−007
	IST	2.146079 s	3.95e−003
16	IAGENR-L0	0.046566 s	1.35e−009
	IST	2.368845 s	9.78e−002
32	IAGENR-L0	0.501139 s	8.24e−003
	IST	1.608879 s	1.14e−001

5. Conclusion

In this paper, we consider an iterative algorithm for solving the generalized elastic net regularization problems with smoothed ℓ_0 penalty for recovering sparse vectors. Then a detailed proof of convergence of the iterative algorithm is given in Section 2 by using the algebraic method. Additionally, the numerical experiments in Section 3 show that our iterative algorithm is convergent and performs better than the IST on recovering sparse vectors.

References

[1] Donoho, D.L. (2006) Compressed Sensing. *IEEE Transactions on Information Theory*, **52**, 1289-1306.

[2] Cands, E.J., Romberg, J. and Tao, T. (2006) Robust Uncertainty Principles: Exagct signal Reconstruction from Highly Incomplete Frequency Information. *IEEE Transactions on Information Theory*, **26**, 489-509.

[3] Duarte, M.F. and Eldar, Y.C. (2011) Structured Compressed Sensing: From Theory to Applications. *IEEE Transactions on Signal Processing*, **59**, 4053-4085.

[4] Lu, Z. (2014) Iterative Hard Thresholding Methods for ℓ_0 Regularized Convex cone Programming. *Mathematical Programming*, **147**, 125-154. https://doi.org/10.1007/s10107-013-0714-4

[5] Cand‘es, E.J. and Tao, T. (2005) Decoding by Linear Programming. *IEEE Transactions on Information Theory*, **51**, 4203-4215.

[6] Chen, S.S., Donodo, D.L. and Saunders, M.A. (1998) Atomic Decomposition by basis Pursuit. *SIAM Journal on Scientific Computing*, **20**, 33-61. https://doi.org/10.1137/S1064827596304010

[7] Daubechies, I., Defries, M. and DeMol, C. (2004) An Iterative Thresholding Algorithm for Linear Inverse Problems with a Sparsity Constraint. *Communications on Pure and Applied Mathematics*, **57**, 1413-1457.

[8] Zou, H. (2006) The Adaptive Lasso and Its Oracle Properties. *Journal of the American Statistical Association*, **101**, 1418-1429. https://doi.org/10.1198/016214506000000735

[9] Meinshausen, N. and Yu, B. (2009) Lasso-Type Recovery of Sparse Representations for High-Dimensional Data. *Annals of Statistics*, **46**, 246-270.

https://doi.org/10.1214/07-AOS582

[10] Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R. (2004) Least Angle Regression. *The Annals of Statistics*, **32**, 407-451.

[11] Tibshirani, R. (1996) Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society Series B* (*Statistical Methodology*), **73**, 273-282.

[12] Zou, H. and Hastie, T. (2005) Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society Series B* (*Statistical Methodology*), **67**, 301-320.

[13] Kordas, G. (2015) A Neurodynamic Optimization Method for Recovery of Compressive Sensed Signals with Globally Converged Solution Approximating to L_0 Minimization. *IEEE Transactions on Neural Networks and Learning Systems*, **26**, 1363-1374.

[14] Natraajan, B.K. (1995) Sparse Approximation to Linear Systems. *SIAM Journal on Computing*, **24**, 227-234. https://doi.org/10.1137/S0097539792240406

[15] Xiao, Y.H. and Song, H.N. (2012) An Inexact Alternating Directions Algorithm for Constrained Total Variation Regularized Compressive Sensing Problems. *Journal of Mathematical Imaging and Vision*, **44**, 114-127.

[16] Daubechies, I., Defrise, M. and Mol, C.D. (2004) An Iterative Thresholding Algorithm for Linear Inverse Problems with a Sparsity Constraint.

[17] Lai, M.J., Xu, Y.Y. and Yin, W.T. (2013) Improved Iteratively Reweighted Least Squares for Unconstrained Smoothed L_q Minimization. *SIAM Journal on Numerical Analysis*, **51**, 927-957.

[18] Garcia, C.B. and Li, T.Y. (1980) On the Number of Solutions to Polynomial Systems of Equations. *SIAM Journal on Numerical Analysis*, **17**, 540-546. https://doi.org/10.1137/0717046