# **ITERATIVE LOG THRESHOLDING**

Dmitry Malioutov, Aleksandr Aravkin

T.J. Watson IBM Research center

# ABSTRACT

Sparse reconstruction approaches using the re-weighted  $\ell_1$ penalty have been shown, both empirically and theoretically, to provide a significant improvement in recovering sparse signals in comparison to the  $\ell_1$ -relaxation. However, numerical optimization of such penalties involves solving problems with  $\ell_1$ -norms in the objective many times. Using the direct link of reweighted  $\ell_1$ -penalties to the concave log-regularizer for sparsity, we derive a simple proximal-like algorithm for the log-regularized formulation. The proximal splitting step of the algorithm has a closed form solution, and we call the algorithm log-thresholding in analogy to soft thresholding for the  $\ell_1$ -penalty. We establish convergence results, and demonstrate that log-thresholding provides more accurate sparse reconstructions compared to both soft and hard thresholding. Furthermore, the approach can be directly extended to optimization over matrices with penalty for rank (i.e. the nuclear norm penalty and its re-weighted version), where we suggest a singular-value log-thresholding approach.

*Index Terms*— sparsity, reweighted  $\ell_1$ , non-convex formulations, proximal methods

## 1. INTRODUCTION

We consider sparse reconstruction problems which attempt to find sparse solutions to under-determined systems of equations. A basic example of such a problem is to recover a sparse vector  $\mathbf{x} \in \mathbb{R}^N$  from measurements  $\mathbf{y} = A\mathbf{x} + \mathbf{n}$ , where  $\mathbf{y} \in \mathbb{R}^M$  with M < N, and  $\mathbf{n}$  captures corruption by noise. Attempting to find the sparsest solutions is known to be NP-hard, so convex relaxations involving  $\ell_1$ -norms have gained unprecedented popularity. Basis pursuit (or LASSO in statistics literature) minimizes the following objective [1]:

$$\min \|\mathbf{y} - A\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1 \tag{1}$$

Here  $\lambda$  is a parameter that balances sparsity versus the norm of the residual error. There is truly a myriad of algorithms to solve (1), e.g. [2, 3, 4, 5]. For large-scale instances, variations of iterative soft thresholding have become very popular [6]:

$$\mathbf{x}^{(n+1)} = \mathcal{S}_{\lambda} \left( \mathbf{x}^{(n)} + A^T (\mathbf{y} - \mathbf{x}^{(n)}) \right)$$
(2)

where  $S_{\lambda}(\mathbf{z})$  applies soft-thresholding for each entry:

$$S_{\lambda}(z_i) = \operatorname{sign}(z_i) \max(0, |z_i| - \lambda).$$
(3)

Based on operator splitting and proximal projection theories, the algorithm in (2) converges if the spectral norm ||A|| < 1 [6, 7]. This can be achieved simply by rescaling *A*. Accelerated versions of iterative thresholding have appeared [8].

An exciting albeit simple improvement over  $\ell_1$ -norms for approximating sparsity involves weighting the  $\ell_1$ -norm:  $\sum_i w_i |x_i|$  with  $w_i > 0$ . Ideal weights require knowledge of the sparse solution, but a practical idea is to use weights based on solutions of previous iterations [9, 10]:

$$w_i^{(n+1)} = \frac{1}{\delta + |\hat{x}_i^{(n)}|} \tag{4}$$

This approach can be motivated as a local linearization of the log-heuristic for sparsity [9]. There is strong empirical [10] and recent theoretical evidence that reweighted  $\ell_1$  approaches improve recovery of sparse signals, in the sense of enabling recovery from fewer measurements [11, 12].

In this paper, we propose a simple proximal algorithm for sparse recovery with the log-penalty and derive a closedform solution for the proximal step, which we call logthresholding. We establish monotone convergence of iterative log-thresholding (ILT) to its fixed points, and conditions relating these fixed points to local minima of the log-penalized objective. Sparse recovery performance of the method on numerical examples surpasses both soft and hard iterative thresholding (IST and IHT). We also extend the approach to minimizing rank for matrix functions via singular value logthresholding. To put this into context of related work, [13] has considered iterative thresholding based on non-convex  $\ell_p$ -norm penalties for sparsity. However, these penalties do not have a connection to re-weighted  $\ell_1$  optimization. Also, [14, 15, 16] have investigated alternative iterative optimization approaches for non-convex penalties including the log penalty, but their solutions do not use closed form log-thresholding.

## 2. ISTA AS PROXIMAL SPLITTING

We briefly review how soft-thresholding can be used to solve the sparse reconstruction problem in (1). Functions of the form f(x) = h(x) + g(x) where h(x) is convex differentiable with a Lipschitz gradient, and g(x) is general convex can be solved by a general proximal splitting method [7]:

$$\hat{\mathbf{x}}^{(n+1)} = \operatorname{prox}_{g} \left( \mathbf{x}^{(n)} - \nabla h(\mathbf{x}^{(n)}) \right).$$
 (5)

The prox-operation is a generalization of projection onto a set to general convex functions:

$$\operatorname{prox}_{h}(\mathbf{x}) = \arg\min_{\mathbf{z}} h(\mathbf{z}) + \frac{1}{2} \|\mathbf{x} - \mathbf{z}\|_{2}^{2}.$$
 (6)

If  $h(\mathbf{x})$  is an indicator function for a convex set, then the proxoperation is equivalent to the projection onto the set, and ISTA itself is equivalent to the projected gradient approach.

Forward-backward splitting can be applied to the sparse recovery problem (1) by deriving the proximal operator for  $\ell_1$ -norm, which is precisely the soft-thresholding operator in (3). The convergence of ISTA in (2) thus follows directly from the theory derived for forward-backward splitting [7].

#### 3. LOG-THRESHOLDING

The reweighted- $\ell_1$  approach can be justified as an iterative upper bounding by a linear approximation to the concave log-heuristic for sparsity (here  $\delta$  is a small positive constant) [9]:

$$\min f(\mathbf{x}) = \min \|\mathbf{y} - A\mathbf{x}\|_2^2 + \lambda \sum_i \log(\delta + |x_i|).$$
(7)

While the log-penalty is not convex (it is in fact concave), we still consider the scalar proximal objective around a fixed x:

$$g_{\lambda}(z) \triangleq (z-x)^2 + \lambda \log(\delta + |z|).$$
 (8)

We note that for  $\delta$  small enough, the global minimum of  $g_{\lambda}(z)$ over z (with x held constant) is always at 0. However, when  $|x| > x_0 \triangleq \sqrt{2\lambda} - \delta$ , the function also exhibits a local minimum. For small x the local minimum disappears. We will argue that it is the local, rather than the global minimum, that provides the link to re-weighted  $\ell_1$  minimization. Now, using first order necessary conditions for optimality we can define the "log-thresholding" operator. For  $|x| > x_0$ , we solve the equation  $\nabla g_{\lambda}(z) = 0$  to find the local minimum in closedform. We call this operation *log-thresholding*,  $\mathcal{L}_{\lambda}(x)$ :

$$\mathcal{L}_{\lambda}(x) = \begin{cases} \frac{1}{2} \left( (x_i - \delta) + \sqrt{(x_i + \delta)^2 - 2\lambda} \right), \ x > x_0 \\ \frac{1}{2} \left( (x_i + \delta) - \sqrt{(x_i - \delta)^2 - 2\lambda} \right), \ x < -x_0 \\ 0, \ \text{otherwise} \end{cases}$$
(9)

where  $x_0 = \sqrt{2\lambda} - \delta$ . We illustrate log-thresholding in Figure 1. The left plot shows  $g_{\lambda}(z)$  as a function of z for several values of x. For large x the function has a local minimum, but for small x the local minimum disappears. For log-thresholding we are specifically interested in the the local minimum: an iterative re-weighted  $\ell_1$  approach with small enough step size starting at x, i.e. beyond the local minimum, will converge to



**Fig. 1**. Illustration of log-thresholding: (a)  $g_{\lambda}(z)$  (b)  $\mathcal{L}_{\lambda}(x)$ .

the local minimum, avoiding the global one. The right plot in Figure 1 shows the log-thresholding operation  $\mathcal{L}_{\lambda}(x)$  with  $x_0 = 1$  as a function of x. It can be seen as a smooth alternative falling between hard and soft thresholding.

In analogy to ISTA, we can now formally define the iterative log-thresholding algorithm:

$$\hat{\mathbf{x}}^{n+1} = \mathcal{L}_{\lambda} \left( \mathbf{x}^n + A^T (\mathbf{y} - A \mathbf{x}^n) \right)$$
(10)

where  $\mathcal{L}_{\lambda}(\mathbf{z})$  applies the element-wise log-thresholding operation we obtained in (9). We establish its convergence next.

## 3.1. Convergence of iterative log-thresholding

The theory of forward-backward splitting does not apply to analyze log-thresholding, as log-penalty is non-convex, and log-thresholding is not not firmly non-expansive. However, we will instead use an approach based on optimization transfer using surrogate functions [17]. At a high-level our analysis for ILT follows that for IHT in [18], but some of the steps are notably different. In Section 6 we establish:

**Proposition 1** *The ILT algorithm in (10) monotonically decreases the objective*  $f(\mathbf{x})$  *in (7), and converges to fixed points if*  $||A||_2 < 1$ .

We also relate ILT fixed points to local minima of the logpenalized objective (7) under additional technical conditions.

### 4. SINGULAR VALUE LOG-THRESHOLDING

A closely related problem to finding sparse solutions to systems of linear equations is finding low-rank matrices from sparse observations, known as matrix completion:

$$\min_{\mathbf{v}} \operatorname{rank}(X) \text{ such that } X_{i,j} = Y_{i,j}, \{(i,j) \in \Omega\}$$
(11)

Similar to sparsity, rank is a combinatorial objective which is typically intractable to optimize directly. However, the nuclear norm  $||X||_* \triangleq \sum_i \sigma_i(X)$ , where  $\sigma_i(X)$  are the singular values of X, serves as the tightest convex relaxation of rank,



Fig. 2. Noiseless sparse recovery: (a) average error-norm (b) probability of exact recovery after 250 iterations over 1000 random trials. M = 100, N = 200.

analogous to  $\ell_1$ -norm being the convex relaxation of the  $\ell_0$ norm. In fact, the nuclear norm is exactly the  $\ell_1$ -norm of the singular value spectrum of a matrix. This connection enables the application of various singular value thresholding algorithms: for instance, the SVT algorithm of [19] alternates soft-thresholding of the singular value spectrum with gradient descent steps. In the experimental section we investigate a simplified singular-value log-thresholding algorithm for matrix completion, where we replace soft thresholding with hard and log-thresholdings. We present very promising empirical results of singular value log-thresholding in Section 5, and a full convergence analysis will appear in a later publication.

## 5. EXPERIMENTS

We investigate the performance of iterative log thresholding via numerical experiments on noiseless and noisy sparse recovery. Intuitively we expect ILT to recover sparser solution than soft-thresholding (IST) due to the connection to reweighted- $\ell_1$  norms, and also to behave better than the non-smooth iterative hard thresholding (IHT).

First we consider sparse recovery without noise, i.e. we would like to find the sparsest solution that satisfies  $\mathbf{y} = A\mathbf{x}$  exactly. One could in principle solve a sequence of problems (1) with decreasing  $\lambda$ , i.e. increasing penalty on  $\|\mathbf{y} - A\mathbf{x}\|_2^2$  via IST, IHT, ILT. However, when we know an upper bound K on the desired number of non-zero coefficients, a more successful approach is to adaptively change  $\lambda$  to eliminate all except the top-K coefficients in each iteration<sup>1</sup> as used e.g. in [20]. We compare the performance of IST, IHT, and the proposed ILT in Figure 2. We use i.i.d. random normal A with N = 200, M = 100 and we vary K. Apart from changing the thresholding operator, all the algorithms are exactly the same. The top plot shows the average reconstruction error from the true sparse solution  $\|\hat{\mathbf{x}} - \mathbf{x}^*\|_2$ . It is averaged over



**Fig. 3**. Sparse recovery with noise. Average error vs. sparsity over 100 trials, after 250 iterations.

1000 trials allowing IST, IHT and ILT to run for up-to 250 iterations. The bottom plot shows probability of recovering the true sparse solution. We can see that ILT is superior in both probability of recovery (higher probability of recovery) and in reconstruction error (lower reconstruction error) over both IST and IHT.

Our next experiment compares the three iterative thresholding algorithms on noisy data. Since regularization parameters have a different meaning for the different penalties, we plot the whole solution path of squared residual error vs. sparsity for the three algorithms in Figure 3. We compute the average residual norm for a given level of sparsity for all three algorithms, averaged over 100 runs. We have M = 100, N =200, K = 10 and a small amount of noise is added. We can see that the iterative log thresholding consistently achieves the smallest error for each level of sparsity.

In our final experiment we consider singular value logthresholding for matrix completion. We study a simplified algorithm that parallels the noiseless sparse recovery algorithm with known number of nonzero-elements K. We alternate gradient steps with steps of eliminating all but the first K singular values by soft, hard and log-thresholding. We have an  $N \times N$  matrix with 30% observed entries, N = 100 and rank, K = 2. We show the average error in Frobenius norm from the true underlying solution as a function of iteration number over 100 random runs in Figure 4. We see that the convergence of log-SV-thresholding to the correct solution is consistently faster. We expect similar improvements to hold for other algorithms involving soft-thresholding, and to other problems beyond matrix completion, e.g. robust PCA.

#### 6. CONVERGENCE OF ILT

Here we establish Proposition 1. We first define a surrogate function for  $f(\mathbf{x})$  in (7):

$$Q(\mathbf{x}, \mathbf{z}) = \|\mathbf{y} - A\mathbf{x}\|_{2}^{2} + \lambda \sum_{i} \log(\delta + |x_{i}|) + \|\mathbf{x} - \mathbf{z}\|_{2}^{2} - \|A(\mathbf{x} - \mathbf{z})\|_{2}^{2}$$
(12)

<sup>&</sup>lt;sup>1</sup>This is easy for IST and IHT by sorting  $|\mathbf{x}|$  in descending order: let  $\mathbf{s} = \text{sort}|\mathbf{x}|$  then  $\lambda = s_{K+1}$ . For ILT we have  $\lambda = \frac{(x_{K+1}+\delta)^2}{4}$  from (9).



**Fig. 4**. Singular-value log-thresholding for matrix completion: Frobenius-norm of error vs. iterations.

Note that  $Q(\mathbf{x}, \mathbf{x}) = f(\mathbf{x})$ . Simplifying (12) we have

$$Q(\mathbf{x}, \mathbf{z}) = \sum_{i} \left( x_i - k_i(\mathbf{z}) \right)^2 + \lambda \log(x_i + |\delta|) + K(\mathbf{z}),$$
(13)

where  $k_i(\mathbf{z}) = z_i + a_i^T \mathbf{y} - a_i^T A \mathbf{z}$  and  $K(\mathbf{z})$  contains terms independent of  $\mathbf{x}$ . The optimization over  $\mathbf{x}$  is now separable, i.e. can be done independently for each coordinate. We can see that finding local minima over  $\mathbf{x}$  of  $Q(\mathbf{x}, \hat{\mathbf{x}}^n)$  to define  $\hat{\mathbf{x}}^{(n+1)}$  corresponds to iterative log-thresholding. Using this motivation for ILT, we can now show:

**Proposition 2**  $f(\hat{\mathbf{x}}^n) = Q(\hat{\mathbf{x}}^n, \hat{\mathbf{x}}^n)$  and  $Q(\hat{\mathbf{x}}^{n+1}, \hat{\mathbf{x}}^n)$ , are monotonically decreasing with iterations *n*, and *ILT* converges to its fixed points, if the spectral norm  $||A||_2 < 1$ .

The proof parallels the IHT proof of [18] using the fact that  $Q(\mathbf{x}^{n+1}, \mathbf{x}^n) = f(\mathbf{x}^n) + \|\mathbf{x}^{n+1} - \mathbf{x}^n\|^2 - \|A(\mathbf{x}^{n+1} - \mathbf{x}^n)\|^2$ , which is independent of the thresholding used. The main difference for ILT is that  $\hat{\mathbf{x}}^{n+1}$  is not the global minimum of  $Q(\mathbf{x}, \hat{\mathbf{x}}^n)$  but it still holds that  $Q(\hat{\mathbf{x}}^{n+1}, \hat{\mathbf{x}}^n) < Q(\hat{\mathbf{x}}^n, \hat{\mathbf{x}}^n)$ . To show convergence to fixed points, we use the proof technique of [Theorem 3] in [18]. Next, we have:

**Proposition 3** Any fixed point of (10) satisfies the following:

$$\begin{cases} \mathbf{a}_{i}^{T}(\mathbf{y} - A\bar{\mathbf{x}}) = \frac{\lambda}{2(\bar{x}_{i} + \delta)} & \text{if} \quad \bar{x}_{i} > x_{0} \\ \mathbf{a}_{i}^{T}(\mathbf{y} - A\bar{\mathbf{x}}) = \frac{\lambda}{2(\bar{x}_{i} - \delta)} & \text{if} \quad \bar{x}_{i} < -x_{0} \\ |\mathbf{a}_{i}^{T}(\mathbf{y} - A\bar{\mathbf{x}})| \le x_{0} & \text{otherwise} \end{cases}$$

In other words, if  $|\bar{x}_i| > x_0$ , then the corresponding gradient component satisfies local stationarity conditions for problem (7), and if  $|\bar{x}_i| < x_0$ , the gradient is bounded. *Proof:* Given a fixed point  $\bar{\mathbf{x}}$  of (10) define

$$s_i = \mathbf{a}_i^T (\mathbf{y} - A\bar{\mathbf{x}}), \tag{14}$$

Suppose first that  $\bar{x}_i + s_i > x_0$ . Explicitly writing (10),

$$\bar{x}_i - s_i + \delta = \sqrt{(\bar{x}_i + s_i + \delta)^2 - 2\lambda}$$

squaring both sides, and simplifying, we have

$$\mathbf{a}_i^T(\mathbf{y} - A\bar{\mathbf{x}}) = \frac{\lambda}{2(\bar{x}_i + \delta)}$$

which is precisely equivalent to local optimality of (7) with respect to the *i*th coordinate. Otherwise, suppose  $0 \le \bar{x}_i + s_i < x_0$ . Then we have  $\bar{x}_i = 0$ , and so  $s_i \le x_0$ .

**Proposition 4** For any fixed point  $\bar{x}$  of the ILT algorithm (7)  $\exists \epsilon > 0 \text{ s.t. for small perturbation } \|\eta\|_{\infty} < \epsilon$ , we have

$$Q(\bar{x}+\eta,\bar{x}) > Q(\bar{x}) + \sum_{i:\bar{x}_i=0} \eta_i^2 + \frac{3}{4} \sum_{i:\bar{x}_i\neq 0} \eta_i^2,$$

if  $\delta$  is small enough. Namely, the condition on  $\delta$  is:

$$\frac{\lambda}{\delta} + 2\delta > 2\sqrt{2\lambda}.\tag{15}$$

*Proof:* This result follows from Proposition 3, together with the proof technique of [18][Lemma 3]. In particular, for any perturbation  $\eta$ , we can write  $Q(\bar{x} + \eta, \bar{x}) - Q(\bar{x}, \bar{x})$  as

$$=\sum_{i}\left(-2\eta_{i}s_{i}+\eta_{i}^{2}+\lambda\log\left(\frac{|\bar{x}_{i}+\eta_{i}|+\delta}{|\bar{x}_{i}|+\delta}\right)\right)$$

Defining now  $\Gamma_0 = \{i : \bar{x}_i = 0\}$  and  $\Gamma_1 = \{i : \bar{x}_i \neq 0\}$ , we can use Proposition 3 to rewrite  $Q(\bar{x} + \eta, \bar{x}) - Q(\bar{x}, \bar{x})$  as:

$$\|\eta\|^{2} + \sum_{i \in \Gamma_{0}} \left( -2\eta_{i}s_{i} + \lambda \log\left(\frac{|\eta_{i}| + \delta}{|\delta|}\right) \right) + \sum_{i \in \Gamma_{1}} \left( -\frac{\eta_{i}\lambda}{\bar{x}_{i} + |\delta|} + \lambda \log\left(\frac{|\bar{x}_{i} + \eta_{i}| + \delta}{|\bar{x}_{i}| + \delta}\right) \right)$$

We now consider lower bounds for each of these two sums taking all  $x_i \ge 0$  WLOG:

$$\sum_{i \in \Gamma_0} -2\eta_i s_i + \lambda \log\left(\frac{|\eta_i| + \delta}{\delta}\right) \ge$$
$$= \sum_{i \in \Gamma_0} \lambda \log\left(1 + \frac{|\eta_i|}{\delta}\right) - 2\eta_i |x_0|$$
$$= \sum_{i \in \Gamma_0} \left(\lambda \frac{|\eta_i|}{\delta} - 2|\eta_i|x_0\right) - O(||\eta||^2)$$

Given that (15) holds, the quantity on the last line is positive for  $\eta$  small enough. For  $\Gamma_1$ , we have

$$\sum_{i\in\Gamma_1} -\frac{\eta_i\lambda}{\bar{x}_i+\delta} + \lambda\log\left(\frac{|\bar{x}_i+\eta_i|+\delta}{|\bar{x}_i|+\delta}\right) \ge \sum_{i\in\Gamma_1} \frac{-\eta_i\lambda+\eta_i\lambda}{\bar{x}_i+\delta} - \frac{1}{4}\frac{2\lambda\eta_i^2}{(|\bar{x}_i|+\delta)^2} \ge -\frac{1}{4}\sum_{i\in\Gamma_1} \eta_i^2$$

where we use  $\frac{(\bar{x}_i + \delta)^2}{2\lambda} \geq 1$  for  $i \in \Gamma_1$  in last line. Finally, using the fact that  $f(\bar{\mathbf{x}} + \eta) = Q(\bar{\mathbf{x}} + \eta, \bar{\mathbf{x}}) - Q(\bar{\mathbf{x}} + \eta, \bar{\mathbf{x}})$ 

Finally, using the fact that  $f(\mathbf{x} + \eta) = Q(\mathbf{x} + \eta, \mathbf{x}) - \|\eta\|^2 + \|A\eta\|^2$  and Proposition 4, we have:

**Proposition 5** If the singular values  $A_{\Gamma_1}$  are greater than  $\frac{1}{2}$ , these fixed points must be the local minima of (7).

We hope to relax this condition in future work.

### 7. REFERENCES

- S.S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.
- [2] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient Projection for Sparse Reconstruction: Application to Compressed Sensing and Other Inverse Problems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 586–597, dec 2007.
- [3] E. Van den Berg and M. P. Friedlander, "Probing the Pareto Frontier for Basis Pursuit Solutions," *SIAM Journal on Scientific Computing*, vol. 31, no. 2, pp. 890–912, jan 2009.
- [4] S. Osher, Y. Mao, B. Dong, and W. Yin, "Fast linearized Bregman iteration for compressive sensing and sparse denoising," *Communications in Mathematical Sciences*, vol. 8, no. 1, pp. 93–111, mar 2010.
- [5] D. M. Malioutov, M. Cetin, and A. S. Willsky, "Homotopy continuation for sparse signal representation," in *Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP'05). IEEE International Conference on.* IEEE, 2005, vol. 5, pp. v–733.
- [6] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Comm. on pure and applied mathematics*, vol. 57, no. 11, pp. 1413–1457, 2004.
- [7] P. L. Combettes and J. C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, pp. 185–212. Springer, 2011.
- [8] A. Beck and M. Teboulle, "A fast iterative shrinkagethresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.
- [9] M. Fazel, H. Hindi, and S. P. Boyd, "A rank minimization heuristic with application to minimum order system approximation," in *IEEE American Control Conf.*, 2001.
- [10] E. J. Candes, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted *l*1 minimization," *Journal* of Fourier Analysis and Applications, vol. 14, no. 5, pp. 877–905, 2008.
- [11] D. Needell, "Noisy signal recovery via iterative reweighted l1-minimization," in Forty-Third Asilomar Conference on Signals, Systems and Computers, 2009. IEEE, 2009, pp. 113–117.

- [12] A. Khajehnejad, W. Xu, S. Avestimehr, and B. Hassibi, "Weighted l1 minimization for sparse recovery with prior information," in *IEEE Int. Symposium on Inf. Theory*, 2009, 2009, pp. 483–487.
- [13] S. Voronin and R. Chartrand, "A new generalized thresholding algorithm for inverse problems with sparsity constraints," in *IEEE Int. Conf. on Acoustics*, *Speech, and Signal Processing*, 2013.
- [14] R. Mazumder, J. H. Friedman, and T. Hastie, "Sparsenet: Coordinate descent with nonconvex penalties," *Journal of the American Statistical Association*, vol. 106, no. 495, 2011.
- [15] G. Gasso, A. Rakotomamonjy, and S. Canu, "Recovering sparse signals with a certain family of nonconvex penalties and DC programming," *Signal Processing*, *IEEE Trans.*, vol. 57, no. 12, pp. 4686–4698, 2009.
- [16] H. Zou and R. Li, "One-step sparse estimates in nonconcave penalized likelihood models," *Annals of Statistics*, vol. 36, no. 4, 2008.
- [17] K. Lange, D. R. Hunter, and I. Yang, "Optimization transfer using surrogate objective functions," *Journal of comput. and graphical statistics*, vol. 9, no. 1, pp. 1–20, 2000.
- [18] T. Blumensath and M. E. Davies, "Iterative thresholding for sparse approximations," *Journal of Fourier Analysis* and Applications, vol. 14, no. 5-6, pp. 629–654, 2008.
- [19] J.F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on Optim.*, vol. 20, no. 4, pp. 1956–1982, 2010.
- [20] A. Maleki, "Coherence analysis of iterative thresholding algorithms," in 47th Annual Allerton Conference on Communication, Control, and Computing, 2009. IEEE, 2009, pp. 236–243.