PENALIZED L1 MINIMIZATION FOR RECONSTRUCTION OF TIME-VARYING SPARSE SIGNALS

Wei Chen[†] *Miguel R. D. Rodrigues*[⋆] *Ian J. Wassell*[†]

[†] Computer Laboratory, University of Cambridge, Cambridge CB3 0FD, UK
[⋆] Instituto de Telecomunicações, Department of Computer Science, University of Porto

ABSTRACT

In this paper, we propose a penalized ℓ_1 minimization algorithm for reconstructing a time-varying signal based on compressive sensing (CS) principles. The time-varying signal can be seen as a sequence of slow-changing frames. In the proposed algorithm, all frames of the sequence are sampled at an equal rate, which makes the encoder simpler than frame-categorized methods. We introduce a specialized Fréchet mean of the target frame and several adjacent frames as the penalty vector to make the algorithm close to ℓ_0 minimization. We prove that the specialized Fréchet mean is a good approximation of the target frame for a sequence of slow time-varying signals. Experimental results demonstrates the superior reconstruction quality of the proposed algorithm.

1. INTRODUCTION

Most practical signals of interest have a sparse representation if they are transformed into a suitable basis. The transformed signal can then be compressed, but this wastes resources as most of the sampled information is discarded after compression. Compressive sensing (CS) leverages the compressibility of the signal by directly acquiring a smaller quantity of random linear measurements that contain a little redundancy in the information level. Thus, CS appears to be an excellent approach for applications in which data acquisition is expensive such as imaging at non-visible wavelengths and sampling made by wireless sensor nodes.

Recently, CS has been studied for recovering a sequence of time-varying sparse signals from linear measurement vectors. This problem arises in the application of environment monitoring by a wireless sensor network using CS video, where the images are taken by a CS camera, for example, a single pixel camera [1] or a random lens imager [2]. There are also related scenarios such as real-time magnetic resonance imaging and channel equalization in communications [3]. Several reconstruction approaches have been proposed for solving the problem of recovering time-varying signals. One natural method involves jointly sampling the sequence and then recovering the sequence at once using higher dimension transformation [4]. This method leads to a large increase in the computational complexity. An alternative method is sampling each frame independently and then recovering the sequence frame by frame. Side information obtained from the first frame, the prior frame or the key frame, can be used to decrease the number of measurements required for recovering the following frames [5–8] or reduce the reconstruction time [3]. However, this method requires the encoder to have the ability to differentiate the key frames in the sequence.

Motivated by the fact that fewer measurements are needed for exact reconstruction with ℓ_0 minimization than with ℓ_1 minimization [9], we propose a penalized CS framework where a prior estimation of the sparse representations is used to make ℓ_1 minimization close to ℓ_0 minimization. The proposed solution differs from the other side information based methods mentioned previously in two aspects. First, although all those methods use side information to improve the performance, the first frame and the key frame in our solution are not necessarily identified and sampled more than the other frames. Second, those methods update the side information using the knowledge of the key frame or the reconstruction result of the previous frame. However, we derive the side information from several adjacent frames. Our penalized CS framework can be viewed as being similar to the updating step of the iteratively reweighted ℓ_1 minimization algorithm (IRL1) [9] where a weight vector is defined by the outcome of the previous iteration and is updated in each iteration. We propose to use a specialized Fréchet mean of several adjacent frames as the penalty vector, and our method only needs to solve the ℓ_1 minimization once for each frame.

2. PENALIZED ℓ_1 MINIMIZATION USING PRIOR ESTIMATION

Suppose that we observe T measurement vectors of a sequence of time-varying sparse signals, which can be written as

$$\mathbf{y}_t = \mathbf{\Phi}_t \mathbf{f}_t + \mathbf{e}_t, \tag{1}$$

where $t = 1, \ldots, T$ is the time index of each frame, $\mathbf{f}_t \in \mathbb{R}^N$ is the tth signal vector, $\mathbf{y}_t \in \mathbb{R}^M (M \ll N)$ is the tth

This work was supported by Fundação para a Ciência e a Tecnologia through the research project PDTC/EEA-TEL/100854/2008.

measurement vector, $\mathbf{\Phi}_t \in \mathbb{R}^{M\times N}$ is a matrix and $\mathbf{e}_t \in \mathbb{R}^M$ is a noise vector. We assume each signal can be transformed to a sparse version $\mathbf{x}_t \in \mathbb{R}^N$ of the form $\mathbf{x}_t = \mathbf{\Psi}^H \mathbf{f}_t$, where $\mathbf{\Psi} \in \mathbb{R}^{N\times N}$ is the orthonormal sparsifying matrix. The vector $\mathbf{f}_t$ can then be written in terms of a sparse representation $\mathbf{x}_t$ as $\mathbf{f}_t = \mathbf{\Psi}\mathbf{x}_t$. Consequently, the vector of observations $\mathbf{y}_t$ can also be written as follows:

$$\mathbf{y}_t = \mathbf{\Phi}_t \mathbf{\Psi} \mathbf{x}_t + \mathbf{e}_t = \mathbf{A}_t \mathbf{x}_t + \mathbf{e}_t, \tag{2}$$

where $\mathbf{A}_t = \mathbf{\Phi}_t \mathbf{\Psi}$. The classical CS approach to recover $\mathbf{x}_t$ involves solving the following ℓ_1 minimization:

$$\min_{\hat{\mathbf{x}}_t} \quad \|\hat{\mathbf{x}}_t\|_{\ell_1} \qquad \text{s.t.} \quad \|\mathbf{A}_t \hat{\mathbf{x}}_t - \mathbf{y}_t\|_{\ell_2} \le \epsilon, \tag{3}$$

where $\epsilon > 0$ relates to an estimate of the noise level.

We assume that the sparse representations change slowly during the observing time of the sequence. Reconstructing the sequence of signals independently is not an efficient way as inter-frame correlation is not exploited. With this idea in mind, we propose a penalized CS framework using a prior estimation $\tilde{\mathbf{x}}_t \in \mathbb{R}^N$ of the representation as a penalty vector in the ℓ_1 minimization algorithm. Our solution can be expressed as:

$$\min_{\hat{\mathbf{x}}_t} \quad \left\| \frac{\hat{\mathbf{x}}_t}{\mathbf{p}_t} \right\|_{\ell_1} \qquad \text{s.t.} \quad \|\mathbf{A}_t \hat{\mathbf{x}}_t - \mathbf{y}_t\|_{\ell_2} \le \epsilon, \tag{4}$$

where $\frac{*}{*}$ denotes array division and $\mathbf{p}_t \in \mathbb{R}^N$ is the penalty of the tth frame with elements $p_{t,i} = |\tilde{x}_{t,i}| + \rho$ $(i = 1, \ldots N)$. Here, ρ is a positive parameter to ensure that the algorithm is well-defined.

The motivation of the approach relates to the fact that the ℓ_1-norm $\left\| \frac{\hat{\mathbf{x}}_t}{\mathbf{p}_t} \right\|_{\ell_1}$ in (4) is a better approximation to the ℓ_0-norm than the ℓ_1-norm $\|\hat{\mathbf{x}}_t\|_{\ell_1}$ in (3) is because $\left\| \frac{\hat{\mathbf{x}}_t}{\mathbf{p}_t} \right\|_{\ell_1} \approx \|\hat{\mathbf{x}}_t\|_{\ell_0}$ for a good prior estimate $\mathbf{p}_t$. Consequently, one expects the algorithm in (4) to outperform the algorithm in (3), in terms of higher estimation quality for a certain number of linear measurements, or, alternatively, a lower number of linear measurements for certain estimation quality. Successful implementation of the penalized CS algorithm depends on the quality of the estimation $\tilde{\mathbf{x}}_t$ of the actual sparse representation $\mathbf{x}_t$.

3. PRIOR ESTIMATION OF TIME-VARYING SPARSE SIGNAL: A FRÉCHET MEAN APPROACH

As the sparse representation changes slowly over time, a natural idea is to use some form of the mean of several adjacent frames as an approximation. However, it appears that we cannot compute the mean vector directly without knowing explicitly the sparse representations. Instead, we propose to use the specialized Fréchet mean as a prior, which can be acquired by solving an ordinary least squares problem.

The Fréchet mean $\tilde{\mathbf{x}}$ of K adjacent frames $\mathbf{y}_k$, $k = 1, \ldots, K$, is defined as follows:

$$\tilde{\mathbf{x}} = \arg\min_{\tilde{\mathbf{x}}} \sum_{k=1}^{K} \lambda_k d^2(\tilde{\mathbf{x}}, \mathbf{y}_k, \mathbf{A}_k), \tag{5}$$

where $\lambda_k > 0$ is the contribution weight of the kth frame and

$$d(\tilde{\mathbf{x}}, \mathbf{y}_k, \mathbf{A}_k) = \|\mathbf{A}_k \tilde{\mathbf{x}} - \mathbf{y}_k\|_{\ell_2}. \tag{6}$$

is a Euclidean distance function. In view of equation (2), we can also rewrite the distance function as $d(\tilde{\mathbf{x}}, \mathbf{y}_k, \mathbf{A}_k) = \|\mathbf{A}_k \tilde{\mathbf{x}} - \mathbf{A}_k \mathbf{x}_k - \mathbf{e}_k\|_{\ell_2}$.

The Fréchet mean minimizes the sum of weighted squared distances between the observations and prediction. The weights allow for the possibility of some frames contributes more than other to the value of the Fréchet mean. For example, consider the reconstruction of some target frame by using a higher weight on the target frame than on other previous frames we expect to acquire a more accurate estimation of the sparse representation of the target frame. This is then used as prior estimate in the previous ℓ_1 minimization algorithm to improve performance.

We compute the specialized Fréchet mean defined in (5) and (6) by solving the following optimization problem:

$$\tilde{\mathbf{x}} = \arg\min_{\tilde{\mathbf{x}}} \qquad \left\| \hat{\mathbf{A}} \tilde{\mathbf{x}} - \hat{\mathbf{y}} \right\|_{\ell_2}^2, \tag{7}$$

where the extended sensing matrix $\hat{\mathbf{A}}$ and the extended measurement vector are given by $\hat{\mathbf{A}} = \left[\sqrt{\lambda_1}\mathbf{A}_1^H \cdots \sqrt{\lambda_K}\mathbf{A}_K^H\right]^H$ and $\hat{\mathbf{y}} = \left[\sqrt{\lambda_1}\mathbf{y}_1^H \cdots \sqrt{\lambda_K}\mathbf{y}_K^H\right]^H$. If $rank(\hat{\mathbf{A}}) = N$, (7) turns out to be an ordinary least squares problem, where the solution can be written explicitly as $\tilde{\mathbf{x}} = (\hat{\mathbf{A}}^H \hat{\mathbf{A}})^{-1} \hat{\mathbf{A}}^H \hat{\mathbf{y}}$ or computed by a conjugate gradient (CG) algorithm [10]. Consequently, for a sequence of independently generated random matrices $\mathbf{A}_k$, at least $K = \lceil \frac{N}{M} \rceil$ frames are needed in the calculation of the prior estimation of a target frame.

3.1. Estimation error

We aim to investigate the error between the target frame estimate given by the specialized Fréchet mean and the real target frame. Before presenting our results, we first give the definition of the restricted isometry property (RIP) [11] which is a widely used tool for analyzing random projections. Formally, a matrix $\mathbf{A}$ of size $M \times N$ is said to satisfy the RIP of order S with restricted isometry constant (RIC) $\delta_S \in (0, 1)$ as the smallest number such that

$$(1 - \delta_S)\|\mathbf{x}\|_{\ell_2}^2 \le \|\mathbf{A}\mathbf{x}\|_{\ell_2}^2 \le (1 + \delta_S)\|\mathbf{x}\|_{\ell_2}^2 \tag{8}$$

holds for all $\mathbf{x}$ with $\|\mathbf{x}\|_{\ell_0} \le S$.

Theorem 1 *Consider the measurement model* $\mathbf{y}_k = \mathbf{A}_k \mathbf{x}_k + \mathbf{e}_k$, $k = 1, \ldots, K$, *where the matrices* $\mathbf{A}_k$ *have*

RIC $\delta_{S,k}$. Fix the integers S_1, S_2 and the positive numbers λ_k, $k = 1, \ldots, K$. If $\|\mathbf{x}_k\|_{\ell_0} \leq S_1$, $k = 1, \ldots, K$, $\|\mathbf{x}_{k_1} - \mathbf{x}_{k_2}\|_{\ell_0} \leq S_2$, $k_1, k_2 \in \{1, \ldots, K\}$ and $\|\mathbf{e}_k\|_{\ell_2}^2 \leq \epsilon^2$, $k \in \{1, \ldots, K\}$. Then

$$\|\tilde{\mathbf{x}} - \mathbf{x}_t\|_{\ell_2}^2 \leq \sum_{k=1}^{K} C_{1,k} \|\mathbf{x}_t - \mathbf{x}_k\|_{\ell_2}^2 + C_{2,k}\epsilon^2, \; t = 1, \ldots, K \tag{9}$$

where $Q = S_1 + (K-1)S_2$, $C_{1,k} = \frac{\lambda_k(1+\delta_{S_2,k})}{\lambda_t(1-\delta_{Q,t})}$ and $C_{2,k} = \frac{\lambda_k + \frac{\lambda_t}{K}}{\lambda_t(1-\delta_{Q,t})}$.

Proof: As the specialized Fréchet mean $\tilde{\mathbf{x}}$ is the point that minimizes the sum of squared distances to all the other points, we have

$$\sum_{k=1}^{K} \lambda_k \|\mathbf{A}_k\tilde{\mathbf{x}} - \mathbf{y}_k\|_{\ell_2}^2 \leq \sum_{k=1}^{K} \lambda_k \|\mathbf{A}_k\mathbf{x}_t - \mathbf{y}_k\|_{\ell_2}^2. \tag{10}$$

Then, according to the RIP (8) and the inequality (10), we can derive

$$\begin{aligned}\|\tilde{\mathbf{x}} - \mathbf{x}_t\|_{\ell_2}^2 &\leq \frac{1}{1-\delta_{Q,t}} \|\mathbf{A}_t(\tilde{\mathbf{x}} - \mathbf{x}_t)\|_{\ell_2}^2 \\ &\leq \frac{1}{1-\delta_{Q,t}} \|\mathbf{A}_t\tilde{\mathbf{x}} - \mathbf{y}_t\|_{\ell_2}^2 + \frac{\|\mathbf{e}_t\|_{\ell_2}^2}{1-\delta_{Q,t}} \\ &\leq \frac{1}{1-\delta_{Q,t}} \sum_{k=1}^{K} \frac{\lambda_k}{\lambda_t} \|\mathbf{A}_k\tilde{\mathbf{x}} - \mathbf{y}_k\|_{\ell_2}^2 + \frac{\epsilon^2}{1-\delta_{Q,t}},\end{aligned} \tag{11}$$

and

$$\begin{aligned}\|\mathbf{A}_k\mathbf{x}_t - \mathbf{y}_k\|_{\ell_2}^2 &\leq \|\mathbf{A}_k(\mathbf{x}_t - \mathbf{x}_k)\|_{\ell_2}^2 + \|\mathbf{e}_k\|_{\ell_2}^2 \\ &\leq (1+\delta_{S_2,k}) \|\mathbf{x}_t - \mathbf{x}_k\|_{\ell_2}^2 + \epsilon^2.\end{aligned} \tag{12}$$

Then, from the inequality (10), (11) and (12), we can deduce that

$$\|\tilde{\mathbf{x}} - \mathbf{x}_t\|_{\ell_2}^2 \leq \sum_{k=1}^{K} \frac{\lambda_k(1+\delta_{S_2,k})}{\lambda_t(1-\delta_{Q,t})} \|\mathbf{x}_t - \mathbf{x}_k\|_{\ell_2}^2 + \frac{\left(\lambda_k + \frac{\lambda_t}{K}\right)\epsilon^2}{\lambda_t(1-\delta_{Q,t})}. \tag{13}$$

Remark: Theorem 1 gives the upper bound of the squared pairwise Euclidean distance between the specialized Fréchet mean and a target frame. The bound is tight in the noiseless case $\epsilon = 0$ and all frames considered are same, i.e. $\mathbf{x}_k = \mathbf{x}_t$. In this case, the specialized Fréchet mean is equal to the target frame. If the signal changes slowly over time, then the squared estimation error bound is equal to a sum of weighed distances between the target frame and other frames plus a noise term. For example, suppose $\delta_{S_2} = 0.4$, $\delta_Q = 0.5$, $\lambda_k = 1$, $\lambda_t = 2$, $\epsilon^2 \leq 0.001\|\mathbf{x}_t\|_{\ell_2}^2$ and the Euclidean distance between the target frame $\mathbf{x_t}$ and any of the $K - 1 = 4$ adjacent frame $\mathbf{x_k}$ satisfies $\|\mathbf{x}_t - \mathbf{x}_k\|_{\ell_2}^2 \leq 0.01\|\mathbf{x}_t\|_{\ell_2}^2$, then the Euclidean distance between the the Fréchet Mean and the target frame satisfies $\|\tilde{\mathbf{x}} - \mathbf{x}_t\|_{\ell_2}^2 \leq 0.06\|\mathbf{x}_t\|_{\ell_2}^2$. Note that the real error will be smaller than the bound given in Theorem 1.

4. ALGORITHM

The pseudo-code for penalized ℓ_1 minimization for a sequence of signals is described as follows:

Algorithm 1 Reconstruction procedure

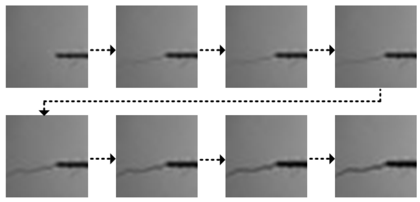
Input: A time sequence of measurement matrices $\mathbf{A}_t$ $(t = 1, 2, \ldots)$, a time sequence of measurement vectors $\mathbf{y}_t$ $(t = 1, 2, \ldots)$, K positive weights λ_k $(k = 1, \ldots, K)$, an estimate of the noise level ϵ and a positive parameter ρ;

Output: An estimate time sequence of signal vector $\hat{\mathbf{x}}_t$ $(t = 1, 2, \ldots)$.

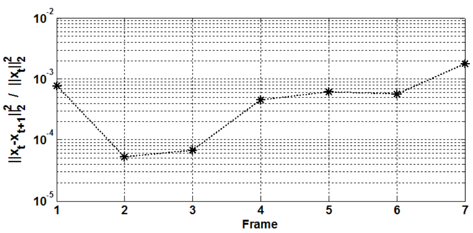
Process: For $t > 0$, do

1. Compute the specialized Fréchet mean $\tilde{\mathbf{x}}_t$ by considering $\mathbf{y}_t$ and $K - 1$ adjacent measurement vectors with weights λ_k;
2. Compute the penalty $\mathbf{p}_t = |\tilde{\mathbf{x}}_t| + \rho$;
3. Compute $\mathbf{x}_t = \arg\min_{\hat{\mathbf{x}}_t} \left\|\frac{\hat{\mathbf{x}}_t}{\mathbf{p}_t}\right\|_{\ell_1}$ $s.t.$ $\|\mathbf{A}_t\hat{\mathbf{x}}_t - \mathbf{y}_t\|_{\ell_2}^2 \leq \epsilon$.

For the first frame, the adjacent frames could be the next $K-1$ frames. For other frames, their adjacent frames could be former frames or later ones. The weights λ_k $(k = 1, \ldots, K)$ are required as the input to the algorithm, which can be simply given the same value. An interesting approach is to use different weights to take account of the time-varying nature of the sequence of signals.

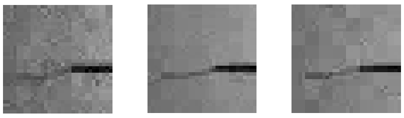


(a) A sequence of crack images



(b) Small varying of wavelet coefficients

Fig. 1. A sequence of crack images and the normalized squared Euclidean distances of wavelet coefficients between adjacent frames.



(a) (b) (c)

Fig. 2. Reconstructed result of the $8th$ crack image. (a) Traditional ℓ_1 minimization with $\frac{M}{N} = 0.3$; (b) Proposed algorithm with $\frac{M}{N} = 0.2$; (c) Proposed algorithm with $\frac{M}{N} = 0.15$

5. EXPERIMENTAL RESULTS

In this section, we test the reconstruction performance of the proposed algorithm with a sequence of crack images, which are taken by a sensor camera for the purpose of monitoring bridge condition. As shown in Fig 1(a), there are 8 frames, each of size 32×32 pixels. We calculate the Daubechies wavelet "db1" coefficients $\mathbf{x}_t$ of each frame at level 2, and then plot the normalized squared Euclidean distance $\frac{\|\mathbf{x}_t - \mathbf{x}_{t+1}\|_{\ell_2}^2}{\|\mathbf{x}_t\|_{\ell_2}^2}$ between any pair of adjacent frames in Fig 1(b). We note that the sparse representations of the sequence are quite close in the Euclidean space, which verifies the assumption of a slow time-varying sparse signal.

The $M \times N$ sensing matrices $\mathbf{\Phi}_t$ $(t = 1, \ldots, 8)$ are generated independently with i.i.d. Gaussian entries $\mathcal{N}(0, \frac{1}{M})$. In Fig 2, we compare the proposed algorithm with traditional ℓ_1 minimization. For the reconstruction of the tth image of the sequence, we let $K = 8$, $\lambda_t = 7$ and $\lambda_k = 1$ $(k = 1, \ldots, 8;\ k \neq t)$, which makes the target frame have a higher impact than other frames in the calculation of the Fréchet mean. We note in Fig 2 that the proposed penalized ℓ_1 minimization outperforms the traditional CS algorithm by visual comparison. We achieve similar performance for the other images of the sequence which are not shown here due to the limited space.

Fig 3 shows the accuracy of reconstruction using the signal-to-noise ratio (SNR) $\frac{\|\mathbf{x}_t\|_{\ell_2}}{\|\mathbf{x}_t - \hat{\mathbf{x}}_t\|_{\ell_2}}$, where $\hat{\mathbf{x}}_t$ is the reconstructed vector. In this experiment, the number of measurements are reduced to 15% of the original except for the first frame of the Modified-CS [6], which has 50% measurements. We calculate the SNR by averaging over 200 experiment trials. The proposed method and the Modified-CS have similar performance except for the first frame. The Modified-CS algorithm requires specific notification of the key frames while the proposed method does not. Thus, the proposed method can be applied in situations where the encoder is ignorant concerning the locations of the key frames.

6. CONCLUSIONS

In this paper, we study the problem of reconstructing a time-varying signal with a small number of linear incoherent measurements. We proposed a penalized ℓ_1 minimization algorithm for reconstruction that even uses fewer measurements than does the traditional CS method. We use a specialized Fréchet mean as the penalty vector, which is shown to be close to the original signal. Experiments demonstrate the advantage of our method in the reconstruction of a sequence of images.

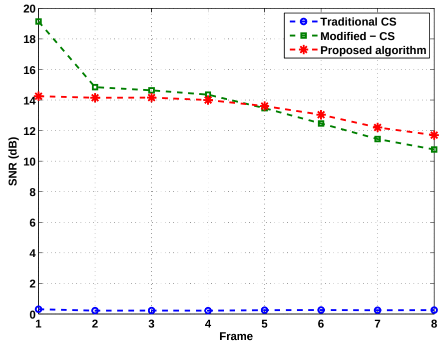


Fig. 3. Comparing the proposed algorithm with traditional CS and the Modified-CS algorithm.

7. REFERENCES

[1] M.F. Duarte, M.A. Davenport, D. Takhar, J.N. Laska, T. Sun, K.F. Kelly, and R.G. Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 83–91, 2008.

[2] R. Fergus, A. Torralba, and W.T. Freeman, "Random lens imaging," *MIT CSAIL TR*, vol. 58, 2006.

[3] M. Salman Asif and J. Romberg, "Dynamic updating for ℓ_1 minimization," *IEEE Journal on selected topics in Signal Processing*, vol. 4, no. 2, pp. 421–434, 2010.

[4] M.B. Wakin, J.N. Laska, M.F. Duarte, D. Baron, S. Sarvotham, D. Takhar, K.F. Kelly, and R.G. Baraniuk, "Compressive imaging for video representation and coding," in *Proceedings of Picture Coding Symposium (PCS*. Citeseer, 2006.

[5] T.T. Do, Y. Chen, D.T. Nguyen, N. Nguyen, L. Gan, and T.D. Tran, "Distributed compressed video sensing," in *Proceedings of the 16th IEEE international conference on Image processing*. IEEE Press, 2009, pp. 1381–1384.

[6] N. Vaswani and W. Lu, "Modified-CS: Modifying compressive sensing for problems with partially known support," *IEEE Transaction on Signal Processing*, vol. PP, no. 99, pp. 1–1, 2010.

[7] D. Angelosante, J.A. Bazerque, and G.B. Giannakis, "Online adaptive estimation of sparse signals: Where rls meets the ell_1 -norm," *Signal Processing, IEEE Transactions on*, vol. 58, no. 7, pp. 3436 –3447, 2010.

[8] Yilun Chen, Yuantao Gu, and A.O. Hero, "Sparse lms for system identification," in *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, 2009, pp. 3125 –3128.

[9] 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.

[10] J. Nocedal and S.J. Wright, *Numerical optimization, 2nd ed.*, Springer verlag, 2006.

[11] E. Candès and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.