# Reconstruction of Sparse Signals by Minimizing a Re-Weighted Approximate $\ell_0$-Norm in the Null Space of the Measurement Matrix

Jeevan K. Pant, Wu-Sheng Lu, and Andreas Antoniou
Department of Electrical and Computer Engineering, University of Victoria
P.O. Box 3055 STN CSC, Victoria, B.C. V8W 3P6, CANADA
Email:{jkpant, wslu}@ece.uvic.ca, aantoniou@ieee.org

***Abstract*—A new algorithm for signal reconstruction in a compressive sensing framework is presented. The algorithm is based on minimizing a re-weighted approximate $\ell_0$-norm in the null space of the measurement matrix, and the unconstrained optimization involved is performed by using a quasi-Newton algorithm. Simulation results are presented which demonstrate that the proposed algorithm yields improved signal reconstruction performance and requires a reduced amount of computation relative to iteratively re-weighted algorithms based on the $\ell_p$-norm with $p < 1$. When compared with a known algorithm based on a smoothed $\ell_0$-norm, improved signal reconstruction is achieved although the amount of computation is increased somewhat.**

## I. Introduction

Compressive sensing (CS) comprises a collection of methods of representing a signal on the basis of a limited number of measurements and then recovering the signal from these measurements. It is now known that if a signal is measured in terms of independent random projections (i.e., inner products of the signal with random waveforms), then the signal can be reconstructed using these measurements as long as a certain condition that involves the dimension and sparsity of the signal and the number of measurements collected is satisfied [1]-[3]. Algorithms for signal reconstruction in a CS framework are referred to as sparse signal reconstruction (SSR) algorithms. One of the most successful of these algorithms, known as *basis pursuit* (BP), is based on constrained $\ell_1$-norm minimization [4]. Several SSR algorithms based on constrained $\ell_p$-norm minimization with $p < 1$ have also been proposed [5], [6]. An SSR algorithm based on the optimization of a smoothed approximate $\ell_0$-norm is studied in [7] where simulation results are compared with corresponding results obtained with several existing SSR algorithms with respect to reconstruction performance and computational complexity. These results favor the use of the approximate $\ell_0$-norm.

In this paper, we present a new signal reconstruction algorithm for CS. Like the algorithm in [7], the proposed algorithm is based on the minimization of a smoothed approximate $\ell_0$-norm but it differs in several aspects. First, the $\ell_0$-norm minimization in our algorithm is carried out in the null space of the measurement matrix. As a result, the constraints on measurements are eliminated and the problem under consideration becomes unconstrained. This opens the door for the use of more efficient algorithms for the optimization. In addition, by working in the null space, the size of the minimization problem is considerably reduced. Second, a re-weighting technique is incorporated into the minimization procedure so as to force the algorithm to reach the desired sparse solution faster. Third, instead of using a steepest-descent algorithm as is done in [7], a quasi-Newton algorithm [8] is used to optimize the unconstrained objective function, which yields better solutions than solutions obtained by using several existing SSR algorithms [6], [7].

## II. Background

A real-valued, discrete-time signal represented by a vector $\boldsymbol{x}$ of size $N$ is said to be $K$-*sparse* if it has $K$ nonzero components with $K \ll N$. Although most real-world signals do not look sparse under the canonical basis, many natural and man-made signals admit sparse representations with respect to an appropriate basis [9]. For this reason, in the rest of the paper we focus on the class of $K$-sparse signals. The acquisition of a sparse signal $\boldsymbol{x}$ in CS theory is carried out by obtaining inner products of $\boldsymbol{x}$ with $M$ different waveforms $\{\boldsymbol{\phi}_1,\ \boldsymbol{\phi}_2,\ \ldots,\ \boldsymbol{\phi}_M\}$, namely, $y_k = \langle \boldsymbol{\phi}_k, \boldsymbol{x}\rangle$ for $k = 1, 2, \ldots, M$. If we let $\boldsymbol{y} = [y_1\ y_2\ \cdots\ y_M]$ and $\boldsymbol{\Phi} = \left[\boldsymbol{\phi}_1^T\ \boldsymbol{\phi}_2^T\ \cdots\ \boldsymbol{\phi}_M^T\right]^T$, then the data acquisition process in a CS framework can be described as

$$\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x} \tag{1}$$

The size of the measurement matrix in (1) is $M \times N$, typically with $M \ll N$. In this way, the signal $\boldsymbol{x}$ is 'sensed' by a reduced or 'compressed' number of measurements, hence the name of compressive sensing.

With $M < N$, (1) is an underdetermined system of linear equations; hence reconstructing signal $\boldsymbol{x}$ from measurement $\boldsymbol{y}$ is in general an *ill-posed* problem [10]. However, the sparsest solution of (1) can be obtained by solving the constrained optimization problem

$$\begin{aligned} &\underset{\boldsymbol{x}}{\text{minimize}} \quad ||\boldsymbol{x}||_0 \\ &\text{subject to:} \quad \boldsymbol{\Phi}\boldsymbol{x} = \boldsymbol{y} \end{aligned} \tag{2}$$

where $||\boldsymbol{x}||_0$ is the $\ell_0$-norm of $\boldsymbol{x}$ defined as $||\boldsymbol{x}||_0 = \sum_{i=1}^{N} |x_i|^0$ which, in effect, counts the number of nonzero components in $\boldsymbol{x}$. Unfortunately, (2) is a combinatorial optimization problem whose computational complexity grows exponentially with the signal size, $N$. A key result in the CS theory is that if $\boldsymbol{x}$ is $K$-sparse, the waveforms in $\{\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \ldots, \boldsymbol{\phi}_M\}$ are independent and identically distributed (i.i.d.) random waveforms, and the number of measurements, $M$, satisfies the condition

$$M \geq c \cdot M \cdot \log(N/K) \tag{3}$$

with $c$ a small constant, then $\boldsymbol{x}$ can be reconstructed by solving the convex problem

$$\begin{aligned} \underset{\boldsymbol{x}}{\text{minimize}} \quad & ||\boldsymbol{x}||_1 \\ \text{subject to:} \quad & \boldsymbol{\Phi x} = \boldsymbol{y} \end{aligned} \tag{4}$$

where $||\boldsymbol{x}||_1$ denotes the $\ell_1$-norm defined as $||\boldsymbol{x}||_1 = \sum_{i=1}^{N} |x_i|$[1]-[3].

For real-valued data $\{\boldsymbol{\Phi}, \boldsymbol{y}\}$, (4) is a *linear programming* (LP) problem whereas for complex-valued $\{\boldsymbol{\Phi}, \boldsymbol{y}\}$ (4) can be cast as a *second-order cone programming* (SOCP) problem [8]. Both the LP and SOCP problems can be solved using reliable and efficient software.

The condition in (3) turns out to be quite restrictive for many practical problems. Several authors have recently studied new algorithms for signal recovery by means of an $\ell_p$ minimization approach where the problem

$$\begin{aligned} \underset{\boldsymbol{x}}{\text{minimize}} \quad & ||\boldsymbol{x}||_p^p \\ \text{subject to:} \quad & \boldsymbol{\Phi x} = \boldsymbol{y} \end{aligned} \tag{5}$$

is solved instead of that in (4) where $||\boldsymbol{x}||_p^p = \sum_{i=1}^{N} |x_i|^p$ with $0 \leq p < 1$ [5], [6]. With $p < 1$, the problem in (5) becomes nonconvex and multiple local solutions exist. However, if the problem is solved with sufficient care, improved results can be obtained relative to those obtained by solving the problem in (4) [6]. In [7], the signal recovery problem is achieved by minimizing a smoothed approximate $\ell_0$-norm of $\boldsymbol{x}$ subject to the condition $\boldsymbol{\Phi x} = \boldsymbol{y}$, namely,

$$\begin{aligned} \underset{\boldsymbol{x}}{\text{minimize}} \quad & F(\boldsymbol{x}) = \sum_{i=1}^{N} \left(1 - e^{-x_i^2/2\sigma^2}\right) \\ \text{subject to:} \quad & \boldsymbol{\Phi x} = \boldsymbol{y} \end{aligned} \tag{6}$$

where $\sigma > 0$ is a parameter. This problem is solved by using an algorithm based on the steepest-descent approach. This algorithm was found to offer improved signal reconstruction performance and computational complexity with respect to several existing algorithms. In the rest of the paper, the algorithm in [7] is referred to as the SL0 algorithm.

## III. Signal Reconstruction by Minimizing a Re-Weighted Approximate $\ell_0$-Norm in Null Space

In this section, we present a method for the reconstruction of signal $\boldsymbol{x}$ using measurement $\boldsymbol{y} = \boldsymbol{\Phi x}$ by minimizing a re-weighted approximate $\ell_0$-norm of $\boldsymbol{x}$ in the null space of $\boldsymbol{\Phi}$.

### A. *Working in the Null Space of* $\boldsymbol{\Phi}$

It is well known that all solutions of $\boldsymbol{\Phi x} = \boldsymbol{y}$ can be parameterized as

$$\boldsymbol{x} = \boldsymbol{x}_s + \boldsymbol{V}_r \xi \tag{7}$$

where $\boldsymbol{x}_s$ is a solution of $\boldsymbol{\Phi x} = \boldsymbol{y}$, $\boldsymbol{V}_r$ is a $N \times (N - M)$ matrix whose columns constitute an orthonormal basis of the null space of $\boldsymbol{\Phi}$, and $\boldsymbol{\xi}$ is a parameter vector of dimension $N - M$. Vector $\boldsymbol{x}_s$ and matrix $\boldsymbol{V}_r$ in (7) can be evaluated by using the singular-value decomposition or, more efficiently, the QR decomposition of matrix $\boldsymbol{\Phi}$ [8],[10]. Using (7), the constrained problem in (6) is reduced to

$$\underset{\boldsymbol{\xi}}{\text{minimize}} \quad F_\sigma(\boldsymbol{\xi}) = \sum_{i=1}^{N} \left\{1 - e^{-[x_s(i) + \boldsymbol{v}_1^T \boldsymbol{\xi}]^2/2\sigma^2}\right\} \tag{8}$$

where $\boldsymbol{v}_i^T$ denotes the $i$th row of matrix $\boldsymbol{V}_r$. The objective function in (8) remains differentiable and its gradient can be obtained as

$$\nabla F_\sigma(\boldsymbol{\xi}) = \frac{\boldsymbol{V}_r^T \boldsymbol{g}}{\sigma^2} \tag{9a}$$

where $\boldsymbol{g} = [g_1 \ g_2 \ \cdots \ g_N]^T$ with

$$g_i = \left[x_s(i) + \boldsymbol{v}_i^T \boldsymbol{\xi}\right] e^{-[x_s(i) + \boldsymbol{v}_i^T \boldsymbol{\xi}]^2/2\sigma^2} \tag{9b}$$

Evidently, working in the null space of $\boldsymbol{\Phi}$ through the parameterization in (7) facilitates the elimination of the constraints in (6) and, furthermore, it reduces the problem size from $N$ to $N - M$. In this way, unconstrained optimization methods that are more powerful than the steepest-descent method can be applied to improve the reconstruction performance, as will be shown in Sec. III-C.

### B. *Re-Weighting the Approximate* $\ell_0$*-Norm*

Signal reconstruction based on the solution of the problem in (8) works well but the technique can be considerably enhanced by incorporating a re-weighting strategy. The re-weighted unconstrained problem is given by

$$\underset{\boldsymbol{\xi}}{\text{minimize}} \quad F_\sigma(\boldsymbol{\xi}) = \sum_{i=1}^{N} w_i \left\{1 - e^{-[x_s(i) + \boldsymbol{v}_i^T \boldsymbol{\xi}]^2/2\sigma^2}\right\} \tag{10}$$

where $w_i$ are positive scalars that form a weight vector $\boldsymbol{w} = [w_1 \ w_2 \ \cdots \ w_N]$. Starting with an initial $\boldsymbol{w}^{(0)} = \boldsymbol{e}_N$ (the all-one vector of dimension $N$), in the $(k+1)$th iteration the weight vector is updated to $\boldsymbol{w}^{(k+1)}$ with its $i$th component given by

$$w_i^{(k+1)} = \frac{1}{|x_i^{(k)}| + \epsilon} \tag{11}$$

where $x_i^{(k)}$ denotes the $i$th component of vector $\boldsymbol{x}^{(k)}$ obtained in the $k$th iteration as $\boldsymbol{x}^{(k)} = \boldsymbol{x}_s + \boldsymbol{V}_r \boldsymbol{\xi}^{(k)}$, and $\epsilon$ is a small positive scalar to prevent numerical instability when $|x_i^{(k)}|$ approaches zero. Evidently, for a small $|x_i^{(k)}|$ the re-weighting strategy in (11) yields a large weight $w_i^{(k+1)}$ and hence solving the problem in (10) tends to reduce $|x_i^{(k)}|$ further thus forcing a sparse solution. The gradient of the re-weighted objective function in (10) is still given by (9a) except that (9b) is slightly modified to

$$g_i = w_i \left[ x_s(i) + \boldsymbol{v}_i^T \boldsymbol{\xi} \right] e^{-[x_s(i)+\boldsymbol{v}_i^T\boldsymbol{\xi}]^2/2\sigma^2} \tag{12}$$

It should be mentioned that various re-weighting techniques have been recently proposed in the literature, see, for example, [6], [11]. In the algorithms presented in these papers, a sequence of optimizations is carried out where the weight calculated in a given optimization is used to re-weight the objective function for the next optimization, i.e., re-weighting is used once in each optimization. In the proposed algorithm, the re-weighting in (11) is used in each iteration.

### C. *Optimization of the Norm Using a Quasi-Newton Method*

It can be readily verified that the region where function $F_\sigma(\boldsymbol{\xi})$ in (10) is convex is closely related to the value of parameter $\sigma$: the greater the value of $\sigma$, the larger the convex region. On the other hand, for $F_\sigma(\boldsymbol{\xi})$ to well approximate the $\ell_0$-norm of $\boldsymbol{x}$, $\sigma$ must be sufficiently small. For this reason, the solution of the optimization problem in (10) is obtained using a relatively large $\sigma = \sigma_0$. This solution is then used as the initial point for minimizing $F_\sigma(\boldsymbol{\xi})$ with a reduced value of $\sigma$, say, $r \cdot \sigma$ with $r < 1$. This procedure is repeated until function $F_\sigma(\boldsymbol{\xi})$ with $\sigma \le \sigma_J$ is minimized where $\sigma_J$ is a prescribed value of $\sigma$. For a fixed value of $\sigma$, the problem in (10) is solved by using a quasi-Newton algorithm where an approximation of the inverse of the Hessian is obtained by using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) update formula [8]. We note that applying a quasi-Newton algorithm is particularly convenient in the present application because the gradient of the objective function can be efficiently evaluated using the closed-form formulas in (9a) and (12). As demonstrated in our simulation studies (see Sec. IV), the application of the BFGS quasi-Newton algorithm to the problem in (10) yields an improved solution relative to that obtained by using the steepest-decent algorithm.

### D. *Algorithm*

The proposed method for reconstructing a sparse signal $\boldsymbol{x}$ using a measurement $\boldsymbol{y} = \boldsymbol{\Phi x}$ can now be implemented in terms of the algorithm in Table I. This will be referred to hereafter as the *null-space re-weighted approximate $\ell_0$-norm* (NRAL0) algorithm.

We conclude this section with a remark concerning the initial value of parameter $\sigma$. It can be shown that function $F_\sigma(\boldsymbol{\xi})$ remains convex in the region where the largest magnitude of the components of $\boldsymbol{x} = \boldsymbol{x}_s + \boldsymbol{V}_r\boldsymbol{\xi}$ is less than $\sigma$. Based on this, a reasonable initial value of $\sigma$ can be chosen as $\sigma_0 = \max|\boldsymbol{x}_s| + \tau$ where $\tau$ is a small positive scalar. As the algorithm starts at the origin $\boldsymbol{\xi}^{(0)} = \boldsymbol{0}$, the above choice of $\sigma_0$ ensures that the optimization starts in a convex region. This greatly facilitates the convergence of the proposed algorithm.

TABLE I
THE NULL-SPACE RE-WEIGHTED APPROXIMATE $\ell_0$-NORM ALGORITHM

| |
|---|
| **Step 1** |
| Input $\boldsymbol{\Phi}$, $\boldsymbol{x}_s$, $\sigma_J$, $r$, $\tau$, and $\epsilon$. |
| **Step 2** |
| Set $\boldsymbol{\xi}^{(0)} = \boldsymbol{0}$, $\boldsymbol{w}^{(0)} = \boldsymbol{e}_N$, $\sigma = \max|\boldsymbol{x}_s| + \tau$, and $k = 0$. |
| **Step 3** |
| Perform the QR decomposition $\boldsymbol{\Phi}^T = \boldsymbol{QR}$ and construct $\boldsymbol{V}_r$ using the last $N - M$ columns of $\boldsymbol{Q}$. |
| **Step 4** |
| With $\boldsymbol{w} = \boldsymbol{w}^{(k)}$ and using $\boldsymbol{\xi}^{(0)}$ as an initial point, apply the BFGS algorithm to solve the problem in (10), where re-weighting with parameter $\epsilon$ is applied using (11) in each iteration. Denote the solution as $\boldsymbol{\xi}^{(k)}$. |
| **Step 5** |
| Compute $\boldsymbol{x}^{(k)} = \boldsymbol{x}_s + \boldsymbol{V}_r\boldsymbol{\xi}^{(k)}$ and update weight vector to $\boldsymbol{w}^{(k+1)}$ using (11). |
| **Step 6** |
| If $\sigma \le \sigma_J$, stop and output $\boldsymbol{x}^{(k)}$ as solution; otherwise, set $\boldsymbol{\xi}^{(0)} = \boldsymbol{\xi}^{(k)}$, $\sigma = r \cdot \sigma$, $k = k + 1$, and repeat from Step 4. |

## IV. EXPERIMENTAL RESULTS

In the first experiment, the signal length and number of measurements were set to $N = 256$ and $M = 100$, respectively. A total of 15 sparse signals with sparsity $K = 5q - 4$, $q = 1, 2, \ldots, 15$ were used. A $K$-sparse signal $\boldsymbol{x}$ was constructed as follows: (1) set $\boldsymbol{x}$ to a zero vector of length $N$; (2) generate a vector $\boldsymbol{u}$ of length $K$ assuming that each component $u_i$ is a random value drawn from a normal distribution $\mathcal{N}(0,1)$; (3) randomly select $K$ indices from the set $\{1, 2, \ldots, N\}$, say $i_1, i_2, \ldots, i_K$, and set $x_{i_1} = u_1, x_{i_2} = u_2, \ldots, x_{i_K} = u_K$. The measurement matrix is of size $M \times N$ and was generated by drawing its elements from $\mathcal{N}(0,1)$, followed by a normalization step so that the $\ell_2$-norm of each column is unity. The measurement is obtained as $\boldsymbol{y} = \boldsymbol{\Phi x}$. The performance of the iteratively re-weighted (IR) algorithm [6] with $p = 0.1$ and $p = 0$, the SL0 algorithm [7], and the proposed NRAL0 algorithm with $\sigma_J = 10^{-4}$, $r = 1/3$, $\tau = 0.01$, and $\epsilon = 0.09$ was measured in terms of number of perfect reconstructions over 100 runs. The results obtained are plotted in Figure 1. It can be observed that the NRAL0 algorithm outperforms the IR algorithm. On comparing NRAL0 with the SL0 algorithm, the two algorithms are comparable for $K$ smaller than 40, but the NRAL0 algorithm performs better for $K$ larger than 40. The mathematical complexity of the four algorithms was measured in terms of the average CPU time over 100 runs for typical instances with $M = N/2$ and $K = \text{round}(M/2.5)$ where $N$ varies in the range between 128 and 512. The CPU time was measured on a PC laptop with a Intel T5750 2 GHz processor using MTLAB commands *tic* and *tac*, and the results are plotted in Figure 2. It is noted that the NRAL0 and SL0 algorithms are more efficient than the IR algorithm, and the complexity of the NRAL0 algorithm is slightly higher than that of the SL0 algorithm. The moderate increase in the mathematical complexity of the NRAL0 algorithm is primarily due to the fact that the objective function in (10) needs to be modified in each iteration using (11).

In the second experiment, the four algorithms were tested by using sparse signals with various values of $N$, $M$, and $K$ so as to examine the algorithms' performance for signals of different lengths, measurement numbers, and sparsity levels.

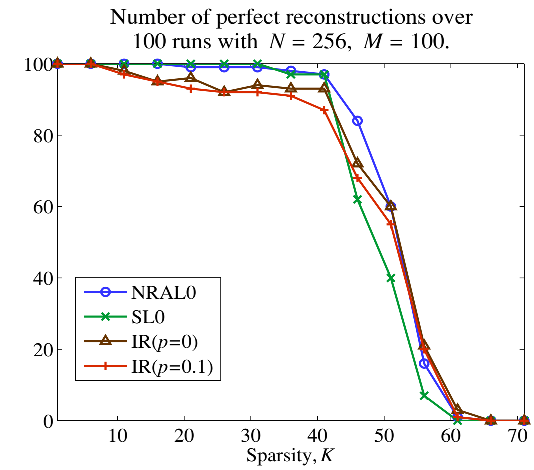


Fig. 1. Number of perfect reconstructions by the IR, SL0, and NRAL0 algorithms over 100 runs.

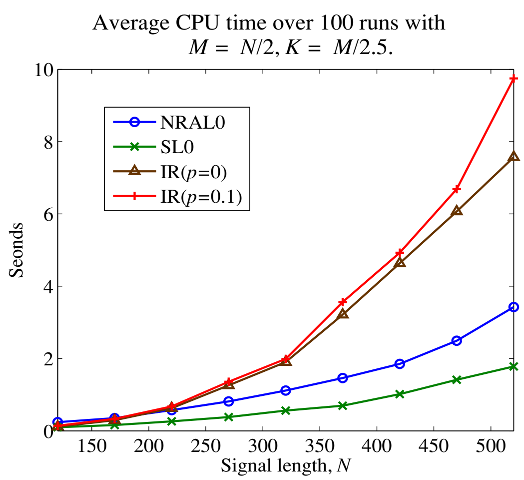


Fig. 2. Average CPU time required by the IR, SL0, and NRAL0 algorithms over 100 runs.

Specifically, the algorithms were tested with $N = 512$ and $M = 200$ using signals with sparsity $K = 70$, 90, and 110; and with $N = 1024$ and $M = 400$, using signals with sparsity $K = 140$, 180, and 220. The results obtained are summarized in Table II. It is observed that the performance of the NRAL0 algorithm is consistently better than those of the IR and SL0 algorithms in most cases.

## V. Conclusion

We have proposed an algorithm, called the null-space re-weighted approximate $\ell_0$-norm algorithm, for the reconstruction of sparse signals using random-projection type of measurements. The algorithm is based on minimizing an approximate $\ell_0$-norm of the signal in the null space of the measurement matrix where a re-weighting technique is used to force the solution's sparsity and a quasi-Newton algorithm is used to accelerate the optimization. Simulation results are presented which demonstrate that the proposed algorithm yields improved signal reconstruction performance and requires a reduced amount of computation relative to iteratively reweighted algorithms based on the $\ell_p$-norm with $p < 1$. When compared with a known algorithm based on a smoothed $\ell_0$-norm, improved signal reconstruction is achieved although the amount of computation is increased somewhat.

TABLE II
Number of Perfect Reconstructions of IR, SL0, and NRAL0 for Various Values of $N$, $M$, and $K$ over 100 Runs.

| $N/M$ | Algorithm | Number of perfect reconstructions | | |
|---|---|---|---|---|
| 512/200 | | $K$=70 | $K$=90 | $K$=110 |
| | IR($p$=0.1) | 77 | 77 | 24 |
| | IR($p$=0) | 85 | 67 | 21 |
| | SL0 | 100 | 91 | 8 |
| | NRAL0 | 100 | 96 | 28 |
| 1024/400 | | $K$=140 | $K$=180 | $K$=220 |
| | IR($p$=0.1) | 65 | 49 | 16 |
| | IR($p$=0) | 75 | 59 | 20 |
| | SL0 | 100 | 94 | 2 |
| | NRAL0 | 97 | 96 | 29 |

## Acknowledgment

The authors are grateful to the Natural Sciences and Engineering Research Council of Canada for supporting this research.

## References

[1] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489-509, Feb. 2006.

[2] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, pp. 1289-1306, Apr. 2006.

[3] E. Candès and T. Tao, "Near-optimal signal recovery from random projections: universal encoding strategies," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406-5425, Dec. 2006.

[4] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Scientif. Comput.*, vol. 20, no. 1, pp. 33-61, 1999.

[5] R. Chartrand, "Exact reconstruction of sparse signals via nonconvex minimization," *IEEE Signal Process. Lett.*, vol. 14, issue 10, pp. 770-710, Oct. 2007.

[6] R. Chartrand and W. Yin, "Iteratively reweighted algorithms for compressive sensing," in Proc. *IEEE Inter. Conf. Acoustics, Speech, Signal Process.*, pp. 3869-3872, 2008.

[7] H. Mohimani, M. Babie-Zadeh, and C. Jutten, "A fast approach for overcomplete sparse decomposition based on smoothed $\ell_0$-norm," *IEEE Trans. Signal Process.*, vol. 57, no. 1, pp. 289-301, Jan. 2009.

[8] A. Antoniou and W.-S. Lu, *Practical Optimization: Algorithms and Engineering Applications,* Springer, 2006.

[9] S. Mallat, *A Wavelet Tour of Signal Processing,* Elsevier, 2nd ed., 1998.

[10] G. H. Golub and C. F. V. Loan, *Matrix Computations,* The Johns Hopkins University Press, 3rd ed., 1996.

[11] E. J. Candès, M. B. Walkin, and S. P. Boyd, "Enhancing sparsity by reweighted $\ell_1$ minimization," *J. Fourier Anal. Appl.*, vol. 14, no. 5, pp. 877-905, special issue on sparisty, Dec. 2008.