

SPARSE LMS FOR SYSTEM IDENTIFICATION

Yilun Chen^{1*}, Yuantao Gu², Alfred O. Hero III¹

¹Department of EECS, University of Michigan, Ann Arbor, MI 48109-2122, USA

²Department of EE, Tsinghua University, Beijing 100084, P.R. China
{yilun,hero}@umich.edu, gyt@tsinghua.edu.cn

ABSTRACT

We propose a new approach to adaptive system identification when the system model is sparse. The approach applies the ℓ_1 relaxation, common in compressive sensing, to improve the performance of LMS-type adaptive methods. This results in two new algorithms, the Zero-Attracting LMS (ZA-LMS) and the Reweighted Zero-Attracting LMS (RZA-LMS). The ZA-LMS is derived via combining a ℓ_1 norm penalty on the coefficients into the quadratic LMS cost function, which generates a zero attractor in the LMS iteration. The zero attractor promotes sparsity in taps during the filtering process, and therefore accelerates convergence when identifying sparse systems. We prove that the ZA-LMS can achieve lower mean square error than the standard LMS. To further improve the filtering performance, the RZA-LMS is developed using a reweighted zero attractor. The performance of the RZA-LMS is superior to that of the ZA-LMS numerically. Experiments demonstrate the advantages of the proposed filters in both convergence rate and steady-state behaviors under sparsity assumptions on the true coefficient vector. The RZA-LMS is also shown to be robust when the number of non-zero taps increases.

Index Terms— LMS, compressive sensing, sparse models, zero-attracting, ℓ_1 norm relaxation

1. INTRODUCTION

The Least Mean Square (LMS) algorithm, introduced by Widrow and Hoff [1], is a popular method for adaptive system identification. Its applications include echo cancellation, channel equalization, interference cancellation and so forth. In many scenarios impulse responses of unknown systems can be assumed to be sparse, containing only a few large coefficients interspersed among many negligible ones. Using such sparse prior information can improve the filtering performance. However, standard LMS filters do not exploit such information. In the past years, many algorithms exploiting sparsity were based on applying a subset selection scheme during the filtering process, which was implemented via statistical detection of active taps [2, 3, 4] or sequential partial updating [5, 6]. Other variants assign proportional step sizes of different taps according to their magnitudes, such as the Proportionate Normalized LMS (PNLMS) and its variations [7, 8].

Motivated by LASSO [9] and recent progress in compressive sensing [10, 11, 12], we propose an alternative approach to identifying sparse systems using LMS filters. The basic idea is to introduce a penalty which favors sparsity in the cost function. We first incorporate a ℓ_1 norm penalty on the coefficients into the quadratic cost

function of the standard LMS. This results in a modified LMS update with a zero attractor for all the taps, naming the Zero-Attracting LMS (ZA-LMS). We analytically demonstrate that the ZA-LMS achieves better steady-state performance than that of the standard LMS for sparse models. To further improve the filtering performance, the Reweighted Zero-Attracting LMS (RZA-LMS) is proposed which employs reweighted step sizes of the zero attractor for different taps, inducing the attractor to selectively promote zero taps rather than uniformly promote zeros on all the taps. Experimental results illustrate that the proposed filters exceed the standard LMS in both transient and steady-state performance for sparse systems; and the RZA-LMS outperforms the ZA-LMS numerically. Furthermore, the RZA-LMS shows robustness when the number of non-zero taps increases, with little loss in performance with respect to the standard LMS in non-sparse situations.

The paper is organized as follows. Section 2 develops the ZA-LMS and RZA-LMS algorithms for sparse systems. In section 3, numerical simulation results are provided. Finally, we conclude the paper and discuss possible future directions in section 4.

Notations: In the following parts of paper, matrices and vectors are denoted by boldface upper case letters and boldface lower case letters, respectively; the superscripts $(\cdot)^T$ and $(\cdot)^{-1}$ denote the transpose and inverse operators, respectively; the operator $\|\cdot\|_1$ denotes the ℓ_1 norm; $\text{tr}(\cdot)$ denotes the trace operator; and $E[\cdot]$ denotes the expectation operator.

2. ALGORITHMS

2.1. Review of the standard LMS

Let $y(n)$ be a sample of an observation of output signal

$$y(n) = \mathbf{w}^T \mathbf{x}(n) + v(n), \quad (1)$$

where $\mathbf{w} = [w_0, w_1, \dots, w_{N-1}]^T$ is the filter coefficient vector, *e.g.*, a FIR channel impulse response; $\mathbf{x}(n) = [x(n), x(n-1), \dots, x(n-N+1)]^T$ denotes the vector of input signal $x(n)$; and $v(n)$ is the observation noise assumed to be independent with $x(n)$.

The goal of LMS-type filters is to sequentially estimate the unknown coefficient vector using the input signal $x(n)$ and the desired output $y(n)$. Let $\mathbf{w}(n)$ be the estimated coefficient vector of the adaptive filter at iteration n . In the standard LMS, the cost function $L(n)$ is defined as

$$L(n) = \frac{1}{2} e^2(n), \quad (2)$$

where $e(n)$ is the instantaneous error:

$$e(n) = y(n) - \mathbf{w}^T(n) \mathbf{x}(n). \quad (3)$$

*This research was partially supported by AFOSR grant FA9550-06-1-0324.

The filter coefficient vector is then updated by

$$\mathbf{w}(n+1) = \mathbf{w}(n) - \mu \frac{\partial L(n)}{\partial \mathbf{w}(n)} = \mathbf{w}(n) + \mu e(n) \mathbf{x}(n), \quad (4)$$

where μ is the step size controlling convergence and the steady-state behavior of the LMS algorithm. Denote \mathbf{R} as the covariance matrix of the input vector $\mathbf{x}(n)$ and λ_{\max} as its maximum eigenvalue. The well-known convergence condition for the LMS is

$$0 < \mu < \frac{1}{\lambda_{\max}}. \quad (5)$$

Under the independence assumption, the steady-state excess MSE is

$$P_{ex}(\infty) = \lim_{n \rightarrow \infty} E \left[\left((\mathbf{w}(n) - \mathbf{w})^T \mathbf{x}(n) \right)^2 \right] = \frac{\eta}{2 - \eta} P_0, \quad (6)$$

where P_0 is the power of observation noise

$$P_0 = E[v^2(n)], \quad (7)$$

and

$$\eta = \text{tr}(\mathbf{R}(\mathbf{I} - \mu\mathbf{R})^{-1}). \quad (8)$$

2.2. The Zero-Attracting LMS algorithm (ZA-LMS)

In the ZA-LMS, a new cost function $L_1(n)$ is defined by combining the instantaneous square error with the ℓ_1 norm penalty of the coefficient vector

$$L_1(n) = \frac{1}{2} e^2(n) + \gamma \|\mathbf{w}(n)\|_1. \quad (9)$$

Using the gradient descent updating, the ZA-LMS filter update is defined as

$$\begin{aligned} \mathbf{w}(n+1) &= \mathbf{w}(n) - \mu \frac{\partial L_1(n)}{\partial \mathbf{w}(n)} \\ &= \mathbf{w}(n) - \rho \text{sgn} \mathbf{w}(n) + \mu e(n) \mathbf{x}(n), \end{aligned} \quad (10)$$

where $\rho = \mu\gamma$ and $\text{sgn}(\cdot)$ is a component-wise sign function defined as

$$\text{sgn}(x) = \begin{cases} x/|x| & x \neq 0 \\ 0 & x = 0 \end{cases}. \quad (11)$$

Comparing the ZA-LMS update (10) to the standard LMS update (4), the ZA-LMS has an additional term $-\rho \text{sgn} \mathbf{w}(n)$ which always attracts the tap coefficients to zero. We call this the zero attractor, whose strength is controlled by ρ . Intuitively, the zero attractor will speed-up convergence when the majority of coefficients of \mathbf{w} are zero, *i.e.*, the system is sparse. The convergence condition of the ZA-LMS is provided in the following theorem.

Theorem 1. *The mean coefficient vector $E[\mathbf{w}(n)]$ converges as $n \rightarrow \infty$ if μ satisfies (5), and the converged vector is*

$$E[\mathbf{w}(\infty)] = \mathbf{w} - \frac{\rho}{\mu} \mathbf{R}^{-1} E[\text{sgn} \mathbf{w}(\infty)]. \quad (12)$$

Proof. Denote $\tilde{\mathbf{w}}(n) = \mathbf{w}(n) - \mathbf{w}$, (4) is equivalent to

$$\tilde{\mathbf{w}}(n) = \left(\mathbf{I} - \mu \mathbf{x}(n) \mathbf{x}(n)^T \right) \tilde{\mathbf{w}}(n-1) - \rho \text{sgn} \mathbf{w}(n-1) + \mu v(n) \mathbf{x}(n). \quad (13)$$

Taking expectations on both sides of (13), there is

$$E[\tilde{\mathbf{w}}(n)] = (\mathbf{I} - \mu\mathbf{R}) E[\tilde{\mathbf{w}}(n-1)] - \rho E[\text{sgn} \mathbf{w}(n-1)]. \quad (14)$$

Note that the vector $\rho E[\text{sgn} \mathbf{w}(n-1)]$ is bounded between $-\rho\mathbf{1}$ and $\rho\mathbf{1}$. Therefore, $E[\tilde{\mathbf{w}}(n)]$ converges if the maximal eigenvalue of $(\mathbf{I} - \mu\mathbf{R})$ is less than 1, which is satisfied by (5). Since $E[\mathbf{w}(n)] = E[\tilde{\mathbf{w}}(n)] + \mathbf{w}$, $E[\mathbf{w}(n)]$ also converges with the limiting vector shown in (12). \square

One can see that the convergence condition of the ZA-LMS and the standard LMS is the same, which is independent with ρ . (12) implies the ZA-LMS filter returns a biased estimate of the true coefficient vector. However, we show that with appropriate ρ the ZA-LMS is able to yield lower MSE than the standard LMS for truly sparse systems.

Theorem 2. *Let NZ denote the index set of non-zero taps, *i.e.*, $w_i \neq 0$ for $i \in NZ$. Assuming ρ is sufficiently small so that for every $i \in NZ$*

$$E[\text{sgn} w_i(\infty)] = \text{sgn} w_i, \quad (15)$$

the excess MSE of the ZA-LMS filter is

$$P_{ex}(\infty) = \frac{\eta}{2 - \eta} P_0 + \frac{\alpha_1}{(2 - \eta)\mu} \rho \left(\rho - \frac{2\alpha_2}{\alpha_1} \right), \quad (16)$$

where

$$\alpha_1 = E \left[\text{sgn} \mathbf{w}(\infty)^T (\mathbf{I} - \mu\mathbf{R})^{-1} \text{sgn} \mathbf{w}(\infty) \right], \quad (17)$$

and

$$\alpha_2 = E[\|\mathbf{w}(\infty)\|_1] - \|\mathbf{w}\|_1. \quad (18)$$

P_0 and η are defined in (7) and (8), respectively.

The proof of Theorem 2 is similar to the derivation of (6) in [1] and is omitted for the lack of space. It is easy to see that α_1 is always positive, and the range of α_1 is

$$0 < \alpha_1 \leq \frac{N}{1 - \mu\lambda_{\max}}. \quad (19)$$

Note that the first term in the RHS of (16) is the excess MSE of the standard LMS filter. Therefore, when $\alpha_2 > 0$, we can expect lower MSE than the standard LMS, *i.e.*, when ρ is selected between 0 and $2\alpha_2/\alpha_1$,

$$P_{ex}(\infty) < \frac{\eta}{2 - \eta} P_0.$$

To further specify α_2 , we have the following result.

Lemma 1. *Let Z and NZ be the index sets of zero taps and non-zero taps respectively. If $\mathbf{w}(n)$ is assumed to be Gaussian distributed, a first order approximation of α_2 is given by*

$$\alpha_2 \simeq \sum_{i \in Z} \sqrt{\frac{2}{\pi} \Phi_{ii}(\infty)} - \frac{\rho}{\mu} \sum_{i \in NZ} |b_i|, \quad (20)$$

where $\Phi_{ii}(\infty)$ and b_i are the i -th element of the diagonal of $\Phi(\infty)$ and \mathbf{b} respectively, defined as

$$\Phi(\infty) = E[(\mathbf{w}(\infty) - \mathbf{w})(\mathbf{w}(\infty) - \mathbf{w})^T]$$

and

$$\mathbf{b} = \mathbf{R}^{-1} E[\text{sgn} \mathbf{w}(\infty)].$$

Proof.

$$\alpha_2 = \sum_{i \in Z} E[|w_i(\infty)|] + \sum_{i \in NZ} (E[|w_i(\infty)|] - |w_i|). \quad (21)$$

We then use the following fact: Assuming a random variable z is Gaussian distributed with mean ν and variance σ^2 ,

$$E[|z|] = \sqrt{\frac{2}{\pi}} E[z^2] + o\left(\frac{|\nu|}{\sigma}\right) \quad \text{when } |\nu| \ll \sigma, \quad (22)$$

and

$$E[|z|] = |\nu| + o\left(\frac{\sigma}{|\nu|}\right) \quad \text{when } |\nu| \gg \sigma. \quad (23)$$

Since $w_i(\infty)$ is a Gaussian random variable, with (12), (22) and (23) the first order approximation of α_2 is

$$\begin{aligned} \alpha_2 &\simeq \sum_{i \in Z} \sqrt{\frac{2}{\pi}} E[w_i(\infty)^2] - \sum_{i \in NZ} |E[w_i(\infty)] - w_i| \\ &= \sum_{i \in Z} \sqrt{\frac{2}{\pi}} \Phi_{ii}(\infty) - \frac{\rho}{\mu} \sum_{i \in NZ} |b_i|. \end{aligned} \quad (24)$$

□

There are two competitive terms on the RHS of (20). The first one varies about zero for the taps associated with zero coefficients of \mathbf{w} . The second term is a bias which is due to the shrinkage of the taps associated with non-zero coefficients of \mathbf{w} . When the zero taps take the majority, the first term dominates the second one and positive α_2 can be therefore obtained.

2.3. The Reweighted Zero-Attracting LMS (RZA-LMS)

Large α_2 is crucial for the ZA-LMS since it results in a greater performance gain and larger margin for choosing ρ . However, the bias term in (20) reduces α_2 and then limits the MSE performance. This behavior comes from the fact that the shrinkage in the ZA-LMS does not distinguish between zero taps and non-zero taps. Since all the taps are forced to zero uniformly, its performance would deteriorate for less sparse systems. Motivated by reweighting in compressive sampling [11], we propose a heuristic approach to refine the zero attractor called the Reweighted Zero-Attracting LMS (RZA-LMS).

The RZA-LMS is derived via the new cost function

$$L_2(n) = \frac{1}{2} e^2(n) + \gamma' \sum_{i=1}^N \log(1 + |w_i|/\varepsilon'). \quad (25)$$

The log-sum penalty $\sum_{i=1}^N \log(1 + |w_i|/\varepsilon')$ has been introduced as it behaves more similarly to the ℓ_0 norm than $\|\mathbf{w}\|_1$. The coefficient vector is then updated by

$$w_i(n+1) = w_i(n) - \rho \frac{\text{sgn } w_i(n)}{1 + \varepsilon|w_i(n)|} + \mu e(n)x_i(n), \quad (26)$$

or equivalently, in vector form

$$\mathbf{w}(n+1) = \mathbf{w}(n) - \rho \frac{\text{sgn}\{\mathbf{w}(n)\}}{1 + \varepsilon|\mathbf{w}(n)|} + \mu e(n)\mathbf{x}(n), \quad (27)$$

where $\rho = \mu\gamma'/\varepsilon'$ and $\varepsilon = 1/\varepsilon'$.

The RZA-LMS selectively shrinks taps with large magnitudes and the ones with small magnitudes. The reweighted zero attractor takes effect only on those taps whose magnitudes are comparable to $1/\varepsilon$; and there is little shrinkage exerted on the taps whose $|w_i(n)| \gg 1/\varepsilon$. In this way, the bias of the RZA-LMS can be reduced.

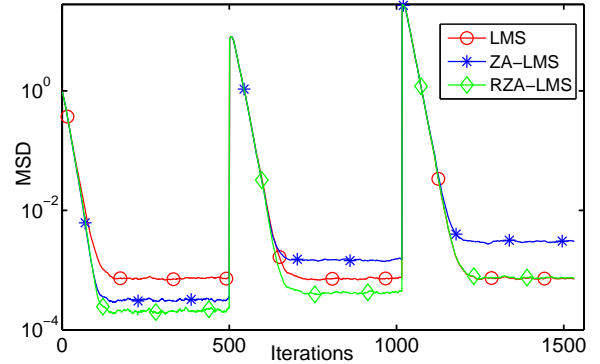


Fig. 1. Tracking and steady-state behaviors of 16-order adaptive filters, driven by white input signal.

3. EXPERIMENTAL RESULTS

In this section, the performance of the ZA-LMS (10) and the RZA-LMS (27) are compared with that of the standard LMS filter. Three experiments are designed to demonstrate their tracking and steady-state performances.

In the first experiment, there are 16 coefficients in the time varying system. Initially, we set the 5th tap with value 1 and others to zero, making the system have a sparsity of 1/16. After 500 iterations, all the odd taps are set to 1 while all the even taps remains to be zero, *i.e.*, a sparsity of 8/16. After 1000 iterations all the even taps are set with value -1 while all the odd taps are maintained to be 1, leaving a completely non-sparse system. The input signal and the observed noise are white Gaussian random sequences with variance of 1 and 10^{-3} , respectively. The three filters (LMS, ZA-LMS and RZA-LMS) are run 200 times. The parameters are set as $\mu = 0.05$, $\rho = 5 \times 10^{-4}$ and $\varepsilon = 10$. Note that we use the same μ and ρ for the three filters. The average estimate of mean square deviation (MSD) is shown in Fig. 1. As we can see from the MSD results, when the system is very sparse (before the 500th iteration), both the ZA-LMS and the RZA-LMS yield faster convergence rate and better steady-state performances than the standard LMS. And the RZA-LMS achieves lower MSD than the ZA-LMS. After the 500th iteration, as the number of non-zero taps increases to eight, the performance of the ZA-LMS deteriorates while the RZA-LMS maintains the best performance among the three filters. After 1000 iterations, the RZA-LMS still performs comparably with the standard LMS even though the system is now completely non-sparse.

The system in the second experiment is the same as the first one, except the switching times are set to the 7000th iteration and the 14000th iteration, respectively. The input signal $x(n)$ is now a correlated signal generated by $x(n) = 0.8x(n-1) + u(n)$ and then normalized to variance 1, where $u(n)$ is a white Gaussian noise. The variance of the observed noise is set to 10^{-3} . The filter parameters are set as $\mu = 0.015$, $\rho = 3 \times 10^{-5}$ and $\varepsilon = 10$. Fig. 2 shows the MSD of the three filters, and similar performance trends are observed as in experiment 1. Observe that at the beginning of the iterations (*e.g.*, from iteration 7000 to 8500), all the three filters converge at a nearly the same rate. After the 8500th iteration, the convergence of the RZA-LMS accelerates due to its selective shrinkage.

The third experiment simulates 256-tap system with 28 non-zero coefficients. The impulse response is shown in Fig. 3. The driving

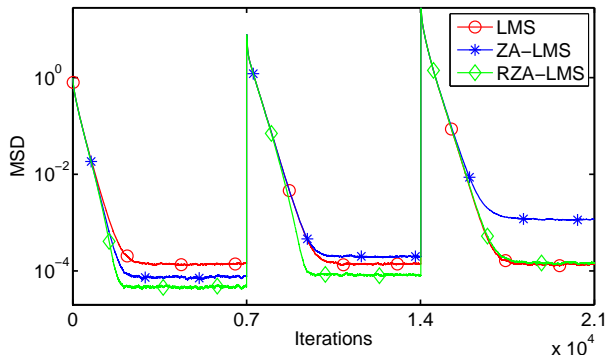


Fig. 2. Tracking and steady-state behaviors of 16-order adaptive filters, driven by correlated input signal.

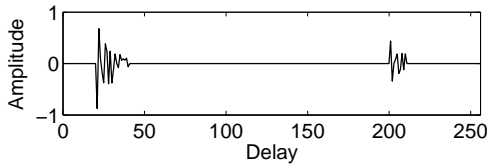


Fig. 3. The impulse response of the system in experiment 3.

signal and observed noise are the same as the first experiment. μ is set to be 5×10^{-3} in the three filters and ε is set to 10. This time we select different values of ρ for the ZA-LMS and the RZA-LMS to yield the best MSE, where ρ is 2.5×10^{-6} for the ZA-LMS and 10^{-5} for the RZA-LMS. The simulations are performed 200 times and the averaged excess MSE is shown in Fig. 4. One sees that for this long sparse system the zero-attracting algorithms continue to outperform the standard LMS as measured by faster convergence rate and lower steady-state MSE.

4. CONCLUSION

In this paper, two novel adaptive filters are proposed for sparse system identification. The ZA-LMS incorporates a ℓ_1 norm penalty of the coefficients into its cost function, which resulted in a shrinkage in the update formula. This shrinkage accelerates the convergence rate when the majority of coefficients are zero. A theorem was given showing that reduced MSE can be obtained by the ZA-LMS. The RZA-LMS was proposed to further improve the filtering performance, where a reweighted zero attractor is devised to perform selective coefficient shrinkage. With the same parameters, the RZA-LMS is superior to the ZA-LMS in both convergence rates and steady-state behaviors. Experiments demonstrate that the ZA-LMS and the RZA-LMS improve on the standard LMS in both transient and steady-state performance when the system is sparse. Furthermore, the RZA-LMS performs robustly under non-sparse systems.

Our future work will include how to choose the parameters of zero-attracting algorithms in a more systematic way. The zero attractor can also be implemented in the NLMS filter for identifying sparse systems. Furthermore, the method of ℓ_1 norm penalization can be further extended to other types of adaptive filters, such as the RLS filter and the adaptive Kalman filter.

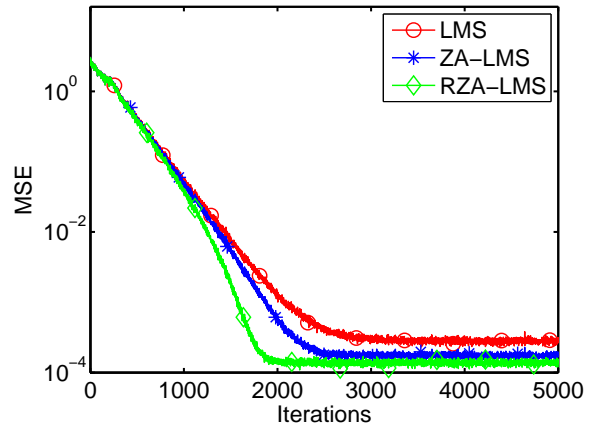


Fig. 4. Tracking and steady-state behaviors of 256-order adaptive filters, driven by white input signal.

5. REFERENCES

- [1] B. Widrow and S.D. Stearns, *Adaptive Signal Processing*, New Jersey: Prentice Hall, 1985.
- [2] S. Kawamura and M. Hatori, "A tap selection algorithm for adaptive filters," in *Proceedings of ICASSP*, 1986, vol. 11, pp. 2979–2982.
- [3] J. Homer, I. Mareels, R.R. Bitmead, B. Wahlberg, and A. Gustafsson, "LMS estimation via structural detection," *IEEE Trans. on Signal Processing*, vol. 46, pp. 2651–2663, October 1998.
- [4] Y. Li, Y. Gu, and K. Tang, "Parallel NLMS filters with stochastic active taps and step-sizes for sparse system identification," in *Proceedings of ICASSP*, 2006, vol. 3, pp. 109–112.
- [5] D.M. Etter, "Identification of sparse impulse response systems using an adaptive delay filter," in *Proceedings of ICASSP*, 1985, pp. 1169–1172.
- [6] M. Godavarti and A. O. Hero, "Partial update LMS algorithms," *IEEE Trans. on Signal Processing*, vol. 53, pp. 2382–2399, 2005.
- [7] S.L. Gay, "An efficient, fast converging adaptive filter for network echocancellation," in *Proceedings of Asilomar*, 1998, vol. 1, pp. 394–398.
- [8] D.L. Duttweiler, "Proportionate normalized least-squares adaptation in echo cancelers," *IEEE Trans. on Speech and Audio Processing*, vol. 8, pp. 508–518, 2000.
- [9] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Royal. Statist. Soc B.*, vol. 58, pp. 267–288, 1996.
- [10] E. Candès, "Compressive sampling," *Int. Congress of Mathematics*, vol. 3, pp. 1433–1452, 2006.
- [11] E. J. Candès, M. Wakin, and S. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *To appear in J. Fourier Anal. Appl.*
- [12] R. Baraniuk, "Compressive sensing," *IEEE Signal Processing Magazine*, vol. 25, pp. 21–30, March 2007.