SPARSE CHANNEL ESTIMATION WITH L_P-NORM AND REWEIGHTED L_1-NORM PENALIZED LEAST MEAN SQUARES

Omid Taheri *Sergiy A. Vorobyov*

Dept. Electrical and Computer Engineering, University of Alberta
Edmonton, AB, T6G 2V4, Canada
Emails: otaheri, vorobyov@ece.ualberta.ca

ABSTRACT

The least mean squares (LMS) algorithm is one of the most popular recursive parameter estimation methods. In its standard form it does not take into account any special characteristics that the parameterized model may have. Assuming that such model is sparse in some domain (for example, it has sparse impulse or frequency response), we aim at developing such LMS algorithms that can adapt to the underlying sparsity and achieve better parameter estimates. Particularly, the example of channel estimation with sparse channel impulse response is considered. The proposed modifications of LMS are the l_p-norm and reweighted l_1-norm penalized LMS algorithms. Our simulation results confirm the superiority of the proposed algorithms over the standard LMS as well as other sparsity-aware modifications of LMS available in the literature.

***Index Terms*—** Compressed sensing, gradient descent, least mean squares, sparsity

1. INTRODUCTION

Depending on the available information about the system and its statistics several recursive parameter estimation methods have been developed in signal processing literature [1]. Among such methods are the least mean squares (LMS) algorithm [2], recursive least squares (RLS) algorithm, Kalman filter and their different modifications. The aforementioned standard recursive parameter estimation methods do not utilize the information about sparsity of the system characteristics if such characteristics are indeed sparse, and thus, may perform poorly [3]. Among the algorithms which exploit the sparsity of a system, the methods based on sequential partial updating [3] and statistical detection of active taps [4] were historically developed first. In parallel, the methods based on the idea of adding a penalty term in the cost function, which enforces the sparse solution, have been developed in geophysics [5].

Supported in part by the Natural Science and Engineering Research Council (NSERC) of Canada and the Alberta Innovates – Technology Futures, Alberta, Canada.

As well as the sparsity-aware estimation method of [5], the recent theory of compressed sensing (CS) deals with the recovery of sparse signals from undersampled and incomplete measurements [6], [7]. Note that a signal is sparse if only a small percentage of its coefficients in a known transform domain is nonzero or significantly different from zero. The CS theory has already been used in a variety of applications such as sensor networks [8], cognitive radios [9], etc. In these applications the underlying problem is the estimation of sparse signal or system parameters. Some efforts have been made for incorporating the information about sparsity of a signal or a system into the estimation problem in order to design more accurate or less complex estimation algorithms. For example, a CS-based Kalman filter algorithm has been developed in [10], where the author considered a time varying sparsity pattern for the signal. The initial support is estimated using CS, and as long as the support does not change, the Kalman filter restricted on the support set is used. A change in the support set leads to an increase in filtering error. In this case, CS is used to estimate the change in the support set. A time- and norm-weighted least-absolute shrinkage and selection operator (Lasso) scheme with l_1-norm weights obtained from the recursive least-squares (RLS) algorithm has been developed in [11] and its consistency has been established. Moreover, a variation of the LMS algorithm with l_1-norm penalty term in the standard LMS cost function has been developed in [12]. It has been also shown that such sparsity-aware LMS algorithm achieves a better performance than the standard LMS algorithm.

The focus of this paper is on LMS, and the main objective is developing and testing the sparsity-aware modifications of LMS, which can improve over the performance of the method of [12]. Particularly, we consider the sparse channel estimation problem and aim at estimating the channel impulse response (CIR) by exploiting the knowledge that CIR is sparse. For example, a time sparse finite impulse response (FIR) channel is the one for which the vector of channel coefficients has only a small number of nonzero taps. We develop the sparsity-aware modifications of LMS by considering the l_p-norm and reweighted l_1-norm penalty terms. For the for-

mer penalty term, the overall cost function becomes nonconvex and the theoretical convergence and consistency results become problematic. However, for the latter penalty term, the corresponding cost function is convex and such results can be developed. We test the proposed sparsity-aware modifications of LMS by simulations and show their superiority to the other LMS-based algorithms for estimating the parameters of sparse systems.

2. SYSTEM MODEL

Consider a communication system shown in Fig. 1. The data sequence $\boldsymbol{x}(n)$ is sent over the FIR channel with CIR $\boldsymbol{h}(n) = (h(n), h(n-1), \ldots, h(n-N+1))^T$ where N is the size of the channel memory and $(\cdot)^T$ denotes the transposition. It is assumed that the CIR is real-valued. In Fig. 1, $v(n)$ denotes the noise at the receiver end, $\hat{\boldsymbol{h}}(n)$ stands for the CIR estimate, $e(n) = d(n) - \hat{\boldsymbol{h}}^T(n)\boldsymbol{x}(n)$ is the instantaneous error with $d(n) = y(n) + v(n)$, and $y(n)$ and $\hat{y}(n)$ denote the system output and its estimate, respectively.

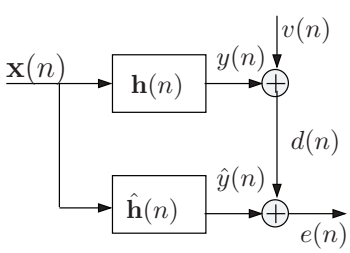

Fig. 1. The communication system.

3. LMS AND SPARSITY-AWARE LMS

The cost function for the standard LMS is given as $L(n) = (1/2)e^2(n)$. The gradient descent method can be used to find the minimum of the cost function $L(n)$. Then the update equation can be written as

$$\hat{\boldsymbol{h}}(n+1) = \hat{\boldsymbol{h}}(n) - \mu\frac{\partial L(n)}{\partial \hat{\boldsymbol{h}}(n)} = \hat{\boldsymbol{h}}(n) + \mu e(n)\boldsymbol{x}(n) \quad (1)$$

where μ is the step size such that $0 < \mu < \lambda_{\max}^{-1}$ with $\lambda_{\max}$ being the maximum eigenvalue of the covariance matrix of $\boldsymbol{x}(n)$. Since the cost function is convex, the gradient descent algorithm is guaranteed to converge to the optimum point under the aforementioned condition on μ.

Assuming that the CIR is sparse, i.e., most of the coefficients in the vector $\boldsymbol{h}(n)$ are zeros or insignificant in value, the following sparsity-aware modifications of LMS have been developed [12]. In order to penalize the non-sparse solutions, the l_1-norm of $\hat{\boldsymbol{h}}(n)$ can be added in the standard LMS cost function so that the new cost function becomes $L_{ZA}(n) = (1/2)e^2(n) + \gamma_{ZA}\|\hat{\boldsymbol{h}}(n)\|_{l_1}$, where $\|\cdot\|_{l_1}$ denotes the l_1-norm of a vector and γ_{ZA} is the weight assigned to the penalty term. Note that this cost function is convex, and therefore, it is guaranteed that the gradient descent method converges under some conditions. The corresponding algorithm is called the zero attracting LMS (ZA-LMS) and its update equation is

$$\hat{\boldsymbol{h}}(n+1) = \hat{\boldsymbol{h}}(n) + \mu e(n)\boldsymbol{x}(n) - \rho_{ZA}\text{sgn}(\hat{\boldsymbol{h}}(n)) \quad (2)$$

where $\rho_{ZA} = \mu\gamma_{ZA}$ and $\text{sgn}(\cdot)$ is the sign function which operates on every component of the vector separately and it is zero for $x = 0$, 1 for $x > 0$, and -1 for $x < 0$.

The other way to penalize the non-sparse solutions is to consider the exact measure of sparsity, that is, the l_0-norm. Since the complexity associated with the use of the l_0-norm is high, a logarithmic penalty that resembles the l_0-norm can be considered and the the cost function becomes

$$L_{RZA}(n) = (1/2)e^2(n) + \gamma_{RZA}\sum_{i=1}^{N}\log\left(1 + \frac{[\hat{\boldsymbol{h}}(n)]_i}{\epsilon'_{RZA}}\right) \quad (3)$$

where $[\hat{\boldsymbol{h}}(n)]_i$ is the i-th element of the vector $\hat{\boldsymbol{h}}(n)$ and γ_{RZA} and ϵ'_{RZA} are some positive numbers. Note that the same penalty term is also used, for example, in [5]. Since the logarithmic penalty in (3) resembles the l_0-norm better than the l_1-norm penalty in ZA-LMS method, one can expect that the corresponding algorithm called in [12] as the reweighted zero attracting LMS (RZA-LMS) will exhibit a better performance than the ZA-LMS. The update equation for the RZA-LMS is

$$\hat{\boldsymbol{h}}(n+1) = \hat{\boldsymbol{h}}(n) + \mu e(n)\boldsymbol{x}(n) - \rho_{RZA}\frac{\text{sgn}(\hat{\boldsymbol{h}}(n))}{1 + \epsilon_{RZA}|\hat{\boldsymbol{h}}(n)|} \quad (4)$$

where $\rho_{RZA} = \mu\gamma_{RZA}\epsilon_{RZA}$, $\epsilon_{RZA} = 1/\epsilon'_{RZA}$, and $|\cdot|$ is the component-wise absolute value operator. However, the cost function (3) is not convex and the convergence and consistency analysis is problematic for (4).

4. NEW SPARSITY-AWARE LMS ALGORITHMS

We develop here two other sparsity-aware modifications of LMS based on the idea of introducing a penalty term which forces the solution to be sparse.

l_p-norm ($0 < p < 1$) penalty: Motivated by the fact that the RZA-LMS shows a better performance than the ZA-LMS [12] because the logarithmic penalty term of the RZA-LMS is closer to the l_0-norm penalty, we consider another function of $\hat{\boldsymbol{h}}(n)$ that is more similar to the l_0-norm. Such function is the l_p-norm of $\hat{\boldsymbol{h}}(n)$ with $0 < p < 1$. The smaller the value of p is the more the l_p-norm resembles the l_0-norm. In this case, the cost function for the l_p-norm penalized method becomes

$$L_{l_p}(n) = (1/2)e^2(n) + \gamma_p\|\hat{\boldsymbol{h}}(n)\|_{l_p} \quad (5)$$

where $\|\cdot\|_{l_p}$ stands for the l_p-norm of a vector and γ_p is the corresponding weight term. Similar to the cost function of

the RZA-LMS, the cost function (5) is not convex and the analysis of the global convergence and consistency of the corresponding algorithm is problematic. However, as it will be seen in the next section, the method based on (5) shows better performance than the RZA-LMS which faces the same problems. Using gradient descent, the update equation based on (5) can be derived as

$$\begin{aligned}\hat{\boldsymbol{h}}(n+1) &= \hat{\boldsymbol{h}}(n) + \mu e(n)\boldsymbol{x}(n) \\ &- \rho_p \frac{\left(\|\hat{\boldsymbol{h}}(n)\|_{l_p}\right)^{1-p} \operatorname{sgn}(\hat{\boldsymbol{h}}(n))}{|\hat{\boldsymbol{h}}(n)|^{(1-p)}} \end{aligned} \quad (6)$$

where $\rho_p = \mu\gamma_p$. Practically, we need to impose an upper bound on the last term in (6) in the situation when an entry of $\hat{\boldsymbol{h}}(n)$ approaches zero, which is the case for a sparse CIR. Then the update equation (6) is modified as

$$\begin{aligned}\hat{\boldsymbol{h}}(n+1) &= \hat{\boldsymbol{h}}(n) + \mu e(n)\boldsymbol{x}(n) \\ &- \rho_p \frac{\left(\|\hat{\boldsymbol{h}}(n)\|_{l_p}\right)^{1-p} \operatorname{sgn}(\hat{\boldsymbol{h}}(n))}{\epsilon_p + |\hat{\boldsymbol{h}}(n)|^{(1-p)}} \end{aligned} \quad (7)$$

where ϵ_p is a value which bounds the last term in (6).

Reweighted l_1-norm penalty: Another way to enforce the sparsity of the solution is to use the reweighted l_1-norm penalty term in the cost function. Note that the reweighted l_1-norm minimization for recovering sparse signals has been recently used in, for example, [13]. This method provides better performance results than the standard l_1-norm minimization. Our reweighted l_1-norm penalized LMS considers a penalty term proportional to the reweighted l_1-norm of the coefficient vector. The corresponding cost function can be written as

$$L_{rl_1}(n) = (1/2)e^2(n) + \gamma_r \|\boldsymbol{w}(n)\hat{\boldsymbol{h}}(n)\|_{l_1} \quad (8)$$

where γ_r is the weight associated with the penalty term and elements of $\boldsymbol{w}(n)$ are set to

$$[\boldsymbol{w}(n)]_i = \frac{1}{\epsilon_r + |[\hat{\boldsymbol{h}}(n-1)]_i|}, \quad i = 1, \ldots, N \quad (9)$$

with ϵ_r being some positive number. The update equation can also be derived from the gradient descent algorithm, and it is

$$\hat{\boldsymbol{h}}(n+1) = \hat{\boldsymbol{h}}(n) + \mu e(n)\boldsymbol{x}(n) - \rho_r \frac{\operatorname{sgn}(\hat{\boldsymbol{h}}(n))}{\epsilon_r + |\hat{\boldsymbol{h}}(n-1)|} \quad (10)$$

where $\rho_r = \mu\gamma_r$. Note that although the weight vector $\boldsymbol{w}(n)$ changes in every stage of this sparsity-aware LMS algorithm, the cost function $L_{rl_1}(n)$ is convex unlike the cost function for the l_p-norm penalized LMS and RZA-LMS of [12]. Therefore, the algorithm is guaranteed to converge to the global minimum under some conditions. The analytical study of (10) will be reported in the consequent journal paper. The choice of γ_p and γ_r in (5) and (8) affects the performance of the penalized LMS methods. Assuming some prior distribution for the data, the optimum values for γ_p and γ_r can be chosen using, for example, the approach of [14]. Here in order to investigate the performance of the proposed algorithms and compare it to that of the ZA-LMS and RZA-LMS algorithms, computer simulations are performed.

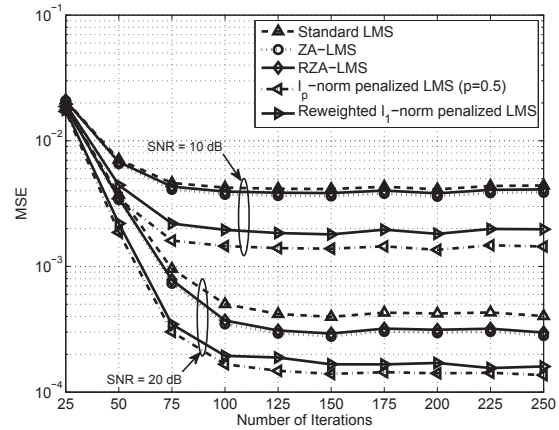

Fig. 2. Example 1, case 1: MSEs of different estimation algorithms vs number of iterations.

5. SIMULATION RESULTS

In our first example, we estimate a CIR of length $N = 16$. Two different cases with different sparsity levels, denoted as case 1 and case 2, are considered. In case 1, only one of 16 taps in the CIR is nonzero, while the position of the nonzero tap is chosen randomly. In case 2, two random taps of the CIR are nonzero. The values of the nonzero taps are chosen from a zero mean Gaussian distribution with a unit variance.

The performance of the proposed l_p-norm penalized and reweighted l_1-norm penalized LMS algorithms is compared to that of the ZA-LMS, RZA-LMS, and standard LMS. For the l_p-norm penalized method, p is set to $1/2$. The other parameters of the proposed algorithms are set to $\rho_p = \rho_r = 5 \times 10^{-4}$ and $\epsilon_p = \epsilon_r = 0.05$. The parameters for the ZA-LMS and RZA-LMS are set to $\rho_{ZA} = \rho_{RZA} = 5 \times 10^{-4}$ and $\epsilon_{RZA} = 10$ as suggested in [12]. The step size is set to $\mu = 0.05$ for all algorithms. Two signal-to-noise ratio (SNR) values of 10 dB and 20 dB are considered. The algorithms tested are compared based on the achievable mean square error (MSE) between the actual and estimated CIR. MSEs are averaged over 1500 simulation runs.

Fig. 2 shows the MSEs of the algorithms tested versus the number of iterations used to estimate the CIR for the sparsity case 1. The MSE results for the sparsity case 2 are shown in Fig. 3. It is expected that as the sparsity level of the CIR increases the MSE performance of the sparsity-aware parameter

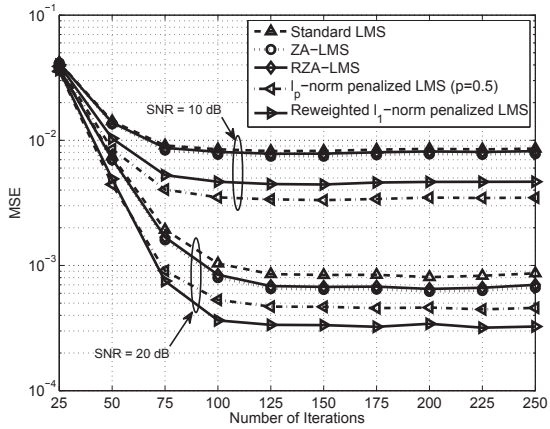

Fig. 3. Example 1, case 2: MSEs of different estimation algorithms vs number of iterations.

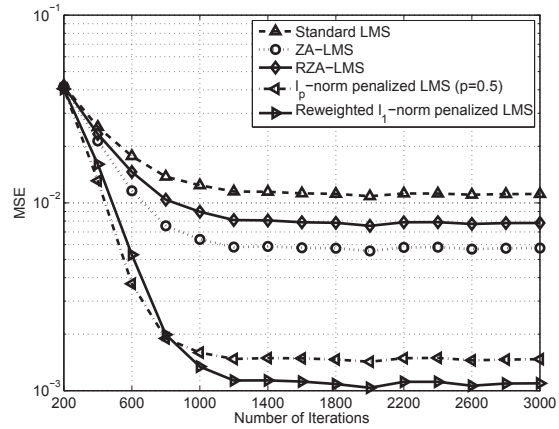

Fig. 4. Example 2: MSEs of different estimation algorithms vs number of iterations, SNR=10 dB.

estimation algorithms degrades. This indeed can be observed by comparing Figs. 2 and 3. For all cases tested, the ZA-LMS and RZA-LMS algorithms exhibit almost the same performance which for the low SNR situation is not better than that of the standard LMS. However, as we increase the SNR, these algorithms outperform the standard LMS. Based on the MSE curves in Figs. 2 and 3 it can be also concluded that among the two proposed algorithms, the l_p-norm penalized LMS has in general better performance than the reweighted l_1-norm penalized LMS. However, in the case SNR = 20 dB in Fig. 3 the reweighted l_1-norm penalized LMS has a better performance. By examining Figs. 2 and 3 it can also be seen that both the reweighted l_1-norm penalized and the l_p-norm penalized LMS algorithms have faster convergence rate and better performance than ZA-LMS and RZA-LMS algorithms.

Our second simulation example considers the case of a CIR of length $N = 256$ with a sparsity level of 16. Therefore, 16 random taps of the CIR are nonzero. Tap values are chosen from a zero mean Gaussian distribution with a variance of 1. SNR is set to 10 dB and the parameter values are chosen as $\mu = 0.005$, $\rho_{ZA} = \rho_{RZA} = 2 \times 10^{-4}$, $\epsilon_{RZA} = 10$, $\rho_p = 5 \times 10^{-6}$, $\rho_r = 5 \times 10^{-5}$, $\epsilon_r = 0.01$, and $\epsilon_p = 0.05$. Fig. 4 shows the MSEs for the algorithms tested. The same conclusions as in the previous example apply here.

6. CONCLUSIONS

The channel estimation problem for channels with sparse CIRs has been considered. The LMS principle has been used as a baseline for parameter estimation. New sparsity-aware l_p-norm penalized and reweighted l_1-norm penalized LMS algorithms have been introduced. The performance of these algorithms has been compared to that of the ZA-LMS and RZA-LMS which are also the examples of sparsity-aware LMS. Simulation results show that the proposed algorithms in general have better performance than the ZA-LMS and RZA-LMS.

7. REFERENCES

[1] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice Hall, 1993.

[2] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*, Prentice Hall, 1985.

[3] M. Godavarti and A. O. Hero, "Partial update LMS algorithms," *IEEE Trans. Signal Processing*, vol. 53, pp. 2382–2399, 2005.

[4] J. Homer, I. Mareels, R. R Bitmead, B. Wahlbertg, and A. Gustafsson, "LMS estimation via structural detection," *IEEE Trans. Signal Processing*, vol. 46, pp. 2651–2663, Oct. 1998.

[5] B. Liu and M. D. Sacchi, "Minimum weighted norm interpolation of seismic records," *Geophysics*, vol. 69, no. 6, pp. 1560–1568, Nov.-Dec. 2004.

[6] E. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, pp. 4203–4215, Dec. 2005.

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

[8] W. Bajwa, J. Haupt, A. Sayeed, and R. Nowak, "Joint source–channel communication for distributed estimation in sensor networks," *IEEE Trans. Inf. Theory*, vol. 53, pp. 3629–3653, Oct. 2007.

[9] Z. Yu, S. Hoyos, and B. M. Sadler, "Mixed–signal parallel compressed sensing and reception for cognitive radio," in *Proc. IEEE ICASSP*, Las Vegas, USA, Apr. 2008, pp. 3861–3864

[10] N. Vaswani, "Kalman filtered compressed sensing," in *Proc. IEEE ICIP*, San Diego, USA, Oct. 2008, pp. 893–896.

[11] D. Angelosante, J. A. Bazerque, and G. B. Giannakis, "Online adaptive estimation of sparse signals: where RLS meets the l_1-norm," *IEEE Trans. Signal Processing*, vol. 58, pp. 3436–3447, July 2010.

[12] Y. Chen, Y. Gu, and A. O. Hero, "Sparse LMS for system identification," in *Proc. IEEE ICASSP*, Taipei, Taiwan, Apr. 2009, pp. 3125–3128.

[13] 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, Dec. 2008.

[14] A. Panahi, M. Lundgern, and M. Viberg, "Maximum a posteriori based regularization parameter selection," in *Proc. IEEE ICASSP*, Prague, Czech Republic, May 2011.