Non-Uniform Norm Constraint LMS Algorithm for Sparse System Identification

F. Y. Wu, *Student Member, IEEE,* and F. Tong, *Member, IEEE*

***Abstract*—Sparsity property has long been exploited to improve the performance of least mean square (LMS) based identification of sparse systems, in the form of l_0-norm or l_1-norm constraint. However, there is a lack of theoretical investigations regarding the optimum norm constraint for specific system with different sparsity. This paper presents an approach by seeking the tradeoff between the sparsity exploitation effect of norm constraint and the estimation bias it produces, from which a novel algorithm is derived to modify the cost function of classic LMS algorithm with a non-uniform norm (p-norm like) penalty. This modification is equivalent to impose a sequence of l_0-norm or l_1-norm zero attraction elements on the iteration according to the relative value of each filter coefficient among all the entries. The superiorities of the proposed method including improved convergence rate as well as better tolerance upon different sparsity are demonstrated by numerical simulations.**

***Index Terms*—p-norm like; non-uniform norm constraint; LMS algorithm; sparsity exploitation.**

I. Introduction

SPARSITY exploitation in adaptive filtering framework has attracted considerable research interest in both theoretical and applied issues for a long time [1,2]. Recently, l_0-LMS [3,4], l_1-LMS [5,6] algorithms are proposed to improve the performance of LMS algorithm for identification of sparse systems by integrating l_0-norm or l_1-norm constraint into the cost function of the standard LMS algorithm to accelerate the convergence of near-zero coefficients. Further analysis of such algorithms was provided in [7,8]. However, for either the l_0-norm or l_1-norm algorithm, there is not any adjustable factor that can adapt the norm constraint itself to the unknown system associated with different sparsity. As a result, the effectiveness of the l_0-LMS [3] or l_1-LMS [7] significantly decreases with the fading of the sparsity of target systems.

Compressive sensing or compressive sampling (CS) methods provide a robust sparse exploitation framework to estimate a sparse signal [9-12]. Accordingly the concept of p-norm like is proposed as an effective diversity measure in [11,12]. Note that minimizing diversity (anti-sparse) is equivalent to maximize the concentration (sparsity) [11,12]. In this letter, a p-norm like constraint is originally adopted to modify the cost function of classic LMS algorithm. Considering that the norm constraint will unavoidably produce an estimation bias at the same time of achieving sparsity exploitation, the classic concept of p-norm like is split to form a new non-uniform norm definition to enable the quantitative adjustment of norm constraint to seek a tradeoff. Thus, optimization analysis of the non-uniform $\mathbf{p}$ vector lead to the derivation of a novel algorithm called non-uniform norm constraint LMS algorithm (NNCLMS), which integrates non-uniform norm constraint into the classic LMS cost function according to the relative value of individual coefficient. Numerical simulations demonstrate that the proposed algorithm outperforms l_0-LMS and l_1-LMS algorithm.

Manuscript received July 26, 2012. The associate editor coordinating the review of this letter and approving it for publication was K. Kuchi.

This work was supported by the funding of the National Natural Science Foundation of China (No.11274259) and the Specialized Research Fund for the Doctoral Program of Higher Education of China (No.20120121110030).

The authors are with the Key Laboratory of Underwater Acoustic Communication and Marine Information Technology of the Minister of Education (Xiamen University), Xiamen, China (e-mail: ftong@xmu.edu.cn).

Digital Object Identifier 10.1109/LCOMM.2013.011113.121586

II. Derivation of the Proposed Algorithm

We consider the following minimization problem:

$$\mathbf{w}(n) = \arg\min_{\mathbf{w}} \mathbf{J}_n(\mathbf{w}) \tag{1}$$

The cost function $\mathbf{J}_n(\mathbf{w})$ is defined as:

$$\mathbf{J}_n(\mathbf{w}) = \left|d(n) - \mathbf{x}^T(n)\mathbf{w}(n)\right|^2 + \gamma||\mathbf{w}(n)||_p^p \tag{2}$$

where, $d(n)$, $\mathbf{w}(n) = [w_0(n), w_1(n), ..., w_{L-1}(n)]^T$ and $\mathbf{x}(n) = [x(n), x(n-1), ..., x(n-L+1)]^T$ denote the desired signal, adaptive filter coefficient vector and input vector respectively, L is the filter length. The estimation square error term $\left|d(n) - \mathbf{x}^T(n)\mathbf{w}(n)\right|^2$ in (2) is commonly used as the cost function in classic LMS-type adaptive methods. The last term $\gamma||\mathbf{w}(n)||_p^p$ is a p-norm like constraint term, where, $\gamma > 0$ is a factor to balance the constraint term and the estimation square error, and $||\mathbf{w}(n)||_p^p$ is called L_p^p norm or "p-norm like" [11,12], which is different from the classic Eulidean norm and defined as[12]:

$$||\mathbf{w}(n)||_p^p = \sum_{i=1}^{n} |w_i|^p, \quad 0 \le p \le 1 \tag{3}$$

Then, we can see that:

$$\lim_{p\to 0} ||\mathbf{w}(n)||_p^p = ||\mathbf{w}(n)||_0 = \#\{i|w_i \neq 0\}, \tag{4}$$

which means counting the number of nonzero coefficients. And,

$$\lim_{p\to 1} ||\mathbf{w}(n)||_p^p = ||\mathbf{w}(n)||_1 = \sum_{i=1}^{n} |w_i|. \tag{5}$$

The norm constraint as (4) or (5) has been utilized and analyzed popularly for the solution of sparse system estimation to derive various l_0-LMS and l_1-LMS algorithm [7,8].

In order to solve the problem in (1) with gradient optimization, we obtain the gradient of the cost function with respect

to filter coefficient vector $\mathbf{w}(n)$ as:

$$\begin{aligned}\hat{\nabla}_n &= \frac{\partial(\left|d(n)-\mathbf{x}^T(n)\mathbf{w}(n)\right|^2)}{\partial\mathbf{w}} + \gamma\frac{\partial(||\mathbf{w}(n)||_p^p)}{\partial\mathbf{w}} \\ &= \mathbf{w}(n) + \gamma\frac{p\mathrm{sgn}[\mathbf{w}(n)]}{||\mathbf{w}(n)||^{1-p}}.\end{aligned} \tag{6}$$

Thus, the gradient descent recursion of the filter coefficient vector is:

$$\begin{aligned}w_i(n+1) =& w_i(n) - \mu\hat{\nabla}_n \\ =& w_i(n) + \mu e(n)x(n-i) \\ &- \frac{\kappa p\mathrm{sgn}[w_i(n)]}{|w_i(n)|^{1-p}}, \quad \forall 0 \le i < L\end{aligned} \tag{7}$$

where an element in the vector $\mathbf{w}(n)$ is defined as $w_i(n)$, μ is the step size parameter of the LMS algorithm, $\kappa = \mu\gamma > 0$ is a parameter combining the contributions of step size and balance factor, $e(n) = d(n) - \mathbf{x}^T(n)\mathbf{w}(n)$ represents estimation error, $\mathrm{sgn}[w_i(n)]$ is the sign function.

However, similar to correction term created by the l_0-LMS and l_1-LMS algorithm [7,8], the last term in (7) caused by the integration of p-norm like constraint will unavoidably cause an estimation bias at the same time of achieving sparsity exploitation [7]. While the introduction of p-norm like enables the optimization of norm constraint via the adjustment of p parameter, as (7) indicates, the p parameter affects the estimation bias as well as the intensity of sparsity correction equally, therefore it cannot be directly adopted to seek a tradeoff between them.

To address this problem, we split the definition of the classic p-norm like in (3) into a non-uniform p-norm like definition which uses a different value of p for each of the L entries in $\mathbf{w}(n)$, as:

$$||\mathbf{w}(n)||_{p,L}^p = \sum_{i=1}^{L}|w_i(n)|_i^p, \quad 0 \le p_i \le 1. \tag{8}$$

Accordingly, the cost function corresponding to the new non-uniform p-norm penalty is:

$$\mathbf{J}_n^{'}(\mathbf{w}) = \left|d(n) - \mathbf{x}^T(n)\mathbf{w}(n)\right|^2 + \gamma||\mathbf{w}(n)||_{p,L}^p \tag{9}$$

The sparsity exploitation optimization problem here becomes:

$$\hat{\nabla}_n = \frac{\partial\mathbf{J}_n^{'}(\mathbf{w})}{\partial\mathbf{w}}, \quad s.t. \quad ||\mathbf{w}(n)||_{p,L}^p = \min. \tag{10}$$

where $s.t.$ means subject to. Instead of (7), the corresponding gradient descent recursion equation is:

$$\begin{aligned}w_i(n+1) =& w_i(n) + \mu e(n)x(n-i) \\ &- \frac{\kappa p_i\mathrm{sgn}[w_i(n)]}{|w_i(n)|^{1-p_i}}, \quad \forall 0 \le i < L\end{aligned} \tag{11}$$

Similar to l_0-LMS and l_1-LMS algorithm [7,8], the intensity of the sparsity correction in (11) is great for small entries of $\mathbf{w}(n)$, but slight for large ones. With the introduction of $\mathbf{p} = [p_0, p_1, ...p_{L-1}]$ vector, it is possible to assign different p for different entries of $\mathbf{w}(n)$ with large or small value to balance the contribution of the sparsity correction and the estimation bias among all of the L entries. To be specific, for large $w_i(n)$, p_i should be optimized to reduce the estimation bias as more as possible, otherwise p_i should be tuned to achieve compromise.

The form of the last term in (11) suggests that a metric of the absolute value can be introduced to quantitatively classify the filter coefficients into the 'large' and 'small' category. Considering the expected range of values of $\mathbf{w}(n)$ vector, we define the absolute value expectation of the vector $\mathbf{w}(n)$ as the classification metric:

$$v(n) = \mathrm{E}[|w_i(n)|], \quad \forall 0 \le i < L \tag{12}$$

Thus, for the 'large' category, the optimization for each entry of $\mathbf{p}$ can be expressed as:

$$\arg\min_{p_i}[p_i|w_i(n)|^{p_i-1}] = 0, \quad i = \arg[w_i(n) > v(n)] \tag{13}$$

For the 'small' category, to avoid the extremely great or slight intensity caused by various value of small $w_i(n)$, as well as to simplify the algorithm implementation, $p = 1$ provides a balanced solution upon different considerations.

Therefore, the comprehensive optimization of the non-uniform constraint will cause p_i taking 0 ($i.e.$, zero constraint) or 1 when $w_i(n) > v(n)$ or $w_i(n) < v(n)$ respectively. In the sense of achieving balance of the sparsity exploitation and estimation bias, we yield the final recursion of filter coefficient vectors:

$$\begin{aligned}w_i(n+1) =& w_i(n) + \mu e(n)x(n-i) \\ &- \kappa f\mathrm{sgn}[w_i(n)], \quad \forall 0 \le i < L\end{aligned} \tag{14}$$

where f can be obtained by:

$$f = \frac{\mathrm{sgn}[v(n) - |w_i(n)|] + 1}{2}, \quad \forall 0 \le i < L \tag{15}$$

The last term in (14) actually impose a non-uniform norm related zero attraction on filter coefficients. Different from the l_0-LMS and l_1-LMS zero attraction term, the exertion of the non-uniform norm term is dependent on the value of each coefficient with respect to the expectation. The non-uniform zero attractor will exert to promote the convergence of small coefficient, and vanish to remove the estimation bias caused by large coefficient, thus seek a tradeoff between these two types of impact on the performance of the algorithm.

Meanwhile, the non-uniform $\mathbf{p}$ vector will exhibit different pattern for target system with different sparse characteristics, it actually provides certain adaptability to the sparsity.

In addition, to further alleviate the residual estimation bias in the proposed algorithm, reweighted zero attraction [7] is introduced to (14). The recursion equation of the proposed algorithm called NNCLMS in this letter can be expressed as:

$$\begin{aligned}w_i(n+1) =& w_i(n) + \mu e(n)x(n-i) \\ &- \frac{\kappa f\mathrm{sgn}[w_i(n)]}{1+\varepsilon|w_i(n)|}, \quad \forall 0 \le i < L\end{aligned} \tag{16}$$

where $\varepsilon > 0$ is a parameter to control the strength of reweighting. With the introduction of reweighting term, the proposed algorithm assigns non-uniform denominator according to the absolute value of individual coefficient to the norm constraint term to reduce the residual estimation bias. The proposed algorithm is described using MATLAB like pseudo-codes as Table I.

We present brief discussions on the impact of algorithm parameters as following:

TABLE I
PSEUDO-CODES OF THE PROPOSED ALGORITHM

Given	μ, κ, ε, L
Initial	$\mathbf{w}$=zeros(L,1)
for	n=1,2...
	Input new $\mathbf{x}(n)$ and $d(n)$;
	$e(n) = d(n) - \mathbf{x}^{\mathrm{T}}(n)\mathbf{w}(n)$;
	$f = \frac{\mathrm{sgn}[\mathrm{E}(\lvert w_i(n)\rvert) - \lvert w_i(n)\rvert]+1}{2}$, $\forall 0 \le i < L$;
	$w_i(n+1) = w_i(n) + \mu e(n) x(n-i) - \frac{\kappa f \mathrm{sgn}[w_i(n)]}{1+\varepsilon\lvert w_i(n)\rvert}$;
end	

According to the above analysis, one can readily accept that parameter κ determines the performance of the proposed algorithm. Reference to [3-8], the parameter denotes the intensity of attraction and results in a faster convergence since the intensity of attraction increases. However, a large κ results in a large steady-state misalignment. So κ is determined by the trade-off between adaptation speed and adaptation quality.

The reweighted parameter ε denotes the strength of reweighing to alleviate the estimation bias. The estimation bias will be reduced when ε increases. Meanwhile, a large ε will result in a weak zero attraction. So the parameter is determined by the trade-off between the estimation bias and the strength of zero attraction.

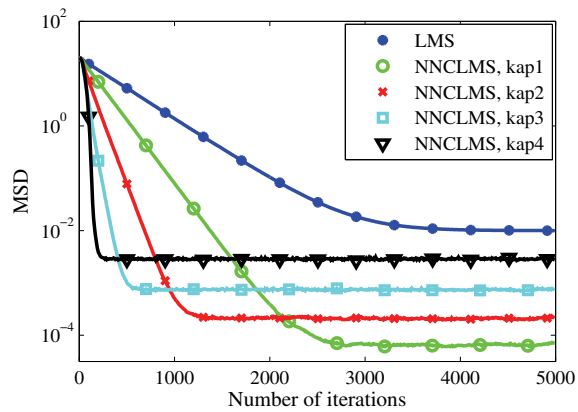

Fig. 1. Learning curves of NNCLMS with different κ, driven by white signal.

TABLE II
PARAMETERS IN THE 1st EXPERIMENT

Algorithms	μ	κ	ε
LMS	0.0012	NA	NA
NNCLMS1	0.0016	0.00016	2
NNCLMS2	0.0024	0.00024	2
NNCLMS3	0.0045	0.00045	2
NNCLMS4	0.0120	0.00120	2

III. SIMULATION

In this section, we will perform numerical simulation studies to analyze the performance of steady state and the convergence rate of NNCLMS, and the performance of NNCLMS will be compared with that of the standard LMS algorithms, l_0-LMS, l_1-LMS, and Reweighted l_1-LMS (Rl_1-LMS),.

The first experiment is designed to test the convergence performance of our method with different value of κ. The unknown system consists of 72 coefficients, in which 2 of them are nonzero ones. The locations of the nonzero taps are uniformly distributed pseudorandom within the global range of the system response, and the values are selected according to zero-mean and unit-variance normally pseudorandom distribution. The driven signal and observed noise are white, Gaussian with variance 1 and 0.01, respectively. The filter length is 72. The parameters of algorithms are provided in Table 2. The analysis results of all algorithms are obtained from 100 times independent simulations. As shown in Fig.1, it is evident that the proposed algorithm converges more rapidly and yields lower MSD than the classic LMS does in a large range of κ. In addition, the Fig.1 also indicates that a larger κ results in a higher convergence rate as well as a larger steady-state misalignment in our method. The above simulation results are consistent with the discussion in the previous section.

The second experiment is designed to test the convergence performance of the proposed algorithm under various sparsities. To evaluate the sparsity quantitatively, the sparse ratio (SR) is defined as the ratio of non-zero tap number to total tap number. A target system consists of 50 coefficients is adopted, with the SR varying from 1/50, to 4/50, and 14/50 at the 3000th, 6000th, and 9000th iterations respectively. The locations and values of the nonzero taps are selected with the same way as the first experiment. The input driven signal and observed noise are the same as that in the first experiment. The filter length is $L = 50$. The parameters of candidate algorithms are carefully chosen to make their steady-state error the same. The parameters of each algorithm are provided in Table 3. All algorithms are simulated 100 times.

The Fig. 2 (a) provides the MSD curves of all algorithms. It is evident that LMS algorithms with different type of norm constraint exhibit different performance enhancement compared to the classic LMS. Specifically, Fig.2 (a) indicates that the proposed NNCLMS achieve the best performance, which is followed by the l_1-LMS, l_0-LMS and Rl_1-LMS respectively.

To test the sensitivity of algorithms to the sparsity of target system, for each algorithm the number of iterations needed to converge from the start point to the MSD level of 10^{-1} under different sparsity is plotted in Fig.2 (b). As can be predicted, Fig.2 (b) reveals that, compared to the classic LMS, the effectiveness of the norm constraint generally degrades with the SR increasing. It is noteworthy that, due to the adaptation capability of non-uniform norm penalty with respect to unknown system, the proposed NNCLMS algorithm significantly outperforms all other algorithms in tolerance upon sparsity, with the number of iterations to achieve convergence only exhibiting tiny deterioration for the SR of 1/50, to 4/50, and 14/50.

IV. CONCLUSION

In order to improve the performance of sparse system estimation, the p-norm like, l_0-norm and l_1-norm are discussed and a new algorithm is derived in this letter by incorporating non-uniform p-norm like constraint to LMS algorithm.

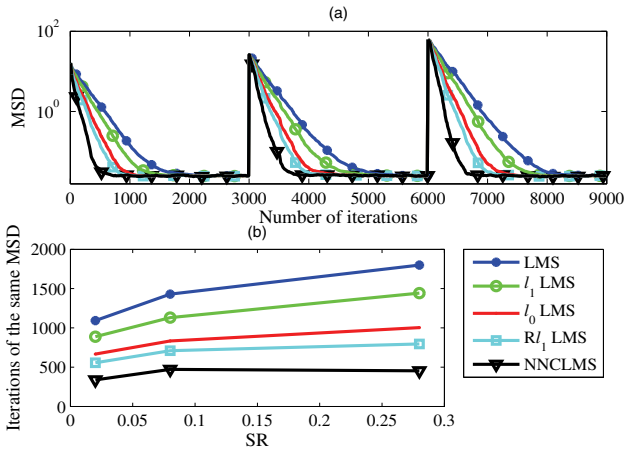

Fig. 2. (a) MSD curves of different algorithms on condition of different SRs. (b) MSD=10^{-1}, iterations of different algorithms at different SR.

TABLE III
PARAMETERS IN THE 2nd EXPERIMENT

Algorithms	μ	κ	ε
LMS	0.0017	NA	NA
l_1-LMS	0.0020	0.00020	2
l_0-LMS	0.0024	0.00024	2
Rl_1-LMS	0.0035	0.00035	2
NNCLMS4	0.0054	0.00054	2

Different from the classic norm constraint algorithms, the proposed algorithm imposes non-uniform zero attraction on the filter coefficient according to the relative value among all entries. Numerical simulations demonstrate that the proposed NNCLMS algorithm has better sparse exploitation performance as well as better tolerance upon sparsity of unknown system.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for improving the quality of this letter.

REFERENCES

[1] R. K. Martin, W. A. Sethares, R. C. Williamson, and C. R. Johnson, Jr., "Exploiting sparsity in adaptive filters," *IEEE Trans. Signal Process.*, vol. 50, no. 8, pp. 1883–1894, Aug. 2002.

[2] B. D. Rao and B. Song, "Adaptive filtering algorithms for promoting sparsity," in *Proc. 2003 Int. Conf. Acoust.*, pp. 361–364.

[3] Y. Gu, J. Jin, and S. Mei, "l_0 norm constraint LMS algorithm for sparse system estimation," *IEEE Signal Process. Lett.*, vol. 16, no. 9, pp. 774–777, Sep. 2009.

[4] K. Shi and P. Shi, "Adaptive sparse Volterra channel estimation with l_0-norm penalty," *Signal Process.*, vol. 91, no. 10, pp. 2432–2436, Oct. 2011.

[5] J. Jin, Y. Gu, and S. Mei, "A stochastic gradient approach on compressive sensing signal reconstruction based on adaptive filtering framework," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 409–420, Apr. 2010.

[6] K. Shi and P. Shi, "Convergence analysis of sparse LMS algorithms with l_1-norm penalty based on white input signal," *Signal Process.*, vol. 90, no. 12, pp. 3289–3293, May 2010.

[7] Y. Chen, Y. Gu, and A. O. Hero, "Sparse LMS for system identification," in *Proc. 2009 Int. Conf. Acoust., Speech, Signal Process.*, pp. 3125–3128.

[8] G. Su, J. Jin, Y. Gu, and J. Wang, "Performance analysis of l_0 norm constraint least mean square algorithm," *IEEE Trans. Signal Process.*, vol. 60, no. 5, pp. 2223–2235, May 2012.

[9] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory.*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[10] R. G. Baraniuk, "Compressive sensing," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 118–122, July 2007.

[11] B. D. Rao and K. K. Delgado, "An affine scaling methodology for best basis selection," *IEEE Trans. Signal Process.*, vol. 47, no. 1, pp. 187–200, Jan. 1999.

[12] R. Chartrand, "Exact reconstruction of sparse signals via nonconvex minimization," *IEEE Signal Process. Lett.*, vol. 14, no. 10, pp. 707–710, Oct. 2007.