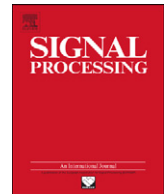




ELSEVIER

Contents lists available at SciVerse ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Fast communication

Gradient optimization p -norm-like constraint LMS algorithm for sparse system estimation

F.Y. Wu, F. Tong*

Key Laboratory of Underwater Acoustic Communication and Marine Information Technology of the Minister of Education, Xiamen University, Xiamen, Fujian 361005, China

ARTICLE INFO

Article history:

Received 19 June 2012

Received in revised form

7 October 2012

Accepted 9 October 2012

Available online 17 October 2012

Keywords:

 p -norm-like constraint

Norm constraint

Least mean square algorithm

Sparsity

Gradient optimization

ABSTRACT

In order to improve the sparsity exploitation performance of norm constraint least mean square (LMS) algorithms, a novel adaptive algorithm is proposed by introducing a variable p -norm-like constraint into the cost function of the LMS algorithm, which exerts a zero attraction to the weight updating iterations. The parameter p of the p -norm-like constraint is adjusted iteratively along the negative gradient direction of the cost function. Numerical simulations show that the proposed algorithm has better performance than traditional l_0 and l_1 norm constraint LMS algorithms.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Aiming to improve the performance of LMS algorithm for identification of sparse systems, Gu and Jin [1] proposed l_0 -LMS, l_1 -LMS [2] algorithms, respectively by integrating l_0 -norm and l_1 -norm constraint into the cost function of the standard LMS algorithm to accelerate the convergence of near-zero coefficients in sparse system. Further analysis concerning the convergence performance of such algorithms was provided in [3–6]. However, as the l_0 -norm or the l_1 -norm constraint itself does not contain any adjustable factor, on the form of norm constraint, it cannot provide adaptability to the characteristics of sparse system when the number of the nonzero taps varies.

Compressive sensing or compressive sampling (CS) methods provide a robust sparsity exploitation framework to estimate a sparse signal [7,8]. The concept of

p -norm-like is proposed as an effective diversity measure in [9,10]. Note that minimizing diversity (antisparsity) is equivalent to maximize the concentration (sparsity) [9,10], in this paper a p -norm-like constraint is introduced to the cost function of classic LMS algorithm.

Considering that the p -norm-like enables the quantitative adjustment of diversity measure by adjusting p , we aim to optimize the p -norm-like constraint with respect to the characteristics (such as the sparsity) of sparse systems by gradient descent tuning of p .

Some gradient iteration algorithms have been proposed to address unit-norm constraints [11] or l_p -norm constraints [12]. But these algorithms are not suitable for solving the nonconvex, non-Lipschitz continuous p -norm-like constraint problem. Inspired by [13], we impose some constraints on the gradient descent iteration of p to derive a novel p -norm-like constraint LMS algorithm. Numerical simulations show that the new algorithm outperforms l_0 -norm and l_1 -norm algorithm. In addition, from the viewpoint of norm based sparsity exploitation, the proposed algorithm provides a formal and systematic way to unify the existing norm constraint LMS algorithms [1–6] into a generalization framework.

* Corresponding author. Tel.: +86 592 2184077;

fax: +86 592 2186397.

E-mail address: ftong@xmu.edu.cn (F. Tong).

2. Derivation of the proposed algorithm

Generally, norm means Euclidean norm which is sometimes called Holder norm noted by L_p or $\|\cdot\|_p$. In [9,10], a general criterion is proposed as an effective measure of diversity:

$$\|\mathbf{x}\|_{p\text{-like}} = \sum_{i=1}^n |x(i)|^p, \quad 0 \leq p \leq 1, \quad (1)$$

where \mathbf{x} is the data vector. As $\|\mathbf{x}\|_{p\text{-like}}$ is different from the classic Euclidean norm, due to its close connection to l_p -norms we call it “ p -norm-like” [10]. It is obvious that “ p -norm-like” is not a classical norm.

For a typical system identification case, the estimation error of the adaptive filter output with respect to the desired signal $d(n)$ is:

$$e(n) = d(n) - \mathbf{x}^T(n)\mathbf{w}(n), \quad (2)$$

where $\mathbf{w}(n) = [w_0(n), w_2(n), \dots, w_{L-1}(n)]^T$, $\mathbf{x}(n) = [x(n), x(n-1), \dots, x(n-L+1)]^T$ denote the coefficient vector and input vector, respectively, and n is discrete time index, L is the filter length. The cost function of the traditional LMS is defined as:

$$\xi(n) = |e(n)|^2, \quad (3)$$

Then, the cost functions of l_0 -LMS and l_1 -LMS are, respectively defined as:

$$\xi_0(n) = |e(n)|^2 + \gamma \|\mathbf{w}(n)\|_0, \quad (4)$$

$$\xi_1(n) = |e(n)|^2 + \gamma \|\mathbf{w}(n)\|_1, \quad (5)$$

where $0 < \gamma < 1$ is a factor to balance the constraint term and the estimation error $e(n)$, the solution of l_0 -norm is an NP hard problem [1], which is usually solved by an approximate function. However, l_1 -norm is a convex function and it is defined by Euclidean norm [2].

For the proposed p -norm-like constraint, we define a cost function $\xi_p(n)$ similar to Eqs. (4) and (5):

$$\xi_p(n) = |e(n)|^2 + \gamma \|\mathbf{w}(n)\|_{p\text{-like}} \quad (6)$$

where $\|\mathbf{w}(n)\|_{p\text{-like}}$ is defined by Eq. (1). The gradient of the p -norm-like constraint cost function Eq. (6) with respect to element $\mathbf{w}(n)$ (defined previously) can be solved as:

$$\begin{aligned} \hat{\nabla}_n &= \frac{\partial \xi_p(n)}{\partial \mathbf{w}(n)} + \gamma \frac{\partial \|\mathbf{w}(n)\|_{p\text{-like}}}{\partial \mathbf{w}(n)} \\ &= -e(n)\mathbf{x}(n-i) + \gamma \frac{p \text{sgn}[w_i(n)]}{|w_i(n)|^{1-p}}, \quad \forall 0 \leq i < L, \end{aligned} \quad (7)$$

Accordingly the filter coefficients are updated iteratively as:

$$\begin{aligned} w_i(n+1) &= w_i(n) - \mu \hat{\nabla}_n \\ &= w_i(n) + \mu e(n)\mathbf{x}(n-i) - \frac{\kappa p \text{sgn}[w_i(n)]}{|w_i(n)|^{1-p}}, \quad \forall 0 \leq i < L, \end{aligned} \quad (8)$$

where $\kappa = \mu\gamma > 0$ and $\text{sgn}[w_i(n)]$ is a sign function defined as:

$$\text{sgn}[w_i(n)] = \begin{cases} 1, & w_i(n) > 0 \\ -1, & w_i(n) < 0, \\ 0, & w_i(n) = 0 \end{cases} \quad \forall 0 \leq i < L. \quad (9)$$

It is necessary to impose an upper bound on the last term in Eq. (8) to avoid divergence when an element of $w_i(n)$ approaches zero, which may exist for a sparse system. Then Eq. (8) is modified as:

$$w_i(n+1) = w_i(n) + \mu e(n)\mathbf{x}(n-i) - \frac{\kappa p \text{sgn}[w_i(n)]}{\varepsilon + |w_i(n)|^{1-p}}, \quad \forall 0 \leq i < L, \quad (10)$$

where $0 < \varepsilon \ll 1$ is a constant set for avoiding the ill-condition calculations in Eq. (8).

As Eq. (10) shows, with the imposing of the p -norm-like constraint, the weight updating of the new algorithm consists of the original LMS updating term and the newly introduced zero-attraction term. Meanwhile, Eq. (10) still contains a variable parameter p . The adjustability of p offers a possible way to adapt the newly introduced p -norm-like constraint to specific sparse system. Thus we hope that p could be adjusted iteratively along the negative gradient direction of the p -norm-like constraint cost function to attain its optimal value. The gradient $G_p(n)$ of the cost function with respect to p can be expressed as:

$$\begin{aligned} G_p(n) &= \frac{\partial \xi_p(n)}{\partial p} = \frac{\partial (\gamma \|\mathbf{w}(n)\|_p^p)}{\partial p} \\ &= \gamma \|\mathbf{w}(n)\|_p^p \ln(\|\mathbf{w}(n)\|_p), \end{aligned} \quad (11)$$

However, there exists a problem that we cannot guarantee the cost function is a concave function with respect to p . It means that the classic gradient descent method may unexpectedly lead to convergence along local optimum value other than the global optimum value. Inspired by [13], we adopt some constraints to impose on the gradient descent derivation to avoid this problem.

First, instead of the exact gradient expressed in Eq. (11), we use the sign of the gradient obtained with a sign function to reduce the possibility at local minimums. The sign function of $G_p(n)$ is given as:

$$\text{sgn}(G_p(n)) = \text{sgn}(\|\mathbf{w}(n)\| - \mathbf{one}(n)), \quad (12)$$

where $\mathbf{one}(n)$ is a unit column vector which has the same size with $\mathbf{w}(n)$.

Furthermore, to alleviate the impact of transient noise caused by stochastic gradient method, smoothed gradient calculated every T iterations are used to guide the descent gradient recursion of p . The iteration formula is thus given by:

$$\begin{aligned} p_{n+T} &= p_n - \delta \text{sgn} \left(\frac{1}{T} \sum_{j=n}^{n+T} G_p(j) \right) \\ &= p_n - \delta \text{sgn} \left\{ \frac{1}{T} \sum_{j=n}^{n+T} [w_j(n)] - 1 \right\}, \end{aligned} \quad (13)$$

where δ is a constant factor to control the step size of descent gradient updating. Thus the proposed algorithm is described using MATLAB like pseudo-codes as Table 1.

3. Brief discussion

The recursion of filter coefficients adopting the sparsity related norm constraint to the classic LMS can be

Table 1
Pseudo-codes of the proposed algorithm.

Given	$\mu, \kappa, p, \varepsilon, \delta, L, T$
Initial for	$\mathbf{w} = \text{zeros}(L, 1), p_{\text{initial}}$ $n = 1, 2, \dots$ Input new $\mathbf{x}(n)$ and $d(n)$; $e(n) = d(n) - \mathbf{x}^T(n)\mathbf{w}(n)$; $p_{n+1} = p_n - \delta \text{sgn} \left\{ \frac{1}{T} \sum_{j=n}^{n+T} [w_j(n)] - 1 \right\}$; $f(n) = p_{n+1} \text{sign}(w_i(n)) / (\varepsilon + \text{abs}(w_i(n))^{(1-p_{n+1})})$; $w_i(n+1) = w_i(n) + \mu e(n) \mathbf{x}(n-i) - \kappa f(n), \forall 0 \leq i < L$
end	

presented as:

$$w_i(n+1) = w_i(n) + \mu e(n) \mathbf{x}(n-i) + Z_{A_{\text{norm-related}}} \\ = w_i(n) + \mu e(n) \mathbf{x}(n-i) - \kappa f(|\cdot|) \quad (14)$$

where $Z_{A_{\text{norm-related}}}$ is specific norm related zero attraction term designed to exert an attraction to zero on small coefficients, $f(|\cdot|)$ of the classic l_0 -norm and l_1 -norm constraint algorithm can be found in [11] and [12], respectively. In the proposed algorithm $f(|\cdot|)$ can be obtained from Eq. (10). The zero attraction term means that new filter weight coefficient will decrease a little when its previous weight coefficient is positive, or increase a little when it is negative. The extent of zero attraction depends on the specific norm constraint adopted, and the algorithm parameters.

In the proposed algorithm, as p can be set to any value between 0 and 1, the p -norm-like is actually degraded to the classic l_0 -norm and l_1 -norm, respectively when $p=0$ and $p=1$. Specifically, when $p=0$, the solution of l_0 -norm is an NP hard problem which is usually solved by an approximate function [1]. When $p=1$, Eq. (8) will be degraded to l_1 -norm constraint LMS algorithm because the p -norm-like constrain has the same gradient function with [2]. Thus, from the viewpoint of generalization, the proposed p -norm-like constraint LMS algorithm offers a unified framework to generalize classic l_0 -norm and l_1 -norm constraint algorithms. Moreover, as p can be adjusted iteratively to attain its optimum value, it is predictable that the proposed p -norm-like constraint LMS algorithm will achieve better sparsity exploitation performance than l_0 -norm or l_1 -norm constraint algorithm does.

Here a brief discussion about the choices of the algorithm parameters, computational complexity as well as the analysis of convergence is given:

The choice of κ : The parameter κ denotes the intensity of zero attraction [1,2], when κ increases, the strength of zero attraction will be promoted to accelerate the convergence. However, a large κ results in a large steady-state misalignment. So the κ is determined by the trade-off between adaptation speed and adaptation quality.

Initialization of p : The negative gradient updating of p is derived in Eq. (13) to attain its optimum value. At the initial stage, generally we can set the original value of p as any value between 0 and 1 to start the descent gradient search.

The choice of δ : As a constant factor to control the step size of p iteration, it may be determined by the trade-off between optimization speed and optimization accuracy.

The computational complexity: The computational complexity per iteration of the four algorithms is listed as Table 2 [2]. For the proposed algorithm, the exponential calculation can be implemented by the form of table look-at to alleviate the complexity at the expense of memory space.

The analysis of convergence: We define the weight error vector $\mathbf{v}(n) = \mathbf{w}(n) - \mathbf{w}_O$, where \mathbf{w}_O is the optimum filter coefficient vector of the classic LMS algorithm. Substituting it to Eq. (10) and taking expectations on both sides of the equation, we obtain:

$$E[\mathbf{v}(n+1)] = (\mathbf{I} - \mu \mathbf{R}) E[\mathbf{v}(n)] - \frac{\kappa p \text{sgn}[\mathbf{w}(n)]}{\varepsilon + |\mathbf{w}(n)|^{1-p}} \quad (15)$$

where $\mathbf{R} = \mathbf{x}(n)\mathbf{x}(n)^T$, \mathbf{I} is the identity matrix, and the vector $(\kappa p \text{sgn}[\mathbf{w}(n)] / \varepsilon + |\mathbf{w}(n)|^{1-p})$ is bounded between $(-\kappa p / \varepsilon + |\mathbf{w}(n)|^{1-p})$ and $(\kappa p / \varepsilon + |\mathbf{w}(n)|^{1-p})$. Therefore, $E[\mathbf{v}(n)]$ converges if μ remains within the range $0 < \mu < (1/\lambda_{\text{max}})$, where λ_{max} is the maximum eigenvalue of $(\mathbf{I} - \mu \mathbf{R})$ [4].

4. Numerical simulation

In this section, the proposed p -norm-like constraint LMS is compared with LMS, l_1 -LMS, l_0 -LMS in the application of sparse system identification. Simulations are designed to test the performance of these algorithms at the presence of unknown system with different sparsity.

The driven signal is zero mean, unit variance white Gaussian signal. The SNR is 40 dB with the length of signal $N=12,000$. The unknown sparse system is a typical underwater acoustic multipath channel [14] which contains 72 coefficients, with the 9th, 11th and 32th dominant coefficients associated with magnitude of 0.5, -0.25 and 1.0, respectively. The other coefficients are all set to zero. We adopt sparsity ratio (SR) to quantitatively evaluate the sparsity of unknown system, which is defined as the ratio of the number of non-zero taps to the total number of coefficients.

During the simulation the SR of the sparse system varies 2 times, taking place at the 4000th and 8000th iterations, respectively. In the first and second variation, two and three new dominant coefficients are added, respectively, producing three sparsity ratios 3/72, 5/72, 8/72 of the target sparse system corresponding to the initial 4000 iterations, the second 4000 iterations and the

Table 2
Computational complexity per iteration of each algorithm.

Algorithms	Addition	Multiplication	Sign	Exponential calculation
LMS	2L	2L+1	NA	NA
l_0 -LMS ($p=0$, [2])	3L	2L	NA	NA
l_1 -LMS ($p=1$, [4])	3L	2L+1	L	NA
p -norm-like LMS	3L+1	2L+1	L	L+1

Table 3
Parameters of candidate algorithms.

Algorithms	μ	κ	δ	T
LMS	0.0034	NA	NA	NA
l_0 -LMS	0.0056	$2.52e-6$	NA	NA
l_1 -LMS	0.0050	$2.80e-6$	NA	NA
p -norm-like LMS	0.0078	$3.92e-6$	0.01	10

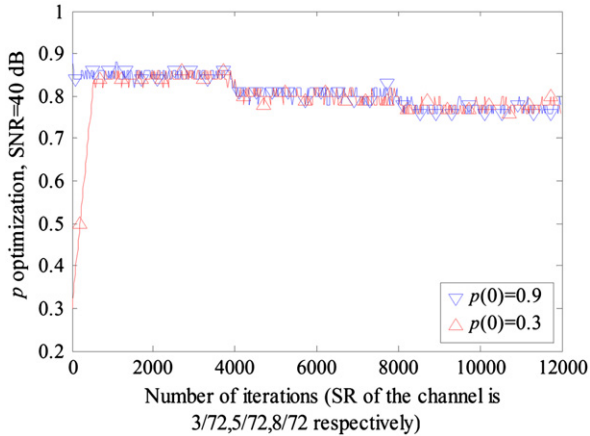


Fig. 1. The iterations of p with different initial value.

third 4000 iterations. The locations of the newly added dominant coefficients are chosen according to Gaussian distribution within the range of global impulse response. The magnitudes of the newly added dominant coefficients are created with Gaussian random value between -1 and 1 . For the selected algorithms, the length of filter L is 72 . The other parameters are carefully selected as listed in Table 3 to make the steady-state errors of each algorithm the same at the initial 4000 iterations. All the algorithms are simulated for 100 times (Monte Carlo, MC simulations).

For the proposed algorithm, initial p is set to 0.3 and 0.9 , respectively to test the performance of the negative gradient p parameter optimization. As shown in Fig. 1, p can iteratively converge starting from both initial value of 0.3 and 0.9 . Moreover, one can also notice that p converges to a new value after each variation of the sparsity. It can be seen from Fig. 1 that high SR may lead to a low p under the gradient guided p -norm-like constraint.

The mean square deviations (MSDs) of each algorithm are shown in Fig. 2. As we can see from Fig. 2, due to performance enhancement of sparsity exploitation, l_1 -norm constraint LMS, l_0 -norm constraint LMS and p -norm-like constraint LMS generally yield faster convergence rate than the standard LMS does. Comparing Fig. 1 with Fig. 2, it can be observed that, under both initial p values of 0.3 and 0.9 , the p -norm-like LMS achieves faster convergence rate than l_1 -norm LMS, l_0 -norm LMS and LMS does.

In addition, from Fig. 2 one may also observe that three sparsity exploitation algorithms show different performance enhancement compared to the standard LMS system when the SR of unknown system varies. For l_0

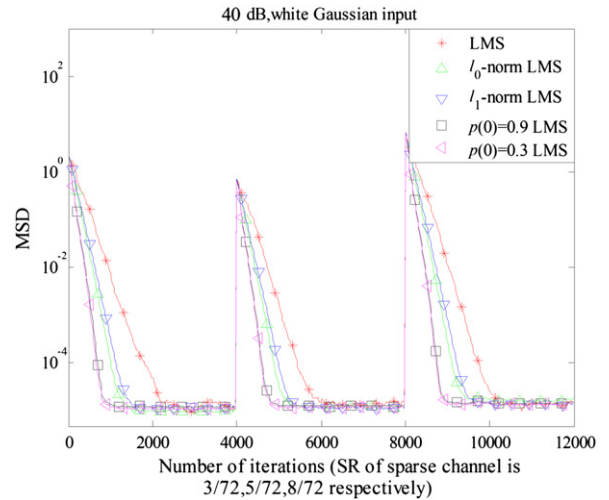


Fig. 2. MSD curves of different algorithms with different SR when SNR=40 dB.

norm and l_1 norm constraint algorithm, Fig. 2 indicates that the performance enhancement with respect to the classic LMS algorithm considerably shrinks with the increasing SR of unknown system. Meanwhile, with the proposed negative gradient tuning of p , the proposed algorithm exhibits better tolerance upon different sparsities, corresponding to slight performance degradation when the SR of unknown system increasing from $3/72$, to $5/72$ and $8/72$.

5. Conclusion

In order to improve the performance of sparse system identification, a new algorithm is derived in this paper by incorporating p -norm-like constraint with LMS algorithm. Different from the classic norm constraint algorithms which use fixed norm constraint such as l_1 -norm or l_0 -norm, p of the proposed algorithm can be iteratively adjusted along the negative gradient to attain its optimum value. Numerical simulations demonstrate that the proposed p -norm-like constraint LMS exhibits better sparsity exploitation performance, as well as better tolerance upon different sparsity than the existing l_1 -norm and l_0 -norm constraint algorithms. Furthermore, from the viewpoint of norm constraint based sparsity exploitation, the proposed algorithm theoretically generalizes existing norm (l_1 -norm and l_0 -norm) constraint LMS algorithms into a unified framework.

Acknowledgment

The authors are grateful for the funding by grants from the National Natural Science Foundation of China (Project No. 11274259) and the Major Project of UIL (University and Industry Liaison) of Fujian province, China (Project No. 2010H6022) in support of the present research. They also thank the anonymous reviewers for improving the quality of this paper.

References

- [1] Y. Gu, J. Jin, S. Mei, l_0 norm constraint LMS algorithm for sparse system identification, *IEEE Signal Processing Letters* 16 (9) (2009) 774–777.
- [2] J. Jin, Y. Gu, S. Mei, A stochastic gradient approach on compressive sensing signal reconstruction based on adaptive filtering framework, *IEEE Journal on Selected Topics in Signal Processing* 4 (2) (2010) 409–420.
- [3] G. Su, J. Jin, Y. Gu, J. Wang, Performance analysis of l_0 norm constraint least mean square algorithm, *IEEE Transactions on Signal Processing* 60 (5) (2012) 2223–2235.
- [4] Y. Chen, Y. Gu, and A.O. Hero, Sparse LMS for system identification, in: *Proc. Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Taiwan, Apr. 2009, pp. 3125–3128.
- [5] K. Shi, P. Shi, Adaptive sparse Volterra system identification with l_0 -norm penalty, *Signal Processing* 91 (10) (2011) 2432–2436.
- [6] K. Shi, P. Shi, Convergence analysis of sparse LMS algorithms with l_1 -norm penalty based on white input signal, *Signal Processing* 90 (12) (2010) 3289–3293.
- [7] D.L. Donoho, Compressed sensing, *IEEE Transactions on Information Theory* 52 (4) (2006) 1289–1306.
- [8] R.G. Baraniuk, Compressive sensing, *IEEE Signal Processing Magazine* 24 (4) (2007) 118–122.
- [9] I.S. Caballero, C.J.P. Prieto, A.A. Rodriguez, Sparse deconvolution using adaptive mixed-Gaussian models, *Signal Processing* 54 (2) (1996) 161–172.
- [10] B.D. Rao, K.K. Delgado, An affine scaling methodology for best basis selection, *IEEE Transactions on Signal Processing* 47 (1) (1999) 187–200.
- [11] S.C. Douglas, S. Amari, S.Y. Kung, On gradient adaptation with unit-norm constraints, *IEEE Transactions on Signal Processing* 48 (6) (2000) 1843–1847.
- [12] W. Huang, D.R. Chen, Gradient iteration with l_p -norm constraints, *Journal on Mathematical Analysis and Applications* 381 (2) (2011) 947–951.
- [13] Y. Gu, K. Tang, H. Cui, LMS algorithm with gradient descent filter length, *IEEE Signal Processing Letters* 11 (3) (2004) 305–307.
- [14] M. Stojanovic, Retrofocusing techniques for high rate acoustic communications, *Journal of the Acoustical Society of America* 117 (3) (2005) 1173–1185.