Two are Better Than One: Adaptive Sparse System Identification using Affine Combination of Two Sparse Adaptive Filters

Guan Gui, Shinya Kumagai, Abolfazl Mehbodniya, and Fumiyuki Adachi
Department of Communications Engineering
Graduate School of Engineering
Tohoku University, Sendai, Japan
{gui, kumagai, method}@mobile.ecei.tohoku.ac.jp; adachi@ecei.tohoku.ac.jp

Abstract—Sparse system identification problems often exist in many applications, such as echo interference cancellation, sparse channel estimation, and adaptive beamforming. One of popular adaptive sparse system identification (ASSI) methods is adopting only one sparse least mean square (LMS) filter. However, the adoption of only one sparse LMS filter cannot simultaneously achieve fast convergence speed and small steady-state mean square deviation (MSD). Unlike the conventional method, we propose an improved ASSI method using affine combination of two sparse LMS filters to simultaneously achieving fast convergence and low steady-state MSD. First, problem formulation and standard affine combination of LMS filters are introduced. Then an approximate optimum affine combiner is adopted for the proposed filter according to stochastic gradient search method. Later, to verify the proposed filter for ASSI, computer simulations are provided to confirm effectiveness of the proposed filter which can achieve better estimation performance than the conventional one and standard affine combination of LMS filters.

Keywords—least mean square, affine combination, sparse adaptive filter, adaptive system identification, $\ell_0$-norm sparse function.

I. INTRODUCTION

A. Background and motivation

Adaptive system identification (ASI) includes many applications such as echo interference cancellation, sparse channel estimation, and adaptive beamforming. One of classical algorithms is called least mean square (LMS) which was first proposed by Widrow and Hoff [1]. The LMS filter is widely used in many applications which require tradeoff between convergence speed and steady-state mean square error (MSE). In other words, a faster (slower) convergence speed of LMS filter often yields a higher (lower) steady-state MSE and/or a higher steady-state mean square deviation (MSD). Unfortunately, LMS filter unable trades off between them due to adopting only one fixed step-size.

To deal with this problem, combination structure of two standard LMS filters, as shown in Fig. 1, is attracting a lot of attention in the last decades. The first adaptive filter $w_1$ uses larger step-size than second filter $w_2$ so that the combination filter can achieve a good/fair tradeoff between convergence speed and steady-state MSE. An improved filter using convex combination of two fixed step-size based standard LMS (CC-LMS) filters was first proposed [2] and later, its steady-state performance was analyzed in [3]. Moreover, an affine combination of two standard LMS (AC-LMS) filters was also proposed and was studied via transient MSE in [4]. However, both CC-LMS and AC-LMS filters do not consider the sparse structure of finite impulse response (FIR) of unknown systems.

In many scenarios, FIR of unknown systems are modeled to be sparse as shown in Fig. 2, containing only a few large coefficients (active) interspersed among many negligible ones (inactive). Taking advantage of such sparse prior information can improve the identifying performance. However, the proposed two combination structure filters [2] [4] do not exploit such information due to the fact that they adopted standard LMS filters. Thus, there is a great interest in exploiting the sparse structure information to improve the filtering performance in sparse systems.

Motivated by the compressive sensing (CS) [5], [6], Chen and his collaborators proposed zero-attracting LMS filter (ZA-LMS) and reweighted ZA-LMS filter (RZA-LMS) using $\ell_1$-norm sparse function.
norm sparse penalty [7]. Based on this work, Taheri and Vorobyov proposed an improved sparse LMS filter using $\ell_p$-norm sparse penalty [8], which is termed as LP-LMS. Gu and his collaborators also proposed an improved sparse LMS filter using approximated $\ell_0$-norm sparse penalty [9], which is termed as L0-LMS. However, the above mentioned adaptive sparse LMS filters adopt only one filter.

To the best of our knowledge, no paper has reported the combined structure of two sparse LMS filters for ASSI. Based on the proposed affine combination filter in [4], in this paper, we propose a novel sparse combination filter adopting two L0-LMS filters to simultaneously achieve three merits: 1) fast convergence speed, 2) low steady-state MSE, and 3) exploiting system sparsity.

![Positions of the FIR coefficients](image)

**Fig. 2.** A typical example of sparse system with FIR length 32 and 5 nonzero coefficients

**B. Main contributions**

In this paper, we propose a novel sparse filter using affine combination of two sparse LMS filters. The proposed filter has two properties: 1) Two LMS filters can provide a good tradeoff between convergence speed and steady-state MSD performance compared with traditional ASI method with only one LMS filter, 2) AC-L0LMS can exploit system sparsity information and then it can further improve the identification performance compared with AC-LMS filter.

The main contribution of this paper is summarized as follows: $\ell_p$-norm sparse function is introduced to cost function of standard AC-LMS filter and then, sparse AC-LMS filter (i.e., AC-L0LMS filter) is proposed for ASSI. Later, several representative experiments are conducted to confirm the effectiveness of our propose methods. In the first experiment, the steady-state MSD performance of the proposed affine combination filter is evaluated with the number of dominant FIR coefficients as parameter. In the second experiment, assuming the constant number of dominant FIR coefficients, the MSD performance of propose algorithms is evaluated with the step-size ratio as a parameter.

**C. Organizations and notations**

The reminder of the rest paper is organized as follows. Section 2 reviews the affine combination of two standard LMS filters and problem formulation. In Section 3, we propose affine combination of two sparse LMS filters to improve system identification performance without sacrificing convergence speed. In section 4, the simulation results via MSD metric are presented to confirm the effectiveness of proposed filters, using a Monte-Carlo technique. Concluding remarks are presented in Section V.

Throughout the paper, matrices and vectors are represented by boldface upper case letters and boldface lower case letters, respectively; the superscripts $(\cdot)^T$, $(\cdot)^H$, and $(\cdot)^{-1}$ denote the transpose, the Hermitian transpose, and the inverse operators, respectively; $E\{\cdot\}$ denotes the expectation operator; $\|\cdot\|_0$ is the $\ell_0$-norm operator that counts the number of nonzero taps in $w$; $w_s$ and $w(n)/\bar{w}(n)$ are the actual system FIR and its $n$-th iterative adaptive FIR estimator, respectively. $E\{\cdot\}$ denotes the expectation operator.

**II. SYSTEM MODEL AND PROBLEM FORMULATION**

Consider a typical sparse finite impulse response (FIR) system as shown in Fig. 2. The input signal $x(t)$ and ideal output signal $d(t)$ are related by

$$d(t) = w^T_s x(t) + z(t),$$

(1)

where $w_s = [w_{s1}, w_{s2}, ..., w_{sN}]^T$ is a $N$-length unknown system FIR vector which is supported only by $K (K \ll N)$ dominant coefficients. $x(t) = [x(t), x(t-1), ..., x(t-(N+1))]$ is $N$-length input signal vector and $z(t)$ is assumed zero-mean and independent and identically distributed (i.i.d.) random noise variable in the system. The object of ASI is to identify the unknown sparse FIR coefficients $w_s$ using the input signal $x(t)$ and output signal $y(t)$. According to Eq. (1), system identification error $e_i(n)$ is given by

$$e_i(n) = d(t) - w^T_s(n)x(t), i = 1, 2,$$

(2)

where $w_s(n)$ is the $n$-th adaptive FIR estimator and $d(n) = d(t)$. Based on Eq. (2), cost function of the $i$-th standard LMS filter [10] is given by

$$J_i(n) = \frac{1}{2} e_i^2(n), i = 1, 2.$$

(3)

With different step-size $\mu_i$, then the $(n+1)$-th update FIR systemvector is updated according to

$$w_i(n+1) = w_i(n) + \mu_i \bar{v}_{w_i(n)} G_i(n)$$

$$= w_i(n) + \mu_i e_i(n)x(t), i = 1, 2,$$

(4)

where $w_i(n), i = 1, 2$ are the $N$-dimensional adaptive vector. Without loss of generality, we assume $\mu_1 \geq \mu_2$ so that $w_1(n)$ achieves faster convergence speed than $w_2(n)$. Note that the steady-state MSD performance of the $w_2(n)$ is better than $w_1(n)$. Also, assuming both $w_1(n)$ and $w_2(n)$ are coupled deterministically and statistically through input signal vector $x(t)$ and additive noise variable $z(t)$. Since $y_i(n) = w_i^T(n)x(t), i = 1, 2,$ denotes individual output signal from $i$-th LMS filter, according to Fig. 2, the system output signal $y(n)$ of the affine combination of the two LMS filters is given by

$$y(n) = \lambda(n)y_1(n) + (1 - \lambda(n))y_2(n)$$

$$= \lambda(n)w_1^T(n)x(n) + (1 - \lambda(n))w_2^T(n)x(t),$$

where $\lambda(n)$ is the time-varying step-size parameter.
\[y(n) = \lambda(n) w_1(n) + w_2(n)\]

where \(w_{12}(n) := w_1(n) - w_2(n)\) is a difference filter and \(\lambda(n)\) is an affine combination parameter to decide final system identification error. In Eq. (5), we can find that \(y(n)\) can be considered as a combination of filter \((w_{12}(n))\) and a weighted filter \((\lambda(n)w_{12}(n))\), the equivalent filter can be given by

\[w_{eq}(n) := \lambda(n)w_{12}(n) + w_2(n).\]  

According to (1) and (5), the overall system error is given by

\[e(n) = d(t) - y(n) = [w_{eq}(n) - \lambda(n)w_{12}(n)]^T x(t) + z(t),\]

where \(w_{eq}(n) := w_o - w_2(n)\) is also considered as a difference filter. This setup generalizes the combination of adaptive filter outputs, and can be used to study the properties of the optimal combination. In [4], the authors proposed the optimal affine combiner

\[\lambda_o(n) = \frac{w_{eq}^T(n)R_{xx}w_{12}(n)}{w_{12}^T(n)R_{xx}w_{12}(n)},\]

which is the expectation for the optimum \(\lambda(n)\), as a function of the unknown weight vector \(w_o\), where \(R_{xx} = E[x(t)x^T(t)]\). This setup generalizes the combination of adaptive filters, and can be used to study the properties of the optimal combination. In [4], the authors proposed the optimal affine combiner

\[\lambda_o(n) = \frac{w_{eq}^T(n)R_{xx}w_{12}(n)}{w_{12}^T(n)R_{xx}w_{12}(n)},\]

which is the expectation for the optimum \(\lambda(n)\), as a function of the unknown weight vector \(w_o\), where \(R_{xx} = E[x(t)x^T(t)]\). However, it cannot be utilized in reality. Employing a stochastic gradient search method, suboptimal affine combiner \(\lambda_s(n)\) [4] was proposed as follows

\[\lambda_s(n + 1) = \lambda_s(n) + \mu_s[d(t) - \hat{w}_{12}^T(n)x(t)]w_{12}(n)x(t),\]

where \(\hat{w}_{12} = \lambda_s(n)w_1(n) + [1 - \lambda_s(n)]w_2(n)\) and \(\mu_s\) are parameters for tracking the adaptation of \(w_1(n)\) and \(w_2(n)\).

### III. AFFINE COMBINATION OF TWO SPARSE LMS FILTERS

Since the affine combination of two standard LMS filters neglects the inherent system sparsity, it often causes performance loss. Unlike the traditional method, we propose an affine combination of two L0LMS filters to exploit the systemsparsity, with individual cost functions [9] given as by,

\[L_i(n) = \frac{1}{2} e_i^2(n) + \beta_i\|w_i(n)\|_0, i = 1, 2.\]

It is well known that solving the \(\|w_i(n)\|_0\) in (10) is a (non-deterministic polynomial-time) NP-hard problem [5]. To deal with this problem, we approximate it by a continuous function

\[\|\hat{w}_i(n)\|_0 \approx \sum_{k=0}^{\infty} (1 - e^{-\alpha|w_{i,k}(n)|}), i = 1, 2.\]

According to (11), cost function of the \(i\)-th L0LMS filter can be changed to

\[L_i(n) = \frac{1}{2} e_i^2(n) + \beta_i \sum_{k=0}^{\infty} (1 - e^{-\alpha|w_{i,k}(n)|}), i = 1, 2.\]

Similarly, with different step-size \(\mu_i\), the \((n+1)\)-th update sparse FIR vector is derived as

\[\hat{w}_i(n + 1) = \hat{w}_i(n) + \mu_i\nabla_w(n)L_i(n)\]

\[= \hat{w}_i(n) + \mu_i e_i(n)x(t) + \alpha \mu_i \beta_i sgn(\hat{w}_i(n))e^{-\alpha|w_i(n)|},\]

for \(i = 1, 2\). It is worth mentioning that the exponential function in Eq. (13) will cause high computational complexity. To reduce the high complexity, the first order Taylor series expansion of exponential function is taken into consideration as

\[e^{-\alpha|w_i(n)|} \approx \begin{cases} 1 - \alpha|w_i(n)|, & \text{when } |\hat{w}_i(n)| \leq \frac{1}{\alpha} \\ 0, & \text{others} \end{cases}.\]

It is worth mentioning that the positive parameter \(\alpha\) controls the system sparseness and identification performance. Though the L0LMS can exploit system sparsity on adaptive system identification, unsuitable parameter \(\alpha\) will cause overall identification performance degradation. In this paper, we adopted \(\alpha = 10\) which is also suggested as in [11]. According to above analysis, the modified update equation of L0LMS can be rewritten as

\[\hat{w}_i(n + 1) = \hat{w}_i(n) + \mu_i e_i(n)x(t) - \mu_s S(\hat{w}_i(n)),\]

where the \(\ell_0\)-norm sparse penalty approximation function \(S(\hat{w}_i(n))\) is defined as

\[S(\hat{w}_i(n)) = \begin{cases} 2\alpha^2 \hat{w}_i(n) - 2\alpha sgn(\hat{w}_i(n)), & \text{if } |\hat{w}_i(n)| \leq \frac{1}{\alpha} \\ 0, & \text{others}. \end{cases}\]

Analog to the Eq. (9), suboptimal affine combiner \(\lambda_s(n)\) can also be given by

\[\hat{\lambda}_s(n + 1) = \hat{\lambda}_s(n) + \tilde{\mu}_s[d(t) - \hat{w}_{12}^T(n)x(t)]\hat{w}_{12}(n)x(t),\]

where \(\hat{w}_{12} = \hat{\lambda}_s(n)\hat{w}_1(n) + [1 - \hat{\lambda}_s(n)]\hat{w}_2(n)\) and \(\tilde{\mu}_s\) are parameters for tracking the adaptation of \(\hat{w}_1(n)\) and \(\hat{w}_2(n)\).

Finally, the overall system error is given by

\[\hat{e}(n) = d(t) - \hat{y}(n) = [\hat{w}_{eq}(n) - \lambda \hat{w}_{12}(n)]^T x(t) + z(t),\]

where \(\hat{w}_{eq}(n) := w_o - \hat{w}_2(n)\) is the difference filter between real and sparse FIR vector.

### IV. COMPUTER SIMULATIONS

To confirm effectiveness of the proposed ASSI method using affine combination of two sparse NLMS filters, we evaluate their steady-state MSD performance which is defined by

\[\text{MSD}(\hat{w}(n)) = E[\|w(n) - \hat{w}(n)\|_2^2].\]

The results are averaged over 1000 independent Monte-Carlo
The signal-to-noise ratio (SNR) is defined as $E_s / \sigma_n^2$, where $E_s$ is the received power of the input signal. Computer simulation parameters are listed in Table I.

| Parameters          | Values                                                                 |
|---------------------|------------------------------------------------------------------------|
| length of FIR       | $N = 32$                                                               |
| no. of nonzero coeffs | $K = 3$ and $6$                                                        |
| positions of nonzero coeffs | Random allocation                                                   |
| distribution of FIR coefficient | random Gaussian $CN(0,1)$            |
| SNR                 | $E_s / \sigma_n^2 = 10$ dB                                           |
| step-size of filter I and filter II | $\mu_1 = \frac{1}{\delta \mu_2}$ and $\mu_2 = \delta \mu_1$ |
| controlling the ratio $\delta = \mu_1 / \mu_2$ | $\delta \in [0,1]$                                                |
| controlling $\mu_1$ of the filter I | $\gamma = 4$                                                             |
| tracking adaptation of $\mathbf{w}_1$ and $\mathbf{w}_2$ in $\lambda_n(n)$ | $\mu_3 = 1$                                                                |
| parameters for $\ell_q$-norm sparse penalty | $\beta = 0.02 \sigma_n^2$ and $\alpha = 10$ |

In the first example, considering two different step-size ratios $\delta \in [0.3,0.5]$, steady-state MSD performance of proposed filter is evaluated for $K = 3$ and $6$, respectively, in Figs. 3-6. To verify the effectiveness of the proposed ASSI method, we compare it with three previous methods, i.e., LMS [10], L0LMS [9] and AC-LMS [4]. For a fair evaluation of these algorithms, same regularization parameter is adopted for L0LMS and AC-L0LMS algorithms, i.e., $\beta = 0.02 \sigma_n^2$, which is also recommended by [12][13]. As Figs. 3-6 show, our proposed filter can achieve lower MSD performance without reducing the convergence speed. Note that choosing smaller (bigger) $\delta$ can achieve lower (higher) MSD and (faster) slower convergence speed. In addition, MSD performance of the proposed filter also depends on FIR sparseness in real systems. For sparser system, the proposed filter can achieve much lower MSD performance by comparing MSD curves of proposed filter in Fig. 3 ($K = 3$) and Fig. 4 ($K = 6$). Hence, the proposed filter can choose different parameters (e.g., $\delta$ and $\beta$) to meet concrete requirements of the system.

![Fig. 3. MSD versus iterations at $K = 3$, $\gamma = 4$ and $\delta = 0.3$ (SNR=10dB).](image)

![Fig. 4. MSD versus iterations at $K = 6$, $\gamma = 4$ and $\delta = 0.3$ (SNR=10dB).](image)

![Fig. 5. MSD versus iterations at $K = 3$, $\gamma = 4$ and $\delta = 0.5$ (SNR=10dB).](image)

![Fig. 6. MSD versus iterations at $K = 6$, $\gamma = 4$ and $\delta = 0.5$ (SNR=10dB).](image)
In this paper, we proposed an affine combination of two sparse LMS filters which can achieve fast convergence and high steady-state MSD performance to improve ASSI performance. First, problem formulation and standard affine combination of LMS filters were introduced. Then, $\ell_p$-norm sparse constraint function based affine combination of two sparse LMS filters for ASSI was presented. System identification performance depends on which affine combiner to choose. The approximate optimum affine combiner was adopted for the proposed filter according to stochastic gradient search method. Later, to verify the effectiveness of the proposed filter for ASSI, selected simulations were provided to confirm the effectiveness of the proposed filter which can achieve better estimation performance than the conventional one and standard affine combination of LMS filters.

**ACKNOWLEDGMENT**

This work was supported by grant-in-aid for the Japan Society for the Promotion of Science (JSPS) fellows grant number 2402366.

**REFERENCES**

1. B. Widrow and D. Stearns, *Adaptive signal processing*, New Jersey: Prentice Hall, 1985.
2. M. Martínez-Ramon, J. Arenas-García, A. Navia-Vazquez, and A. R. Figueiras-Vidal, “An adaptive combination of adaptive filters for plant identification,” in DSP2002, Santorini, Greece, 1-3 July 2002, pp. 1195–1198.
3. J. Arenas-García, “Mean-square performance of a convex combination of two adaptive filters,” IEEE Transactions on Signal Processing, vol. 54, no. 3, pp. 1078–1090, 2006.
4. N. Bershad, “An affine combination of two LMS adaptive filters—transient mean-square analysis,” IEEE Transactions on Signal Processing, vol. 56, no. 5, pp. 1853–1864, 2008.
5. D. L. Donoho, “Compressed Sensing,” IEEE Transactions on Information Theory, vol. 52, no. 4, pp. 1289–1306, 2006.
6. E. J. Candès, J. Romberg, and T. Tao, “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,” IEEE Transactions on Information Theory, vol. 52, no. 2, pp. 489–509, 2006.
7. Y. Chen, Y. Gu, and A. O. Hero III, “Sparse LMS for System Identification,” in ICASSP, Taipei, Taiwan, 19-24 April 2009, pp. 3125–3128.
8. O. Taheri and S. A. Vorobyov, “Sparse channel estimation with $\ell_p$-norm and reweighted $l_1$-norm penalized least mean squares,” in ICASSP, Prague, Czech Republic, 22-27 May 2011, pp. 2864–2867.
9. Y. Gu, J. Jin, and S. Mei, “$L_0$-norm constraint LMS algorithm for sparse system identification,” IEEE Signal Processing Letters, vol. 16, no. 9, pp. 774–777, 2009.
10. S. Haykin, *Adaptive filter theory*, New Jersey: Prentice–Hall, 1996.
11. G. Gui and F. Adachi, “Improved adaptive sparse channel estimation using least mean square algorithm,” EURASIP Journal on Wireless Communications and Networking, vol. 2013, no. 1, pp. 1–18, 2013.
12. Z. Huang, G. Gui, A. Huang, D. Xiang, and F. Adachi, “Regularization selection methods for LMS-Type sparse multipath channel estimation,” in APCC2013, Bali island, Indonesia, 29-31 Aug 2013, pp. 1–5.
13. G. Gui, A. Mehtbhuya, and F. Adachi, “Least mean square/fourth algorithm for adaptive sparse channel estimation,” in PIMRC2013, London, UK, 8-11 Sept 2013, pp. 291–295.