

Convex Combination of LMF and ZA-LMF for Variable Sparse System Identification

Naveed Iqbal, Murwan Bashir, Azzedine Zerguine, Abdeldjalil Aissa El Bey

▶ **To cite this version:**

Naveed Iqbal, Murwan Bashir, Azzedine Zerguine, Abdeldjalil Aissa El Bey. Convex Combination of LMF and ZA-LMF for Variable Sparse System Identification. Asilomar 2019 : Conference on Signals, Systems, and Computers, Nov 2019, Pacific Grove, United States. 10.1109/IEEECONF44664.2019.9048873 . hal-02380181

HAL Id: hal-02380181

https://imt-atlantique.hal.science/hal-02380181

Submitted on 26 Nov 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Convex Combination of LMF and ZA-LMF for Variable Sparse System Identification

Naveed Iqbal *, Murwan Bashir *, Azzedine Zerguine *, and Abdeldjalil Aïssa-El-Bey ‡
*Electrical Engineering Department, King Fahd University of Petroleum & Minerals
Dhahran, 31261, Saudi Arabia
e-mail: {naveediqbal, murwan, azzedine}@kfupm.edu.sa
‡IMT Atlantique, UMR CNRS 6285 Lab-STICC, F-29238 Brest, France
e-mail: {abdeldjalil.aissaelbey@imt-atlantique.fr}

***Abstract*—The objective of this work is to introduce a convex combination of two filters to perform a variable sparse system identification. The first filter is based on the Least Mean Fourth algorithm (LMF), whereas the second is based on its sparse aware version, i.e., the Zero-attractor-LMF (ZA-LMF) algorithm. The convex combination is proposed to solve the sparsity problem under non-Gaussian noise environments. The universality study of the filter indicates that the convex combination always chooses the component filter that offers the lowest Excess Mean Square Error (EMSE) possible. Computer Simulations are performed to confirm the theoretical findings.**

***Index Terms*—Least Mean Fourth (LMF), Transform Domain (TD), Zero-Attractor ZA, Weighted Zero Attractor WZA, L_p norm, L_0 norm, Sparse solution**

I. Introduction

Since the introduction of the Wiener filter theory, the full density of the optimum solution (in a system identification setting) was taken for granted as a fact. However, with time, it is found that the sparse system has noticeable frequent occurrence in the nature and practice. For example, in an acoustic echo cancellation scenario, it is found that the impulse response has small significant elements, whereas the rest are negligibly small. The explanation of this phenomena is due to the large delays introduced in these type of systems [1]. Another interesting prominence is in wireless communication multi-path channels, where the environment contains few dominant signal reflectors implying that the impulse response of the channel is sparse, which is similar to the acoustic echo cancellation system [2]. Large reflectors are also one of the features of the under sea communication channels, where the channel is found to be sparse [3]. Other types of systems have been found to have special structure as well, for example the envelope of the impulse response is exponential in special cases of the acoustic echo cancellation system, which is an additional fact that helps the designer of the adaptive algorithm to achieve better design. The sparsity of the optimal solution is considered even as a stronger feature and exploiting this information is a must for the algorithms' designer [4]. This remarkable structure has triggered the research for the sparse system identification, which nevertheless its novelty, has already produced a plethora of algorithms.

Before understanding the problem from sparsity structure vantage point, the solution is customized to the problem under consideration. This leads to the introduction of family of algorithms based on Proportionate Update (PN) concept. PN algorithms assign adaption rate (step size) for each elements that is proportionate to its value. Proportionate Update Least Mean Square (PN-LMS) and its variants, μ-law MPNLMS [5] and Improved PN-LMS [6], were introduced in beginning for solving the echo cancellation problem [7]. The PN-LMF algorithm has been recently proposed and analyzed in [8]. The PN techniques modify the sparse-agnostic algorithms (like LMS [9] and LMF [10]) to sparse aware. However their performance begins to deteriorate as the number of the significant elements increases.

Another area of research emerged after the advent of the Compressive Sensing (CS) and the LASSO operator techniques [11]. The CS proposed solutions to the problem by using sparsity-recognizing norm l_0. However, because of the difficulty in optimizing this norm and its mathematical non-convexity properties, the l_1 norm is ascribed instead to yield a real time solution to the problem. The LASSO technique is implemented with the Mean Square Error (MSE) function of the LMS, and the resultant algorithm is called as Zero Attractor LMS (LMS) [13]. The ZA-LMF algorithm is introduced in [14]. This LASSO technique results in lower steady state error and also reduces the computational complexity. In order to overcome the deterioration of the EMSE of the ZA-LMS when sparsity increases, the Re-weighted ZA-LMS (RZA-LMS) is proposed. The RZA-LMS [13] assigns weights for elements before taking the l_1 norm. The weights are assigned according to the elements which are to be included in applying the attraction-to-zero force. The RZA-LMS and RZA-LMF [14] are found to be very sensitive to the choice of parameters, which implies that they cannot be considered as final solution to the variable sparsity problem.

The convex combination filter, a filter that chooses between two component filters, is considered as a solution to the trade-off between the two algorithms [15]. For example, the tracking ability of the LMS is superior to that of the RLS under specific conditions, and the converse is true for the RLS. By using a convex combination of the RLS and LMS [16], the trade-off is removed and the convex filter will always insure the best tracking ability, and the convex is called universal — it always chooses the best performing component filter. By looking at the variable sparsity identification as a trade-off between the

sparse-aware algorithm and the sparse-agnostic algorithm, a convex combination is proposed with the LMS and the ZA-LMS as component filters [17].

The LMF is known to be superior to the LMS algorithm when the Signal-to-Noise Ratio (SNR) is low and the measurement noise is non-Gaussian. In this work we propose a convex combination of LMF and ZA-LMF algorithms to solve the problem of variable sparse system identification under these conditions.

II. LMS and LMF algorithms: performance investigation in low SNR environments

The Wiener filter problem depends heavily on the assumption of the Gaussianty of the adaptation error. The Gaussian distribution is known to defined by its first and second only moments, i.e., the mean and the variance. Assuming the mean of the distribution is zero, the task of minimizing the error becomes minimizing the second moment. This results in the famous LMS algorithm [18], which minimizes the variance of the adaption error. The LMS performs exceptionally well when both the input regressors and the measurement noise are Gaussian under high SNR environments. These conditions, insures from the first iterations that the adaptation error is a Gaussian process.

The distribution of the error, however, is a function of SNR. Figure 1 depicts the relationship between the excess kurtosis and SNR for the adaption error at the first iteration in the LMS algorithm. We use the excess kurtosis because it measures how much the distribution in hand is similar to the normal distribution. If the excess kurtosis is zero, we have a perfect normal distribution, where negative and positive distributions correspond to platykurtic and leptokurtic distributions, respectively. As it can be seen from this figure, as the SNR decreases the distribution of the error changes to platykurtic distribution, which is known for their large variations of observations. Hence, we infer from this experiment that the LMS at low SNR suffers slow convergence because it ignores to minimize the excess kurtosis. On the other hand, the LMF succeed in minimizing both the variance and the excess kurtosis simultaneously. We conclude from this observation that the LMF acknowledges the platykurtic distribution of the error in a low SNR environment more than does the LMS, which explains the faster convergence and better MSE level of the LMF described in [10].

III. Problem Formulation

The objective of this work is to introduce a solution to the problem of variable sparse system identification in low signal to noise ratio environments. The LMF is a sparse agnostic algorithm and for sparse systems it has higher excess MSE compared to its sparse aware companion, ZA-LMF, defined by the recurrence:

$$\mathbf{w}_i = \mathbf{w}_{i-1} + \mu \mathbf{u}_i^T e^3(i) - \rho sign(\mathbf{w}_i). \quad (1)$$

It introduces lower EMSE in case of sparse optimum filter, however, the EMSE starts to increase as the sparsity rate

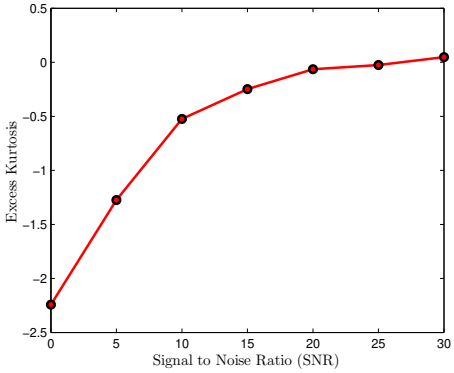


Fig. 1: Excess Kurtosis versus SNR.

increases. Hence, both of the algorithms are not suitable (by itself) for the variable sparse system identification problem.

By combining the two algorithms, with the convex filter technique introduced in [15], the resultant algorithm is universal compared to its components. The universality here means that the convex filter performs as better as its best component elements (ZA-LMF and LMF in this case), in the steady state.

The solution offered by the convex filter is given by the following equation:

$$\mathbf{w}(i) = \lambda(i)\mathbf{w}_1(i) + [1 - \lambda(i)]\mathbf{w}_2(i). \quad (2)$$

The convex combination filter is described in Fig. 2, is an aggregation of two component filters, in our case one is the LMF and the second is the ZA-LMF filter.

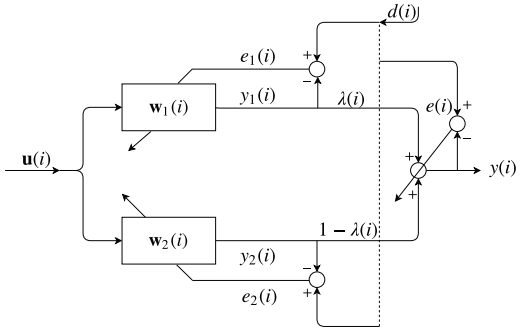


Fig. 2: Diagram of the convex combination filter.

In order to insure that the combination is convex, we have to find a function of $\lambda(i)$ such that it is bounded from zero to one. Sigmoid function has this property, and it is smooth as well (perfect for derivative). The combination factor $\lambda(i)$ is given by the following equation:

$$\lambda(i) = \frac{1}{1 + e^{-a(i)}}. \quad (3)$$

From (3), it is obvious that the update recursion of the convex factor will be on $a(i)$ rather than $\lambda(i)$ directly. The recursion equation is given by

$$\begin{aligned} a(i+1) &= a(i) - \frac{\mu_a}{2}\frac{\partial e^2(i)}{\partial a(i)} \\ &= a(i) + \mu_a e(i)[y_1(i) - y_2(i)]\lambda(i)[1 - \lambda(i)]. \end{aligned} \quad (4)$$

The universality, mathematically speaking, is to see where the combination goes as the nature of the system changes from sparse through semi-sparse to completely non-sparse. The combination should follow the change, and offer the lowest possible EMSE. In order to do this, we study the behavior of (4) at steady state. At steady state we assume that $\lambda(i)$ is independent of $e_{a,k}(i)$, we have the following equation that describes the steady state behavior of $a(i)$:

$$\begin{aligned} E[a(i+1)] &= E[a(i)] + \mu_a E[\lambda(i)[1-\lambda(i)]^2]\Delta J_2 \\ &- \mu_a E[\lambda(i)^2[1-\lambda(i)]]\Delta J_1 \end{aligned} \quad (5)$$

where $\Delta J_1 = J_{ex,1}(\infty) - J_{ex,12}(\infty)$ and $\Delta J_2 = J_{ex,2}(\infty) - J_{ex,12}(\infty)$, and $J_{ex,1}(\infty)$ and $J_{ex,2}(\infty)$ denote the steady-state EMSE of the LMF and the ZA-LMF filters, respectively.

Equation (5) describes the evolution of the $E[a(i)]$ and assumes that both the algorithms are converged. Next, we study (5) for the three cases of systems, considering that the two filers use the same step size.

IV. The universality of the combination filter

The universality in the EMSE sense means that the convex filter always performs as good as its best component filter. Let us denote the EMSE of the LMF filter as $J_{ex,1}(i)$ and the ZA-LMF filter as $J_{ex,2}(i)$. To decide the universality we use the cross error between the two filters which is defined by:

$$J_{ex,12}(i) = E\{e_{a,1}(i)e_{a,2}(i)\} \quad (6)$$

We use this cross error as a metric to decide what the filter chooses in the specific condition. The conditions we have here are full sparse, semi-sparse or completely non-sparse system (dense). Before we study the universality, we find the excess mean square errors, $J_{ex,1}(i), J_{ex,1}(i)$ and $J_{ex,cs}(i)$.

1) EMSE of LMF and ZA-LMF algorithms: We use the following general LMF recursion to start the analysis of the EMSE:

$$\mathbf{w}_i = \mathbf{w}_{i-1} - \mathbf{s}_i + \mu e^3(i)\mathbf{u}_i^T. \quad (7)$$

When the sparsity aware term $\mathbf{s}_i = 0$, we have the ordinary LMF and $\mathbf{s}_i = \rho sign(\mathbf{w}_i)$ results in ZA-LMF algorithm. We customize the analysis for the white Gaussian input, which dictates a linear relation between the EMSE and Mean Square Deviation (MSD), reads as follows:

$$MSD(\infty) = \sigma_u^2 EMSE(\infty). \quad (8)$$

By defining the misalignment error as $\mathbf{z}_i = \mathbf{w}_o - \mathbf{w}_i$. The energy conservation is then derived as follows:

$$\begin{aligned} \|\mathbf{z}_i\|_2^2 &= \{\mathbf{z}_{i-1} + \mathbf{s}_i - \mu e^3(i)\mathbf{u}_i^T\}^T\{\mathbf{z}_{i-1} + \mathbf{s}_i - \mu e^3(i)\mathbf{u}_i^T\} \\ &= \|\mathbf{z}_{i-1}\|_2^2 - \mathbf{z}_{i-1}^T\mathbf{s}_i + \mu e^3(i)\mathbf{z}_{i-1}^T\mathbf{u}_i^T \\ &+ \mathbf{s}_i^T\mathbf{z}_{i-1} + \mathbf{s}_i^2 - \mu e^3(i)\mathbf{s}_i\mathbf{u}_i \\ &+ \mu e^3(i)\mathbf{u}_i\mathbf{z}_{i-1} - \mu e^3(i)\mathbf{u}_i\mathbf{s}_i + \mu^2 e^6(i)\|\mathbf{u}_i\|_2^2. \end{aligned} \quad (9)$$

assuming here that all the signals are real. The Energy relation then summarizes to

$$\begin{aligned} \|\mathbf{z}_i\|_2^2 &= \|\mathbf{z}_{i-1}\|_2^2 + \|\mathbf{s}_i\|_2^2 - 2\mathbf{z}_{i-1}^T\mathbf{s}_i + 2\mu e^3(i)\mathbf{z}_{i-1}\mathbf{u}_i^T \\ &- 2\mu e^3(i)\mathbf{u}_i\mathbf{s}_i + \mu^2 e^6(i)\|\mathbf{u}_i\|_2^2. \end{aligned} \quad (10)$$

The adaption error is assumed to have a symmetrical pdf, and invoking the following assumptions:

1) The input is white Gaussian with autocorrelation matrix $\sigma_u^2\mathbf{I}$.
2) $E\|\mathbf{z}_i\|_2^2 = E\|\mathbf{z}_{i-1}\|_2^2$ as $i \to \infty$. (steady state definition).
3) The sparse enforcing term $\mathbf{s}_i$ is independent from the input regressor and adaption error (this assumption follows directly from the independent assumption).
4) The weight vector $\mathbf{w}_i$ is Gaussian at steady state.
5) The measurement noise is white Gaussian with power σ_v^2.

Now, define the following terms as:

$$\underbrace{2\mu E e^3(i)\mathbf{z}_{i-1}\mathbf{u}_i^T}_{A} = \underbrace{\mu^2 E e^6(i) E\|\mathbf{u}_i\|_2^2}_{B} + \underbrace{E\|\mathbf{s}_i\|_2^2}_{C} + \underbrace{2E\mathbf{z}_{i-1}^T\mathbf{s}_i}_{D} \quad (11)$$

We tackle the terms as follows:
Term A:

$$\begin{aligned} A &= 2\mu E e^3(i)\mathbf{z}_{i-1}\mathbf{u}_i^T = 2\mu E e^3(i)e_a(i) \\ &= 2\mu E e_a^4(i) + 6\mu\sigma_v^2 E e_a^2(i) \end{aligned} \quad (12)$$

Term B:

$$\begin{aligned} B &= \mu^2 M\sigma_u^2 E e^6(i) \\ &= \mu^2 M\sigma_u^2\{E e_a^6(i) + 15\sigma_v^2 E e_a^4(i) + 15\eta_v^4 E e_a^2(i) + \eta_v^6\} \end{aligned} \quad (13)$$

where $\eta_v^j = E|v(i)|^j$
Term C:

$$C = E\|\mathbf{s}_i\|_2^2 = M\rho^2 \quad (14)$$

and $\rho = 0$ for LMF.
Term D: Clearly this term vanishes, when the algorithm is sparse agnostic (LMF). For the case of the ZA-LMF algorithm, we segment the elements of the sparse optimum filter into two groups, the zero and the non-zero elements (i.e., significant). For the non-zeros elements (k $\in NZ$), the sparsity acknowledging term $\mathbf{s}_i$ becomes independent of the filter update vector $\mathbf{w}_i$, while this fact does not hold for the zero elements (k $\in$ Z). The term D then can be expanded into its components as following:

$$\begin{aligned} D &= E\sum_{k\in Z} z(i-1,k)s(i,k) + \sum_{k\in NZ} E\{z(i-1,k)\}s(i,k) \\ &= -\rho E\sum_{k\in Z} w(i-1,k)sgn\{w(i-1,k)\} = \eta_z \end{aligned} \quad (15)$$

By invoking assumption 3 and Price's theorem, we deduce the following:

$$Ew(i-1,k)sgn(w(i-1,k)) = aEw(i-1,k)^2 > 0 \quad (16)$$

$$a = \sqrt{\{(\frac{2}{\pi\sigma_{w,k}^2})\}} \quad (17)$$

Substituting (12)-(15) into (11), an expression for the EMSE is given by

$$Ee_a(i)^2 = \underbrace{\frac{\mu M\sigma_u^2\eta_v^6}{6\sigma_v^2 - 15\mu M\sigma_u^2\eta_v^4}}_{\zeta_{LMF}} + \underbrace{\frac{M\rho^2+\eta_z}{6\mu\sigma_v^2 - 15\mu^2 M\sigma_u^2\eta_v^4}}_{\zeta_{ZA-LMF}} \quad (18)$$

For the case of $\rho = 0$, (18) falls back into the EMSE expression of the LMF described in [19]. Note that $M\rho^2 + \eta_z = \rho\{M\rho - \sqrt{\frac{2}{\pi}}(M-N)\sigma_w\} < 0$ [1], when the number of Non-zeros (N) is very small. From (17), we can see that it is always beneficial to apply a zero attractor to the sparse system identification problem, since it will always insures lower EMSE given that we have chosen the suitable attraction factor value ρ. Now. it remains to find cross excess mean square error for the combination filter defined in (6).

2) Cross Excess Mean Square Error, CEMSE: Starting with the following two recursions of the LMF and ZA-LMF, respectively[2],

$$\begin{aligned} \mathbf{w}_1(i) &= \mathbf{w}_1(i-1) + \mu e_1^3(i)\mathbf{u}_i^T & (19)\\ \mathbf{w}_2(i) &= \mathbf{w}_2(i-1) + \mu e_2^3(i)\mathbf{u}_i^T - \mathbf{s}(i), & (20)\end{aligned}$$

and the relation between the CEMSE and the CMSD, for Gaussian input defined by

$$CEMSE(i) = \sigma_u^4 CMSD(i). \quad (21)$$

Now, we can follow an energy conservation argument approach as follows:

$$\begin{aligned} \mathbf{z}_1(i)^T\mathbf{z}_2(i) &= \mathbf{z}_1(i-1)^T\mathbf{z}_2(i-1) - \mu e_2^3(i)\mathbf{u}_i\mathbf{z}_1(i-1)\\ &+ \mathbf{z}_1^T(i-1)\mathbf{s}_i - \mu e_1^3(i)\mathbf{u}_i\mathbf{z}_2(i-1)\\ &+ \mu^2 e_1^3(i)e_2^3(i)\|\mathbf{u}_i\|_2^2 - \mu e_1^3(i)\mathbf{u}_i\mathbf{s}_i \quad (22)\end{aligned}$$

where $\mathbf{z}_i = \mathbf{w}_o - \mathbf{w}_i$. Similar to the component filter, at steady state the following condition is satisfied [15]:

$$E\{\mathbf{z}_1(i)^T\mathbf{z}_2(i)\} = E\{\mathbf{z}_1(i-1)^T\mathbf{z}_2(i-1)\} \quad (23)$$

then (22) reduces to

$$\begin{aligned} \mu E e_1^3(i)e_{a,1}(i) + \mu E e_2^3(i)e_{a,2}(i) &= \mu^2 M\sigma_u^2 E e_1^3(i)e_2^3(i)\\ &+ E\mathbf{z}_1(i-1)^T\mathbf{s}(i) \quad (24)\end{aligned}$$

By ignoring the higher power errors and recalling assumption 4, (24) further reduces to

$$\begin{aligned} 6\mu\sigma_v^2 J_{ex,12}(i) &= 9\mu^2 M\sigma_u^2\eta_v^4 J_{ex,12}(i) + 9\mu^2 M\sigma_u^2\eta_v^6\\ &+ E\underbrace{\mathbf{z}_1(i-1)^T\mathbf{s}(i)}_{\zeta_z} \quad (25)\\ J_{ex,12}(\infty) &= \frac{\mu M\sigma_u^2\eta_v^6}{6\sigma_v^2 - 9\mu M\sigma_u^2\eta_v^4}\\ &+ \frac{\zeta_s}{6\mu\sigma_v^2 - 9\mu^2 M\sigma_u^2\eta_v^6} \quad (26)\end{aligned}$$

By following the same argument in Term D (that leads (15), we conclude

$$\begin{aligned} \zeta_z &= -\rho\sum_{k\in Z} w_1(\infty,k)sgn(w_2(\infty,k))\\ &= -a\rho\sum_{k\in Z} E\{w_1(\infty,k)w_2(\infty,k)\}\\ &= -a\rho\sum_{k\in Z} E\{z_1(\infty,k)z_2(\infty,k) \quad (27)\end{aligned}$$

where $a = \sqrt{\{(\frac{2}{\pi\sigma_{w,k}^2})\}}$. In order to investigate the sign of ζ_z, we study the behavior of the cross miss-adjustment as $i \to \infty$. The next step is to prove that the term $E\{w_1(\infty,k)w_2(\infty,k)\}$ is positive at steady state.

3) Sign of $E\{w_1(\infty,k)w_2(\infty,k)\}$ at steady state: Starting with the matrix version of (22):

$$\begin{aligned} \mathbf{z}_1(i+1) &= \mathbf{z}_1(i) - \mu e_1^3(i)\mathbf{x}_i^T & (28)\\ \mathbf{z}_2(i+1) &= \mathbf{z}_2(i) - \mu e_2^3(i)\mathbf{x}_i^T\\ &+ \rho sgn\{\mathbf{w}_2(i)\} & (29)\\ E\{\mathbf{z}_2(i+1)\mathbf{z}_1^T(i+1)\} &= E\{\mathbf{z}_2(i) - \mu e_2^3(i)\mathbf{x}_i^T\\ &+ \rho sgn\{\mathbf{w}_2(i)\}\}\{\mathbf{z}_1(i) - \mu e_1^3(i)\mathbf{x}_i^T\}^T\\ &= E\{\mathbf{z}_2(i)\mathbf{z}_1^T(i)\} - \mu E\{\mathbf{z}_2(i)\mathbf{x}_i e_1^3(i)\}\\ &- \mu E\{\mathbf{x}_i^T\mathbf{z}_1(i)e_2^3(i)\}\\ &+ \mu^2 E\{e_1^3(i)e_2^3(i)\mathbf{x}_i\mathbf{x}_i^T\}\\ &+ \rho E\{sgn\{\mathbf{w}_2(i)\}\mathbf{z}_1(i-1)\}\\ &- \mu\rho E\{sgn\{\mathbf{w}_2(i)\}\mathbf{x}_i e_1^3(i)\} & (30)\end{aligned}$$

Ignoring the higher power errors, and evoking assumption (5), and the famous independence assumption, recursion (22) becomes

$$\begin{aligned} \mathbf{c}(i+1) &= \mathbf{c}(i) - 3\mu\sigma_v^2\mathbf{c}(i)\mathbf{R}_x - 3\mu\sigma_v^2\mathbf{R}_x\mathbf{c}(i) + \rho\mathbf{b}(i)\\ &- \mu\rho\sigma_v^2\mathbf{b}(i)\mathbf{R}_x + 9\mu^2\eta_v^6\mathbf{R}_x\\ &+ 18\mu^2\eta_v^4\sigma_x^4\mathbf{c}(i) + 9\mu^2\eta_v^4\sigma_x^4 E\{\mathbf{z}_1^T(i)\mathbf{z}_2(i)\}\mathbf{1} \quad (31)\end{aligned}$$

Recalling that the input correlation matrix of the input is $\sigma_x^2\mathbf{I}$, we finally have:

$$\begin{aligned} \mathbf{c}(i+1) &= \mathbf{c}(i)[1 - 6\mu\sigma_v^2\sigma_x^2 + 18\mu^2\eta_v^4\sigma_x^4] + \rho(1 - \mu\sigma_v^2\sigma_x^2)\mathbf{b}(\mathbf{i})\\ &+ 9\mu^2\sigma_x^2\eta_v^6\mathbf{1} + 9\mu^2\eta_v^4\sigma_x^4 E\{\mathbf{z}_1^T(i)\mathbf{z}_2(i)\}\mathbf{1} \quad (32)\end{aligned}$$

where $\mathbf{c}(i) = diag[E\{\mathbf{z}_2(i)\mathbf{z}_1^T(i)\}]$ and $\mathbf{b}(\mathbf{i}) = diag[E\{sgn\{\mathbf{w}_2(i)\}\mathbf{z}_1(i)\}]$. We pursue (32) more by grouping the elements of $\mathbf{c}(i)$ and $\mathbf{b}(i)$ into the zero $[k \in Z]$ and non-zero $[k \in NZ]$ elements.

For the $\mathbf{b}(i) = [\mathbf{b}_{NZ}(i), \mathbf{b}_Z(i)]^T$, we notice that for the non-zero (significant) elements as $i \to \infty$, $E[sign\{w_{k,2}(i)\}z_{k,1}] = sign\{w_{k,2}^o\}E[z_{k,1}](i) = 0$, whereas, for the zero elements, $E[sign\{w_{k,2}(i)\}z_{k,1}] = -aE[w_{k,2}(i)w_{k,1}] = -aE[z_{k,2}(i)z_{k,1}]$ where $a = \sqrt{\{(\frac{2}{\pi\sigma_{w,k}^2})\}}$.

Hence, $\mathbf{b}(i) = -a\mathbf{c}(i)$, when $i \to \infty$. Next, the evolution of the cross weight error vectors, $E[\mathbf{z_1^T(i)z_2(i)}]\mathbf{1} = \lambda(i)$is given

[1]Practically, $\sqrt{\frac{2}{\pi}} >> \rho$

[2]we assume equal step sizes for the two recursions.

by

$$\lambda(i+1) = (1 - 6\mu\sigma_v^2\sigma_x^2 + 18\mu^2\eta_v^4\sigma_x^4)\mathbf{1}^T\mathbf{c}_{NZ}(i) + (1 - 6\mu\sigma_v^2\sigma_x^2 + 18\mu^2\eta_v^4\sigma_x^4 - a\rho(1 - \mu\sigma_v^2\sigma_x^2))\mathbf{1}^T\mathbf{c}_z(i) + 9\mu^2 M\sigma_x^2\eta_v^6 + 9M\mu^2\eta_v^4\sigma_x^4\lambda(i). \quad (33)$$

The convergence of (33) depends almost on $(1 - 6\mu\sigma_v^2\sigma_x^2 + 18\mu^2\eta_v^4\sigma_x^4) = \alpha$, since $\rho \approx 0$. The attraction factor is normally chosen to be small, since it introduces high EMSE for high ρ for the non-zero elements. For the convergence condition given in [10], i.e., $0 < \mu < \frac{1}{6\sigma_v^2 Tr(\mathbf{R}_x)} = \frac{1}{6M\sigma_v^2\sigma_x^2}$. $\alpha < 1$. Equation (33) is descriptive because it gives the convergence of the cross weight errors in terms of the cross weight error of the zero and non-zeros elements $\mathbf{c}_Z(i)$ and $\mathbf{c}_{NZ}(i)$, respectively. By neglecting $\rho(.)$, we have:

$$\lambda(i+1) = \lambda(i)[1 - 6\mu\sigma_v^2\sigma_x^2 + (9M+18)\mu^2\eta_v^4\sigma_x^4] + 9M\mu^2\sigma_x^2\eta_v^6 \quad (34)$$

$$\lambda(\infty) = \frac{9M\mu^2\sigma_x^2\eta_v^6}{\mu\sigma_x^2(6\sigma_v^2 - (9M+18)\mu\eta_v^4\sigma_x^2)} > 0 \quad (35)$$

for large M and Gaussian noise [3].

For $k \in Z$ (33) becomes:

$$c_k(i+1) = c_k(i)[1 - 6\mu\sigma_v^2\sigma_x^2 + 18\mu^2\eta_v^4\sigma_x^4] + 9\mu^2\sigma_x^2\eta_v^6 + 9\mu^2\eta_v^4\sigma_x^4\lambda(i) \quad (36)$$

$$c_k(\infty) = \frac{9M\mu^2\sigma_x^2(\eta_v^4 + \sigma_x^2\eta_v^6\lambda(\infty))}{\mu\sigma_x^2(6\sigma_v^2 - (9M+18)\mu\eta_v^4\sigma_x^2)} > 0 \quad (37)$$

It follows from (37) that $\zeta_z < 0$ in (27).

In the ensuing, we study the combination behavior in the three states under consideration.

Non-sparse system (dense): In this case the number of significant elements in the optimal vector solution is very large. Hence, $S_{12}(\infty) \approx 0$ and $S(\infty) > 0$. which means $J_{ex,1} > J_{ex,2}$ for this case. The error difference quantities in (5) summarizes as follows:

$$\Delta J_2 = J_{ex,2}(\infty) - J_{ex,12}(\infty) = \frac{6M^2\mu^2\sigma_u^4\eta_v^4\eta_v^6}{(+)} \approx 0 \quad (38)$$

$$\Delta J_1 \approx \frac{\mu M\sigma_u^2\eta_v^6}{6\sigma_v^2 - 15\mu M\sigma_u^2\eta_v^4} - \frac{M\rho^2}{6\sigma_v^2 - 9\mu M\sigma_u^2\eta_v^4} > 0 \quad (39)$$

Note that we have used the small step size approximation, which is apropos for the LMF filters, since we use very small step size to annihilate the probability of divergence. This case corresponds to the second case introduced in [15]. The evolution of $E\{a(i)\}$ is described by:

$$E\{a(i+1)\} \leq [E\{a(i)\} - C]_{a^-}^{a^+} \quad (40)$$

where $C = \lambda^+(1-\lambda^+)^2(\Delta J_1 - \Delta J_2)$, indicating that the asymptotic value of $Ea(i) = -a^+$. In another words, the combination chooses the LMF over the ZA-LMF.

Semi-Sparse Systems:

In this case the following condition is satisfied:

$$\Delta J_1 > 0, \Delta J_2 > 0 \quad (41)$$

since $\Delta J_{ex,1}(\infty) > \Delta J_{ex,2}(\infty) > \Delta J_{ex,12}$ (because the number of zero elements increased dramatically for this case compared to the dense systems), the semi-sparse case then corresponds to the third case in [15], where the asymptotic combination factor is given by

$$E[\lambda(\infty)] = \frac{\Delta J_2}{\Delta J_1 + \Delta J_2} > 0.5 \quad (42)$$

In this case the combination filter is completely optimal and generates EMSE lower than the EMSE of both components filter, i.e.:

$$J_{ex}(\infty) \leq min\{J_{ex,1}(\infty), J_{ex,2}(\infty)\} \quad (43)$$

Sparse Systems:

For this case, $J_{ex,2}(\infty) > J_{ex,1}(\infty)$, since the $S(\infty) < 0$ because of the dominance of the zero terms.

$J_{ex,2}(\infty) > J_{ex,1}(\infty)$ is implied because $S_{12}(\infty) < 0$. Hence, $\Delta J_2 > 0$. What remains is ΔJ_1, which is given by

$$\Delta J_1 = J_{ex,1}(\infty) - J_{ex,12}(\infty) \approx \frac{M\rho}{6\sigma_v^2}\left[\frac{\rho}{\mu} - \sqrt{\frac{2}{\pi}}(1 - \frac{N}{M})\frac{\sigma_w}{\mu} + \sqrt{\frac{2}{\pi}}\frac{9}{2}\frac{\sigma_v^2}{\sigma_w}\right] = k\rho^2 + l\rho \quad (44)$$

and $k = \frac{M}{6\sigma_v^2\mu} > 0$ and $l = \sqrt{\frac{2}{\pi}}[\frac{\sigma_w}{\sigma_v^2} - \frac{9\sigma_v^2}{2\sigma_w}]$. Assuming that A.5 is satisfied, the behavior of ΔJ_1 is a function of the sign of l. Here, then we consider the following two cases of sign:

$\underline{l > 0}$

In this case $\Delta J_1 > 0$, and the sparse case matches again with the third case introduced in [15], i.e., the combination filter introduces lower EMSE compared to the LMF and ZA-LMF.

$\underline{l < 0}$

For this case $\Delta J_1 < 0$, for the range of $0 < \rho < \frac{l}{k}$. This case corresponds to the first case in [15]. In which, the evolution of the combination factor $a(i)$ is described by:

$$E\{a(i+1)\} \geq [E\{a(i)\} + C]_{a^-}^{a^+} \quad (45)$$

which leads, eventually, to the asymptotic value of $\lambda(i) = \lambda^+$, which implies that the combination filter switches to the ZA-LMF completely.

V. Simulation Results

The convex combination is tested under low SNR (= 10 dB) environment with measurement noise modeled with uniform process. The optimum system with number of taps $M = 80$, is first set to be very sparse with fixed value (= 1) at support $S = \{5\}$, then changed to semi sparse system with $S = \{1, 3, 5, 7, 9, 11, 13, 15, 17, 19\}$, and completely non-sparse (dense) system for the last stage. The step size of the algorithms is 0.002, the zero-attractor factors are 2×10^{-6}, 1×10^{-4} for LMF and LMS based algorithms, respectively. The step size of the convex combination filters (i.e. μ_a) has the value of 50 and 10 for LMF and LMS respectively[4]. It should be noted that the MSD as a measure of performance

[3] Sparse filter are naturally large.

[4] Unlike the conventional very high values of μ_a (in terms of 1000,10000), to avoid instability under the conditions of the experiment— this values are found by experiment

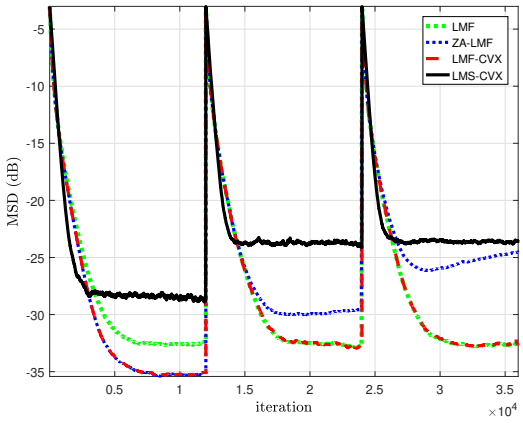


Fig. 3: MSD curves for LMF, ZA-LMF algorithms and their convex combination. SNR level is 10 dB with uniform measurement noise.

is equivalent to the EMSE, especially when we are using Gaussian with unity input power (this can clearly be seen from (21). The results are depicted in Figs. 3-4. For the first stage, the sparse system, the LMS convex combination converged faster to steady state, but with much higher misadjustment compared to the convex combination of the LMF. This trend continued over the three stages of the experiment, proving that the LMF is superior to the LMS under low SNR with uniform noise process. The LMF convex filter followed the ZA-LMF completely, which reaches the lowest steady state level and it is also fastest. However, the MSD level of the ZA-LMF deteriorates as the density of the significant elements in the optimal vector increased. We conclude from the experiment that not only the LMF convex filter insured lower MSD level compared to the LMS combination, but it also reduced the dynamic range of the MSD compared to the LMS.

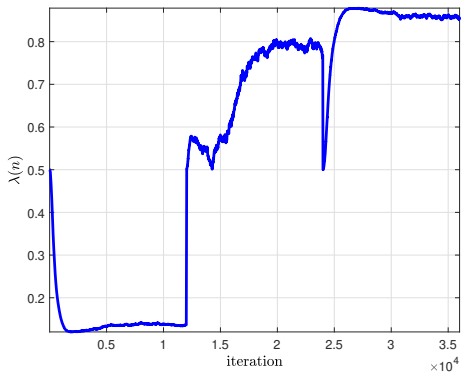


Fig. 4: Combination factor evolution for the first experiment

VI. Conclusion

In this work, we introduced the convex combination filter of LMF and ZA-LMF algorithms to solve the problem of variable sparse system identification under non-Gaussian measurement environments. The steady state universality is conducted, and its outcomes is confirmed with computer simulations, where we found that the convex insures of having the best performance among its components. The convex is compared with its LMS version and found to have excellent performance in this comparison.

Acknowledgment

The authors acknowledge the support provided by the Deanship of Scientific Research at KFUPM under Research Grant **IN161057**.

References

[1] E. Hänsler, "The hands-free telephone problem-an annotated bibliography," *Signal Processing*, vol. 27, no. 3, pp. 259–271, 1992.

[2] W. U. Bajwa, J. Haupt, G. Raz, and R. Nowak, "Compressed channel sensing," in *42nd Annual Conference on Information Sciences and Systems, 2008. CISS 2008.*. IEEE, 2008, pp. 5–10.

[3] V. Kocic, D. Brady, and M. Stojanovic, "Sparse equalization for real-time digital underwater acoustic communications," in *OCEANS'95. MTS/IEEE. Challenges of Our Changing Global Environment. Conference Proceedings.*, vol. 3. IEEE, 1995, pp. 1417–1422.

[4] Y. Gu, K. Tang, H. Cui, and W. Du, "Convergence analysis of a deficient-length lms filter and optimal-length sequence to model exponential decay impulse response," *IEEE Signal Processing Letters*, vol. 10, no. 1, pp. 4–7, 2003.

[5] H. Deng and M. Doroslovacki, "Improving convergence of the pnlms algorithm for sparse impulse response identification," *IEEE Signal Processing Letters*, vol. 12, no. 3, pp. 181–184, 2005.

[6] J. Benesty and S. L. Gay, "An improved pnlms algorithm," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2002* , vol. 2. IEEE, 2002, pp. II–1881.

[7] D. L. Duttweiler, "Proportionate normalized least-mean-squares adaptation in echo cancelers," *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 5, pp. 508–518, 2000.

[8] M. O. Sayin, Y. Yilmaz, A. Demir, and S. S. Kozat, "The krylov-proportionate normalized least mean fourth approach: Formulation and performance analysis," *Signal Processing*, vol. 109, no. 0, pp. 1–13, 2015.

[9] B. Widrow, M. E. Hoff *et al.*, "Adaptive switching circuits." 1960.

[10] E. Walach and B. Widrow, "The least mean fourth (lmf) adaptive algorithm and its family," *IEEE Transactions on Information Theory*, vol. 30, no. 2, pp. 275–283, 1984.

[11] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[12] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.

[13] Y. Chen, Y. Gu, and A. O. Hero, "Sparse lms for system identification," in *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2009.* . IEEE, 2009, pp. 3125–3128.

[14] G. Gui and F. Adachi, "Sparse least mean fourth algorithm for adaptive channel estimation in low signal-to-noise ratio region," *International Journal of Communication Systems*, vol. 27, no. 11, pp. 3147–3157, 2014.

[15] J. Arenas-García, A. R. Figueiras-Vidal, and A. H. Sayed, "Mean-square performance of a convex combination of two adaptive filters," *IEEE Transactions on Signal Processing*, vol. 54, no. 3, pp. 1078–1090, 2006.

[16] M. Silva and V. H. Nascimento, "Improving the tracking capability of adaptive filters via convex combination," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 3137–3149, 2008.

[17] B. K. Das and M. Chakraborty, "Sparse adaptive filtering by an adaptive convex combination of the LMS and the ZA-LMS Algorithms," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 61, no. 5, pp. 1499–1507, 2014.

[18] J. C. Principe, D. Xu, and J. Fisher, "Information theoretic learning," *Unsupervised adaptive filtering*, vol. 1, pp. 265–319, 2000.

[19] A. H. Sayed, *Fundamentals of adaptive filtering*. John Wiley & Sons, 2003.