

A Transform Domain Function Controlled VSSLMS Algorithm for Sparse System Identification

Cemil Turan, Mohammad Shukri Salman

Abstract—The convergence rate of the least-mean-square (LMS) algorithm deteriorates if the input signal to the filter is correlated. In a system identification problem, this convergence rate can be improved if the signal is white and/or if the system is sparse. We recently proposed a sparse transform domain LMS-type algorithm that uses a variable step-size for a sparse system identification. The proposed algorithm provided high performance even if the input signal is highly correlated. In this work, we investigate the performance of the proposed TD-LMS algorithm for a large number of filter tap which is also a critical issue for standard LMS algorithm. Additionally, the optimum value of the most important parameter is calculated for all experiments. Moreover, the convergence analysis of the proposed algorithm is provided. The performance of the proposed algorithm has been compared to different algorithms in a sparse system identification setting of different sparsity levels and different number of filter taps. Simulations have shown that the proposed algorithm has prominent performance compared to the other algorithms.

Keywords—Adaptive filtering, sparse system identification, VSSLMS algorithm, TD-LMS algorithm.

I. INTRODUCTION

LEAST-MEAN-SQUARE (LMS)-type algorithms are very popular due to their simplicity and robustness in adaptive filtering technology [1], [2]. Since they are stochastic gradient based algorithms, they usually have a trade-off between the convergence rate and the misadjustment because of the constant step-size [3]. To enhance the performance of the LMS algorithm, several variable step-size algorithms have been developed [4]-[6]. In [7], authors proposed a function controlled variable step-size LMS algorithm. The algorithm is, basically, based on appropriately selecting a function to control the value of the step-size. It has a high performance compared to many similar algorithms.

Recently, many proposals have shown that the performances of such algorithms can be improved further in system identification settings when the system is sparse (digital tv transmission channel [8], echo paths [9], etc.). In [10], the authors have proposed sparse LMS algorithms that exploit the sparsity of the system. However, still these algorithms suffer from the constant step-size problem. In [11], we have proposed a sparse function controlled variable step-size LMS (SFC-VSSLMS) algorithm. The algorithm takes the advantages of sparsity and variable step-size, and provides prominent results, when the additive noise is white. However, similar to the other algorithms, the performance of the

Cemil Turan is with the Computer Engineering Department, Suleyman Demirel University, Kaskelen, Almaty, Kazakhstan (e-mail: cemil.turan@sdu.edu.kz).

Mohammad Shukri Salman is with the Electrical Engineering Department, American University of the Middle East, Kuwait (e-mail: mohammad.salman@aum.edu.kw).

SFC-VSSLMS algorithm deteriorates when the input signal and/or the additive noise are/is correlated (i.e., the eigenvalue spread of the input correlation matrix is relatively high [12]).

Many proposals appeared to deal with the problem of the high eigenvalue spread of the correlation matrix [13]-[16]. For example, the authors in [15] used the transformed input signal in another domain to reduce the eigenvalue spread of its correlation matrix and as a result the performance of the adaptive filter was enhanced. In order to exploit sparsity on top of the transformation, authors in [16] propose a transform domain reweighted zero attracting LMS (TD-RZALMS) algorithm. Still this algorithm suffers from the constant step-size.

Up to our knowledge, there was no algorithm that exploits the sparsity of the system, uses a variable step-size and transformation of the input signal to reduce the eigenvalue spread of the autocorrelation matrix at the same time. Therefore, in [17], we proposed a transform domain sparse function controlled variable step-size algorithm that combines all of the above mentioned properties. The proposed algorithm imposes an approximate l_0 -norm penalty in the cost function of the transform domain FC-VSSLMS algorithm in order to enhance its performance when the unknown system is sparse. In this paper, we examine the convergence rate and mean-square deviation (MSD) performance of the proposed TD-LMS algorithm for a higher number of filter taps. In addition, the most critical parameter of the algorithm is optimized for better results. Furthermore, the convergence analysis in the mean sense is presented.

The next sections of the paper provides the followings: In Section II, short reviews of the TD-LMS and SFC-VSSLMS algorithms are explained, and the proposed algorithm is presented. In Section III, the convergence analysis of the proposed algorithm is carried out in the mean sense. In Section IV, the procedures of the experiments are explained by comparing the performance of the proposed algorithm to other algorithms by means of simulations in MATLAB and the results are provided and discussed. Finally, conclusions and recommendations are drawn.

II. THE PROPOSED ALGORITHM

A. Review of the Transform Domain LMS Algorithm

Consider a linear system with input-tap vector $\mathbf{x}(k) = [x_0, \dots, x_{M-1}]^T$ and output $d(k)$ related by

$$d(k) = \mathbf{h}^T \mathbf{x}(k) + n(k) \quad (1)$$

where $\mathbf{h} = [h_0, \dots, h_{M-1}]^T$ is the unknown system coefficients with length M , T is the transposition operator

and $n(k)$ is the observation noise. In order to use the input vector $\mathbf{x}(k)$ in the TD-LMS algorithm, it can be processed by a unitary transform such as discrete Fourier transform (DFT) or discrete cosine transform (DCT). If the filter length is specified as (M) then the transform matrix \mathbf{T} will be in a dimension of $M \times M$ with orthonormal rows. So the transformed vector may be calculated as

$$\mathbf{X}(k) = \mathbf{T}\mathbf{x}(k), \quad (2)$$

where \mathbf{T} is a unitary matrix (i.e., $\mathbf{T}^T\mathbf{T} = \mathbf{I}$). The filter output is given as

$$y(k) = \mathbf{W}^T(k)\mathbf{X}(k) \quad (3)$$

and the corresponding estimation error is

$$e(k) = d(k) - y(k) \quad (4)$$

where $\mathbf{W}(k)$ is the transform domain filter coefficients vector. It is important to note that although $\mathbf{X}(k)$ is in the transform domain, $y(k)$ and $e(k)$ are both in time domain. The filter coefficients of TD-LMS are then updated as

$$\mathbf{W}(k+1) = \mathbf{W}(k) + \mu\mathbf{D}^{-1}e(k)\mathbf{X}(k), \quad (5)$$

where \mathbf{D} is an $M \times M$ diagonal matrix whose elements are the transform signal power components ($E[|X_i|^2]$) [16]. It is clear that the convergence speed of TD-LMS algorithm depends on $\mathbf{D}^{-1}\mathbf{R}_{XX}$. Appendix A shows that, with a proper orthogonal transformation, the eigenvalue spread of the autocorrelation matrix of the input signal can be reduced.

B. Proposed Algorithm

In the SFC-VSSLMS algorithm, the aim is to find the optimum vector of \mathbf{h} as $\mathbf{h}(\mathbf{k})$,

$$\mathbf{h}(\mathbf{k}) = \arg \min_{\mathbf{w}(\mathbf{k})} \left\{ \frac{1}{2} |e(k)|^2 + \xi \|\mathbf{w}(k)\|_0 \right\}, \quad (6)$$

where $e(k)$ is defined in (4), ξ is a small positive constant and $\|\cdot\|_0$ denotes the l_0 -norm of the weights vector. Since (6) is an NP-hard problem, $\|\mathbf{w}(k)\|_0$ is approximated in [18] as

$$\|\mathbf{w}(k)\|_0 \simeq \sum_{k=0}^{M-1} (1 - e^{-\lambda|\mathbf{w}(k)|}), \quad (7)$$

where $\lambda > 0$. The update equation of the SFC-VSSLMS algorithm is written as:

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu(k)e(k)\mathbf{x}(k) - \rho(k)\text{sgn}[\mathbf{w}(k)]e^{-\lambda|\mathbf{w}(k)|}, \quad (8)$$

where $\rho(k) = \xi\lambda\mu(k)$ and $\mu(k)$ is the variable step-size parameter and given by [7] as,

$$\mu(k+1) = \alpha\mu(k) + \gamma_s f(k) \frac{|e(k)|^2}{\hat{e}_{ms}^2(k)}, \quad (9)$$

where $0 < \alpha < 1$, $\gamma_s > 0$ is an updating parameter, $f(k) = 1/k$ if $k < L$ else $f(k) = 1/L$ and $\hat{e}_{ms}^2(k)$ is the estimated MSE defined as,

$$\hat{e}_{ms}^2(k) = \beta\hat{e}_{ms}^2(k-1) + (1-\beta)|e(k)|^2. \quad (10)$$

where β is the weighting factor $0 \ll \beta < 1$.

In this paper, we propose a new cost function using inverse TD coefficient vector $\mathbf{W}(k)$ obtained by TD input vector $\mathbf{X}(k)$, hence

$$\mathbf{H}(\mathbf{k}) = \arg \min_{\mathbf{W}(k)} \left\{ \frac{1}{2} |e(k)|^2 + \xi \|\mathbf{T}^T\mathbf{W}(k)\|_0 \right\}. \quad (11)$$

Deriving (11) with respect to $\mathbf{W}(k)$ and substituting in $\mathbf{W}(k+1) = \mathbf{W}(k) - \mu(k) \frac{\partial J(\mathbf{W}(k))}{\partial \mathbf{W}(k)}$ yields

$$\mathbf{W}(k+1) = \mathbf{W}(k) + \mu(k)\mathbf{D}^{-1}e(k)\mathbf{X}(k) - \rho(k)\mathbf{D}^{-1}\mathbf{T}^T \text{sgn}[\mathbf{T}^T\mathbf{W}(k)]e^{-\lambda|\mathbf{T}^T\mathbf{W}(k)|} \quad (12)$$

where $\mathbf{X}(k)$ is given in (2) and $\mathbf{W}(k)$ is the TD vector of $\mathbf{w}(k)$. In this paper we use the DCT due to its real valued components.

III. CONVERGENCE ANALYSIS OF THE PROPOSED ALGORITHM

In this section we present the convergence analysis of the proposed algorithm assuming that the input signal and noise are statistically mutually independent. Denoting \mathbf{H} to be the transformed optimal filter coefficients, as

$$\mathbf{H} = \mathbf{T}\mathbf{h}. \quad (13)$$

Substituting (13) and (2) into (1) yields

$$d(k) = \mathbf{H}^T\mathbf{X}(k) + n(k). \quad (14)$$

The TD misalignment vector of the LMS algorithm is defined as

$$\boldsymbol{\theta}(k) = \mathbf{W}(k) - \mathbf{H}, \quad (15)$$

combining (3), (4), (14) and (15) gives

$$e(k) = -\mathbf{X}^T(k)\boldsymbol{\theta}(k) + n(k), \quad (16)$$

substituting the results of (15) and (16) in (12) provide,

$$\boldsymbol{\theta}(k+1) = [\mathbf{I}_M - \mu(k)\mathbf{D}^{-1}\mathbf{X}(k)\mathbf{X}^T(k)]\boldsymbol{\theta}(k) + \mu(k)\mathbf{D}^{-1}\mathbf{X}(k)n(k) - \rho(k)\mathbf{D}^{-1}\mathbf{T}^T \text{sgn}[\mathbf{T}^T\mathbf{W}(k)]e^{-\lambda|\mathbf{T}^T\mathbf{W}(k)|}, \quad (17)$$

taking the expectation of (17) with the independence assumption we obtain,

$$E[\boldsymbol{\theta}(k+1)] = [\mathbf{I}_M - \mu(k)\mathbf{D}^{-1}\mathbf{R}_{XX}]E[\boldsymbol{\theta}(k)] - \rho(k)\mathbf{D}^{-1}\mathbf{T}^T E[\text{sgn}[\mathbf{T}^T\mathbf{W}(k)]e^{-\lambda|\mathbf{T}^T\mathbf{W}(k)|}]. \quad (18)$$

In (18), $\rho(k)\mathbf{D}^{-1}\mathbf{T}^T E[\text{sgn}[\mathbf{T}^T\mathbf{W}(k)]e^{-\lambda|\mathbf{T}^T\mathbf{W}(k)|}]$ is bounded and hence $E[\boldsymbol{\theta}(k)]$ converges if the maximum eigenvalue of

$[\mathbf{I}_M - \mu(k)\mathbf{D}^{-1}\mathbf{R}_{XX}]$ in-between $(-1, 1)$ and this, in turn, guarantees the convergence of the algorithm in the mean sense.

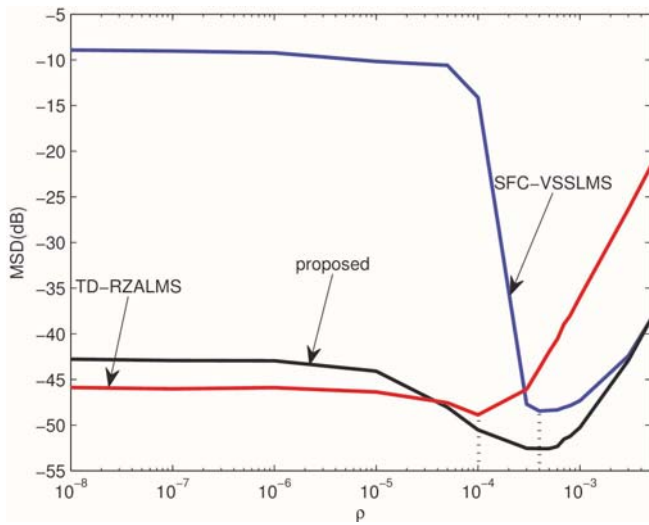


Fig. 1 Optimum ρ for all algorithms by extensive simulations in the first experiment

IV. SIMULATIONS AND RESULTS

In this section, to verify the performance of the proposed algorithm, we compare the results with those of the SFC-VSSLMS and TD-RZALMS algorithms (see Table I) in sparse system identification settings where the colored input signal is used. All the experiments are implemented with 300 independent runs.

In the first experiment, three unknown time varying systems of 16 coefficients with different sparsity levels are assumed. In the first 15000 iterations, 1 randomly placed coefficient with value “1” is assumed (15/16 sparsity). In the next 15000 iterations, 4 randomly placed coefficients with values “1” per each are used (12/16 sparsity). In the last 15000 iterations, 8 randomly distributed coefficients with “1” value per each (8/16 sparsity) are used. The tap-input vector $\mathbf{x}(k)$ is considered to be an autoregressive process of order four obtained by filtered white Gaussian noise from a fourth order system such that $x(k) = 1.79x(k-1) - 1.85x(k-2) + 1.27x(k-3) - 0.41x(k-4) + n_0(k)$, where $n_0(k)$ is a zero mean white Gaussian sequence with 0.15 variance. The eigenvalue spread ratio of the input correlation matrix is measured to be 944. The observed noise is assumed to be a white Gaussian random sequence with a variance regulated for a signal-to-noise ratio (SNR) of 30 dB. The performance criteria for comparisons is mean-square-deviation ($MSD = E\{\|\mathbf{H} - \mathbf{W}(k)\|^2\}$). The parameters used in simulations are given in the following: For SFC-VSSLMS: $\alpha = 0.999$, $\beta = 0.75$, $\gamma = 0.004$, $L = 200$, $\lambda = 8$ and $\rho = 5 \times 10^{-4}$. For TD-RZALMS: $\rho = 10^{-5}$, $\epsilon = 10$ and $\mu = 0.003$. For the proposed algorithm: $\alpha = 0.997$ and the other parameters are the same that of SFC-VSSLMS. The most important parameter selection is the sparsity-aware parameter

ρ . We select ρ by assuming 1/16 sparsity of the unknown system and find ρ that gives minimum MSD as shown in Fig. 1 and generalized to the other parts of the experiment. However, for TD-RZALMS, we found that ρ needs to be regularized if the sparsity changes, so we have selected different ρ than the found optimum one in order to guarantee the convergence of the algorithm in the other sparsity regions. Fig. 2 shows the tracking versus MSD curves for all algorithms. It is seen that the proposed algorithm provides a higher convergence speed and a lower MSD than those of the other algorithms in all regions. However, it should be noted that, in region 3; where the sparsity is relatively low, the performance of the TD-RZALMS and SFC-VSSLMS algorithms severely deteriorate whereas the proposed algorithm is not so affected.

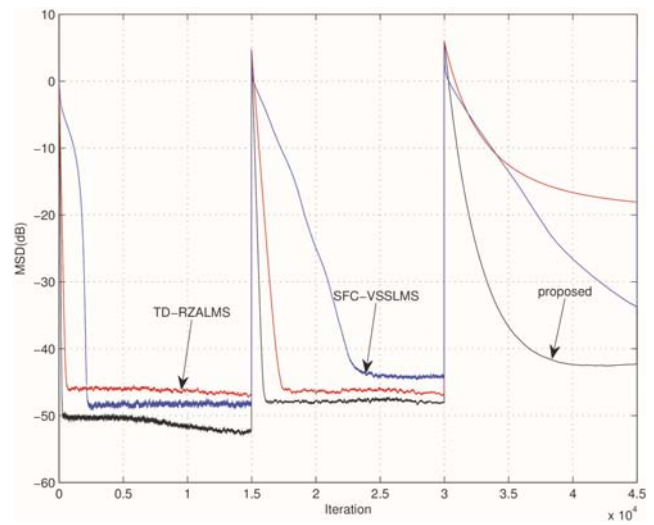


Fig. 2 Tracking and steady state behaviors of the algorithms for 94%, 75% and 50% sparsity with an adaptive filter driven by a colored input signal

In the second experiment, in order to observe the performance of the algorithms for a higher number of filter taps and with a correlated Gaussian noise, their performances are compared for a 150 taps filter with thirty randomly distributed coefficients with value “1” (80 % sparsity) and the SNR 30 dB. The algorithms are simulated with the following parameters. For the TD-RZALMS: $\rho = 10^{-5}$, $\epsilon = 10$ and $\mu = 0.003$. For SFC-VSSLMS and the proposed algorithms: $\alpha = 0.99$, $\beta = 0.75$, $\gamma = 0.003$, $L = 200$, $\lambda = 8$ and $\rho = 5 \times 10^{-4}$. Note that ρ is selected in the same way explained in experiment I (please see Fig. 3). Fig. 4 shows that the convergence of SFC-VSSLMS is very slow (here the advantage of the transformation is very clear). Whereas, the proposed algorithm has a convergence than that of the TD-RZALMS algorithm by almost 1000 iteration and 6 dB lower MSD.

From experiments I and II we see the virtues of combining variable step-size (faster convergence) and transform domain (lower MSD) very clearly.

V. CONCLUSIONS

In this paper, the performance of recently proposed TD-LMS algorithm is investigated for a sparse system

TABLE I
SUMMARY OF THE COMPARED ALGORITHMS

TD-RZALMS	SFC-VSSLMS	Proposed
Initialize: μ, ρ for $k = 1, \dots, N$ $e(k) = d(k) - \mathbf{w}(k)^T \mathbf{x}(k)$ $\mathbf{X}(k) = \mathbf{T}\mathbf{x}(k)$ and $\mathbf{W}(k) = \mathbf{T}\mathbf{w}(k)$ $\mathbf{W}(k+1) = \mathbf{W}(k) + \mu \mathbf{D}^{-1}$ $e(k)\mathbf{X}(k) - \rho \mathbf{D}^{-1} \mathbf{T}^T \text{sgn}(\mathbf{T}^T \mathbf{W}(k))$	Initialize: $ems(-1) = 0, \mu(-1) = 0, \rho$ for $k = 1, \dots, N$ $ems(k) = \beta ems(k-1) + (1-\beta) e(k) ^2$ $\mu(k) = \alpha \mu(k-1) + \gamma f(k) e(k) ^2 / ems(k)$ where $e(k) = d(k) - \mathbf{w}(k)^T \mathbf{x}(k)$ $f(k) = 1/k$ if $k < L$ else $f(k) = 1/L$ $\mathbf{w}(k+1) = \mathbf{w}(k) + \mu(k)e(k)\mathbf{x}(k) - \rho(k)\text{sgn}(\mathbf{w}(k))e^{-\lambda \mathbf{w}(k) }$	Initialize: $ems(-1) = 0, \mu(-1) = 0, \rho$ for $k = 1, \dots, N$ $ems(k) = \beta ems(k-1) + (1-\beta) e(k) ^2$ $\mu(k) = \alpha \mu(k-1) + \gamma f(k) e(k) ^2 / ems(k)$ where $e(k) = d(k) - \mathbf{w}(k)^T \mathbf{x}(k)$ $f(k) = 1/k$ if $k < L$ else $f(k) = 1/L$ $\mathbf{X}(k) = \mathbf{T}\mathbf{x}(k)$ and $\mathbf{W}(k) = \mathbf{T}\mathbf{w}(k)$ $\mathbf{W}(k+1) = \mathbf{W}(k) + \mu(k)\mathbf{D}^{-1}e(k)\mathbf{X}(k) - \rho(k)\mathbf{D}^{-1}\mathbf{T}^T \text{sgn}(\mathbf{T}^T \mathbf{W}(k))e^{-\lambda \mathbf{T}^T \mathbf{W}(k) }$

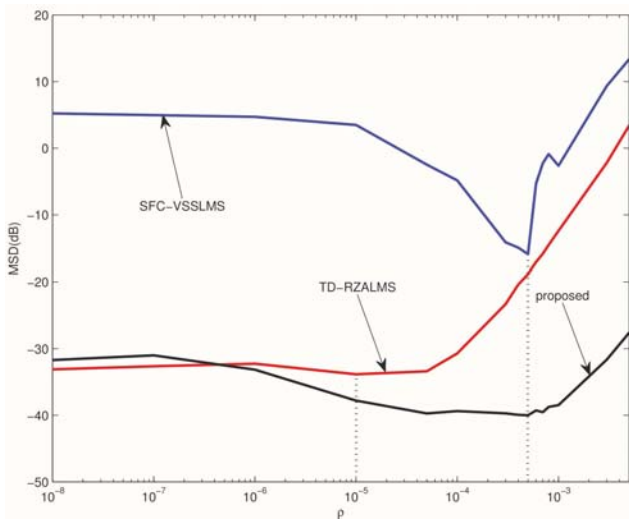


Fig. 3 Optimum ρ for all algorithms by extensive simulations in the second experiment

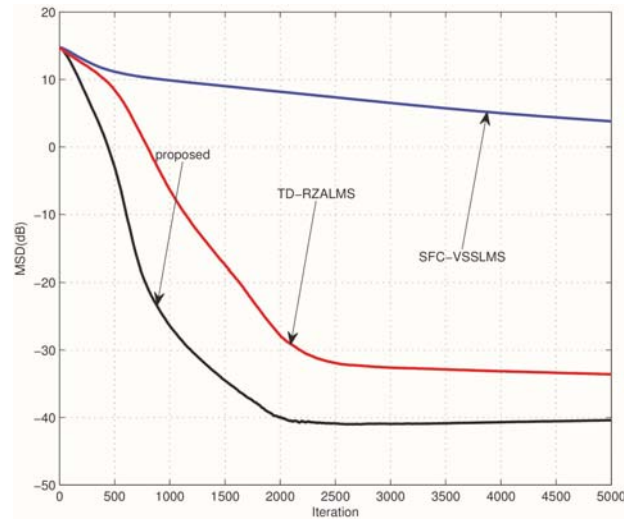


Fig. 4 MSD learning curves of the algorithms for a 150 taps filter with 30 random coefficients as "1"

having a large number of filter taps. An optimum value is searched out for the critical parameter and employed for an enhanced steady state performance. The convergence analysis provided in the mean sense guarantees the convergence having the similar conditions of those in conventional LMS algorithm. The results show that the proposed algorithm has a prominent performance compared to the SFC-VSSLMS and TD-RZALMS in sparse system identification settings of different sparsity levels and different number of filter taps with correlated input and/or noise.

APPENDIX

Without loss of generality, assume that the power of the input signal is unity, i.e., $E(x_n^2) = 1$. From matrix theory [19], for any square matrix A with size $N \times N$, a maximum eigenvalue (λ_{max}) and a minimum eigenvalue (λ_{min}),

$$\lambda_{max} \leq Tr(A) \quad (19)$$

and for $N \geq 2$

$$\lambda_{min} \geq Det(A) \quad (20)$$

where Tr and Det are trace and determinant operators, respectively. Therefore the ratio of

$$\psi(A) = \frac{Tr(A)}{Det(A)} \geq \frac{\lambda_{max}}{\lambda_{min}} \quad (21)$$

Defining

$$R_{XX} = E[\mathbf{X}\mathbf{X}^T] = \mathbf{T}R_{xx}\mathbf{T}^T \quad (22)$$

where R_{XX} and R_{xx} are the autocorrelation matrices of the transformed and non-transformed input signals, respectively.

$$Tr(D^{-1}R_{XX}) = Tr(R_{xx}) = N \quad (23)$$

and

$$Det(D^{-1}R_{XX}) = Det(D^{-1})Det(R_{xx}) \quad (24)$$

Therefore, dividing (23) by (24)

$$\begin{aligned} \psi(D^{-1}R_{XX}) &= \frac{N}{Det(D^{-1})Det(R_{xx})} \\ &= Det(D)\psi(R_{xx}) \end{aligned} \quad (25)$$

Since the $Det(D)$ is always assured to be less than or equal to unity, hence

$$\psi(D^{-1}R_{XX}) \leq \psi(R_{xx}) \quad (26)$$

In other words, (26) shows that, with a proper orthogonal transform, eigenvalue spread can be reduced.

REFERENCES

- [1] Sayed, A. H.: Fundamentals of Adaptive Filtering, Wiley, New York, USA, 2003.

- [2] Haykin, S.: Adaptive filter theory, Prentice Hall, Upper Saddle River, NJ, 2002.
- [3] Bismor, D.: "Extension of LMS stability condition over a wide set of signals," International Journal of Adaptive Control and Signal Processing, 2014, DOI: 10.1002/acs.2500.
- [4] Kwong, R. H.: "A Variable Step-Size LMS Algorithm," IEEE Transaction on Signal Processing, 40 (7): 1633-1641, 1992.
- [5] Jamel, T. M. and Al-Magazachi, K. K.: "Simple variable step size LMS algorithm for adaptive identification of IIR filtering system," IEEE International Conference on Communications, Computers and Applications, 23-28, 2012.
- [6] Wu, X., Gao, L. and Tan, Z.: "An improved variable step size LMS algorithm," IEEE International Conference on Measurement, Information and Control (ICMIC), 01: 533-536, 2013.
- [7] Li, M., Li, L. and Tai, H-M.: "Variable Step Size LMS Algorithm Based on Function Control," Springer, Circuits, Systems and Signal processing, 32 (6): 3121-3130, 2013.
- [8] Schreiber, W. F.: "Advanced television systems for terrestrial broadcasting: some problems and some proposed solutions," Proceedings of the IEEE, 83: 958-981, 1995.
- [9] Duttweiler, D. L.: "Proportionate normalized least-mean-squares adaptation in echo cancelers," IEEE Transaction on Speech Audio Processing, 8(5): 508-518, 2000.
- [10] Chen, Y., Gu, Y. and Hero, A. O.: "Sparse LMS For System Identification," IEEE International Conference Acoustic, Speech and Signal Processing, 3125-3128, 2009.
- [11] Turan, C. and Salman, M. S.: "A sparse function controlled variable step-size LMS algorithm for system identification," Signal Processing and Communications Applications Conference, 329-332, 2014.
- [12] Madisetti, V. K. and Williams, D. B.: Digital Signal Processing Handbook, CRC Press, 1999.
- [13] Jahromi, M. N. S., Salman, M. S., Hocanin, A. and Kukrer, O.: "Convergence Analysis of the Zero-Attracting Variable Step-Size LMS Algorithm for Sparse System Identification," Signal, Image & Video Processing, Springer, 2013, DOI: 10.1007/s11760-013-0580-9.
- [14] Salman, M. S.: "Sparse leaky-LMS algorithm for system identification and its convergence analysis," International Journal of Adaptive Control and Signal Processing, 28: 1065-1072, 2014.
- [15] Narayan, S., Peterson, A. M. and Narasimha, M. J.: "Transform domain LMS algorithm," IEEE Transactions on Acoustics, Speech and Signal Processing, 31(3): 609-615, 1983.
- [16] Shi, K. and Ma, X.: "Transform domain LMS algorithms for sparse system identification," IEEE International Conference on Acoustics, Speech and Signal Processing, 3714-3717, 2010.
- [17] Turan, C., Salman, M. S. and Haddad, H.: "A Transform Domain Sparse LMS-type Algorithm for Highly Correlated Biomedical Signals in Sparse System Identification" in IEEE 35th International Conference on Electronics and Nanotechnology (ELNANO15), Kyiv, Ukraine, 711-714, 2015.
- [18] Sing-Long, C. A., Tejos, C. A. and Irrarazaval, P.: "Evaluation of continuous approximation functions for the l_0 - norm for compressed sensing" Proceedings of the International Society for Magnetic Resonance in Medicine, 17, 2009.
- [19] Strang, G.: Linear Algebra and its Applications, New York, Academic Press, 1976.