# Mean-Squared Analysis of the Partial-Update NLMS Algorithm

Stefan Werner, Marcello L. R. de Campos, and Paulo S. R. Diniz

Helsinki University of Technology, Universidade Federal de Rio de Janeiro

*Abstract*—In this paper, we present mean-squared convergence analysis for the partial-update normalized least-mean square (PU-NLMS) algorithm with closed-form expressions for the case of white input signals. The analysis used order statistics and the formulas presented here are more accurate than the ones found in the literature for the PU-NLMS algorithm. Simulation results shows excellent agreement with the results predicted by the analysis of the PU-NLMS algorithm.

## I. INTRODUCTION

WHEN implementing an adaptive-filtering algorithm, the affordable number of coefficients that can be used will depend on the application in question, the adaptation algorithm, and the hardware chosen for implementation. With the choice of algorithms ranging from the simple least-mean square (LMS) algorithm to the more complex recursive least squares (RLS) algorithm, tradeoffs between performance criteria such as, e.g., computational complexity and convergence rate, have to be made. In certain applications, the use of the RLS algorithm is prohibitive due to the high computational complexity and in such cases we must resort to simpler algorithms. As an example, consider the acoustic echo cancellation application where the adaptive filter may require thousands of coefficients [1]. This large number of filter coefficients may impair even the implementation of low computational complexity algorithms, such as the normalized least-mean square (NLMS) algorithm [1]. As an alternative, instead of reducing filter order, one may choose to update only part of the filter coefficient vector at each time instant. Such algorithms, referred to as partial-update (PU) algorithms, can reduce computational complexity while performing close to their full-update counterparts in terms of convergence rate and final mean-squared error (MSE). In the literature one can find several variants of the LMS and the NLMS algorithms with partial updates [2]–[8], as well as more computationally complex variants based on the affine projection algorithm [9], [8].

The objective of this paper is to analyze one particular case of the partial-update NLMS (PU-NLMS) algorithm introduced in [9], [8] that obeys the principle of minimum disturbance [9], [8]. The results from our analysis, which is based on order statistics, yield more accurate bounds on step size and on the prediction of excess MSE when compared to the results presented in [9], [8]. We also clarify the relationship between the PU-NLMS and M-Max NLMS [3], [4] algorithms, whereby we show that the M-Max NLMS algorithm uses an instantaneous

S. Werner is with the Signal Processing Laboratory, Helsinki University of Technology, P.O. Box 3000, FIN-02015, Espoo, Finland (e-mail: stefan.werner@hut.fi).

M. L. R. de Campos and P. S. R. Diniz are with COPPE/EE-Universidade Federal do Rio de Janeiro, Caixa Postal 68504, RJ, Brazil, 21945-970 (e-mails: {campos,diniz}@lps.ufrj.br).

estimate of the step size that achieves the fastest convergence in the MSE.

## II. THE PARTIAL-UPDATE NLMS ALGORITHM

This section reviews the partial-update NLMS (PU-NLMS) algorithm proposed in [9], [8]. The approach taken here is slightly different from that in [9], [8], but the final algorithm is the same as the one that satisfies the minimum disturbance criterion. We also provide analysis in the mean-squared sense with new bounds on the step size to be used in the PU-NLMS algorithm that are more accurate than the one given in [9], [8].

The objective in PU adaptation is to derive an algorithm that only updates $L$ out of the $N + 1$ filter coefficients. Let the $L$ coefficients to be updated at time instant $k$ be specified by an index set $\mathcal{I}_L(k) = \{i_0(k), , \ldots, i_{L-1}(k)\}$ with $\{i_j(k)\}_{j=0}^{L-1}$ taken from the set $\{0, \ldots, N\}$. Note that $\mathcal{I}_L(k)$ depends on the time instant $k$. As a consequence, the $L$ coefficients to be updated can change between consecutive time instants. A question that naturally arises is "Which $L$ coefficients should be updated?" The answer to this question can be related to the optimization criterion chosen for the algorithm derivation.

In the conventional NLMS algorithm, the new coefficient vector can be obtained as the vector $\mathbf{w}_{k+1}$ that minimizes the Euclidean distance $\|\mathbf{w}_{k+1} - \mathbf{w}_k\|^2$ subject to the constraint of zero *a posteriori* error. Applying the same idea for the partial update of vector $\mathbf{w}_k$, we take the updated vector $\mathbf{w}_{k+1}$ as the vector minimizing the Euclidean distance $\|\mathbf{w}_{k+1} - \mathbf{w}_k\|^2$ subject to the constraint of zero *a posteriori* error *with the additional constraint of updating only $L$ coefficients*. For this purpose, we introduce the diagonal matrix $\mathbf{A}_{\mathcal{I}_{L(k)}}$ having $L$ elements equal to one in the positions indicated by $\mathcal{I}_{L(k)}$ and zeros elsewhere. Defining the complementary matrix $\tilde{\mathbf{A}}_{\mathcal{I}_{L(k)}} = \mathbf{I} - \mathbf{A}_{\mathcal{I}_{L(k)}}$ will give $\tilde{\mathbf{A}}_{\mathcal{I}_{L(k)}} \mathbf{w}_{k+1} = \tilde{\mathbf{A}}_{\mathcal{I}_{L(k)}} \mathbf{w}_k$, which means that only $L$ coefficients are updated. Assuming a sequence of input vectors $\{\mathbf{x}_k\}_{k=1}^{\infty}$ and a sequence of desired signals $\{d_k\}_{k=1}^{\infty}$, we can write the sequence of output errors $\{e_k\}_{k=1}^{\infty}$ as

$$e_k = d_k - \mathbf{w}^\mathrm{T} \mathbf{x}_k$$

where $\mathbf{x}_k$ and $\mathbf{w} \in \mathbb{R}^{N+1}$, and $d_k$ and $e_k \in \mathbb{R}$. With this notation the optimization criterion for the partial update can be formulated as

$$\mathbf{w}_{k+1} = \min_{\mathbf{w}} \|\mathbf{w} - \mathbf{w}_k\|^2 \text{ subject to}$$

$$\begin{cases} \mathbf{x}_k^\mathrm{T} \mathbf{w} = d_k \\ \tilde{\mathbf{A}}_{\mathcal{I}_{L(k)}} (\mathbf{w} - \mathbf{w}_k) = \mathbf{0} \end{cases} \quad (1)$$

Applying the method of Lagrange multipliers onto the objective function $f(\mathbf{w}, \lambda_1, \lambda_2) = \|\mathbf{w} - \mathbf{w}_k\|^2 + \lambda_1(d_k - \mathbf{x}_k^\mathrm{T}\mathbf{w}) +$

$\boldsymbol{\lambda}_2^{\mathrm{T}} \tilde{\mathbf{A}}_{\mathcal{I}_L(k)} (\mathbf{w} - \mathbf{w}_k)$ where $\lambda_1$ is a scalar and $\boldsymbol{\lambda}_2$ is an $(N + 1) \times 1$ vector, gives

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \frac{e_k \mathbf{A}_{\mathcal{I}_L(k)} \mathbf{x}_k}{\|\mathbf{A}_{\mathcal{I}_L(k)} \mathbf{x}_k\|^2}. \tag{2}$$

We see from (2) that only the coefficients of $\mathbf{w}_k$ indicated by the index set $\mathcal{I}_L(k)$ are updated, whereas the remaining coefficients are not changed from iteration $k$ to iteration $k + 1$.

We now concentrate on the choice of the index set $\mathcal{I}_L(k)$. Substituting the recursions in (2) into (1) we get the Euclidean distance as

$$E(k) = \|\mathbf{w}_{k+1} - \mathbf{w}_k\|^2 = \frac{e_k^2}{\|\mathbf{A}_{\mathcal{I}_L(k)} \mathbf{x}_k\|^2}$$

For a given value of $e_k^2$, we can conclude that $E(k)$ achieves its minimum when $\|\mathbf{A}_{\mathcal{I}_L(k)} \mathbf{x}_k\|$ is maximized. In other words, we should update the $L$ coefficients of $\mathbf{w}_k$ related to the elements of $\mathbf{x}_k$ with the largest norm.

In order to control stability, convergence speed, and error in the mean-squared sense a step size is required, leading to the following final recursion for the PU-NLMS algorithm

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \mu \frac{e_k \mathbf{A}_{\mathcal{I}_L(k)} \mathbf{x}_k}{\|\mathbf{A}_{\mathcal{I}_L(k)} \mathbf{x}_k\|^2} \tag{3}$$

The bound on the step size is given by (see Appendix I)

$$0 < \mu < \frac{2}{\mathrm{E}\left[\frac{r_k^2}{\tilde{r}_k^2}\right]} \approx \frac{2\mathrm{E}\left[\tilde{r}_k^2\right]}{(N+1)\sigma_x^2}$$

where $\tilde{r}_k^2$ has the same probability distribution as $\|\mathbf{A}_{\mathcal{I}_L(k)} \mathbf{x}_k\|^2$, and $r_k^2$ has the same probability distribution as $\|\mathbf{x}_k\|^2$, which in this particular case is a sample of an independent process with chi-distribution with $(N + 1)$ degrees of freedom, $\mathrm{E}\left[r_k^2\right] = (N + 1)\sigma_x^2$. For given $N$ and $L$, $\mathrm{E}\left[\tilde{r}_k^2\right]$ can be evaluated numerically, as shown in Appendix II. It can also be shown that $L\sigma_x^2 \leq \mathrm{E}\left[\tilde{r}_k^2\right] \leq (N + 1)\sigma_x^2$ for white Gaussian input signals (see Lemma 3 in Appendix II). A more pessimistic bound on the step size, $0 \leq \mu \leq 2L/(N + 1)$, was given in [8] as a consequence of the crude approximation $\mathrm{E}\left[\tilde{r}_k^2\right] \approx L\sigma_x^2$.

In Appendix II it is shown that if order statistics is used, the final excess MSE after convergence is given by

$$\Delta\xi_{exc} \approx (N+1) \frac{\mu\sigma_n^2\sigma_x^2}{2 - \mu\mathrm{E}\left[\frac{r_k^2}{\tilde{r}_k^2}\right]} \mathrm{E}\left[\frac{1}{\tilde{r}_k^2}\right]$$

$$\approx (N+1) \frac{\mu\sigma_n^2\sigma_x^2}{2\mathrm{E}\left[\tilde{r}_k^2\right] - \mu(N+1)\sigma_x^2} \tag{4}$$

When $L = N + 1$, Equation (4) is consistent with the results obtained for the conventional NLMS algorithm in [10]. The algorithm presented in this section is identical to the partial-update NLMS algorithm with multiple blocks of contiguous coefficients to be updated proposed in [9], [8], for the case of unity block size and $L$ blocks. Choosing blocks of filter coefficients rather than the $L$ coefficients corresponding to the elements with largest magnitude in the input-signal vector can
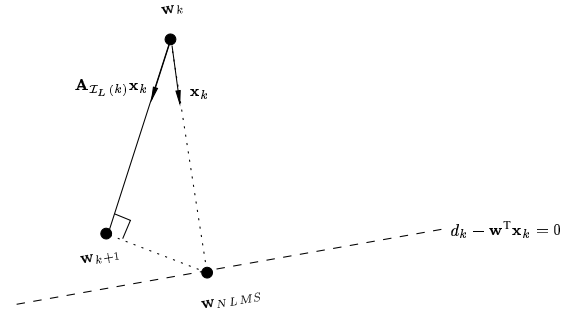


Fig. 1. The solution $\mathbf{w}_{k+1}$ is the PU-NLMS algorithm update obtained with a time-varying step size $\mu_k = \|\mathbf{A}_{\mathcal{I}_L(k)} \mathbf{x}_k\|^2/\|\mathbf{x}_k\|^2$, or equivalently, the M-Max NLMS algorithm [3] with unity step size.

reduce the amount of memory required for implementation [6]. However, such an approach will no longer perform an update that minimizes the criterion in (1), resulting in slower convergence speed.

For a step size $\mu_k = \bar{\mu}\|\mathbf{A}_{\mathcal{I}_L(k)} \mathbf{x}_k\|^2/\|\mathbf{x}_k\|^2$, the PU-NLMS in (3) becomes identical to the M-Max NLMS algorithm of [3]. For $\bar{\mu} = 1$, the solution is the projection of the solution of the NLMS algorithm with unity step size onto the direction of $\mathbf{A}_{\mathcal{I}_L(k)} \mathbf{x}_k$, as illustrated in Figure 1. Furthermore, $\mu = \|\mathbf{A}_{\mathcal{I}_L(k)} \mathbf{x}_k\|^2/\|\mathbf{x}_k\|^2$ corresponds to the instantaneous estimate of $\mathrm{E}[r_k^2/\tilde{r}_k^2]$ which gives the fastest convergence, as observed in Appendix I.

## III. SIMULATION RESULTS

In this subsection, our analysis of the PU-NLMS algorithm is validated using a system-identification setup. The order of the plant chosen was $N = 50$, and the input signal was zero-mean Gaussian noise with $\sigma_x^2 = 1$. The signal-to-noise ratio ($SNR$) was set to 60 dB.

Figure 2 shows the learning curves for the case of $L = 5$, $L = 10$, and $L = 25$ coefficients in the partial update. The curves were obtained through averaging 100 trials. The step size for each value of $L$ was chosen such that convergence to the same level of misadjustment was achieved. The corresponding theoretical learning curves obtained from evaluating Equation (7) in the Appendix I were also plotted. As can be seen from the figure, the theoretical curves are very close to the simulations. Figure 3 shows the excess MSE as a function of $\mu$ ranging from $0.1\mu_{max}$ to $0.8\mu_{max}$ for different values of $L$, where $\mu_{max}$ is given by Equation (9) in Appendix I. Note that the axis is normalized with respect to the maximum step size $\mu_{max}$, which is different for each value $L$. The quantity $\mathrm{E}\left[\tilde{r}_k^2\right]$ needed for the calculation of $\mu_{max}$ was obtained through numerical integration. For $L = 5$, $L = 10$, and $L = 25$ the corresponding values were $\mathrm{E}\left[\tilde{r}_k^2\right] = 21.438$, $\mathrm{E}\left[\tilde{r}_k^2\right] = 32.232$, and $\mathrm{E}\left[\tilde{r}_k^2\right] = 43.860$, respectively. As can be seen from Figure 3, the theoretical results are very close to the simulations within the range of step sizes considered. Using step sizes larger than $0.8\mu_{max}$, resulted in poor accuracy or caused divergence. This is expected due to the approximations made in the analysis. However, only step sizes in the range $\mu \leq 0.5\mu_{max}$ are of practical interest because larger values will neither increase

convergence speed nor decrease misadjustment. This fact is illustrated in Figure 4, where the theoretical convergence curves were plotted for different values of $\mu$ using $L = 10$ and $N = 50$. Therefore, we may state that our theoretical analysis is able to predict very accurately the excess MSE for the whole range of practical step sizes.

In Figure 5 we compare our results (solid lines) with those provided by [8] (dashed lines) for the particular case where their algorithm is equal to the one presented in Section II of this paper. As seen from Figure 5, the results presented in [8] are not accurate even for reasonably high values of $L$, whereas Figure 3 shows that our analysis is accurate for a large range of $L$. This comes from the fact that in [8] order statistics was not applied in the analysis, resulting in poor estimates of $\mathrm{E}[\|\mathbf{A}_{\mathcal{I}_L(k)}\mathbf{x}_k\|^2]$ for most values of $L < (N + 1)$.
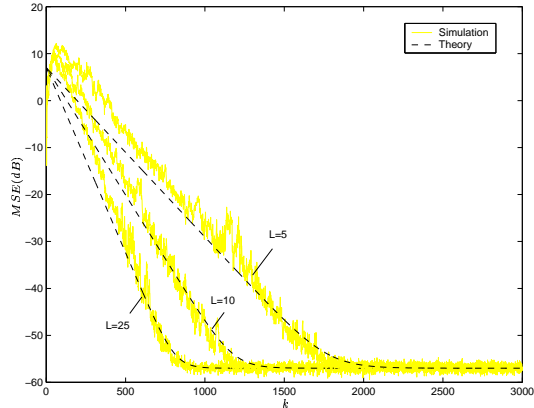


Fig. 4. Theoretical learning curves for different choice of step size in the PU-NLMS algorithm for $N = 50$ and $L = 10$, $SNR = 60$ dB.



Fig. 2. Learning curves for the PU-NLMS algorithm for $N = 50$, $L = 5$, $L = 10$ and $L = 25$, $SNR = 60$ dB.



Fig. 5. Comparison of Equation (7) (solid lines) with the excess MSE formula obtained from [8] (dashed lines).
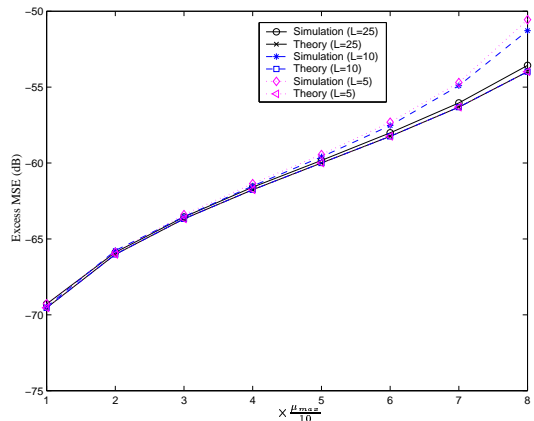


Fig. 3. Excess MSE for the PU-NLMS algorithm versus the step size $\mu$ for $N = 50$, $L = 5$, $L = 10$ and $L = 25$, $SNR = 60$ dB.

## IV. CONCLUSIONS

This paper studied normalized partial-update adaptation algorithms. Convergence analysis for the conventional partial-update NLMS (PU-NLMS) algorithm was presented, which gave further insight to the algorithm in terms of stability, transient and steady-state performances. The analysis was validated through simulations showing excellent agreement. New stabil-
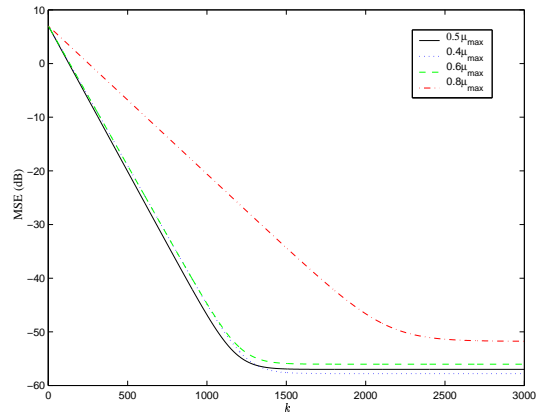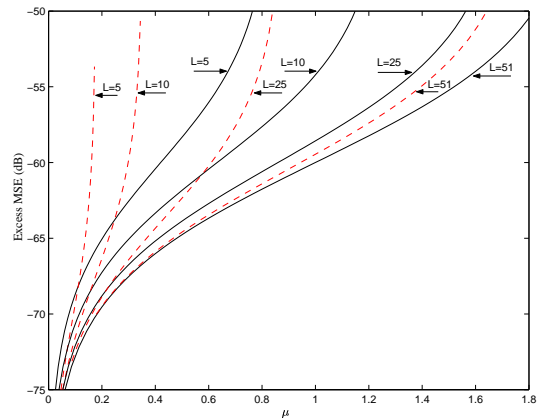
ity bounds were given for the step size that controls the stability, convergence speed, and final excess MSE of the PU-NLMS algorithm. It was shown that the step size giving the fastest convergence could be related to the time-varying step size of the M-Max NLMS algorithm. These results extend and improve in accuracy previous results reported in the literature. The excellent agreement between the theory and the simulations presented here for the PU-NLMS algorithm has advanced significantly the study of order-statistic-based adaptive filtering algorithms.

## APPENDIX I

In this appendix, the PU-NLMS algorithm is analyzed in the mean-squared sense.

### A. Coefficient error vector

In order to derive expressions for the second-order statistics of the PU-NLMS algorithm we will first derive an expression for the evolution of the coefficient-error vector. Assuming that the desired signal is given by $d_k = \mathbf{x}_k^{\mathrm{T}}\mathbf{w}_{\mathrm{opt}} + n_k$ and defining the coefficient error vector as $\Delta\mathbf{w}_k = \mathbf{w}_k - \mathbf{w}_{\mathrm{opt}}$, we can express the error as

$$e_k = n_k - \mathbf{x}_k^{\mathrm{T}}\Delta\mathbf{w}_k \tag{5}$$

Therefore, from Equations (3) and (5) we have

$$\Delta \mathbf{w}_{k+1} = \left[\mathbf{I} - \mu \frac{\mathbf{A}_{\mathcal{I}_L(k)}\mathbf{x}_k\mathbf{x}_k^\mathsf{T}}{\|\mathbf{A}_{\mathcal{I}_L(k)}\mathbf{x}_k\|^2}\right] \Delta \mathbf{w}_k + \mu \frac{n_k \mathbf{A}_{\mathcal{I}_L(k)}\mathbf{x}_k}{\|\mathbf{A}_{\mathcal{I}_L(k)\mathbf{x}_k}\mathbf{x}_k\|^2}$$

### B. Excess MSE for white input signals

For the MSE analysis, we assume that the vectors are excited in a discrete number of directions. This model was used to analyze the NLMS algorithm in [10] and is consistent with the first- and second-order statistics of the original input signal. The model was also successfully used to analyze the quasi-Newton (QN) [12] and the binormalized data-reusing LMS (BNDRLMS) [13] algorithms.

The following assumptions are made:

- Independence between $\mathbf{x}_k$ and $\Delta \mathbf{w}_k$.
- The vectors $\mathbf{x}_k$ and $\mathbf{A}_{\mathcal{I}_L(k)}\mathbf{x}_k$ are modeled by $\mathbf{x}_k = s_k r_k \mathbf{v}_k$ and $\mathbf{A}_{\mathcal{I}_L(k)}\mathbf{x}_k = \tilde{s}_k \tilde{r}_k \tilde{\mathbf{v}}_k$, respectively, where:
  - $\star$ $s_k$ and $\tilde{s}_k$ take on values $\pm 1$ with probability $1/2$.
  - $\star$ $r_k$ and $\tilde{r}_k$ are positive real valued stochastic variables such that $r_k^2$ and $\tilde{r}_k^2$ have the same probability distribution functions as $\|\mathbf{x}_k\|^2$ and $\|\mathbf{A}_{\mathcal{I}_L(k)}\mathbf{x}_k\|^2$, respectively.
  - $\star$ $\mathbf{v}_k$ is equal to one of the $N+1$ orthonormal eigenvectors of $\mathbf{R} = \mathrm{E}\left[\mathbf{x}_k\mathbf{x}_k^\mathsf{T}\right]$ denoted as $\mathcal{V}_k$, and $\tilde{\mathbf{v}}_k$ is equal to one of the $N+1$ orthonormal eigenvectors of $\tilde{\mathbf{R}} = \mathrm{E}\left[\mathbf{A}_{\mathcal{I}_L(k)}\mathbf{x}_k\mathbf{x}_k^\mathsf{T}\mathbf{A}_{\mathcal{I}_L(k)}^\mathsf{T}\right]$ denoted as $\tilde{\mathcal{V}}_k$. For white Gaussian input signals $\mathbf{v}_k$ and $\tilde{\mathbf{v}}_k$ are uniformly distributed and $\mathbf{R}$ and $\tilde{\mathbf{R}}$ share the same eigenvectors, i.e., $\mathcal{V}_k = \tilde{\mathcal{V}}_k$. Therefore,

$$P\left(\mathbf{v}_k = \mathcal{V}_k\right) = P\left(\tilde{\mathbf{v}}_k = \mathcal{V}_k\right) = \frac{1}{N+1} \qquad (6)$$

Notice that for any value of $L$ we have $s_k = \tilde{s}_k$ since the angle between $\mathbf{x}_k$ and $\mathbf{A}_{\mathcal{I}_L(k)}\mathbf{x}_k$ is always smaller than $90°$.

For white input signals, the excess MSE is given by $\Delta \xi_{k+1} = \sigma_x^2 \mathrm{tr}\left[\mathrm{cov}\left(\Delta \mathbf{w}_{k+1}\right)\right]$ [1], where

$$\mathrm{cov}\left[\Delta \mathbf{w}_{k+1}\right] = \mathrm{E}\left[\Delta \mathbf{w}_{k+1}\Delta \mathbf{w}_{k+1}^\mathsf{T}\right]$$

$$= \mathrm{E}\left[\left(\mathbf{I} - \mu \frac{\mathbf{A}_{\mathcal{I}_L(k)}\mathbf{x}_k\mathbf{x}_k^\mathsf{T}}{\|\mathbf{A}_{\mathcal{I}_L(k)}\mathbf{x}_k\|^2}\right)\Delta \mathbf{w}_k \right.$$
$$\left. \Delta \mathbf{w}_k^\mathsf{T}\left(\mathbf{I} - \mu \frac{\mathbf{A}_{\mathcal{I}_L(k)}\mathbf{x}_k\mathbf{x}_k^\mathsf{T}}{\|\mathbf{A}_{\mathcal{I}_L(k)}\mathbf{x}_k\|^2}\right)^\mathsf{T}\right]$$

$$+ \mathrm{E}\left[\mu^2 n_k^2 \frac{\mathbf{A}_{\mathcal{I}_L(k)}\mathbf{x}_k\mathbf{x}_k^\mathsf{T}\mathbf{A}_{\mathcal{I}_L(k)}^\mathsf{T}}{\|\mathbf{A}_{\mathcal{I}_L(k)}\mathbf{x}_k\|^4}\right]$$

$$= \mathrm{E}\left[\Delta \mathbf{w}_k \Delta \mathbf{w}_k^\mathsf{T}\right] - \mathrm{E}\left[\mu \frac{\Delta \mathbf{w}_k \Delta \mathbf{w}_k^\mathsf{T}\mathbf{x}_k\mathbf{x}_k^\mathsf{T}\mathbf{A}_{\mathcal{I}_L(k)}^\mathsf{T}}{\|\mathbf{A}_{\mathcal{I}_L(k)}\mathbf{x}_k\|^2}\right]$$

$$- \mathrm{E}\left[\mu \frac{\mathbf{A}_{\mathcal{I}_L(k)}\mathbf{x}_k\mathbf{x}_k^\mathsf{T}\Delta \mathbf{w}_k \Delta \mathbf{w}_k^\mathsf{T}}{\|\mathbf{A}_{\mathcal{I}_L(k)}\mathbf{x}_k\|^2}\right]$$

$$+ \mathrm{E}\left[\mu^2 \frac{\mathbf{A}_{\mathcal{I}_L(k)}\mathbf{x}_k\mathbf{x}_k^\mathsf{T}\Delta \mathbf{w}_k \Delta \mathbf{w}_k^\mathsf{T}\mathbf{x}_k\mathbf{x}_k^\mathsf{T}\mathbf{A}_{\mathcal{I}_L(k)}^\mathsf{T}}{\|\mathbf{A}_{\mathcal{I}_L(k)}\mathbf{x}_k\|^4}\right]$$

$$+ \mathrm{E}\left[\mu^2 n_k^2 \frac{\mathbf{A}_{\mathcal{I}_L(k)}\mathbf{x}_k\mathbf{x}_k^\mathsf{T}\mathbf{A}_{\mathcal{I}_L(k)}^\mathsf{T}}{\|\mathbf{A}_{\mathcal{I}_L(k)}\mathbf{x}_k\|^4}\right]$$

Let us analyze each term separately:

$$\psi_1 = \sigma_x^2 \mathrm{tr}\left\{\mathrm{E}\left[\Delta \mathbf{w}_k \Delta \mathbf{w}_k^\mathsf{T}\right]\right\} = \Delta \xi_k$$

$$\psi_2 = \sigma_x^2 \mathrm{tr}\left\{\mathrm{E}\left[\mu \frac{\Delta \mathbf{w}_k \Delta \mathbf{w}_k^\mathsf{T}\mathbf{x}_k\mathbf{x}_k^\mathsf{T}\mathbf{A}_{\mathcal{I}_L(k)}^\mathsf{T}}{\|\mathbf{A}_{\mathcal{I}_L(k)}\mathbf{x}_k\|^2}\right]\right\}$$

$$= \sigma_x^2 \mathrm{tr}\left\{\mathrm{E}\left[\mu \frac{\Delta \mathbf{w}_k \Delta \mathbf{w}_k^\mathsf{T} s_k r_k \tilde{s}_k \tilde{r}_k \mathbf{v}_k \tilde{\mathbf{v}}_k^\mathsf{T}}{\tilde{s}_k^2 \tilde{r}_k^2}\right]\right\}$$

$$= \sigma_x^2 \mathrm{E}\left[\mu \Delta \mathbf{w}_k^\mathsf{T}\mathbf{v}_k \tilde{\mathbf{v}}_k^\mathsf{T}\Delta \mathbf{w}_k\right]\mathrm{E}\left[\frac{s_k r_k \tilde{s}_k \tilde{r}_k}{\tilde{r}_k^2}\right]$$

$$= \frac{\mu}{N+1}\mathrm{E}\left[\frac{r_k}{\tilde{r}_k}\right]\Delta \xi_k$$

where we used Equation (6). Since $\mathrm{tr}\{\mathbf{AB}\} = \mathrm{tr}\{\mathbf{BA}\}$ we will have $\psi_2 = \psi_3$.

$$\psi_4 = \sigma_x^2 \mu^2 \mathrm{tr}\left\{\mathrm{E}\left[\frac{\mathbf{A}_{\mathcal{I}_L(k)}\mathbf{x}_k\mathbf{x}_k^\mathsf{T}\Delta \mathbf{w}_k \Delta \mathbf{w}_k^\mathsf{T}\mathbf{x}_k\mathbf{x}_k^\mathsf{T}\mathbf{A}_{\mathcal{I}_L(k)}^\mathsf{T}}{\|\mathbf{A}_{\mathcal{I}_L(k)}\mathbf{x}_k\|^4}\right]\right\}$$

$$= \sigma_x^2 \mu^2 \mathrm{E}\left[\frac{\Delta \mathbf{w}_k^\mathsf{T}\mathbf{x}_k\mathbf{x}_k^\mathsf{T}\mathbf{A}_{\mathcal{I}_L(k)}^\mathsf{T}\mathbf{A}_{\mathcal{I}_L(k)}\mathbf{x}_k\mathbf{x}_k^\mathsf{T}\Delta \mathbf{w}_k}{\|\mathbf{A}_{\mathcal{I}_L(k)}\mathbf{x}_k\|^4}\right]$$

$$= \sigma_x^2 \mu^2 \mathrm{E}\left[\frac{\Delta \mathbf{w}_k^\mathsf{T}\mathbf{x}_k\mathbf{x}_k^\mathsf{T}\Delta \mathbf{w}_k}{\|\mathbf{A}_{\mathcal{I}_L(k)}\mathbf{x}_k\|^2}\right]$$

$$= \sigma_x^2 \mu^2 \mathrm{E}\left[\Delta \mathbf{w}_k^\mathsf{T}\mathbf{v}_k \mathbf{v}_k^\mathsf{T}\Delta \mathbf{w}_k \frac{r_k^2}{\tilde{r}_k^2}\right]$$

$$= \mu^2 \frac{1}{N+1}\mathrm{E}\left[\frac{r_k^2}{\tilde{r}_k^2}\right]\Delta \xi_k$$

$$\psi_5 = \sigma_x^2 \mu^2 \mathrm{tr}\left\{\mathrm{E}\left[n_k^2 \frac{\mathbf{A}_{\mathcal{I}_L(k)}\mathbf{x}_k\mathbf{x}_k^\mathsf{T}\mathbf{A}_{\mathcal{I}_L(k)}^\mathsf{T}}{\|\mathbf{A}_{\mathcal{I}_L(k)}\mathbf{x}_k\|^4}\right]\right\}$$

$$= \sigma_x^2 \sigma_n^2 \mu^2 \mathrm{E}\left[\frac{1}{\tilde{r}_k^2}\right]$$

Finally we obtain the expression for the excess MSE

$$\Delta \xi_{k+1} \approx \psi_1 - \psi_2 - \psi_3 + \psi_4 + \psi_5$$

$$= \left\{1 - \frac{\mu}{N+1}\left(2\mathrm{E}\left[\frac{r_k}{\tilde{r}_k}\right] - \mu \mathrm{E}\left[\frac{r_k^2}{\tilde{r}_k^2}\right]\right)\right\}\Delta \xi_k$$

$$\mu^2 \sigma_x^2 \sigma_n^2 \mathrm{E}\left[\frac{1}{\tilde{r}_k^2}\right]$$

which can be approximated as

$$\Delta \xi_{k+1} \approx \left\{1 - \frac{\mu}{N+1}\left(2 - \mu \mathrm{E}\left[\frac{r_k^2}{\tilde{r}_k^2}\right]\right)\right\}\Delta \xi_k$$

$$+ \mu^2 \sigma_x^2 \sigma_n^2 \mathrm{E}\left[\frac{1}{\tilde{r}_k^2}\right] \qquad (7)$$

where the conservative approximation $\mathrm{E}\left[\frac{r_k}{\tilde{r}_k}\right] \approx 1$ was used. The stability region in the mean-squared sense for $\mu$ is

$$0 < \mu < \frac{2}{\mathrm{E}\left[\frac{r_k^2}{\tilde{r}_k^2}\right]} \qquad (8)$$

where the step size $\mu = 1/\mathrm{E}\left[\frac{r_k^2}{\tilde{r}_k^2}\right]$ yields maximum reduction of $\Delta\xi_k$ in (7). Further simplifications with $\mathrm{E}\left[\frac{r_k^2}{\tilde{r}_k^2}\right] \approx \frac{\mathrm{E}\left[r_k^2\right]}{\mathrm{E}\left[\tilde{r}_k^2\right]}$ give us

$$0 < \mu < \frac{2\mathrm{E}\left[\tilde{r}_k^2\right]}{(N+1)\sigma_x^2} \tag{9}$$

where $\mathrm{E}\left[r_k^2\right] = (N+1)\sigma_x^2$ and $\mathrm{E}\left[\tilde{r}_k^2\right]$ can be calculated using knowledge of $L$ and $N$ using order statistics (see also Appendix II). A more pessimistic bound can be obtained by using the relation $\mathrm{E}\left[\tilde{r}_k^2\right] \geq L\sigma_x^2$ (see Appendix II) giving

$$0 < \mu < 2 \cdot \frac{L}{N+1} \tag{10}$$

which corresponds to the bound given in [8]. We stress that the analysis presented in this appendix shows that step sizes larger than the ones indicated by Equation (10) may be used according to Equation (9).

For $k \to \infty$ we have

$$\Delta\xi_{exc} \approx (N+1)\frac{\mu\sigma_n^2\sigma_x^2}{2 - \mu\mathrm{E}\left[\frac{r_k^2}{\tilde{r}_k^2}\right]}\mathrm{E}\left[\frac{1}{\tilde{r}_k^2}\right]$$

$$\approx (N+1)\frac{\mu\sigma_n^2\sigma_x^2}{2\mathrm{E}\left[\tilde{r}_k^2\right] - \mu\mathrm{E}\left[r_k^2\right]}$$

$$= (N+1)\frac{\mu\sigma_n^2\sigma_x^2}{2\mathrm{E}\left[\tilde{r}_k^2\right] - \mu(N+1)\sigma_x^2}.$$

### APPENDIX II

In this Appendix it is shown how to obtain numerically $\mathrm{E}\left[\tilde{r}_k^2\right]$ used in the step size bound derived in Appendix I. In addition, a lower bound on $\mathrm{E}\left[\tilde{r}_k^2\right]$ is provided. This parameter was also required in the analysis of the M-Max NLMS algorithm [4], which used the approach as presented here.

The basic problem here is to calculate the second moment of ordered statistics. This problem has received much attention in the past, see, e.g., [14]–[16], where recursion formulas and tables were produced for expected values and moments of ordered statistics for various different distributions.

Let $\mathbf{y} = [y_1\ y_2\ \ldots\ y_{N+1}]^\mathrm{T}$ be a vector containing the elements of vector $\mathbf{x}_k = [x_k\ x_{k-1}\ \ldots\ x_{k-N}]^\mathrm{T}$ ordered in value, i.e, $y_1 \leq y_2 \leq \ldots \leq y_j \leq \ldots \leq y_{N+1}$. The probability density function $f_j(y)$ of the $j$th element in $\mathbf{y}$ is given by [17]

$$f_j(y) = \frac{(N+1)!F_x^{j-1}(y)\left[1 - F_x(y)\right]^{N+1-j}f_x(y)}{(j-1)!(N+1-j)!}$$

where $f_x(x)$ is the density of the unsorted random variables in vector $\mathbf{x}_k$ and $F_x^{j-1}(x)$ is their cumulative distribution to the power of $j - 1$. The second moment of the $j$th element is given by

$$\mathrm{E}\left[y_j^2\right] = \int_{-\infty}^{\infty} y_j^2 f_j(y)dy$$

$$= \frac{(N+1)!}{(j-1)!(N+1-j)!} \times$$
$$\int_{-\infty}^{\infty} y^2 F_x^{j-1}(y)\left[1 - F_x(y)\right]^{N+1-j}f_x(y)dy \tag{11}$$

The PU-NLMS algorithm chooses the $L$ elements in $\mathbf{x}_k$ of largest magnitude. Therefore, if we order the values in $\mathbf{A}_{\mathcal{I}_L(k)}\mathbf{x}_k$ in magnitude their second moments can be found by evaluating (11) for $j = N + 2 - L,\ \ldots,\ N + 1$. For the case of Gaussian input signals and using the cumulative distribution and density functions for the magnitude of a Gaussian variable, we have

$$F_x(y) = \begin{cases} 2\Phi_x(y) - 1 & \text{for } y \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{12}$$

and

$$f_x(y) = \begin{cases} 2\phi_x(y) & \text{for } y \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{13}$$

where $\Phi_x(y)$ and $\phi_x(y)$ are the cumulative distribution function and the density function, respectively, of a Gaussian variable. The density function $f_x(y)$ in (13) is in fact the probability density function for a random variable from a chi-distribution with one degree of freedom. The problem of calculating moments of order statistics in samples from the chi-distribution (1 degree of freedom) was considered in [16], where a recursion formula was developed. The quantity $\mathrm{E}\left[\tilde{r}_k^2\right]$ is given by

$$\mathrm{E}\left[\tilde{r}_k^2\right] = \sum_{j=N+2-L}^{N+1} \frac{2(N+1)!}{(j-1)!(N+1-j)!} \times$$
$$\int_0^\infty y^2 \left(2\Phi_x(y) - 1\right)^{j-1}\left(2 - 2\Phi_x(y)\right)^{N+1-j}\phi_x(y)dy$$

which for given $N$ and $L$ can be evaluated numerically.

With the aid of the previous results we are able to calculate bounds for $\mathrm{E}[\tilde{r}_k^2]$, as stated in the following lemma.

**Lemma 1.** *If the input signal $x_k$ is Gaussian with zero mean and variance $\sigma_x^2$, then $\mathrm{E}[\tilde{r}_k^2] = \sum_{j=N+2-L}^{N+1}\mathrm{E}[y_j^2]$, where $L \leq N + 1$, is bounded as follows:*

$$L\sigma_x^2 \leq \mathrm{E}[\tilde{r}_k^2] \leq (N+1)\sigma_x^2$$

*with equality iff $L = N + 1$.*

*Proof:* In the proof we need the following relations

$$\mathrm{E}[y_1^2] \leq \mathrm{E}[y_2^2] \leq \cdots \leq \mathrm{E}[y_k^2] \leq \cdots \leq \mathrm{E}[y_{N+1}^2] \tag{14}$$

$$\sum_{k=1}^{N+1}\mathrm{E}[y_k^2] = (N+1)\mathrm{E}[x_k^2] \tag{15}$$

$$\mathrm{E}[y_1^2] < \sigma_x^2, \text{ for } N > 0 \tag{16}$$

Relation (14) holds true by definition, and (15) holds true for an arbitrary distribution for which the integral in (11) con-

verges [16], as shown below

$$\sum_{j=1}^{N+1} \mathrm{E}[y_j^2] = \sum_{j=1}^{N+1} \frac{(N+1)!}{(j-1)!(N+1-j)!} \times$$

$$\int_{-\infty}^{\infty} y^2 F_x^{j-1}(y) \left[1 - F_x(y)\right]^{N+1-j} f_x(y) dy$$

$$= \int_{-\infty}^{\infty} \left\{ \sum_{j=1}^{N+1} \frac{(N+1)!}{(j-1)!(N+1-j)!} F_x^{j-1}(y) \right.$$

$$\left. \left[1 - F_x(y)\right]^{N+1-j} \right\} y^2 f_x(y) dy$$

$$= \int_{-\infty}^{\infty} \left\{ (N+1) \sum_{k=0}^{N} \frac{N!}{k!(N-k)!} F_x^k(y) \times \right.$$

$$\left. \left[1 - F_x(y)\right]^{N-k} \right\} y^2 f_x(y) dy$$

$$= \int_{-\infty}^{\infty} (N+1) y^2 f_x(y) dy$$

$$= (N+1)\mathrm{E}[x_k^2]$$

where we used $\sum_{k=0}^{N} \binom{N}{k} p^k q^{N-k} = (p+q)^N$. The relation (16) can be shown for Gaussian input signals by evaluating (11) for $j = 1$ with $F_x(y)$ and $f_x(y)$ given by Equations (12) and (13):

$$\mathrm{E}[y_1^2] = \int_0^{\infty} 2(N+1) y^2 \left[2 - 2\Phi_x(y)\right]^N \phi_x(y) dy$$

$$= \int_0^{\infty} 2^{N+1}(N+1) y^2 \left[1 - \Phi_x(y)\right]^N \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{\frac{-y^2}{2\sigma_x^2}} dy$$

$$\leq \int_0^{\infty} 2^{N+1}(N+1) y^2 \left[\frac{1}{2} e^{\frac{-y^2}{2\sigma_x^2}}\right]^N \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{\frac{-y^2}{2\sigma_x^2}} dy$$

$$= \int_0^{\infty} 2(N+1) y^2 \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{\frac{-(N+1)y^2}{2\sigma_x^2}} dy$$

$$= \frac{\sigma_x^2}{\sqrt{N+1}} < \sigma_x^2 \text{ for } N > 0.$$

where we used $1 - \Phi_x(y) \leq \frac{1}{2} e^{\frac{-y^2}{2\sigma_x^2}}$ for $y \geq 0$ [17], and $\int_0^{\infty} y^2 e^{-ay^2} dy = \frac{1}{4a}\sqrt{\frac{\pi}{a}}$. From relations (14) and (16) it follows that $L\sigma_x^2 \leq \mathrm{E}[\tilde{r}_k^2] \leq (N+1)\sigma_x^2$ holds true for $L \leq N+1$. Relation (16) gives us strict inequality for $L < N+1$ when $N > 0$, and consequently equality holds true only when $L = N+1$. $\qquad\square$

## REFERENCES

[1] P. S. R. Diniz, *Adaptive Filtering — Algorithms and Practical Implementation*, Boston, MA: Kluwer Academic Press, 1997.

[2] S. C. Douglas, "Adaptive filters employing partial updates," *IEEE Transaction Circuits and Systems-II*, vol. 44, pp. 209–216, Mar. 1997.

[3] T. Aboulnasr and K. Mayyas, "Complexity reduction of the NLMS algorithm via selective coefficient update," *IEEE Transaction Signal Processing*, vol. 47, pp. 1421–1427, May 1999.

[4] M. I. Haddad, K. Mayyas, and M. A. Khasawneh, "Analytical development of the MMAXNLMS algorithm," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP'99*, Phoenix, USA, Mar. 1999, vol. 4, pp. 1853–1856.

[5] M. Godavarti and A. O. Hero III, "Stochastic partial update LMS algorithm for adaptive arrays," *IEEE Sensor Array and Multichannel Signal Processing Workshop*, Boston, USA, pp. 322-326, Mar. 2000.

[6] T. Schertler, "Selective block update of NLMS type algorithms," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP'98*, Seattle, USA, May 1998, vol. 3, pp. 1717–1720.

[7] S. Attallah and S. W. Liaw, "DCTLMS algorithm employing partial coefficient updates," *IEEE Adaptive Systems for Signal Processing, Communications, and Control Symposium 2000*, pp. 218–223, Oct. 2000.

[8] K. Doğançay and O. Tanrıkulu, "Adaptive Filtering algorithms with selective partial update," *IEEE Transaction Circuits and Systems-II*, vol. 48, pp. 762–769, Aug. 2001.

[9] K. Doğançay and O. Tanrıkulu, "Selective-partial-update NLMS and affine projection algorithms for acoustic echo cancellation," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP'2000*, Istanbul, Turkey, Jun. 2000, vol. 1, pp. 448–451.

[10] D. T. M. Slock, "On the convergence behavior of the LMS and the normalized LMS algorithms," *IEEE Transactions on Signal Processing*, vol. 41, pp. 2811–2825, Sep. 1993.

[11] J. Pitas, "Fast algorithms for running ordering and max/min calculation," *IEEE Transaction Circuits and Systems-II*, vol. 36, pp. 795–804, Jun. 1989.

[12] M. L. R. de Campos and A. Antoniou, "A new quasi-Newton adaptive filtering algorithm," *IEEE Transactions on Circuits and Systems — Part II*, vol. 44, pp. 924–934, Nov. 1997.

[13] J. A. Apolinário, M. L. R. de Campos and P. S. R. Diniz, "Convergence analysis of the binormalized data-reusing LMS algorithm," *IEEE Transaction on Signal Processing*, vol. 48, pp. 3235–3242, Nov. 2000.

[14] H. J. Godvin "Some low moments of order statistics," *Annals of Mathematical Statistics*, vol. 20, pp. 279–285, Jun. 1949.

[15] H. L. Harter "Expected values of normal order-statistics," *Biometrika*, vol. 48, pp. 151–165, Jun. 1961.

[16] Z. Govindarajulu, "Exact moments of order statistics in samples from the chi-distribution (1 d.f)," *Annals of Mathematical Statistics*, vol. 33, pp. 1292–1305, Dec. 1963.

[17] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, 1991.