DATA-SELECTIVE LMS-NEWTON AND LMS-QUASI-NEWTON ALGORITHMS

*Christos G. Tsinos** and *Paulo S. R. Diniz*⋆

* SnT-Interdisciplinary Centre for Security, Reliability and Trust, Univ. of Luxembourg, Luxembourg

⋆ Signals, Multimedia, and Telecommun. Lab, Univ. Federal do Rio de Janeiro DEL/Poli & PEE/COPPE/UFRJ, Rio de Janeiro, Brazil.

e-mails: christos.tsinos@uni.lu, diniz@smt.ufrj.br

ABSTRACT

The huge volume of data that are available today requires data-selective processing approaches that avoid the costs in computational complexity via appropriately treating the non-innovative data. In this paper, extensions of the well-known adaptive filtering LMS-Newton and LMS-Quasi-Newton Algorithms are developed that enable data selection while also addressing the censorship of outliers that emerge due to high measurement errors. The proposed solutions allow the prescription of how often the acquired data are expected to be incorporated into the learning process based on some a priori information regarding the environment. Simulation results on both synthetic and real-world data verify the effectiveness of the proposed algorithms that may achieve significant reductions in computational costs without sacrificing estimation accuracy due to the selection of the data.

Index Terms— Adaptive Signal Processing, LMS-Newton, Adaptive Filters, Learning Systems, Data Processing

1. INTRODUCTION

In the era of big data, the processing of the acquired datasets will demand huge computational load if an effective strategy is not followed. To that end, data-selective processing techniques have been recently developed with the view to reduce the required computational complexity by enabling the processing for only innovation bearing new data. Towards that direction, a number of different traditional online learning and adaptive processing schemes have been enhanced with data-selective capabilities [1–8]. The proposed solutions may achieve similar performance to the one of the non-data-selective versions by processing only a very small portion of the available data. Solutions that treat the presence of outliers have also been considered.

In this work, data-selective algorithms are developed for the Least Mean Square-Newton (LMSN) and LMS-Quasi-Newton (LMSQN) ones that have yet to appear in the literature. A number of different LMSN-based algorithms have been developed throughout the past 25 years [9–16]. They are powerful alternatives to the classical LMS algorithm whose performance is highly dependent on the statistics of the input signal. Thus, despite their increased complexity (compared to the LMS one), they offer significantly improved performance in cases, as an example, where the spread of the eigenvalues of the input-signal correlation matrix is large, similarly to the well-known Recursive Least Squares (RLS) algorithm [11]. Furthermore, they give rise to the LMSQN counterparts that exhibit the desirable characteristics of the LMSN/RLS algorithms with reduced requirements in Floating Point Operations (FLOPS) per update iteration [10, 12, 13] and improved robustness to quantization errors [14].

To that end, a data-selective version of the classical LMSN [11] and the LMSQN in [14] are developed in the present work. The proposed algorithms incorporate only innovative data to the adaptation process while discarding possible outliers and non-innovative information. In literature so far there are works that study the robustness behavior of some adaptive algorithms [17–21], though the proposed approaches are not related to the one presented in the paper. Here, the aim is to discard the outliers from the learning process, as discussed above, and not to provide solutions that help the adaptive algorithms to recover from outlier samples, as in [17–21].

The data selection process is achieved by establishing a connection between two thresholds, used to classify the data quality as non-innovative/innovative/outliers and a prescribed probability of update. The latter probability of update is further connected to the Mean Square Error (MSE) performance metric [22] of the algorithm under consideration. The performance of the proposed approaches is evaluated via simulations on synthetic and real-world data, as well. The results show on both the cases that the data-selective solutions achieve very close performance to the ones of the corresponding non-data-selective counterparts while being much more efficient due to the computational complexity reduction from their data selection capability. Furthermore, they do not allow the incorporation of outliers in the learning process, and thus, avoiding a possible performance degradation.

The rest of this paper is organized as follows. Sec. 2 describes the considered system model. In Sec. 3, the concept of LMSN and LMSQN algorithms are described. Sec. 4 presents the data-selective LMSN and LMSQN algorithms. Sec. 5 presents some indicative simulation results, and Sec. 6 concludes this work.

2. SYSTEM MODEL

In this paper, we consider learning algorithms for the problem of linear system identification. The unknown system is excited with a known real signal $x(k)$, where k is the time index. The observed signal at the output of the system is given by,

$$d(k) = \mathbf{w}_o^T \mathbf{x}(k) + n(k), \tag{1}$$

where $\mathbf{w}_o \in \mathbb{R}^{L+1}$ is the $(L+1) \times 1$ unknown system to be identified, $\mathbf{x}(k) = [x(k)\, x(k-1)\, \ldots\, x(k-L+1)]^T$, $n(k)$ is a Gaussian noise sample of zero mean and variance σ_n^2, superscript T stands for the transpose of vector/matrix and $\mathbb{R}$ is the set of real numbers.

This work was supported by FNR, Luxembourg under the project CORE ECLECTIC. Furthermore, this study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nivel Superior - Brasil (CAPES) - Finance Code 001 and also supported by the research councils: CNPq (Conselho Nacional de Desenvolvimento Cientfico e Tecnolgico), and FAPERJ (Fundao de Amparo Pesquisa do Estado do Rio de Janeiro). The authors would like to thank Ms. M. Mendoça and Mr. J. Ferreira of COPPE/UFRJ for their help in generating the simulation results based on real-world data-set.

Let us assume that the filtering algorithm generates an output signal estimation through $\mathbf{w}^T(k)\mathbf{x}(k)$, where $\mathbf{w}(k)$ is the $(L+1)\times 1$ vector whose entries are the estimation of the unknown system's parameters at time index k. Then, the error estimation sequence $e(k)$ is defined as

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

for $k = 0, 1, \ldots, \infty$.

The mean square error for the discussed model at time instance k can be expressed as [22] ,

$$\begin{aligned}\xi(k) =&\mathbb{E}\{e^2(k)\} = \mathbb{E}\{n^2(k)\} - 2\mathbb{E}\{n(k)\Delta\mathbf{w}^T(k)\mathbf{x}(k)\}+\\ &\mathbb{E}\{\Delta\mathbf{w}^T(k)\mathbf{x}(k)\mathbf{x}^T(k)\Delta\mathbf{w}(k)\},\end{aligned} \tag{3}$$

where $\Delta\mathbf{w}(k) = \mathbf{w}(k) - \mathbf{w}_o$ and $\mathbb{E}\{\cdot\}$ denotes the expectation operator. By assuming that the additive noise variables and the coefficients are uncorrelated results in

$$\begin{aligned}\xi(k) =&\sigma_n^2 + \mathbb{E}\{\Delta\mathbf{w}^T(k)\mathbf{x}(k)\mathbf{x}^T(k)\Delta\mathbf{w}(k)\}\\ =&\sigma_n^2 + \xi_{exc}(k),\end{aligned} \tag{4}$$

where $\xi_{exc}(k)$ denotes the excess MSE and $\mathbb{E}\{n^2(k)\} = \sigma_n^2$. Eq. (4) is important for the data-selective algorithms since it is used for prescribing the desired probability of update, as shown in Sec. 4.

3. LMSN AND LMSQN ALGORITHMS

In this section, the LMSN and LMSQN algorithms are described. Such algorithms aim at minimizing the cost function

$$J(\mathbf{w}(k)) = \frac{1}{2}|e(k)|^2, \tag{5}$$

where $|\cdot|$ denotes the absolute value. This is achieved by performing recursive coefficient updates in the direction of the negative estimated gradient $\nabla_{\mathbf{w}(k)}J(\mathbf{w}(k)) = -\mathbf{x}(k)\tilde{e}(k)$ pre-multiplied by the inverse of the estimated input signal autocorrelation matrix $\hat{\mathbf{R}}^{-1}(k)$ as [11]

$$\mathbf{w}(k) = \mathbf{w}(k-1) + \frac{\mu}{\mathbf{x}^T(k)\hat{\mathbf{R}}^{-1}(k)\mathbf{x}(k)}\hat{\mathbf{R}}^{-1}(k)\mathbf{x}(k)\tilde{e}(k), \tag{6}$$

where μ is a step-size parameter, $\tilde{e}(k) = d(k) - \mathbf{w}^T(k-1)\mathbf{x}(k)$ is the a priori estimation error and $\hat{\mathbf{R}}(k)$ is the estimation of the input signal autocorrelation matrix, defined as $\mathbf{R} = \mathbb{E}\{\mathbf{x}(k)\mathbf{x}^T(k)\}$.

The difference between the LMSN and the LMSQN algorithms considered here is related to the way the estimate $\hat{\mathbf{R}}^{-1}(k)$ is calculated. For the LMSN method, matrix $\hat{\mathbf{R}}(k)$ is estimated via a Robbins-Monro procedure that results in the following update of its inverse, given by [11]

$$\begin{aligned}\hat{\mathbf{R}}^{-1}(k) =&\frac{1}{1-\alpha}\Bigg\{\hat{\mathbf{R}}^{-1}(k-1)\\ &-\frac{\hat{\mathbf{R}}^{-1}(k-1)\mathbf{x}(k)\mathbf{x}^T(k)\hat{\mathbf{R}}^{-1}(k-1)}{\frac{1-a}{a} + \mathbf{x}(k)^T\hat{\mathbf{R}}^{-1}(k)\mathbf{x}(k)}\Bigg\},\end{aligned} \tag{7}$$

where α is a step-size parameter. For the LMSQN algorithm considered in the present paper, the approach in [14] is used to update $\hat{\mathbf{R}}^{-1}(k)$, that is

$$\begin{aligned}\hat{\mathbf{R}}^{-1}(k) =&\frac{1}{1-\alpha}\Bigg\{\hat{\mathbf{R}}^{-1}(k-1) + \left(\frac{\mu}{2\mathbf{x}(k)^T\hat{\mathbf{R}}^{-1}(k)\mathbf{x}(k)} - 1\right)\\ &\times\frac{\hat{\mathbf{R}}^{-1}(k-1)\mathbf{x}(k)\mathbf{x}^T(k)\hat{\mathbf{R}}^{-1}(k-1)}{\mathbf{x}(k)^T\hat{\mathbf{R}}^{-1}(k)\mathbf{x}(k)}\Bigg\}.\end{aligned} \tag{8}$$

The performance of each one of the approaches has been theoretically analyzed and extensively studied via simulations in [11] and [14], respectively. It is noteworthy to mention that the LMSQN approach appears to be very robust to quantization errors that in general tend to degrade the performance of the LMSN and RLS algorithms significantly [14].

4. DATA-SELECTIVE APPROACHES

A way to introduce the data selection property to the LMSN and LMSQN algorithms is to perform an update of the coefficients only if $|e(k)|^2$ is greater than a scaled noise power level, i.e., $\tau(k)\sigma_n^2$ for some threshold $\tau(k)$. Correspondingly, if $|e(k)|^2$ is greater than $\tau_{max}\sigma_n^2$ for again some threshold τ_{max}, an outlier is identified and thus, no update is either performed. These can be translated to the minimization of the function

$$J'(\mathbf{w}(k)) = \begin{cases}\frac{1}{2}|e(k)|^2, \text{ if } \sqrt{\tau(k)} \leq \frac{|e(k)|}{\sigma_n} < \sqrt{\tau_{max}}\\ 0, \text{ otherwise.}\end{cases} \tag{9}$$

The negative subgradient of this nondifferentiable objective function, is given by,

$$\partial J'(\mathbf{w}(k)) = \begin{cases}-x(k)e(k),\ \sqrt{\tau(k)} < \frac{|e(k)|}{\sigma_n} < \sqrt{\tau_{max}}\\ \{-\theta x(k)e(k) : 0 \leq \theta \leq 1\},\ \frac{|e(k)|}{\sigma_n} = \sqrt{\tau(k)}\\ 0,\ \frac{|e(k)|}{\sigma_n} < \sqrt{\tau(k)}\\ \{-\phi x(k)e(k) : 0 \leq \phi \leq 1\},\ \frac{|e(k)|}{\sigma_n} = \sqrt{\tau_{max}}\\ 0,\ \frac{|e(k)|}{\sigma_n} > \sqrt{\tau_{max}}.\end{cases} \tag{10}$$

By using (10), it can be shown that the updates of the coefficients for the data-selective LMSN and LMSQN algorithms are given by,

$$\mathbf{w}(k) = \begin{cases}\mathbf{w}(k-1) + \mu\frac{\hat{\mathbf{R}}^{-1}(k)\mathbf{x}(k)\tilde{e}(k)}{\mathbf{x}^T(k)\hat{\mathbf{R}}^{-1}(k)\mathbf{x}(k)},\ \sqrt{\tau(k)} \leq \frac{|e(k)|}{\sigma_n} < \sqrt{\tau_{max}}\\ \mathbf{w}(k-1),\ \text{otherwise.}\end{cases} \tag{11}$$

The data-selective strategy may also be adopted for the update of matrix $\hat{\mathbf{R}}^{-1}$, as well. That is, $\hat{\mathbf{R}}^{-1}$ is updated via (7) and (8) , respectively provided that $\sqrt{\tau(k)} < \frac{|e(k)|}{\sigma_n} < \sqrt{\tau_{max}}$, holds. Alternatively, if outliers are not in the input signal, it is possible to update the $\hat{\mathbf{R}}^{-1}$ in all iterations, as long as the computational burden related to it is not an issue. In this paper, the former approach is followed.

Let us now assume, that the desired probability of coefficient updates based on the upper branch in (11) is denoted by $P_{up}(k)$. This probability can be further modeled as

$$P_{up}(k) = P\left\{\frac{|e(k)|}{\sigma_n} > \sqrt{\tau(k)}\right\} - P\left\{\frac{|e(k)|}{\sigma_n} > \sqrt{\tau_{max}}\right\}, \tag{12}$$

where $P\{\cdot\}$ is the probability of an event. In the steady state performance of the algorithm and under the assumption of white Gaussian input signals, (12) may be written as,

$$P_{up} = 2Q\left(\frac{\sigma_n\sqrt{\tau}}{\sigma_e}\right) - 2Q\left(\frac{\sigma_n\sqrt{\tau_{max}}}{\sigma_e}\right), \tag{13}$$

where $Q(\cdot)$ is the complementary Gaussian cumulative distribution function, given by $Q(x) = 1/(2\pi)\int_x^\infty exp(-t^2/2)dt$ [23] and σ_e^2 represents the error signal variance. Note that in (13), the dependence on index k was silently dropped under the assumption of stationarity. In other words, in the steady state, it is now assumed that

Algorithm 1 Data-selective LMSN and LMSQN Algorithms

1: Inputs: $0 < \mu \leq 1$, $0 < \alpha \leq 1$ (for LMSN), γ small positive value, P_{up} and τ_{max}
2: Initialize $\mathbf{w}(0) = \mathbf{0}_{L+1}$ and $\hat{\mathbf{R}}^{-1}(0) = \gamma \mathbf{I}_{L+1}$
3: Set $\beta = \frac{\mu P_{up}}{2 - \mu P_{up}}$
4: Calculate τ from (13), if outliers are present or from (16), otherwise
5: **for** $k = 1, 2, \ldots$ **do**
6: Acquire $\mathbf{x}(k)$ and $d(k)$
7: $e(k) = d(k) - \mathbf{w}^T(k)\mathbf{x}(k)$
8: **if** $\sqrt{\tau}\sigma_n \leq |e(k)| \leq \sqrt{\tau_{max}}\sigma_n$ **then**
9: $\mathbf{t}(k) \leftarrow \hat{\mathbf{R}}^{-1}(k)\mathbf{x}(k)$
10: $\psi(k) \leftarrow \mathbf{x}^T(k)\mathbf{t}(k)$
11: $\mathbf{w}(k+1) \leftarrow \mathbf{w}(k) + \mu \frac{\mathbf{t}(k)e(k)}{\psi(k)}$
12: $\hat{\mathbf{R}}^{-1}(k+1) \leftarrow \frac{1}{1-a}\left[\hat{\mathbf{R}}^{-1}(k) - \frac{\mathbf{t}(k)\mathbf{t}^T(k)}{\frac{1-\alpha}{\alpha} + \psi(k)}\right]$, for LMSN
$\hat{\mathbf{R}}^{-1}(k+1) \leftarrow \hat{\mathbf{R}}^{-1}(k) + \frac{\frac{\mu}{2\psi(k)} - 1}{\psi(k)}\mathbf{t}(k)\mathbf{t}^T(k)$, for LMSQN
13: **else if** $|e(k)| \leq \sqrt{\tau}\sigma_n$ **then**
14: $\mathbf{w}(k+1) \leftarrow \mathbf{w}(k)$
15: **else if** $|e(k)| \geq \sqrt{\tau_{max}}\sigma_n$ **then**
16: $\mathbf{w}(k+1) \leftarrow \mathbf{w}(k)$, $e(k) = 0$, $d(k) = 0$
17: **end if**
18: **end for**

$\tau(k)$ is fixed at τ. For the case that no outliers are expected, the second term of the difference in (13) is dropped.

It remains to calculate σ_e^2 in order to prescribe the desired probability of update through (13). It can be seen that under the considered assumptions $\mathbb{E}\{e(k)\} = 0$ and thus, $\sigma_e^2 = \xi(\infty)$ where $\xi(\infty)$ is the steady-state MSE for the algorithms under consideration. By (4), in order to calculate $\xi(\infty)$, we need the expression of the steady-state excess MSE, $\xi_{exc}(\infty)$, specified by the following proposition.

Proposition 1: The excess mean square error at the steady-state of data-selective LMSN and LMSQN algorithms that follow the update rule in (11) can be approximated by

$$\xi_{exc}(\infty) = \frac{\mu P_{up}}{2 - \mu P_{up}} \sigma_n^2. \tag{14}$$

Sketch of the proof: Both data-selective algorithms, the LMSN and LMSQN, utilize an estimate of the inverse correlation matrix that should converge to the actual $\mathbf{R}^{-1}$. In both cases, the general expression for their coefficient update can be represented by an equivalent expression given by (15)

$$\mathbf{w}(k) = \mathbf{w}(k-1) + \frac{\mu P_{up}}{\mathbf{x}^T(k)\hat{\mathbf{R}}^{-1}(k)\mathbf{x}(k)} \hat{\mathbf{R}}^{-1}(k)\mathbf{x}(k)\tilde{e}(k). \tag{15}$$

For this type of update equation, using similar analysis procedures, the result of (14) can be derived, see [11] and [24], for details. $\square$

Having specified the excess MSE of the data-selective algorithms, we may calculate the required thresholds by plugging its expression in (13). If no outliers are present in the data, the value of parameter τ that prescribes the desired P_{up} can be calculated by

$$\sqrt{\tau} = \sqrt{1 + \beta}\, Q^{-1}(0.5 P_{up}), \tag{16}$$

where $\beta = \frac{\mu P_{up}}{2 - \mu P_{up}}$ and $Q^{-1}(\cdot)$ is the inverse of the $Q(\cdot)$ function. For the case of outliers, the threshold should be selected according to some prior information regarding the signal sources and supporting circuitry such as overflow levels [25]. The steps of both the algorithms are given in Algorithm 1, where the quantities $\mathbf{t}(k)$ and $\psi(k)$ in lines 9 and 10, respectively are used for minimizing the required computations. As it is evident also from the analysis so far, the proposed data-selective solutions require knowledge of the noise power. Such information may be estimated as discussed in [25].

We close this section with a discussion regarding the complexity of the proposed data-selective approaches. By analyzing the complexity of the original algorithms that do not possess data selection capabilities, it can be shown that the LMSN algorithm requires $C_{LMSN} = 6L^2 + 19L + 15$ FLOPS while the LMSQN one requires $C_{LMSQN} = 5L^2 + 17L + 15$ FLOPS per iteration, respectively. The average computational complexity per iteration of the data-selective LMSN and LMSQN algorithms is $P_{up}C_{LMSN}$ and $P_{up}C_{LMSQN}$, respectively. This is the case since the data-selective algorithms are performing updates with average probability equal to P_{up}.

5. SIMULATIONS

In this section, simulations are presented on synthetic and real-world data in order to evaluate the performance of the data-selective LMSN and LMSQN algorithms. In all the synthetic experiments, a system identification problem is examined for an unknown system with impulse response given by,

$$[0.1010\ 0.3030\ 0\ -0.2020\ -0.4040\\ -0.7071\ -0.4040\ -0.2020].$$

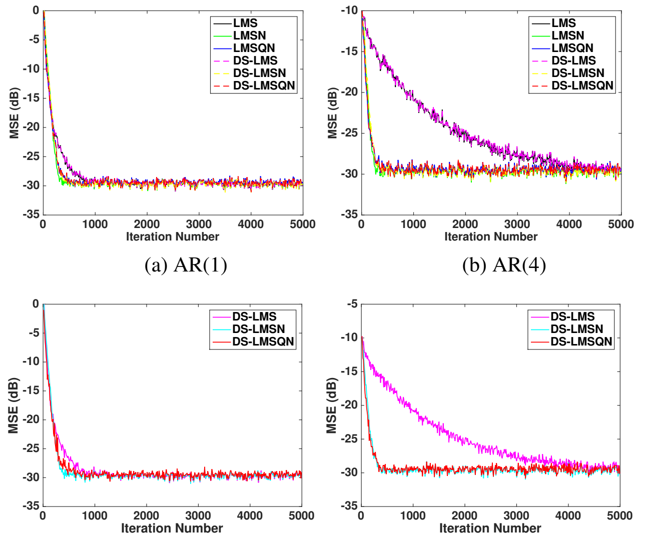
Two cases of input signals are examined, a first-order and a fourth-order AR process, given by

$$\begin{aligned} x(k) =& 0.88x(k-1) + n_1(k), \\ x(k) =& -0.55x(k-1) - 1.221x(k-2) - 0.49955x(k-3) \\ & - 0.4536x(k-1) + n_2(k), \end{aligned}$$

where $n_1(k)$ and $n_2(k)$ are samples from a Gaussian noise uncorrelated with the additional noise whose variance is set such as the input signal is of unit variance. The additional noise is of variance $\sigma_n^2 = 0.001$. The parameters of the LMSN and LMSQN are set to $\mu = 0.1$ and $\alpha = 0.1$ and $\mu = 0.08$ for both the data-selective and the original versions, respectively. In the examined cases, the performance of the LMS and its data-selective counterpart [6] is also presented for comparison purposes. Both the step-size parameter of the data-selective and the original LMS is set to $\mu = 0.08$.

In Fig. 1.a, the MSE learning curves of the data-selective LMSN ("DS-LMSN") and LMSQN ("DS-LMSQN") are compared to the ones of the original non-data-selective approaches and the ones of the data-selective ("DS-LMS") and the original LMS for the first-order AR input signal. The probability of update is set to $P_{up} = 0.4$. As can been seen, the performance of the data-selective approaches is almost identical to the original ones, even if the former are performing approximately 60% fewer updates in the mean than the latter. Furthermore, the LMSN and LMSQN based approaches perform better than the LMS-based one since they converge faster to their steady-state performance. Similar conclusions can be reached by observing the corresponding results when the input signal is a fourth-order AR process in Fig. 1.b. Though, we observe that the gap on the performance between the LMSN and LMSQN-based approaches and the LMS-based one is even bigger. This is due to fact that the former approaches appear to be less dependent on the statistics of the input signal, as discussed on the Introduction.

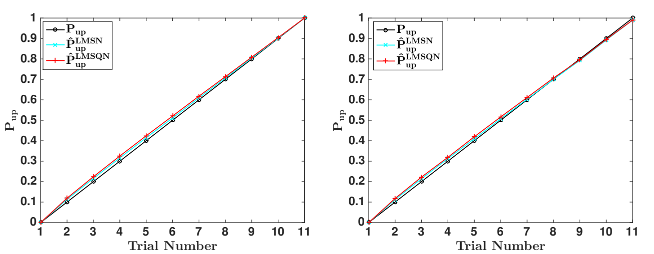
In Figs. 1.c - 1.d, the experimental setup of Figs. 1.a - 1.b is repeated, though now it is assumed that outliers are also present in the output signal $d(k)$. The outliers addition to the 1% of the output signal is done via a Bernouli process multiplying a random binary



(a) AR(1) (b) AR(4)

(c) AR(1) + outliers (d) AR(4) + outliers

Fig. 1. MSE learning curves for the data-selective and the original approaches.



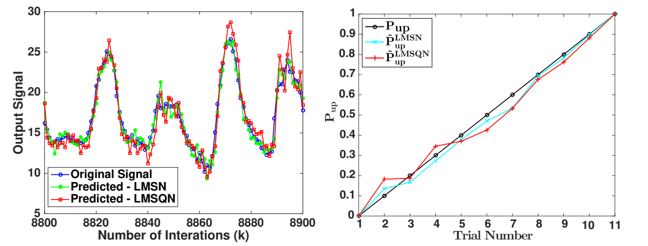
(a) No outliers (b) Outliers

Fig. 2. Comparison between prescribed P_{up} and the achieved $\hat{P}_{up}^{LMSN}$ and $\hat{P}_{up}^{LMSQN}$ by the data-selective LMSN and LMSQN algorithms.

signal with an amplitude equal to ± 50. In Figs. 1.c - 1.d, the MSE learning curves are shown for the data-selective LMSN, LMSQN and LMS algorithms for the first-order and fourth-order AR input signals, respectively. The proposed solutions are performing satisfactorily for the prescribed probability of update while they maintain their gap on the performance compared to the LMS based solution.

In Figs. 2.a - 2.b, the accuracy of the proposed method in prescribing the desired probability of update to the considered algorithms is examined. To that end, we plot the achieved probability of update along with the desired one ("P_{up}") for both the data-selective LMSN ("$\hat{P}_{up}^{LMSN}$") and LMSQN ("$\hat{P}_{up}^{LMSQN}$") algorithms. An AR(1)-modeled input signal is considered for both the cases where the output signal is corrupted by outliers or not. The results show that the achieved probability of update for both the proposed data-selective algorithms follows the desired one closely.

Let us move now to the results based on the real-world data. The data-set used is provided by the University of California at Irvine and is constituted by temperature recordings from a significantly polluted area in Italy [26]. The aim is to employ the adaptive filter in order to predict the temperature values based on past measurements. To that end, the values of the predicting filter must be identified. It is evident that in order to apply a data selection strategy, the value of the prediction error variance σ_e^2 must also be derived for the considered prediction setup. This can be done by following the procedure



(a) (b)

Fig. 3. Simulation results in real world data-set. a) Comparison between the original signal and the predicted one by the data-selective LMSN and LMSQN algorithms. b) Comparison between prescribed P_{up} and the achieved $\hat{P}_{up}^{LMSN}$ and $\hat{P}_{up}^{LMSQN}$ by the data-selective algorithms.

in [22]. By assuming that the past samples $\mathbf{x}(k)$ are used for an one-step-ahead prediction setup, the MSE may be written as,

$$\xi(k) = \mathbb{E}\{(x(k+1) - \mathbf{w}^T(k)\mathbf{x}(k))^2\}. \tag{17}$$

From (17), it can be shown that the minimum MSE is given by

$$\xi_{min}(k) = r(0) - \mathbf{w}_o^T \begin{bmatrix} r(1) \\ r(2) \\ \vdots \\ r(N+1) \end{bmatrix}, \tag{18}$$

where $\mathbf{w}_o$ is the optimal coefficient of the prediction filter and $r(l) = \mathbb{E}\{x(k)x(k-l)\}$ for a stationary process. An estimate of σ_e^2 may be derived at the kth iterate by using the current coefficient estimate $\mathbf{w}(k)$ in place of $\mathbf{w}_o$ in (18). The value of $r(l)$ is estimated via

$$r(l) = \beta r(l-1) + (1-\beta)x(k)x(k-l), \tag{19}$$

where β is a forgetting factor. The value used for the simulations is $\beta = 0.99$. The step size parameter is set to $\mu = 0.1$ for both the algorithms. For the LMSN one, we set $\alpha = 0.05$, as well. The length of the prediction filter is set to $L + 1 = 15$.

In Fig 3.a, the output of the data-selective prediction filter for $P_{up} = 0.4$ is compared to the original signal values. As it is shown, both the data-selective LMSN and LMSQN approaches predict quite accurately the original signal while they perform a reduced number of updates. In Fig. 3.b, the achieved probability of the data-selective LMSN and LMSQN algorithms is examined when the prescribed probability is set to a value within the range [0 1]. The results show that the achieved probability for both the LMSN and LMSQN algorithms is close to the prescribed one. The achieved probability of LMSN is more close to the prescribed one compared to one of the LMSQN algorithm, though that gap is negligible for higher values of the prescribed probability.

6. CONCLUSION

In this work, data-selective versions of the well known LMSN and LMSQN algorithms were developed. The proposed solutions aim at the computational complexity reduction by avoiding update iterations when non-innovative data are acquired. Furthermore, the proposed solutions are also avoiding updates when outliers are detected, thus enhancing the learning performance of the original schemes. Simulation results verify the efficiency of the proposed approaches over the original ones that have not data selection capabilities and over the data-selective LMS algorithm.

7. REFERENCES

[1] S. Werner and P. S. R. Diniz, "Set-membership affine projection algorithm," *IEEE Signal Process. Lett.*, vol. 8, no. 8, pp. 231–235, Aug. 2001.

[2] P. S. R. Diniz and S. Werner, "Set-membership binormalized data-reusing LMS algorithms," *IEEE Trans. Signal Process.*, vol. 51, no. 1, pp. 124–134, Jan. 2003.

[3] J. F. Galdino, J. A. Apolinario, and M. L. R. De Campos, "A set-membership NLMS algorithm with time-varying error bound," in *IEEE Int. Symposium Circuits and Syst. (ISCS*, May 2006, pp. 4 pp.–280.

[4] D. Berberidis, V. Kekatos, and G. B. Giannakis, "Online censoring for large-scale regressions with application to streaming big data," *IEEE Trans. Signal Process.*, vol. 64, no. 15, pp. 3854–3867, Aug. 2016.

[5] Z. Wang, Z. Yu, Q. Ling, D. Berberidis, and G. B. Giannakis, "Distributed recursive least-squares with data-adaptive censoring," in *IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, March 2017, pp. 5860–5864.

[6] P. S. R. Diniz, "On data-selective adaptive filtering," *IEEE Trans. Signal Process.*, vol. 66, no. 16, pp. 4239–4252, Aug. 2018.

[7] S. Gollamudi, S. Nagaraj, S. Kapoor, and Yih-Fang Huang, "Set-membership filtering and a set-membership normalized LMS algorithm with an adaptive step size," *IEEE Signal Process. Lett.*, vol. 5, no. 5, pp. 111–114, May 1998.

[8] N. Takahashi and I. Yamada, "Steady-state mean-square performance analysis of a relaxed set-membership NLMS algorithm by the energy conservation argument," *IEEE Trans. Signal Process.*, vol. 57, no. 9, pp. 3361–3372, Sep. 2009.

[9] P. S. R. Diniz and L. W. P. Biscainho, "Optimal variable step size for the LMS/Newton algorithm with application to subband adaptive filtering," *IEEE Trans. Signal Process.*, vol. 40, no. 11, pp. 2825–2829, Nov. 1992.

[10] D. F. Marshall and W. K. Jenkins, "A fast quasi-Newton adaptive filtering algorithm," *IEEE Trans. Signal Process.*, vol. 40, no. 7, pp. 1652–1662, July 1992.

[11] P. S. R. Diniz, M. L. R. de Campos, and A. Antoniou, "Analysis of LMS-Newton adaptive filtering algorithms with variable convergence factor," *IEEE Trans. Signal Process.*, vol. 43, no. 3, pp. 617–627, March 1995.

[12] G. Glentis, K. Berberidis, and S. Theodoridis, "Efficient least squares adaptive algorithms for FIR transversal filtering," *IEEE Signal Process. Mag.*, vol. 16, no. 4, pp. 13–41, July 1999.

[13] B. Farhang-Boroujeny, "Fast LMS/Newton algorithms based on autoregressive modeling and their application to acoustic echo cancellation," *IEEE Trans. Signal Process.*, vol. 45, no. 8, pp. 1987–2000, Aug. 1997.

[14] M. L. R. De Campos and A. Antoniou, "A new quasi-Newton adaptive filtering algorithm," *IEEE Trans. Circuits Syst. II. Analog Digit. Signal Process.*, vol. 44, no. 11, pp. 924–934, Nov. 1997.

[15] K. Berberidis, S. Rantos, and J. Palicot, "A step-by-step quasi-Newton algorithm in the frequency domain and its application to adaptive channel equalization," *IEEE Trans. Signal Process.*, vol. 52, no. 12, pp. 3335–3344, Dec. 2004.

[16] M. S. Salman, O. Kukrer, and A. Hocanin, "A fast implementation of quasi-Newton LMS algorithm using FFT," in *Int. Conf. Digit. Inf. Commun. Technol. Appl. (DICTAP)*, May 2012, pp. 510–513.

[17] M. Z. A. Bhotto and A. Antoniou, "Robust set-membership affine-projection adaptive-filtering algorithm," *IEEE Trans. Signal Process.*, vol. 60, no. 1, pp. 73–81, Jan 2012.

[18] S. Zhang and J. Zhang, "Set-membership NLMS algorithm with robust error bound," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 61, no. 7, pp. 536–540, July 2014.

[19] W. Mao, "Robust set-membership filtering techniques on GPS sensor jamming mitigation," *IEEE Sensors J.*, vol. 17, no. 6, pp. 1810–1818, March 2017.

[20] M. Rupp, "Pseudo affine projection algorithms revisited: Robustness and stability analysis," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2017–2023, May 2011.

[21] H. Yazdanpanah, M.V.S. Lima, and P.S.R. Diniz, "On the robustness of set-membership adaptive filtering algorithms," *EURASIP J. Advances Signal Process.*, vol. 2017, no. 1, pp. 72, 2017.

[22] P.S.R Diniz, *Adaptive Filtering: Algorithms and Practical Implementation*, New York, NY, USA, Springer, 4th edition, 2013.

[23] Athanasios Papoulis and S Unnikrishna Pillai, *Probability, random variables, and stochastic processes*, Tata McGraw-Hill Education, 2002.

[24] M.S.V. Lima and P.S.R. Diniz, "Steady-state MSE performance of the set-membership affine projection algorithm," *Circuits Sys. Signal Process.*, vol. 32, pp. 1811–1837, Aug. 2013.

[25] M.O.K. Mendonça, J. O. Ferreira, C. G. Tsinos, P.S.R. Diniz, and T. N. Ferreira, "On fast converging data-selective adaptive filtering," *Algorithms*, vol. 12, no. 1, pp. 4, 2019.

[26] UC Irvine, "Air quality data set, machine learning repository, [online]," https://archive.ics.uci.edu/ml/datasets/Air+quality, Accessed: 2017-09.