Estimation of Measurement-Noise Variance for Variable-Step-Size NLMS Filters

Tilo Strutz

Deutsche Telekom, Leipzig University of Telecommunications (HfTL), Institute of Communications Engineering
Gustav-Freytag-Str. 43–45, 04277 Leipzig, Germany

Abstract—**Least-mean-square (LMS) filters are a well-studied processing technique that adapts iteratively to an unknown process. It has been proven that the parameters of the LMS filter converge to the optimum (Wiener) solution. Unfortunately, this is only possible if the adaptation steps are infinitely small. Small steps, however, result in slow convergence. The challenge is to vary the step size such that they are large, when the LMS filter coefficients are far from their optimal values and to lower the step size, when the adaptive system is approaching the optimum. State-of-the-art approaches to variable-step-size determination take estimates of the measurement noise into account for optimal performance.**

This paper proposes a new technique for the estimation of the measurement-noise variance, which also can deal with sudden changes of the unknown system. Based on investigations with a broad range of experimental conditions in terms of test signals and different measurement-noise levels, it is shown that the proposed estimation technique is robust to changes of the unknown system and outperforms other methods.

Index Terms—**least mean squares, NLMS, measurement-noise estimation, adaptive systems, change detection**

I. INTRODUCTION

Adaptive filters based on the least-mean-squares (LMS) principle are a mature and well-studied optimisation technique [1] which allows to iteratively adjust the parameters of finite-impulse-response (FIR) filters based on a stochastic approximation of the gradient-descent method [2]. The general idea can be explained in application to system identification, **Figure 1**. An unknown system having the impulse response[1] $\mathbf{h} = (h_0 \quad h_1 \quad \dots \quad h_{L-1})$ processes an input signal $x[n]$ and outputs a corresponding signal $d'[n]$. Typically this output is not directly accessible, but it is disturbed by noise. This could be additive noise as represented by $v[n]$ in Figure 1 leading to $d[n]$. The latter is the *desired* signal and acts as reference. The impulse response $\mathbf{h}$ has to be estimated using an adaptive system that works in parallel. It is fed with the same input signal and outputs signal $y[n]$. The adaptive system is defined by its own, time-variant impulse response $\mathbf{a}[n] = (a_0 \quad a_1 \quad \dots \quad a_{M-1})$ with a length of M. The output of the adaptive system at time step n

$$y[n] = \mathbf{a}[n] \cdot \mathbf{x}[n] \tag{1}$$

is based on the tap input vector $\mathbf{x}[n] = (x[n] \ x[n{-}1] \ \cdots \ x[n{-}M{+}1])^{\mathrm{T}}$ which requires the choice of a suitable number M of filter taps allowing the emulation of $\mathbf{h}$.

The idea is to use a time-dependent vector $\mathbf{a}[n]$ such that the energy of the difference signal $e[n] = d[n] - y[n]$ approaches

---

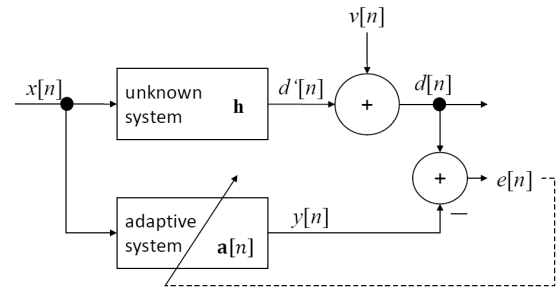[1]For a system with infinite impulse response, L is equal to infinity.

Fig. 1. Principle of estimating the impulse response $\mathbf{h}$ of an unknown system using an adaptive system

its minimum. The required derivations (see [3] or [4]) lead to incremental updates for the LMS filter parameters:

$$\mathbf{a}[n + 1] = \mathbf{a}[n] + \mu' \cdot e[n] \cdot \mathbf{x}[n] \ . \tag{2}$$

The correction term contains three values: the current error $e[n]$, the input signal vector $x[n]$, and a so-called learning rate μ'. This learning rate (or 'step size') influences the speed of adaptation, and also the quality of the adaptation after convergence.

The basic convergence properties of the LMS filter are understood for a long time [5], [6] and have been analysed under specific conditions [7], [8]. Different techniques have been proposed to accelerate the speed of convergence [9], [10]. The iterative estimation process introduces noise into the vector of filter coefficients that is proportional to the speed of adaptation and is also proportional to the number of filter taps. This adaptation noise prevents the LMS filter to reach the optimal (Wiener) solution. With other words: the choice of μ' controls the compromise between fast convergence (and low accuracy of estimation, *steady-state misalignment*) and highly accurate estimation of filter coefficients (and very slow convergence).

One problem with the adaptation rule (2) is that if the energy of $\mathbf{x}[n]$ is large, the correction term becomes large as well and introduces noise into the weights vector. This problem can be alleviated by proper normalisation [11], [12]. In [3], the corresponding relation for normalised LMS (NLMS) filters has been derived as *principle of minimal disturbance*. Due to the constraint that the adaptation step leads to a filter output identical to the reference signal value, the change in weights should be minimal. The derivations lead to a modified version

of (2):

$$\mathbf{a}[n+1] = \mathbf{a}[n] + \mu \cdot e[n] \cdot \frac{\mathbf{x}[n]}{||\mathbf{x}[n]||^2 + \varepsilon} \qquad \varepsilon > 0 . \quad (3)$$

The constant ε prevents possible division by zero. In [6], [13]–[15], the convergence properties of the NLMS filters have been evaluated for zero-mean Gaussian distributed input signals. The proposal of [14] has been analysed more in detail in [16]. Other approaches with variable learning rates have been discussed in [17]–[19], for example, while adapting ε as a regularisation term. Valin has adopted this idea for non-stationary sources in application to double-talk [20]. The earliest attempts to variable step sizes were based on LMS without normalisation [21], [22], which also inspired [23]. All these variable-step-size (VSS) approaches try to use a large adaptation step when the estimation is far away from the true impulse response (in order to achieve fast convergence) and to decrease the steps when the estimated impulse response is getting closer to the desired one. This is typically achieved by tuning one or more parameters. Their optimal values depend on the signal characteristics, so that the methods mostly are not suitable for any application. Benesty *et al* [24] showed a nice derivation towards a non-parametric algorithm, however the result requires the knowledge of the true measurement noise variance σ_v^2. In application to acoustic echo cancellation, they proposed to estimate the noise during silent sections. As in [24], also in [25] the relation between the variance of measurement noise and the error signal is exploited in an effective manner.

More recently, a method has been derived in [26] based on a joint optimisation of μ and ε minimising the system misalignment. The same approach has been presented in [27] in application to acoustic echo cancellation. However, like [24], the algorithm does not include the required estimation of the measurement-noise variance σ_v^2, but refer to [28], which has already addressed the problem of measurement-noise estimation and change detection.

In combination with the methods of [24] and [26], this paper proposes a new approach to robust estimation of measurement noise that can reliably handle sudden changes in the unknown system and is capable of responding accordingly and allowing rapid adaptation convergence even after such changes. The investigations have been performed under various conditions making statements about the general applicability possible and thus serve the comprehensive experimental validation.

The paper is organised as follows. Section II first discusses two state-of-the-art approaches to step-size adaptation that will be used later in the investigations. Section III presents two methods for measurement-noise estimation; the first is state of the art from [28] and the second is a new low-complexity proposal. Section IV describes the investigations and presents results. The paper finishes with a summary in Section V.

II. VARIABLE STEP SIZE METHODS BASED ON ESTIMATED MEASUREMENT NOISE

A. Noise-Error relation

Benesty *et al* [24] developed a method which is based on the knowledge of the true measurement noise power σ_v^2. They have

derived the step size as $\mu = (1 - \sigma_v/\sigma_e)$ changing equation (3) to

$$\mathbf{a}[n+1] = \mathbf{a}[n] + \left(1 - \frac{\sigma_v}{\sigma_e}\right) \cdot e[n] \cdot \frac{\mathbf{x}[n]}{||\mathbf{x}[n]||^2 + \varepsilon} , \quad (4)$$

where σ_e is the standard deviation of the error signal $\{e[n]\}$. While σ_e could be easily estimated from the recorded error signal, the standard deviation σ_v of the measurement noise is not directly accessible.

B. Joint step size and regularisation optimisation

In [26], a method has been proposed aiming at the minimisation of system misalignment. The adaptation process requires two predefined variables: m, which is initialised with a small positive number, and $\hat{\sigma}_w^2[n-1] = 0$, which expresses the variance of the filter updates.

During the adaptation process following is calculated:

$$p = m + M \cdot \sigma_w^2[n-1] \quad (5)$$

$$\hat{\sigma}_x^2 = \frac{1}{M} \cdot ||\mathbf{x}[n]||^2$$

$$\mu' = \frac{p}{(M+2) \cdot p \cdot \hat{\sigma}_x^2 + M \cdot \hat{\sigma}_v^2} , \quad (6)$$

where M is the number of filter taps in $\mathbf{a}$. In comparison to (3), it can be seen that the strength of normalisation is controlled by p, while the regularisation term is defined depending on the estimated measurement-noise variance $\hat{\sigma}_v^2$.

After adjusting the estimated impulse response $\mathbf{a}$ with

$$\mathbf{u}[n] = \mu' \cdot \mathbf{x}[n] \cdot e[n]$$

$$\mathbf{a}[n] = \mathbf{a}[n-1] + \mathbf{u}[n] ,$$

the control parameters m and $\hat{\sigma}_w^2$ must be newly determined:

$$m = (1 - \mu' \cdot \hat{\sigma}_x^2) \cdot p \qquad \text{and}$$

$$\hat{\sigma}_w^2[n] = \frac{1}{M} \cdot \mathbf{u}[n]^T \cdot \mathbf{u}[n] . \quad (7)$$

III. METHODS FOR MEASUREMENT-NOISE ESTIMATION

Unfortunately, the estimation of σ_v^2 has not been specified in [24] and [26]. The authors of the latter paper referred to different estimation methods in the literature and did not analyse its influence on the algorithms performance. We found that the measurement noise must in fact be carefully estimated because reliable estimates distinctly benefit the steady-state performance.

This section first repeats a state-of-the-art method for measurement-noise estimation. Afterwards a new robust estimation technique is introduced.

A. Estimation based on crosscorrelation of input and error signal

In application to echo cancellation, Iqbal and Grant estimated the measurement noise in [28] with

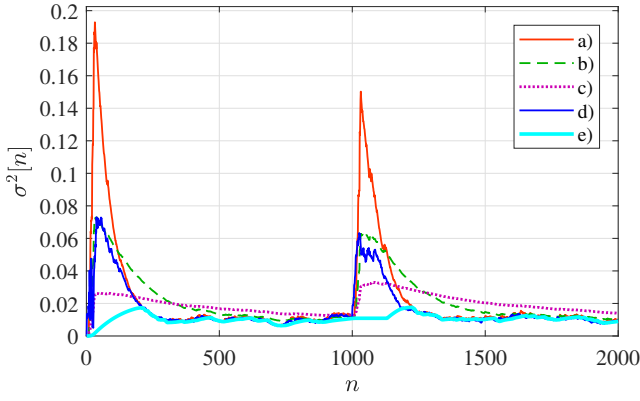$$\hat{\sigma}_v^2 = \hat{\sigma}_e^2 - \frac{\mathbf{r}_{ex} \cdot \mathbf{r}_{ex}^T}{\hat{\sigma}_x^2} . \quad (8)$$

Fig. 2. Curse of variance estimates: a) s_f^2, b) s_m^2, c) s_s^2, d) $\hat{\sigma}_v^2$ based on (8)-(10), e) $\hat{\sigma}_v^2$ based on (11)-(15). The true variance is $\sigma_v^2 = 0.01$.

Under the premise that the input signal $\{x[n]\}$ and the error signal $\{e[n]\}$ have a mean equal to zero, the variances on the right side of (8) can recursively be estimated by

$$\hat{\sigma}_e^2 \leftarrow w_i \cdot \hat{\sigma}_e^2 + (1 - w_i) \cdot (e[n])^2$$
$$\hat{\sigma}_x^2 \leftarrow w_i \cdot \hat{\sigma}_x^2 + (1 - w_i) \cdot (x[n])^2 , \quad (9)$$

with $w_i = 1 - 1/(6M)$ as smoothing factor as suggested in [29]. The crosscorrelation vector is calculated in a similar manner

$$\mathbf{r}_{ex} \leftarrow w_i \cdot \mathbf{r}_{ex}(n) + (1 - w_i) \cdot \mathbf{x}[n] \cdot e[n] . \quad (10)$$

B. New measurement-noise estimation

With regard to equation (4), the basic idea of the proposal is not to focus on the most accurate estimates σ_e and σ_v, but rather to address the relationship between them. Under stationary conditions, σ_v is constant and the value of σ_e converges to σ_v with the progress of adaptation. Hence, if σ_e is estimated based on recent samples, it is closer to σ_v compared to the case where it is estimated including earlier samples. That is why we apply a fast (f), a medium (m) and a slow (s) estimator working in parallel:

$$s_f^2 \leftarrow w_f \cdot s_f^2 + (1 - w_f) \cdot (e[n])^2$$
$$s_m^2 \leftarrow w_m \cdot s_m^2 + (1 - w_m) \cdot (e[n])^2$$
$$s_s^2 \leftarrow w_s \cdot s_s^2 + (1 - w_s) \cdot (e[n])^2 , \quad (11)$$

with weights of $w_s = 1 - 1/(45M)$, $w_m = 1 - 1/(15M)$, and $w_f = 1 - 1/(5M)$. In contrast to the statement in [29] we need $w_f < w_s$ in order to make s_f^2 the better estimate of σ_v^2. The initial values are set to $s_f^2 = s_m^2 = s_s^2 = 0$. The values of the weights have empirically been determined leading to good performance under all tested conditions. **Figure 2** shows a typical course of these estimates (a, b, c) if the unknown system has been changed at $n = 1000$.

It is assumed that the minimum estimate is the most suitable one

$$s_{min}^2 = \min(s_f^2, s_m^2, s_s^2) \quad (12)$$

and the measurement noise is rapidly adapted to this value

$$\hat{\sigma}_v^2 \leftarrow w_f \cdot \hat{\sigma}_v^2 + (1 - w_f) \cdot s_{min}^2 . \quad (13)$$

This procedure describes the normal mode and has to be supplemented with a change-detection mechanism. As soon as $s_f^2 > s_s^2$ is observed, the estimator activates the change mode, which is entered at next time step. Arrived in change mode, $\hat{\sigma}_v^2$ is modified only if this mode is entered for the first time[2] or s_f^2 has dropped below s_m^2:

$$\hat{\sigma}_v^2 \leftarrow w_m \cdot \hat{\sigma}_v^2 + (1 - w_m) \cdot s_{min}^2 . \quad (14)$$

Otherwise the estimate of the measurement noise is kept fix as can be seen in Figure 2, curve e) at $1000 < n < 1100$. The procedure returns to the normal state as soon $s_m^2 < s_s^2$ holds. Independent on the current mode, the final check is

$$\hat{\sigma}_v^2 \leftarrow \min(\hat{\sigma}_v^2, s_f^2) . \quad (15)$$

Curve e) of Figure 2 shows the resulting estimates of $\hat{\sigma}_v^2$. It can be clearly seen that these estimate are more accurate than the estimates based on (8), curve d), after starting the adaptation process and, even more important, after changes of the unknown system.

While the estimation according to Subsection III-A requires $5 + 2 \cdot M$ multiplications, the complexity of the proposal is reduced to eight multiplications in maximum as the computation of the crosscorrelation vector is not required.

All thresholds and weights used in this approach have been determined empirically and work well for the broad range of different scenarios. In real application with some pre-knowledge about the characteristics of the unknown system or the input signals, these parameters can probably be optimised.

IV. INVESTIGATIONS

A. Experimental set-up

The two VSS techniques described in Section II ('Bene', [24] and 'Cioc', [26]) have been combined with either the estimation method of [28] ('_Iqba', see Subsection III-A) or with the proposed approach as described in Subsection III-B ('_new'). These four combinations have been compared with each other and with the standard NLMS method (equation (3), $\mu = 1$) using different test signals $x[n]$ of length $N = 2000$ each. According to the block diagram in Figure 1, the signals pass a simulated unknown system with an impulse response **h**. White Gaussian measurement noise ($\sigma_v^2 = 0.01$) is added afterwards:

$$d[n] = v[n] + \sum_{j=0}^{L-1} h_j \cdot x[n - j] . \quad (16)$$

The investigations aim at general statements about the performance of the different methods independent on any previous knowledge about the unknown system or the properties of the input signals. So, the results shall be valid in general case. Therefore, a broad range of possible combinations have been simulated. The intensive performance study serves as indispensable validation of the mathematically derived approaches. Intermediate filter taps h_j', $j = 0, 1, \ldots, L - 1$, are randomly

---

[2]This condition is needed after initialisation of the adaptation process.

TABLE I
MISALIGNMENT (MSE IN dB) AFTER 2000 ADAPTATION STEPS OF
DIFFERENT APPROACHES DEPENDENT ON USED TEST SIGNAL (x_1 .. WHITE
NOISE; x_2 AR(3); x_3 .. NON-LINEAR; x_4 .. HARMONIC COMPONENT; x_5 ..
SPEECH SIGNAL). BEST RESULTS ARE SHOWN BOLD FACE; SECOND BEST
ARE PRINTED IN ITALICS. WORSE RESULTS THAN FOR NLMS ARE
UNDERLINED.

method	$x_1[n]$	$x_2[n]$	$x_3[n]$	$x_4[n]$	$x_5[n]$
NLMS	-16.80	-17.04	-16.49	*-17.12*	-15.32
Bene_Iqba	**-19.84**	-14.39	-19.28	-7.12	-13.92
Bene_new	-19.76	**-17.55**	-19.29	-16.61	*-16.25*
Cioc_Iqba	*-19.79*	-16.44	-19.05	-12.79	-14.63
Cioc_new	-19.76	*-17.51*	**-19.76**	**-17.26**	**-16.52**

drawn from a uniform distribution with $[-0.5\ldots 0.5]$. The final impulse response is achieved by normalisation:

$$\mathbf{h} = \mathbf{h}'/\sqrt{\mathbf{h}' \cdot \mathbf{h}'^{\mathrm{T}}} . \tag{17}$$

This vector is perturbed after $N/2$ steps. For evaluation purposes, the adaptive system has the same model as the unknown system and $M = L = 10$ is used. The different approaches have been applied to following input signals:

- $x_1[n]$: random sequence, zero-mean, normal distribution $\sigma = 1$,
- correlated input, autoregressive process of order 3, [19]:

$$x_2[n] = \frac{3}{2} \cdot x_2[n-1] - x_2[n-2] + \frac{1}{4} \cdot x_2[n-3] + x_1[n]$$

- $x_3[n] = \frac{x_3[n-1]}{1+(x_3[n-1])^2} + (x_1[n])^3$ (nonlinear, [17]),
- $x_4[n] = \sin\left(80\pi \cdot \frac{n}{N} + \phi\right) + x_1[n]$ (harmonic, random ϕ),
- $x_5[n]$: snippet from a speech signal,

The start positions of $x_5[n]$ randomly vary within a range of 400 samples. During the processing of the entire signal $x_i[n]$, the squared error $e_i^2[n], \forall n = 1, 2, \ldots, N$ is recorded.

For each test signal, the procedure is repeated $T = 400$ times in order to alleviate the influence of different impulse responses and different signal realisation caused by the random elements. The squared error signals are averaged over all these T trials and the mean squared error is computed for each position n in decibel:

$$MSE[n] = 10 \cdot \log_{10}\left(\frac{1}{T} \cdot \sum_{i=1}^{T} e_i^2[n]\right) . \tag{18}$$

The final performance measure for steady-state misalignment is based on the last two-hundred MSE values

$$MSE = \frac{1}{200} \cdot \sum_{n=N-199}^{N} MSE[n] . \tag{19}$$

In correspondence to the settings explained above, this implies that the convergence should have finished within 800 steps after the last perturbation of the unknown system. As this implication is not always given, the MSE values indirectly also express the convergence rate of the approaches.

B. Results

The performances in terms of final MSE are listed in **Table I**. While for the white-noise signal $x_1[n]$ the performance is almost the same for all investigated VSS methods, the
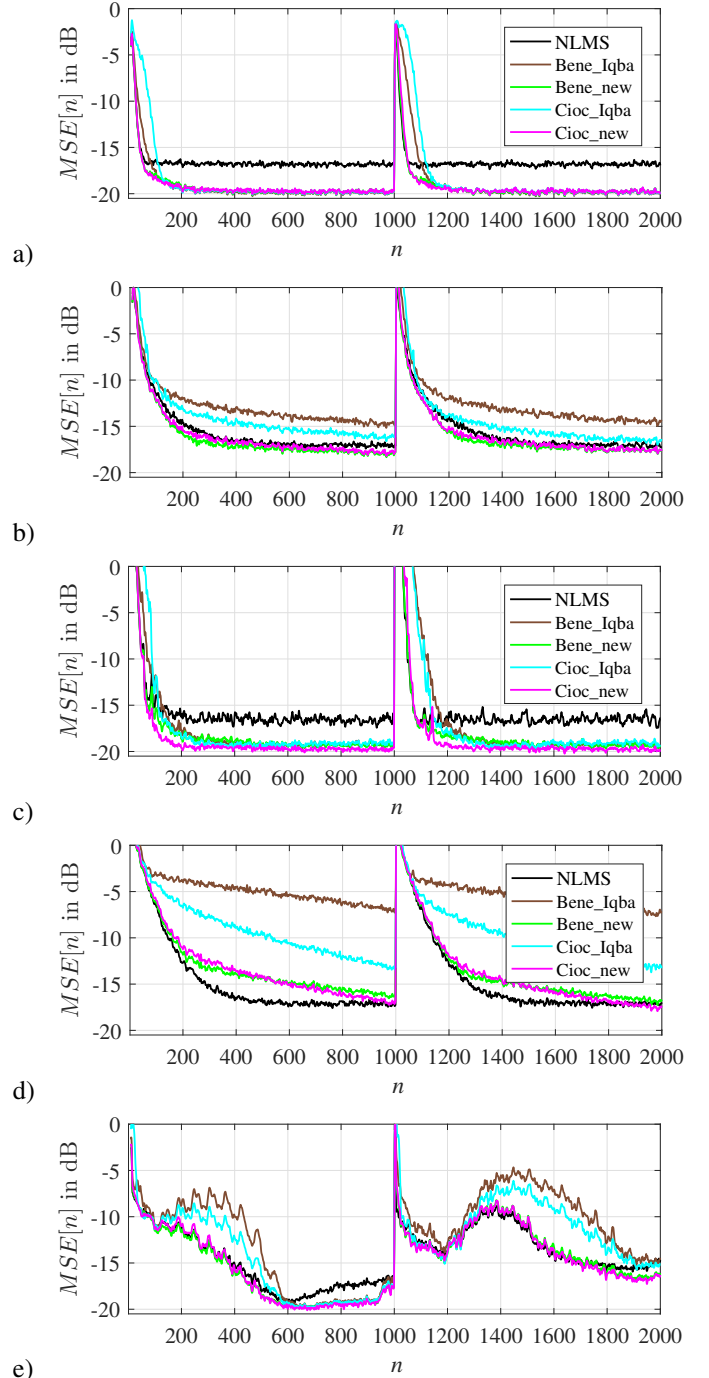

a)

b)

c)

d)

e)

Fig. 3. Course of mean squared error applied to: a) $x_1[n]$; b) $x_2[n]$; c) $x_3[n]$; d) $x_4[n]$; e) $x_5[n]$

proposed estimation of $\hat{\sigma}_{\mathrm{v}}^2$ shows distinct advantages for all other test signals. Inspecting the MSE graphs in **Figure 3**, it is obvious that the proposed method for measurement-noise estimation benefits also the speed of convergence.

When reducing the simulated measurement noise to $\sigma_{\mathrm{v}}^2 = 0.001$, none of the VSS approaches can beat NLMS in application to $x_2[n]$ and $x_4[n]$, **Table II**. Nevertheless, the proposed method still improves the adaptation compared to the estimation according to [28]. **Table III** lists the results for $\sigma_{\mathrm{v}}^2 = 0.1$. As expected, the MSE values are much higher.

TABLE II
MISALIGNMENT (MSE IN DB) WHEN USING $\sigma_{\mathrm{v}}^2 = 0.001$

method	$x_1[n]$	$x_2[n]$	$x_3[n]$	$x_4[n]$	$x_5[n]$
NLMS	-26.80	-26.86	-26.49	-27.09	*-19.12*
Bene_Iqba	**-29.84**	*-15.84*	-29.27	-7.39	-15.44
Bene_new	*-29.78*	*-21.79*	-29.28	-19.73	-18.92
Cioc_Iqba	-29.77	*-18.95*	-29.10	-13.78	-16.36
Cioc_new	-29.76	-21.33	**-29.73**	-20.72	**-19.24**

TABLE III
MISALIGNMENT (MSE IN DB) WHEN USING $\sigma_{\mathrm{v}}^2 = 0.1$

method	$x_1[n]$	$x_2[n]$	$x_3[n]$	$x_4[n]$	$x_5[n]$
NLMS	-6.80	-7.06	-6.49	-7.12	-6.99
Bene_Iqba	**-9.84**	-8.63	*-9.28*	-5.16	-8.04
Bene_new	-9.73	*-9.27*	*-9.27*	*-9.05*	*-8.97*
Cioc_Iqba	*-9.80*	-9.26	-9.05	-8.10	-8.49
Cioc_new	-9.74	**-9.33**	**-9.77**	**-9.28**	**-9.14**

However, the relation between the values is pretty much the same as in Table I.

V. SUMMARY

The paper has presented a new technique estimating the measurement-noise variance more accurately and with lower complexity than a state-of-the-art method. The proposed approach has been combined with two variable-step-size NLMS-filter methods leading to superior performance in terms of convergence speed and steady-state misalignment, especially in changing environments, regardless of the chosen simulated measurement noise.

The investigations have comprised many different conditions in terms of impulse responses for the unknown system, input signals, and measurement noise. This experimental validation allows general statements about the performance of the tested approaches without restriction to a particular application.

The proposed method relies on some parameters that have been set empirically. The recursive estimation of different variances, for example, is based on weighing factors. Here, the relation of these factors is of higher importance than the single values themselves.

In this work, the proposed method has only been investigated in application to system identification. Its applicability to other tasks has to be tested in future research. Supporting reproducible research, the software used for the investigations can be downloaded from [30].

VI. ACKNOWLEDGEMENTS

REFERENCES

[1] J. Zeidler, "Performance analysis of the lms adaptive prediction filters," *Proc. of the IEEE*, vol. 78, no. 12, pp. 1781 – 1806, December 1990.

[2] B. Widrow, "Thinking about thinking: The discovery of the lms algorithm," *IEEE Signal Processing Magazine*, vol. 22, no. 1, pp. 100–106, Jan. 2005.

[3] S. Haykin, *Adaptive Filter Theory*, 4th ed. Prentice Hall, 2002.

[4] T. Strutz, *Data Fitting and Uncertainty*, 2nd ed. SpringerVieweg, 2016.

[5] B. Widrow, J. McCool, M. Larimore, and C. Johnson, "Stationary and nonstationary learning characteristics of the lms adaptive filter," *Proc. of the IEEE*, vol. 64, no. 8, pp. 1151–1162, August 1976.

[6] A. Feuer and E. Weinstein, "Convergence analysis of lms filters with uncorrelated gaussian data," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. 33, no. 1, pp. 222–230, January 1985.

[7] V. Nascimento and A. Sayed, "On the learning mechanism of adaptive filters," *IEEE Transactions on Signal Processing*, vol. 48, no. 6, pp. 1609–1625, 2000.

[8] 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.

[9] S. Reddi, "A time-domain adaptive algorithm for rapid convergence," *Proc. of the IEEE*, vol. 72, no. 4, pp. 533–535, 1984.

[10] J. Benesty and P. Duhamel, "A fast exact least mean square adaptive algorithm," *IEEE Transactions on Signal Processing*, vol. 40, no. 12, pp. 2904–2920, 1992.

[11] J. Nagumo and A. Noda, "A learning method for system identification," *IEEE Transactions on Automatic Control*, vol. 12, no. 3, pp. 282–287, June 1967.

[12] A. Albert and L. Gardner, *Stochastic Approximation and Nonlinear Regression*, 1st ed. MIT Press, Cambridge, 1967.

[13] M. Tarrab and A. Feuer, "Convergence and performance analysis of the normalized lms algorithm with uncorrelated gaussian data," *IEEE Trans. on Information Theory*, vol. 34, no. 4, pp. 680–691, 1988.

[14] V. Mathews and Z. Xie, "A stochastic gradient adaptive filter with gradient adaptive step size," *IEEE Transactions on Signal Processing*, vol. 42, no. 6, pp. 2075–2087, 1993.

[15] P. An, B. Brown, and C. Harris, "On the convergence rate performance of the normalized least-mean-square adaptation," *IEEE Transactions on Neural Networks*, vol. 8, no. 5, pp. 1211–1214, 1997.

[16] A. Sulyman and A. Zerguine, "Convergence and steady-state analysis of a variable step-size nlms algorithm," *Signal Processing*, vol. 83, no. 6, pp. 1255–1273, 2003.

[17] D. Mandic, "A generalized normalized gradient descent algorithm," *IEEE Signal Processing Letters*, vol. 11, no. 2, pp. 115–118, 2004.

[18] Y.-S. Choi, H.-C. Shin, and W.-J. Song, "Robust regularization for normalized lms algorithms," *IEEE Trans. on Circuits and Systems II: Express Briefs*, vol. 53, no. 8, pp. 627–631, 2006.

[19] J.-K. Hwang and Y.-P. Li, "Variable step-size lms algorithm with a gradient-based weighted average," *IEEE Signal Processing Letters*, vol. 16, no. 12, pp. 1043–1046, 2009.

[20] J.-M. Valin and I. Collings, "Interference-normalized least mean square algorithm," *IEEE Signal Processing Letters*, vol. 14, no. 12, pp. 988–991, 2007.

[21] R. Harris, D. Chabries, and F. Bishop, "A variable step (vs) adaptive filter algorithm," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. 34, no. 4, pp. 309–316, April 1986.

[22] R. Kwong and E. Johnston, "A variable step size lms algorithm," *IEEE Transactions on Signal Processing*, vol. 40, no. 7, pp. 1633–1642, 1992.

[23] T. Aboulnasr and K. Mayyas, "A robust variable step-size lms-type algorithm: Analysis and simulations," *IEEE Transactions on Signal Processing*, vol. 45, no. 3, pp. 631–639, 1997.

[24] J. Benesty, H. Rey, L. Vega, and S. Tressens, "A nonparametric vss nlms algorithm," *IEEE Signal Processing Letters*, vol. 13, no. 10, pp. 581–584, 2006.

[25] H.-C. Huang and J. Lee, "A new variable step-size nlms algorithm and its performance analysis," *IEEE Transactions on Signal Processing*, vol. 60, no. 4, pp. 2055–2060, 2012.

[26] S. Ciochină, C. Paleologu, and J. Benesty, "An optimized nlms algorithm for system identification," *Signal Processing*, vol. 118, no. 1, pp. 115–121, January 2016.

[27] C. Paleologu, S. Ciochină, J. Benesty, and S. Grant, "An overview on optimized nlms algorithms for acoustic echo cancellation," *EURASIP Journal on Advances in Signal Proc.*, no. 97, pp. 1–19, January 2016.

[28] M. Iqbal and S. Grant, "Novel variable step size nlms algorithm for echo cancellation," in *Proc. of Int. Conf. on Acoustics, Speech and Signal Processing*, 31 March-4 April 2008, pp. 241–244.

[29] C. Paleologu, J. Benesty, S. Grant, and C. Osterwise, "Variable step-size nlms algorithms designed for echo cancellation," in *Proceedings of of 43rd Asilomar Conf. on Signals, Systems and Computers*, November 2009, pp. 633–637.

[30] T. Strutz, "Noise estimation, source code," http://www1.hft-leipzig.de/strutz/Papers/NoiseEstimation-resources/Strutz-NoiseEst.zip last visited: 04.02.2019.