The NLMS Algorithm with Time-Variant Optimum Stepsize Derived from a Bayesian Network Perspective

Christian Huemmer, Roland Maas, and Walter Kellermann, Fellow, IEEE

Abstract-In this letter, we derive a new stepsize adaptation for the normalized least mean square algorithm (NLMS) by describing the task of linear acoustic echo cancellation from a Bayesian network perspective. Similar to the well-known Kalman filter equations, we model the acoustic wave propagation from the loudspeaker to the microphone by a latent state vector and define a linear observation equation (to model the relation between the state vector and the observation) as well as a linear process equation (to model the temporal progress of the state vector). Based on additional assumptions on the statistics of the random variables in observation and process equation, we apply the expectation-maximization (EM) algorithm to derive an NLMS-like filter adaptation. By exploiting the conditional independence rules for Bayesian networks, we reveal that the resulting EM-NLMS algorithm has a stepsize update equivalent to the optimal-stepsize calculation proposed by Yamamoto and Kitayama in 1982, which has been adopted in many textbooks. As main difference, the instantaneous stepsize value is estimated in the M step of the EM algorithm (instead of being approximated by artificially extending the acoustic echo path). The EM-NLMS algorithm is experimentally verified for synthesized scenarios with both, white noise and male speech as input signal.

Index Terms-Adaptive stepsize, EM algorithm, NLMS.

I. INTRODUCTION

ACHINE learning techniques have been widely applied to signal processing tasks since decades [1], [2]. For example, directed graphical models, termed Bayesian networks, have shown to provide a powerful framework for modeling causal probabilistic relationships between random variables [3]–[7]. In previous work, the update equations of the Kalman filter and the normalized least mean square (NLMS) algorithm have already been derived from a Bayesian network perspective based on a linear relation between the latent room impulse response (RIR) vector and the observation [8], [9].

The NLMS algorithm is one of the most-widely used adaptive algorithms in speech signal processing and a variety of stepsize adaptation schemes has been proposed to improve its

The authors are with the Chair of Multimedia Communications and Signal Processing (LMS), University of Erlangen-Nuremberg, Erlangen, Germany (e-mail:huemmer@lnt.de; maas@lnt.de; wk@lnt.de).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2015.2439392



Fig. 1. System model for linear AEC with RIR vector \mathbf{h}_n .

system identification performance [10]–[21]. In this article, we derive a novel NLMS-like filter adaptation (termed EM-NLMS algorithm) by applying the expectation maximization (EM) algorithm to a probabilistic model for linear system identification. Based on the conditional independence rules for Bayesian networks, it is shown that the normalized stepsize of the EM-NLMS algorithm is equivalent to the one proposed in [10], which is now commonly accepted as optimum NLMS stepsize rule, see e.g. [22]. As the main difference relative to [10], the normalized stepsize is here estimated as part of the EM algorithm instead of being approximated by artificially extending the acoustic echo path. For a valid comparison, we review the algorithm of [10] for the linear acoustic echo cancellation (AEC) scenario shown in Fig. 1. The acoustic path between loudspeaker and microphone at time n is modeled by the linear finite impulse response (FIR) filter

$$\mathbf{h}_{n} = [h_{0,n}, h_{1,n}, \dots, h_{M-1,n}]^{T}$$
(1)

with time-variant coefficients $h_{\kappa,n}$, where $\kappa = 0, \ldots, M - 1$. The observation equation models the microphone sample d_n :

$$d_n = \mathbf{x}_n^T \mathbf{h}_n + v_n, \tag{2}$$

with the additive variable v_n modeling near-end interferences and the observed input signal vector $\mathbf{x}_n = [x_n, x_{n-1}, \dots, x_{n-M+1}]^T$ capturing the time-domain samples x_n . The iterative estimation of the RIR vector by the adaptive FIR filter $\hat{\mathbf{h}}_n$ is realized by the update rule

$$\hat{\mathbf{h}}_n = \hat{\mathbf{h}}_{n-1} + \lambda_n \mathbf{x}_n e_n, \tag{3}$$

with the stepsize λ_n and the error signal

$$\mathbf{e}_n = d_n - \mathbf{x}_n^T \hat{\mathbf{h}}_{n-1} \tag{4}$$

relating the observation d_n and its estimate $\hat{d}_n = \mathbf{x}_n^T \hat{\mathbf{h}}_{n-1}$. In [10], the optimal choice of λ_n has been approximated as:

$$\lambda_n \approx \frac{1}{M} \frac{\mathcal{E}\{\|\mathbf{h}_n - \mathbf{h}_{n-1}\|_2^2\}}{\mathcal{E}\{e_n^2\}},\tag{5}$$

where $\|\cdot\|_2$ denotes the Euclidean norm and $\mathcal{E}\{\cdot\}$ the expectation operator. As the true echo path \mathbf{h}_n is unobservable, so that

1070-9908 © 2015 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.

Manuscript received April 27, 2015; accepted May 26, 2015. Date of publication June 01, 2015; date of current version June 11, 2015. The work was supported by the Deutsche Forschungsgemeinschaft (DFG) under Contract KE 890/4-2. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Michael Rabbat.

See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

 TABLE I

 Relation Between the NLMS Algorithm Following [10]

 AND THE PROPOSED EM-NLMS Algorithm

	NLMS algorith	um [10] I	EM-NLMS algorithm
Norm. stepsize λ_n	(5)	equivalent (Section II	to E step
Estimation of λ_n	(6)	replaced b (Subsec. II-	Py M step

the numerator in (5) cannot be computed, λ_n is further approximated by introducing a delay of N_T coefficients to the echo path \mathbf{h}_n . Moreover, a recursive approximation of the denominator in (5) is applied using the smoothing factor η [22], [23]. The resulting stepsize approximation

$$\lambda_n \approx \frac{1}{N_T} \frac{\sum_{\kappa=0}^{N_T - 1} \hat{h}_{k,n-1}^2}{(1 - \eta)e_n^2 + \eta \mathcal{E}\{e_{n-1}^2\}}$$
(6)

leads to oscillations which have to be addressed by limiting the absolute value of λ_n [24]. In this article, we derive the EM-NLMS algorithm which applies the filter update of (3) using the stepsize in (5), where λ_n is estimated in the M Step of the EM algorithm instead of being approximated by using (6).

This article is structured as follows: In Section II, we propose a probabilistic model for the linear AEC scenario of Fig. 1 and derive the EM-NLMS algorithm, which is revealed in Section III to be similar to the NLMS algorithm proposed in [10]. As main difference (cf. Table I), the stepsize is estimated in the M Step of the EM algorithm instead of being approximated by artificially extending the acoustic echo path. In Section IV, the EM-NLMS algorithm is experimentally verified for synthesized scenarios with both, white noise and male speech as input signal. Finally, conclusions are drawn in Section V.

II. THE EM-NLMS ALGORITHM FOR LINEAR AEC

Throughout this article, the Gaussian probability density function (PDF) of a real-valued length-M vector \mathbf{z}_n with mean vector $\mu_{\mathbf{z},n}$ and covariance matrix $\mathbf{C}_{\mathbf{z},n}$ is denoted as

$$p(\mathbf{z}_n) = \mathcal{N}\{\mathbf{z}_n | \mu_{\mathbf{z},n}, \mathbf{C}_{\mathbf{z},n}\}, \qquad \mathbf{z}_n \sim \mathcal{N}\{\mathbf{z}_n | \mu_{\mathbf{z},n}, \mathbf{C}_{\mathbf{z},n}\}.$$
(7)

Furthermore, $\mathbf{C}_{\mathbf{z},n} = C_{\mathbf{z},n}\mathbf{I}$ (with identity matrix **I**) implies the elements of \mathbf{z}_n to be mutually statistically independent and of equal variance $C_{\mathbf{z},n}$.

A. Probabilistic AEC Model

To describe the linear AEC scenario of Fig. 1 from a Bayesian network perspective, we model the acoustic echo path as a latent state vector \mathbf{h}_n identically defined as in (1) and capture uncertainties (e.g. due to the limitation to a linear system with a finite set of coefficients) by the additive uncertainty \mathbf{w}_n . Consequently, the linear process equation and the linear observation equation,

$$\mathbf{h}_n = \mathbf{h}_{n-1} + \mathbf{w}_n \quad \text{and} \quad d_n = \mathbf{x}_n^T \mathbf{h}_n + v_n,$$
(8)

can be jointly represented by the graphical model shown in Fig. 2. The directed links express statistical dependencies between the nodes and random variables are marked as circles.



Fig. 2. Bayesian network for linear AEC with latent state vector \mathbf{h}_n .

We make the following assumptions on the PDFs of the random variables in Fig. 2:

• The uncertainty \mathbf{w}_n is normally distributed with mean vector **0** (of zero-valued entries) and variance $C_{\mathbf{w},n}$:

$$\mathbf{w}_n \sim \mathcal{N}\{\mathbf{w}_n | \mathbf{0}, \mathbf{C}_{\mathbf{w},n}\}, \quad \mathbf{C}_{\mathbf{w},n} = C_{\mathbf{w},n}\mathbf{I}.$$
 (9)

• The microphone signal uncertainty v_n is assumed to be normally distributed with variance $C_{v,n}$ and zero mean:

$$v_n \sim \mathcal{N}\{v_n | 0, C_{v,n}\}.$$
(10)

• The initial state vector \mathbf{h}_1 is normally distributed. Due to the linear relations of the random variables in (8), also $p(\mathbf{h}_n|\mathbf{h}_{n-1}), p(d_n|\mathbf{h}_n)$ and $p(\mathbf{h}_n|d_{1:n})$ are Gaussian probability distributions, where the latter is denoted as:

$$p(\mathbf{h}_n|d_{1:n}) = \mathcal{N}\{\mathbf{h}_n|\mu_{\mathbf{h},n}, \mathbf{C}_{\mathbf{h},n}\},\tag{11}$$

with $d_{1:n} = d_1, \ldots, d_n$. Finally, we restrict the covariance matrix of the posterior distribution to be [16]

$$\mathbf{C}_{\mathbf{h},n} = C_{\mathbf{h},n}\mathbf{I}$$
 with $C_{\mathbf{h},n} = \operatorname{tr}\{\mathbf{C}_{\mathbf{h},n}\}/M,$ (12)

where $tr\{\cdot\}$ represents the trace of a matrix.

Based on this probabilistic AEC model, we apply the EM algorithm where the filter update of (3) is derived in the E Step (Section II-B). As main innovation, the optimal stepsize λ_{n+1} in (5) is estimated using the model parameters $C_{v,n+1}$ and $C_{w,n+1}$ predicted in the M step (Section II-C).

B. E step: Inference of the State Vector

The minimum mean square error (MMSE) estimation of the state vector identifies the mean vector of the posterior distribution as estimate $\hat{\mathbf{h}}_n$:

$$\hat{\mathbf{h}}_n = \operatorname*{arg\,min}_{\tilde{\mathbf{h}}_n} \, \mathcal{E}\{\|\tilde{\mathbf{h}}_n - \mathbf{h}_n\|_2^2\} = \mathcal{E}\{\mathbf{h}_n | d_{1:n}\} = \mu_{\mathbf{h},n}. \tag{13}$$

Due to the linear relations between the variables in (2) and (8), and under the restrictions to a linear estimator of $\hat{\mathbf{h}}_n$ and normally distributed random variables, the MMSE estimation is analytically tractable [9]. Exploiting the product rules for linear Gaussian models and conditional independence of the Bayesian network in Fig. 2, the filter update can be derived as a special case of the Kalman filter equations [9, p. 639]:

$$\ddot{\mathbf{h}}_n = \ddot{\mathbf{h}}_{n-1} + \Lambda_n \mathbf{x}_n e_n, \tag{14}$$

where the stepsize matrix Λ_n is calculated as follows:

$$\Lambda_n = \frac{\mathbf{C}_{\mathbf{h},n-1} + \mathbf{C}_{\mathbf{w},n}}{\mathbf{x}_n^T (\mathbf{C}_{\mathbf{h},n-1} + \mathbf{C}_{\mathbf{w},n}) \mathbf{x}_n + C_{v,n}},$$
(15)

$$\mathbf{C}_{\mathbf{h},n} = \left(\mathbf{I} - \Lambda_n \mathbf{x}_n \mathbf{x}_n^T\right) \left(\mathbf{C}_{\mathbf{h},n-1} + \mathbf{C}_{\mathbf{w},n}\right).$$
(16)

By inserting (9) and (12), we can rewrite the filter update of (14) to the filter update defined in (3) with scalar stepsize λ_n :

$$\lambda_n \stackrel{\text{from}}{=} \frac{(15)}{\mathbf{x}_n^T \mathbf{x}_n (C_{\mathbf{h},n-1} + C_{\mathbf{w},n}) + C_{v,n}}, \qquad (17)$$

$$C_{\mathbf{h},n} \stackrel{\text{from (16)}}{=} \left(1 - \lambda_n \frac{\mathbf{x}_n^T \mathbf{x}_n}{M}\right) (C_{\mathbf{h},n-1} + C_{\mathbf{w},n}). \quad (18)$$

Before showing the equality of the stepsize updates in (17) and (5) in Section III, we propose a new alternative to estimate λ_n in (17) by deriving the updates of the model parameters $C_{\mathbf{w},n}$ and $C_{v,n}$ in the following section.

C. M step: Online Learning of the Model Parameters

In the M step, we predict the model parameters for the following time instant. Although the maximum likelihood estimation is analytically tractable, we apply the EM algorithm to derive an online estimator: In order to update $\theta_n = \{C_{v,n}, C_{\mathbf{w},n}\}$ to the new parameters $\{C_{v,n}^{\text{new}}, C_{\mathbf{w},n}^{\text{new}}\}$, the lower bound [9]

$$\mathcal{E}_{\mathbf{h}_{1:n}|\theta_{1:n}}\{\ln\left(p(d_{1:n},\mathbf{h}_{1:n}|\theta_{1:n})\right)\} \le \ln p(d_{1:n}|\theta_{1:n}), \quad (19)$$

is maximized, where $\theta_{1:n} = \{C_{v,1:n}, C_{w,1:n}\}$. For this, the PDF $p(d_{1:n}, \mathbf{h}_{1:n} | \theta_{1:n})$ is determined by applying the decomposition rules for Bayesian networks [9]:

$$p(d_{1:n}, \mathbf{h}_{1:n} | \boldsymbol{\theta}_{1:n}) = p(\mathbf{h}_n | \mathbf{h}_{n-1}, C_{\mathbf{w},n} \mathbf{I}) p(d_n | \mathbf{h}_n, C_{v,n})$$

$$\cdot \prod_{m=1}^{n-1} p(\mathbf{h}_m | \mathbf{h}_{m-1}, C_{\mathbf{w},m} \mathbf{I}) p(d_m | \mathbf{h}_m, C_{v,m}).$$
(20)

Next, we take the natural logarithm ln (·) of $p(d_{1:n}, \mathbf{h}_{1:n}|\theta_{1:n})$, replace θ_n by $\{C_{v,n}^{\text{new}}, C_{\mathbf{w},n}^{\text{new}}\}$ and maximize the left-hand side of (19) with respect to the new parameters:

$$C_{\mathbf{w},n}^{\text{new}} = \underset{C_{\mathbf{w},n}^{\text{new}}}{\arg\max} \, \mathcal{E}_{\mathbf{h}_{1:n}|\theta_n} \{ \ln \left(p(\mathbf{h}_n | \mathbf{h}_{n-1}, C_{\mathbf{w},n}^{\text{new}} \mathbf{I}) \right) \} (21)$$

$$C_{v,n}^{\text{new}} = \underset{\substack{C_{v,n}^{\text{new}}\\v,n}}{\arg\max} \ \mathcal{E}_{\mathbf{h}_{1:n}|\theta_n}\{\ln\left(p(d_n|\mathbf{h}_n, C_{v,n}^{\text{new}})\right)\}, \quad (22)$$

starting with the estimation of $C_{v,n}^{\text{new}}$ by inserting

$$\ln(p(d_n|\mathbf{h}_n, C_{v,n}^{\text{new}})) \stackrel{(8)}{=} -\frac{\ln(2\pi C_{v,n}^{\text{new}})}{2} - \frac{(d_n - \mathbf{x}_n^T \mathbf{h}_n)^2}{2C_{v,n}^{\text{new}}}$$
(23)

into (22). This leads to the instantaneous estimate:

$$C_{v,n}^{\text{new}} = \mathcal{E}_{\mathbf{h}_{1:n}|\theta_n} \{ (d_n - \mathbf{x}_n^T \mathbf{h}_n)^2 \}$$
(24)

$$= d_n + \mathbf{x}_n^T (C_{\mathbf{h},n} \mathbf{I} + \mathbf{h}_n \mathbf{h}_n^T) \mathbf{x}_n - 2\mathbf{x}_n^T \mathbf{h}_n \qquad (25)$$

$$= (d_n - \mathbf{x}_n^T \mathbf{h}_n)^2 + \mathbf{x}_n^T \mathbf{x}_n C_{\mathbf{h},n}.$$
 (26)

The variance (of the microphone signal uncertainty) $C_{v,n}^{\text{new}}$ in (26) consists of two components, which can be interpreted as follows [25]: The first term in (26) is given as the squared error signal after filter adaptation and is influenced by near-end interferences like background noise. The second term in (26) depends on the signal energy $\mathbf{x}_n^T \mathbf{x}_n$ and the variance $C_{\mathbf{h},n}$ which implies that it considers uncertainties in the linear echo path model. Similar to the derivation for $C_{v,n}^{\text{new}}$, we insert

$$\frac{\ln(p(\mathbf{h}_{n}|\mathbf{h}_{n-1}, C_{\mathbf{w},n}^{\mathrm{new}}\mathbf{I}))}{= -\frac{M\ln(2\pi C_{\mathbf{w},n}^{\mathrm{new}})}{2} - \frac{(\mathbf{h}_{n} - \mathbf{h}_{n-1})^{T}(\mathbf{h}_{n} - \mathbf{h}_{n-1})}{2C_{\mathbf{w},n}^{\mathrm{new}}} (27)$$

into (21), to derive the instantaneous estimate of $C_{\mathbf{w},n}^{\text{new}}$:

$$C_{\mathbf{w},n}^{\text{new}} = \frac{1}{M} \mathcal{E}_{\mathbf{h}_{1:n}|\theta_n} \{ (\mathbf{h}_n - \mathbf{h}_{n-1})^T (\mathbf{h}_n - \mathbf{h}_{n-1}) \}$$
(28)
$$\stackrel{(12)}{=} C_{\mathbf{h},n} - C_{\mathbf{h},n-1} + \frac{1}{M} \left(\hat{\mathbf{h}}_n^T \hat{\mathbf{h}}_n - \hat{\mathbf{h}}_{n-1}^T \hat{\mathbf{h}}_{n-1} \right),$$
(29)

where we employed the statistical independence between \mathbf{w}_n and \mathbf{h}_{n-1} . Equation (29) implies the estimation of $C_{\mathbf{w},n}^{\text{new}}$ as difference of the filter tap autocorrelations between the time instants n and n - 1. Finally, the updated parameter values are used as initialization for the following time step, so that

$$C_{\mathbf{w},n+1} := C_{\mathbf{w},n}^{\text{new}}, \quad C_{v,n+1} := C_{v,n}^{\text{new}}.$$
 (30)

III. COMPARISON BETWEEN THE EM-NLMS ALGORITHM AND THE NLMS ALGORITHM PROPOSED IN [10]

In this part, we compare the proposed EM-NLMS algorithm to the NLMS algorithm reviewed in Section I and show the equality between the stepsizes in (5) and (17). We reformulate the update in (17) by applying the conditional independence rules for Bayesian networks [9]: First, we exploit the equalities

$$\mathbf{C}_{\mathbf{h},n} \stackrel{(12)}{=} C_{\mathbf{h},n} \mathbf{I} \stackrel{(13)}{=} \mathcal{E}\{(\mathbf{h}_n - \hat{\mathbf{h}}_n)(\mathbf{h}_n - \hat{\mathbf{h}}_n)^T\},\\ \mathbf{C}_{\mathbf{w},n} \stackrel{(9)}{=} C_{\mathbf{w},n} \mathbf{I} = \mathcal{E}\{\mathbf{w}_n \mathbf{w}_n^T\},$$
(31)

which lead to the following relations:

$$C_{\mathbf{h},n} = \frac{\mathcal{E}\{(\mathbf{h}_n - \hat{\mathbf{h}}_n)^T(\mathbf{h}_n - \hat{\mathbf{h}}_n)\}}{M} = \frac{\mathcal{E}\{\|\mathbf{h}_n - \hat{\mathbf{h}}_n\|_2^2\}}{M},$$
$$C_{\mathbf{w},n} = \frac{\mathcal{E}\{\mathbf{w}_n^T \mathbf{w}_n\}}{M} = \frac{\mathcal{E}\{\|\mathbf{w}_n\|_2^2\}}{M}.$$
(32)

Second, it can be seen in Fig. 2 that the state vector \mathbf{h}_{n-1} and the uncertainty \mathbf{w}_n are statistically independent as they share a head-to-head relationship with respect to the latent vector \mathbf{h}_n . As a consequence, the numerator in (17) can be rewritten as

$$C_{\mathbf{h},n-1} + C_{\mathbf{w},n} \stackrel{(32)}{=} \frac{\mathcal{E}\{\|\mathbf{h}_{n-1} - \mathbf{h}_{n-1}\|_{2}^{2}\}}{M} + \frac{\mathcal{E}\{\|\mathbf{w}_{n}\|_{2}^{2}\}}{M}$$
$$\stackrel{(8)}{=} \frac{\mathcal{E}\{\|\mathbf{h}_{n} - \hat{\mathbf{h}}_{n-1}\|_{2}^{2}\}}{M}.$$
(33)

Finally, we consider the mean of the squared error signal

$$\mathcal{E}\{e_n^2\} \stackrel{(2),(4)}{=} \mathcal{E}\{(\mathbf{x}_n^T(\mathbf{h}_n - \hat{\mathbf{h}}_{n-1}) + v_n)^2\}, \qquad (34)$$

which is not conditioned on the microphone signal d_n . By applying the conditional independence rules to the Bayesian network in Fig. 2, the head-to-head relationship with respect to d_n implies v_n to be statistically independent from \mathbf{h}_{n-1} and \mathbf{w}_n , respectively. Consequently, we can rewrite (34) as:

$$\mathcal{E}\left\{e_{n}^{2}\right\}^{(10)} = \mathbf{x}_{n}^{T} \mathcal{E}\left\{\left(\mathbf{h}_{n} - \hat{\mathbf{h}}_{n-1}\right)\left(\mathbf{h}_{n} - \hat{\mathbf{h}}_{n-1}\right)^{T}\right\} \mathbf{x}_{n} + C_{v,n}$$

$$\stackrel{(8),(31)}{=} \mathbf{x}_{n}^{T} \mathbf{x}_{n} (C_{\mathbf{h},n-1} + C_{\mathbf{w},n}) + C_{v,n}.$$
(35)

The insertion of (33) and (35) into the stepsize defined in (17) yields the identical expression for λ_n as in (5). Thus, the main contribution of the proposed EM-NLMS algorithm is that the model parameters $C_{\mathbf{h},n}$ and $C_{\mathbf{w},n}$ (and consequently the normalized stepsize λ_n) are estimated in the M step of the EM algorithm instead of being approximated using (6).

This paper previously published in IEEE Signal Processing Letters

TABLE II Realizations of the EM-NLMS Algorithm ("EM-NLMS"), the NLMS Algorithm due to [10] ("Adapt. NLMS") and the Conventional NLMS Algorithm ("Conv. NLMS")

	$\mathbf{\hat{h}}_n = \mathbf{\hat{h}}_{n-1} + \lambda_n \mathbf{x}_n e_n$
EM-NLMS	$\lambda_n = \frac{C_{\mathbf{h},n-1} + C_{\mathbf{w},n}}{\mathbf{x}_n^T \mathbf{x}_n (C_{\mathbf{h},n-1} + C_{\mathbf{w},n}) + C_{v,n} + \epsilon}$
	$C_{\mathbf{h},n} = \left(1 - \lambda_n \frac{\mathbf{x}_n^T \mathbf{x}_n}{M}\right) \left(C_{\mathbf{h},n-1} + C_{\mathbf{w},n}\right)$
	$C_{v,n+1} = \left(d_n - \mathbf{x}_n^T \mathbf{\hat{h}}_n\right)^2 + \mathbf{x}_n^T \mathbf{x}_n C_{\mathbf{h},n}$
	$C_{\mathbf{w},n+1} = C_{\mathbf{h},n} - C_{\mathbf{h},n-1} + \frac{\hat{\mathbf{h}}_n^T \hat{\mathbf{h}}_n - \hat{\mathbf{h}}_{n-1}^T \hat{\mathbf{h}}_{n-1}}{M}$
	$C_{\mathbf{h},0} = C_{\mathbf{w},0} = C_{v,0} = 0.1, \ \epsilon = 0.01$
Adapt. NLMS	$\lambda_n \approx \frac{1}{N_T} \frac{\sum\limits_{\kappa=0}^{N_T-1} \hat{h}_{k,n-1}^2}{(1-\eta)e_n^2 + \eta \mathcal{E}\{e_{n-1}^2\} + \epsilon}$
	$N_T = 5, \ \eta = 0.9, \ e_0^2 = 0.1, \ \epsilon = 0.01$
Conv. NLMS	$\lambda_n = \frac{0.5}{\mathbf{x}_n^T \mathbf{x}_n + \epsilon}, \ \epsilon = 0.01$

IV. EXPERIMENTAL RESULTS

This section focuses on the experimental verification of the EM-NLMS algorithm ("EM-NLMS") in comparison to the adaptive stepsize-NLMS algorithm described in Section I ("Adapt. NLMS") and the conventional NLMS algorithm ("Conv. NLMS") with a fixed stepsize. An overview of the algorithms including the individually tuned model parameters is shown in Table II. Note the regularization of all three stepsize updates by the additive constant $\epsilon = 0.01$ to avoid a division by zero. For the evaluation, we synthesize the microphone signal by convolution of the loudspeaker signal with an RIR vector measured in a room with $T_{60} = 100$ ms. This is realized at a sampling rate of 16 kHz (M = 512) for both white noise and a male speech signal as loudspeaker signals. Furthermore, background noise is simulated by adding Gaussian white noise at a global signal-to-noise ratio of 20 dB. The comparison is realized in terms of the stepsize α_n and the system distance Δh_n as a measure for the system identification performance:

$$\Delta h_n = 10 \log_{10} \frac{\|\mathbf{h}_n - \mathbf{h}_n\|_2^2}{\|\mathbf{h}_n\|_2^2} \, \mathrm{dB}, \quad \alpha_n = \lambda_n(\mathbf{x}_n^T \mathbf{x}_n). \tag{36}$$

The results for white noise as input signal are shown in Fig. 3. Note that in Fig. 3(a) the EM-NLMS shows the best system identification compared to the Adapt. NLMS and the Conv. NLMS. As depicted in Fig. 3(b), the stepsize α_n of the EM-NLMS is not affected by oscillations as the Adapt. NLMS.

For male speech as input signal, we improve the convergence of the Conv. NLMS by setting a fixed threshold to stop adaptation $(\alpha_n = 0)$ in speech pauses. Furthermore, the absolute value of λ_n for the Adapt. NLMS is limited to 0.5 (for a heuristic justification see [24]). As illustrated in Fig. 4(a), the EM-NLMS shows again the best system identification compared to the Adapt. NLMS and the Conv. NLMS. By focusing on a small time frame, we can see in Fig. 4(b) that the stepsize α_n of the EM-NLMS algorithm is not restricted to the values of 0 and 0.5 (as Conv. NLMS) and not affected by oscillations (as Adapt. NLMS).

Due to the instantaneous parameter updates in Table I (we apply one EM iteration at each time step), the number of real-valued multiplications per iteration step is only increased from 2M + 3 (Conv. NLMS) to 4M + 8 (EM-NLMS). Note that



Fig. 3. Comparison of the EM-NLMS algorithm ("EM-NLMS"), the NLMS algorithm due to [10] ("Adapt. NLMS") and the conventional NLMS algorithm ("Conv. NLMS") in terms of the system distance Δh_n and the stepsize α_n for white Gaussian noise as input signal.



Fig. 4. Comparison of the EM-NLMS algorithm ("EM-NLMS"), the NLMS algorithm due to [10] ("Adapt. NLMS") and the conventional NLMS algorithm ("Conv. NLMS") in terms of the system distance Δh_n and the stepsize α_n (short time frame for visualization purposes) for male speech as input signal (see the microphone signal d_n in Fig. 4(c)).

the derived stepsize rule is also promising for applications like drift tracking, that the estimation of $C_{v,n}$ and $C_{w,n}$ converges to a locally optimum solution and that multiple EM iterations (at each time step) might also be of interest.

V. CONCLUSION

A probabilistic model for linear system identification has been defined to address the task of linear AEC from a Bayesian network perspective. We apply the EM algorithm and derive an NLMS-like filter adaptation. As main innovation, the stepsize is shown to be equivalent to the commonly accepted optimum stepsize in [10] and estimated in the M step of the EM algorithm (instead of being approximated by artificially extending the acoustic echo path). The resulting approach is denoted as EM-NLMS algorithm and experimentally verified for the task of linear AEC with simulated scenarios. By deriving an improved estimate for the optimum NLMS stepsize, this article exemplifies the benefit of applying machine learning techniques to classical signal processing tasks.

This paper previously published in IEEE Signal Processing Letters

REFERENCES

- B. J. Frey and N. Jojic, "A comparison of algorithms for inference and learning in probabilistic graphical models," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 27, no. 9, pp. 1392–1416, Sep. 2005.
- [2] T. Adali, D. Miller, K. Diamantaras, and J. Larsen, "Trends in machine learning for signal processing [in the spotlight]," *IEEE Signal Process. Mag.*, vol. 28, no. 6, pp. 193–196, Nov. 2011.
- [3] J. A. Bilmes and C. Bartels, "Graphical model architectures for speech recognition," *IEEE Signal Process. Mag.*, vol. 22, no. 5, pp. 89–100, Sep. 2005.
- [4] S. J. Rennie, P. Aarabi, and B. J. Frey, "Variational probabilistic speech separation using microphone arrays," *IEEE Trans. Audio, Speech Lang. Process.*, vol. 15, no. 1, pp. 135–149, Jan. 2007.
- [5] M. J. Wainwright and M. I. Jordan, "Graphical models, exponential families, and variational inference," *Found. Trends Mach. Learn.*, vol. 1, no. 1–2, pp. 1–305, Dec. 2008.
- [6] D. Barber and A. Cemgil, "Graphical models for time-series," *IEEE Signal Process. Mag.*, Nov. 2010.
- [7] C. W. Maina and J. M. Walsh, "Joint speech enhancement and speaker identification using approximate Bayesian inference," *IEEE Trans. Audio, Speech Lang. Process.*, vol. 19, no. 6, pp. 1517–1529, Aug. 2011.
- [8] R. Maas, C. Huemmer, A. Schwarz, C. Hofmann, and W. Kellermann, "A bayesian network view on linear and nonlinear acoustic echo cancellation," in *IEEE China Summit Int. Conf. Signal Inform. Process.* (*ChinaSIP*), Xi'an, China, Jul. 2014, pp. 495–499.
- [9] C. M. Bishop, *Pattern Recognition and Machine Learning*, 8th ed. Berlin, Germany: Springer, 2006.
- [10] S. Yamamoto and S. Kitayama, "An adaptive echo canceller with variable step gain method," *Trans. IECE Jpn.*, vol. E65, no. 1, pp. 1–8, Jan. 1982.
- [11] H.-C. Huang and J. Lee, "A variable step size LMS algorithm," *IEEE Trans. Signal Process.*, vol. 40, no. 7, pp. 1633–1642, Jul. 1992.
- [12] T. Aboulnasr and K. Mayyas, "A robust variable step-size LMS-type algorithm: Analysis and simulations," *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 631–639, Jul. 1992.

- [13] A. Mader, H. Puder, and G. U. Schmidt, "Step-size control for acoustic echo cancellation filters—an overview," *Signal Process.*, vol. 80, no. 9, pp. 1697–1719, Sep. 2000.
- [14] H.-C. Shin, A. H. Sayed, and W.-J. Song, "Variable step-size NLMS and affine projection algorithms," *IEEE Signal Process. Lett.*, vol. 11, no. 2, pp. 132–135, Feb. 2004.
- [15] J. Benesty, H. Rey, L. R. Vega, and S. Tressens, "A nonparametric VSS NLMS algorithm," *IEEE Trans. Signal Process.*, vol. 13, no. 10, pp. 581–584, Oct. 2006.
- [16] P. A. C. Lopes and J. B. Gerald, "New normalized LMS algorithms based on the Kalman filter," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Process. (ICASSP)*, May 2007, pp. 117–120.
- [17] M. Asif Iqbal and S. L. Grant, "Novel variable step size NLMS algorithms for echo cancellation," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Process. (ICASSP)*, Apr. 2008, pp. 241–244.
- [18] C. Paleologu, J. Benesty, S. L. Grant, and C. Osterwise, "Variable step-size NLMS algorithms designed for echo cancellation," in *IEEE Rec. 43rd Asilomar Conf. Signals, Syst. and Comput.*, Nov. 2009, pp. 633–637.
- [19] J.-K. Hwang and Y.-P. Li, "Variable step-size LMS algorithm with a gradient-based weighted average," *IEEE Signal Process. Lett.*, vol. 16, no. 12, pp. 1043–1046, Dec. 2009.
- [20] H.-C. Huang and J. Lee, "A new variable step-size NLMS algorithm and its performance analysis," *IEEE Trans. Signal Process.*, vol. 60, no. 4, pp. 2055–2060, Apr. 2012.
- [21] H. Zhao and Y. Yu, "Novel adaptive VSS-NLMS algorithm for system identification," in *IEEE 4th Int. Conf. Intell. Control Inform. Process.* (ICICIP), Jun. 2013, pp. 760–764.
- [22] S. Haykin, Adaptive Filter Theory. Upper Saddle River, NJ, USA: Prentice-Hall, 2002.
- [23] C. Breining, P. Dreiseitel, E. Hänsler, A. Mader, B. Nitsch, H. Puder, T. Schertler, G. Schmidt, and J. Tilp, "Acoustic echo control," *Signal Process.*, vol. 16, no. 4, pp. 42–69, Jul. 1999.
- [24] U. Schultheiß, "Über die Adaption eines Kompensators f
 ür akustische Echos," Ph.D. thesis, TU Darmstadt, Darmstadt, Germany, 1988.
- [25] R. Maas, C. Huemmer, C. Hofmann, and W. Kellermann, "On Bayesian networks in speech signal processing," in *ITG Conf. Speech Commun.*, Erlangen, Germany, Sep. 2014.