IDENTIFICATION OF LINEAR SYSTEMS WITH ADAPTIVE CONTROL OF THE CROSS-MULTIPLICATIVE TRANSFER FUNCTION APPROXIMATION

Yekutiel Avargel and Israel Cohen

Department of Electrical Engineering, Technion - Israel Institute of Technology
Technion City, Haifa 32000, Israel
{kutiav@tx,icohen@ee}.technion.ac.il

ABSTRACT

In this paper, we extend the cross-multiplicative transfer function (CMTF) approach for improved system identification in the short-time Fourier transform (STFT) domain. The proposed algorithm adaptively controls the number of cross-terms in the CMTF approximation to achieve the minimum mean-square error (mmse) at each iteration. A small number of cross-terms is initially used to achieve fast convergence, and as the adaptation process proceeds, the algorithm gradually increases this number to enhance the steady-state performance. When compared to the conventional multiplicative transfer function (MTF) approach, the resulting algorithm achieves a substantial improvement in steady-state performance, without compromising for slower convergence. Experimental results validate the theoretical derivations and demonstrate the advantage of the proposed approach to acoustic echo cancellation.

Index Terms— System identification, time-frequency analysis, multiplicative transfer function, subband adaptive filtering.

1. INTRODUCTION

Linear systems in the short-time Fourier transform (STFT) domain are often modeled by multiplicative transfer functions (MTFs) (e.g., [1–4]). The MTF approximation relies on the assumption that the support of the STFT analysis window is sufficiently large compared to the duration of the system impulse response. Recently, we proposed a cross-MTF (CMTF) approximation for representing linear systems in the STFT domain by introducing cross-multiplicative terms between distinct subbands [5]. We showed that compared to the MTF approximation, the CMTF approximation is associated with slower convergence, but smaller steady-state mean-square error (mse). However, since this algorithm employs a fixed number of cross-terms during the adaptation process, it may suffer from either slow convergence in case the number of cross-terms is large, or relatively high steady-state mse in case the number of cross-terms is small.

In this paper, we extend the CMTF approach and propose to adaptively control the number of cross-terms. The proposed algorithm finds the optimal number of cross terms and achieves the minimum mse (mmse) at each iteration. At the beginning of the adaptation process, the proposed algorithm is initialized by a small number of cross-terms to achieve fast convergence, and as the adaptation process proceeds, it gradually increases this number to improve the steady-state performance. This is done by simultaneously updating three system models, each consisting of different (but consecutive) number of cross-terms, and determining the optimal number using an appropriate decision rule. When compared to the conventional MTF approach, the resulting algorithm achieves a substantial improvement in steady-state performance, without degrading its convergence rate. Experimental results validate the theoretical derivations and demonstrate the advantage of the proposed approach for acoustic echo cancellation.

The paper is organized as follows. In Section 2, we introduce the CMTF approximation for system identification in the STFT domain. In Section 3, we present an CMTF adaptation procedure using a fixed number of cross-terms. In Section 4, we adaptively control the number of cross-terms. Finally, in Section 5, we present experimental results which verify the theoretical derivations.

2. CROSS-MTF APPROXIMATION

Let an input $x(n)$ and output $y(n)$ of an unknown linear time-invariant (LTI) system be related by

$$y(n) = h(n) * x(n) + \xi(n) \triangleq d(n) + \xi(n)\,, \tag{1}$$

where $h(n)$ represents the impulse response of the system, $\xi(n)$ is an additive noise signal, $d(n)$ is the signal component in the system output, and $*$ denotes convolution. Applying the STFT to $y(n)$, we have in the time-frequency domain

$$y_{p,k} = d_{p,k} + \xi_{p,k}\,, \tag{2}$$

where p is the frame index and k represents the frequency-bin index ($0 \le k \le N-1$). To perfectly represent an LTI system in the STFT domain, crossband filters between subbands are generally required [1, 6]. The widely-used MTF approximation [2] avoids these crossband filters by assuming that the STFT analysis window is long and smooth relative to the impulse response $h(n)$, so that the transfer function is approximated as multiplicative in the STFT domain:

$$d_{p,k} \approx h_k\, x_{p,k}\,, \tag{3}$$

This research was supported by the Israel Science Foundation (grant no. 1085/05).

where $h_k \triangleq \sum_{m=0}^{N_h-1} h(m)\exp(-j2\pi mk/N)$ and N_h is the length of $h(n)$. In case of finite length input signals, the MTF approximation is insufficient, since a longer analysis window comes at the expense of fewer observations that become available in each frequency bin [2].

An CMTF approximation for modeling an LTI system in the STFT domain is obtained by including cross-multiplicative terms between distinct subbands. Let $h_{k,k'}$ denote a cross-term from frequency bin k' to frequency bin k. Then an CMTF approximation of $d_{p,k}$ by $2K+1$ cross-terms around frequency bin k is given by

$$d_{p,k} \approx \sum_{k'=k-K}^{k+K} h_{k,k' \bmod N}\, x_{p,k' \bmod N}\,. \tag{4}$$

Note that for $K=0$, (4) reduces to the MTF approximation (3).

3. CONVENTIONAL CMTF ADAPTATION

In this section, we present an LMS-based adaptive algorithm for estimating the cross-terms in each frequency bin. Let $\hat{d}_{p,k}$ be an estimate of $d_{p,k}$ with $2K+1$ cross-terms:

$$\hat{d}_{p,k} = \sum_{k'=k-K}^{k+K} x_{p,k'}\hat{h}_{k,k'}(p)\,, \tag{5}$$

where $\hat{h}_{k,k'}(p)$ is an adaptive cross-term that represents an estimate of $h_{k,k'}$ at frame index p (recall that due to periodicity of the frequency bins, the summation index k' is related to frequency bin $k' \bmod N$). Let $\hat{\mathbf{h}}_k(p) = [\ \hat{h}_{k,k-K}(p) \quad \cdots \quad \hat{h}_{k,k+K}(p)\ ]^T$ denote $2K+1$ adaptive cross-terms at the kth frequency bin, and let $\mathbf{x}_k(p) = [\ x_{p,k-K} \quad \cdots \quad x_{p,k+K}\ ]^T$ be the input data vector corresponding to $\hat{\mathbf{h}}_k(p)$. Then (5) can be rewritten as

$$\hat{d}_{p,k} = \mathbf{x}_k^T(p)\hat{\mathbf{h}}_k(p)\,. \tag{6}$$

The $2K+1$ cross-terms are updated using the LMS algorithm by

$$\hat{\mathbf{h}}_k(p+1) = \hat{\mathbf{h}}_k(p) + \mu e_{p,k}\mathbf{x}_k^*(p) \tag{7}$$

where $e_{p,k} = y_{p,k} - \hat{d}_{p,k}$ is the error signal in the kth frequency bin, $y_{p,k}$ is defined in (2), and μ is a step-size. Let

$$\epsilon_k(p) = E\{|e_{p,k}|^2\} \tag{8}$$

denote the transient mse in the kth frequency bin. Then, assuming that $x_{p,k}$ and $\xi_{p,k}$ are uncorrelated zero-mean white Gaussian signals, the mse can be expressed recursively as [5]

$$\epsilon_k(p+1) = \alpha(K)\,\epsilon_k(p) + \beta_k(K)\,, \tag{9}$$

where $\alpha(K)$ and $\beta_k(K)$ depend on the step-size μ and the number of cross-terms K. Accordingly, it can be shown [5] that the optimal step-size that results in the fastest convergence for each K is given by

$$\mu_{\text{opt}} = \frac{1}{2\sigma_x^2(K+1)}\,, \tag{10}$$

where σ_x^2 is the variance of $x_{p,k}$. Equation (10) indicates that as the number of cross-terms increases (K increases), a smaller step-size has to be utilized. Consequently, the MTF approximation ($K=0$) is associated with faster convergence, but suffers from higher steady-state mse $\epsilon_k(\infty)$. Estimation of additional cross-terms results in a slower convergence, but improves the steady-state mse. Since the number of cross-terms is fixed during the adaptation process, this algorithm may suffer from either slow convergence (typical to large K) or relatively high steady-state mse (typical to small K). To improve both the convergence rate and the steady-state mse, the number of cross-terms at each iteration should be adaptively controlled, as discussed in the following section.

4. ADAPTIVE CONTROL OF CROSS-TERMS

In this section, we adaptively control the number of cross-terms to achieve both faster convergence and smaller steady-state mse, compared to using a fixed number of cross-terms. The strategy of controlling the number of cross-terms is related to filter-length control (e.g., [7, 8]). However, existing length-control algorithms operate in the time domain, focusing on linear FIR adaptive filters. Here, we extend the approach presented in [7] to construct an adaptive control procedure for CMTF adaptation implemented in the STFT domain.

4.1. Proposed Algorithm Description

The main objective of the proposed algorithm is to find the optimal number of cross-terms that achieves the mmse at each iteration. Let

$$K_{\text{opt}}(p) = \arg\min_{K} \epsilon_k(p)\,. \tag{11}$$

Then, $2K_{\text{opt}}(p)+1$ denotes the optimal number of cross-terms at iteration p. It was shown in the previous section that as more data is employable in the adaptation process (i.e., the frame index p increases), we expect to attain a lower mse by increasing the number of cross-terms. Therefore, the proposed algorithm should initially select a small number of cross-terms (usually $K=0$) to achieve initial fast convergence, and then, as the adaptation process proceeds, it should gradually increase this number to achieve the desired steady-state performance. This is done by simultaneously updating three system models, each consists of different number of cross-terms. Specifically, let $\hat{\mathbf{h}}_{1k}(p)$, $\hat{\mathbf{h}}_{2k}(p)$ and $\hat{\mathbf{h}}_{3k}(p)$ denote three vectors of $2K_1(p)+1$, $2K_2(p)+1$ and $2K_3(p)+1$ adaptive cross-terms, respectively. At the beginning of the adaptation ($p=0$), the number of cross-terms in each vector is initialized to $K_1(0) = K_0 - 1$, $K_2(0) = K_0$ and $K_3(0) = K_0 + 1$, where K_0 is a constant integer. Then, these vectors are updated simultaneously at each iteration using the normalized LMS (NLMS) algorithm

$$\hat{\mathbf{h}}_{ik}(p+1) = \hat{\mathbf{h}}_{ik}(p) + \frac{\mu_i(p)}{\|\mathbf{x}_{ik}(p)\|^2} e_{p,k}^i \mathbf{x}_{ik}^*(p) \tag{12}$$

where $i = 1, 2, 3$, $\mathbf{x}_{ik}(p) = [\ x_{p,k-K_i(p)} \quad \cdots \quad x_{p,k+K_i(p)}\ ]^T$, $e_{p,k}^i = y_{p,k} - \mathbf{x}_{ik}^T(p)\hat{\mathbf{h}}_{ik}(p)$ is the resulting error signal, and $\mu_i(p)$ is the relative step-size. Since the step-size should be inversely proportional to the number of cross-terms [see (10)], we choose $\mu_i(p) = M/(K_i(p)+1)$, with M being a constant parameter. The second adaptive vector $\hat{\mathbf{h}}_{2k}(p)$ is the vector of interest as its coeffi-

cients are used for estimating the desired signal $d_{p,k}$, i.e.,

$$\hat{d}_{p,k} = \mathbf{x}_{2k}^T(p)\hat{\mathbf{h}}_{2k}(p)\,. \tag{13}$$

Therefore, the dimension of $\hat{\mathbf{h}}_{2k}(p)$, $2K_2(p)+1$, should represent the optimal number of cross-terms in each iteration. For this purpose, we define the following averages

$$\epsilon_{ik}(p) = \frac{1}{P}\sum_{q=p-P+1}^{p} |e_{q,k}^i|^2\,,\ i=1,2,3 \tag{14}$$

for the mse estimate at the pth iteration, where P is a constant parameter. These averages are computed every P frames, and the value of $K_2(p)$ is then determined by the following decision rule:

$$K_2\,(p+1) = \begin{cases} K_2(p)+1 & ;\text{if } \epsilon_{1k}(p) > \epsilon_{2k}(p) > \epsilon_{3k}(p) \\ K_2(p) & ;\text{if } \epsilon_{1k}(p) > \epsilon_{2k}(p) \le \epsilon_{3k}(p) \\ K_2(p)-1 & ;\text{otherwise} \end{cases}. \tag{15}$$

Accordingly, $K_1(p+1)$ and $K_3(p+1)$ are updated by

$$\begin{aligned} K_1(p+1) &= K_2\,(p+1)-1\,, \\ K_3(p+1) &= K_2\,(p+1)+1\,, \end{aligned} \tag{16}$$

and the adaptation proceeds by updating the resized vectors $\hat{\mathbf{h}}_{ik}(p)$ using (12). Note that the parameter P should be sufficiently small to enable tracking during variations in the optimal number of cross-terms, and sufficiently large to achieve an efficient approximation of the mse by (14).

The decision rule in (15) can be explained as follows. When the optimum number of cross-terms is equal or larger than $K_3(p)$, then $\epsilon_{1k}(p) > \epsilon_{2k}(p) > \epsilon_{3k}(p)$ and all values are increased by one. In this case, the vectors are reinitialized by $\hat{\mathbf{h}}_{1k}(p+1) = \hat{\mathbf{h}}_{2k}(p)$, $\hat{\mathbf{h}}_{2k}(p+1) = \hat{\mathbf{h}}_{3k}(p)$, and $\hat{\mathbf{h}}_{3k}(p+1) = \left[0\ \hat{\mathbf{h}}_{3k}^T(p)\ 0\right]^T$. When $K_2(p)$ is the optimum number, then $\epsilon_{1k}(p) > \epsilon_{2k}(p) \le \epsilon_{3k}(p)$ and the values remain unchanged. Finally, when the optimum number is equal or smaller than $K_1(p)$, we have $\epsilon_{1k}(p) \le \epsilon_{2k}(p) < \epsilon_{3k}(p)$ and all values are decreased by one. In this case, we reinitialize the vectors by $\hat{\mathbf{h}}_{3k}(p+1) = \hat{\mathbf{h}}_{2k}(p)$, $\hat{\mathbf{h}}_{2k}(p+1) = \hat{\mathbf{h}}_{1k}(p)$, and $\hat{\mathbf{h}}_{1k}(p+1)$ is obtained by eliminating the first and last elements of $\hat{\mathbf{h}}_{1k}(p)$. The decision rule is aimed at reaching the minimal mse for each frequency bin separately. That is, distinctive frequency bins may have different values of $K_2(p)$ at each frame index p. Clearly, this decision rule is unsuitable for applications where the error signal to be minimized is in the time domain. In such cases, the optimal number of cross-terms is the one that minimizes the time-domain mse $E\{|e(n)|^2\}$ [contrary to (11)]. Therefore, we use the following averages

$$\epsilon_i(n) = \frac{1}{\tilde{P}}\sum_{m=n-\tilde{P}+1}^{n} |e_i(m)|^2\,,\ i=1,2,3 \tag{17}$$

for estimating the time-domain mse, where $e_i(n)$ is the inverse STFT of $e_{p,k}^i$, $\tilde{P} \triangleq (P-1)\,L+N$, and L is the translation factor of the STFT. Then, as in (14), these averages are computed every P frames (corresponding to PL time-domain iterations), and $K_2(n)$ is determined similarly to (15) by substituting $\epsilon_i(n)$ for $\epsilon_{ik}(p)$ and n for p. Note that now all frequency bins have the same number of cross-terms $[2K_2(p)+1]$ at each frame. The two proposed decision rules, for both time and STFT domains adaptation, will be further demonstrated in the next section.

4.2. Computational Complexity

Updating $2K+1$ cross-terms using the NLMS adaptation formula (12), requires $8K+6$ arithmetic operations for every L input samples [5]. Therefore, since three vectors of cross-terms are updated simultaneously in each frame, the adaptation process of the proposed approach requires $8\,[K_1(p)+K_2(p)+K_3(p)]+6$ arithmetic operations. Using (16) and computing the desired signal estimate (6), the overall complexity of the proposed approach is given by $28K_2(p)+7$ arithmetic operation for every L input samples and each frequency bin. The computations required for updating $K_2(p)$ [see (14)-(16)] are relatively negligible, since they are carried out only once every P iterations. When compared to the conventional MTF approach ($K=0$), the proposed approach involves an increase of $28K_2(p)+1$ arithmetic operations for every L input samples and every frequency bin.

5. EXPERIMENTAL RESULTS

In this section, we present experimental results which verify the theoretical analysis and demonstrate the effectiveness of the proposed approach. In the first experiment, we examine the proposed approach performance in the STFT domain for white Gaussian signals. That is, the input signal $x(n)$ and the additive noise signal $\xi(n)$ are uncorrelated zero-mean white Gaussian processes with variances $\sigma_x^2 = 1$ and $\sigma_\xi^2 = 0.001$, respectively. We model the impulse response as a stochastic process with an exponential decay envelope, i.e., $h(n) = u(n)\beta(n)e^{-0.02n}$, where $u(n)$ is the unit step function and $\beta(n)$ is a unit-variance zero-mean white Gaussian noise. The impulse response length is set to $N_h = 16$, and a Hamming synthesis window of length $N = 128$ with 50% overlap is employed. Figure 1 shows the transient mse curves $\epsilon_k(p)$ of both the CMTF approach with fixed number of cross-terms, and the proposed approach with variable number of cross-terms. The cross-terms in the first approach are updated by the NLMS adaptation formula (12) using $M=0.1$. For the proposed approach, we use $K_0=0$, $P=30$ and $M=0.1$. Results are averaged out over 2000 independent runs. The results confirm that when the number of cross-terms is fixed during the adaptation process, a lower steady-state mse is achieved with increasing K, but at the expense of a slower convergence. Contrarily, the proposed algorithm achieves the lowest steady-state mse with a convergence rate comparable to that of the conventional MTF approach ($K=0$). In particular, a decrease of 13 dB in the mse is obtained by the proposed approach, when compared to the MTF approach. The bottom of Fig. 1 compares $K_2(p)$, which determines the number of cross-terms selected by the proposed algorithm at iteration p, to the optimal number of cross-terms $K_{\text{opt}}(p)$ [see (11)]. Clearly, the number of estimated cross-terms increases as more data is available in the adaptation process. The proposed algorithm well predicts the optimal value $K_{\text{opt}}(p)$, which enables to achieve the minimal mse at each iteration.

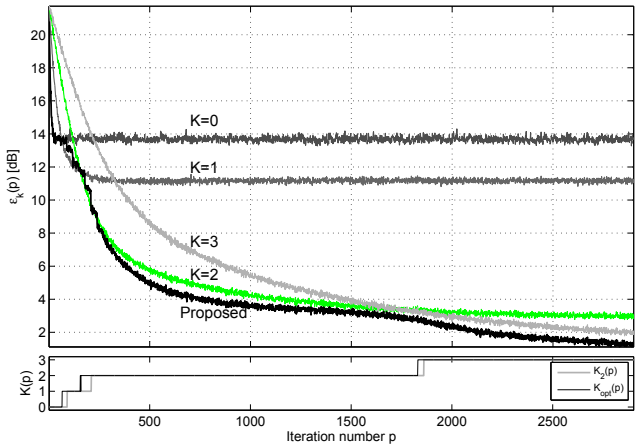

Fig. 1. Transient mse curves for white Gaussian signals, obtained by adaptively updating a fixed number of cross-terms ($K = 0, 1, 2$ and 3), and by using the proposed approach. $K_2(p)$ and $K_{\text{opt}}(p)$ are compared at the bottom.

In the second experiment, we demonstrate the proposed approach in an acoustic echo cancellation application using real speech signals. We use an ordinary office with a reverberation time T_{60} of about 100 ms. In this experiment, the signals are sampled at 16 kHz. A far-end speech signal $x(n)$ is generated by a loudspeaker and received by a microphone as an echo signal $d(n)$ together with a near-end speech signal and local noise [collectively denoted by $\xi(n)$]. The distance between the near-end source and the microphone is 1 m. The effective length of the echo path is 100 ms ($N_h = 1600$). The STFT is implemented with a Hamming synthesis window of length $N = 3200$ and 50% overlap. The acoustic echo canceller (AEC) performance is evaluated by the echo-return loss enhancement (ERLE), defined in dB by

$$\text{ERLE} = 10\log_{10}\frac{E\{y^2(n)\}}{E\{e^2(n)\}}, \tag{18}$$

where $e(n)$ is the inverse STFT of $e_{p,k}$. Figures 2(a)–(b) show the far-end and microphone signals, respectively, where a double-talk situation (simultaneously active far-end and near-end speakers) occurs between 3.4 s and 4.4 s (indicated by two vertical dotted lines). Figures 2(c)–(d) show the error signal $e(n)$ obtained by the CMTF approach with a fixed number of cross-terms ($K = 0$ and $K = 2$, respectively), and Fig. 2(e) shows the error signal obtained by the proposed approach. Other simulation parameters are $K_0 = 0$, $P = 5$ and $M = 1$. In this case, the time-domain decision rule, based on the mse estimate in (17), is employed. The ERLE values of the corresponding error signals were computed after convergence of the algorithms, and are given by 12.8 dB ($K = 0$), 16.5 dB ($K = 2$), and 18.6 dB (proposed). Clearly, the proposed algorithm achieves both fast convergence as the MTF approach and high ERLE as the CMTF approach, while adaptively controlling the number of cross-terms.

6. CONCLUSIONS

We have introduced a new algorithm for system identification in the STFT domain, which relies on the recently proposed CMTF approximation. Instead of using a fixed number of cross-terms, the proposed algorithm adaptively controls the number of cross-terms in each iteration, and enables to achieve faster convergence without compromising for higher steady-state mse.

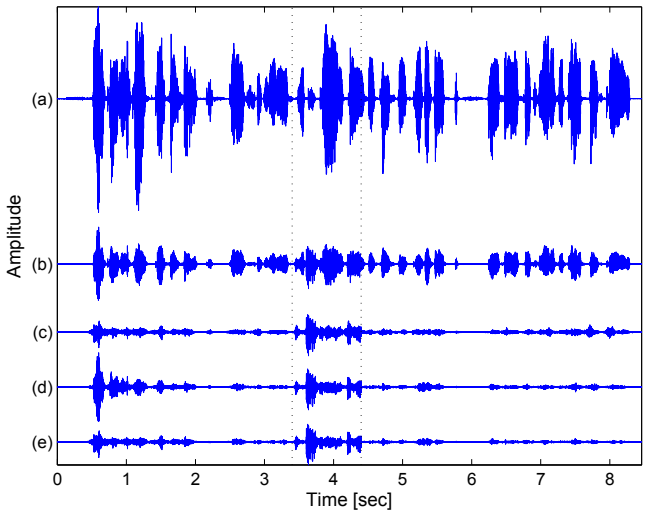

Fig. 2. Speech waveforms and error signals. A double-talk situation is indicated by vertical dotted lines. (a) Far-end signal (b) Microphone signal. (c)–(d) Error signals obtained by using the CMTF approach with fixed number of cross-terms: $K = 0$ and $K = 2$, respectively. (e) Error signal obtained by the proposed algorithm.

7. REFERENCES

[1] Y. Avargel and I. Cohen, "System identification in the short-time Fourier transform domain with crossband filtering," *IEEE Trans. Audio, Speech, Language Process.*, vol. 15, no. 4, pp. 1305–1319, May 2007.

[2] ——, "On multiplicative transfer function approximation in the short-time Fourier transform domain," *IEEE Signal Process. Lett.*, vol. 14, no. 5, pp. 337–340, May 2007.

[3] I. Cohen, "Relative transfer function identification using speech signals," *IEEE Trans. Speech Audio Process.*, vol. 12, no. 5, pp. 451–459, Sept. 2004.

[4] P. Smaragdis, "Blind separation of convolved mixtures in the frequency domain," *Neurocomputing*, vol. 22, no. 1-3, pp. 21–34, Nov. 1998.

[5] Y. Avargel and I. Cohen, "Adaptive system identification in the short-time Fourier transform domain using cross-multiplicative transfer function approximation," *IEEE Trans. Audio, Speech, Language Process.*, vol. 16, no. 1, pp. 162–173, Jan. 2008.

[6] A. Gilloire and M. Vetterli, "Adaptive filtering in subbands with critical sampling: Analysis, experiments, and application to acoustic echo cancellation," *IEEE Trans. Signal Process.*, vol. 40, no. 8, pp. 1862–1875, Aug. 1992.

[7] R. C. Bilcu, P. Kuosmanen, and K. Egiazarian, "On length adaptation for the least mean square adaptive filters," *Signal Process.*, vol. 86, pp. 3089–3094, Oct. 2006.

[8] Y. Gong and C. F. N. Cowan, "An LMS style variable tap-length algorithm for structure adaptation," *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2400–2407, July 2005.