A STATE-SPACE PARTITIONED-BLOCK ADAPTIVE FILTER FOR ECHO CANCELLATION USING INTER-BAND CORRELATIONS IN THE KALMAN GAIN COMPUTATION

María Luis Valero, Edwin Mabande and Emanuël A. P. Habets

International Audio Laboratories Erlangen*
Am Wolfsmantel 33, 91058 Erlangen, Germany
{maria.luis.valero,emanuel.habets}@audiolabs-erlangen.de

ABSTRACT

A partitioned-block-based architecture for a model-based acoustic echo canceller in the frequency domain was recently presented. Partitioned-block-based frequency domain adaptive filters provide a lower algorithmic delay compared to the non-partitioned formulations, which is achieved by partitioning the acoustic echo path and hence shortening the time-frequency transforms. Under these circumstances, avoiding the linearisation of the involved circular convolutions can significantly impair the performance. This paper extends the already proposed diagonalization of the Kalman gain matrix by taking into account the neglected inter-band correlations to improve the performance of the acoustic echo canceller. This comes at the cost of a moderate increase in the algorithmic complexity.

***Index Terms*—** Acoustic Echo Cancellation, Frequency Domain Adaptive Filters, Kalman Filter

1. INTRODUCTION

In a loudspeaker-enclosure-microphone environment, the signal reproduced by the loudspeaker propagates through the near-end room and is acquired by the microphone. Acoustic echo cancellation (AEC), [1, 2], is the most commonly used technique to prevent that the acoustic echo signal is transmitted back to the far-end. AEC uses adaptive filtering techniques, [3], to estimate the acoustic echo signal, which is subtracted from the microphone signal prior to transmission. Frequency domain adaptive filters (FDAF), [4], are known to present less algorithmic complexity as their time domain counterparts. This is the consequence of using fast convolutions instead of the highly complex time domain ones.

It is known that fast convolutions deliver only a limited number of coefficients that correspond to a linear convolution [4, 5]. Consequently, only the linear components have to be selected to ensure the optimal performance of the FDAF. The selection is implemented using a constraint, which is computationally complex as it involves transforming the signal blocks back and forth several times per frame or performing highly time-consuming matrix multiplications in the frequency domain. In the past, several approaches were presented which addressed the simplification of the constraints, [6–11]. Their aim was to reduce the algorithmic complexity by either reducing the number of constraints or simplifying them in the frequency domain.

A model-based AEC was proposed in [12, 13], which uses a Kalman filter, [14], in the frequency domain for the echo path estimation. The main advantages of the model-based AEC are the robustness against doubletalk, [15], and that it inherently estimates the residual echo after AEC, which can be used in a post-filter. A partitioned-block (PB) based formulation, based on [16], was presented in [17] which reduces the overall delay, enabling its use in modern communication devices. In [17] it was proposed to simplify the constraints in the frequency domain by using only the band-to-band correlations. Yet, this approximation is only valid if the time-frequency transforms are long enough [8], and this requirement is not always met by PB-FDAF. In [11] an analysis on the use of cross-band filters for AEC in the short-time Fourier transform domain was presented, which showed that using inter-band correlations enhances the AEC performance if the signal-to-noise ratio (SNR) is sufficiently high. In the following, an alternative formulation to the one in [17] is presented, which uses inter-band correlations only in the Kalman gain matrix computation. Thus, all involved covariance matrices are still formulated using only band-to-band correlations. The proposed Kalman gain matrix computation increases the algorithmic complexity moderately compared to the simplification in [17].

2. PARTITIONED-BLOCK-BASED FORMULATION

In a hands-free communication scenario, the signal acquired by the microphone, $y(n)$, at discrete time index n is usually described by

$$y(n) = x(n) * h(n) + s(n) + u(n) = d(n) + r(n), \quad (1)$$

where $d(n)$ is the acoustic echo signal, which is the result of the propagation of the far-end signal, $x(n)$, through the near-end room. In the following, it is assumed that the acoustic echo path, $h(n)$, can be modelled by a finite impulse response (FIR) filter of length L. In addition, only one interference signal, $r(n)$, is taken into account which comprises both the near-end speech, $s(n)$, and the background noise, $u(n)$. In AEC, the adaptive algorithm is driven by the error signal after adaptation,

$$e(n) = y(n) - x(n) * \hat{h}(n) = d(n) - \hat{d}(n) + r(n), \quad (2)$$

where estimation is denoted by the superscript $\hat{\cdot}$. In order to obtain a PB-based formulation, $h(n)$ is partitioned into B blocks of length $N = L/B$. Hence, at frame k for block b

$$\mathbf{h}^b(k) = [h_k(bN), \ldots, h_k(bN + N - 1)]^{\mathrm{T}}, \quad (3)$$

$$\mathbf{x}^b(k) = [x(kR - bN - M + 1), \ldots, x(kR - bN)]^{\mathrm{T}}, \quad (4)$$

where $M > N$ is the length of the input signal frame if the problem is formulated using the overlap-save method, [4, 5], and R denotes the frame shift. Consequently, the length of the zero padding needed to extend $\mathbf{h}^b(k)$ to match the length of the discrete Fourier transform (DFT), M, is $V = M - N$. Hereafter, the time-frequency domain counterparts of the time domain signal blocks in (3) and (4) are

$$\mathbf{H}^b(k) = \mathbf{F} \begin{bmatrix} \mathbf{h}^b(k) \\ \mathbf{0}_{V\times 1} \end{bmatrix} \text{ and } \mathbf{X}^b(k) = \mathrm{diag}\{\mathbf{F}\mathbf{x}^b(k)\}, \quad (5)$$

*A joint institution of the Friedrich-Alexander-University Erlangen-Nürnberg (FAU) and Fraunhofer IIS, Germany.

respectively, where $\mathrm{diag}\{\mathbf{a}\}$ generates a diagonal matrix with the elements of $\mathbf{a}$ on its main diagonal. In (5) $\mathbf{F}$ denotes the $M \times M$ DFT matrix, whose elements are $F(a,b) = W_M^{ab} = \exp(-j2\pi ab/M)$ with $j = \sqrt{-1}$. The major benefit of FDAF algorithms is the complexity reduction obtained from the frequency domain convolutions. Yet, these are equivalent to circular convolutions in the time domain. Thus, to obtain an optimal behaviour of the algorithm, the wrap-around components have to be rejected. In the time-frequency domain (2) is equivalent to

$$\mathbf{E}(k) = \mathbf{Y}(k) - \mathbf{F}\mathbf{g}\mathbf{F}^{-1}\left(\sum_{b=0}^{B-1} \mathbf{X}^b(k)\hat{\mathbf{H}}^b(k)\right), \quad (6)$$

in which the linearisation of the circular convolution is achieved by applying the constraining window, $\mathbf{g} = \mathrm{diag}\{[\mathbf{0}_{1\times N}, \mathbf{1}_{1\times V}]\}$. Consequently $\mathbf{E}(k) = \mathbf{F}[\mathbf{0}_{1\times N}, \mathbf{e}^{\mathrm{T}}(k)]^{\mathrm{T}}$ contains only linear components. It must be mentioned that the result of the circular convolution in (6) contains $M - N + 1$ linear components. However, for consistency, all signal blocks are defined to be of length V, i.e.,

$$\mathbf{q}(k) = [q(kR - V + 1), \ldots, q(kR)]^{\mathrm{T}},$$
$$\mathbf{Q}(k) = \mathbf{F}[\mathbf{0}_{1\times N}, \mathbf{q}^{\mathrm{T}}(k)]^{\mathrm{T}},$$

with $\mathbf{q} \in \{\mathbf{y}, \mathbf{d}, \mathbf{r}, \mathbf{e}\}$ and $R \leq V$, which provides the relation between the frame shift and the partition length.

Now it is possible to define $\mathbf{G} = \mathbf{F}\mathbf{g}\mathbf{F}^{-1}$ which is the frequency domain constraining matrix, whose elements are

$$G(m, m') = \frac{1}{M}\sum_{l=N}^{M-1} W_M^{\Delta m \cdot l} = \frac{1}{M}\frac{W_M^{\Delta m \cdot N} - W_M^{\Delta m \cdot M}}{1 - W_M^{\Delta m}}, \quad (7)$$

where $\Delta m = m - m'$, and m and m' are the discrete frequency indexes. Eq. (7) shows that constraining is similar to decimating in the frequency domain, [18]. The structure of $\mathbf{G}$ depends on both M and N. If M is sufficiently larger than N, the main diagonal is dominant, while the values on the off-diagonals decrease quickly [8]. As a consequence, it is valid to simplify $\mathbf{G} = \frac{V}{M}\mathbf{I}$ if the DFT is long enough [8], where $\mathbf{I}$ is the $M \times M$ identity matrix. However, for shorter DFT lengths or larger frame overlaps, the error introduced by neglecting the inter-band correlations is no longer negligible; a thorough analysis for $M = 2N$ is given in [8]. The error can be reduced by taking a limited number of off-diagonals into account, which depend on the relation V/M and on the SNR. An analysis for arbitrary analysis and synthesis windows is presented in [11]. Finally, it is important to note that $\mathbf{G}\mathbf{Y}(k) \equiv \mathbf{Y}(k)$, as the linearisation of an already constrained signal block is redundant.

3. MODEL-BASED STATE-SPACE ALGORITHM

In this section, the partitioned-block model-based AEC is described using a modified but equivalent formulation as in [17]. The main difference is that the constraining matrix is applied to the sum of the convolutions, as in (6), and not to the individual convolutions using $\mathbf{C}^b(k) = \mathbf{G}\mathbf{X}^b(k)$ as in [17]. In the following section, the Kalman gain matrix simplification proposed in [17] is briefly described, followed by the proposed simplification. An efficient implementation of the proposed Kalman gain matrix computation, including a complexity analysis and a performance evaluation, is also provided.

The model-based AEC algorithm, [13], is developed based on the assumption that the acoustic echo path varies only gradually, which can be described using a first-order Markov model, as depicted in Fig. 1. The acoustic echo path at frame k is then defined by

$$\mathbf{H}^b(k) = \mathbf{A}\mathbf{H}^b(k-1) + \Delta\mathbf{H}^b(k-1), \quad (8)$$

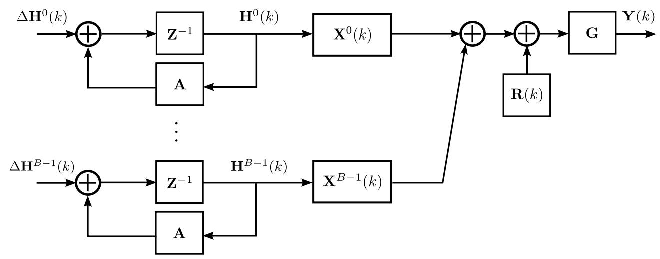


Fig. 1: Partitioned-block-based state-space architecture

where $\Delta\mathbf{H}^b(k-1)$ is the zero-mean process noise, which is assumed to be uncorrelated, and $\mathbf{A}$ is the time-invariant transition matrix. Assuming that the different echo path partitions, and their updates, are mutually uncorrelated and have zero-mean, one can now develop the state-space algorithm.

Given (8), it is reasonable to describe the adaptation of the filter coefficients in two steps, i.e., transition and update,

$$\hat{\mathbf{H}}^b(k|k-1) = \mathbf{A}\hat{\mathbf{H}}^b(k-1) \quad (9a)$$
$$\hat{\mathbf{H}}^b(k) = \hat{\mathbf{H}}^b(k|k-1) + \tilde{\mathbf{G}}\mathbf{K}^b(k)\mathbf{E}(k|k-1), \quad (9b)$$

where $\mathbf{K}^b(k)$ is the Kalman gain matrix (KGM), [13], and $\tilde{\mathbf{G}}$ is the fast correlation constraining matrix. It is assumed that the filter update vector is perfectly linearised as described in [4], and hence $\tilde{\mathbf{G}}$ is omitted in the remainder of the paper for brevity. It is necessary to first define the system distance vector,

$$\mathbf{W}^b(k|k-1) = \mathbf{H}^b(k) - \hat{\mathbf{H}}^b(k|k-1) \quad (10a)$$
$$\mathbf{W}^b(k) = \mathbf{H}^b(k) - \hat{\mathbf{H}}^b(k) = \mathbf{W}^b(k|k-1) - \mathbf{K}^b(k)\mathbf{E}(k|k-1), \quad (10b)$$

whose covariance matrix, $\mathbf{\Psi}_{WW}^b(k) = E\{\mathbf{W}^b(k)\mathbf{W}^{b\mathrm{H}}(k)\}$, is used to obtain the KGM; where $E\{\cdot\}$ denotes mathematical expectation and $\cdot^{\mathrm{H}}$ denotes hermitian transpose. Using (10a) and (6) it is possible to define,

$$\mathbf{E}(k|k-1) = \mathbf{G}\sum_{b=0}^{B-1} \mathbf{X}^b(k)\mathbf{W}^b(k|k-1) + \mathbf{R}(k). \quad (11)$$

Under the assumption that $\mathbf{W}^b(k)$ and $\mathbf{Y}(k)$ are orthogonal, the KGM is obtained by equating $E\{\mathbf{W}^b(k)\mathbf{Y}^{\mathrm{H}}(k)\}$ to zero. Which results in

$$\mathbf{K}^b(k) = \mathbf{\Psi}_{WW}^b(k|k-1)\mathbf{X}^{b\mathrm{H}}(k)\mathbf{G}^{\mathrm{H}}\mathbf{\Psi}_{EE}^{-1}(k|k-1), \quad (12)$$

where the covariance matrix of the error signal, $\mathbf{\Psi}_{EE}(k|k-1)$, is equivalent to

$$\mathbf{G}\left(\sum_{b=0}^{B-1} \mathbf{X}^b(k)\mathbf{\Psi}_{WW}^b(k|k-1)\mathbf{X}^{b\mathrm{H}}(k) + \mathbf{\Psi}_{RR}(k)\right)\mathbf{G}^{\mathrm{H}}, \quad (13)$$

where $\mathbf{\Psi}_{RR}(k)$ is the covariance matrix of the interference signal. Finally, the covariance matrix of the system distance is also updated

$$\mathbf{\Psi}_{WW}^b(k) = (\mathbf{I} - \mathbf{K}^b(k)\mathbf{G}\mathbf{X}^b(k))\mathbf{\Psi}_{WW}^b(k|k-1) \quad (14a)$$
$$\mathbf{\Psi}_{WW}^b(k+1|k) = \mathbf{A}\mathbf{\Psi}_{WW}^b(k)\mathbf{A}^{\mathrm{H}} + \mathbf{\Psi}_{\Delta\Delta}^b(k), \quad (14b)$$

where the covariance matrix of the process noise is

$$\mathbf{\Psi}_{\Delta\Delta}^b(k) = \mathbf{\Psi}_{\hat{H}\hat{H}}^b(k-1) - \mathbf{A}\mathbf{\Psi}_{\hat{H}\hat{H}}^b(k-1)\mathbf{A}^{\mathrm{H}}, \quad (15)$$

where $\mathbf{\Psi}_{\hat{H}\hat{H}}^b(k)$ is the covariance matrix of the estimated acoustic echo path. Eq. (15) holds under the assumption of stationarity of the acoustic echo path and the noise vectors in (8). As stated in [13], a covariance matrix in the DFT domain is related to a diagonal matrix containing the power spectral density (PSD) of the transformed signal blocks. Hence, it is reasonable to approximate the covariance matrices by diagonal ones.

4. SIMPLIFICATION OF THE KALMAN GAIN MATRIX

The implementation of the model-based AEC algorithm as described in the previous section is not feasible due to the high computational cost of the $M \times M$ matrix multiplications. Moreover, the error covariance matrix given by (13) is usually ill-conditioned or even singular. In this section, first the simplification of the KGM proposed in [17] is described, followed by the description of the proposed method, which takes into account the inter-band correlations.

4.1. Diagonalization of the Kalman gain matrix

In [13,15,17] it is proposed to diagonalize the KGM, allowing to formulate the proposed algorithm analogously to a classical least mean squares (LMS) type algorithm in the frequency domain, i.e., with a frequency dependent step-size vector. Using the approximation $\mathbf{G}\mathbf{X}^b(k) \approx \frac{V}{M}\mathbf{X}^b(k)$ as proposed in [8], (13) can be written as

$$\frac{V}{M}\sum_{b=0}^{B-1}\mathbf{X}^b(k)\mathbf{\Psi}_{WW}^b(k|k-1)\mathbf{X}^{b\mathrm{H}}(k)+\mathbf{\Psi}_{RR}(k). \tag{16}$$

In addition, in [17] it is proposed to directly estimate the covariance matrix of the error signal instead of using (13), avoiding the estimation of the interference signal. The KGM is then obtained by

$$\mathbf{K}^b(k) = \frac{V}{M}\mathbf{\Psi}_{WW}^b(k|k-1)\mathbf{\Psi}_{EE}^{-1}(k|k-1)\circ\mathbf{X}^{b\mathrm{H}}(k), \tag{17}$$

where $\circ$ denotes an element-wise or Hadamard matrix multiplication. Eq. (17) is valid under the assumption that $\mathbf{\Psi}_{EE}(k|k-1)$ is diagonal, and consequently its inverse is also diagonal. Hereafter, by initializing $\mathbf{\Psi}_{WW}^b(0) = \mathbf{I}$, both the KGM and the system distance covariance matrix are diagonal matrices.

4.2. Improved implementation using inter-band correlations

The diagonalization of the KGM delivers the lowest-possible algorithmic complexity, yet the performance can be reduced if compared to taking the inter-band correlations into account. The proposed method is also based on the assumption that the error covariance matrix can be approximated by a diagonal matrix. In contrast to the derivation in Sec. 4.1, the diagonalization of the KGM is not forced, whose elements can be expressed as

$$\begin{aligned}K^b(m,m',k) =&\Psi_{WW}^b(m,m,k|k-1)X^{b^*}(m,m,k)\\ &\times G^{\mathrm{H}}(m,m')\Psi_{EE}^{-1}(m',m',k|k-1),\end{aligned} \tag{18}$$

with $K^b(m,m',k) \neq 0$ for $((m-m'))_M \leq T$, where $((\cdot))_M$ denotes modulo M operation and $2T$ is the number of off-diagonals of $\mathbf{G}$ used for the simplification. This is analogous to taking $2T$ cross-bands around the frequency band m, [11]. As it can be observed in (18), only the elements on the main diagonal of $\mathbf{\Psi}_{WW}^b(k)$ are used. Hence, the calculation of (14a) can be simplified to

$$\mathbf{\Psi}_{WW}^b(k) = (\mathbf{I}-\mathbf{K}^b(k)\mathbf{G}\mathbf{X}^b(k))\circ\mathbf{\Psi}_{WW}^b(k|k-1), \tag{19}$$

which ensures that $\mathbf{\Psi}_{WW}^b(k)$ is diagonal if $\mathbf{\Psi}_{WW}^b(0) = \mathbf{I}$. Now, given (18) and (19), it can be observed that it is possible to take into account a limited number of off-diagonals of the KGM, without having to perform the complete matrix multiplications.

The implementation can be optimized by implicitly applying the Kalman gains in (9b),

$$\begin{aligned}\Delta\hat{H}^b(m,k) = \Psi_{WW}^b(m,m,k|k-1)X^{b^*}(m,m,k)\times\\ \sum_{l=m-T}^{m+T}G^{\mathrm{H}}(m,l_M)\Psi_{EE}^{-1}(l_M,l_M,k|k-1)E(l_M,k|k-1),\end{aligned} \tag{20}$$

where $l_M = ((l))_M$ and the second part of the equation can be calculated only once per frame as it is block independent. In addition, (19) can be simplified by

$$\begin{aligned}\Psi_{WW}^b(m,m,k) = (1-\Psi_{WW}^b(m,m,k|k-1)U(m,m,k)\times\\ |X^b(m,m,k)|^2)\Psi_{WW}^b(m,m,k|k-1),\end{aligned} \tag{21}$$

as $\mathbf{X}^b(k)$ and $\mathbf{U}(k)$ are diagonal matrices; being

$$U(m,m,k) = \sum_{l=m-T}^{m+T}|G(m,l_M)|^2\Psi_{EE}^{-1}(l_M,l_M,k|k-1), \tag{22}$$

which is block independent and has to be calculated only once per frame. In addition, due to the symmetry of $\mathbf{G}$ it holds that $|G(m,((m+c))_M)|^2 \equiv |G(m,((m-c))_M)|^2$, $\forall c < M$, which enables the further simplification of (22). The implementation using (20) and (21) does not explicitly calculate the Kalman gains, but it is completely equivalent to making use of (18) and (19).

5. COMPLEXITY ANALYSIS

A complexity analysis is provided to evaluate the increase in algorithmic complexity caused by the utilization of inter-band correlations as described in the previous section. It must be noted that the proposed implementation with $T = 0$ is approximately as complex as the proposed implementation in [17]. The common complexity per frame to all methods, taking into account the constraint of the filter update in (9b), is approximately

$$C_{\mathrm{FDAF}} \approx (3+2B)\mathrm{O(FFT)}+B\mathrm{O(CpxM)}+B(4M), \tag{23}$$

where FFT denotes fast Fourier transform. The computational complexity of a FFT can be approximated by $\mathrm{O(FFT)} \approx 2M\log_2(M) - 4M$ and the one of a complex multiplication by $\mathrm{O(CpxM)} \approx 6M$, see [10] and the references therein. Furthermore, the complexity of (20) using $2T$ off-diagonals can be approximated by

$$C_{\Delta\hat{\mathbf{H}}} \approx B(2M+\mathrm{O(CpxM)})+(3+7T)M, \tag{24}$$

if the symmetry of $\mathbf{G}$ is exploited in the implementation. Finally, the complexity of the update of the system distance covariance matrix, (21) and (22), is approximately

$$C_{\mathbf{\Psi}_{WW}} \approx B(6M)+(1+2T)M. \tag{25}$$

The total complexity overhead, introduced by (24) and (25), is of $9TM$ operations per frame per additional pair of off-diagonals. Hence, it can be expected that the increase in complexity will be small, as the overhead does not depend on the number of partitions. In Fig. 2 the complexity ratios between the diagonalized and the proposed method with $2T$ off-diagonals, denoted as $C_{T|0}$, are depicted for different number of partitions and frame lengths.

It can be concluded that the complexity increases only moderately if B is small, while the complexity ratio tends to one for an increasing number of partitions. For instance, even using 4 additional off-diagonals, the complexity ratio is below 1.25 for $M = 128$ and only 1 partition, which is the worst-case in the analysis.

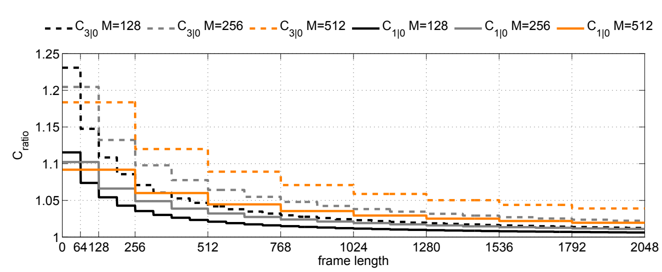


Fig. 2: Complexity ratio

6. PERFORMANCE EVALUATION

In this section, some simulation results are presented to evaluate the performance of the proposed algorithm. The simulations were run using both speech and unit-variance white Gaussian noise with zero mean as excitation signals, which were convolved with the first $L = 1024$ coefficients of a room impulse response (RIR) of total length 4096 taps at a sampling frequency of 16 kHz. The RIR was generated using the image method, [19], for a room size of $5 \times 4 \times 3$ (width $\times$ length $\times$ height) cubic meters and a reverberation time, T_{60}, of 350 ms. The distance between loudspeaker and microphone was set to $d = 1$ m. Two different FFT lengths were used, $M = 256$ and $M = 512$, with $M = 2N$ and a frame overlap of 50%; i.e., $V = R$. The performance of the algorithm was tested under different noise conditions by adding white Gaussian noise to the microphone signal. Every simulation was run 50 times and the outcomes were averaged.

Concerning the model-based AEC parameters, the transition matrix $\mathbf{A} = 0.999 \cdot \mathbf{I}$ was used, which is known to deliver the best performance, [15, 17]. The error covariance matrix was estimated using a first-order recursive filter, i.e.,

$$\hat{\mathbf{\Psi}}_{EE}(k) = \beta \hat{\mathbf{\Psi}}_{EE}(k-1) + (1-\beta)\mathrm{diag}\{\mathbf{E}(k) \circ \mathbf{E}^{\mathrm{H}}(k)\}, \quad (26)$$

with a forgetting factor $\beta = 0.91$. The estimation of $\hat{\mathbf{\Psi}}^{b}_{\hat{H}\hat{H}}(k)$ was carried out analogously. The normalized misalignment (NMA) of the estimated acoustic echo path with respect to the generated RIR,

$$\mathrm{NMA}(k) = 10 \log_{10} \left(\frac{||\mathbf{h}(k) - \hat{\mathbf{h}}(k)||_2^2}{||\mathbf{h}(k)||_2^2} \right), \quad (27)$$

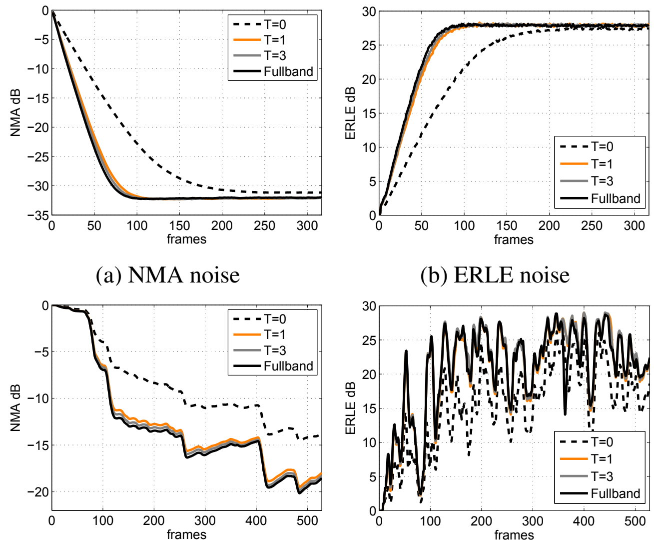
where $|| \cdot ||_2$ denotes the l^2-norm, and the echo return loss enhancement (ERLE),

$$\mathrm{ERLE}(k) = 10 \log_{10} \left(\frac{||\mathbf{d}(k)||_2^2}{||\mathbf{y}(k) - \hat{\mathbf{d}}(k)||_2^2} \right), \quad (28)$$

were used as performance measures.

Two sets of simulations were run, on one hand, for $M = 256$ and a segmental echo-to-noise ratio (ENR) of 30 dB, the performance of the proposed method using a different number of off-diagonals was compared. The diagonalization of the KGM, which is equivalent to the proposed method with $T = 0$, was compared against $T = 1$, $T = 3$ and the non-simplified, or full-band, KGM computation. The condition $T = 2$ was not included in the simulations as for $M = 2N$, $G(m, m \pm \dot{2}) = 0$, where $\dot{2}$ denotes multiple of 2. Figs. 3a and 3b show the misalignment and the ERLE, respectively, if the excitation signal is white Gaussian noise. It can be observed that increasing the number of off-diagonals not only increases the convergence speed but, in addition, the algorithm converges to a lower steady-state misalignment and a higher final ERLE. It can also be seen that adding one pair of off-diagonals already performs comparably to the full-band method. Figs. 3c and 3d depict the outcome of the simulation for speech. Similar conclusions as before can be drawn from the misalignment results. Fig. 3d shows that the diagonalization of the KGM performs worse than the other methods under test. However, it is not possible to conclude that using all the inter-band correlations outperforms the proposed method with $T = 1$ and $T = 3$.

On the other hand, the proposed algorithm was evaluated under different noise conditions for two transform lengths using white Gaussian noise as excitation signal. In Fig. 4 both the steady state ERLE and the convergence speed of the algorithm, until the steady-state is reached, are shown. It can be observed that for a longer DFT length all conditions reach a higher final ERLE. In addition, for high ENR levels the conditions with $T > 0$ outperform the algorithm with $T = 0$, and they all converge to a similar ERLE level. Yet, for lower ENR levels, $T = 0$ performs comparably or even slightly better than the proposed algorithm with $T > 0$. This was expected based on the analysis in [11]. The analysis of the convergence speed shows that using the inter-band correlations accelerates the adaptation process; for instance, the convergence speed is even doubled for ENR $\geq$ 20 dB.



(a) NMA noise (b) ERLE noise

(c) NMA speech (d) ERLE speech

Fig. 3: Performance, $M = 256$ and segmental ENR $= 30$ dB

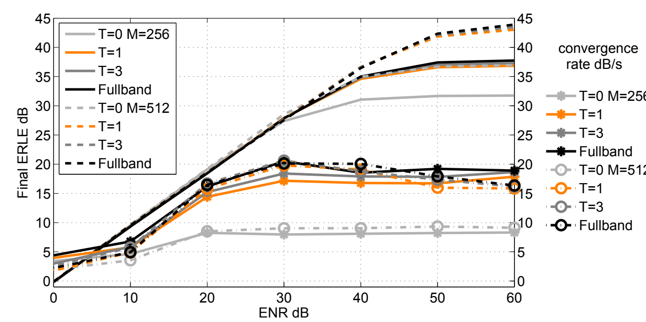


Fig. 4: ERLE as a function of the input ENR

7. CONCLUSIONS

A simplification of the Kalman gain matrix for the sate-space PB-FDAF for AEC was presented. The proposed method takes into account the inter-band correlations to improve the performance of the final algorithm. To this end, first the expressions of the state-space PB-FDAF algorithm in [17] were reformulated to highlight the convolution constraining matrix. Secondly, the proposed simplification and its optimized implementation were described, whose complexity analysis demonstrated that the complexity overhead can be kept moderately low. Finally, the performance evaluation showed that the use of inter-band correlations in the Kalman gain matrix computation results in an improvement of the AEC performance for short DFT lengths if the ENR level is sufficient. Moreover, during the early adaptation stage, the convergence rate is visibly increased.

8. REFERENCES

[1] E. Hänsler and G. Schmidt, *Acoustic Echo and Noise Control: A practical Approach*, Wiley, New Jersey, USA, 2004.

[2] G. Schmidt, "Applications of acoustic echo control – an overview," in *Proc. European Signal Processing Conf. (EUSIPCO)*, Vienna, Austria, 2004, pp. 9–16.

[3] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, fourth edition, 2002.

[4] J. J. Shynk, "Frequency-domain and multirate adaptive filtering," *IEEE Signal Process. Mag.*, vol. 9, no. 1, pp. 14–37, Jan. 1992.

[5] A. Oppenheim and R. W. Schafer, *Digital Signal Processing*, Prentice-Hall Inc., Englewood Cliff, NJ, second edition, 1993.

[6] D. Mansour and A. J. Gray, Jr, "Unconstrained frequency-domain adaptive filter," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 30, no. 5, pp. 726–734, Oct. 1982.

[7] J. S. Soo and K. K. Pang, "Multidelay block frequency domain adaptive filter," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, pp. 373–376, Feb. 1990.

[8] J. Benesty and D. R.Morgan, "Frequency-domain adaptive filtering revisited, generalization to the multi-channel case, and application to acoustic echo cancellation," in *Proc. IEEE ICASSP*, 2000, vol. 2, pp. 289–292.

[9] M. Joho and G. S. Moschytz, "Connecting partitioned frequency-domain filters in parallel or in cascade," *IEEE Trans. Circuits Syst. II*, vol. 47, no. 8, pp. 685–697, Aug. 2000.

[10] R. M. M. Derkx, G. P. M. Engelmeers, and P. C. W. Sommen, "New constraining method for partitioned block frequency-domain adaptive filters," *IEEE Trans. Signal Process.*, vol. 50, no. 3, pp. 2177–2186, 2002.

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

[12] G. Enzner and P. Vary, "Frequency-domain adaptive Kalman filter for acoustic echo control in hands-free telephones," *Signal Processing*, vol. 86, no. 6, pp. 1140–1156, 2006.

[13] G. Enzner, *A Model-Based Optimum Filtering Approach to Acoustic Echo Control: Theory and Practice*, Ph.D. thesis, RWTH Aachen University, Wissenschaftsverlag Mainz, Aachen, Germany, Apr. 2006, ISBN 3–86130-648–4.

[14] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Trans. of the ASME Journal of Basic Engineering*, vol. 82, no. Series D, pp. 35–45, 1960.

[15] S. Malik and G. Enzner, "Model-based vs. traditional frequency-domain adaptive filtering in the presence of continuous double-talk and acoustic echo path variability," in *Proc. Intl. Workshop Acoust. Echo Noise Control (IWAENC)*, Seattle, WA, Sept. 2008.

[16] P. C. W. Sommen, "Partitioned frequency-domain adaptive filters," in *Proc. Asilomar Conf. on Signals, Systems and Computers*, 1989, pp. 677–681.

[17] F. Küch, E. Mabande, and G. Enzner, "State-space architecture of the partitioned-block based acoustic echo controller," in *Proc. IEEE ICASSP*, Florence, IT, May 2014.

[18] L. Pelkowitz, "Frequency domain analysis of wraparound error in fast convolution algorithms," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 29, no. 3, pp. 413–422, 1981.

[19] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.*, vol. 65, no. 4, pp. 943–950, Apr. 1979.