Statistical Analysis of a GSC-based Jointly Optimized Beamformer-Assisted Acoustic Echo C canceler

Marcos H. Maruo, Student Member, IEEE, José C. M. Bermudez, Senior Member, IEEE and Leonardo S. Resende, Member, IEEE

Abstract—This work presents a statistical analysis of a class of jointly optimized beamformer-assisted acoustic echo cancelers (AEC) with the beamformer (BF) implemented in the Generalized Sidelobe C canceller (GSC) form and using the least-mean square (LMS) algorithm. The analysis considers the possibility of independent convergence control for the BF and the AEC. The resulting models permit the study of system performance under typical handling of double-talk and channel changes. We show that the joint optimization of the BF-AEC is equivalent to a linearly-constrained minimum variance problem. Hence, the derived analytical model can be used to predict the transient performance of general adaptive wideband beamformers. We study the transient and steady-state behaviors of the residual mean echo power for stationary Gaussian inputs. A convergence analysis leads to stability bounds for the step-size matrix and mean echo power for stationary Gaussian inputs. A convergence analysis leads to stability bounds for the step-size matrix and mean echo power for stationary Gaussian inputs. A convergence analysis leads to stability bounds for the step-size matrix and mean echo power for stationary Gaussian inputs. A convergence analysis leads to stability bounds for the step-size matrix and mean echo power for stationary Gaussian inputs. A convergence analysis leads to stability bounds for the step-size matrix and mean echo power for stationary Gaussian inputs. A convergence analysis leads to stability bounds for the step-size matrix and mean echo power for stationary Gaussian inputs.

Index Terms—Acoustic echo cancellation, adaptive filtering, beamforming, generalized sidelobe canceller, statistical analysis

I. INTRODUCTION

Acoustic echoes arise in hands free communications when a microphone picks up both the signal radiated in a direct path by a loudspeaker and its reflections at the borders of a reverberant environment. Acoustic echoes tend to degrade intelligibility and listening comfort [1], [2]. Modern solutions incorporate adaptive echo cancellers. However, typical room reverberation times require adaptive acoustic echo cancellers with very long responses [1], [2]. Also, signal contamination by speech from other talkers, noise and their reflections in the acoustic environment make it difficult to obtain fast convergence and satisfactory echo cancellation with such long cancellers [1]–[3]. Moreover, conventional acoustic echo cancellation also requires a complex control logic to avoid divergence during double-talk periods [3], [4]. Few recent work [5] proposes the use of blind source separation techniques. Though promising, such technique still lacks computationally efficient solutions.

Assuming it is possible to estimate the direction of arrival (DOA) of the desired speaker, spatial filtering (beamforming) can help attenuate interfering signals in other directions than the desired one. Beamformers (BFs) have limited echo suppression capacity due to limits in the array directivity [5] and the large number of microphones necessary to suppress all reflections outside the desired DOA [6].

Acoustic echo cancellation solutions in which BFs and acoustic echo cancellers (AECs) have complementary functions have raised a lot of interest recently [7]–[12]. BFs and AECs contribute by different means to reduce the residual echo. Hence, using both techniques in a synergistic way can improve the acoustic echo cancellation performance [13]–[22]. BFs and AECs are usually combined by means of two basic structures [11], [13]. The AEC first structure (AEC-BF) employs one AEC per microphone [19]–[21]. The BF then processes the AEC outputs for spatial filtering. It requires several long AECs, leading to very high computational costs [19]. Moreover, signals outside the desired DOA must be treated as double talk, complicating the design. The BF first (BF-AEC) structure does the spatial filtering first, leaving basically the echo in the desired DOA to be canceled by a single AEC [16]–[18]. This structure presents a significantly lower computational complexity when compared to the AEC-BF structure, even considering that the BF impulse response adds to the length of the response to be identified by the AEC [11]. However, as a single AEC has to cancel echoes arriving at many microphones and its desired signal is affected by the BF state, the plant identification model is not valid. Therefore, previous theoretical work has to be used carefully when this structure is studied. In addition, since the AEC solution depends on the BF state, an abrupt change in the desired DOA can lead to a degraded performance until the AEC tracks the new solution.

Alternative structures that have been proposed include the use of polynomial approximations in delay-and-sum beamformers [24], [25], the Transfer-Function Generalized Sidelobe C canceller (TF-GSC) [26]–[29], AEC sub-modeling [30], mutually exclusive adaptation of the BF and AEC [31] and wave-domain filtering [32], [33].

Optimization of BF-assisted acoustic echo cancellation systems can be based on different performance surfaces, depending on how the BF and the AEC are optimized. One may define the beamformer performance surface from its own local error [12] or use a joint optimization scheme [4], [17] in
which the global cancellation error is used to jointly optimize
the BF and the AEC. The joint optimization scheme was
first proposed in [14]. It was later applied to a robot speech
recognition system [15]. Joint BF-AEC optimization leads to
an optimal solution with better echo cancellation performance
than separate BF and AEC optimizations [22].

Despite the possibilities of combined BF and AEC acoustic
echo cancellation systems, we find only few analyses of their
transient behavior in the literature. The AEC-BF structure
has been studied for the acoustic echo cancellation problem
in [19]–[21] and for the acoustic feedback cancellation in
[34]. A stochastic model has been derived using the power
transfer function method for the case of a fixed BF, where
just the AEC is adapted. More recently, the transient behavior
of a system where a direct-form BF and an AEC are jointly
adapted using equal and fixed step-sizes was analyzed in [17],
[18]. The derived analytical model was shown to accurately
predict the adaptive system behavior and corroborated previous
experimental findings that the same cancellation performance
of a single-microphone AEC can be achieved with a shorter
AEC when the possibility of spatial filtering is available [35].
The model, based on the equivalence to a conventional Linearly
Constrained Minimum Variance (LCMV) optimization, allows
the use of previous analytical results [36], [37].

Adaptive LCMV beamforming may be implemented in
many different forms and by using different algorithms [36],
[38]–[43]. The direct and GSC forms are equivalent in that
both lead to the same optimal solution [44]. For some algo-
rithms and under specific conditions they are equivalent even
in their transient behavior [38], [41]–[43], [45]. Both forms
tend to have comparable computational complexities for a
small number of constraints. However, the GSC form offers
greater design flexibility due to the possibility of choosing the
block matrix. Good choices may lead reduced computational
complexity [38, p. 31]. Also, robust GSC implementations
with an adaptive block matrix have been proposed to account
for small changes in the desired signal DOA [15], [46]–[49].
Therefore, it is of interest to study the behavior of the GSC
form of the BF-AEC structure.

This work extends the analysis in [17], [18] to the study
of the transient behavior of the jointly optimized BF-AEC
structure in the GSC form. We formulate the joint optimization
as a single constrained optimization problem, what simplifies
the statistical analysis. Moreover, the analysis incorporates
the case of a positive-definite step-size matrix [50]–[54]. The
incorporation of this extra flexibility to the model is particu-
larly interesting for BF-assisted echo cancelers, as their AEC
adaptation control logic stops AEC adaptation during double-
talk periods [6], [7], while the BF continues adapting using
the AEC output as the reference signal. The problem of designing
an adaptive filter with step-size matrices was studied in [50]–[54].
The derived analytical model was shown to accurately
predict the adaptive system behavior during different control
logic states (different step-sizes) is of undiscussible practical
relevance. The analysis model uses a positive-definite step-
size matrix.

Using the proposed formulation, we derive a statistical model
of the behavior of the BF-AEC system implemented in the
GSC form with a positive-definite matrix step-size. The model
also allows the derivation of a high convergence rate algorithm
based on a quasi-Newton adaptation scheme in which the step-
size matrices are designed to whiten the combined input vector
to accelerate convergence.

This paper is organized as follows. Section II formulates the
problem addressed. Up to Section II-C the material is basically
the same as in [18] and is necessary to establish the notation
used in the rest of the paper. Section II-D introduces the GSC
formulation for the problem studied. Section III describes the
analysis structure that allows the analysis of the adaptation
using different step-sizes and the quasi-Newton algorithm
using the same mathematical framework. Section IV describes
the adaptive solution. Section V derives the statistical model
for the adaptive solution. The statistical model convergence
is analyzed in Section VI. Based on the results in section VI,
the new quasi-Newton adaptation is derived in Section VII.
Section VIII validates the proposed model using simulation
examples. Finally, conclusions are presented in Section IX.
In this paper, plain lowercase or uppercase letters denote
scalars, lowercase boldface letters denote column vectors and
uppercase boldface letters denote matrices.

II. Problem Formulation

Fig. 1 shows the BF-AEC structure with $M$ echo impulse
response vectors $h_{m}$ of length $N_{h}$, $M$ microphone signals
$x_{m}[n]$, one adaptive wideband beamformer composed of $M$
filters $b_{m}[n]$ of length $N_{BF}$ and an adaptive AEC filter $h_{A}[n]$.
of length $N_{AEC}$. We assume impulse responses $h_m$ constant and stationary signals for mathematical tractability [5], pp. 348–351]. The analysis for a time variant echo path becomes specially challenging in this case even for the simple random walk system nonstationarity model [3]. [5]. This is because a time variant loudspeaker-enclosure-microphone (LEM) model would lead to a nonstationary beamformer input signal. Moreover, the statistically independent increments to the channel response vectors $h_m$ due to the random walk model would be time-correlated by the BF filters. This would render the analysis too complex even for such simple nonstationarity model, making it very hard to study fundamental properties of the algorithm behavior. The study for nonstationary input signals requires a specific model for the input nonstationarity. To the best of our knowledge, there is no generally accepted model for signal nonstationarity. On the other hand, predictions derived under stationarity assumptions can still show tendencies of the algorithm behavior for reasonably small degrees of nonstationarity [4, p. 595]. Simulation results in Section VIII-C will illustrate that this is the case for the present study.

It has been conjectured that the spatial filtering realized by the BF reduces the required AEC length, as compared to the conventional finite impulse response (FIR) AEC structure [35]. Hence, our analysis considers the possibility of an AEC shorter than the LEM impulse responses by admitting $N_{AEC} \leq N_b$.

Grouping the LEM responses as columns of the matrix

$$H = [h_0, h_1, \ldots, h_{M-1}]$$

and defining the echo snapshot vector as

$$e_s[n] = [e_0[n], e_1[n], \ldots, e_{M-1}[n]]^T$$

(1) leads to the linear mapping

$$e_s[n] = H^T u_b[n].$$

(5)

The $m$th microphone signal $x_m[n]$ is the sum of a near-end signal $r_m[n]$ and an echo $e_m[n]$:

$$x_m[n] = e_m[n] + r_m[n], \quad m = 0, \ldots, M - 1.$$  (6)

Each signal $r_m[n]$ is composed of local speech, local interferences and random noise. We define the microphone array snapshot $x_s[n]$ as the vector composed by all $x_m[n]$:

$$x_s[n] = [x_0[n], x_1[n], \ldots, x_{M-1}[n]]^T.$$  (7)

Then, combining (4), (6) and (7) yields

$$x_s[n] = e_s[n] + r_s[n]$$

where $r_s[n] = [r_0[n], r_1[n], \ldots, r_{M-1}[n]]^T$ is the near-end signal component snapshot.

A. The Beamformer Input Vector

Each of the $M$ LEM impulse responses $h_m$, $m = 0, \ldots, M - 1$, models the transmission of the far-end signal $u[n]$ from the speaker to one of the $M$ microphones. The adaptive wideband beamformer is composed by $M$ FIR filters with impulse responses $b_m[n]$, $m = 0, \ldots, M - 1$, each of length $N_{BF}$ [57]. The echo signal at the $m$th microphone is given by [1]

$$e_m[n] = h_m^T u_h[n]$$

(1)

where

$$u_h[n] = [u[n], u[n-1], \ldots, u[n - (N_h + N_{BF} - 2)]]^T$$

(2)

is the LEM plant input vector.

Fig. 1. BF-AEC system configuration in the direct-form structure [22].

We now define the extended far-end sample vector as

$$u[n] = [u[n], u[n-1], \ldots, u[n - (N_h + N_{BF} - 2)]]^T$$

(8)

where the dimension of $u[n]$ is the length of the convolution of $h_m[n]$ and $b_m[n]$. Then to express the microphone array input signals (the echo signals) as functions of $u[n]$ we rewrite (5) as

$$e_s[n-k] = \begin{bmatrix} 0_{M \times k} & H^T & 0_{M \times N_{BF} - (k+1)} \end{bmatrix} u[n]$$

(9)

where $0_{N_1 \times N_2}$ denotes the null matrix with dimension $N_1$ lines and $N_2$ columns. Then, defining the $M.N_{BF} \times 1$ stacked echo vector

$$e[n] = [e_s^T[n], e_s^T[n-1], \ldots, e_s^T[n - (N_{BF} - 1)]]^T$$

(10)

we can write

$$e[n] = H^T u[n]$$

(11)

where

$$H = \begin{bmatrix} H & 0_{1 \times M} & \cdots & 0_{N_h-1 \times M} \\ 0_{N_{BF} - 1 \times M} & H & \cdots & 0_{N_{BF} - 2 \times M} \\ \vdots & \vdots & \ddots & \vdots \\ 0_{N_{BF} - (N_h+N_{BF} - 1) \times M} & 0_{N_{BF} - (N_h+N_{BF} - 2) \times M} & \cdots & H \end{bmatrix}$$

(12)

is the $N_b + N_{BF} - 1 \times M.N_{BF}$ modified echo channel matrix. Note that $e[n]$ contains the echo signals for the time window corresponding to the length of the BF impulse response.

Using (6), (7), (10) and (11), and defining the $M.N_{BF} \times 1$ near-end vector component (without echo) as

$$r_b[n] = [r_s^T[n], r_s^T[n-1], \ldots, r_s^T[n - (N_{BF} - 1)]]^T$$

(13)

and the $M.N_{BF} \times 1$ combined beamformer input regressor
as [30]
\[ x_b[n] = [x_b^T[n], x_b^T[n-1], \cdots, x_b^T[n-(N_{BF} - 1)]]^T \] (14)
we write the beamformer input vector as
\[ x_b[n] = \mathcal{H}^T u[n] + r_b[n]. \] (15)

**B. The Residual Echo**

Define the vector \( b_{si}[n] \) of the \( \ell \)th components of all vectors \( b_m[n], m = 0, \ldots, M-1, \) at time \( n \) as
\[ b_{si}[n] = [b_{0i}[n], \ldots, b_{M-1i}[n]]^T, \ell = 0, \ldots, N_{BF} - 1. \]

We then write the beamformer output \( y[n] \) as
\[ y[n] = \sum_{\ell=0}^{N_{BF}-1} x_b^T[n - \ell] b_{si}[n]. \] (16)

Now, defining the stacked beamformer weight vector
\[ b[n] = [b_0^T[n], b_1^T[n], \ldots, b_{N_{BF}-1}^T[n]]^T \] (17)
we can write \( y[n] \) as the inner product
\[ y[n] = x_b^T[n] b[n]. \] (18)

Next, defining the AEC weight vector
\[ \hat{h}[n] = [\hat{h}_0[n], \hat{h}_1[n], \ldots, \hat{h}_{N_{AEC}-1}[n]]^T \] (19)
and the AEC input vector
\[ u_{\psi}[n] = [u[n], u[n-1], \ldots, u[n-(N_{AEC} - 1)]]^T \] (20)
we can write the AEC output as
\[ \hat{y}[n] = \hat{h}^T[n] u_{\psi}[n]. \] (21)

Using (18) and (21) we write the residual echo \( d[n] \) as the inner product
\[ d[n] = -u_{\psi}^T[n] \hat{h}[n] + x_b^T[n] b[n]. \] (22)

### III. The Analysis Structure

With the problem formulation presented in Section II we can define an analysis problem that corresponds to the study of a single GSC structure that combines the beamformer and the AEC adaptations. To this end, we define the \( N_{AEC} + M \cdot N_{BF} \times 1 \) stacked input vector
\[ s[n] = [-u_{\psi}^T[n], x_b^T[n]]^T \] (23)
and, from (17) and (19), the stacked coefficient vector
\[ w[n] = [\hat{h}^T[n], b^T[n]]^T. \] (24)

Then, we can write the residual echo \( d[n] \) as the inner product
\[ d[n] = s^T[n] w[n]. \] (25)

This simple model will permit to relate the study of the BF-AEC structure to that of the LCMV problem.

Interestingly, input vectors \( u[n] \) in (5) and \( u_{\psi}[n] \) in (20) are related by
\[ u_{\psi}[n] = \begin{bmatrix} I_{N_{AEC}} & 0_{N_{AEC} \times (N_b + N_{BF} - N_{AEC} - 1)} \end{bmatrix} u[n] \] (26)
where we use the notation \( I_K \) to denote the \( K \times K \) identity matrix. Hence, (15) and (26) permit to write \( s[n] \) in (23) as a function of the input vectors \( u[n] \) and \( r_b[n] \). Equation (26) also allows to study the algorithm performance for \( N_{\psi} > N_{AEC} \), and thus verifies the possibility of reducing \( N_{AEC} \) by increasing the number of microphones.

**A. Performance Surface**

The mean output power (MOP) performance surface \( J \) is defined as the mean value of \( d^2[n] \) conditioned on \( w[n] = w \). From (25),
\[ J = E \{ d^2[n] \} w[n] = w = \begin{bmatrix} w^T s[n] s^T[n] w \end{bmatrix} = w^T R_{ss} w. \] (27)

where \( R_{ss} = E \{ s[n] s^T[n] \} \) is the input autocorrelation matrix. A set of \( N_f \) linear constraints on the beamformer coefficients implements the spatial filtering. Usually, an \( MN_{BF} \times N_f \) constraint matrix \( C \) and an \( N_f \times 1 \) response vector \( f \) jointly define the frequency response in the desired DOA [38, 39].

To formulate the linear constraints as a function of the combined coefficient vector, we define the extended constraint matrix [14]
\[ C_e = \begin{bmatrix} 0_{N_f \times N_{AEC}} & C \end{bmatrix}^T. \] (28)

Finally, the joint optimization problem can be formulated as
\[ w_{opt} = \arg \min_w w^T R_{ss} w \] (29a)
such as \( C_e^T w = f \) (29b)
and the optimal solution is given by [36] \( w_{opt} = R_{ss}^{-1} C_e \left( C_e^T R_{ss}^{-1} C_e \right)^{-1} f \).

**B. Implementation using the GSC Form**

In the GSC form [38], the dashed square of Fig. 2 is replaced by the dashed square of Fig. 2. Feasible solutions to (29) are

![Fig. 2. BF-AEC system in the GSC configuration.](image-url)
orthogonal to $C_e$ $(C_e^T B_e = 0)$, $\psi$ is an $N_e$-dimensional vector and $N_e = N_{AEC} + MN_{BF} - N_f$. The minimum norm solution to (29b) is
\begin{equation}
q_e = C_e (C_e^T C_e)^{-1} f. \tag{31}
\end{equation}

1) Optimal Solution: As $C_e^T B_e = 0$, $w$ in (30) satisfies (29b) for any $\psi$, and (29) becomes an unconstrained optimization problem in $\psi$ with solution (41)
\begin{equation}
\psi_{opt} = \arg \min \psi \quad q_e^T R_{ss} q_e - 2\psi^T B_e^T R_{ss} q_e + \psi^T R_{blocl} \psi \tag{32}
\end{equation}
where $R_{blocl} = B_e^T R_{ss} B_e$ denotes the blocked input autocorrelation matrix, and from (30)\]
\begin{equation}
w_{opt} = q_e - B_e \psi_{opt}. \tag{33}
\end{equation}

Defining the cost function of (32)
\begin{equation}
C(\psi) = q_e^T R_{ss} q_e - 2\psi^T B_e^T R_{ss} q_e + \psi^T R_{blocl} \psi \tag{34}
\end{equation}
its gradient with respect to $\psi$ is
\begin{equation}
\nabla_{\psi} C(\psi) = -2B_e^T R_{ss} (q_e - B_e \psi) \tag{35}
\end{equation}
Setting (35) equal to the null vector yields (41)
\begin{equation}
\psi_{opt} = R_{blocl}^{-1} B_e^T R_{ss} q_e. \tag{36}
\end{equation}

IV. THE WEIGHT ADAPTATION EQUATION

To obtain a model flexible enough to allow the study of the system performance with independent BF and AEC adaptations we choose the following block diagonal form for the blocking matrix
\begin{equation}
B_e = \begin{bmatrix} -I_{N_{AEC}} & 0_{N_{AEC} \times (MN_{BF} - N_f)} \\ 0_{MN_{BF} \times N_{AEC}} & B \end{bmatrix} \tag{37}
\end{equation}
and split
\begin{equation}
\psi = \begin{bmatrix} \psi_{h}^T, \psi_{b}^T \end{bmatrix}^T \tag{38}
\end{equation}
where $[\psi_{h}]_i = [\psi]_i, i = 1, \ldots, N_{AEC}$ and $[\psi_{b}]_i = [\psi]_{i+N_{AEC}}, i = 1, \ldots, MN_{AEC} - N_f$. The same block matrix structure has been used in [14, 15] for the implementation of the GSC-based BF-AEC acoustic echo canceler.

Using (30) in (35) and noting from (25) that $R_{ss} w = E\{s[n]d[n]\}$ we have
\begin{equation}
\nabla_{\psi} C(\psi) = -2B_e^T E\{s[n]d[n]\}. \tag{39}
\end{equation}

Splitting the gradient vector according to (38) yields
\begin{equation}
\nabla_{\psi} C(\psi) = \begin{bmatrix} \nabla_{\psi_{h}} C(\psi), \nabla_{\psi_{b}} C(\psi) \end{bmatrix}^T \tag{40}
\end{equation}
where, from (43a) and (44b)
\begin{equation}
\nabla_{\psi_{h}} C(\psi) = 2E\{u_{h}[n]d[n]\} \tag{40a}
\end{equation}
\begin{equation}
\nabla_{\psi_{b}} C(\psi) = -2B_e^T E\{x_{b}[n]d[n]\}. \tag{40b}
\end{equation}

Additionally, setting $d_e = [0_{1 \times N_{AEC}} \quad q^T]^T$ and using (37) and (41) in (30) yields
\begin{equation}
w = \begin{bmatrix} \psi_{h}^T, (q - B \psi_{b})^T \end{bmatrix}^T \tag{42}
\end{equation}
where $q = C(C_e^T C)^{-1} f$.

Comparing (24) and (41) we conclude that $\psi_{h} = \hat{h}$ and $b = q - B \psi_{b}$. Hence, the steepest-descent algorithms for $\hat{h}[n]$ and $\psi_{b}[n]$ with the gradients in (40a) and (40b) respectively are
\begin{equation}
\hat{h}[n+1] = \hat{h}[n] - \mu_{AEC} E\{u_{h}[n]d[n]\} \tag{42a}
\end{equation}
\begin{equation}
\psi_{b}[n+1] = \psi_{b}[n] + \mu_{BF} B_e^T E\{x_{b}[n]d[n]\}. \tag{42b}
\end{equation}

where $\mu_{AEC}$ and $\mu_{BF}$ are the step-size parameters. Note that (42) is different from the steepest descent algorithm for $\psi[n]$ unless $\mu_{AEC} = \mu_{BF}$. This also makes this analysis different from [18] by using the equivalence derived on [38], [41]. However, this extra degree of flexibility is necessary to analyze the behavior of the BF-AEC system under different control logic states that usually act on $(\mu_{AEC}, \mu_{BF})$ to avoid divergence. The stochastic approximations of (42a) and (42b) yield
\begin{equation}
\hat{h}[n+1] \approx \hat{h}[n] - \mu_{AEC} u_{h}[n]d[n] \tag{43a}
\end{equation}
\begin{equation}
\psi_{b}[n+1] \approx \psi_{b}[n] + \mu_{BF} B_e^T x_{b}[n]d[n]. \tag{43b}
\end{equation}

Implementation of (43) has almost the same computational cost of the separate implementation of an LMS implementation of an AEC and a BF demanding only an extra subtraction in the computation of $d[n]$. It also requires only one extra memory allocation to account for the second scalar step-size. Despite its simplicity, (43) can model the BF-AEC system behavior under most control logic states. Implementation of (43) is shown in Fig. 3.

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{fig3.png}
\caption{Implementation of (43)}
\end{figure}

Finally, the recursive weight update equation is obtained defining the diagonal step-size matrix
\begin{equation}
\mathcal{M} = \begin{bmatrix} \mu_{AEC} I_{N_{AEC}} & 0_{N_{AEC} \times (MN_{BF} - N_f)} \\ 0_{(MN_{BF} - N_f) \times N_{AEC}} & \mu_{BF} I_{MN_{BF} - N_f} \end{bmatrix} \tag{44}
\end{equation}
then (43) can be written as
\begin{equation}
\psi[n+1] = \psi[n] + \mathcal{M} B_e^T s[n]d[n]. \tag{45}
\end{equation}

Note that (45) has the exact same behavior of (44), which can be used to study the performance of the practical implementation.

In the following we perform the analysis of an even more general form of (45), in which the only requirements on $\mathcal{M}$ and $B_e$ are that $\mathcal{M}$ is symmetric positive-definite and $B_e$ is a full column-rank matrix that satisfies $C_e^T B_e = 0$. The typical implementation described above will correspond to a particular case of the more general analysis.
A. Weight Error Vector

Define the weight error vector \( v[n] = w[n] - w_{\text{opt}} \). From (33),
\[
v[n] = q_e - B_e \psi[n] - (q_e - B_e \psi_{\text{opt}}) = -B_e \vartheta[n]
\]  
(46)
where
\[
\vartheta[n] = \psi[n] - \psi_{\text{opt}}
\]  
(47)
denotes the weight error vector of the unconstrained filter conditioned on \( B_e \) and \( q_e \). From (46), \( v[n] \) is in the range of \( B_e \). Hence, \( v[n] \) is completely determined by \( \vartheta[n] \) conditioned on \( B_e \). We then study the behavior of \( \vartheta[n] \).

Subtracting \( \psi_{\text{opt}} \) from both sides of (45), using (40) with (33) and (30) we obtain a recursive update equation for \( \vartheta[n] \):
\[
\vartheta[n+1] = (I_{N_e} - \mathcal{M} B_e^T s[n] s^T[n] B_e) \vartheta[n] + \mathcal{M} B_e^T s[n] s^T[n] w_{\text{opt}}.
\]  
(48)

V. STATISTICAL ANALYSIS

A. Simplifying Assumptions

We now study the behavior of BF-assisted GSC-form echo canceler using (45) under the following typical simplifying assumptions required for mathematical tractability [5]:

A1 \( s[n] \) is a zero-mean Gaussian vector;
A2 \( u[n] \) and \( r[n] \) are statistically independent;
A3 \( R_e \) is positive-definite and both \( C_e \) and \( B_e \) have full column rank;
A4 The statistical dependence between \( s[n] s^T[n] \) and \( \psi[n] \) can be neglected;
A5 The DOA does not change during adaptation.

Though not always valid in practice, these assumptions make analysis viable and frequently lead to results that retain sufficient information to serve as reliable design guidelines [5, 12], [14]. Simulation results will confirm their reasonableness for this analysis. A1 simplifies the evaluation of fourth order moments of \( s[n] \). These moments are dependent on the distribution of \( s[n] \), and the Gaussian distribution combines the advantages of being a good model for several physical processes and simplifying the required mathematical derivations. A2 is physically reasonable, as \( u[n] \) and \( r[n] \) are generated at different sides of the communications channel by independent speakers. A3 is reasonable in practice, as \( s[n] \) always has some uncorrelated noise component and both \( C_e \) and \( B_e \) are under reasonable control of the designer. A4 is required to estimate moments involving the input signal and the weight vector, as the statistical distribution of the latter is unknown. This assumption is in fact less restrictive than the usually employed independence assumption, which requires \( s[n] \) and \( \psi[n] \) to be independent, as discussed in detail in [50]. A5 is employed for mathematical tractability and because the main goal of the present analysis is to determine fundamental properties of the adaptive system.

B. Mean Weight Error Vector Behavior

Taking the expected value of (48) under A4 and using (30) and (45) leads to
\[
E[\vartheta[n+1]] = (I_{N_e} - \mathcal{M} B_e^T s[n] s^T[n] B_e) E[\vartheta[n]] + \mathcal{M} B_e^T s[n] s^T[n] w_{\text{opt}}
\]  
(49)

since
\[
E[B_e^T s[n] s^T[n] w_{\text{opt}}] = B_e^T R_e s[n] - R_{\text{bloc}} R_e^{-1} B_e^T R_{\text{bloc}} q_e = 0_{N_e \times 1}.
\]  
(50)

Hence, the mean weights converge asymptotically to the optimal solution if all eigenvalues of \( I_{N_e} - \mathcal{M} R_{\text{bloc}} \) are inside the unit circle. In this case, (45) results in asymptotically unbiased solutions in the mean.

C. Mean Output Power (MOP)

To determine the MOP we use (23) with \( w[n] = v[n] + w_{\text{opt}} \) and (45). Defining \( R_{\vartheta \vartheta}[n] = E[\vartheta[n] \vartheta^T[n]] \) we obtain
\[
J[n] = E\{w_{\text{opt}} - B_e \vartheta[n]\}^T s[n] s^T[n] (w_{\text{opt}} - B_e \vartheta[n])
\]  
(51)
where we have used (50) and A4 to obtain the second line. A recursive expression for the \( N_{\vartheta} \times N_{\vartheta} \) matrix \( R_{\vartheta \vartheta}[n] \) is derived in the next section to complete the model (51).

D. Correlation Matrix of \( \vartheta[n] \)

Post-multiplying (45) by its transpose, taking the expected value, using A1–A5 and (50) yields
\[
R_{\vartheta \vartheta}[n+1] = R_{\vartheta \vartheta}[n] - \mathcal{M} R_{\text{bloc}} R_{\vartheta \vartheta}[n] - R_{\vartheta \vartheta}[n] R_{\text{bloc}} \mathcal{M}
\]  
\[+ \mathcal{M} E\{B_e^T s[n] s^T[n] B_e \vartheta[n] \vartheta^T[n] B_e^T s[n] s^T[n] B_e\} \mathcal{M}
\]  
\[+ J_{\text{min}} \mathcal{M} R_{\text{bloc}} \mathcal{M}.
\]  
(52)

Using A1, A4 and the Gaussian moment factoring theorem, the expectation in (52) is given by
\[
E\{B_e^T s[n] s^T[n] B_e \vartheta[n] \vartheta^T[n] B_e^T s[n] s^T[n] B_e\} = 2 R_{\text{bloc}} R_{\vartheta \vartheta}[n] R_{\text{bloc}} + R_{\text{bloc}} R_{\vartheta \vartheta}[n] R_{\text{bloc}} \mathcal{M}
\]  
(53)
Finally, substituting (53) into (52) yields
\[
R_{\vartheta \vartheta}[n+1] = R_{\vartheta \vartheta}[n] - \mathcal{M} R_{\text{bloc}} R_{\vartheta \vartheta}[n] - R_{\vartheta \vartheta}[n] R_{\text{bloc}} \mathcal{M}
\]  
\[+ J_{\text{min}} \mathcal{M} R_{\text{bloc}} \mathcal{M} + 2 R_{\text{bloc}} R_{\vartheta \vartheta}[n] R_{\text{bloc}} \mathcal{M}.
\]  
(54)

Equation (54) completes the MOP model in (51).

VI. CONVERGENCE ANALYSIS

Classical convergence analysis of (54) would project \( R_{\vartheta \vartheta}[n] \) into the eigenspace of \( R_{\text{bloc}} \) and study the convergence of the diagonal entries of the transformed matrix [4]. The presence of \( \mathcal{M} \), however, requires a different approach. As \( \mathcal{M} R_{\text{bloc}} \neq R_{\text{bloc}} \mathcal{M} \), (54) is not entirely diagonalizable by the same projection [59, p. 558]. Nevertheless, it is still possible to diagonalize both \( \mathcal{M} \) and \( R_{\text{bloc}} \) through contragradient diagonalization [59, p. 465], [60, p. 466]. As \( \mathcal{M} \) is positive definite, Cholesky decomposition yields \( \mathcal{M} = L^T L \) with \( L \) nonsingular. Then, we can transform the vector space into \( \xi[n] = L^{-1} \vartheta[n], R_{\xi \xi}[n] = E\{\xi[n] \xi^T[n]\} = L^{-1} R_{\vartheta \vartheta}[n] L^{-T} \) and
\[
R_{\text{mod}} = E\{L^T B_e^T s[n] (L^T B_e^T s[n])^T\} = L^T R_{\text{bloc}} L.
\]  
(55)
Hence, pre-multiplying (54) by $L^{-1}$, post-multiplying by $L^{-T}$ and using $\text{tr}(R_{\xi\xi}[n]R_{\text{mod}}) = \text{tr}(R_{\theta\theta}[n]R_{\text{bloc}})$ yields

\[ R_{\xi\xi}[n + 1] = R_{\xi\xi}[n] - R_{\text{mod}}R_{\xi\xi}[n] - R_{\xi\xi}[n]R_{\text{mod}} + R_{\text{mod}}[J_{\text{min}} + \text{tr}(R_{\xi\xi}[n]R_{\text{mod}})] + 2R_{\text{mod}}R_{\xi\xi}[n]R_{\text{mod}} \]  

(56)

where $R_{\text{mod}}$ is symmetric and positive definite. Hence, it is diagonalizable as $R_{\text{mod}} = QAQ^T$ with $Q^TQ = I_{N_v}$ and $\Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_{N_v})$.

(57)

Pre-multiplying (56) by $Q^T$ and post-multiplying by $Q$ yields

\[ \mathbf{\Omega}_{\xi\xi}[n + 1] = \mathbf{\Omega}_{\xi\xi}[n] - \Lambda\mathbf{\Omega}_{\xi\xi}[n] - \mathbf{\Omega}_{\xi\xi}[n]\Lambda + \Lambda(J_{\text{min}} + \text{tr}(\mathbf{\Omega}_{\xi\xi}[n]\Lambda)) + 2\Lambda\mathbf{\Omega}_{\xi\xi}[n]\Lambda \]  

(58)

where $\mathbf{\Omega}_{\xi\xi}[n] = Q^TR_{\xi\xi}[n]Q$.

$\mathbf{\Omega}_{\xi\xi}[n]$ is an autocorrelation matrix. Then $[\mathbf{\Omega}_{\xi\xi}[n]]_{i,j} \leq [\mathbf{\Omega}_{\xi\xi}[n]]_{i,i} [\mathbf{\Omega}_{\xi\xi}[n]]_{j,j}$, $[\mathbf{\Omega}_{\xi\xi}[n]]_{i,j} \geq 0$ (61 p. 251). (62), and convergence of (58) can be observed studying only the diagonal elements of $\mathbf{\Omega}_{\xi\xi}[n]$. Let $\nu[n]$ be the vector of diagonal entries of $\mathbf{\Omega}_{\xi\xi}[n]$ and $\lambda = [\lambda_1, \lambda_2, \ldots, \lambda_{N_v}]^T$ be the vector of the eigenvalues of $R_{\text{mod}}$. Then, from (58)

\[ [\nu[n + 1]]_i = \left(1 - \lambda_i^2\right) [\nu[n]]_i + \lambda_i \left(\lambda_i^T\nu[n] + J_{\text{min}}\right) \]  

(59)

and

\[ \nu[n + 1] = \Phi^\nu[n] + J_{\text{min}}\lambda \]  

(60)

where $\lambda^T\nu[n] = \text{tr}(\Lambda\mathbf{\Omega}_{\xi\xi}[n])$, $\Phi = \text{diag}(\rho_1, \rho_2, \ldots, \rho_{N_v}) + \lambda\lambda^T$ and $\rho_k = \left(1 - \lambda_k^2\right) + \lambda_k^2$.

The matrix $\Phi$ is symmetric and positive definite, as for any nonzero vector $\alpha$ we have

\[ \alpha^T \Phi \alpha = \alpha^T \text{diag}(\rho_1, \rho_2, \ldots, \rho_{N_v}) \alpha + \alpha^T \lambda\lambda^T \alpha = \sum_{k=1}^{N_v} \left(1 - \lambda_k^2\right) \left(\alpha_k\right)^2 + \left(\lambda^T\alpha\right)^2 > 0. \]

The solution to (60) is (63)

\[ \nu[n] = \Phi^\nu[0] + J_{\text{min}} \sum_{j=0}^{n-1} \Phi^j\lambda. \]  

(62)

Using (62) we now study the stability conditions and the steady-state behavior of (45).

A. Mean Weight Error Revisited

Pre-multiplying (59) by $L^{-1}$ the mean weight error recursion is transformed to (51)

\[ E\{\xi[n + 1]\} = (I_{N_v} - R_{\text{mod}})E\{\xi[n]\}. \]  

(63)

Hence, analysis can be restricted to the eigenspace of $R_{\text{mod}}$ defined in (55).

B. Stability Conditions

Recursion (62) is a state-space equation whose stability is determined exclusively by the eigenvalues $\lambda_{\Phi_k}$, $k = 1, \ldots, N_v$, of $\Phi$ (63). From Gershgorin’s theorem (64),

\[ \lambda_{\Phi_k} < \rho_k + \lambda_k \sum_{j=1}^{N_v} \lambda_j \]  

(64)

\[ < 1 - 2\lambda_k + 2\lambda_k^2 + \lambda_k \text{tr}(R_{\text{mod}}), \forall k \]  

and (62) is stable if $\lambda_{\Phi_k} < 1$ for all $k$. Then, (64) leads to the sufficient condition

\[ -2\lambda_k + 2\lambda_k^2 + \lambda_k \text{tr}(R_{\text{mod}}) < 0, \forall k \]  

(65)

which implies that $\lambda_k \neq 0$ and

\[ 2 \max\{\lambda_k\} + \text{tr}(R_{\text{mod}}) < 2. \]  

(66)

In most practical cases, a reliable estimate of the eigenvalues of $R_{\text{mod}}$ is not available a priori and the upper bound in (66) cannot be used. However, using the inequality

\[ \max\{\lambda_k\} \leq \text{tr}(R_{\text{mod}}) \]

it is possible to derive a tighter upper bound (65), (66)

\[ \text{tr}(R_{\text{mod}}) = \sum_{k=1}^{N_v} \lambda_k < 2 \frac{2}{3}. \]  

(67)

In the particularly important implementation using (43), $B_c$ is given by (37) and $\mathbf{M}$ by (44). Hence, if we write $\text{tr}(R_{\text{mod}}) = \text{tr}(L^T R_{\text{bloc}} L) = \text{tr}(\mathbf{M} R_{\text{bloc}})$ with the matrices in the partitioned form using (23), (37), (44) and (55) yields

\[ \text{tr}(R_{\text{mod}}) = \mu_{\text{AEFC}} \text{tr}(R_{ukuk}) + \mu_{\text{BF}} \text{tr}(B^T R_{ukuk} B) \]  

(68)

where $R_{ukuk} = E\{u_k[n]u_k^T[n]\}$ and $R_{ukuk} = E\{x_k[n]x_k^T[n]\}$. Hence, (67) becomes

\[ \mu_{\text{AEFC}} \text{tr}(R_{ukuk}) + \mu_{\text{BF}} \text{tr}(B^T R_{ukuk} B) < 2 \frac{2}{3}. \]  

(69)

C. Excess MOP

Using $\text{tr}(R_{\theta\theta} R_{\text{bloc}}) = \text{tr}(R_{\xi\xi}[n]R_{\text{mod}}) = \text{tr}(\mathbf{\Omega}_{\xi\xi}[n]\Lambda)$ we write the MOP as a function of $\lambda$ and $\nu[n]$ in (51) we write the MOP as a function of $\lambda$ and $\nu[n]$

\[ J[n] = J_{\text{min}} + \lambda^T \nu[n]. \]  

(70)

Thus, the excess MOP is given by (5) p. 302

\[ J_{\text{ex}}[n] = \lambda^T \nu[n]. \]  

(71)

D. Steady-State Excess MOP

When (66) holds, $\lim_{n \to \infty} \nu[n + 1] = \nu[n] = \nu[\infty]$ and from (69) and (71) we have

\[ \nu[\infty]]_i = \left(1 - \lambda_i^2\right) + \lambda_i \left(J_{\text{ex}}[\infty] + J_{\text{min}}\right) \]  

(72)

which solved for $[\nu[\infty]]_i$ yields

\[ [\nu[\infty]]_i = (J_{\text{ex}}[\infty] + J_{\text{min}}) \frac{1}{2 - 2\lambda_i}. \]  

(73)

Using this result in (71) as $n \to \infty$ and solving for $J_{\text{ex}}[\infty]$ yields

\[ J_{\text{ex}}[\infty] = J_{\text{min}} \frac{1}{2} \sum_{i=1}^{N_v} \frac{1 - \lambda_i}{1 - \lambda_i}. \]  

(74)
From (55), $R_{\text{bloc}} = B_e^T R_e B_e$ and $A_3$, $R_{\text{mod}}$ is symmetric and positive definite. Hence, its largest eigenvalue is related to its largest singular value through $\max\{\lambda\} = \sqrt{\max\{\sigma\}}$ where $\sigma$ denotes the vector of singular values of $R_{\text{mod}}$. The largest singular value of a matrix is equal to its 2-norm [67]. Then, using the Cauchy-Schwarz inequality [60 pg. 291]

$$\max\{\lambda\} \leq \left( \left\| L^T \right\|_2 \| R_{\text{bloc}} \|_2 \right)^{\frac{1}{2}} \left( \left\| M \right\|_2 \| R_{\text{bloc}} \|_2 \right)^{\frac{1}{2}}\ (75)$$

where both $M$ and $R_{\text{bloc}}$ are symmetric positive definite matrices. Hence, for $\max\{\lambda_M\} \max\{\lambda_{R_{\text{bloc}}}\} \ll 1$, where $\lambda_M$ and $\lambda_{R_{\text{bloc}}}$ are vectors containing the eigenvalues of $M$ and $R_{\text{bloc}}$, respectively, we conclude that $\max\{\lambda\} \ll 1$ and (74) reduces to

$$J_{\text{ex}}[\infty] \approx J_{\text{min}} \frac{\frac{1}{2} \text{tr}(R_{\text{mod}})}{1 - \frac{1}{2} \text{tr}(R_{\text{mod}})}.$$ (76)

For implementations using (43), substituting (68) in (76) yields

$$J_{\text{ex}}[\infty] = J_{\text{min}} \frac{\mu_{\text{AEC}} \text{tr}(R_{x_u\nu}) + \mu_{\text{BF}} \text{tr}(B_e^T R_{x_u\nu} B_e)}{2 - \frac{\mu_{\text{AEC}} \text{tr}(R_{x_u\nu}) - \mu_{\text{BF}} \text{tr}(B_e^T R_{x_u\nu} B_e)}{2}}.$$ (77)

Further assuming $\mu_{\text{AEC}} \text{tr}(R_{x_u\nu}) + \mu_{\text{BF}} \text{tr}(B_e^T R_{x_u\nu} B_e) \ll 2$ we have

$$J_{\text{ex}}[\infty] \approx J_{\text{min}} \frac{1}{2} \left[ \mu_{\text{AEC}} \text{tr}(R_{x_u\nu}) + \mu_{\text{BF}} \text{tr}(B_e^T R_{x_u\nu} B_e) \right].$$ (78)

### VII. A NEW JOINT ADAPTATION ALGORITHM

The analysis results derived in sections VI and VII are valid for the general weight update equation (45). At the same time, (45) in its general form where $M$ and $B_e$ satisfy only the criteria of $M$ being positive-definite and $C_e^T B_e = 0$ can be considered a new adaptive algorithm that allows weight updating in directions that do not correspond to the stochastic gradient.

Next, we discuss one possibility of taking advantage of the more flexible structure, namely, designing for a faster convergence speed.

The simplest way to guarantee asymptotic convergence to $\psi_{\text{opt}}$ is set $M = \mu I_{N_\psi}$ in (45) where $\mu = \mu_{\text{AEC}} = \mu_{\text{BF}}$, which is the standard LMS update. However, it is known that LMS presents a low rate of convergence when the gradient of the performance surface has a low magnitude in the direction of at least one eigenvector of $R_{\text{mod}}$. To alleviate this issue, one may use the step-matrix in (44) or a stochastic approximation of the Newton method. The idea underlying quasi-Newton methods is to use an approximation to the inverse Hessian. The form of the approximation varies among different methods — ranging from the simplest, where it remains fixed throughout the iterative process, to the more advanced where improved approximations are built up on the basis of information gathered during the descent process [63].

#### A. High Convergence Rate Block Matrix and Step-Matrix Pair

Theoretical results show that the rate of convergence of (45) is increased with the reduction eigenvalue spread of $L^T B_e^T R_e B_e L$ [41, 5, 69] reaching its maximum when all eigenvalues are equal. This is because the MOP of an adaptive filter trained with an algorithm of the LMS family decreases over time as a sum of exponentials whose time constants are inversely proportional to the eigenvalues of the autocorrelation matrix of the filter inputs [18]. Hence, small eigenvalues create slow convergence modes while large eigenvalues limit the maximum step-size that can be chosen without encountering stability problems as observed in [60, 70]. When all $R_{\text{mod}}$ eigenvalues are equal, we have

$$Q^T R_{\text{mod}} Q = \lambda I_{N_\psi}$$ (79)

where $\lambda = \text{tr}(R_{\text{mod}})/N_\psi \approx 2 \frac{\text{tr}(\psi)}{N_\psi}$ from (76). Pre-multiplying (79) by $Q$, post-multiplying by $Q^T$ and noting that $Q Q^T = I_{N_\psi}$ we conclude that $Q R_{\text{mod}} Q = I_{N_\psi}$. Observing $R_{\text{mod}}$ structure from (55) yields $L^T \left( \frac{1}{\lambda} R_{\text{bloc}} L \right) = I_{N_\psi}$ and observing that both $L^T$ and $\frac{1}{\lambda} R_{\text{bloc}} L$ are square matrices, we conclude that

$$L^T \left( \frac{1}{\lambda} R_{\text{bloc}} L \right)^{-1} = \lambda L^{-1} \frac{1}{\lambda} R_{\text{bloc}} L.$$ (80)

Finally, pre-multiplying (80) by $L$ and substituting $M = L L^T$ yields

$$M = \frac{2}{N_\psi} J_{\text{ex}}[\infty] \left( B_e^T R_e B_e \right)^{-1}.$$ (81)

Direct use of (81) would require prior knowledge of second-order statistics of $s[n]$ and $J_{\text{min}}$. Nevertheless, the transient behavior when using (81) is a useful measure of the upper bound on the convergence speed. A compromise solution would be the estimation of $M$ every few iterations and the use of update (45). In this case, (81) becomes a quasi-Newton adaptive filter [71].

### VIII. SIMULATION EXAMPLES

This section presents simulation and design examples to verify the accuracy of the derived model and to illustrate its use in design. In all simulations, except when explicitly stated, the far-end signal was drawn from an autoregressive process AR1(a1) given by $u[n] = -a_1 u[n-1] + z[n]$, with $z[n]$ a white Gaussian noise with variance $\sigma_z^2$ such that $\sigma_z^2 = 1$. For the accuracy tests, adaptation of the coefficients is assumed to be done during a single-talk period. Simulations under different control logic states are considered in subsection VIII-F, where only the BF is adapted during a double-talk period and the convergence of both filters is accelerated after a LEM plant change is detected. The LEM plants are designed following the theoretical results derived in sections V and VI are valid for the general weight update equation (45). At the same time, (45) in its general form where $M$ and $B_e$ satisfy only the criteria of $M$ being positive-definite and $C_e^T B_e = 0$ can be considered a new adaptive algorithm that allows weight updating in directions that do not correspond to the stochastic gradient.

Next, we discuss one possibility of taking advantage of the more flexible structure, namely, designing for a faster convergence speed.

The simplest way to guarantee asymptotic convergence to $\psi_{\text{opt}}$ is set $M = \mu I_{N_\psi}$ in (45) where $\mu = \mu_{\text{AEC}} = \mu_{\text{BF}}$, which is the standard LMS update. However, it is known that LMS presents a low rate of convergence when the gradient of the performance surface has a low magnitude in the direction of at least one eigenvector of $R_{\text{mod}}$. To alleviate this issue, one may use the step-matrix in (44) or a stochastic approximation of the Newton method. The idea underlying quasi-Newton methods is to use an approximation to the inverse Hessian. The form of the approximation varies among different methods — ranging from the simplest, where it remains fixed throughout the iterative process, to the more advanced where improved approximations are built up on the basis of information gathered during the descent process [63].

\footnote{$F$ is the ratio between the temporal oversampling factor and the spatial oversampling factor used to generate spatially correlated LEM impulse responses [13 Appendix B].}
A. Model Verification 1

The accuracy of the derived model has been verified through Monte-Carlo simulations using several different parameter sets. To conserve space, Fig. 4 shows a few of these simulations for $M = 2$ and LEM responses $h_1$ and $h_2$ with $N_h = 128$. The beamformer filters had $N_{BF} = 16$ and linear phase in the look direction. The AEC length was $N_{AEC} = 128$. The noise variance at each microphone was $10^{-2}$. The theoretical predictions (smooth red curves) are in very good agreement with the Monte Carlo simulations (300 runs). Values of $J[\infty]$ using (76) are shown by the red horizontal dotted lines.

(b) AR1(-0.9) $f(\mu_{AEC}, \mu_{BF}) = 2/3$

(c) AR1(-0.5) $f(\mu_{AEC}, \mu_{BF}) = 2/3$

(d) AR1(-0.5) $f(\mu_{AEC}, \mu_{BF}) = 2/3$

Fig. 4. Proposed model and Monte-Carlo simulation results based on 300 runs for different far-end signal statistics ($M = 2$, $N_h = 128$, $F = 4$, $N_{BF} = 16$, $N_{AEC} = 128$)

B. Model Verification 2

Consider a unit power first order autoregressive AR1(-0.9) far-end signal, 2 microphones, $h_0$ and $h_1$ with 500 taps each, generated according to the model in [18]. The desired DOA was assumed orthogonal to the microphone array. We assumed the absence of double-talk, and noises $r_0[n]$ and $r_1[n]$ were zero-mean white Gaussian with variance $10^{-2}$. The adaptive BF was designed with $N_{BF} = 16$, linear phase, and all-pass frequency response with $N_f = 16$. The AEC used $N_{AEC} = N_h + N_{BF} - 1$. Fig. 5 shows the predicted and simulated transient MOP. We tested 2 scenarios: $[\mu_{AEC}, \mu_{BF}] = [2.6191 \times 10^{-4}, 0.0262]$ and $[\mu_{AEC}, \mu_{BF}] = [3.9840 \times 10^{-4}, 0.0028]$. Fig. 5 shows excellent agreement between theory and predictions in both cases. Counterintuitively, the results show that a larger convergence speed does not necessarily imply a higher steady-state error.

C. Model Verification 3

The model in this paper is derived under stationarity assumptions for the input signals. Nevertheless, for nonstationary signals, it preserves sufficient information about the adaptation process to derive useful design guidelines. It is important to stress the fundamental difference between design guidelines and design rules. Design guidelines are not rules that should be followed to design the system with a desired exact performance. No stochastic model can provide such rules as analytical models for adaptive algorithm behavior always rely on assumptions needed for mathematical tractability. Though the stationarity assumption is not satisfied in most practical systems, it is largely recognized that the models derived using them can still show tendencies of the algorithm behavior for reasonably small degrees of nonstationarity. To illustrate the validity of the model even for a nonstationary input signal, 3 simulation scenarios were tested for a GSC-AEC system with $M = 2$, $N_h = 1024$, $N_{BF} = 16$, $N_{AEC} = 1039$, $\mu_{AEC} = \mu_{BF} = 3.1612 \times 10^{-4}$ and different degrees of nonstationarity $\eta$ [4]. Results are shown in Fig. 6. For higher the degrees of nonstationarity, the behavior of the system diverges from the theoretical prediction. However, the system performance is still close enough to the MC simulation to jumpstart the design choices.

D. Design Example 1

Consider an acoustic echo cancellation system with a reverberation time $T_R(60) = 60$ ms (typical of a car cabin) and a background noise level of $-20$ dB. We assume the frequency response in the desired DOA (broadside) can be modeled by a linear phase, delayless all-pass filter with $N_f = 16$ coefficients, i.e. $f = [1, 0_{1 \times N_f - 1}]^T$ [36]. The design goals are convergence of $|\hat{r}|_{dB} < -20$ in less than 2 s (evaluated at $n = 1.5E4$ with (62) and (70)).
We consider the design using the step-size matrix in (44). The frequency response model requires \( N_{BF} = N_f = 16 \). The LEM plants have length \( N_h > f_s T_R \) (60) = 480, and we thus set \( N_h = 500 \). The free parameters are then \( M \) and \( N_{AEC} \). The choices of \( M \) and \( N_{AEC} \) affect the computational complexity per iteration and the convergence speed.

Fig. 7 was produced evaluating (62) and (70) for \( M = \{1, 2\} \), \( N_{AEC} \) from 290 to 515, and \( \text{tr}(MR_{mod}) = \{2/300, \ldots , 2/3\} \) to compute \( J[\infty] \). If \( J[\infty] > -20 \) dB this configuration is discarded. (62) and (70) are evaluated for \( \mu_{AEC} = \{0.01, \ldots , 0.99\} \times 2(J[\infty] - J_{\min})/(J[\infty] \text{tr}(R_{u0u0})) \).

If more than one combination of \((\mu_{AEC}, \mu_{BF})\) is capable of reaching the desired cancellation at \( n = 1.5E4 \) then only the one with the lower \( J[\infty] \) is considered.

From Fig. 7, 2 candidate solutions were selected with \( J[\infty] \approx -21.5 \) dB and simulated using real speech signals (with pauses removed). Average results for an ensemble of 50 runs are shown in Fig. 8.

**TABLE I**

| \( M \) | \( N_{AEC} \) | \( \mu_{AEC} \) | \( \mu_{BF} \) | \( J[\infty] \) |
|---|---|---|---|---|
| 2 | 405 | 9.777E-04 | 6.460E-04 | -21.54 dB |
| 4 | 290 | 9.103E-04 | 1.696E-04 | -21.5 dB |

E. Design Example 2

To illustrate the use of the derived optimal step-matrix derived in (81), consider an acoustic echo canceler with \( M = 2 \) microphones for a large conference room with reverberation...
time \( T_R(60) = 0.45 \) s. Assuming a sampling rate of 8 kHz, the LEM plant length is about \( N_h \approx 0.45 \times 10^3 = 3600 \) coefficients. For this sampling rate, we also consider the frequency response in the desired DOA can be guaranteed with \( N_f = 16 \) constraints. In this simulation we consider only design choices with \( N_{AEC} = N_h + N_{BF} - 1 \). The far-end signal is modeled by a unity variance AR1(-0.9) random process and the noises in each microphone are assumed independent and modeled by Gaussian i.i.d. variables with variance 10^{-2}. In this design we desire a steady-state MOP of \(-22 \) dB and a \(-10 \) dB MOP after 1 second of convergence (\( n = 8000 \)). To verify the feasibility of this design we used the proposed statistical model to predict the transient behavior using the optimal step-matrix derived in [81]. Results are shown in Fig. 9. We observe that the whitening transformation is capable of achieving a MOP around \(-8 \) dB at \( n = 8000 \). As this is the design with the optimal convergence speed, it is not possible to achieve, at the same time, the desired steady-state MOP of \(-22 \) dB and less than \(-8 \) dB at \( n = 8000 \). Therefore it is not possible to design an BF-AEC system in the GSC form with \( T_R(60) = 0.45 \) and these performance requirements.

**F. Design Example 3**

Consider an BF-AEC system designed to work in a room with reverberation time \( T_R(60) = 100 \) ms. To guarantee a Public Switched Telephone Network (PSTN) quality signal, the sampling rate was chosen \( f_s = 8000 \). To model the LEM plant impulse responses we have chosen \( N_h = 1000 \) coefficients. Assume there are \( M = 2 \) microphones available and a reasonable frequency response in the DOA is achieved with \( N_{BF} = 16 \). Finally, an AEC length \( N_{AEC} = N_h + N_{BF} - 1 = 1015 \) coefficients was used. For this example, we assume an ideal double-talk detector is available. The initial DOA is assumed initially at a \( \pi/4 \) angle in relation to the broadside of the microphone array. During the first \( 10^6 \) samples, the adaptation occurs in the absence of near-end speech with equal step-sizes \( \mu_{AEC} = \mu_{BF} = 8.2147 \times 10^{-5} \). Then, a double-talk period occurs. The unitary power near-end speech, modelled as an AR1(-0.9) process, arrives from the broadside of the microphone array. We assume double-talk control logic, constraint and block matrix correction act instantly. The adaptation of the AEC is frozen by the double-talk detector \( (\mu_{AEC} = 0) \), and during the next \( 5.10^8 \) samples only the BF is adapted with \( \mu_{BF} = 8.2147 \times 10^{-5} \). During this period, the convergence is significantly faster as the BF-AEC structure is not jointly-optimized and the effective adaptive filter length is reduced to \( MN_{BF} - N_f \). During the next \( 10^6 \) samples, the near-end speech is removed and, considering the BF state from the double-talk period is a good initial solution, step-sizes are set to \( \mu_{AEC} = 9.3243 \times 10^{-5} \) and \( \mu_{BF} = 10^{-7} \). Finally, the LEM plant is subjected to an abrupt change in which a completely new \( H \) is used. In this configuration the new step-sizes are set to \( \mu_{AEC} = \mu_{BF} = 9 \times 10^{-5} \) to accelerate convergence after an abrupt LEM plant change [7]. The model predictions and Monte Carlo Simulations (ensemble of 50 runs) are compared on Fig. 10.

**IX. Conclusion**

This work presented a statistical analysis of a class of jointly optimized beamformer-assisted AEC. The analysis was performed for systems with the BF implemented in the GSC form and using the LMS algorithm. The analysis considered convergence control using a step size matrix to accommodate typical control logic implementations. We have shown that the joint optimization of the BF-AEC is equivalent to a LCMV problem. Thus, the derived analytical models can be used to predict the transient performance of general adaptive wideband beamformers. The stochastic model was determined for the transient and steady-state behaviors of the residual mean echo power for stationary Gaussian inputs. Convergence analysis lead to stability bounds for the step-size matrix. Design guidelines were derived from the analytical models. Monte Carlo simulations illustrated the accuracy of the theoretical models and the applicability of the proposed design guidelines. Finally, it was shown how a high convergence rate can be achieved using a quasi-Newton adaptation scheme in which the step-size matrix is designed to whiten the combined input vector.
ACKNOWLEDGEMENTS

The authors would like to thank the invaluable help of Prof. Abraham Alcaim, from CETUC at PUC-Rio, who generously provided the speech signal database from which the real speech signals used in the described experiments were taken.

REFERENCES

[1] C. Breining, P. Dreiseitl, E. Hänsler, A. Mader, B. Nitsch, H. Puder, T. Schertler, G. Schmidt, and J. Tilp. "Acoustic echo control. An application of very-high-order adaptive filters," Signal Processing Magazine, IEEE, vol. 16, no. 4, pp. 42–69, Jul. 1999.

[2] E. Hänsler and G. Schmidt, Acoustic Echo and Noise Control: A Practical Approach. Wiley-Interscience, 2004.

[3] B. Widrow, J. M. McCool, M. G. Larimore, and C. R. Johnson, "Stationary and nonstationary learning characteristics of the LMS adaptive filter," Proceedings of the IEEE, vol. 64, no. 8, pp. 1151–1162, 1976.

[4] D. G. Manolakis, V. K. Ingle, and S. M. Kogon, Statistical and adaptive signal processing: spectral estimation, signal modeling, adaptive filtering, and array processing. New York, NY, USA: McGraw-Hill, 2000.

[5] S. Haykin, Adaptive Filter Theory (2nd Edition). Prentice Hall, Sep. 1996.

[6] N. Bershad and J.-Y. Tourneret, "Echo cancellation – a likelihood ratio test for double-talk versus channel change," Signal Processing, IEEE Transactions on, vol. 54, no. 12, pp. 4572–4581, 2006.

[7] J.-Y. Tourneret, N. Bershad, and J. Bermudez, "Echo cancellation – the generalized likelihood ratio test for double-talk versus channel change," Signal Processing, IEEE Transactions on, vol. 57, no. 3, pp. 916–926, 2009.

[8] J. Gunther, "Learning echo paths during continuous double-talk using semi-blind source separation," Audio, Speech, and Language Processing, IEEE Transactions on, vol. 20, no. 2, pp. 646–660, Feb. 2012.

[9] B. Van Veen and K. Buckley, "Beamforming: a versatile approach to spatial filtering," ASSP Magazine, IEEE, vol. 5, no. 2, pp. 4–24, Apr. 1988.

[10] H. L. Van Trees, Optimum Array Processing (Detection, Estimation, and Modulation Theory, Part IV), 1st ed. Wiley-Interscience, Mar. 2002.

[11] W. Kellermann, "Strategies for combining acoustic echo cancellation and adaptive beamforming microphone arrays," in Acoustics, Speech, and Signal Processing, 1997. ICASSP'97., 1997 IEEE International Conference on, vol. 1, Apr. 1997, pp. 219–222.

[12] W. Herboldt and W. Kellermann, "GSAEC - acoustic echo cancellation embedded into the generalized sidelobe canceller," in Proc. European Signal Processing Conference (EUSIPCO), vol. 3, Sep. 2000, pp. 1843–1846.

[13] W. Herboldt, "Limits for generalized sidelobe cancellers with embedded acoustic echo cancellation," in Acoustics, Speech, and Signal Processing, 2001. Proceedings. (ICASSP ‘01). 2001 IEEE International Conference on, vol. 5, 2001, pp. 3241–3244.

[14] W. Herboldt, W. Kellermann, and S. Nakamura, "Joint optimization of LCMV beamforming and acoustic echo cancellation," in Proc. European Signal Processing Conference (EUSIPCO), 2004, pp. 2003–2006.

[15] M. Guo, S. Jensen, and S. Grant, "Analysis of closed-loop acoustic feedback cancellation systems," in Acoustics, Speech, and Signal Processing, 2013. ICASSP 2013. Proceedings. 2013 IEEE International Conference on, vol. 3, Mar. 2013, pp. ii/77 – ii/80 Vol. 3.

[16] K.-D. Kammeyer, M. Kallinger, and A. Mertins, "New aspects of combining echo cancellers with beamformers," in Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing, vol. 3, Philadelphia, PA, USA, Mar. 2005, pp. 137–140.

[17] M. H. Maruo, J. C. M. Bermudez, and L. S. Resende, "Statistical analysis of the jointly-optimized acoustic echo cancellation BF-AEC structure," in Acoustics, Speech, and Signal Processing, 2013. ICASSP 2013. Proceedings. 2013 IEEE International Conference on, pp. 5840–5844.

[18] M. H. Maruo, J. C. M. Bermudez, and L. S. Resende, "Statistical analysis of a jointly optimized beamformer-assisted acoustic echo canceler," Signal Processing, IEEE Transactions on, vol. 62, no. 1, pp. 252–265, Jan. 2014.

[19] M. Guo, T. B. Elmedyeb, S. H. Jensen, and J. Jensen, "Analysis of acoustic feedback/echo cancellation in multiple-microphone and single-loudspeaker systems using a power transfer function method," Signal Processing, IEEE Transactions on, vol. 59, no. 12, pp. 5774–5788, Dec. 2011.
