EFFICIENT MINIMUM NORM MULTI-CHANNEL ACOUSTIC ECHO CANCELLATION

Jeroen Veen and Piet Sommen

Eindhoven University of Technology, the Netherlands.
J.Veen@tue.nl, P.C.W.Sommen@tue.nl .

ABSTRACT

In multi-channel acoustic echo cancellation a problem of non-uniqueness occurs. Due to the high correlation of the recorded signals the system of normal equations is rank-deficient and there is an infinite number of solutions. In this paper, we propose an unambiguous solution to this problem by choosing the minimum norm solution to the system of normal equations, thus avoiding preprocessing of the recorded signals. Furthermore, an efficient implementation of the solution is presented using a fast eigenvalue decomposition of the autocorrelation matrix.

1. INTRODUCTION

Multi-channel acoustic echo cancellation (MCAEC) is typically intended for use in high-quality teleconferencing systems. In such a teleconferencing system, acoustic echoes emerge as a result of acoustic coupling of a microphone and a loudspeaker within the same room. In order to allow for transparent and full-duplex communication, an echo canceler is employed to remove all acoustic echoes.

In case of multiple audio channels, if a single person is speaking, then the sounds that are recorded by the microphones are highly correlated. As a consequence of this correlation a problem of non-uniqueness occurs [1] [5], and straightforward generalization of the single-channel case does not lead to the desired echo cancellation in multi-channel audio systems. In literature, a number of methods is available that deal with the non-uniqueness problem by processing the recorded signals before they are locally reproduced by the loudspeakers, *e.g.* [1] [5]. Preprocessing is done in order to reduce the high correlation of the signals, such that multi-channel generalizations of single-channel acoustic echo cancelers can be applied to the processed signals. Evidently, these methods change the problem in favor of the tool, which is not the approach chosen in this paper. Furthermore, since transparent audio channels are desired in high-quality teleconferencing systems, preprocessing is only possible to a limited extent.

The objective of the work presented in this paper is to provide a solution to the non-uniqueness problem without preprocessing the signals. To that end, a new approach to MCAEC is chosen.

The paper is organized as follows. In Section II, MCAEC is classified as a rank-deficient system identification problem. In Section III a solution to such a problem is formulated. Furthermore, an efficient method is presented for computation of this solution. In Section IV this method is conveniently fit into an existing multi-channel frequency domain adaptive filter (MCFDAF) algorithm. Subsequently in Section V, simulation results of the application of these concepts to MCAEC are shown. Finally, in Section VI conclusions are drawn.

In this paper, the following notations are used: Matrices and vectors are boldfaced and underlined, respectively. The l, m-th element of matrix $\mathbf{A}$ is denoted as $(\mathbf{A})_{lm}$. The superscript ^{t} denotes the transpose operator, the superscript ^{h} denotes the hermitian operator.

2. RANK-DEFICIENT SYSTEM IDENTIFICATION

Consider a schematic diagram for stereophonic acoustic echo cancellation (SAEC) which is a special case of MCAEC, as shown in Figure 1. SAEC is easily generalized to acoustic echo cancellation with an arbitrary number of channels, and serves as a guideline throughout this paper. Echo cancellation for a single microphone in the near-end room is discussed, a similar approach applies to the other microphones. The canceler estimates the contribution of the loudspeaker signals to the microphone signal by identifying the corresponding echo paths, therefore MCAEC is regarded as a straightforward multi-input, single-output (MISO) system identification problem.

More often than not, people within the same room are not speaking simultaneously, and there is only a single acoustic source in the far-end room. Therefore a causal relationship between the input signals exist [1],

$$\underline{g}_2^t \underline{x}_1[k] = \underline{g}_1^t \underline{x}_2[k], \tag{1}$$

where $\underline{g}_j = (g_{j,N-1}, \ldots, g_{j,1}, g_{j,0})^t$ for $j = 1, 2$ are FIR models of the acoustic impulse responses in the far-end room and $\underline{x}_j[k] = (x_j[k-N+1], \ldots, x_j[k-1], x_j[k])^t$ are discrete-time versions of the loudspeaker signals.

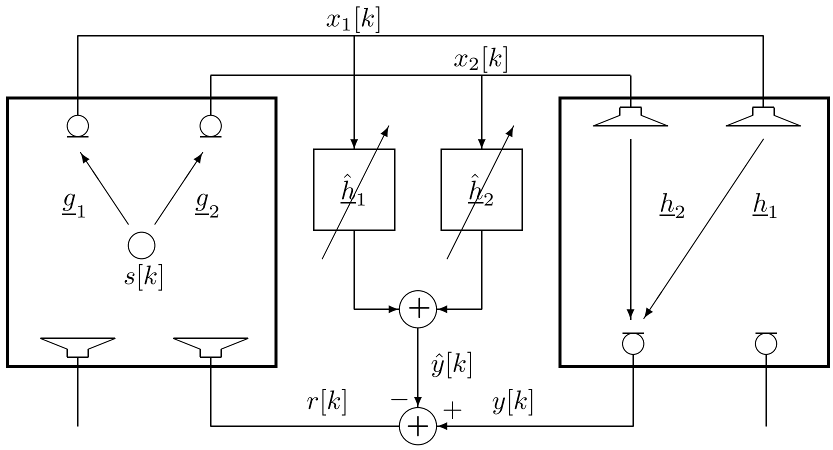

Figure 1: Schematic diagram for SAEC.

An FIR filter $\underline{\hat{h}}_j[k] = (\hat{h}_{j,N-1}[k], \ldots, \hat{h}_{j,1}[k], \hat{h}_{j,0}[k])^t$ is used to identify the jth acoustic echo path. Using the definitions $\underline{x}^t[k] = (\underline{x}_1^t[k], \underline{x}_2^t[k])$ and $\underline{\hat{h}}^t[k] = (\underline{\hat{h}}_1^t[k], \underline{\hat{h}}_2{}^t[k])$, the following system of normal equations is formulated,

$$\mathbf{R}_x[k]\underline{\hat{h}}[k] = \underline{r}_{xy}[k], \tag{2}$$

where $\mathbf{R}_x[k] = \frac{1}{k}\sum_{i=1}^{k} \underline{x}[i]\underline{x}^t[i]$ is the autocorrelation matrix of $\underline{x}[k]$ and $\underline{r}_{xy}[k] = \frac{1}{k}\sum_{i=1}^{k} \underline{x}[i]y[i]$ is the crosscorrelationvector of $\underline{x}[k]$ and $y[k]$. If expression (1) holds exactly then a number of equations in the system (2) is linearly dependent. Consequently, the nullspace of $\mathbf{R}_x[k]$ is non-trivial, which implies that the autocorrelation matrix is rank-deficient, and there is no unique solution to the system (2).

3. MINIMUM NORM SOLUTION

We propose to choose the minimum norm solution to (2) since it is unique and well-defined. Choice for the minimum norm solution is not arbitrary since it avoids projection onto the nullspace of $\mathbf{R}_x[k]$. The solution is formulated using the eigenvalue decomposition (EVD) of $\mathbf{R}_x[k]$, given by $\mathbf{R}_x[k] = \mathbf{Q}[k]\mathbf{\Lambda}[k]\mathbf{Q}^t[k]$. Since $\mathbf{R}_x[k]$ is rank-deficient, a number of eigenvalues are zero, *i.e.*

$$\mathbf{\Lambda}[k] = \begin{pmatrix} \mathbf{\Lambda}_s[k] & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}, \quad \mathbf{Q}[k] = \begin{pmatrix} \mathbf{Q}_s[k] & \mathbf{Q}_n[k] \end{pmatrix},$$

where the eigenvectors in $\mathbf{Q}_s[k]$ and $\mathbf{Q}_n[k]$ span the column space and the nullspace of $\mathbf{R}_x[k]$ respectively. The minimum norm solution is now given by

$$\underline{\hat{h}}_{min}[k] = \mathbf{R}_x^{\dagger}[k]\underline{r}_{xy}[k], \tag{3}$$

where $\mathbf{R}_x^{\dagger}[k] = \mathbf{Q}_s[k]\mathbf{\Lambda}_s^{-1}[k]\mathbf{Q}_s^t[k]$ is referred to as the pseudo-inverse of $\mathbf{R}_x[k]$.

Computation of an EVD has been a matter of considerable study; also quite some effort has been spent on obtaining fast methods for updating an EVD, *e.g.* [2]. However, even the most sophisticated methods have excessive computational complexity in comparison with efficient implementations of adaptive filters as applied to MCAEC, *e.g.* [4]. Due to the high computational complexity of the EVD of $\mathbf{R}_x[k]$, real-time computation of the minimum norm solution (3) is not feasible. In order to reduce complexity, we propose an alternative.

The autocorrelation matrix $\mathbf{R}_x[k]$ is partitioned as

$$\mathbf{R}_x[k] = \begin{pmatrix} \mathbf{R}_{x_1x_1}[k] & \mathbf{R}_{x_1x_2}[k] \\ \mathbf{R}_{x_2x_1}[k] & \mathbf{R}_{x_2x_2}[k] \end{pmatrix},$$

where $\mathbf{R}_{x_ax_b}[k] = \frac{1}{k}\sum_{i=1}^{k} \underline{x}_a[i]\underline{x}_b^t[i]$ for $a, b = 1, 2$ is the crosscorrelation matrix of $\underline{x}_a[k]$ and $\underline{x}_b[k]$. We assume that all signals are weakly stationary, therefore each crosscorrelation matrix has Toeplitz structure. It is well-known [3] that Toeplitz matrices are approximately diagonalized by the discrete Fourier transform (DFT) matrix. The $N \times N$ DFT matrix $\mathbf{F}$ is defined as

$$(\mathbf{F})_{ab} = e^{-\jmath 2\pi \frac{(a-1)\cdot(b-1)}{N}} \quad \text{for } 1 \leq a, b < N$$

The autocorrelation matrix is approximately block diagonalized as follows

$$\frac{1}{N}(\mathbf{I} \otimes \mathbf{F})^h \, \mathbf{R}_x[k] \, (\mathbf{I} \otimes \mathbf{F}) \approx \mathbf{P}_x[k], \tag{4}$$

where $\otimes$ symbolizes the Kronecker product and, in case of SAEC, $\mathbf{I}$ is the 2×2 identity matrix.

The power matrix $\mathbf{P}_x[k]$ contains the diagonalized blocks,

$$\mathbf{P}_x[k] = \begin{pmatrix} \mathbf{P}_{x_1x_1}[k] & \mathbf{P}_{x_1x_2}[k] \\ \mathbf{P}_{x_2x_1}[k] & \mathbf{P}_{x_2x_2}[k] \end{pmatrix}.$$

Since the power matrix is Hermitian its EVD equals $\mathbf{P}_x[k] = \mathbf{V}[k]\mathbf{\Gamma}[k]\mathbf{V}^h[k]$. It is easily shown that the EVD of a $2N \times 2N$ matrix with 2^2 diagonal blocks is identical to the distinct EVDs of N 2×2 matrices. Consequently, the computational complexity of the EVD of $\mathbf{P}_x[k]$ is dramatically reduced. For example, if orthogonal iteration is used to calculate the EVDs, then the complexity of the EVD of $\mathbf{R}_x[k]$ and $\mathbf{P}_x[k]$ equal $\mathcal{O}(N^3)$

and $\mathcal{O}(N)$, respectively.
Note that the matrices $\mathbf{\Lambda}[k]$ and $\mathbf{\Gamma}[k]$ are asymptotically equivalent, yet ordered differently. There are no practical consequences involved in this matter; nevertheless, in order to keep consistency with previous definitions, it is assumed that the EVD of $\mathbf{P}_x[k]$ is reordered such that $\mathbf{\Gamma}[k] \approx \mathbf{\Lambda}[k]$. Consequently $\mathbf{V}[k]$ is partitioned as

$$\mathbf{V}[k] = \begin{pmatrix} \mathbf{V}_s[k] & \mathbf{V}_n[k] \end{pmatrix} .$$

The pseudo-inverse of $\mathbf{R}_x[k]$ is approximated by

$$\mathbf{R}_x^\dagger[k] \approx \frac{1}{N}(\mathbf{I} \otimes \mathbf{F})\ \mathbf{P}_x^\dagger[k]\ (\mathbf{I} \otimes \mathbf{F})^h ,$$

where $\mathbf{P}_x^\dagger[k] = \mathbf{V}_s[k]\mathbf{\Lambda}_s^{-1}[k]\mathbf{V}_s^h[k]$ is the pseudo-inverse of $\mathbf{P}_x[k]$. Note that $\mathbf{\Lambda}_s[k]$ is a diagonal matrix and is inverted efficiently. In addition, since $\mathbf{V}[k]$ and $\mathbf{P}_x[k]$ are isomorphic the complete computation of $\mathbf{P}_x^\dagger[k]$ requires $\mathcal{O}(N)$ multiplications.
The approximated minimum norm solution (3) equals

$$\underline{\hat{h}}_{min}[k] \approx \frac{1}{N}(\mathbf{I} \otimes \mathbf{F})\ \mathbf{P}_x^\dagger[k]\ (\mathbf{I} \otimes \mathbf{F})^h\ \underline{r}_{xy}[k] . \qquad (5)$$

4. APPLICATION TO MCFDAF

The concept of the previous section is conveniently fit into existing MCFDAF algorithms - for more details on efficient MCFDAF algorithms see *e.g.* [4]. As an example, in this section a variant of the recursive least-squares (RLS) algorithm is applied to SAEC, given by

$$\begin{aligned} \mathbf{R}_x[k] &= \alpha\mathbf{R}_x[k-1] + \frac{\underline{x}[k]\underline{x}^t[k]}{2N} \\ r[k] &= y[k] - \underline{x}^t[k]\underline{\hat{h}}[k] \\ \underline{\hat{h}}[k+1] &= \underline{\hat{h}}[k] + \beta\mathbf{R}_x^\dagger[k]\underline{x}[k]r[k] \end{aligned}$$

where α is an exponential forgetting factor and β is the stepsize. The first line shows the update of an approximation of the autocorrelation matrix. The second line shows the filter part of the algorithm - which is not of interest in this paper. The last line shows the update of the model filter coefficients. Subsequently, the matrix $(\mathbf{I} \otimes \mathbf{F})$ is introduced into the algorithm, resulting in

$$\begin{aligned} \mathbf{P}_x[k] &= \alpha\mathbf{P}_x[k-1] + \frac{\underline{X}[k]\underline{X}^*[k]}{2N} \\ \underline{\hat{h}}[k+1] &\approx \underline{\hat{h}}[k] + \beta(\mathbf{I} \otimes \mathbf{F})\mathbf{P}_x^\dagger[k]\underline{X}^*[k]r[k] \end{aligned}$$

where $\underline{X}[k] = (\mathbf{I} \otimes \mathbf{F})\underline{x}[k]$ is the frequency domain data vector and $\mathbf{P}_x[k]$ is an approximation of the power matrix. Since $\mathbf{P}_x[k]$ has diagonal blocks, it is updated efficiently [4]. Moreover, the EVD is computed efficiently by calculating the separate EVDs of N different 2×2 matrices. To that end, for each frequency bin p the matrix $\mathbf{\Theta}_p[k]$ is constructed

$$\mathbf{\Theta}_p[k] = \begin{pmatrix} (\mathbf{P}_{x_1x_1}[k])_{pp} & (\mathbf{P}_{x_1x_2}[k])_{pp} \\ (\mathbf{P}_{x_2x_1}[k])_{pp} & (\mathbf{P}_{x_2x_2}[k])_{pp} \end{pmatrix} ,$$

where $(\mathbf{P}_{x_ix_j}[k])_{pp}$ is the p-th element on the diagonal of the i,j-th block of the power matrix $\mathbf{P}_x[k]$, *i.e.* the power in the p-th frequency bin of the cross-power matrix of $x_i[k]$ and $x_j[k]$. Subsequently, the EVD $\mathbf{\Theta}_p[k] = \mathbf{V}_p[k]\mathbf{\Lambda}_p[k]\mathbf{V}_p^h[k]$ is calculated. The matrix $\mathbf{\Lambda}_p[k]$ contains 2 eigenvalues of the power-matrix $\mathbf{P}_x[k]$. By executing this process for each frequency bin $p = 0, 1, \ldots, N-1$, all $2N$ eigenvalues and eigenvectors of the power-matrix are obtained. Subsequently, the non-zero eigenvalues from $\mathbf{\Lambda}[k]$ are selected and placed on the diagonal of the matrix $\mathbf{\Lambda}_s[k]$. The corresponding eigenvectors are stacked side by side in the matrix $\mathbf{V}_s[k]$. Finally the pseudo-inverse of the power matrix is calculated using the EVD of $\mathbf{P}_x[k]$.

5. SIMULATIONS

Simulations were done with a block-version of the algorithm presented in the previous section that has been implemented in Matlab [4]. The source signal $s[k]$ in the far-end room is zero-mean white noise. The source signal is filtered by two synthesized far-end room impulse responses $\underline{g}_1$ and $\underline{g}_2$ each of length 1024, resulting in two near-end loudspeaker signals $x_1[k]$ and $x_2[k]$. White measurement noise is added to the loudspeaker signals, such that the signal to noise ratio (SNR) equals 30 dB. Subsequently, the loudspeaker signals are filtered by two synthesized near-end room impulse responses $\underline{h}_1$ and $\underline{h}_2$ each of length 1024, and added, resulting in a near-end microphone signal $y[k]$. White measurement noise is added to the microphone signal, such that SNR = 30 dB. In the simulations we will compare our approach with the conventional MCFDAF algorithm. In either case the FIR model filters $\underline{\hat{h}}_1$ and $\underline{\hat{h}}_2$ each have $N = 1024$ tabs. The forgetting factor in the power estimation is $\alpha = 0.9$ and the stepsize in the update is $\beta = 0.001$.
With regard to this simulation setup it should be noted that the zero eigenvalues of the autocorrelation matrix are not actually zero but very small. On one hand this discrepancy is caused by perturbations in relationship (1) that arise as a result of the finite filter length of the model filters. On the other hand the discrepancy is due to the measurement noise that has been added to the signals. In order to distinguish between 'zero' and non-zero eigenvalues we introduce a predefined SNR. Using this predefined SNR we can define the minimum signal power with respect to the overall maximum signal power. Eigenvalues that do not exceed the minimum signal power are now defined as being zero. In the simulations the predefined SNR is set to 30 dB.
In the first part of the simulations we show the consistency of our approach. Figure 2 shows the mean squared error (MSE) of the conventional algorithm and the MSE of the proposed algorithm. In order to smooth the curves, the signals are averaged over 512 samples.

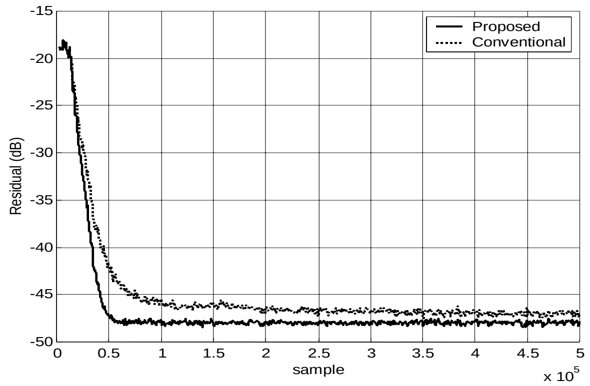

Figure 2: Mean squared error.

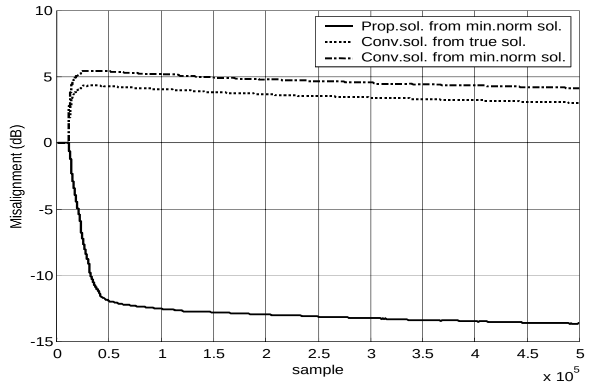

Figure 3: Misalignment from solution.

Evidently, the echo cancellation properties of both algorithms are similar. Subsequently, for both algorithms the misalignment from the desired solution is computed, which is defined as $\|\underline{h} - \hat{\underline{h}}[k]\|/\|\underline{h}\|$ for the conventional and as $\|\underline{h}_{min} - \hat{\underline{h}}[k]\|/\|\underline{h}_{min}\|$ for the proposed algorithm; where $\underline{h} = (\underline{h}_1^t, \underline{h}_2^t)^t$ is the true solution and $\underline{h}_{min}$ is obtained by means of expression (3) using all available data. Figure 3 shows the misalignment of both algorithms; it is concluded that the new algorithm converges to the minimum norm solution, while the conventional algorithm does not converge to the true solution. Thus the new algorithm is consistent, while the conventional algorithm is not. In order to show that the conventional algorithm does not converge to the minimum norm solution either a third curve is plotted in Figure 3 representing the misalignment of the conventional solution from the minimum norm solution.

In the second part of the simulations we show that the proposed algorithm converges to an unambiguous solution, while the conventional algorithm converges to an arbitrary solution. In this experiment the previous simulation is repeated without measurement noise added to the signals. Thus the identification problem itself remains unaltered, just the conditions have changed. We use the misalignment to compare the solutions from the two simulations. In Figure 4 the misalignment $\|\hat{\underline{h}}_B[k] - \hat{\underline{h}}_A\|/\|\hat{\underline{h}}_A\|$ is plotted for both algorithms. Here

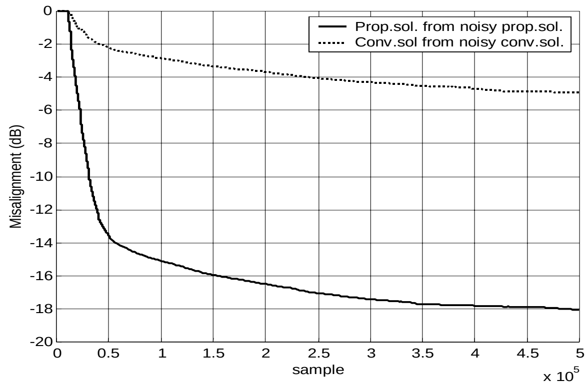

Figure 4: Misalignment from previous solution.

$\hat{\underline{h}}_A$ refers to the final solution that is obtained in the first simulation and $\hat{\underline{h}}_B[k]$ refers to the second simulation, the noiseless case. It is concluded that the proposed algorithm converges to a well-defined solution, while the conventional algorithm does not.

6. CONCLUSIONS

In this paper we have proposed a new solution to the non-uniqueness problem that occurs in MCAEC, by choosing the minimum norm solution to the system of normal equations. This approach provides an unambiguous solution, yet does not rely upon preprocessing. Furthermore, our method is conveniently applied to existing MCFDAF algorithm with small additional computational load. Since the minimum norm solution is well-defined, it is expected that tracking properties and robustness of MCFDAF algorithms will improve.

7. REFERENCES

[1] J. Benesty, D. Morgan, and M. Sondhi. A better understanding and an impoved solution to the specific prolems of stereophonic acoustic echo cancellation. *IEEE Trans. Speech Audio Processing*, vol. 6:156–165, Mar. 1998.

[2] G. Golub and C. van Loan. *Matrix computations.* The Johns Hopkins university press, Baltimore and London, third edition, 1996.

[3] R. Gray. On the asymptotic eigenvalue distribution of toeplitz matrices. *IEEE Trans. IT*, vol. IT-18:725–730, Nov. 1972.

[4] D. Schobben. *Efficient adaptive multi-channel concepts in acoustics.* PhD thesis, Eindhoven university of technology, the Netherlands, 1999.

[5] M. Sondhi, D. Morgan, and J. Hall. Stereophonic acoustic echo cancellation-an overview of the fundamental problem. *IEEE Signal Processing Letters*, vol. 2:148–151, Aug. 1995.