A New Auxiliary-Vector Algorithm with Conjugate Orthogonality for Speech Enhancement

Shengkui Zhao¹, Douglas L Jones²

¹Advanced Digital Sciences Center, Illinois at Singapore, Singapore
²Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Illinois, USA

shengkui.zhao@adsc.com.sg, jones@ifp.illinois.edu

Abstract

In this paper, we propose a new auxiliary-vector (AV) algorithm using the conjugate orthogonality for speech enhancement. When only a limited data record is available, the AV algorithm is the state-of-the-art for obtaining the minimum-variance-distortionless (MVDR) filter. However, the current AV algorithms suffer from convergence problems when applied to the speech enhancement. Based on the conjugate Gram-Schmidt process, we develop new auxiliary vectors that are conjugate orthogonal and apply them to the AV algorithm. The proposed conjugate AV algorithm converges to the optimal MVDR solution within finite steps no greater than the filter dimension. Theoretical analysis establishes formal convergence of the proposed conjugate AV algorithm. Our experiments using the synthetic and real speech data show favorites of the new proposal over the state-of-the-art approaches.

Index Terms: speech enhancement, microphone arrays, correlation, convergence, adaptive signal processing

1. Introduction

Microphone arrays for speech enhancement have drawn many interests in the applications such as teleconferencing, hearing aids, voice communication and automatic speech recognition. In the literature, many algorithms were developed using microphone arrays to recover speech signals from noisy backgrounds [1, 2, 3, 4, 5, 6]. One of the most popular techniques is known as the linearly constrained minimum variance (LCMV) beamformer [1]. It uses constraints to preserve the desired signal while minimizing power due to interfering and noise signals arriving from directions other than the direction of interest. When it preserves a distortionless response to the direction of interest, the beamformer is known as the minimum-variance-distortionlessness-response (MVDR) filter or Capon filter [8].

Conventionally, the computation of the MVDR filter begins with the calculation of the inverse of the input correlation matrix. To avoid any explicit correlation matrix inversion, the auxiliary-vector (AV) algorithms were proposed by Podos and Karystinos in [9] and by Pados and Batalama in [16]. Computationally, the AV algorithms are simple, noninvasive and recursive procedures. The AV algorithms start from the matched filter and generate a sequence of filters that converges to the MVDR solution. When only a limited data record is available, the AV algorithms are the state-of-the-art for obtaining the MVDR filter. For the direct-sequence/code-division multiple-access (DS/CDMA) communication systems, the AV algorithms are shown to be superior than the Frost algorithm [1], the recursive-least-squares (RLS) algorithm [1] and the sample-matrix inversion (SMI) estimator [9].

However, for speech enhancement there often presents multiple interfering sources together with the target speech source. For such nonstationary speech inputs, the above AV algorithms suffer from the intrinsic convergence problem [7]. Specifically, the AV algorithm proposed in [9] uses pairwise orthogonal auxiliary vectors. The updating direction of the current step contains the similar directions as the earlier steps. Therefore, the convergence of the sequence of estimates to the MVDR solution highly depends on the eigenstructure of the input correlation matrix and the convergence is usually very slow for speech signals. On the other hand, the AV algorithm proposed in [16] uses mutually orthogonal auxiliary vectors. However, the proposed step-size factors in [16] cannot guarantee the convergence of the algorithm. We found that the optimal step-size factors that allow the AV algorithm to converge to the MVDR solution need the calculation of the inverse of the input correlation matrix. It was also verified in our simulated experimental results that the AV algorithm of [16] stops the estimate updating before it converges to the MVDR solution. To solve for the convergence problem of the AV algorithms, we propose to use the conjugate auxiliary vectors that are derived from the conjugate Gram-Schmidt process [14] and mutually orthogonal with respect to the input correlation matrix. Moreover, we derive the step-size factors that guarantee a finite convergence of the new AV algorithm to the optimal MVDR solution. We will provide a theoretical analysis to establish the finite convergence and demonstrate the superior performance of the new AV algorithm for speech enhancement.

2. The Previous Auxiliary Vector Algorithms

2.1. Problem Formulation

Let’s consider an M-microphone system and the M microphones receive the target speech from the look direction contaminated by noise and interfering speech signals. The frequency-domain microphone signals \(X_m(k,l), m = 1,2,\ldots,M\) are represented as

\[
X_m(k,l) = e_m(k,\theta)S(k,l) + V_m(k,l)
\]

where \(k\) is the frequency bin index and \(l\) the frame index of a short-time segment of input data, \(e_m(k,\theta)\) denotes the time-invariant frequency response of the \(m\)th microphone to the direction-of-arrival (DOA), \(\theta\) of the target source, \(S(k,l)\) and \(V_m(k,l)\) denote the short-time Fourier transformed speech sig-
nal and interference-plus-noise signal, respectively. The multi-channel speech enhancement methods aim to estimate the desired signal $S(k,l)$ from the multichannel noisy microphone signals. Let $x(k,l), w(k,l) \in \mathbb{C}^{M \times 1}$ be the microphone signal vector and the MVDR filter weight vector, respectively. The beamformer output spectrum, $Y(k,l)$, is formulated as

$$Y(k,l) = w^H(k,l)x(k,l).$$  \hspace{1cm} (2)

The time-domain beamformer output is therefore obtained by taking the short-time inverse Fourier transform of $Y(k,l)$. To find the weight vector, the MVDR filter was designed to minimize the total output signal power at each frequency bin while constraining the filtering response of the speech signal from the look direction to be unity [8]. The cost function is formulated by the method of Lagrange multipliers as

$$\mathcal{F}(w) = \frac{1}{2} w^H R_{xx} w + \Re\{\lambda (w^H e - 1)\}$$  \hspace{1cm} (3)

where $\lambda \in \mathbb{C}$ is the Lagrange’s multiplier, $R_{xx}$ denotes the input correlation matrix, $e$ denotes the steering vector, and $\Re\{\cdot\}$ denotes the real value. The frequency bin index $k$, the frame index $l$ and the DOA $\theta$ are omitted for the sake of conciseness. Minimizing the cost function $\mathcal{F}(w)$ with respect to $w$ leads to the well-known MVDR solution [8]

$$w_{MVDR} = R_{xx}^{-1} e.$$  \hspace{1cm} (4)

The conventional computation of the MVDR solution is often involved with explicit or implicit matrix inversion [13]. In this paper, we are interested in investigating the iterative algorithms for the direct calculation of the MVDR solution since the speech enhancement adopts a short-time processing approach.

### 2.2. The Previous Auxiliary Vector Algorithms

To iteratively solve for the MVDR solution, the auxiliary vector (AV) algorithms [9, 16] are the state-of-the-art approaches that avoid the calculation of the inverse of the correlation matrix. In this section, we briefly introduce the two popular AV algorithms.

The Pados and Karysinos’s AV (PKA V) algorithm [9] starts from the conventional matched filter $w_1 = \frac{e}{|e|^2}$ and takes the following updating form for $w_{n+1}$:

$$w_{n+1} = w_n - \mu_n g_n, n = 1, 2, 3...$$  \hspace{1cm} (5)

where the vector $g_n$ is termed as the auxiliary vector. The auxiliary vectors were developed to be orthogonal to the steering vector $e$ and to maximize the magnitude of the cross correlation between $w_n^H x$ and $g_n^H x$, with the constraints $g_n^H e = 0$. It is mathematically represented as

$$g_n = \arg \max_{g} |w_n^H R_{xx} g|, \text{ subject to } g^H e = 0.$$  \hspace{1cm} (6)

For speech processing, the equation (6) avoids a normalization compared to [9]. The final auxiliary vector is given as

$$g_n = R w_n$$  \hspace{1cm} (7)

where $R \triangleq \left( I - \frac{e e^H}{|e|^2} \right) R_{xx}$ is a projection matrix of $R_{xx}$.

The step-size scaler $\mu_n$ was derived using the criterion that minimizes the variance at the output of $w_{n+1}$

$$\frac{\partial (w_{n+1}^H R_{xx} w_{n+1})}{\partial \mu_n} = 0.$$  \hspace{1cm} (8)

where $\partial$ stands for the partial derivative and $*$ denotes the complex conjugate. Substituting the weight vector update equation (5) into (7) results in

$$\mu_n = \frac{g_n^H R_{xx} w_n}{g_n^H R_{xx} g_n}.$$  \hspace{1cm} (9)

It was shown that the successive auxiliary vectors in the PKA V algorithm are pairwisely orthogonal. It implies that the updating directions are not optimal as the current auxiliary vector can be overlapped with the earlier auxiliary vectors.

Considering the above problem, the Pados and Batallana’s AV (PBA V) algorithm used mutually orthogonal auxiliary vectors as proposed in [16]. For a given correlation matrix of finite data size, the orthogonal auxiliary vector $g_n$ maximizes the cross-correlation magnitude $|w_n^H R_{xx} g_n|$, subject to the orthogonal constraints $g_n^H e = 0, g_n^H g_k = 0, i < n$. The auxiliary vector $g_n$ is given by [16]

$$g_n = R w_n - \sum_{j=1}^{n-1} g_j R_{xx} w_n g_j.$$  \hspace{1cm} (10)

where $g_1 = R_{xx} w_1 - (w_1^H R_{xx} w_1) w_1$. The corresponding complex scalar $\mu_n$ is computed by minimizing the output variance of the filter and it is given as the same form as (9).

Since there are only $N \leq M$ mutually orthogonal auxiliary vectors available, the PBA V algorithm has at most $M$ updating steps. Next, we show that the step-size factor $\mu_n$ cannot guarantee finite convergence of the PBAV algorithm. If we assume the PBAV algorithm converges to MVDR solution after $M$ iterations, we will have

$$w_{MVDR} = w_1 - \sum_{i=1}^{M} \mu_i g_i.$$  \hspace{1cm} (11)

The optimal complex scalar $\mu_n$ that satisfies above equation (11) can be easily solved. Substituting the MVDR solution (4) into (11) and multiplying $g_n$ to both sides of (11), the complex scalar $\mu_n$ is given as

$$\mu_{opt,n} = \frac{g_n^H R_{xx}^{-1} w_1}{w_1^H R_{xx}^{-1} w_1}. $$  \hspace{1cm} (12)

Here we used the orthogonality of $g_n$. Eq.(12) implies that the optimal value of the complex scalar $\mu_{opt,n}$ is a function of the inverse of the correlation matrix. Without computing the inverse of the correlation matrix, this optimal complex scalar cannot be obtained using (9). Using the optimal complex scalar in (12), the PBAV algorithm converges to the MVDR solution within $M$ iterations. While using the complex scalar in (9), the PBAV algorithm does not converge to the MVDR solution. The PBAV algorithm stops updating after $M$ iterations and introduces a higher mean-square (MS) estimation error.

### 3. The Proposed Conjugate Auxiliary Vector Algorithm

In this section, we derive the conjugate auxiliary vectors and prove that the AV algorithm with the conjugate auxiliary vectors can converge in at most $M$ steps.
3.1. Updating form of the proposed algorithm

Starting from the same matched filter as used in the PKAV and PBAV algorithms, we propose to update the estimates using a set of conjugate auxiliary vectors \( \{ p_1, p_2, \ldots, p_N \} \), \( 1 \leq N \leq M \). The conjugate auxiliary vectors are defined to be \( R - \text{orthogonal} \) [10]:

\[
p_n^H R p_j = 0, \quad n \neq j
\]

where the matrix \( R \) is the projection matrix of the correlation matrix \( R_{xx} \) on the space \( \{ I - \| e \|^2 \} \).

Let \( \{ w_1, w_2, \ldots, w_N \} \) be iterative estimates of the MVDR solution and the updating form is given as:

\[
w_{n+1} = w_1 - \sum_{i=1}^{n} \alpha_i p_i = w_n - \alpha_n p_n, \quad (n = 1, 2, \ldots, N)
\]

where \( w_1 = e/\| e \|^2 \), and \( n \) is the iteration index. The factor \( \alpha_n \) can be found by setting the directional derivative to zero:

\[
\frac{\partial F(w_{n+1})}{\partial \alpha_n} = 0
\]

By manipulating (15), the step size factor \( \alpha_n \) is given as:

\[
\alpha_n = -\frac{p_n^H R g_n}{p_n^H R p_n}
\]

We give the following Lemma for the proof of finite convergence of the proposed AV algorithm. Here we assume that the conjugate auxiliary vectors are available. We will provide the procedures to compute the conjugate auxiliary vectors in the next subsection.

**Lemma:** The iterative procedure (14) and (16) computes the MVDR solution \( w_{MVDR} \) within \( M \) steps. That is:

\[
w_{MVDR} = w_1 - \sum_{n=1}^{N} \alpha_n p_n, \quad 1 \leq N \leq M.
\]

**Proof:** To prove the above procedure really does compute \( w_{MVDR} \) in \( N \) steps. Let us express the error term \( \tilde{w}_i = w_{MVDR} - w_1 \) as a linear combination of search directions:

\[
\tilde{w}_i = -\sum_{n=1}^{N} \delta_n p_n.
\]

We only need to prove that the value of \( \delta_n \) is actually equal to \( \alpha_n \). Premultiplying the expression (18) by \( p_i^H R \) and using the \( R - \text{orthogonality} \), we simplify that:

\[
p_i^H R \tilde{w}_i = -\sum_{n=1}^{N} \delta_n p_i^H R p_n
\]

Therefore, we compute \( \delta_n \) as:

\[
\delta_n = -\frac{p_i^H R \tilde{w}_i}{p_i^H R p_i} = -\frac{p_i^H R (\sum_{n=1}^{N-1} \alpha_n p_n)}{p_i^H R p_i}
\]

\[
= -\frac{p_i^H R \tilde{w}_i}{p_i^H R p_i} = \alpha_i
\]

where we used the expressions \( \tilde{w}_i = w_{MVDR} - w_1 = \tilde{w}_1 + \sum_{n=1}^{N-1} \alpha_n p_n \) and \( R_{MVDR}, R_p = 0 \). Therefore, we have the expression (17).

3.2. Derivation of the conjugate auxiliary vectors

The conjugate auxiliary vector is set to \( p_1 = -R w_0 \) at \( n = 1 \) and then it is iteratively constructed by conjugation of the residuals following the conjugate Gram-Schmidt process [14] as follows:

\[
p_n = g_n + \sum_{j=1}^{n-1} \beta_n p_j, \quad (n = 2, 3, \ldots, N)
\]

where \( \beta_n \) is defined as the residual vector and \( \beta_n \) is defined for \( n > j \). Next, we need to find the step size factor \( \beta_n \).

Premultiplying the expression (23) by \( p_i^H R \) and using the \( R - \text{orthogonality} \), we have:

\[
\beta_n = -\frac{p_i^H R g_n}{p_i^H R p_i}.
\]

Using the orthogonal properties of \( p_i^H g_n = 0 \) and \( g_{n+1}^H g_n = 0, \quad (i < n) \) [11], we simplify the numerator term of (24) as:

\[
p_i^H R g_n = \begin{cases} -\frac{1}{\beta_{n+1}} g_n^H g_{n+1}, & n = i, \\
\frac{1}{\beta_{n-1}} g_{n-1}^H g_n, & n = i + 1, \\
0, & \text{otherwise}. \end{cases}
\]

Using (25) into (24), we simplify the parameter \( \beta_n \) as:

\[
\beta_{n-1} = -\frac{g_{n+1}^H g_n}{\alpha_n p_i^H R p_n}.
\]

Therefore, the updating of conjugate vector (23) can be rewritten as:

\[
p_{n+1} = g_{n+1} + \beta_n p_n.
\]

4. Experimental Results

In this section, we give the experimental study of the proposed conjugate AV algorithm for speech enhancement using both synthetic and real-world speech data. The performance is compared with the PKAV and PBAV algorithms.

4.1. Evaluation on Synthetic Data

We consider a simulated uniform linear array (ULA) of 10 elements with element distance of 4 cm. There are one target source and four interfering sources. The source signal processes are generated using a second-order Autoregressive (AR) process [12]:

\[
S_i(f, l) = \frac{1}{1 + a_1 e^{-j2\pi f} + a_2 e^{-j4\pi f}} V_i(f, l)
\]

where \( V_i(f, l) \) is a complex white noise process. The constants \( a_1 \) and \( a_2 \) are the AR parameters. The input signal vector of the ULA is modeled as:

\[
x(f, l) = \sum_{i=1}^{5} e(f, \theta_i) S_i(f, l) + v(f, l)
\]
where $\mathbf{e}(f, \theta_i)$ denotes steering vector of the source at $\theta_i$, and $\mathbf{v}(f, l)$ is an additive complex white-noise vector with correlation matrix of $\sigma_v^2 \mathbf{I}$. In this case, the input correlation matrix is given by

$$R_{xx} = \sum_{i=1}^{5} \sigma_e^2 \mathbf{e}(\theta_i) \mathbf{e}^H(\theta_i) + \sigma_v^2 \mathbf{I}$$  \hspace{1cm} (31)

Fig. 1 compares convergence of the sequence of estimates to the MVDR solution generated by the PKAV, PBAV and proposed AV algorithms using the correlation matrix $R_{xx}$ computed from (31). The convergence of the algorithms to the MVDR solution (4) is illustrated by the norm-square metric $\|\mathbf{w}_n - \mathbf{w}_{MVDR}\|^2$ as a function of the iteration step $n$. It is observed that the proposed AV algorithm converges to the MVDR solution at the iteration $n = 5$. The PKAV algorithm converges to the MVDR solution in an asymptotical manner. The PBAV algorithm stops updating at the iteration $n = 5$ before it converges to the MVDR solution.

4.2. Evaluation on Real Data

4.2.1. Experimental setup

In this experimental setup, we used a four-channel miniature acoustic-vector-sensor (AVS) microphone array [17, 18]. Our recordings were made in an office room with a reverberation time of RT60 = 300 ms. The ambient noises came from the air-conditioner and PC. The AVS microphone array and the loudspeakers were set at the same height. The target source was from the loudspeaker in front of the array ($0^\circ$) at a distance of 2 m. The two interfering sources were located at an incident angle of $60^\circ$ and $-60^\circ$. For each source, we selected 20 sentences of 3 to 5 seconds long from the Aurora 4 database. The recordings were sampled at 16 kHz. All the algorithms used a 512-point FFT and a Hann window with a length of 512 samples and 75% overlap.

4.2.2. Experimental results

We assess the performance of the various algorithms in terms of the Perceptual Evaluation of Speech Quality (PESQ) measure [15] and the speech spectrograms. Table 1 shows the PESQ scores obtained by the various algorithms at input SINRs 0 dB and 5 dB. It is seen that the proposed algorithm achieved the highest PESQ scores among the algorithms. The PBAV algorithm performed the worst. Example of spectrograms of unprocessed target and recorded noisy speech signals and also those of the outputs of the various algorithms are presented in Fig. 2. The superiority of the proposed AV algorithm over the PKAV and PBAV algorithms is apparent by comparing the residual interfering speech signals.

5. Conclusions

We have presented a new auxiliary-vector (AV) algorithm using the conjugate orthogonality for speech enhancement. The proposed conjugate AV algorithm achieves faster convergence to the MVDR solution compared to the state-of-the-art approaches. Our experimental results on the synthetic and real speech data demonstrated the better PESQ scores and spectrogram of the proposed AV approach for speech enhancement.

6. Acknowledgements

This study is supported by the research grant for the Human Sixth Sense Programme at the Advanced Digital Sciences Center from Singapore’s Agency for Science, Technology and Research (A*STAR).

Table 1: PESQ Score Obtained by the Various Processing Algorithms for Speech Captured in an Office Environment with Two Interfering Speech Signals and Ambient Noise for Different Input SINRs.

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>PESQ (SINR = 0 dB)</th>
<th>PESQ (SINR = 5 dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MVDR</td>
<td>2.30</td>
<td>2.61</td>
</tr>
<tr>
<td>Proposed</td>
<td>2.69</td>
<td>2.98</td>
</tr>
<tr>
<td>PKAV</td>
<td>2.45</td>
<td>2.73</td>
</tr>
<tr>
<td>PBAV</td>
<td>2.05</td>
<td>2.33</td>
</tr>
</tbody>
</table>
7. References


