DOA ESTIMATION METHOD FOR WIDEBAND COLOR SIGNALS BASED ON LEAST-SQUARES JOINT APPROXIMATE DIAGNOSALIZATION

L. Yip, C. E. Chen, R. E. Hudson, and K. Yao

Electrical Engineering Department, UCLA
Los Angeles, CA 90095-1594, USA
leanyip2000@yahoo.com; chiaoen@ucla.edu; ralph@ucla.edu; yao@ee.ucla.edu

ABSTRACT

Direction of Arrival (DOA) estimation is one of the major tasks in many acoustic array signal processing applications. In this work, we show that a Modified Joint Approximate Diagonlization (MJAD) method utilizing temporal structure of the wideband color signals can outperform the conventional wideband DOA estimation method such as Approximate Maximum Likelihood (AML) and Coherent Signal-Subspace Processing (CSSP).

1. INTRODUCTION

In various acoustic array signal processing applications such as acoustic sensor network for monitoring and automatic camera steering in teleconference, DOA estimation of the speech signals is a central task. In the past, various parametric methods have been proposed, which including Multiple Signal Classification (MUSIC)[1], Maximum Likelihood (ML)[2], and Weighted Subspace Fitting (WSF)[3]. These methods have been shown to outperform the conventional beamforming method assuming if the observation is generated by the underlying model [4]. Most of these methods are applied to the narrow band system. Various extension to the wideband system were also proposed, for example, the Coherent Signal-Subspace Processing (CSSP) [5] and Wideband ESPRIT [6]. In [7] and [8], a frequency domain algorithm based on ML criterion called AML method has been proposed for source localization and DOA estimation. The AML algorithm has been successfully implemented in a real-time testbed [8].

Conventional DOA estimation methods does not consider the temporal correlations. However, by the assumptions of statistical independence among sources and sources being either non-Gaussian or color, correlation matrix at different time-delays or fourth order cumulant matrix can be exploited to develop various Blind Source Separation (BSS) algorithms [9][10][11]. In our work, instead of focusing on signal separation, we try to utilize this property to improve DOA estimation of acoustic signals. By the assumption of the steering matrix is a diagonal matrix, we develop a modified form of Joint Approximate Diagonalization method [12] for DOA estimation of acoustic signals by using the temporal structure of the signal. Simulation results show that the proposed method can outperform AML method for DOA estimation of acoustic signals.

This paper is organized as follows. In section 2, we describe the data model and assumptions. We show that the statistical independence of the sources in the time domain still holds in the time-frequency domain. In section 3, we present our MJAD method for DOA estimation of wideband color signals. We carry out some simulation study in section 4 to evaluate the performance of the MJAD method. Finally, we draw our conclusions.

2. DATA MODEL AND ASSUMPTION

In our data model, the acoustic signals are assumed to be wideband and lie in the far field of the sensor array. In addition, we make the same assumption as BSS which state as the following.

A1) Sources are spatially uncorrelated with different autocorrelation functions but are temporally correlated (color) stochastic signals with zero mean.

A2) Sources are stationary signals or their variances maybe time varying.

A3) Noises are white Gaussian.

The arriving wavefront to the array is assumed to be planer and only the angle of arrival can be estimated. For simplicity, we assume both the source and sensor array lie in the same plane (a 2-D scenario). Let there be $M$ wideband sources, each at an angle $\theta_m$ from the array. The angle convention is such that north is 0° and east is 90°. The sensor array consists of $N$ sensors, each at position $r_r = [x_r, y_r]^T$. The sensors are assumed to be omni-directional and have identical response. The array centroid position is given by $r_c = \frac{1}{N} \sum_{r=1}^{N} r_r = [x_c, y_c]^T$. We use the array centroid as the reference point and define a signal model based on the relative time-delays from this position. The relative time-
delay of the \( m \)-th source is given by
\[
t_{cr}^{(m)} = t_{cr}^{(m)} - t_{cr}^{(r)} = [(x_c - x_r) \sin \theta_m + (y_c - y_r) \cos \theta_m] / v,
\]
where \( t_{cr}^{(m)} \) and \( t_{cr}^{(r)} \) are the absolute time-delays from the \( m \)-th source to the centroid and the \( r \)-th sensor, respectively, and \( v \) is the speed of propagation in length unit per sample. The data received by the \( r \)-th sensor at time \( n \) is then:
\[
x_r(n) = \sum_{m=1}^{M} s_r^{(m)}(n - t_{cr}^{(m)}) + w_r(n),
\]
for \( n = 0, \ldots, N - 1 \), \( r = 1, \ldots, R \), and \( m = 1, \ldots, M \), where \( S_r^{(m)} \) is the \( m \)-th source signal arriving at the array centroid position, \( t_{cr}^{(m)} \) is allowed to be any real-valued number, and \( w_r \) is the zero mean white Gaussian noise with variance \( \sigma^2 \).

For the ease of derivation and analysis, the received wideband signal can be transformed into the frequency domain via the FFT, where a narrowband model can be given for each frequency bin. For \( N \)-point FFT transformation, the array data model in the frequency domain is given by
\[
X(k) = A(k)S_r(k) + \eta(k),
\]
for \( k = 0, \ldots, N - 1 \), where the received data spectrum is
\[
X(k) = [X_1(k), \ldots, X_R(k)]^T,
\]
the steering matrix \( A(k) = [d^{(1)}(k), \ldots, d^{(M)}(k)] \), the steering vector is given by
\[
d^{(m)}(k) = [d_1^{(m)}(k), \ldots, d_R^{(m)}(k)]^T, d_r^{(m)} = e^{-j2\pi k t_{cr}^{(m)}} / N,
\]
and the source spectrum is given by
\[
S_r(k) = [S_1^{(1)}(k), \ldots, S_U^{(m)}(k)]^T.
\]
The noise spectrum vector \( \eta(k) \) is zero mean complex white Gaussian distributed with variance \( N \sigma^2 \). It can be shown that Assumptions A1) to A3) still hold in the time-frequency domain. Note, due to the transformation to the frequency domain, \( \eta(k) \) asymptotically approaches a Gaussian distribution by the central limit theorem even if the actual time-domain noise has an arbitrary i.i.d. distribution (with bounded variance) which is possibly even non-Gaussian. This asymptotic property in the frequency domain provides a more reliable noise model than the time-domain model in some practical cases. Since the FFT of the source signals is a linear combination of the time domain signal, it is easy to verify that the A2) is still valid for the time-frequency domain data. The key assumption of BSS is A1), and we give the proof that it is still valid in the time-frequency domain.

Let us assume there are two sources \( S_1 \) and \( S_2 \) and considering at time frame \( l \) and angular frequency \( \omega_k \)
\[
E[S_1^*(\omega_k, l) S_1(\omega_k, l)] = E[\sum_{i=0}^{N-1} s_1(i + (l - 1)N)]
\]
\[
e^{-j\omega_k(i+(l-1)N)} E[\sum_{i=0}^{N-1} s_1(i + (l - 1)N)e^{-j\omega_k(i+(l-1)N)}]
\]
\[
= N r_1(0) + \sum_{i=1}^{N-1} 2(N - i) \cos \omega_k r_1(i)
\]
\[
E[S_1^*(\omega_k, l) S_2(\omega_k, l)] = E[\sum_{i=0}^{N-1} s_1(i + (l - 1)N)]
\]
\[
e^{-j\omega_k(i+(l-1)N)} E[\sum_{i=0}^{N-1} s_2(i + (l - 1)N)e^{-j\omega_k(i+(l-1)N)}]
\]
\[
= N r_2(0) + \sum_{i=1}^{N-1} 2(N - i) \cos \omega_k r_2(i) = 0
\]
Thus, \( E[S(\omega_k, l) S^*(\omega_k, l)] = R_{S_{\omega_k}}(0) \) is a diagonal matrix. Similarly, \( E[S_1^*(\omega_k, l) S_2(\omega_k, l + p)] \) is a linear combination of \( r_1(p), \ldots, r_1(p + N - 1) \), \( E[S_1^*(\omega_k, l) S_2(\omega_k, l + p)] = 0 \), and therefore \( E[S(\omega_k, l) S(\omega_k, l+p)] = R_{S_{\omega_k}}(p) \) is a diagonal matrix.

From (2) and Assumption A1) to A3) in the time-frequency domain, it can be shown that the correlation matrix of the received sensor data in the time-frequency domain satisfies the following equations.
\[
R_{X_{\omega_k}}(0) = E[X_{\omega_k}(l) X^*_{\omega_k}(l)]
\]
\[
= A(\omega_k) R_{S_{\omega_k}}(0) A(\omega_k)^* + R_{\eta}(0),
\]
\[
R_{X_{\omega_k}}(p) = E[X_{\omega_k}(l) X^*_{\omega_k}(l + p)]
\]
\[
= A(\omega_k) R_{S_{\omega_k}}(p) A(\omega_k)^*,
\]
where \( R_{S_{\omega_k}}(0) \) and \( R_{S_{\omega_k}}(p) \) are the correlation matrices of the sources at zero time delay and time delay \( p \) respectively. These matrices are diagonal by assumption A1). Thus, in addition to the correlation matrix at zero time delay, the temporal structure of the speech signals results in different correlation matrix corresponding to different time delays. In the following, we try to exploit these additional information to improve the DOA estimation using an MJAD method.

3. MODIFIED JOINT APPROXIMATE DIAGNOLIZATION METHOD FOR DOA ESTIMATION OF WIDEBAND COLOR SIGNALS

In this section, we describe a JAD method for DOA estimation of speech signals in the time-frequency domain. The main idea of this method lies in the joint processing of correlation matrices with different time delays. From (3) and (4), each correlation matrix of the received data can be decomposed into the steering matrix and a diagonal matrix. The DOA estimation can then be obtain by simultaneously diagonalizing different \( R_{X_{\omega_k}}(p) \)s. Our new JAD method adopt the Joint Approximative Diagonalization (JAD) method based on the least-squares (LS) criterion in [12]. However, we restrict the diagonalizing matrix to be a steering matrix instead of an orthogonal matrix, which allow us to estimate the source DOAs. From the JAD approach based on LS criterion, the penalty function that we
try to minimize can be formulated as

\[
J(\mathbf{A}(\omega_k), \mathbf{R}_{\mathbf{S}(\omega_k)}(p)) = \sum_{k=1}^{K} \sum_{p=0}^{P} \| \mathbf{R}_{\mathbf{X}(\omega_k)}(p) - \mathbf{A}(\omega_k) \mathbf{R}_{\mathbf{S}(\omega_k)}(p) \mathbf{A}(\omega_k)^H \|^2.
\]

By the assumption of each \( \mathbf{R}_{\mathbf{S}(\omega_k)}(p) \)’s are independent of each other, the above minimization can be obtain by individual minimization of each term inside the summation. Using the technique of separating variables, we first fix the term \( \mathbf{A}(\omega_k) \), and minimize the following function with respect to \( \mathbf{R}_{\mathbf{S}(\omega_k)}(p) \), thus

\[
\tilde{J}(\mathbf{R}_{\mathbf{S}(\omega_k)}(p)) = \| \text{vec}(\mathbf{R}_{\mathbf{X}(\omega_k)}(p)) - \mathbf{U} \mathbf{d} \|^2,
\]

where \( \mathbf{a}_m \) is the \( m \) column of the steering matrix \( \mathbf{A} \), and \( \lambda_1,...,\lambda_M \) are diagonal elements of \( \mathbf{R}_{\mathbf{S}(\omega_k)}(p) \). The least-squares solution of the above minimization can be found by the matrix vectorizing, that is

\[
\tilde{J}(\mathbf{R}_{\mathbf{S}(\omega_k)}(p)) = \| \text{vec}(\mathbf{R}_{\mathbf{X}(\omega_k)}(p)) - \mathbf{U} \mathbf{d} \|^2,
\]

where \( \mathbf{U} \) is the \( R^2 \times M \) matrix

\[
\mathbf{U} = [\text{vec}(\mathbf{a}_1\mathbf{a}_1^H), ..., \text{vec}(\mathbf{a}_M\mathbf{a}_M^H)],
\]

and \( \mathbf{d} \) is the \( M \times 1 \) vector

\[
\mathbf{d} = [\lambda_1, ..., \lambda_M]^T.
\]

The least-squares solution of vector \( \mathbf{d} \) is then

\[
\mathbf{d} = (\mathbf{U}^H \mathbf{U})^{-1} \mathbf{U}^H \text{vec}(\mathbf{R}_{\mathbf{S}}).
\]

Substitute the above equation into (7) and then into (5), the estimation criterion with respect to the estimated angle is

\[
\hat{\Theta} = \arg\min_{\Theta} \sum_{k=1}^{K} \sum_{p=0}^{P} \| \mathbf{I} - \mathbf{P}_U(\omega_k) \text{vec}(\mathbf{R}_{\mathbf{X}(\omega_k)}(p)) \|^2,
\]

where \( \mathbf{P}_U = \mathbf{U}(\mathbf{U}^H \mathbf{U})^{-1} \mathbf{U}^H \mathbf{P} \) is the projection matrix of \( \mathbf{U} \). The resulting function is highly non-linear in terms of the searching angles \( \Theta \) which require \( M \times D \) exhaustive search. We apply the alternating projection (AP) technique proposed by [13] which reduces the \( M \times D \) search to a sequence of 1-D search. The outline of the MJAD algorithm is given as follows

1) Setting parameters: the number of estimate source \( M \), the sensor array geometry, the number of selective frequency bins \( K \), the FFT size \( N_t \),

2) FFT the received signal at each time frame

3) Estimate the correlation matrix for a set of pre-selected time delays \( \mathbf{R}_{\mathbf{X}(\omega_k)}(p) \) for selected frequency bins \( \omega_k, k = 1, ..., K \)

4) Using the AP method to search the angle which maximize the function at (11).

Although the MJAD algorithm is applied mainly for DOA estimation of wideband signals in the time-frequency domain here, it is easy to apply this method to narrow band communication systems[14]. In wireless communication systems, the signals are considered to be non-Gaussian and the matrix to be diagonalized is the fourth-order cumulant matrices.

4. SIMULATION RESULTS

In our simulation, we consider two independent speech signals provided by [9]. The source signals are considered to be far-field sources which are located at 30° and 45° from the sensor array respectively. The sensor array is a uniform linear array (ULA) with 4 sensors and intersensor displacement is 10.8 cm. We select three different FFT size (1024, 2048 and 4096) for each time frame and the received signals are divided into multiple time frames. We select 60 frequency bins corresponding to the maximum magnitude to perform DOA estimation in the time-frequency domain. For MJAD algorithm, we select 0, 1, and 2 time delays in digital domain to perform joint diagonalization. For CSSP algorithm, we set the initial angle at 1 degree offset from the true angle, and select the focusing frequency at the frequency bin corresponding to the largest receive signal magnitude. We vary the SNR range from -10dB to 20dB and calculate the root mean square (rms) error via 100 simulation runs at each SNR. The rms error of MJAD, AML and CSSP is plotted at Fig. 1. It can be seen that the MJAD algorithm outperforms the AML and CSSP algorithm in all SNR range. The threshold SNR for good angle estimation is improved by almost 10dB. The second simulation changes the array geometry to a uniform circular array (UCA) with 4 sensors, the other parameters such as source arrival angles, intersensor displacement, and FFT size remain the same. The CSSP can not resolve two sources for this array geometry. Fig. 2 shows the rms performance of the AML and MJAD. Both AML and MJAD show degraded performance when comparing to the ULA case since the aperture of UCA is less than ULA. However, MJAD still performs better than AML at the low SNR range while the performance are comparable at the high SNR range.
5. CONCLUSION

We have presented an MJAD algorithm to estimate the DOAs of acoustic signals. Our simulation results show that the performance of this method outperforms the AML method which has been shown in [7]-[8] to be quite effective for wideband source DOA estimation and localization. It remains to see the theoretical justification of the good performance of the MJAD method. In the future, we are interested in applying this method to real life test data.

6. REFERENCES


