Passive Direction-of-Arrival and Range Estimation for Near-Field Sources

A. L. Swindlehurst and T. Kailath
Information Systems Laboratory
Stanford University
Stanford, CA 94305

Abstract—The majority of model-based techniques proposed to date for solving the multiple source direction-of-arrival (DOA) estimation problem operate under the assumption that the wavefronts impinging on the array are planar. However, when sources are located close to the array (i.e., in the near-field), the inherent curvature of the wavefront is no longer negligible. In this paper, we propose the use of high-resolution signal subspace methods in conjunction with a spatial version of the Wigner-Ville distribution (WVD) to solve the near-field problem. Advantages of this approach are that accurate source range/DOA estimates are obtained with relatively good resolution, and without computation or search of a three-dimensional spectral surface. The effectiveness of the algorithm is demonstrated by extensive simulations, and its performance relative to the Cramér-Rao bound is also presented.

1. Introduction

In recent years, the focus of attention in emitter location research has been on model-based estimation of the directions-of-arrival (DOAs) of multiple plane-wave signals received by an array of sensors. Virtually all of the techniques developed to date rely heavily on the plane-wave assumption; i.e., that the signal sources are located in the far-field of the antenna array aperture, and hence that the signal wavefronts are spatially described by a single location parameter (e.g., azimuth). However, when sources are located close to the array (i.e., in the near-field), the inherent curvature of the wavefront is no longer negligible. Since the wavefront in such cases is a spatial function of two independent location parameters (e.g., range and azimuth), algorithms based on the plane-wave assumption are no longer directly applicable and alternatives must be found.

Traditional approaches to the near-field problem are based on either triangulation, simultaneous beamforming over range and azimuth, or time-delay estimation (cf. [1, 2]). To improve the resolution performance of such methods, parametric range/DOA estimators have been proposed [3, 4], though they suffer from the high computational load required in creating and searching a two-dimensional range/bearing spectrum. The desire to exploit the computational advantages of the recently developed total-least-squares (TLS)-ESPRIT algorithm [5, 6], together with recent discussions involving Dr. H. Whitehouse, has led us to consider a different approach, one based on a spatial analog to the Wigner-Ville distribution (WVD). The motivation behind this approach is that for sources located in the Fresnel region of the antenna array aperture, the combined range/DOA estimation problem is similar in form to that of estimating the parameters of chirp signals in noise. This is a straightforward generalisation of the isomorphism between sinusoidal frequency estimation and DOA estimation problems. In this paper, the concept of a spatial WVD is described and used to develop an algorithm that provides estimates of the range and azimuth of multiple sources located in either the near- or far-field. The algorithm is currently limited in applicability to narrowband uncorrelated sources and a uniform linear array of sensors.

We gratefully acknowledge the assistance of Dr. J. Speiser in bringing to our attention a recent paper by Breed and Posch [7] which describes a spatial WVD method for combined range/azimuth estimation. Our algorithm differs from that of [7] in that signal subspace methods are used in place of a discrete Fourier transform to generate the desired estimates. This approach offers the following advantages: the source locations are estimated more accurately, closely-spaced sources are more easily resolved, the algorithm is more easily extended to the multiple-source situation, and no computation or search of a spectral surface is required. In the sections that follow, the signal model and algorithm steps are explained in detail, and the results of simulations are presented for a variety of scenarios. Comparisons are also made with the Cramér-Rao bound for the single-source case.

2. Signal Modeling Assumptions

Consider the scenario depicted in Figure 1, where the signals from d sources located at range/bearing coordinates $(R_i, \theta_i)$ relative to some reference point impinge upon an array of m sensors. The array is assumed to be uniform and linear, with unity-gain omnidirectional sensors and an interelement spacing of $\delta$. The signals are assumed to be
uncorrelated (i.e., the signal covariance matrix is diagonal) and narrowband. As shown in the figure, we will define the array aperture $A$ to be $A = (m-1)d$. If the reference point is chosen to be at one of the sensors on the end of the array, the distance from the $i$th source to the $k$th sensor is given by a simple application of the law of cosines:

$$r_{ik} = \sqrt{R_i^2 + k^2d^2 - 2kdR_i \sin \theta_i}.$$  

In this notation, $R_i \equiv R_{1i}$. Using the narrowband assumption, the signal at the $k$th sensor, $z_k(t)$, may thus be written as

$$z_k(t) = \sum_{i=1}^{d} a_i(t) e^{j\omega t} e^{j(2\pi/\lambda)(R_{ik}-R_i)} + n_k(t),$$  

where $a_i(t)$ is the complex envelope of the $i$th signal and $n_k(t)$ is the additive noise present at sensor $k$. It will be assumed throughout that the noise at each sensor has equal but unknown variance $\sigma^2$, and that it is independent from sensor to sensor. This assumption can be relaxed in cases where the noise spatial covariance matrix is known to within a scale factor.

For sources in the far-field, $R_i \gg \Delta$, and $r_{ik}$ is adequately approximated by taking only the first two terms of the binomial expansion of (1):

$$r_{ik} \approx R_i - k\delta \sin \theta_i.$$  

Equation (3) treats the spherical wavefronts as planar wavefronts as they propagate across the array, and hence it is referred to as the plane-wave approximation. When this approximation is substituted into Equation (2), the phase of the delay term becomes a linear function of $k$, the sensor index, and the signal part of $z_k(t)$ is seen to be a sum of sinusoids in the variable $k$. This is, of course, the well-known connection between the time-series frequency estimation problem and DOA estimation with a uniform linear array (ULA).

In many applications, the distances from the array to the emitters are on the order of only a few array apertures, and the approximation of (3) is no longer valid. Retaining an additional term from the binomial expansion of (1) leads to an expression that is quadratic in $k$:

$$\tilde{r}_{ik} = R_i - k\delta \sin \theta_i + k^2(\delta \cos \theta_i)^2.$$

Equation (4) is sometimes referred to as the Fresnel approximation. The significance of this approximation is that in the vicinity of the array the spherical wavefronts are modeled as quadratic, rather than planar, surfaces. The error introduced by use of (4) is illustrated in Figure 2. The contours of constant % error shown on the plot correspond to the error in measuring $R_{im}$ using (4) relative to the array aperture $\Delta$ (i.e., $100|\tilde{R}_{im} - R_{im}|/\Delta$). We will define the Fresnel region of the array to be roughly that region beyond the 5% contour.

As a consequence of Equation (4), sources in the Fresnel region of the antenna array are modeled as producing a quadratic, rather than linear, phase variation as the wave propagates across the array:

$$z_k(t) \approx \sum_{i=1}^{d} a_i(t) e^{j\omega t} e^{-j(2\pi/\lambda)(\alpha_i \delta - \beta_i k^2/2)} + n_k(t),$$

where

$$\alpha_i = \delta \sin \theta_i, \quad \beta_i = (\delta \cos \theta_i)^2 / R_i.$$

Instead of a sum of sinusoids in $k$, $z_k(t)$ is now modeled as a sum of chirp or linear-frequency-modulated (LFM) signals in $k$, parameterized by the 2d source location parameters $(R_i, \theta_i), 1 \leq i \leq d$. The problem of estimating these parameters is thus similar in form to the time-series problem of chirp signal parameter estimation. This connection does not seem to have been exploited in the research literature and used to develop algorithms for the near-field problem that parallel those already developed for its time-series counterpart. In the following section, this space/time duality is used to develop an algorithm for the near-field problem based on a spatial analog to the Wigner-Ville distribution (see also [7]).
3. The Spatial Wigner-Ville Algorithm

The Wigner distribution was first described in 1932 by Wigner [8] as a space-momentum representation for use in the field of quantum mechanics. Ville reintroduced the concept in 1948 [3] in the context of signal analysis, and it is typically referred to today as the Wigner-Ville distribution (WVD). The WVD has been the subject of a great deal of interest in the signal processing community over the past several years, primarily in the area of time-varying signal analysis and synthesis. In the time-series context, the WVD can be thought of as a representation of how signal energy is distributed jointly over time and frequency. This interpretation is not strictly true since the WVD cannot be guaranteed to always be positive, but the WVD possesses a large number of properties that one would expect from a "true" time-varying spectrum. An excellent review of these properties can be found in the series of papers by Clenon [9].

For a discrete stochastic signal \( z(t) \), the WVD is defined as follows [10]:

\[
W_z(t, \omega) = \sum_{r=-\infty}^{\infty} K_z(t, r)e^{-j2\pi r},
\]

where the Wigner-Ville kernel \( K_z(t, r) \) is given by

\[
K_z(t, r) = E\{z(t + r)z^*(t - r)\},
\]

In the above equation, \( * \) denotes complex conjugation and \( E\{\cdot\} \) denotes expectation. When viewed as a function of \( r \), the WVD kernel \( K_z(t, r) \) is made up of the even-numbered lags of the signal's autocorrelation at time \( t \). It is a simple matter to incorporate the odd autocorrelation lags into \( W_z(t, \omega) \) as well, though to simplify the derivation of the algorithm we will postpone doing so until section 4.1. The WVD is seen to be the frequency scaled Fourier transform of the even autocorrelation lags of \( z(t) \). The frequency scaling by a factor of 2 is necessary to ensure that frequency components of \( z(t) \) at \( \omega \) occur in \( W_z(t, \omega) \) at \( \omega \) as well. One consequence of defining the WVD in this fashion is that \( W_z(t, \omega) \) is periodic in \( \omega \) with period \( \pi \) instead of \( 2\pi \). Thus, to avoid aliasing, \( z(t) \) must either be oversampled (at, say, twice the Nyquist frequency) or the frequency support of \( z(t) \) must be entirely contained in \( 2\pi \) periodic intervals of length \( \pi \) (e.g., \( z(t) \) analytic). This issue will be raised in later discussions of the spatial WVD.

It is well known that the WVD is particularly useful in applications involving LFM, or chirp, signals. To see why this is so, consider the following (deterministic) discrete-time chirp signal:

\[
z(t) = A e^{j\pi t^2/2}.
\]

The WVD of this signal is

\[
W_z(t, \omega) = \pi|A|^2 \sum_{k=-\infty}^{\infty} \delta(\omega - \beta t - k\pi),
\]

where \( \delta(\cdot) \) represents the Dirac delta function. This is an intuitively appealing result since it shows that for any time \( t \), the WVD of a chirp signal is highly concentrated around the signal's instantaneous frequency. In other words, for fixed \( t \), the WVD kernel \( K_z(t, r) \) for a chirp signal is sinusoidal in \( r \) with frequency \( \beta t \). The motivation for using a spatial version of the WVD to solve the near-field problem stems primarily from this ability of the WVD to convert the chirp parameter estimation problem to one of sinusoidal frequency estimation 2.

With this idea in mind, consider the signal of Equation (5) and switch the time and space indices by rewriting \( z(t) \) as \( z(k) \), thus fixing time and viewing the signal as a function of the spatial index \( k \). Next, define a spatial WVD kernel as follows:

\[
K_z(k, r) = E\{z_k(k + r)z_k^*(k - r)\},
\]

where \( r \) is now a sensor index. Substitution of (5) into (10) leads to the following expression for \( K_z(k, r) \):

\[
K_z(k, r) = \sum_{i=1}^{d} \sigma_i^2 e^{j2\pi \beta_i(k - r)} + \sigma^2 \delta_r,
\]

where \( \alpha_i \) and \( \beta_i \) are as defined in (5), \( \sigma_i^2 \) is the power of the \( i^{th} \) source, \( \sigma^2 \) is the noise power, and \( \delta_r \) is the Kronecker delta function. In deriving (11), use is made of the fact that the sources are completely uncorrelated. If this assumption were not made, cross-product terms would appear in the result.

As in the time-series case, the signal part of \( K_z(k, r) \) for each \( k \) is sinusoidal in \( r \). Furthermore, the sinusoidal frequencies are linear functions of \( k \), with slope and intercept given by non-linear functions of the signal location parameters \( (R_i, \theta_i) \), \( 1 \leq i \leq d \). This fact suggests that the following high-level algorithm could be used to estimate the source locations:

1. Compute the WV kernels for a central subset of the sensors in the array.
2. Obtain an estimate of the number of signals \( d \) and the noise variance \( \sigma^2 \).
3. Estimate \( d \) frequencies from each WV kernel.
4. Fit \( d \) lines to the frequency data and obtain range and bearing estimates as follows:

\[
\hat{\theta}_i = \sin^{-1}\left(\frac{\lambda_i}{4\pi R_i^2}\right), \quad \hat{R}_i = -\frac{4\pi(\delta \cos \theta_i)^2}{\lambda_i},
\]

where \( \alpha_i \) and \( \beta_i \) are respectively the slope and intercept of the \( i^{th} \) line.

In [7], step 2 was not an issue since only the noise-free, single source case was considered, and step 3 was solved using the 2The "rank-reducing" effect of the non-linear transformation from \( z \) to \( K_z \) is precisely that noted by Kumaresan and Verma in [11], though no connection with the WVD was made.
peaks of the Fourier transform for each WV kernel. In the following section, each step is considered in detail, and the following improvements to the algorithm described in [7] are proposed:

- signal-subspace ideas are used to extend application of the algorithm to cases where both noise and multiple signals are present.
- the total-least-squares (TLS)-ESPRIT [5, 6] algorithm is used in place of a Fourier transform in order to increase the accuracy and resolution of the resulting parameter estimates.

4. Algorithm Details

4.1 Computing the WV kernels

Since our array contains a finite number of sensors and since we will later require kernels of length at least greater than \( d + 1 \), the functions \( K_d(k, r) \) will only be computed for some subset of sensors near the center of the array. Examination of Equation (11) reveals that \( K_d(k, r) \) is not a function of time (due to the assumption that the noise is stationary over the data collection interval), and hence sample averages can be employed to approximate the expectation operation. In fact, \( K_d(k, r) \) is nothing more than the anti-diagonal containing the \( k^{th} \) diagonal element of the array covariance matrix. Thus, computation of the spatial WV kernels is embedded in the formation of the sample covariance. The desire to use all the anti-diagonals of the sample covariance (i.e., even those that do not possess an element on the main diagonal) leads to an expanded symmetric definition of the discrete WVD incorporating both the even and odd lags of the autocorrelation kernel:

\[
W_k(k, \omega) = \begin{cases} 
\sum_{r \text{ even}} K_d(k, r) e^{-j\omega r}, & \text{integer } k \\
\sum_{r \text{ odd}} K_d(k, r) e^{-j\omega r}, & \text{half-integer } k 
\end{cases}
\]

(13)

where by half-integer \( k \) we mean \( n/2 \) for some integer \( n \). The definition given for integer \( k \) corresponds to the definition of Equation (6) through a simple change of the summation index. The half-integer terms are "intersensor" WV kernels; i.e., instead of being associated with a specific sensor, they correspond to a point between two sensors. Note that the expanded set of WV kernels is also sinusoidal in \( r \) with linearly increasing frequency in \( k \). The introduction of the half-integer terms effectively doubles the amount of information used in forming the parameter estimates. This is especially critical for the case of the spatial WVD presented here since typically the number of sensors in an array is relatively small.

The previous discussion on aliasing for the time-series WVD applies as well in the computation of the spatial WV kernels. To avoid spatial aliasing, \( z_2(t) \) must either be spatially oversampled or the sources must lie within a restricted region in space ("spatial bandlimiting"). In the absence of a priori knowledge about source locations, it is sufficient to set \( \delta = \lambda/4 \).

4.2 Estimating the number of signals

Steps 3 and 4 of the above algorithm require that the number of signals \( d \) and the noise variance \( \sigma^2 \) be known, though typically this is not the case in practice. The problem of estimating these parameters for the situation considered here is the same as the regular far-field DOA detection problem. The covariance matrix in each case has the following form:

\[
R = SSA^* + \sigma^2 I \]

where \( S \) is the signal covariance matrix and where the \( i^{th} \) column of \( A \) is given by the steering vector \( a_i \):

\[
a_i = \begin{bmatrix} 1 e^{j(2\pi \lambda/\lambda)(R_{11} - R_i)} & \ldots & e^{j(2\pi \lambda/\lambda)(R_{m1} - R_i)} \end{bmatrix}^T.
\]

Estimation of \( d \) is usually accomplished by exploiting the underlying multiplicity of the minimal eigenvalue of \( R \). Since only an estimate of \( R \) is available in practice, reliable statistical procedures have been developed to estimate the eigenvalue multiplicity (cf [12, 13]). Therefore, in the simulations of the next section, \( d \) is assumed to have been correctly determined. Performing an eigendecomposition of \( R \) in order to estimate \( d \) also allows one to use the following simple estimator for \( \sigma^2 \):

\[
\hat{\sigma}^2 = \frac{1}{m-d} \sum_{i=1}^{m-d} \lambda_i,
\]

where \( \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_m \) are the eigenvalues of the sample covariance matrix.

4.3 Estimating the kernel frequencies

The finite length of the array aperture leads to a rectangular windowing of the WV kernels, and hence for a given \( k \), \( W_k(k, r) \) is defined only for relatively few values of \( r \). As such, there is little hope of accurately estimating or resolving the frequencies of the underlying sinusoids using Fourier transform based methods. Parametric (specifically autoregressive) modeling of the WV has previously been proposed [14, 15, 16] and in [14, 15] applied to the chirp signal estimation problem. The use of eigenstructure based methods has also been proposed in [14, 17], though simulation results using these methods were not given.

We will also take a parametric approach in estimating the kernel frequencies, using the recently developed total-least-squares (TLS)-ESPRIT algorithm [5, 6]. There are two reasons for choosing this algorithm. First, it directly estimates the sinusoidal frequencies, and thus eliminates the need to search for peaks in some spectral measure. Second, simulations have shown its performance to be superior in terms of both accuracy and robustness when compared with other high-resolution spectral estimators [5]. The displacement invariance required by ESPRIT is satisfied in this application since the samples of the WV kernels are uniformly spaced in \( r \). As alluded to earlier, the noise variance term
of Equation (11) must be subtracted from the \( \tau = 0 \) sample of the even-lag WV kernels before ESPRIT may be applied.

4.4 Determining source locations

ESPRIT provides frequency estimates for each WV kernel. The source location estimates are formed from the slope and intercept of lines fit to the frequency estimates as functions of \( k \), the sensor index. It is thus necessary to associate the individual estimates with one of \( d \) (or more) lines. If the sources are assumed to be located such that the frequency tracks in \( k \) do not cross, this association can be trivially accomplished by sorting the estimates for each kernel; however, more complicated clustering procedures are required both to detect and to separate overlapping lines [18]. In the application considered here, this difficulty arises only when two sources, one near the array and one in the far-field, possess nearly coincident DOAs.

5. Simulation Results

For all the simulations presented in this section, a 16-element ULA with interelement spacing \( \delta = \lambda/4 \) was assumed. In each case nine WV kernels were computed, corresponding to sensor and inter-sensor locations between the sixth and eleventh array elements (i.e., \( k \in \{6.5, 7, \ldots, 10.5\} \)). Tables 1 and 2 present results for the single and multiple source scenarios, respectively. In these tables, \( R \) is the range in wavelengths, \( \theta \) the DOA in degrees, SNR the signal-to-noise ratio in dB, and \( N \) the number of snapshots used in forming the WV kernels. For each case, 100 independent trials were conducted, and the estimates shown represent the corresponding sample mean and standard deviation.

The performance of the algorithm is particularly impressive for cases involving only one source. When multiple sources are present, reasonable results are obtained only for relatively large values of \( N \). This is a result of the fact that the sources are, strictly speaking, only asymptotically uncorrelated; when the WV kernels are computed from time intervals of finite length, the cross-product terms have not completely decorrelated and are present in the result. It is interesting, however, to observe that even in situations where the range estimate is degraded due to the cross-product phenomenon or low SNR, the DOA estimate retains a high degree of accuracy.

Further simulations have shown that the ranging error of Figure 2 may be as large as 5% before the performance of the algorithm is seriously degraded. This was the reason for the definition of the Fresnel region as given in Section 2. Thus, according to Figure 2, the algorithm may be applied to estimate source locations beyond about three array apertures. If the sources are located near broadside, the algorithm is applicable at even closer ranges.

In Figures 3 and 4, the performance of the algorithm relative to the Cramér-Rao bound (CRB) is presented for the case \( R = 75 \lambda, \theta = 10^\circ \), and \( N = 250 \). The DOA estimate is seen to be near this performance limit in all cases considered, while the range estimate appears to have a relative efficiency of about three or four compared to the CRB.

6. Concluding Remarks

An algorithm for simultaneously estimating the range and bearing of multiple near-field sources has been presented. The method is based on the application of signal-subspace ideas to the spatial Wigner-Ville distribution approach originally presented in [7]. The principal advantages of using signal-subspace methods are that the range/DOA estimates are obtained with improved precision and resolution, and without computation or search of a complicated spectral surface. Additionally, these methods allow a simple and more effective extension of the spatial Wigner-Ville approach to cases in which noise and/or multiple signals are present. Simulations have shown that the algorithm performs well for a wide variety of test cases, and comparisons with the Cramér-Rao bound indicate near optimal DOA estimates.

References


---

Table 1: Results for Single Source Cases

<table>
<thead>
<tr>
<th>Actual</th>
<th>Estimated</th>
</tr>
</thead>
<tbody>
<tr>
<td>( R )</td>
<td>( \theta )</td>
</tr>
<tr>
<td>80</td>
<td>0</td>
</tr>
<tr>
<td>80</td>
<td>-30</td>
</tr>
<tr>
<td>80</td>
<td>60</td>
</tr>
<tr>
<td>80</td>
<td>24</td>
</tr>
<tr>
<td>80</td>
<td>24</td>
</tr>
<tr>
<td>80</td>
<td>24</td>
</tr>
<tr>
<td>80</td>
<td>24</td>
</tr>
<tr>
<td>100</td>
<td>24</td>
</tr>
<tr>
<td>60</td>
<td>24</td>
</tr>
<tr>
<td>20</td>
<td>24</td>
</tr>
</tbody>
</table>

Table 2: Results for Multiple Sources

<table>
<thead>
<tr>
<th>Actual</th>
<th>Estimated</th>
</tr>
</thead>
<tbody>
<tr>
<td>( R )</td>
<td>( \theta )</td>
</tr>
<tr>
<td>80</td>
<td>10</td>
</tr>
<tr>
<td>100</td>
<td>25</td>
</tr>
<tr>
<td>80</td>
<td>10</td>
</tr>
<tr>
<td>100</td>
<td>25</td>
</tr>
<tr>
<td>80</td>
<td>10</td>
</tr>
<tr>
<td>100</td>
<td>25</td>
</tr>
<tr>
<td>75</td>
<td>10</td>
</tr>
<tr>
<td>63</td>
<td>14</td>
</tr>
<tr>
<td>44</td>
<td>-25</td>
</tr>
<tr>
<td>65</td>
<td>40</td>
</tr>
<tr>
<td>80</td>
<td>0</td>
</tr>
</tbody>
</table>
Figure 3: Performance vs. Cramèr-Rao Bound: DOA Estimate

Figure 4: Performance vs. Cramèr-Rao Bound: Range Estimate


