Estimation of Multipath Parameters in Wireless Communications

Michaela C. Vanderveen[1], Alle-Jan van der Veen[2], and Arogyaswami Paulraj[3]

[1] Stanford University, Scientific Computing Program, Stanford, CA 94305
[2] Delft Univ. of Technology, Dept. Elec. Eng./DIMES, 2628 CD Delft, The Netherlands
[3] Stanford University, Information Systems Laboratory, Stanford, CA 94305

Abstract

In a parametric multipath propagation model, a source is received by an antenna array via a number of rays, each described by an arrival angle, a delay and a fading parameter. Unlike the fading, the angles and delays are stationary over long time intervals. This fact is exploited in a new subspace-based high-resolution method for simultaneous estimation of the angle/delay parameters from *multiple* estimates of the channel impulse response. A computationally expensive optimization search can be avoided by using an ESPRIT-like algorithm. Finally, we investigate certain resolution issues that take the fact that the source is bandlimited into account.

1. Introduction

Source localization is an issue of interest in wireless communications. A major motivation is personal safety, such as in the emergency localization (E-911 service), which will be a system requirement for wireless operators in the near future [2]. Some other applications are accident reporting, automatic billing, fraud detection, cargo tracking, and intelligent transportation systems. Methods for wireless position localization are based on direction of arrival (DOA) and/or time difference of arrival (TDOA) estimation of signals. A second type of application in wireless communication is the estimation of a parametric propagation channel, to assist equalization and directive transmission in the downlink. Estimating propagation parameters from measurements at a phased antenna array also has applications in radar, sonar, and seismic exploration.

This paper focuses on the *joint* estimation of angles and relative delays of multipath propagation signals emanating from a single source and received by a single antenna array. In comparison to "classical" disjoint techniques which first estimate delays and subsequently the angle corresponding to each delay, joint estimation has an advantage in cases where multiple rays have approximately equal delays (or angles). Parametric joint angle/delay estimation has received increased research interest lately [1, 3–8]. Many of the proposed algorithms are based on maximum-likelihood or multidimensional MUSIC, which is unattractive for online estimation. The approach in [3] is an off-line channel sounding method using a range of unmodulated carriers, the other algorithms are based on knowledge of the source signal and in particular the modulation waveform. The algorithm in [8] is based on a multi-dimensional version of ESPRIT. Only a few of the proposed methods allow for the simultaneous estimation of more paths than the number of antennas available [5, 6, 8].

One aspect of this multipath estimation problem which has received little attention so far is that of stationarity. In mobile communication, arrival angles and time delays are relatively stationary. In contrast, the amplitude and relative phase of each path is highly nonstationary and subject to (Rayleigh) fading. The stationarity of the fading is related to the speed of the mobile: its coherence time is roughly the inverse of the Doppler shift, or given by

$$t_{\mathrm{coh}} = \frac{c}{v f_c},$$

where c is the speed of light, v the speed of the mobile, and f_c the carrier frequency. At 1800 MHz and

[0]Manuscript submitted to IEEE Trans. Signal Processing, June. 1997. Part of this paper appeared as [1]

a mobile speed of 1 m/s (walking speed), the coherence time is 160 ms, at 30 m/s (100 km/h), this is 5.6 ms. A typical TDMA system such as GSM or DCS1800 has a slot length of order 0.6 ms and a spacing between slots belonging to the same source in the order of 5 ms. Thus the fading within a single time slot is stationary; at 30 m/s it is also uncorrelated among slots, while at 1 m/s it is uncorrelated every 30 slots. We also note the stationarity of the angles and delays of arrival paths, even for fast-moving users. Over 40 slots, a mobile moving at 30 m/s changes angular position only by 0.1° with respect to a base station 3 km away.

The method presented herein is an effective way to exploit the stationarity of angles and delays, as well as the independence of fading over many time slots. At each time slots a channel estimate is obtained. Multiple channel estimates can be combined if they have the same angles and delays. In comparison to algorithms that are based only on a single channel estimate, the estimation is (of course) more accurate, but also the number of rays that can be resolved is larger. Two short versions of various parts of this work appeared as [1, 7].

In summary, the major assumptions we make on the multipath scenario are: (*i*) the number of rays is small and discrete (i.e., specular multipath), (*ii*) the signals are narrowband with respect to the array aperture, (*iii*) the antenna array response is unambiguous and of known structure, (*iv*) the modulation waveform is known, (*v*) the radio channel is time-slotted.

The outline of the paper is as follows. Section 2 introduces the data model, section 3 outlines the basic technique and section 4 presents an ESPRIT-based algorithm. Resolution issues are treated in section 5, while section 6 discusses the Cramer-Rao bound. We close with simulations in section 7 and conclusions in section 8.

Notation * denotes matrix complex conjugate transpose, $^{\mathrm{T}}$ the matrix transpose, and $\dagger$ is the matrix pseudo-inverse (Moore-Penrose inverse). $\mathbf{I}_m$ is the $m \times m$ identity matrix, 0_1 is a column vector of zeros. $\otimes$ is the Kronecker product, $\odot$ the element-wise product, $\circ$ is the Khatri-Rao product [9], which is a column-wise Kronecker product:

$$\mathbf{A} \circ \mathbf{B} = [\mathbf{a}_1 \otimes \mathbf{b}_1 \quad \mathbf{a}_2 \otimes \mathbf{b}_2 \quad \cdots].$$

2. Data model

Consider the case of a single user transmitting a digital signal in a specular multipath environment. As discussed in the introduction, we regard the channel to be fading but stationary over short time intervals; the impulse response in the n-th interval is denoted by $\mathbf{h}^{(n)}(t)$. Typically, such an interval would coincide with a single time slot in a TDMA system. At each time interval, we collect data over N symbol periods; in total, there are S time slots.

The received baseband signal at an M-element antenna array at time t in the n-th interval, $\mathbf{x}^{(n)}(t) = [x_1^{(n)}(t) \ \ldots \ x_M^{(n)}(t)]^T$, can be written as the convolution of the transmitted digital sequence $\{s_l^{(n)}\}$ with the channel $\mathbf{h}^{(n)}(t)$,

$$\mathbf{x}^{(n)}(t) = \sum_l s_l^{(n)} \mathbf{h}^{(n)}(t - lT) + \mathbf{n}^{(n)}(t), \tag{1}$$

where T is the symbol period and $\mathbf{n}(t)$ is the additive noise. Let Q be the number of paths in this specular multipath environment. Each path is parameterized by a DOA θ_i, time delay τ_i (measured in symbol periods T), and complex path attenuation (fading) $\beta_i(n)$ which is varying between time slots but not within a symbol period. The channel can thus be modeled as

$$\mathbf{h}^{(n)}(t) = \sum_{i=1}^{Q} \mathbf{a}(\theta_i)\beta_i(n)g(t - \tau_i) \tag{2}$$

where $\mathbf{a}(\theta_i)$ is the array response to a path from direction θ_i, and $g(t)$ is the known modulation pulse shape. It is reasonable to assume that $g(t)$ has finite support, say on $t \in [0, L_g T)$. With $\tau_{max} = \lceil \max_{1 \le i \le Q} \tau_i \rceil$, the channel length is $LT = L_g T + \tau_{max}$, i.e., $\mathbf{h}^{(n)}(t)$ is nonzero for $t \in [0, LT)$.

The antenna outputs are sampled at a rate of P times the symbol rate. If we stack each set of P samples of $\mathbf{x}^{(n)}(\cdot)$ and collect the samples in an $MP \times N$ matrix $\bar{\mathbf{X}}$, then (1) leads to

$$\bar{\mathbf{X}}^{(n)} = \bar{\mathbf{H}}^{(n)}\bar{\mathbf{S}}^{(n)} + \bar{\mathbf{N}}^{(n)}, \quad n = 1, \cdots, S$$

where $\bar{\mathbf{H}}^{(n)}$ is an $MP \times L$ matrix of samples of $\mathbf{h}(\cdot)$ and $\bar{\mathbf{S}}^{(n)}$ is an $L \times N$ Toeplitz matrix of data symbols. If $\bar{\mathbf{S}}^{(n)}$ is known from training, then a least-squares estimate of $\bar{\mathbf{H}}^{(n)}$ can be obtained as $\bar{\mathbf{H}}^{(n)}_{est} = \bar{\mathbf{X}}^{(n)}\bar{\mathbf{S}}^{(n)\dagger}$, provided $N \geq MP$. Alternatively, $\bar{\mathbf{H}}^{(n)}$ can be obtained from a blind channel estimation. The noisy channel estimates are

$$\bar{\mathbf{H}}^{(n)}_{est} = \bar{\mathbf{H}}^{(n)} + \bar{\mathbf{V}}^{(n)}_{est} \tag{3}$$

After the channel matrix $\bar{\mathbf{H}}^{(n)}$ has been estimated, it will be convenient to rearrange it into

$$\mathbf{H}^{(n)} = [\mathbf{h}^{(n)}(0) \quad \mathbf{h}^{(n)}(\tfrac{T}{P}) \ \cdots \ \mathbf{h}^{(n)}((L - \tfrac{1}{P})T)], \qquad (M \times LP)$$

which satisfies the factorization (viz. (2))

$$\mathbf{H}^{(n)} = [\mathbf{a}(\theta_1) \ \cdots \ \mathbf{a}(\theta_Q)] \begin{bmatrix} \beta_1(n) & & 0 \\ & \ddots & \\ 0 & & \beta_Q(n) \end{bmatrix} \begin{bmatrix} \mathbf{g}^T(\tau_1) \\ \vdots \\ \mathbf{g}^T(\tau_Q) \end{bmatrix} =: \mathbf{A}(\boldsymbol{\theta})\ \mathrm{diag}[\boldsymbol{\beta}(n)]\ \mathbf{G}^T(\boldsymbol{\tau}). \tag{4}$$

Here, $\mathbf{g}(\tau_i) = [g(kT - \tau_i)]_{k=0,1/P,\cdots,L-1/P}$ is an LP-dimensional column vector containing the samples of $g(t - \tau_i)$, $\boldsymbol{\theta} = [\theta_1 \ldots \theta_Q]$, and $\boldsymbol{\tau} = [\tau_1 \ldots \tau_Q]$.

3. Joint Angle and Delay Estimation

3.1. Space-Time Manifold

Similar to the array response vector $\mathbf{a}(\theta)$, we can define the *space-time response vector* $\mathbf{u}(\theta, \tau)$ as as

$$\mathbf{u}(\theta, \tau) = \mathbf{g}(\tau) \otimes \mathbf{a}(\theta)\,.$$

This MPL-dimensional vector is the response of the array to a channel with a single path with direction θ and delay τ, and includes the pulse shape function. As θ and τ vary over the space of DOAs and delays, the $MLP \times 1$ vector $\mathbf{u}(\theta, \tau)$ traces a multidimensional *space-time manifold*. The space-time response matrix with Q paths is defined as

$$\mathbf{U}(\boldsymbol{\theta}, \boldsymbol{\tau}) = \mathbf{G}(\boldsymbol{\tau}) \circ \mathbf{A}(\boldsymbol{\theta}) = [\mathbf{u}(\theta_1, \tau_1), \cdots, \mathbf{u}(\theta_Q, \tau_Q)]\,. \tag{5}$$

The space-time manifold is a combination of the array manifold $\mathbf{a}(\theta)$, which is determined by the array geometry, and the delay manifold $\mathbf{g}(\tau)$, which depends on the pulse-shape function and the sampling phase. Since the array manifold $\mathbf{a}(\theta)$ and pulse shape function $\mathbf{g}(\tau)$ are both assumed known functions, this will enable to extract the desired parameters $\boldsymbol{\theta}$ and $\boldsymbol{\tau}$ from knowledge of the column span of $\mathbf{U}(\boldsymbol{\theta}, \boldsymbol{\tau})$.

3.2. Method outline

The key observation to be made at this point is that the space-time matrix $\mathbf{U}(\boldsymbol{\theta}, \boldsymbol{\tau})$ can be assumed time-invariant over the observation interval, since the angle/delay parameters are stationary. Thus, we have

$$\mathrm{vec}(\mathbf{H}^{(n)}) = (\mathbf{G}(\boldsymbol{\tau}) \circ \mathbf{A}(\boldsymbol{\theta}))\boldsymbol{\beta}(n) = \mathbf{U}(\boldsymbol{\theta}, \boldsymbol{\tau})\boldsymbol{\beta}(n). \tag{6}$$

We have used the general relation $\mathrm{vec}(\mathbf{A}\ \mathrm{diag}[\mathbf{b}]\ \mathbf{C}) = (\mathbf{C}^T \circ \mathbf{A})\mathbf{b}$ as it applies to (4).

As a first step we estimate $\mathbf{H}^{(n)}$. Applying the vec operation to (3) and using (5) and (6), we arrive at

$$\mathbf{y}(n) = \mathbf{U}(\boldsymbol{\theta}, \boldsymbol{\tau})\boldsymbol{\beta}(n) + \mathbf{v}(n), \qquad n = 1, \ldots, S,$$

where $\mathbf{y}(n) = \mathrm{vec}(\mathbf{H}^{(n)}_{est})$, and $\mathbf{v}(n) = \mathrm{vec}(\mathbf{V}^{(n)}_{est})$. In matrix form, the above equation is

$$\mathbf{Y} := [\mathbf{y}(1) \ \cdots \ \mathbf{y}(S)] = \mathbf{U}(\boldsymbol{\theta}, \boldsymbol{\tau})\mathbf{B} + \mathbf{V} \tag{7}$$

where $\mathbf{B} = [\boldsymbol{\beta}(1) \ \ldots \ \boldsymbol{\beta}(S)]$, and similarly for $\mathbf{V}$.

The joint angle/delay estimation (JADE) problem is, for given channel estimates $\mathbf{Y}$, to find the angles $\boldsymbol{\theta}$ and delays $\boldsymbol{\tau}$ using the model in (7). As an aside, note the resemblance of the JADE model to the familiar DOA model

$$\mathbf{X} = \mathbf{A}(\boldsymbol{\theta})\mathbf{S} + \mathbf{N} \tag{8}$$

where $\mathbf{X}$ is the array output measurements, $\mathbf{S}$ is the matrix of signals and $\mathbf{N}$ the additive noise. The difference to (7) is (*i*) the "data" are the channel estimates, not the array outputs, (*ii*) the manifold matrix is parametrized by both angles and delays, (*iii*) the path fadings play the role of the signals.

The next and last step of the method thus consists of jointly estimating the parameters $(\boldsymbol{\theta}, \boldsymbol{\tau})$ that satisfy the model (7). Many of the well known methods such as maximum likelihood (ML), subspace fitting (SF), and MUSIC (see [10] for an overview) that have been developed for the DOA model (8) are applicable to the JADE problem. These algorithms estimate the desired parameters by solving an optimization problem of the general form

$$[\hat{\mathbf{U}}, \hat{\mathbf{T}}] = \arg\min_{\mathbf{U},\mathbf{T}} ||\mathbf{M} - \mathbf{U}(\boldsymbol{\theta}, \boldsymbol{\tau})\mathbf{T}||_F^2 \tag{9}$$

where $\mathbf{M}$ is a matrix obtained from the data $\mathbf{Y}$ (typically a basis for its column span) and $\mathbf{T}$ is any square invertible matrix. These algorithms as they apply to the JADE model are described in some detail in [7]. Identifiability of the parameters using such algorithms is addressed in section 3.3. For the solution to (9) to be ML, however, requires additional conditions to be satisfied: first, that the noise $\mathbf{V}$ is Gaussian and second, that the path fadings are uncorrelated from time slot to time slot (i.e., $\mathbf{B}\mathbf{B}^* \sim \mathbf{I}$). These two assumptions are discussed in in section 6. All methods to solve (9) without using further structure of $\mathbf{A}(\boldsymbol{\theta})$ or $\mathbf{G}(\boldsymbol{\tau})$ require multi-dimensional searches. A closed-form ESPRIT-like algorithm is possible if the array manifold has a doublet structure, and will be discussed in section 4.

3.3. Conditions for identifiability

To be able to identify $\boldsymbol{\theta}$, $\boldsymbol{\tau}$ using (7), we need that $\mathbf{Y}$ is a rank deficient matrix (under noise-free conditions). This implies that

1. $\mathbf{U}(\boldsymbol{\theta}, \boldsymbol{\tau})$ *be strictly tall and full column rank.* Thus it is necessary that $Q < MLP$. For $\mathbf{U}(\boldsymbol{\theta}, \boldsymbol{\tau})$ to be full column rank it is neither necessary nor sufficient that $\mathbf{A}(\boldsymbol{\theta})$ and/or $\mathbf{G}(\boldsymbol{\tau})$ be full rank. Even if a few angles or delays are identical, so that $\mathbf{A}(\boldsymbol{\theta})$ or $\mathbf{G}(\tau)$ are rank deficient, $\mathbf{U}$ may still be full rank, and it is easy to find counterexamples for the contrary (see section 5). Sufficient conditions occur in the simple case where $Q \leq \max\{M, LP\}$: by Proposition 2 proved in section 5, $\mathbf{U}$ is guaranteed to be full rank even if there are up to Q identical angles or delays.

2. $\mathbf{B}$ *be wide and full row rank.* This implies $S \geq Q$, i.e., we need to collect at least as many channel estimates as the number of multipaths. The full rank condition will be satisfied when the channel estimates are taken during a time interval larger than the coherence time of the channel. Note that it is not necessary that the fadings be uncorrelated from slot to slot (although this would improve the conditioning of $\mathbf{B}$).

To summarize, the most important condition for parameter identifiability is that the number of paths satisfies $Q < MPL$. Thus, JADE in general does not require more antennas than paths present (as is needed for identifiability in DOA models).

The rank condition on $\mathbf{B}$ can be alleviated by introducing spatial and temporal smoothing techniques, i.e., by constructing Hankel matrices out of each of the channel impulse response matrices. Ultimately, all rank conditions on $\mathbf{B}$ can be lifted, which brings us back to the single channel estimate version of the JADE algorithm considered in [8]. The penalty for smoothing is a reduced number of paths that can be estimated.

3.4. Multiple users

When there are more than one user in the same time slot, we can independently estimate the channel matrices $\mathbf{H}$ using each user's unique (usually orthogonal) training signal. We can then proceed as in the

single-user case, but with decoupled problems. If no training signals are available, we can still find the space-time channel $\mathbf{H}$ for each user using blind methods, which exploit finite alphabet structures and oversampling [11].

If the training sequences from each user are perfectly orthogonal, then ML is optimal, but since the estimation noise is independent from user to user, the ML problem "decouples", such that applying the JADE method to each user *separately* is still optimal. If the users' training sequences are not perfectly orthogonal, then the estimation noise is not independent from user to user. When the training signals are correlated between users, the optimal approach is to do parameter estimation *jointly* for all users. However, this is not computationally feasible, since the optimal approach ML requires that the noise be independent in order to arrive at a computationally tractable optimization problem such as (9). In summary, it is best in all cases of multi-user non-blind channel estimation to apply JADE to each user separately.

3.5. Estimating the number of paths

If not known in advance, the number of paths present can be estimated. Given the similarity of the JADE model to the angle-spread only model (8), many of the signal detection methods are applicable to our model. The most direct approach involves finding the multiplicity of the smallest eigenvalue of the covariance matrix of the channel estimates. Also the detection method based on the AIC and MDL principle outlined in [12] can be used, since the path fadings are normally distributed. The effects of angle spread and delay spread on the performance of these two methods need further study.

4. JADE-ESPRIT

We now discuss a computationally attractive closed-form algorithm to solve the JADE problem (7) in an approximate sense. A Fourier transform on the channel estimates is used to map $\mathbf{G}$ to a Vandermonde matrix $\mathbf{F}$, after which the ESPRIT algorithm [13] can be used to estimate the delays. If the antenna array is uniform linear or otherwise has the ESPRIT doublet-structure, then the angles and delays can be estimated jointly in a closed form, as follows.

We use the fact that a delay in the time domain maps to a pointwise multiplication by a phase progression in the frequency domain. This property is to a very good approximation true also for sampled signals in case the Nyquist condition is satisfied. Thus, by taking the Fourier transform of the rows of $\mathbf{H}$, we effectively perform a Fourier transform on every column $\mathbf{g}(\tau_i)$ of $\mathbf{G}$. We can then divide out the known Fourier transform $\mathbf{f}$ of the sampled pulse shape function $\mathbf{g}$ on its nonzero support. After some rearrangements, the channel model in (4) thus becomes (see [8] for details):

$$\tilde{\mathbf{H}}^{(n)} = [\mathbf{a}(\psi_i) \ \cdots \ \mathbf{a}(\psi_Q)] \begin{bmatrix} \beta_1^{(n)} & & 0 \\ & \ddots & \\ 0 & & \beta_Q^{(n)} \end{bmatrix} \begin{bmatrix} \mathbf{f}^T(\phi_1) \\ \vdots \\ \mathbf{f}^T(\phi_Q) \end{bmatrix} =: \mathbf{A}(\boldsymbol{\psi}) \ \mathrm{diag}[\boldsymbol{\beta}(n)] \ \mathbf{F}^T(\boldsymbol{\phi}) \qquad (M \times LP')) \tag{10}$$

where P' is the Nyquist rate,

$$\mathbf{f}(\phi_i) = [1 \ \phi_i \ \ldots \ \phi_i^{LP'-1}]^T, \quad \phi_i = e^{-j2\pi\tau_i/L}$$

and similarly, for a uniform linear array with sensor spacing Δ wavelengths,

$$\mathbf{a}(\psi_i) = [1 \ \psi_i \ \ldots \ \psi_i^{M-1}]^T, \quad \psi_i = e^{j2\pi\Delta \sin\theta_i} \, .$$

The JADE model (7) becomes, with obvious notation,

$$\tilde{\mathbf{Y}} = \mathbf{U}(\boldsymbol{\psi}, \boldsymbol{\phi})\mathbf{B} + \tilde{\mathbf{V}}, \qquad \mathbf{U}(\boldsymbol{\psi}, \boldsymbol{\phi}) = \mathbf{F}(\boldsymbol{\phi}) \circ \mathbf{A}(\boldsymbol{\psi}) \, . \tag{11}$$

Obviously, $\mathbf{U}(\boldsymbol{\psi}, \boldsymbol{\phi})$ has a double Vandermonde structure, similar to the situation of 2-D DOA estimation using a uniform rectangular array. A matrix $\mathbf{E}$ containing a basis of the column span of $\mathbf{U}$ can be estimated by taking the left singular vectors corresponding to the largest Q singular values of $\tilde{\mathbf{Y}}$. Without noise,

$\mathbf{E}\mathbf{T}^{-1} = \mathbf{U}(\boldsymbol{\psi}, \boldsymbol{\phi})$, where $\mathbf{T}$ is a square invertible $Q \times Q$ matrix. To estimate $(\boldsymbol{\psi}, \boldsymbol{\phi})$ using the Vandermonde structure of $\mathbf{U}(\boldsymbol{\psi}, \boldsymbol{\phi})$, define the following selection matrices:

$$\begin{array}{rclcrcl} \mathbf{J}_\psi &=& \mathbf{I}_{LP'} \otimes [\mathbf{I}_{M-1}\ 0_1], & & \mathbf{J}'_\psi &=& \mathbf{I}_{LP'} \otimes [0_1\ \mathbf{I}_{M-1}], \\ \mathbf{J}_\phi &=& [\mathbf{I}_{LP'-1}\ 0_1] \otimes \mathbf{I}_M, & & \mathbf{J}'_\phi &=& [0_1\ \mathbf{I}_{LP'-1}] \otimes \mathbf{I}_M \end{array}$$

and let $\mathbf{V}_\psi = \mathbf{J}_\psi \mathbf{U}$ and similarly for $\mathbf{V}'_\psi, \mathbf{V}_\phi, \mathbf{V}'_\phi$. The shift invariance structure in $\mathbf{U}$ gives that $\mathbf{V}'_\psi = \mathbf{V}_\psi \boldsymbol{\Psi}$, $\mathbf{V}'_\phi = \mathbf{V}_\phi \boldsymbol{\Phi}$ where $\boldsymbol{\Psi} = \mathrm{diag}[\psi_\mathbf{1} \ldots \psi_\mathbf{Q}]$ and $\boldsymbol{\Phi} = \mathrm{diag}[\phi_\mathbf{1} \ldots \phi_\mathbf{Q}]$. Finally let

$$\left\{ \begin{array}{rcl} \mathbf{E}_\psi &:=& \mathbf{J}_\psi \mathbf{E} \\ \mathbf{E}'_\psi &:=& \mathbf{J}'_\psi \mathbf{E} \end{array} \right. \qquad \left\{ \begin{array}{rcl} \mathbf{E}_\phi &:=& \mathbf{J}_\phi \mathbf{E} \\ \mathbf{E}'_\phi &:=& \mathbf{J}'_\phi \mathbf{E} \end{array} \right.$$

These data matrices have the structure

$$\left\{ \begin{array}{rcl} \mathbf{E}_\psi &=& \mathbf{V}_\psi \mathbf{T} \\ \mathbf{E}'_\psi &=& \mathbf{V}_\psi \boldsymbol{\Psi} \mathbf{T} \end{array} \right. \qquad \left\{ \begin{array}{rcl} \mathbf{E}_\phi &=& \mathbf{V}_\phi \mathbf{T} \\ \mathbf{E}'_\phi &=& \mathbf{V}_\phi \boldsymbol{\Phi} \mathbf{T} \end{array} \right. . \tag{12}$$

Since $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ are diagonal, the ψ_i's and ϕ_i's are given by the rank-reducing numbers of the matrix pencils $(\mathbf{E}'_\psi, \mathbf{E}_\psi)$ and $(\mathbf{E}'_\phi, \mathbf{E}_\phi)$ respectively. To obtain them one needs to jointly diagonalize $\mathbf{E}_\psi^\dagger \mathbf{E}'_\psi = \mathbf{T}^{-1} \boldsymbol{\Psi} \mathbf{T}$ and $\mathbf{E}_\phi^\dagger \mathbf{E}'_\phi = \mathbf{T}^{-1} \boldsymbol{\Phi} \mathbf{T}$. The reader is referred to [8, 14] for details of joint diagonalization methods. The connection of the θ_i's and τ_i's is provided by the fact that they have the same eigenvectors, the columns of $\mathbf{T}^{-1}$.

The algorithm could be extended with ideas from Unitary-ESPRIT: the number of columns can be doubled and all computations can be kept in the real domain [14, 15].

The above algorithm will be referred to as JADE-ESPRIT. It is quite similar to the recently developed SI-JADE algorithm [8, 11]. SI-JADE, however, only acts on a single channel estimate $\mathbf{H}^{(1)}$, so that the vectorization into $\mathbf{y}$ becomes useless. To enable identification, a block Hankel matrix $\mathcal{H}$ has to be constructed, by stacking horizontal shifts of $\mathbf{H}^{(1)}$. This reduces the number of columns and may also introduce an undesired coupling between the angle and delay estimates in case the model is not perfectly satisfied.

It is also possible to combine both algorithms. In this case every $\mathbf{H}^{(n)}$ is replaced by a block Hankel matrix $\mathcal{H}^{(n)}$ based on it. This would be useful in situations where sufficient fading diversity cannot be guaranteed so that $\mathbf{B}$ can be rank deficient.

Note that if only delay estimates are of interest (as is the case for TDOA-based triangularization methods for localization), then these can be estimated *without knowledge of the array manifold*, by just using the second set of equations in (12). This is attractive when the array manifold is unknown or not reliably calibrated.

The JADE-ESPRIT algorithm requires stricter identifiability conditions. That is, to uniquely identify $\boldsymbol{\Psi}$, $\boldsymbol{\Phi}$ using (12), $\mathbf{V}_\psi$ and $\mathbf{V}_\phi$ need to be tall and full rank, which implies that $Q < \min(M(LP' - 1), (M-1)LP')$. If $Q \leq \max\{M-1, LP-1\}$, then by Proposition 2 proved in section 5, $\mathbf{V}_\psi$ and $\mathbf{V}_\phi$ are guaranteed to be full rank even if there are up to Q identical angles (or delays).

5. Number of resolvable paths

We will assume that the identifiability requirements are satisfied. Therefore, $MLP - 1$ is an upper bound on the maximum number of resolvable paths Q_{max}. Furthermore, for any fixed M, L, P, we have $Q_{max} = \mathrm{rank}\, \mathbf{U}$. To see why, note that $\mathbf{U}$ has only $\rho := \mathrm{rank}\, \mathbf{U}$ linearly independent columns, and thus the angle-delay subspace (i.e., column span of $\mathbf{U}$) is ρ-dimensional. Thus there can be at most ρ paths that will yield unique parameter estimates.

The following result shows that the resolution power in space and in time each affect the resolution power in space-time:

Proposition 1. *Let Q be the dimension of the parameter vectors $(\boldsymbol{\theta}, \boldsymbol{\tau})$, and let $\mathbf{U}(\boldsymbol{\theta}, \boldsymbol{\tau}) = \mathbf{G}(\tau) \circ \mathbf{A}(\boldsymbol{\theta})$. Then*

$$\mathrm{rank}\, \mathbf{U}(\boldsymbol{\theta}, \boldsymbol{\tau}) \leq \min\{Q,\ \mathrm{rank}\, \mathbf{G}(\tau) \cdot \mathrm{rank}\, \mathbf{A}(\boldsymbol{\theta})\} \tag{13}$$

A sufficient condition for equality is to have $\mathbf{G}$ and/or $\mathbf{A}$ tall and full rank.

PROOF Define the singular value decompositions $\mathbf{A} = \mathbf{U}_A \mathbf{\Sigma}_A \mathbf{V}_A^*, \mathbf{G} = \mathbf{U}_G \mathbf{\Sigma}_G \mathbf{V}_G^*$ (the matrices $\mathbf{U}_{A,G}$ and $\mathbf{V}_{A,G}$ are not related to $\mathbf{U}$ and $\mathbf{V}$) and note that

$$\mathbf{G} \circ \mathbf{A} = (\mathbf{U}_G \otimes \mathbf{U}_A)(\mathbf{\Sigma}_G \otimes \mathbf{\Sigma}_A)(\mathbf{V}_G^* \circ \mathbf{V}_A^*),$$

where we have used the general relations proved in [9] $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC} \otimes \mathbf{BD})$ and $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \circ \mathbf{D}) = (\mathbf{AC}) \circ (\mathbf{BD})$. Observing that $\mathbf{U}_G \otimes \mathbf{U}_A$ is a unitary matrix, that $\operatorname{rank}(\mathbf{V}_G^* \circ \mathbf{V}_A^*) = \operatorname{rank} \mathbf{V}_G^* = Q$, and that $\operatorname{rank}(\mathbf{\Sigma}_G \otimes \mathbf{\Sigma}_A) = \operatorname{rank} \mathbf{G} \cdot \operatorname{rank} \mathbf{A}$, we have

$$\operatorname{rank}(\mathbf{G} \circ \mathbf{A}) = \operatorname{rank}\{(\mathbf{\Sigma}_G \otimes \mathbf{\Sigma}_A)(\mathbf{V}_G^* \circ \mathbf{V}_A^*)\} \leq \min\{\operatorname{rank}(\mathbf{\Sigma}_G \otimes \mathbf{\Sigma}_A), \operatorname{rank}(\mathbf{V}_G^* \circ \mathbf{V}_A^*)\} = \min\{\operatorname{rank} \mathbf{G} \cdot \operatorname{rank} \mathbf{A}, Q\}$$

To see that a sufficient condition for equality is that $\mathbf{G}$ is tall ($Q \leq LP$) and full rank, we note that by results in [9], all Q columns of $\mathbf{U}$ are linearly independent, and thus $\operatorname{rank} \mathbf{U} = Q = \min\{Q,\ Q \cdot \operatorname{rank} \mathbf{A}\}$. The argument for the case when $\mathbf{A}$ is tall and full rank is similar. □

A simple counterexample of matrices where both $\mathbf{A}$ and $\mathbf{G}$ are full rank (but "wide"), and $\mathbf{G} \circ \mathbf{A}$ is rank deficient (even though it is "tall"), is provided by

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 2 & 1 \\ 0 & 0 & 1 & 3 & 1 \end{bmatrix}, \qquad \mathbf{G} = \begin{bmatrix} 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 2 & 1 & 0.5 \end{bmatrix}.$$

$\mathbf{G} \circ \mathbf{A}$ has size 6×5, but is of rank 4. Note that any 3-pair of columns of $\mathbf{A}$ are linearly independent (and the same with 2-pairs of columns of $\mathbf{G}$), so that the array and time manifolds can be considered unambiguous for this instance. It follows that distinct angles and delays, and unambiguous space and time manifolds are not sufficient to have an unambiguous space-time manifold. (For practical purposes, however, such counterexamples can be considered contrived: equality holds in (13) for almost every $(\boldsymbol{\theta}, \boldsymbol{\tau})$.)

We continue to show that the number of resolvable paths is not determined by the oversampling factor P, if we sample at or beyond the Nyquist rate. If we increase P, then the number of rows of the space-time manifold matrix increases, so it appears that we can also tolerate more paths and expect to resolve them, since Q is limited only by MLP. Yet oversampling past a certain rate should not increase the resolution power when the signal has finite bandwidth [16]. In typical communication systems, the modulation waveform $g(t)$ is bandlimited, with spectrum nonzero only for $|\omega| \leq \frac{\pi}{T}(1+\alpha)$, where $0 \leq \alpha \leq 1$ is the excess bandwidth with respect to the symbol rate. As far as the rank of $\mathbf{U}$ is concerned, this implies that P is effectively replaced by $(1+\alpha)$:

Proposition 2. *The effective rank of* $\mathbf{U}$ *is at most* $M(1+\alpha)L$.

PROOF The spectrum of $g(t)$ is nonzero for $|\omega| \leq \frac{\pi}{T}(1+\alpha)$. Consider the discrete Fourier transform $\mathbf{f}$ of the sampled pulse shape function $\mathbf{g}$, consisting of LP samples. If we sample faster than the Nyquist rate $(1+\alpha)/T$, i.e., $P \geq 1+\alpha$, then $\mathbf{f}$ has only $\gamma := (1+\alpha)L$ nonzero elements out of a total of LP — all others are *effectively*[1] zero. Thus $\mathbf{f}$ has the form

$$\mathbf{f} = [f_1\ f_2\ \ldots\ f_{\gamma/2}\ 0\ \ldots\ 0\ f_{\gamma/2+1}\ \ldots\ f_\gamma]^T.$$

Moreover, since a delay in the time domain is equivalent to pointwise multiplication with a phase progression in the frequency domain,

$$\mathbf{g}(\tau) \quad \leftrightarrow \quad \mathbf{f}(\tau) = [1 \quad \phi\ \cdots\ \phi^{LP-1}] \odot \mathbf{f} \qquad \phi = e^{-2\pi\tau/L}, \tag{14}$$

sampling $g(t-\tau)$ rather than $g(t)$ does not change the above argument, and the zero entries do not change location.

With Q paths, we can construct a matrix $\mathbf{F}$ consisting of Q columns of this form. Thus, it cannot have rank bigger than $(1+\alpha)L$. The same must hold for $\mathbf{G}$, since $\mathbf{G} = \mathcal{F}_{LP}^{-1}\mathbf{F}$, where $\mathcal{F}_{LP}$ is the $LP \times LP$ DFT matrix, which is a scaling of a unitary matrix. Applying proposition 1, we obtain that $\operatorname{rank} \mathbf{U} \leq \min\{Q, M(1+\alpha)L\}$. □

[1] Not precisely, since $g(t)$ was supposed to have finite support.

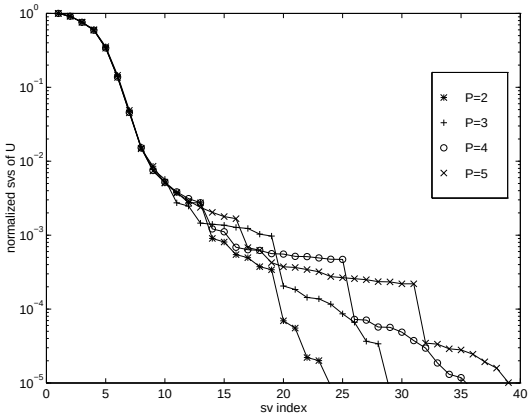

Figure 1. Singular values of $\mathbf{U}$ for various oversampling rates P, with number of rays $Q = MPL - 1$.

Numerical example. As an example of proposition 2, consider the case of a signal arriving at an $M = 2$-element antenna via Q paths with total delay spread $\delta = 3$ symbol periods ($T = 1$) and total angle spread 120°. The paths are considered equispaced in time and space: the first path has delay 0, the second $\frac{\delta}{Q-1}$, the third $\frac{2\delta}{Q-1}$, and so on. The modulation waveform is a raised-cosine pulse with excess bandwidth $\alpha = 0.25$, truncated to zero outside the interval $(-3, 3]$. We observe the rank of $\mathbf{U}$ as the oversampling rate increases. Taking $P = 2, 3, 4$ and 5, and Q to be the maximum number of paths resolvable if the pulse shape would be not bandlimited, $Q = MLP - 1$, we note that even though we simulate more paths with higher oversampling rates, the singular value plots in Figure 1 are overlapping: the large singular values are all the same and the increase in rank is only noticeable from the addition of small singular values, which is unobservable in practice. Thus, the effective rank of $\mathbf{U}$ does not increase. □

Remarks on rank. So far we assumed that $\mathbf{A}$ has maximum attainable rank M. But its rank is affected by angle-spread and beamwidth, such that, for example, if the beamwidth is large compared with the angle-spread, then the number of resolvable paths in space is far from reaching M. As an illustration, say we have 4 equispaced antennas and 4 angles equally spaced in the interval $[0°, 90°]$. Then rank $\mathbf{A}$ depends on the spacing of the sensors, i.e., on the beamwidth. With a spacing of $\lambda/100$, where λ is the wavelength, the beamwidth is large compared to the angle spread of 90°, and the numerical rank of $\mathbf{A}$ is only 2. For $\lambda/5$, the beamwidth decreases to $\approx 75°$ and rank $\mathbf{A}$ increases to 3. Finally for half-wavelength spacing the beamwidth is $\approx 30°$ and rank $\mathbf{A}$ reaches 4.

By looking at the structure of $\mathbf{A}$ for fixed M and Q, it is noted that its condition improves if the angle spread increases, since the vector-norm separation of its columns increases; the same improvement in condition is expected if the sensor spacing increases (and thus beamwidth decreases), since this causes the separation of the rows of $\mathbf{A}$ to increase. We can thus conclude that the number of resolvable paths is directly proportional with the total angle spread and inversely proportional with the array beamwidth. An exact relationship does not exist, because the rank of $\mathbf{A}$ depends in a very complex manner on the array geometry, exact definition of beamwidth and angle spread, as well as the tolerance for numerical rank determination (recall that $\mathbf{A}$ is always of full theoretical rank for distinct paths). However, a rule-of-thumb can be given: The effective rank of $\mathbf{A}$ is $\min(M, Q, 1 + \text{Angle Spread/ Beamwidth})$. We emphasize that the last argument of min is just a "ball-park estimate", inspired by the Rayleigh resolution criterion.

A similar reasoning applies in the delay-spread only model, such that the rank of $\mathbf{G}$ is not always LP or even $L(1 + \alpha)$ but depends on the delay-spread and pulse-width: with a very wide pulse shape we cannot distinguish multiple paths with a short total delay-spread. Again, an exact relationship does not exist, but a heuristic can be given: the approximate number of resolvable paths is $\min(L(1 + \alpha), Q, 1 +$ Delay Spread/ Pulsewidth).

6. The Cramer-Rao bound

The Cramer-Rao bound (CRB) provides a lower bound on the variance of any unbiased estimator. We derive the lower bound on the covariance matrix of $[\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\tau}}]$ for the JADE problem (7). To this end we assume the estimation noise $\mathbf{v}(n),\ n = 1, \ldots, S$ is a complex, stationary, zero-mean Gaussian random process, uncorrelated from slot to slot and from path to path. Formally, the last two assumptions are (i) $E[\mathbf{v}(n)\mathbf{v}(m)^*] = \sigma_h^2 \delta_{n,m}\mathbf{I},\ E[\mathbf{v}(n)\mathbf{v}(m)^{\mathrm{T}}] = 0$, and (ii) $E[\mathbf{v}(n)\mathbf{v}(n)^*] = \sigma_h^2\mathbf{I},\ E[\mathbf{v}(n)\mathbf{v}(n)^{\mathrm{T}}] = 0$, where σ_h^2 is the variance of the entries of the estimation noise matrix $\mathbf{V}$ and $\delta_{n,m}$ denotes the Kronecker delta. For the case of non-blind channel estimation, it can be shown that, if the entries of the noise $\mathbf{N}$ are i.i.d. complex normally distributed with variance σ^2, and if the training sequences are designed to have perfect autocorrelation properties (i.e., $\mathbf{SS}^* = N\mathbf{I}_L$), then the estimation noise $\mathbf{V}$ is indeed white and Gaussian, with its columns satisfying (i) and (ii) above with $\sigma_h^2 = \frac{\sigma^2}{N}$. For blind channel estimation techniques, the estimation noise distribution is difficult to assess and may not be white and Gaussian.

The CRB depends on whether the path fadings are modeled as unknown deterministic (i.e., fixed) quantities or as random variables with a known distribution. The deterministic CRB appeared (without proof) in [1]:

$$CRB(\boldsymbol{\theta}, \boldsymbol{\tau}) = \frac{\sigma_h^2}{2} \left\{ \sum_{n=1}^{S} \mathrm{real}\left(\mathcal{B}(n)^* \mathbf{D}^* \mathbf{P}_{\mathbf{U}}^{\perp} \mathbf{D} \mathcal{B}(n) \right) \right\}^{-1} \tag{15}$$

where $\mathcal{B}(n) = \mathbf{I}_2 \otimes \mathrm{diag}[\boldsymbol{\beta}(n)]$, $\mathbf{P}_{\mathbf{U}}^{\perp} = \mathbf{I} - \mathbf{U}\mathbf{U}^{\dagger}$, and $\mathbf{D} = [\mathbf{G} \circ \mathbf{A}',\ \mathbf{G}' \circ \mathbf{A}]$. Here, prime denotes differentiation where each column is differentiated with respect to the corresponding parameter and all matrices are evaluated at the true parameter values:

$$\mathbf{A}' = \mathbf{A}'(\boldsymbol{\theta}) = \left[\frac{\mathrm{d}\mathbf{a}}{\mathrm{d}\theta_1}(\theta_1),\ \cdots,\ \frac{\mathrm{d}\mathbf{a}}{\mathrm{d}\theta_Q}(\theta_Q) \right],$$

and similarly for $\mathbf{G}' = \mathbf{G}'(\boldsymbol{\tau})$. The proof of this claim is similar to the one in [17] and left to the appendix.

For a Rayleigh-fading channel, the path fadings $\boldsymbol{\beta}$ have a zero-mean complex Gaussian distribution, with some covariance matrix $\mathbf{R}_B$. A key assumption needed here is that the fadings are uncorrelated from time slot to time slot (otherwise the joint likelihood function will not be the product of the individual likelihoods from each slot). This assumption is only valid when the radio channel is sufficiently fast fading (coherence time shorter than the frame length). For this case, the Cramer-Rao bound has been developed in [7]:

$$CRB(\boldsymbol{\theta}, \boldsymbol{\tau}) = \frac{\sigma_h^2}{2} \left\{ \mathrm{real}\left(\mathbf{D}^* \mathbf{P}_U^{\perp} \mathbf{D} \odot (\mathbf{1}_{2\times 2} \otimes \mathbf{R}_B \mathbf{U}^* \mathbf{R}_Y^{-1} \mathbf{U} \mathbf{R}_B)^T \right) \right\}^{-1} \tag{16}$$

Here $\mathbf{1}_{2\times 2}$ is a 2×2 matrix of ones, and $\mathbf{R}_Y$ is the covariance matrix of the channel estimates, given by $\mathbf{R}_Y = \mathbf{U}\mathbf{R}_B\mathbf{U}^* + \sigma_h^2\mathbf{I}$.

7. Simulations results

The performance of the algorithm is assessed by the following computer simulations. We initially assume one source emitting signals that arrive at an array of $M = 2$ half-wavelength spaced sensors via $Q = 3$ paths. The additive noise at the array has a variance of -15 dB. The angles-of-arrival are $[-5,\ 0,\ 10]^\circ$ and time delays $[0,\ 0.8,\ 1.0]T$. The path fadings are generated from a complex Gaussian distribution with mean zero and variance [1, .9, .8] respectively for the three rays. We also assume the communication systems uses training bits, from which the channel is estimated using least squares. We collect samples of $\mathbf{x}(t)$ during $N = 40$ symbol periods and a total of $S = 20$ time slots. The pulse shape function is a raised cosine with 0.35 excess bandwidth, truncated to a length of $L_g = 6$ symbols. The sampling rate is $P = 2$ and the symbol rate is normalized to $T = 1$. The experimental standard deviation of the estimates is averaged over 500 Monte-Carlo runs of JADE-ESPRIT, and is compared against the deterministic CRB.

Basic performance of JADE-ESPRIT. Figure 2 shows the experimental standard deviation of the angle and delay estimates as a function of the noise power σ^2. The three curves correspond to each of the three paths. Compared to the CRB, the difference is approximately 3 to 5 dB. The bias of the estimates was at least an order of magnitude smaller than their standard deviation.

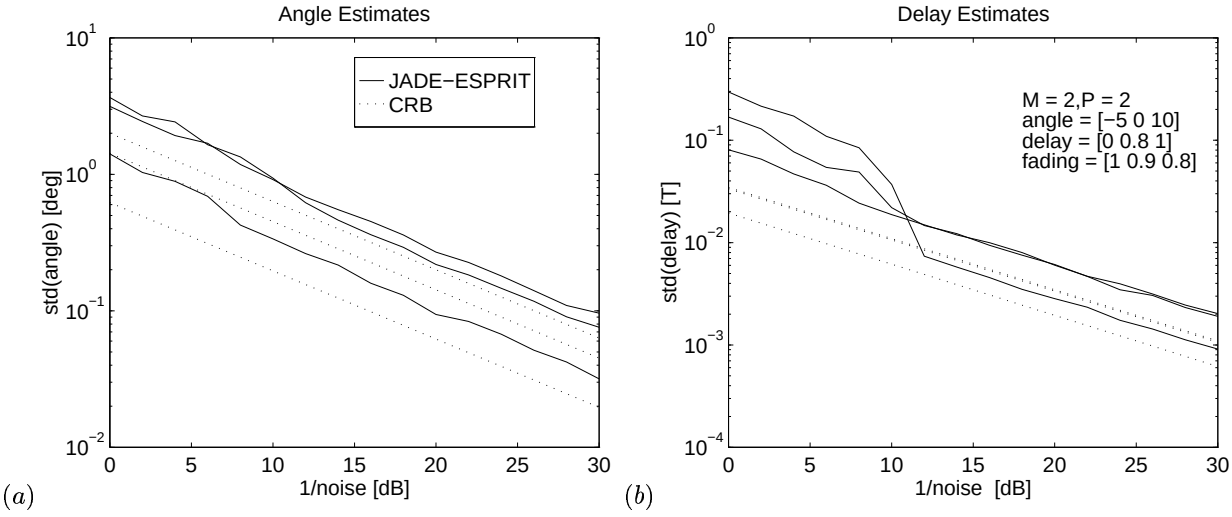

Figure 2. Standard deviation of JADE-ESPRIT estimates vs. noise

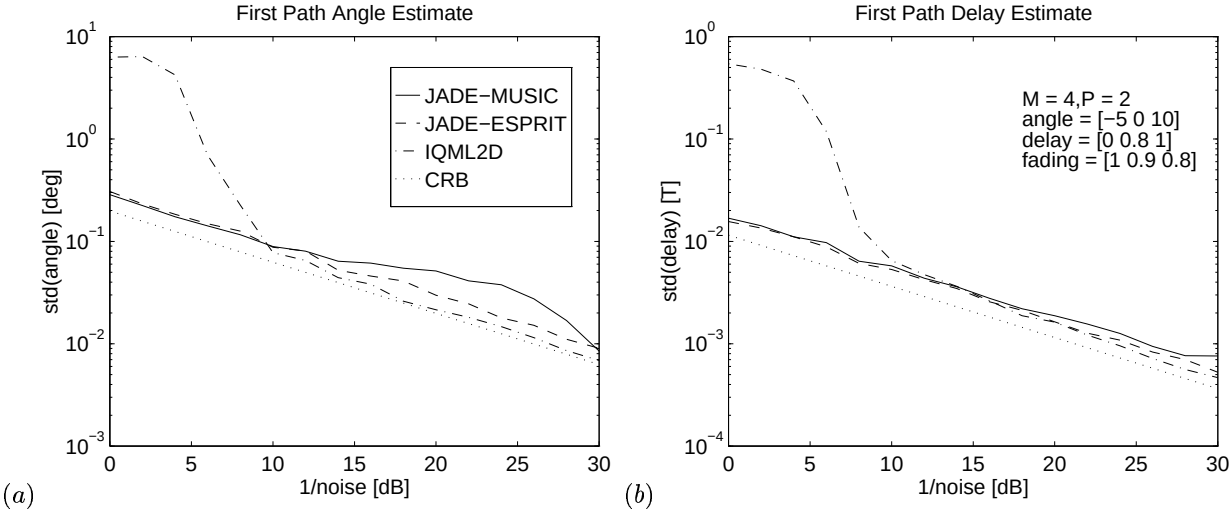

Figure 3. Standard deviation of estimates vs. noise

Comparison to other algorithms. To fairly compare the JADE approach to other algorithms, we need algorithms that use multiple channel estimates. The only obvious choice is the optimal IQML-2D method of [18]. IQML-2D was originally developed for estimating the two-dimensional modes of sinusoids in Gaussian noise, being based on ML. It can be used to determine angles and delays if both manifolds have Vandermonde structure. Thus IQML-2D can use the same model as JADE-ESPRIT (11), but it requires more antennas than paths. We thus let $M = 4$ and $Q = 3$ for the comparison. The results are shown in Figure 3; for readability, only the statistics of the first path are shown, the other paths exhibiting similar behavior. "JADE-MUSIC" is the MUSIC-based JADE algorithm [1]. We note that IQML-2D approaches the CRB as the noise decreases, as expected, while the JADE-based algorithms do not. However at lower SNR's JADE outperforms IQML-2D.

8. Conclusions

In this work we have presented a unified view of a parametric, subspace-based method to jointly estimate multipath parameters. The main novel parts of this work are (i) an algorithm that uses shift-invariance techniques arising from exploiting uniform sampling in space and time; and (ii) an analysis of the resolution power of the approach. We have exploited both spatial and temporal information, in particular the known pulse-shape function. Relying on the spatial manifold alone to extract the desired parameters has

proved to be of limited value in practice, because of the expense of array calibration and variability of the array response due to changes in the environment. The JADE approach relies on both the spatial and the temporal manifold, resulting in increased robustness, since the temporal manifold is known exactly and does not change with the environment.

9. Appendix

To prove the deterministic Cramer-Rao bound, let the parameter vector be defined as $[\sigma_h^2 \ \boldsymbol{\eta}^T]^T$, where $\boldsymbol{\eta} := [\boldsymbol{\theta}^T, \boldsymbol{\tau}^T]^T$. The likelihood function of the data, which is the channel estimates $\mathbf{y}$, i.i.d. Gaussian, satisfying (7) is:

$$L(\mathbf{y}(1), \ldots \mathbf{y}(S)) = \frac{1}{(2\pi)^{MPLS}(\sigma_h^2/2)^{MPLS}} \exp\left\{ -\frac{1}{\sigma_h^2} \sum_{n=1}^{S} [\mathbf{y}(n) - \mathbf{U}\boldsymbol{\beta}(n)]^* \cdot [\mathbf{y}(n) - \mathbf{U}\boldsymbol{\beta}(n)] \right\}$$

Thus, the log-likelihood function is

$$\ln L = \text{const} - MPLS \ln \sigma_h^2 - \frac{1}{\sigma_h^2} \sum_{n=1}^{S} [\mathbf{y}(n) - \mathbf{U}\boldsymbol{\beta}(n)]^* \cdot [\mathbf{y}(n) - \mathbf{U}\boldsymbol{\beta}(n)]. \tag{17}$$

We compute the derivatives of the above with respect to σ_h^2, $\{\bar{\boldsymbol{\beta}}(n) := \text{real } \boldsymbol{\beta}(n)\}$, $\{\tilde{\boldsymbol{\beta}}(n) := \text{imag } \boldsymbol{\beta}(n)\}$, and $\boldsymbol{\eta}$.

$$\frac{\partial \ln L}{\partial(\sigma_h^2)} = -\frac{MPLS}{\sigma_h^2} + \frac{1}{\sigma_h^2} \sum_{n=1}^{S} \mathbf{v}^*(n)\mathbf{v}(n)$$

$$\frac{\partial \ln L}{\partial \bar{\boldsymbol{\beta}}(n)} = \frac{2}{\sigma_h^2} \text{real}[\mathbf{U}^* \mathbf{v}(n)], \qquad \frac{\partial \ln L}{\partial \tilde{\boldsymbol{\beta}}(n)} = \frac{2}{\sigma_h^2} \text{imag}[\mathbf{U}^* \mathbf{v}(n)],$$

The derivatives of (17) with respect to $\boldsymbol{\eta}$ are different from the ones in the angle-spread only model:

$$\begin{aligned} \frac{\partial \ln L}{\partial \theta_i} &= \frac{1}{\sigma_h^2} \sum_{n=1}^{S} \frac{\partial}{\partial \theta_i} \{2 \text{ real } \{\boldsymbol{\beta}^*(n)\mathbf{U}^*\mathbf{y}(n)\} - \boldsymbol{\beta}^*(n)\mathbf{U}^*\mathbf{U}\beta(n)\} \\ &= \frac{2}{\sigma_h^2} \sum_{n=1}^{S} \text{real } \{\beta_i(n)\mathbf{d}_{\theta_i}^* \mathbf{v}(n)\} \quad i = 1, \ldots, Q, \end{aligned}$$

and similarly,

$$\frac{\partial \ln L}{\partial \tau_i} = \frac{2}{\sigma_h^2} \sum_{n=1}^{S} \text{real } \{\beta_i(n)\mathbf{d}_{\tau_i}^* \mathbf{v}(n)\} \quad i = 1, \ldots, Q,$$

where $\mathbf{d}_{\theta_i}$ is the derivative with respect to θ_i of the i-th column of $\mathbf{U}$, $\mathbf{d}_{\theta_i} = \mathbf{g}(\tau_i) \otimes \partial \mathbf{a}(\theta_i)/\partial\theta_i$ similarly for $\mathbf{d}_{\tau_i}$. Written more compactly,

$$\frac{\partial \ln L}{\partial \boldsymbol{\theta}} = \frac{2}{\sigma_h^2} \sum_{n=1}^{S} \text{real } \{(\text{diag}[\boldsymbol{\beta}])^*(n) D_{\boldsymbol{\theta}}^* \mathbf{v}(n)\}, \qquad \frac{\partial \ln L}{\partial \boldsymbol{\tau}} = \frac{2}{\sigma_h^2} \sum_{n=1}^{S} \text{real } \{(\text{diag}[\boldsymbol{\beta}])^*(n) D_{\boldsymbol{\tau}}^* \mathbf{v}(n)\},$$

where $\mathbf{D}_{\boldsymbol{\theta}} = [\mathbf{d}_{\theta_1} \ldots \mathbf{d}_{\theta_Q}] := \mathbf{G} \circ \mathbf{A}'$ and $\mathbf{D}_{\boldsymbol{\tau}} = [\mathbf{d}_{\tau_1} \ldots \mathbf{d}_{\tau_Q}] := \mathbf{G}' \circ \mathbf{A}$. Finally, since $\mathbf{D} = [\mathbf{D}_{\boldsymbol{\theta}} \ \mathbf{D}_{\boldsymbol{\tau}}]$ $(MPL \times 2Q)$, we arrive at

$$\frac{\partial \ln L}{\partial \boldsymbol{\eta}} = \frac{2}{\sigma_h^2} \sum_{n=1}^{S} \text{real } \{\mathcal{B}^*(n)\mathbf{D}^*\mathbf{v}(n)\}$$

Using results proven in [17], we first note that $\frac{\partial \ln L}{\partial(\sigma_h^2)}$ is not correlated with any of the other derivatives computed above. Then we obtain

$$E\left[\left(\frac{\partial \ln L}{\partial(\sigma_h^2)}\right)^2\right] = \frac{MPLS}{\sigma_h^4},$$

$$E\left[\left(\frac{\partial \ln L}{\partial \bar{\boldsymbol{\beta}}(n)}\right)\left(\frac{\partial \ln L}{\partial \bar{\boldsymbol{\beta}}(m)}\right)^T\right] = \frac{2}{\sigma_h^2}\text{real}\left[\mathbf{U}^*\mathbf{U}\right]\delta_{n,m},$$

$$E\left[\left(\frac{\partial \ln L}{\partial \bar{\boldsymbol{\beta}}(n)}\right)\left(\frac{\partial \ln L}{\partial \tilde{\boldsymbol{\beta}}(m)}\right)^T\right] = -\frac{2}{\sigma_h^2}\text{imag}\left[\mathbf{U}^*\mathbf{U}\right]\delta_{n,m},$$

$$E\left[\left(\frac{\partial \ln L}{\partial \tilde{\boldsymbol{\beta}}(n)}\right)\left(\frac{\partial \ln L}{\partial \tilde{\boldsymbol{\beta}}(m)}\right)^T\right] = \frac{2}{\sigma_h^2}\text{real}\left[\mathbf{U}^*\mathbf{U}\right]\delta_{n,m},$$

$$E\left[\left(\frac{\partial \ln L}{\partial \bar{\boldsymbol{\beta}}(n)}\right)\left(\frac{\partial \ln L}{\partial \boldsymbol{\eta}}\right)^T\right] = \frac{2}{\sigma_h^2}\text{real}\left[\mathbf{U}^*\mathbf{D}\mathcal{B}(n)\right],$$

$$E\left[\left(\frac{\partial \ln L}{\partial \tilde{\boldsymbol{\beta}}(n)}\right)\left(\frac{\partial \ln L}{\partial \boldsymbol{\eta}}\right)^T\right] = \frac{2}{\sigma_h^2}\text{imag}\left[\mathbf{U}^*\mathbf{D}\mathcal{B}(n)\right],$$

$$E\left[\left(\frac{\partial \ln L}{\partial \boldsymbol{\eta}}\right)\left(\frac{\partial \ln L}{\partial \boldsymbol{\eta}}\right)^T\right] = \frac{2}{\sigma_h^2}\sum_{n=1}^{S}\text{real}\left[\mathcal{B}^*(n)\mathbf{D}^*\mathbf{D}\mathcal{B}(n)\right].$$

Finally the Fisher Information Matrix (FIM) for the parameters is given by $E[\boldsymbol{\omega}\boldsymbol{\omega}^T]$, where $\boldsymbol{\omega} := \partial \ln L/\partial[\sigma_h^2\ \bar{\boldsymbol{\beta}}^T(1)\ \tilde{\boldsymbol{\beta}}^T(1)\ \ldots\ \bar{\boldsymbol{\beta}}^T(S)\ \tilde{\boldsymbol{\beta}}^T(S)\ \boldsymbol{\eta}^T]$. Again using results in [17] we can show that the CRB for the parameters of interest is

$$CRB(\boldsymbol{\eta})^{-1} = \frac{2}{\sigma_h^2}\left\{\sum_{n=1}^{S}\text{real}\left(\mathcal{B}(n)^*\mathbf{D}^*\mathbf{P}_{\mathbf{U}}^{\perp}\mathbf{D}\mathcal{B}(n)\right)\right\},$$

which completes the proof.

References

[1] M.C. Vanderveen, C.B. Papadias, and A. Paulraj, "Joint angle and delay estimation (JADE) for multipath signals arriving at an antenna array," *IEEE Communications Letters*, vol. 1, Jan. 1997.

[2] T.S. Rappaport, J.H.Reed, and B.D. Woerner, "Position location using wireless communications on highways of the future," *IEEE Communications Magazine*, Oct. 1996.

[3] Y. Ogawa, N. Hamaguchi, K. Ohshima, and K. Itoh, "High-resolution analysis of indoor multipath propagation structure," *IEICE Transactions on Communications*, vol. E78-B, pp. 1450–1457, Nov. 1995.

[4] M. Wax and A. Leshem, "Joint estimation of time delays and directions of arrival of multiple reflections of a known signal," in *Proc. IEEE ICASSP*, (Atlanta, GA), pp. 2622–2625, IEEE, May 1996.

[5] A.L. Swindlehurst, "Time delay and spatial signature estimation using known asynchronous signals," *preprint*, 1996.

[6] M.C. Cedervall and A. Paulraj, "Joint channel and space-time parameter estimation," in *Proc. 30th Asilomar Conf. Circuits, Syst. Comp.*, Nov. 1996.

[7] M.C. Vanderveen, B.C. Ng, C.B. Papadias, and A. Paulraj, "Joint angle and delay estimation (JADE) for signals in multipath environments," in *Proc. 30th Asilomar Conf. Circuits, Syst. Comp.*, Nov. 1996.

[8] A.J. van der Veen, M.C. Vanderveen, and A. Paulraj, "Joint angle and delay estimation using shift-invariance properties," *IEEE Signal Processing Letters*, May 1997.

[9] C. G. Khatri and C. R. Rao, "Solutions to some functional equations and their applications to characterizations of probability distributions," *Sankhya*, vol. 30, pp. 167–180, 1968.

[10] A. Paulraj, B. Ottersten, R. Roy, L. Swindlehurst, G. Xu, and T. Kailath, *Handbook of Statistics, Signal Processing and Applications, Editors: C.R. Rao and N.K. Bose*, vol. 10. Elsevier Press, 1993.

[11] A.J. van der Veen, S. Talwar, and A. Paulraj, "A subspace approach to blind space-time signal processing for wireless communications systems," *IEEE Trans. Signal Processing*, vol. 45, pp. 173–190, Jan. 1997.

[12] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 33, pp. 387–392, April 1985.

[13] A. Paulraj, R. Roy, and T. Kailath, "Estimation of signal parameters via rotational invariance techniques – ESPRIT," in *Proc. Nineteenth Asilomar Conf. on Circuits, Systems and Comp.*, (San Jose, CA), Maple Press, November 1985.

[14] A.J. van der Veen, M.C. Vanderveen, and A. Paulraj, "Joint angle and delay estimation using shift-invariance techniques." submitted to IEEE Trans. Signal Processing, Mar. 1997.

[15] M.D. Zoltowski, M. Haardt, and C.P. Mathews, "Closed-form 2-D angle estimation with rectangular arrays in element space or beamspace via Unitary ESPRIT," *IEEE Trans. Signal Processing*, vol. 44, pp. 316–328, Feb. 1996.

[16] A.J. van der Veen, "Resolution limits of blind multi-user multi-channel identification schemes—the bandlimited case," in *Proc. IEEE ICASSP*, (Atlanta, GA), pp. 2722–2725, May 1996.

[17] P. Stoica and A. Nehorai, "MUSIC, Maximum Likelihood and Cramér-Rao bound," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 720–741, May 1989.

[18] M.P. Clark and L.L. Scharf, "Two-dimensional modal analysis based on maximum likelihood," *IEEE Trans. Signal Processing*, vol. 42, pp. 1443–52, June 1994.