

Joint angle and delay estimation using shift-invariance techniques

Alle-Jan van der Veen, Michaela C. Vanderveen, and Arogyaswami Paulraj

Abstract—In a multipath communication scenario, it is often relevant to estimate the directions and relative delays of each multipath ray. We derive a closed-form subspace-based method for the simultaneous estimation of these parameters from an estimated channel impulse response, using knowledge of the transmitted pulse shape function. The algorithm uses a 2-D ESPRIT-like shift-invariance technique to separate and estimate the phase shifts due to delay and direction-of-incidence, with automatic pairing of the two parameter sets. Improved resolution is obtained by enlarging the data matrix with shifted and conjugated copies of itself.

I. INTRODUCTION

Source localization is one of the recurring problems in signal processing. In general, it can involve the joint estimation of frequencies, Doppler shifts, directions of arrival (azimuth/elevation) and time/time-difference of arrival, and as such it is the central issue in many radar or sonar applications. In mobile communications, source localization by the base station is of interest for advanced handover schemes, emergency localization, and potentially many user services for which a GPS receiver is impractical (see [2] for a recent discussion in this area).

In a multipath scenario, source localization by the base station involves the estimation of the directions and relative delays of each multipath ray. It is often assumed that the directions and delays of the paths do not change quickly, as fading affects only their powers, so that it makes sense to estimate these parameters. This information can then be used to adjust a space-time (RAKE) receiver in the uplink, although for this purpose it is not really necessary to determine the parameters themselves: estimation of the independent space-time manifold components suffices (see e.g. [3,4]). However, the parameters are essential for space-time selective transmission in the downlink, especially in FDD systems: the space-time manifold is frequency-dependent and quickly varies because of independent fading of the rays, but the angle and delay parameters are relatively stationary. Knowledge of these parameters can be used for effective transmit diversity as well.

⁰Manuscript submitted to IEEE Trans. Signal Processing, March 1997, revised July 1997. A preliminary short version of this paper appeared as [1]. A.J. van der Veen is with the Department of Electrical Engineering/DIMES, Delft University of Technology, Delft, The Netherlands. M.C. Vanderveen is with the Scientific Computing Program, Stanford University, Stanford, CA 94305. A. Paulraj is with the Information Systems Laboratory, Stanford University, Stanford, CA 94305. This research was supported by the National Science Foundation and by the Department of the Army, Army Research Office, under Grant No. DAAH04-95-1-0249. The work of A.J. van der Veen was performed in the context of the project ACTS AC090 "FRAMES", which is partly funded by the European Community.

In this paper, we derive an algorithm for the joint high-resolution estimation of multipath angles and delays, assuming linearly modulated sources with a known pulse shape function and no appreciable Doppler shifts. An extensive literature on source localization exists, and hence it is essential to list the conditions on the mobile communications scenario that we shall consider:

1. The number of sources is small. For convenience, we consider only one source in a multipath environment, but this is no limitation.
2. The multipath environment is modeled by a discrete number of rays, each parameterized by a delay, complex amplitude (fading), and angle. This leaves out diffuse scattering.
3. A channel estimate is available. For communication applications, this typically implies that the source signals are (known) digital sequences, linearly modulated by known pulse shape functions.
4. Doppler shifts and residual carriers of sources are neglected.
5. The source signals are received by a narrowband phased array consisting of at least two antennas spaced at half-wavelength or closer.
6. The data received by the antennas is sampled at or above the Nyquist rate. For digital sources, this typically implies fractional sampling by a factor of two.

The Doppler condition restricts the delay estimation by placing a limit on the number of samples that can be processed in one batch. Similarly, the narrowband assumption in 5 entails that a delay across the array can be modeled as a phase shift. This puts a limit on the processing bandwidth: it should be much less than the carrier frequency so that the wavelengths at both ends of the band are roughly the same. In communication applications, the condition that the sources are known is not strict since the algorithm can be used in tandem with a blind channel estimator. In the case of multiple sources with training available, we can get independent channel estimates of each source, which reduces the problem to the single-source case.

Active radar applications are viable as well, provided again that the Doppler shifts are small in comparison to the pulse bandwidth, which in turn should be much smaller than the carrier frequency. If both azimuth and elevation are to be estimated, then the antenna array must have a two- (or three-)dimensional configuration.

As mentioned, the angle/delay estimation problem is in fact a radar problem, and as such, many algorithms have already been proposed. Here, we are interested in high-resolution joint estimation algorithms, which work in situations where the number of parameters to be estimated

is relatively small. Joint estimation can resolve a larger number of rays than 1-D estimation, and is preferred to avoid exceptions and maintain resolution in cases where two or more rays have equal DOAs or delays. Various approaches to similar joint estimation problems with known pulse shape have recently been proposed [5–10]. These approaches often require computationally unattractive ML searches and/or need accurate initial points, and do not always work properly for rays with nearly equal directions or delays. The method proposed by Ogawa e.a. is a two-dimensional (windowed) MUSIC algorithm, the method by Wax e.a. [8] performs a successive ML optimization for an increasing number of rays, using lower-order results as initial points. The method by Swindlehurst e.a. [6, 7] that is applicable to our scenario consists of an iterative ML scheme (IQML) which requires initialization. For this, an ESPRIT-type harmonic retrieval algorithm was proposed, which bears some resemblance to the algorithm in this paper, but it is non-joint and restricts the total number of rays to be less than the number of antennas.

The algorithm we develop herein transforms the data by a DFT and a deconvolution by the known pulse shape function, which maps delays into phase shifts in the frequency domain. This is of course a classical approach and has been considered e.g., in [6–8] as well. New here is the observation that by stacking the result into a Hankel matrix, the problem is reduced to one that can be solved using 2-D ESPRIT techniques [11, 12], which were developed for joint azimuth-elevation estimation. Thus, the algorithm is closed-form and computationally attractive, and angles and delays are jointly estimated and automatically paired. Many of the tricks developed for ESPRIT and DOA estimation, such as forward-backward averaging, spatial smoothing [13], and real processing [12, 14], are readily incorporated into the current algorithm. The number of rays may be larger than the number of antennas, which overcomes a limitation of the non-joint 1-D ESPRIT method mentioned in [6] for initialization of a joint iterative ML optimization.

A second difference to several other approaches is that we propose to first estimate the channel matrix and subsequently deconvolve the pulse-shape function via the DFT, rather than directly deconvolve the observed modulated data. This should lead to better results if the number of samples is small because there are no edge effects. Finally, the algorithm has an elegant extension to the estimation of delays and both azimuth and elevation angles. This results in a joint diagonalization problem of *three* matrices. Similar generalizations occur if we have a non-uniform array with multiple baselines.

The structure of the paper is as follows. The data and channel model are described in section II. Section III contains a detailed derivation of the basic steps of the algorithm, including various processing techniques to improve the numerical accuracy and to deal with closely spaced rays. Identifiability of the DOAs and delays using the proposed technique is addressed in section IV, while section V briefly explores how the algorithm can be adapted to use antenna arrays other than ULAs. The Cramer-Rao bound

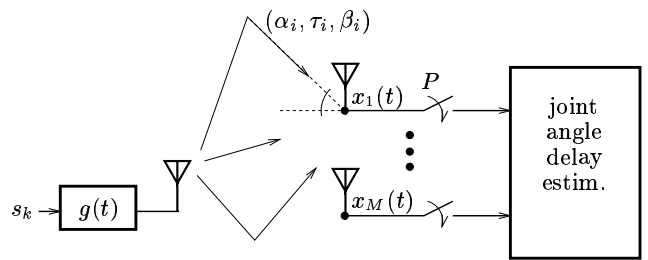


Fig. 1. Multiray propagation channel

of the estimates is given in section VI, while section VII illustrates the performance using computer simulations.

Notation

Vectors are denoted by boldface, matrices by capitals. $*$ denotes matrix complex conjugate transpose, \dagger is the matrix pseudo-inverse (Moore-Penrose inverse), $(\cdot)^c$ denotes complex conjugate. I_m is the $m \times m$ identity matrix, 0_m a zero matrix with m columns and an appropriate number of rows. \otimes is the Kronecker product, \diamond is the “Khatri-Rao” product, which is a column-wise Kronecker product: $A \diamond B = [\mathbf{a}_1 \otimes \mathbf{b}_1 \quad \mathbf{a}_2 \otimes \mathbf{b}_2 \quad \dots]$.

II. DATA MODEL

A. Channel model estimation

We derive a data model for the reception of a single source in a multipath scenario. Assume that we transmit a digital sequence $\{s_k\}$ over a channel, and measure the response using M antennas (cf. figure 1).¹ The noiseless received data in general has the form

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ \vdots \\ x_M(t) \end{bmatrix} = \sum_k s_k \mathbf{h}(t - kT) \quad (1)$$

where T is the symbol rate, which will be normalized to $T = 1$ from now on. A commonly used multiray propagation model, for specular multipath, writes the $M \times 1$ impulse response as

$$\mathbf{h}(t) = \sum_{i=1}^r \mathbf{a}(\alpha_i) \beta_i g(t - \tau_i) \quad (2)$$

where $g(t)$ is a known pulse shape function by which $\{s_k\}$ is modulated. In this model, there are r distinct propagation paths, each parameterized by a triple $(\alpha_i, \tau_i, \beta_i)$, where α_i is the direction-of-arrival (DOA), τ_i is the path delay, and $\beta_i \in \mathbb{C}$ is the complex path attenuation (fading). The vector-valued function $\mathbf{a}(\alpha)$ is the array response vector for an array of M antenna elements to a signal from direction α .

Suppose that $\mathbf{h}(t)$ has finite duration and is zero outside an interval $[0, L)$, where L is the (integer) channel

¹The assumption of digital sources is not at all essential but chosen because it gives a useful normalization to several parameters in the time/frequency domain.

length measured in symbol periods. We assume that the received data $\mathbf{x}(t)$ is sampled at a rate of P times the symbol rate. Using either training sequences (known $\{s_k\}$) or blind channel estimation techniques (e.g., [15, 16]), it is possible to estimate $\mathbf{h}(k)$, $k = 0, \frac{1}{P}, \dots, L - \frac{1}{P}$, at least up to a scalar. Specifically, suppose we start sampling at $t = 0$ and collect samples of $\mathbf{x}(t)$ during N symbol periods, then (1) implies that

$$X = HS,$$

where

$$X := \begin{bmatrix} \mathbf{x}(0) & \mathbf{x}(1) & \cdots & \mathbf{x}(N-1) \\ \mathbf{x}(\frac{1}{P}) & \mathbf{x}(1 + \frac{1}{P}) & & \vdots \\ \vdots & \vdots & & \vdots \\ \mathbf{x}(1 - \frac{1}{P}) & \mathbf{x}(2 - \frac{1}{P}) & \cdots & \mathbf{x}(N - \frac{1}{P}) \\ \mathbf{h}(0) & \mathbf{h}(1) & \cdots & \mathbf{h}(L-1) \\ \mathbf{h}(\frac{1}{P}) & \mathbf{h}(1 + \frac{1}{P}) & & \vdots \\ \vdots & \vdots & & \vdots \\ \mathbf{h}(1 - \frac{1}{P}) & \mathbf{h}(2 - \frac{1}{P}) & \cdots & \mathbf{h}(L - \frac{1}{P}) \end{bmatrix},$$

$$H := \begin{bmatrix} s_0 & s_1 & \cdots & s_{N-1} \\ s_{-1} & s_0 & s_1 & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ s_{-L+1} & s_{-L+2} & \cdots & s_{N-L} \end{bmatrix}.$$

If $\{s_k\}$ is known for $k = -L + 1, \dots, N - 1$, and $N \geq L$, then a least-squares estimate of H is $H = XS^\dagger$, where $S^\dagger = (SS^*)^{-1}S^*$. If L is unknown, then it can be estimated from the rank of X if $MP < L$, or else from the rank of a Hankel matrix constructed from X (see [15]). As long as $\{s_k\}$ is known, it is safe to overestimate L as this will simply extend H by zero columns. Alternatively, if $\{s_k\}$ is unknown, then blind channel estimation techniques can estimate L and both H and S , up to a scalar, by using the Toeplitz structure of S , under certain additional conditions that are not of interest to us here.

It is at this point convenient to recollect the estimated impulse response samples into an $M \times PL$ matrix,

$$H' := [\mathbf{h}(0) \quad \mathbf{h}(\frac{1}{P}) \quad \cdots \quad \mathbf{h}(L - \frac{1}{P})].$$

The channel model (2) can then be written as

$$H' = [\mathbf{a}_1 \quad \cdots \quad \mathbf{a}_r] \begin{bmatrix} \beta_1 & & 0 \\ & \ddots & \\ 0 & & \beta_r \end{bmatrix} \begin{bmatrix} \mathbf{g}_1 \\ \vdots \\ \mathbf{g}_r \end{bmatrix} =: ABG \quad (3)$$

where $\mathbf{a}_i = \mathbf{a}(\alpha_i)$, and $\mathbf{g}_i = [g(k - \tau_i)]_{k=0,1/P,\dots,L-1/P}$ is an LP -dimensional row vector containing the samples of $g(t - \tau_i)$.

B. Channel model transformation

In the delay estimation algorithm, we make use of the fact that a Fourier transform maps a delay to a certain

phase progression. Collect the samples of the known waveform $g(t)$ into a row vector

$$\mathbf{g} = [g(0) \quad g(\frac{1}{P}) \quad \cdots \quad g(L - \frac{1}{P})]$$

and let $\tilde{\mathbf{g}} = \mathbf{g}F$ where F denotes the DFT matrix of size $LP \times LP$, defined by

$$F := \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & \phi & \cdots & \phi^{LP-1} \\ \vdots & \vdots & & \vdots \\ 1 & \phi^{LP-1} & \cdots & \phi^{(LP-1)^2} \end{bmatrix}, \quad \phi = e^{-j\frac{2\pi}{LP}}. \quad (4)$$

If τ is an integer multiple of $\frac{1}{P}$, then it is straightforward to see that the Fourier transform $\tilde{\mathbf{g}}_\tau$ of the sampled version of $g(t - \tau)$ is given by

$$\tilde{\mathbf{g}}_\tau = [1 \quad \phi^{\tau P} \quad (\phi^{\tau P})^2 \quad \cdots \quad (\phi^{\tau P})^{LP-1}] \text{diag}(\tilde{\mathbf{g}}). \quad (5)$$

The same holds true for any τ if $g(t)$ is bandlimited and sampled at or above the Nyquist rate. This is not in full agreement with the FIR assumption we made earlier, which requires that $g(t)$ is nonzero only in a finite interval $[0, L_g]$, so that $L = L_g + \lceil \max(\tau_i) \rceil$ is finite. Because of this truncation, the spectrum of $g(t)$ widens and sampling at a rate $\frac{1}{P}$ introduces some aliasing due to spectral folding. This gives extra terms in (5) that will eventually lead to a bias in the delay estimate. In typical situations, however, the extra terms are small. For example, for $P = 2$, a raised-cosine pulse shape with roll-off factor (excess bandwidth) $\rho = 0.10$ truncated at a length $L_g = 4$ leads to a model mismatch in (5) of maximally 6% for any τ , with a corresponding delay estimation error of less than 0.002. The error becomes even smaller for larger P , L_g or ρ . Hence, in comparison to estimation errors that will occur in the presence of noise, this bias will not be of any significance.

Thus, we will assume that $g(t)$ is bandlimited and sampled at such a rate that (5) is "valid" even if τ is not an integer multiple of $\frac{1}{P}$. We can then write the Fourier-transformed data model

$$\tilde{H} := H'F$$

as

$$\tilde{H} = ABF \text{diag}(\tilde{\mathbf{g}}),$$

where F is the following Vandermonde matrix:

$$F_{LP} := \begin{bmatrix} 1 & \phi_1 & \phi_1^2 & \cdots & \phi_1^{LP-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \phi_r & \phi_r^2 & \cdots & \phi_r^{LP-1} \end{bmatrix}$$

$$\phi_i := \phi^{\tau_i P} = e^{-j\frac{2\pi}{LP}\tau_i}.$$

(With abuse of notation, we usually omit the size index of F : although it will vary at several places, its size is always clear from context.)

The next step is to do a deconvolution of $g(t)$ by dividing \tilde{H} by $\text{diag}(\tilde{\mathbf{g}})$. Obviously, this can be done only on intervals where $\tilde{\mathbf{g}}$ is nonzero. To be specific, assume that $g(t)$ is

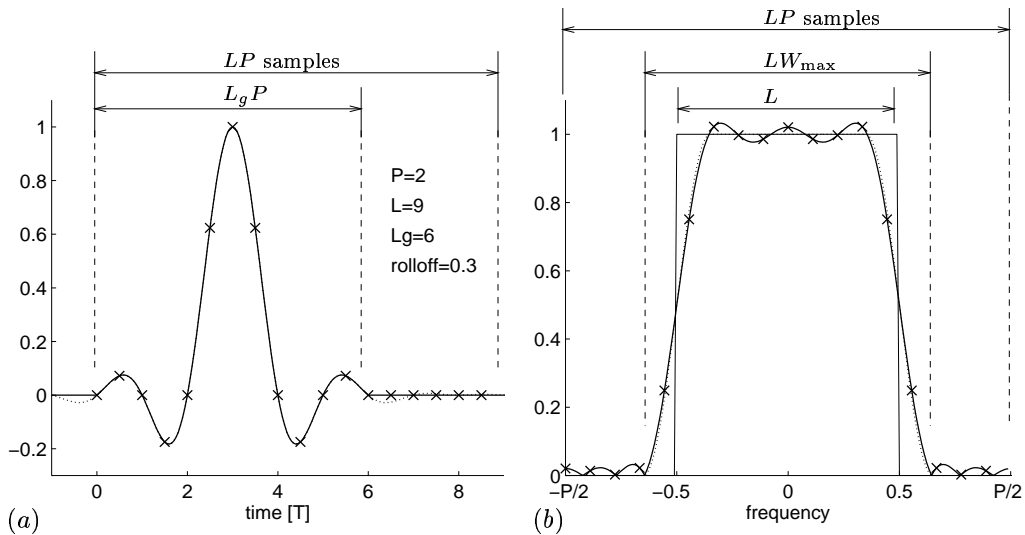


Fig. 2. Definition of parameters: (a) time domain, (b) frequency domain.

bandlimited with normalized bandwidth W_{\max} (that is, its Fourier transform is nonzero only for angular frequencies $|\omega| \leq \frac{1}{2}W_{\max}$), and assume that $P > W_{\max}$. Then $\tilde{\mathbf{g}}$ has at most LW_{\max} nonzero entries, and we can limit our attention to this interval. For a raised-cosine pulse shape with roll-off factor (excess bandwidth) ρ , we have $W_{\max} = 1 + \rho$, see figure 2. Usually, however, we would select a somewhat smaller number, W say, since the entries at the border can be relatively small as well, and their inversion can blow up the noise. Indeed, in the case of a raised-cosine pulse, we advise to set $W = 1$ and select only the L center frequency samples.

Let $J_{\tilde{\mathbf{g}}} : LP \times LW$ be the corresponding selection matrix for $\tilde{\mathbf{g}}$. For later use, we require that the selected frequencies appear in increasing order, which with the definition of the DFT in (4) usually means that the final $\lceil LW/2 \rceil$ samples of $\tilde{\mathbf{g}}$ should be moved up front: $J_{\tilde{\mathbf{g}}}$ has the form

$$J_{\tilde{\mathbf{g}}} = \begin{bmatrix} 0 & I_{\lceil LW/2 \rceil} \\ 0 & 0 \\ I_{\lceil LW/2 \rceil} & 0 \end{bmatrix} : LP \times LW.$$

If there are no other (intermittent) zero entries, we can factor $\text{diag}(\tilde{\mathbf{g}}J_{\tilde{\mathbf{g}}})$ out of $\tilde{H}J_{\tilde{\mathbf{g}}}$ and obtain

$$\tilde{H} := \tilde{H} J_{\tilde{\mathbf{g}}} \{\text{diag}(\tilde{\mathbf{g}}J_{\tilde{\mathbf{g}}})\}^{-1}, \quad (M \times LW) \quad (6)$$

which satisfies the model

$$\tilde{H} = ABF. \quad (7)$$

If the number of multipaths is not larger than the number of antennas ($r \leq M$), then it is possible to estimate the ϕ_i 's and hence the delays $\{\tau_i\}$ from the shift-invariance structure of F , independent of the structure of A . This is essentially the ESPRIT algorithm [17] as applied to harmonic analysis. However, in general, the number of antennas is limited and might not satisfy the condition $M > r$.

We can avoid this problem by constructing a Hankel matrix out of \tilde{H} . It then becomes sufficient to have $M = 2$, as we will explain in section III.

To estimate the DOAs $\{\alpha_i\}$, we need to know the array manifold structure. For simplicity, we will assume a uniform linear array (ULA) consisting of omnidirectional elements with equal interelement spacings, so that we can use the ESPRIT algorithm to estimate the DOAs as well. Any other array configuration on which the ESPRIT algorithm works can be used here. A two-dimensional configuration is considered in section V below.

It is clear that angles and delays can be estimated independently of each other, by directly working on the rows and columns of \tilde{H} . However, this does not give a pairing of angles to corresponding delays, and might result in poor resolution for closely spaced angles and delays. The algorithm derived in section III provides a joint estimate, using ideas from 2-D DOA estimation (viz. [11, 12]).

C. Remarks

In the above deconvolution approach, we first estimate H from X and S , then do a DFT and divide out the pulse shape in the frequency domain. A small model mismatch occurs because of the spectral aliasing after truncation of the pulse shape (governed by L_g, ρ and P). Another method of deconvolution would be to do a Fourier transform directly on X and on the remodulated source $\sum s_k g(t - kT)$, and divide these to obtain \tilde{H} . This is the approach followed in [6]. This direct method is computationally cheaper and does not need an estimate of L , but the accuracy is limited by the fact that the first or last L symbols are not taken into account correctly. This effect is averaged out as more sample periods are taken into account.

Note that it is safe to overestimate L . This will extend $\mathbf{h}(t)$ with additional zero columns. After Fourier transformation, $\tilde{\mathbf{g}}$ has more nonzero samples than before, but

satisfies the same model: the new samples interpolate the old ones. Hence, (7) is still valid for a larger L .

III. JOINT ANGLE AND DELAY ESTIMATION

A. Algorithm outline

Our objective is to estimate $\{(\alpha_i, \tau_i)\}$ from the shift-invariance properties present in the data model $\bar{H} = ABF$. We first outline the procedure and then introduce improvements to arrive at the final algorithm.

Let us assume that our antenna array is a uniform linear array consisting of M omnidirectional antennas spaced at a distance of Δ wavelengths. Then $A = A_\theta$, where

$$\Theta = \text{diag}[\theta_1 \cdots \theta_r], \quad A_\theta = \begin{bmatrix} 1 & \cdots & 1 \\ \theta_1 & \cdots & \theta_r \\ \vdots & & \vdots \\ \theta_1^{M-1} & \cdots & \theta_r^{M-1} \end{bmatrix},$$

$$\theta_i := e^{j2\pi\Delta \sin \alpha_i}.$$

In analogy, we can define (for some m)

$$\Phi = \text{diag}[\phi_1 \cdots \phi_r], \quad A_\phi = \begin{bmatrix} 1 & \cdots & 1 \\ \phi_1 & \cdots & \phi_r \\ \vdots & & \vdots \\ \phi_1^{m-1} & \cdots & \phi_r^{m-1} \end{bmatrix},$$

$$\phi_i = e^{-j\frac{2\pi}{L}\tau_i}.$$

The main idea now is as follows. From \bar{H} , we construct a Hankel matrix \mathcal{H} by left-shifting and stacking m copies of \bar{H} . In particular, for a $1 \leq i \leq m$, define the left-shifted matrix $\bar{H}^{(i)} := \bar{H}_{\bullet, i \cdots LW-m+i}$. (The notation $\bullet, i \cdots j$ indicates taking columns i through j of a matrix.) Then we define \mathcal{H} by

$$\mathcal{H} := \begin{bmatrix} \bar{H}^{(1)} \\ \vdots \\ \bar{H}^{(m)} \end{bmatrix} \quad (mM \times LW - m + 1). \quad (8)$$

The motivation for this step is that \mathcal{H} has a factorization

$$\mathcal{H} = ABF, \quad \mathcal{A} := \begin{bmatrix} A_\theta \\ A_\theta \Phi \\ \vdots \\ A_\theta \Phi^{m-1} \end{bmatrix} = A_\theta \diamond A_\phi \quad (9)$$

where \diamond denotes the Khatri-Rao product, i.e., a column-wise Kronecker product. Hence, if we can choose the stacking parameter m such that both $mM \geq r$ and $LW - m + 1 \geq r$, and if all factors are full rank, then \mathcal{H} has rank r , which means that we can estimate \mathcal{A} up to an $r \times r$ factor at the right. Detection of r is possible if there is an m such that \mathcal{H} becomes singular, which requires that at least one of these inequalities is a strict inequality.

The estimation of Φ and Θ from \mathcal{H} is based on exploiting the various shift-invariant structures present in \mathcal{A} . Define the selection matrices

$$J_{x\phi} := [I_{m-1} \ 0_1] \otimes I_M, \quad J_{x\theta} := I_m \otimes [I_{M-1} \ 0_1],$$

$$J_{y\phi} := [0_1 \ I_{m-1}] \otimes I_M, \quad J_{y\theta} := I_m \otimes [0_1 \ I_{M-1}],$$

To estimate Φ , we take submatrices consisting of the first and respectively last $M(m-1)$ rows of \mathcal{H} , i.e., $X_\phi = J_{x\phi}\mathcal{H}$, $Y_\phi = J_{y\phi}\mathcal{H}$, whereas to estimate Θ we stack, for all m blocks $\bar{H}^{(i)}$, its first and respectively last $M-1$ rows: $X_\theta = J_{x\theta}\mathcal{H}$, $Y_\theta = J_{y\theta}\mathcal{H}$. These data matrices have the structure

$$\begin{cases} X_\phi &= A_{x\phi}BF \\ Y_\phi &= A_{x\phi}\Phi BF \end{cases} \quad \begin{cases} X_\theta &= A_{x\theta}BF \\ Y_\theta &= A_{x\theta}\Theta BF \end{cases} \quad (10)$$

where $A_{x\phi} = J_{x\phi}\mathcal{A}$, $A_{x\theta} = J_{x\theta}\mathcal{A}$. If dimensions are such that these are low-rank factorizations, then we can apply the 2-D ESPRIT algorithm [11, 12] to estimate Φ and Θ . In particular, since

$$\begin{aligned} Y_\phi - \lambda X_\phi &= A_{x\phi}[\Phi - \lambda I_r]BF \\ Y_\theta - \lambda X_\theta &= A_{x\theta}[\Theta - \lambda I_r]BF \end{aligned}$$

the ϕ_i are given by the rank reducing numbers of the pencil (Y_ϕ, X_ϕ) , whereas the θ_i are the rank reducing numbers of (Y_θ, X_θ) . These are the same as the nonzero eigenvalues of $X_\phi^\dagger Y_\phi$ and $X_\theta^\dagger Y_\theta$.

The correct pairing of each ϕ_i with its corresponding θ_i follows from the fact that $X_\phi^\dagger Y_\phi$ and $X_\theta^\dagger Y_\theta$ have the same eigenvectors, which is caused by the common factor F (the eigenvectors are in fact scalings of the columns of F^\dagger). In particular, these matrices satisfy a model of the form

$$\begin{aligned} X_\phi^\dagger Y_\phi &= T^{-1}\Phi T \\ X_\theta^\dagger Y_\theta &= T^{-1}\Theta T \end{aligned}$$

The challenge of joint estimation is to find a matrix T that best diagonalizes *both* matrices. Various (suboptimal) approaches are possible, which are described in section III-E.

B. Data extensions

If two rays have the same delays, then F in (9) becomes rank deficient. As a result, the rank of \mathcal{H} is $r-1$ instead of r . The two corresponding \mathcal{A} -vectors will be combined and the angles cannot be identified correctly. This is entirely similar to the problem with coherent signals in the usual DOA problem, where it was solved using ‘‘spatial smoothing’’. This technique can be nicely integrated with our approach, as follows.

For integers $2 \leq m_1 \leq LW$, $1 \leq m_2 \leq M-1$, define the following equal-sized submatrices

$$\bar{H}^{(i,j)} = \begin{bmatrix} \bar{H}_{i,j} & \cdots & \bar{H}_{i,LW-m_1+j} \\ \vdots & & \vdots \\ \bar{H}_{M-m_2+i,j} & \cdots & \bar{H}_{M-m_2+i,LW-m_1+j} \end{bmatrix}$$

$$1 \leq i \leq m_2, \quad 1 \leq j \leq m_1.$$

As a generalization of (8), redefine \mathcal{H} as

$$\mathcal{H} = \begin{bmatrix} \bar{H}^{(1,1)} & \cdots & \bar{H}^{(m_2,1)} \\ \vdots & & \vdots \\ \bar{H}^{(1,m_1)} & \cdots & \bar{H}^{(m_2,m_1)} \end{bmatrix}$$

$$m_1(M - m_2 + 1) \times m_2(LW - m_1 + 1).$$

\mathcal{H} has a factorization

$$\mathcal{H} = \mathcal{A}\mathcal{B}\mathcal{F} := \begin{bmatrix} A_\theta \\ A_\theta\Phi \\ \vdots \\ A_\theta\Phi^{m_1-1} \end{bmatrix} B[F \quad \Theta F \quad \dots \quad \Theta^{m_2-1}F]. \quad (11)$$

(A_θ is similar as before, but now has only $M - m_2 + 1$ rows.) The computation of Φ and Θ from \mathcal{H} proceeds as before. If $m_2 = 2$, then we can have two rays with equal delays and still \mathcal{F} will be full rank. More in general, if d rays can have equal delays, then we need to set $m_2 \geq d$ to ensure the rank of \mathcal{F} . Naturally, this will reduce the number of rows of A_θ accordingly. To be able to estimate the corresponding angles using shift-invariance of \mathcal{A} , we need A_θ to have at least $d + 1$ rows, because the multiplication by A_ϕ does not add resolution in this case. Hence, $M \geq 2d$ antennas are needed (but see below).

A second technique to extend the data matrix is known as forward-backward averaging and uses the fact that the eigenvalues (ϕ_i, θ_i) are on the unit circle, along with the symmetric structures of A_θ and A_ϕ . Let J denote the exchange matrix which reverses the ordering of rows, and define

$$\mathcal{H}_e = [\mathcal{H} \quad J\mathcal{H}^{(c)}], \quad m_1(M - m_2 + 1) \times 2m_2(LW - m_1 + 1), \quad (12)$$

where (c) indicates complex conjugate. Since $J\mathcal{A}^{(c)} = (JA_\phi^{(c)}) \diamond (JA_\theta^{(c)}) = (A_\phi\Phi^{-(m_1-1)}) \diamond (A_\theta\Theta^{-(M-m_2)}) = \mathcal{A}\Phi^{-m_1+1}\Theta^{-M+m_2}$, it follows that \mathcal{H}_e has a factorization

$$\mathcal{H}_e = \mathcal{A}\mathcal{F}_e := \mathcal{A}[B\mathcal{F}, \quad \Phi^{-m_1+1}\Theta^{-M+m_2}B^{(c)}\mathcal{F}^{(c)}]. \quad (13)$$

The computation of Φ and Θ proceeds as before. Thus, we can double the number of columns of the data, which gives a significant improvement in accuracy. It also provides some protection against loss of rank in case of equal delays: even if $m_2 = 1$, the multiplication by $\Theta^{-M+m_2}B^{(c)}$ ensures that usually we can tolerate two rays with equal delays. It is not entirely sufficient that the two angles are different, because of the phase aliasing that occurs in Θ^{-M+m_2} . However, since also B and $B^{(c)}$ play a role (this includes initial phase offsets), we may assume in practice that \mathcal{F}_e has full rank if two delays are the same. In general, to be able to identify d rays with equal delays, we need $m_2 \geq \frac{1}{2}d$ and $M - m_2 \geq d$, i.e., $M \geq \frac{3}{2}d$.

We will assume from now on that m_1 and m_2 are selected such that we can identify all r rays. The implied conditions on M and LW are discussed in section IV.

C. Rank reduction

In the presence of noise, \mathcal{H}_e will be of full rank rather than rank r . As most subspace-based parameter estimation algorithms, we first have to reduce \mathcal{H}_e to its r -dimensional principal column span. This reduction constitutes the main computational expense, but is necessary because noise increases the rank of the matrix pencils in (10), and thus spurious eigenvalues are introduced. To avoid this, it is

standard practice to modify the algorithm such that the pencil problem involves matrices of size $r \times r$.

There are several techniques to do the rank reduction. The most accurate is perhaps to ensure that $X_\phi, Y_\phi, X_\theta, Y_\theta$ in (10) have the same row span $\text{row}(F)$, and pairwise the same column span $\text{col}(A_{x\phi})$ and $\text{col}(A_{x\theta})$. This corresponds to a Total Least Squares solution, and requires three subspace estimates. Computationally less demanding, and almost equally accurate in practice, is the following LS procedure which forces a common row span but implicitly projects Y_ϕ and Y_θ onto the column spans of X_ϕ and X_θ , respectively.

Let

$$\mathcal{H}_e = U\Sigma V^*$$

be a singular value decomposition (SVD) of \mathcal{H}_e , where U and V are unitary matrices, and Σ is a diagonal matrix containing the singular values of \mathcal{H}_e in nonincreasing order [18]. The number of rays r can be estimated from this SVD as the number of singular values larger than a certain noise level. Let \hat{U} be the first r columns of U : it forms a basis of the estimated r -dimensional column span of \mathcal{A} : without noise, we have $\hat{U} = AT$, for some non-singular $r \times r$ matrix T .

Define

$$U_{x\phi} = J_{x\phi}\hat{U}, \quad U_{y\phi} = J_{y\phi}\hat{U}, \quad U_{x\theta} = J_{x\theta}\hat{U}, \quad U_{y\theta} = J_{y\theta}\hat{U}.$$

These matrices have noise-perturbed models $U_{x\phi} = A_{x\phi}T$, $U_{y\phi} = A_{y\phi}\Phi T$, etc. To arrive at an $r \times r$ pencil problem, the number of rows of each of these matrices has to be reduced to r as well. Again, this can be done in various ways. A numerically pleasing way is to compute a QR factorization of $U_{x\phi}$, and to apply the Q -factor to $U_{y\phi}$ as well: this corresponds to a LS projection of the column span of $U_{y\phi}$ onto that of $U_{x\phi}$. Thus introduce the QR-factorizations

$$Q'^*[U_{x\phi} \quad U_{y\phi}] =: \begin{bmatrix} r & r \\ E_{x\phi} & E_{y\phi} \\ 0 & * \end{bmatrix},$$

$$Q''*[U_{x\theta} \quad U_{y\theta}] =: \begin{bmatrix} r & r \\ E_{x\theta} & E_{y\theta} \\ 0 & * \end{bmatrix}.$$

The four original data matrices have now been reduced to equivalent $r \times r$ data matrices, satisfying

$$\begin{cases} E_{x\phi} = S'T \\ E_{y\phi} = S'\Phi T \end{cases} \quad \begin{cases} E_{x\theta} = S''T \\ E_{y\theta} = S''\Theta T \end{cases} \quad (14)$$

for certain nonsingular $r \times r$ matrices S' , S'' , T .

D. Real processing

Similar as in [12, 14], we can use the structure of \mathcal{H}_e to do a transformation to a real matrix, which allows to keep the SVD of \mathcal{H}_e and all subsequent operations in the real domain, with obvious computational and numerical advantages. This is possible because every entry in \mathcal{H}_e can be

combined with its complex conjugate by simple unitary column and row transformations which do not depend on the data. The details are in [12, 14] and need not be repeated here.

Suffice it to say that we can transform \mathcal{H}_e to a real matrix and compute a real-transformation of \hat{U} from this matrix. We can also transform $(E_{x\phi} + E_{y\phi})$ and $j(E_{x\phi} - E_{y\phi})$ to real matrices $\tilde{E}_{x\phi}$ and $\tilde{E}_{y\phi}$, and compute these in terms of the real-transformed \hat{U} using transformed selection matrices and QR factorizations as before. The real-valued matrices that we obtain this way satisfy the following model:

$$\begin{cases} \tilde{E}_{x\phi} &= S'T \\ \tilde{E}_{y\phi} &= S'\tilde{\Phi}T \end{cases} \quad \begin{cases} \tilde{E}_{x\theta} &= S''T \\ \tilde{E}_{y\theta} &= S''\tilde{\Theta}T \end{cases} \quad (15)$$

for certain new real-valued non-singular $r \times r$ matrices S' , S'' , T . Here, $\tilde{\Phi} = j(I + \Phi)^{-1}(I - \Phi)$, $\tilde{\Theta} = j(I + \Theta)^{-1}(I - \Theta)$: the generalized eigenvalues are real-valued and in fact the Cayley transformations of $\{\phi_i\}$, $\{\theta_i\}$.

E. Joint diagonalization

The final step of the algorithm is to estimate Φ and Θ from (14) or (15). In essence, we have to find a square invertible matrix T to simultaneously diagonalize two matrices,

$$\begin{aligned} E_{x\phi}^{-1}E_{y\phi} &= T^{-1}\Phi T \\ E_{x\theta}^{-1}E_{y\theta} &= T^{-1}\Theta T \end{aligned} \quad (16)$$

and similar for the real-valued decompositions. The fact that the same matrix T is used gives a coupling between the two eigenvalue problems: the i -th entry of $\tilde{\Phi}$ corresponds to the i -th entry of $\tilde{\Theta}$. Numerically more favorable is to determine Schur decompositions:

$$\begin{aligned} E_{x\phi}^{-1}E_{y\phi} &= QR_{\Phi}Q^* \\ E_{x\theta}^{-1}E_{y\theta} &= QR_{\Theta}Q^* \end{aligned} \quad (17)$$

where Q is a unitary matrix and R_{Φ} and R_{Θ} are upper triangular. $\tilde{\Phi}$ and $\tilde{\Theta}$ are given by the diagonals of these matrices.

A problem with using the Cayley transformation of unit-modulus to real-valued eigenvalues is that eigenvalues close to -1 are mapped to $\pm\infty$. This occurs for relatively large delays τ_i , or angles α_i close to $\pm 90^\circ$ for half-wavelength spaced antenna elements. Hence, unlike the complex-valued case, $\tilde{E}_{x\phi}$ and $\tilde{E}_{x\theta}$ can be badly conditioned. In view of this, it is numerically advisable *not* to form $(\tilde{E}_{x\phi}^{-1}\tilde{E}_{y\phi})$ and $(\tilde{E}_{x\theta}^{-1}\tilde{E}_{y\theta})$, but to work on the original pencils and to compute the generalized Schur decompositions:

$$\begin{cases} \tilde{E}_{x\phi} &= Q'R_{x\phi}Z \\ \tilde{E}_{y\phi} &= Q'R_{y\phi}Z \end{cases} \quad \begin{cases} \tilde{E}_{x\theta} &= Q''R_{x\theta}Z \\ \tilde{E}_{y\theta} &= Q''R_{y\theta}Z \end{cases} \quad (18)$$

where Q' , Q'' and Z are unitary and all R -matrices are upper triangular. After these generalized Schur decompositions have been determined, we obtain estimates of the parameter values by setting $\tilde{\Phi} = \text{diag}(R_{y\phi})\text{diag}^{-1}(R_{x\phi})$, $\tilde{\Theta} = \text{diag}(R_{y\theta})\text{diag}^{-1}(R_{x\theta})$.

A number of techniques have been proposed to solve such problems in the presence of noise. All of these are suboptimal.

Method “T”

The method proposed in [12] is the easiest to describe. After real-transformations, both $\tilde{\Phi}$ and $\tilde{\Theta}$ are real. Hence, we can diagonalize

$$(\tilde{E}_{x\phi}^{-1}\tilde{E}_{y\phi}) + j(\tilde{E}_{x\theta}^{-1}\tilde{E}_{y\theta}) = T^{-1}(\tilde{\Phi} + j\tilde{\Theta})T.$$

(Alternatively, we can work with the Schur decomposition in (17).) Thus, the real part of the eigenvalues gives $\tilde{\Phi}$, the imaginary part gives $\tilde{\Theta}$. This method works usually fine and guarantees that only a single T is used. The problem is that it does not guarantee that T is a real matrix. Hence, in critical cases it might happen that T becomes complex, so that $T(\tilde{E}_{x\phi}^{-1}\tilde{E}_{y\phi})T^{-1}$ is complex. The imaginary part of this term gives a contribution to the estimate of $\tilde{\Theta}$. Similarly, the real part of $jT(\tilde{E}_{x\theta}^{-1}\tilde{E}_{y\theta})T^{-1}$ gives a contribution to $\tilde{\Phi}$. A second problem is that the method cannot be extended to cases with more than two matrices, as occurs e.g., in the joint estimation of delays and both azimuth and elevation. Finally, it is not easily modified to work with generalized eigenvalues: clearly, it is not correct to compute a generalized eigenvalue decomposition of $(\tilde{E}_{x\phi} + j\tilde{E}_{x\theta}, \tilde{E}_{y\phi} + j\tilde{E}_{y\theta})$.

Method “Q”

A second method is described in [11] and was later called ACMP (“algebraically coupled matrix pencil”). It works with the Schur decomposition (17) or generalized Schur decomposition (18). In essence, the method first computes a Schur decomposition of the first matrix: $Q(E_{x\phi}^{-1}E_{y\phi})Q^* = R_{\Phi}$. When we apply Q to $(E_{x\theta}^{-1}E_{y\theta})$, this produces in the presence of noise an “almost-upper” matrix, approximately equal to R_{Θ} . The method then continues to compute the exact eigenvalue decomposition of the second matrix using 2×2 Jacobi rotations [19]. Since the matrix is already close to upper, each Jacobi rotation is either close to an identity matrix or to a permutation. This is easily detected, and in the latter case, a permutation is also applied to the estimate of Φ . The algorithm works well, although it would fail in the theoretical case of *exactly* repeated eigenvalues of the first matrix. It is easily extended to more than two matrix pencils. The method has as feature that the true eigenvalues of each matrix pencil are obtained as parameter estimates. It is suboptimal, because slightly *different* matrices are used to diagonalize each pencil.

An alternative but similar way to couple the two eigenvalue problems is to independently compute diagonalizing matrices T_{θ} and T_{ϕ} for $\tilde{E}_{x\theta}^{-1}\tilde{E}_{y\theta}$ and $\tilde{E}_{x\phi}^{-1}\tilde{E}_{y\phi}$ respectively, and follow the observation that $T_{\theta}^{-1}T_{\phi}$ should be close to a permutation matrix (in the case of distinct eigenvalues).

Method “QZ”

A third method is similar to the one proposed in [20] and can be called a “super”-generalized Schur method, as it tries to compute a QZ (Schur) decomposition for more than two matrix pencils. It is an attempt to find unitary Q' , Q'' and Z to make all four E -matrices in (18) as much upper triangular as possible, by a straightforward extension

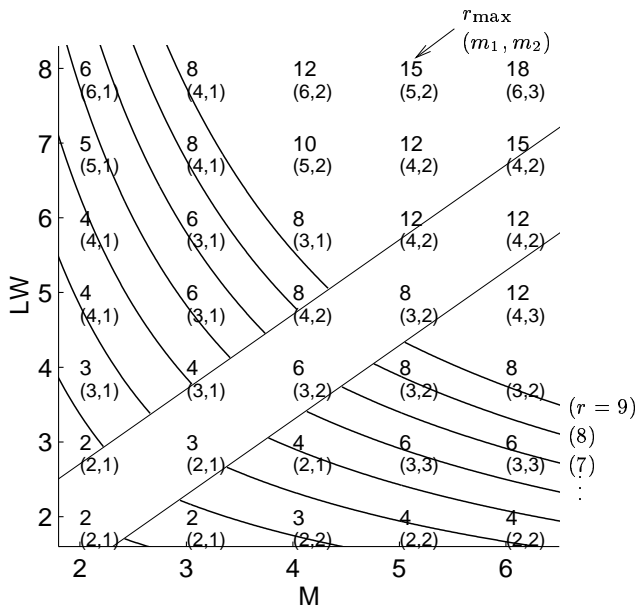


Fig. 3. Maximal number of rays that can be identified. The curved lines are the analytically computed contours of the maxima, assuming continuous parameters

of the usual QZ iteration [18]. This leads to the following algorithm outline (cf. [20]):

Initialize $Z^{(0)}$ by a Schur decomposition of $E_{x\phi}$, $E_{y\phi}$ for $k = 1, 2, \dots$

- Find $Q^{(k)}$ (unitary) to minimize $\|Q^{(k)}E_{x\phi}Z^{(k-1)}\|_{LF}^2 + \|Q^{(k)}E_{y\phi}Z^{(k-1)}\|_{LF}^2$
- Find $Q''^{(k)}$ (unitary) to minimize $\|Q''^{(k)}E_{x\theta}Z^{(k-1)}\|_{LF}^2 + \|Q''^{(k)}E_{y\theta}Z^{(k-1)}\|_{LF}^2$
- find $Z^{(k)}$ (unitary) to minimize $\|Q^{(k)}E_{x\phi}Z^{(k)}\|_{LF}^2 + \|Q^{(k)}E_{y\phi}Z^{(k)}\|_{LF}^2 + \|Q''^{(k)}E_{x\theta}Z^{(k)}\|_{LF}^2 + \|Q''^{(k)}E_{y\theta}Z^{(k)}\|_{LF}^2$

Here, $\|\cdot\|_{LF}$ denotes the Frobenius norm of the strictly lower triangular part of a matrix. It is hard to find the exact minimizers in every step, but this is also not necessary: we can find good approximate solutions and rely on the outer loop to provide convergence. There are several approximate solutions, e.g., based on Householder rotations [20] or Givens rotations [21].

It should be said that, unlike in [20], the convergence properties of this iteration have been disappointing so far. Without accurate starting point (as obtained by the initial Schur decomposition), the convergence can be very slow and usually stalls in a local minimum. This is probably because it corresponds to an *unshifted* QZ-iteration. With accurate starting point, the performance of this method is slightly better than the previous two methods, which indicates that there is some advantage in a truly joint estimation approach.

IV. IDENTIFIABILITY

To identify Φ and Θ from (13) and (10), necessary conditions are that the submatrices $A_{x\phi}$ and $A_{x\theta}$ of \mathcal{A} are “tall”,

while \mathcal{F}_e is “wide”, i.e.,

- $r \leq (m_1 - 1)(M - m_2 + 1)$
- $r \leq m_1(M - m_2)$
- $r \leq 2m_2(LW - m_1 + 1)$.

Subject to these conditions, we can try to maximize the number of rays that can be identified for given M and LW , by optimizing over m_1 and m_2 . This is analytically feasible only if we assume continuous parameters, which after some calculations then produces

$$\begin{aligned} \text{if } LW \geq M + \frac{1}{2}\sqrt{2} : & \begin{cases} r_{\max} = (LW + 1)M(2 - \sqrt{2})^2 \\ m_{1,\text{opt}} = (LW + 1)(2 - \sqrt{2}) \\ m_{2,\text{opt}} = M(\sqrt{2} - 1) \end{cases} \\ \text{if } LW \leq M - \frac{1}{2}\sqrt{2} : & \begin{cases} r_{\max} = LW(M + 1)(2 - \sqrt{2})^2 \\ m_{1,\text{opt}} = 1 + LW(2 - \sqrt{2}) \\ m_{2,\text{opt}} = (M + 1)(\sqrt{2} - 1) \end{cases} \end{aligned} \quad (19)$$

The first range corresponds to a region where conditions (a) and (c) are satisfied with equality, whereas the second range has conditions (b) and (c) satisfied with equality. There is a small third range in between, where all three conditions are satisfied with equality. Although this can be solved analytically, the equations are awkward and hence omitted. The actual maxima are slightly smaller because m_1 and m_2 can take only integer values: see figure 3. The general behavior however is quite accurately captured by the equations.

For identifiability, it is not sufficient that the dimensions of the factors $A_{x\phi}$, $A_{x\theta}$ and \mathcal{F}_e are at least r : they should also be of full rank r . The Vandermonde structure of these factors ensures that this is the case if all angles and all delays are different. As discussed before, if d delays are equal (but with different angles), then the corresponding columns of A_ϕ are all the same, so that the corresponding columns of $A_{x\theta}$ have to span a rank- d subspace: its size should be at least d rows. Hence we need $d \leq M - m_2$ to enable identification. A similar rank condition holds on \mathcal{F}_e : here we need $d \leq 2m_2$ to ensure the rank (almost—we assume that the $\theta_i^{-M+m_2} \beta_i^{(c)} \beta_i^{-1}$ are different). Altogether, this gives the following necessary and (almost) sufficient conditions for identifiability of r rays with at most d equal delays: (19) holds, and

$$\begin{cases} m_2 \geq \frac{1}{2}d \\ M \geq \frac{3}{2}d \end{cases}$$

With these additional constraints, can we still identify the same maximal number of rays? Suppose $M \geq \frac{3}{2}d$, then the smaller of the two $m_{2,\text{opt}}$ in (19) is $m_{2,\text{opt}} = (\sqrt{2} - 1)M \geq \frac{3}{2}(\sqrt{2} - 1)d > \frac{1}{2}d$. Hence, $m_{2,\text{opt}}$ can be reached and the maxima reported in figure 3 do not change provided $M \geq \frac{3}{2}d$.

Similarly, suppose d rays have equal angles but different delays. Then the corresponding columns of A_θ are the same. To ensure the rank of $A_{x\phi}$, we need $d \leq m_1 - 1$. For \mathcal{F}_e , we need $d \leq 2(LW - m_1 + 1)$. Hence, r rays with

at most d equal angles are identifiable if and (almost) only if (19) holds, and

$$\begin{cases} m_1 & \geq d+1 \\ LW & \geq \frac{3}{2}d \end{cases}$$

Suppose $LW \geq \frac{3}{2}d$, then the smaller of the two $m_{1,\text{opt}}$ in (19) is $m_{1,\text{opt}} = (2 - \sqrt{2})LW + 1 \geq (2 - \sqrt{2})\frac{3}{2}d + 1 > d + 1$. Again, we can still reach the same maximal number of rays as before, provided $LW \geq \frac{3}{2}d$.

One way to increase the maximal number of rays could be to increase the pulse shape length L_g , since $L \geq L_g + \max(\tau_i)$. Another way to increase L is to zero-pad $\mathbf{h}(t)$ prior to the DFT. This results in extra samples in the frequency domain that fit the same model as before and interpolate the previous samples. In principle, this means that even with $M = 2$ antennas we can estimate an arbitrarily large number of rays! In actuality, however, the amount of noise that can be tolerated becomes exceedingly small. The underlying problem is that increasing L in this way does not improve the conditioning of \mathcal{H}_e , whereas every additional ray within the same resolution cell deteriorates the conditioning of \mathcal{H}_e roughly by an order of magnitude. Very quickly, even the modeling errors (i.e., aliasing due to the FIR assumption on $g(t)$) destroy the required accuracy. As another way to view this, note that for increasing L the eigenvalues Φ are compressed to a decreasingly small sector of the unit circle.

V. OTHER ARRAY CONFIGURATIONS

A. Two-dimensional angle estimation

In this section we indicate an extension to delay plus both azimuth and elevation estimation using a two-dimensional antenna array. For simplicity of exposition, we will consider only one type of array, consisting of two ULAs oriented in two different directions, e.g., in an L-shape or a +-shape. Extensions to more general 2-D arrays on which the ESPRIT algorithm works are straightforward to derive, see e.g., [12]. The main issues are the preservation of shift-invariance properties, and the correct pairing of the estimated path parameters using a coupled eigenvalue method.

Thus consider a sensor array consisting of a ULA with M_1 elements spaced Δ_1 wavelengths in one direction, and a ULA of M_2 elements spaced Δ_2 wavelengths in a complementary direction. The two arrays should be close to each other, preferably in a centro-symmetric +-shape. The array response matrix A_θ is replaced by

$$A = \begin{bmatrix} A_\Theta \\ A_\Psi \end{bmatrix}, \quad A_\theta = \begin{bmatrix} 1 & \cdots & 1 \\ \theta_1 & \cdots & \theta_r \\ \vdots & & \vdots \\ \theta_1^{M_1-1} & \cdots & \theta_r^{M_1-1} \end{bmatrix}, \quad A_\psi = \begin{bmatrix} 1 & \cdots & 1 \\ \psi_1 & \cdots & \psi_r \\ \vdots & & \vdots \\ \psi_1^{M_2-1} & \cdots & \psi_r^{M_2-1} \end{bmatrix} \Psi_0.$$

where $\theta_i = e^{j2\pi\Delta_1 u_i}$ and $\psi_i = e^{j2\pi\Delta_2 v_i}$, with u_i and v_i are the direction cosine variables relative to the orientations of

the two arrays. Ψ_0 is a diagonal unimodular matrix that accounts for the phase offset between the first elements of both arrays.

In the construction of \mathcal{H} , we need to keep track of the partitioning of A . The shifts in frequency-domain are unaltered, but the spatial shifts lead to significant complications that we wish to avoid here. Hence, let us assume that only the frequency shifts are taken. With m shifts, this then leads to \mathcal{H} as in (8) before, satisfying the model²

$$\mathcal{H} = ABF, \quad A = A_\phi \diamond A. \quad (20)$$

Forward-backward averaging can be applied if the array is centro-symmetric. In that case, we can define

$$\mathcal{H}_e = [\mathcal{H} \quad J'\mathcal{H}^{(c)}], \quad J' = J_\phi \diamond \begin{bmatrix} J_\theta & 0 \\ 0 & J_\psi \end{bmatrix}.$$

The selection matrices need re-definition, as follows:

$$\begin{aligned} J_{x\phi} &:= [I_{m-1} \quad 0_1] \otimes I_{M_1+M_2}, \\ J_{y\phi} &:= [0_1 \quad I_{m-1}] \otimes I_{M_1+M_2}, \\ J_{x\theta} &:= I_m \otimes [I_{M_1-1} \quad 0_1 \quad 0_{M_2}], \\ J_{y\theta} &:= I_m \otimes [0_1 \quad I_{M_1-1} \quad 0_{M_2}], \\ J_{x\psi} &:= I_m \otimes [0_{M_1} \quad I_{M_2-1} \quad 0_1], \\ J_{y\psi} &:= I_m \otimes [0_{M_1} \quad 0_1 \quad I_{M_2-1}]. \end{aligned}$$

Proceeding similar to the ULA case, we let $X_\phi = J_{x\phi}\mathcal{H}_e$, $Y_\phi = J_{y\phi}\mathcal{H}_e$, etc. Due to the preserved shift-invariance structure, these data matrices then admit factorizations similar to (10):

$$\begin{cases} X_\phi = A_{x\phi}BF \\ Y_\phi = A_{y\phi}\Phi BF \\ X_\psi = A_{x\psi}BF \\ Y_\psi = A_{y\psi}\Psi BF \end{cases} \quad \begin{cases} X_\theta = A_{x\theta}BF \\ Y_\theta = A_{y\theta}\Theta BF \end{cases} \quad (21)$$

where $\Psi = \text{diag}[\psi_1 \dots \psi_r]$. The parameter triples $\{(\phi_i, \theta_i, \psi_i)\}$ are given by the rank-reducing numbers of each of the pencils (Y_ϕ, X_ϕ) , (Y_θ, X_θ) and (Y_ψ, X_ψ) respectively. Proceeding as before, we eventually we have to solve three coupled eigenvalue problems that share the same (right) eigenvectors. Either joint eigenvalue estimation method ‘‘Q’’ or ‘‘QZ’’ in section III-E is applicable.

B. General arrays with identical elements

Because the above even works for $M_1 = 2$ and $M_2 = 2$, the array geometry is essentially arbitrary. Indeed, for a general array, every available baseline pair generates a block in the A -matrix and can be used to estimate a direction cosine with respect to this baseline, provided the antenna elements of the pair are identical. If there are multiple identical baseline pairs (the array is redundant), then these can be combined in larger blocks.

To illustrate this with an example, consider a uniform hexagonal array (figure 4). For the construction

²If an array element is shared by both sub-arrays, then in the above equations it appears repeated. In computing the SVD of \mathcal{H} , the duplicate rows should be suppressed.

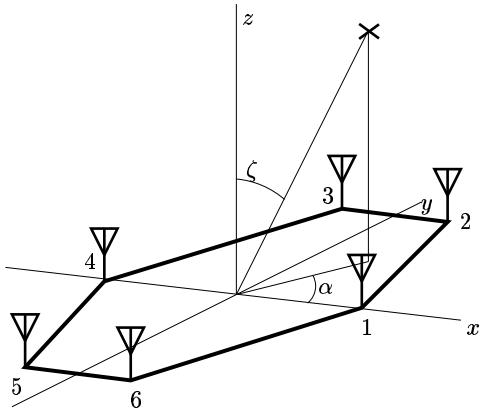


Fig. 4. Uniform circular array

of \mathcal{H} , we simply work with the array response vector $\mathbf{a} = [a_1 \cdots a_6]^T$ and its corresponding matrix A , which gives (20) as before. Since the array is centro-symmetric, we can define

$$J'' = \begin{bmatrix} 0_{3 \times 3} & I_3 \\ I_3 & 0_{3 \times 3} \end{bmatrix}$$

so that $J''\mathbf{a}^{(c)} = \mathbf{a}$, assuming zero phase at the center of the array. As before, we set

$$\mathcal{H}_e = [\mathcal{H} \quad J'\mathcal{H}^{(c)}], \quad J' = J \diamond J''.$$

The difference is in the definition of the selection matrices. There are 9 different baselines. The corresponding baseline blocks pairs are defined by

$$\begin{aligned} \mathbf{a}_{x1} &= \begin{bmatrix} a_1 \\ a_5 \end{bmatrix}, \quad \mathbf{a}_{y1} = \begin{bmatrix} a_2 \\ a_4 \end{bmatrix}, \quad \mathbf{a}_{x2} = \begin{bmatrix} a_2 \\ a_6 \end{bmatrix}, \quad \mathbf{a}_{y2} = \begin{bmatrix} a_3 \\ a_5 \end{bmatrix}, \\ \mathbf{a}_{x3} &= \begin{bmatrix} a_3 \\ a_1 \end{bmatrix}, \quad \mathbf{a}_{x3} = \begin{bmatrix} a_4 \\ a_6 \end{bmatrix}, \quad \mathbf{a}_{x4} = \begin{bmatrix} a_1 \\ a_6 \end{bmatrix}, \quad \mathbf{a}_{y4} = \begin{bmatrix} a_3 \\ a_4 \end{bmatrix}, \\ \mathbf{a}_{x5} &= \begin{bmatrix} a_2 \\ a_1 \end{bmatrix}, \quad \mathbf{a}_{y5} = \begin{bmatrix} a_4 \\ a_5 \end{bmatrix}, \quad \mathbf{a}_{x6} = \begin{bmatrix} a_3 \\ a_2 \end{bmatrix}, \quad \mathbf{a}_{y6} = \begin{bmatrix} a_5 \\ a_6 \end{bmatrix}, \\ \mathbf{a}_{x7} &= a_1, \quad \mathbf{a}_{y7} = a_4, \quad \mathbf{a}_{x8} = a_2, \quad \mathbf{a}_{y8} = a_5, \\ \mathbf{a}_{x9} &= a_3, \quad \mathbf{a}_{y9} = a_6. \end{aligned}$$

Along with a selection matrix for Φ , we finally obtain up to 10 coupled matrix pencils, all having the same right eigenvectors. Obviously, a truly joint procedure such as method “QZ” can now have definite advantages.

There are several issues remaining here. The number of rays that can be estimated is limited by the smallest submatrix that we take. If we discard baseline pairs 7 to 9, then we obtain $r \leq 2(LW - m_1 + 1)$. Since we didn't consider spatial smoothing but did do forward-backward averaging, we can have no more than two rays with equal delays. Finally, a discussion on how to combine the various phase difference estimates into a single azimuth-elevation estimate is omitted here. For each ray, it involves the least-squares estimation of the direction vector $[\sin \zeta \cos \alpha, \sin \zeta \sin \alpha]^T$, which is linearly related to the phase differences expressed in radians. Combinatorial problems due to aliasing arise if some baselines have a larger length than half the wavelength.

VI. CRAMER-RAO BOUND

The Cramer-Rao bound (CRB) provides a lower bound on the variance of any unbiased estimator. The CRB depends on whether the path fading is modeled as unknown deterministic quantities or as random variables with a known distribution. We first consider the deterministic assumption, which will produce a bound on the angle/delay estimates in the presence of random noise but constant (unknown) fading. In this case we do not have to assign a distribution to the fading. If we apply the “vec” operation to the noise-perturbed model in (3), and with some abuse of notation call the resulting vector “ \mathbf{h} ”, we obtain $\mathbf{h} := \text{vec}(H') = U\boldsymbol{\beta} + \mathbf{v}$, where $U := G^T \diamond A$, $\boldsymbol{\beta} := [\beta_1 \cdots \beta_r]^T$ and \mathbf{v} is the noise on the channel estimates. To relate this to the noise on the data matrix X , let us assume that the latter noise has a zero-mean white complex Gaussian distribution with variance σ^2 . The channel estimates are given by $H = XS^*(SS^*)^{-1}$, and if the training signals have perfect autocorrelation properties ($\frac{1}{N}SS^* = I_L$) then the estimation noise on the channel is zero-mean white complex Gaussian with variance $\sigma_v^2 = \sigma^2/N$.

The CRB for DOA estimation without delay spread was derived in [22]. The model therein is the familiar $\mathbf{x}(t) = A\mathbf{s}(t) + \mathbf{n}(t)$, $t = 1, \dots, N$, with the noise $\mathbf{n}(t)$ assumed Gaussian and spatially and temporally uncorrelated and $\mathbf{s}(t)$ some unknown deterministic sequence. We can readily adapt the results in [22] to our model. Omitting the details, we have

$$CRB(\boldsymbol{\alpha}, \boldsymbol{\tau}) = \frac{\sigma_v^2}{2} \{ \text{real} (\mathcal{B}^* D^* P_U^\perp D \mathcal{B}) \}^{-1} \quad (22)$$

where $\mathcal{B} = I_2 \otimes B$, $P_U^\perp = I - U(U^*U)^{-1}U^*$, and $D = [G^T \diamond A', (G')^T \diamond A]$. 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:

$$A' = A'(\boldsymbol{\alpha}) = \left[\frac{d\mathbf{a}}{d\alpha_1}(\alpha_1), \dots, \frac{d\mathbf{a}}{d\alpha_r}(\alpha_r) \right],$$

and similar for $G' = G'(\boldsymbol{\tau})$. The diagonal elements of the CRB define the minimal variance that can be obtained for an unbiased estimate of the corresponding parameter.

For a Rayleigh-fading channel, the path fading $\boldsymbol{\beta}$ have a zero-mean complex Gaussian distribution, with some covariance matrix R_β . The Cramer-Rao bound in this case has been developed in [5]:

$$CRB(\boldsymbol{\alpha}, \boldsymbol{\tau}) = \frac{\sigma_v^2}{2} \{ \text{real} (D^* P_U^\perp D \odot (\mathbf{1}_{2 \times 2} \otimes R_\beta U^* R_{\mathbf{h}}^{-1} U R_\beta)^T) \}^{-1} \quad (23)$$

Here \odot denotes the element-wise matrix product (Schur-Hadamard product), $\mathbf{1}_{2 \times 2}$ is a 2×2 matrix of ones, and $R_{\mathbf{h}}$ is the covariance matrix of the channel estimates \mathbf{h} , given by $R_{\mathbf{h}} = U R_\beta U^* + \sigma_v^2 I$.

VII. SIMULATION RESULTS

In this section we illustrate the performance of the algorithm, referred to as SIJADE (Shift-Invariance Joint Angle and Delay Estimation). We assume one source emitting signals that arrive at an array of $M = 2$ sensors via $r = 2$ paths. We also assume the communication protocol uses training bits, from which the channel is estimated using least squares. We collect samples of $\mathbf{x}(t)$ during $N = 40$ symbol periods. The pulse shape function is a raised cosine with 0.35 excess bandwidth, truncated to a length of $L_g = 6$ symbols. In the basic set-up, we choose angles of arrival $[-10, 20]^\circ$, time delays $[0, 1.1]T$, constant fading amplitudes $[1, .8]$, and randomly selected but constant fading phases. The stacking parameters are $m_1 = 5$ and $m_2 = 1$, the oversampling rate is $P = 2$, and the noise power is -15 dB (which translates to an SNR of roughly 16–18 dB, varying with fading phases). In subsequent simulations, some of these parameters are varied. The experimental standard deviation of the estimates is based on 500 Monte-Carlo runs, and is compared against the deterministic CRB. As explained in section II-B, the estimates of the algorithm are slightly biased, but since this is usually an order of magnitude smaller than the standard deviation of the estimates, a comparison to the CRB is meaningful. The diagonalization method used in the final step is either method “T” or method “Q”: the results of these methods are almost always indistinguishable except in critical cases, as demonstrated later in this section.

Basic SIJADE performance.

Figure 5 shows the experimental variance of the angle and delay estimates as a function of the noise power σ . All parameters are as listed above. The two curves in each figure correspond to each of the two rays (for the CRB as well). It is seen that the difference in performance compared to the CRB is approximately 3 to 5 dB. The bias of the estimates was at least an order of magnitude smaller than their standard deviation.

Resolution of SIJADE.

The achievable resolution is illustrated in figure 6 by varying the angle and delay of the second ray, keeping the angle and delay of the first ray fixed at $(-10^\circ, 0T)$. The same parameters as before were used. As expected, for “well-separated” delays the resolution of the angles is only limited by the amount of noise, whereas the resolution of the delays suffers whenever the τ ’s are close, since with two antennas we cannot resolve two equal-delay rays using ESPRIT.

Influence of the stacking parameters m_1 and m_2 .

For well-conditioned scenarios, the precise values of m_1 and m_2 are not critical, as long as they allow the estimation of r rays. The choice of the stacking parameters matters more when the paths are not well separated in space or time.

For close angles, such as in figure 7(a), the delay estimation is much influenced by the choice of m_1 , while the angle estimation is not (and thus is not shown here). Increasing m_1 usually results in better performance. This is because the angle-delay matrix \mathcal{A} ($m_1(M - m_2 + 1) \times r$) in (11) or (13) gets taller, so that its columns are more linearly independent from each other, resulting in increased accuracy of the angle and delay estimates. However, the delay matrix \mathcal{F}_e in (13) gets less wide, and as we increase m_1 past a certain point the performance starts to degrade again because the rows are insufficiently independent. (This doesn’t show very well in figure 7(a) because the delays are well separated.) Note from the graph that the optimal value of m_1 is 6. Interestingly, figure 3 predicts that precisely this setting allows a maximal number of rays to be estimated in the case $LW = 8$ and $M = 2$.

Similar conclusions follow from figure 7(b), where the delays are close. In this case the angle estimation is affected by the choice of m_2 , while the delay estimation is not. Increasing m_2 from 1 to 2 marginally improves the accuracy of angle estimation, but too large an m_2 will take a toll on the conditioning of \mathcal{A} , since its columns become shorter and will not be as “independent”. Note from the graph that the optimal value of m_2 is 2, which is also the optimal setting in figure 3 for $LW = 5$ and $M = 7$.

Choice of channel length L .

Figure 8 looks into the effect of estimating the channel length wrongly. The true channel length is 8. It is seen that when we increase the estimate of the channel length L , the performance slowly degrades. This is because more noise enters into the tail of the channel estimate, and hence into the data matrix \mathcal{H}_e . The performance degradation is smooth: a small overestimation of L is certainly acceptable. Underestimating L is much more serious since this affects the validity of the model.

Influence of joint diagonalization method.

Usually, the performance of the algorithm is not influenced very much by the choice of the joint diagonalization method in the final estimation of the parameters (methods “T”, “Q”, “QZ” in section III-E). However, there are critical cases where the difference is clear. One such case, where the delays are very closely spaced, is depicted in figure 9. Method “T” performs poorly in estimating the delays. This is because in the critical case the eigenvector matrix T becomes complex, so that $T(\tilde{E}_{x\phi}^{-1}\tilde{E}_{y\phi})T^{-1}$ and $T(\tilde{E}_{x\theta}^{-1}\tilde{E}_{y\theta})T^{-1}$ are no longer real-valued. Hence the real and imaginary parts of the eigenvalues are mixed, causing the angle and delay estimates to be mixed as well. (Since the delays are quite close, this does not show up in the angle estimates.) The difference between method “Q” and “QZ” is usually not significant.

Comparison to other algorithms.

Finally, we compare SIJADE to a few other angle-delay estimation algorithms. We will focus on a variant of the algorithm by Swindlehurst [7], which is based on IQML (cf.

[23]) and the method in [24], which is known as 2-D IQML. In both cases, we start from the model $\bar{H} = ABF$ as derived in section II. ³ The IQML method of [7] estimates the delays from the Vandermonde structure of F , while disregarding the structure in A . After F has been found, A can be computed (up to scaling by B) as $\bar{H}F^\dagger$, and the angle (or angles) corresponding to each delay are estimated by an ML search on each column of A , by fitting to the closest vectors in the array manifold. 2-D IQML is a generalization of IQML, originally for joint azimuth-elevation estimation, and uses the model $\mathbf{h} = (F^T \diamond A)\boldsymbol{\beta}$. It tries to simultaneously fit the Vandermonde structures of both F^T and A . 2-D IQML requires that $M > r$, so that we let the number of antennas be $M = 3$ in the comparison. Results as a function of SNR are shown in figure 10. The performance of all three algorithms is comparable in this well-conditioned case. The computational complexity of IQML-based algorithms is quite high, since they are iterative and need to compute eigenvalues at each iteration.

Figure 11 compares the three algorithms in a more critical case where the delay spread is small. For the chosen parameter values, the signal singular values of \bar{H} are only slightly above the noise level. Because the IQML and 2-D IQML algorithms as described in [7, 24] do not work with an extended \mathcal{H}_e -matrix, the fact that F is almost singular makes the algorithms fail: they lack sufficient resolution for delays that are closely spaced.

VIII. CONCLUSION

We have described an algorithm that jointly estimates the directions of arrival and time delays of multiple paths using an estimate of the channel impulse response. The SIJADE algorithm is closed form and can estimate the parameters of more paths than the number of antennas. The usual "smoothing" techniques available to ESPRIT can be elegantly incorporated and improve the resolution of closely separated rays.

A limitation of the algorithm is that it starts from impulse response data. If we have input-output data available, as is typically the case, then the estimation of first the impulse response and second the parameters is inherently suboptimal.

REFERENCES

- [1] 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.
- [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] B.H. Khalaj, A. Paulraj, and T. Kailath, "Spatio-temporal channel estimation techniques for multiple access spread spectrum systems with antenna arrays," in *Proc. IEEE Int. Conf. Communications*, (Seattle, WA), pp. 1520-1524 vol.3, IEEE, June 1995.
- [4] M.D. Zoltowski and J. Ramos, "Blind adaptive beamforming for narrowband co-channel digital communications signals in a multipath environment," in *Proc. IEEE ICASSP*, (Detroit, MI), pp. 1745-1748, 1995.
- [5] M.C. Vanderveen, B.C. Ng, C. 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.
- [6] J. Gunther and A.L. Swindlehurst, "Algorithms for blind equalization with multiple antennas based on frequency domain subspaces," in *Proc. IEEE ICASSP*, (Atlanta, GA), pp. 2421-2424, 1996.
- [7] A.L. Swindlehurst, "Time delay and spatial signature estimation using known asynchronous signals," *submitted to IEEE Trans. Signal Proc.*, May 1996.
- [8] 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, May 1996.
- [9] Y. Ogawa, N. Hamaguchi, K. Ohshima, and K. Itoh, "High-resolution analysis of indoor multipath propagation structure," *IEICE Trans. Communications*, vol. E78-B, pp. 1450-1457, Nov. 1995.
- [10] N. Kikuma, H. Ino, T. Kitamura, and N. Inagaki, "Parameter estimation of indoor multipath waves using ESPRIT with rotatory scanning of antenna," in *IEEE Antennas and Propagation Soc. Int. Symp.*, (Chicago, IL), pp. 1023-1026 vol.2, IEEE, July 1992.
- [11] A.J. van der Veen, P.B. Ober, and E.F. Deprettere, "Azimuth and elevation computation in high resolution DOA estimation," *IEEE Trans. Signal Processing*, vol. 40, pp. 1828-1832, July 1992.
- [12] 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.
- [13] T.J. Shan, M. Wax, and T. Kailath, "On spatial smoothing for direction-of-arrival estimation of coherent signals," *IEEE Trans. Acoust. Speech Signal Proc.*, vol. 33, pp. 806-811, Apr. 1985.
- [14] M. Haardt and J.A. Nossek, "Unitary ESPRIT: how to obtain increased estimation accuracy with a reduced computational burden," *IEEE Trans. Signal Processing*, vol. 43, pp. 1232-1242, May 1995.
- [15] A.J. van der Veen, S. Talwar, and A. Paulraj, "Blind estimation of multiple digital signals transmitted over FIR channels," *IEEE Signal Processing Letters*, vol. 2, pp. 99-102, May 1995.
- [16] E. Moulines, P. Duhamel, J.F. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Tr. Signal Processing*, vol. 43, pp. 516-525, Feb. 1995.
- [17] R. Roy, A. Paulraj, and T. Kailath, "ESPRIT - A subspace rotation approach to estimation of parameters of cisoids in noise," *IEEE Trans. on Acoust., Speech, Signal Proc.*, vol. 34, pp. 1340-1342, October 1986.
- [18] G. Golub and C.F. Van Loan, *Matrix Computations*. The Johns Hopkins University Press, 1989.
- [19] P.J. Eberlein, "On the Schur decomposition of a matrix for parallel computation," *IEEE Trans. Comp.*, vol. 36, pp. 167-174, 1987.
- [20] A.J. van der Veen and A. Paulraj, "An analytical constant modulus algorithm," *IEEE Trans. Signal Processing*, vol. 44, pp. 1136-1155, May 1996.
- [21] L. De Lathauwer, B. De Moor, and J. Vandewalle, "Independent component analysis based on higher-order statistics only," in *Proc. IEEE SP Workshop on Stat. Signal Array Proc.*, (Corfu, Greece), pp. 356-359, June 1996.
- [22] 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.
- [23] Y. Bresler and A. Macovski, "Exact maximum likelihood parameter estimation of superimposed exponential signals in noise," *IEEE Trans. Acoust. Speech Signal Proc.*, vol. 34, pp. 1081-1089, Oct. 1986.
- [24] 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.

³The way to arrive at this model is not the same in all algorithms, but here we do not wish to compare on discrepancies caused by different methods of deconvolution.

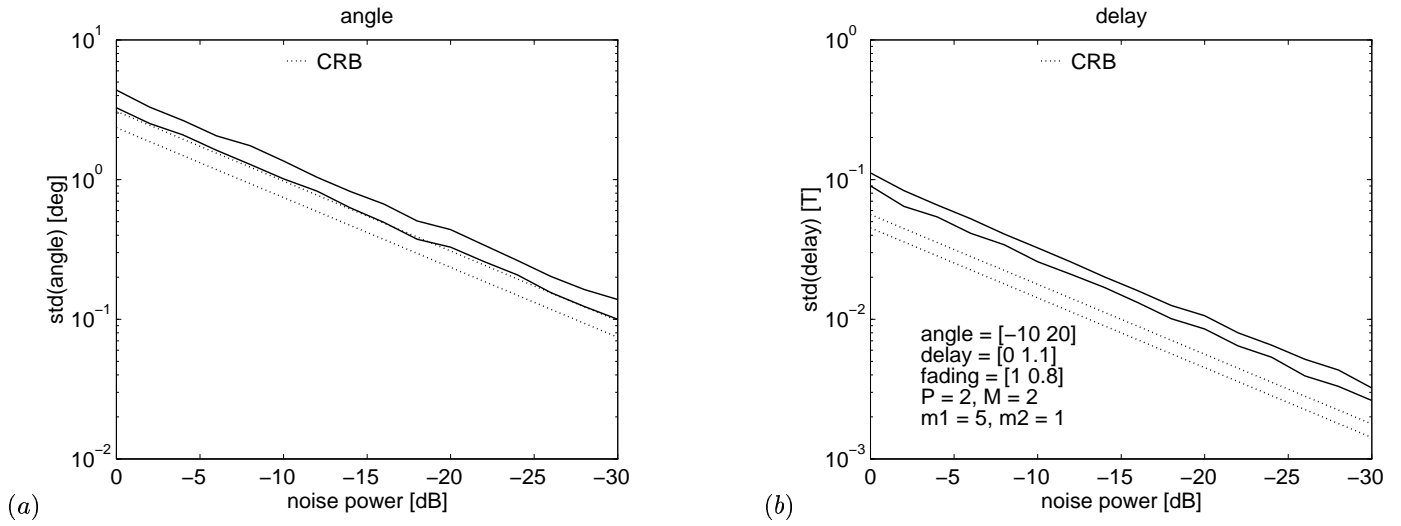


Fig. 5. Standard deviation of estimates vs. noise

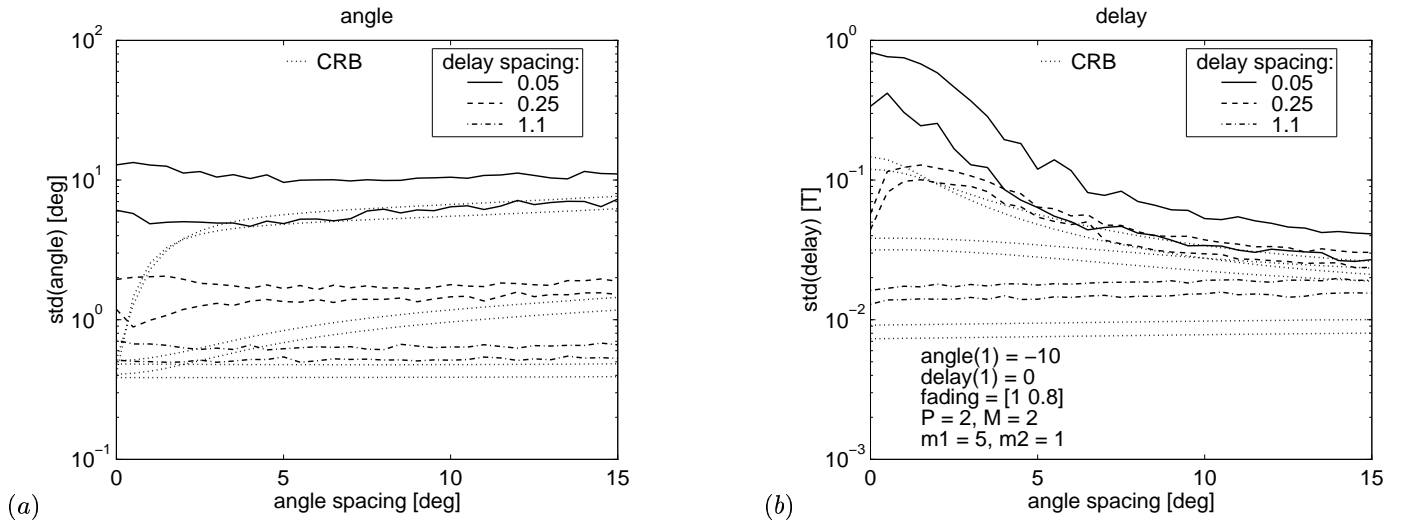


Fig. 6. Standard deviation of estimates vs. angle/delay separation

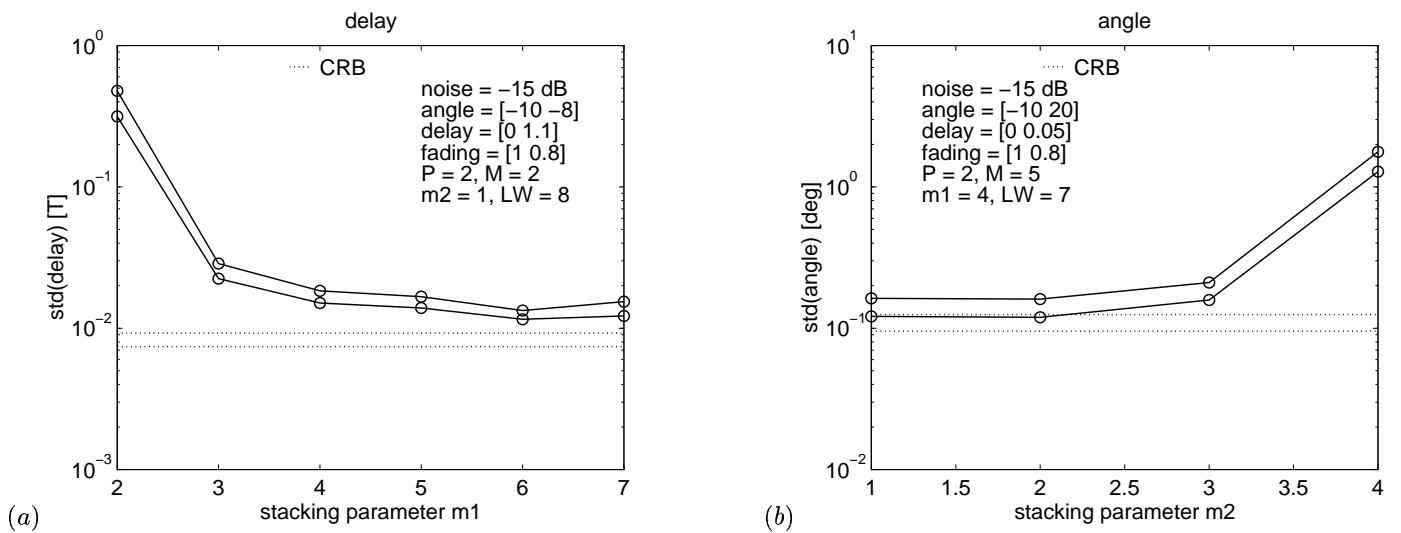


Fig. 7. Standard deviation of estimates vs. choice of m_1, m_2

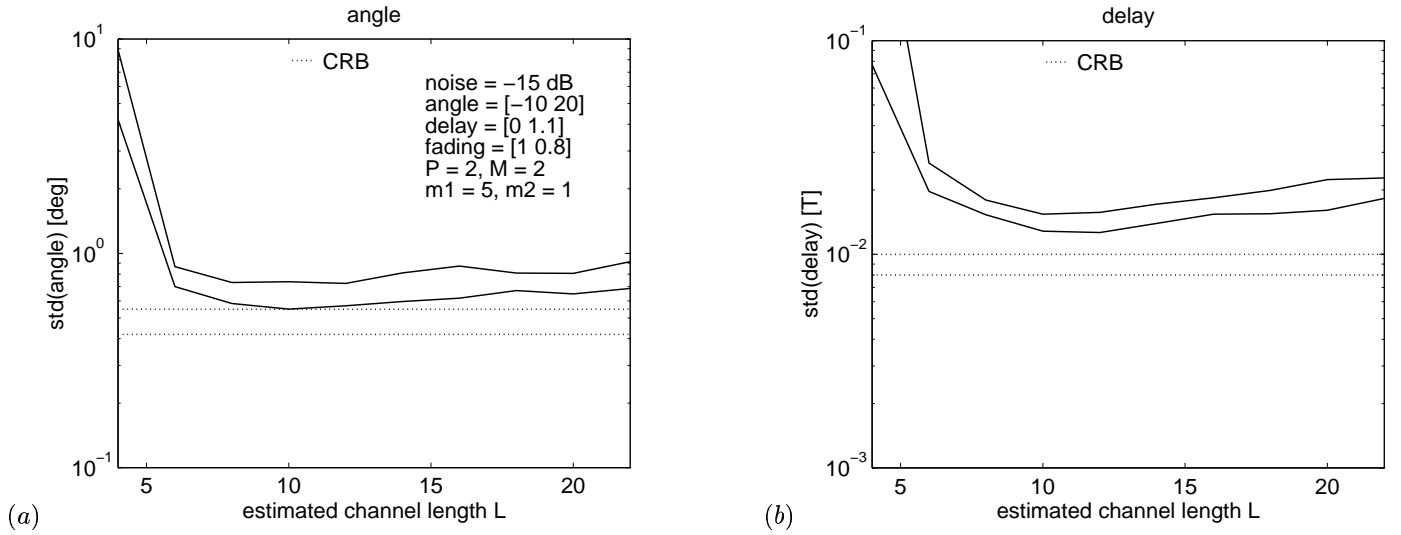


Fig. 8. Standard deviation of estimates vs. estimated channel length

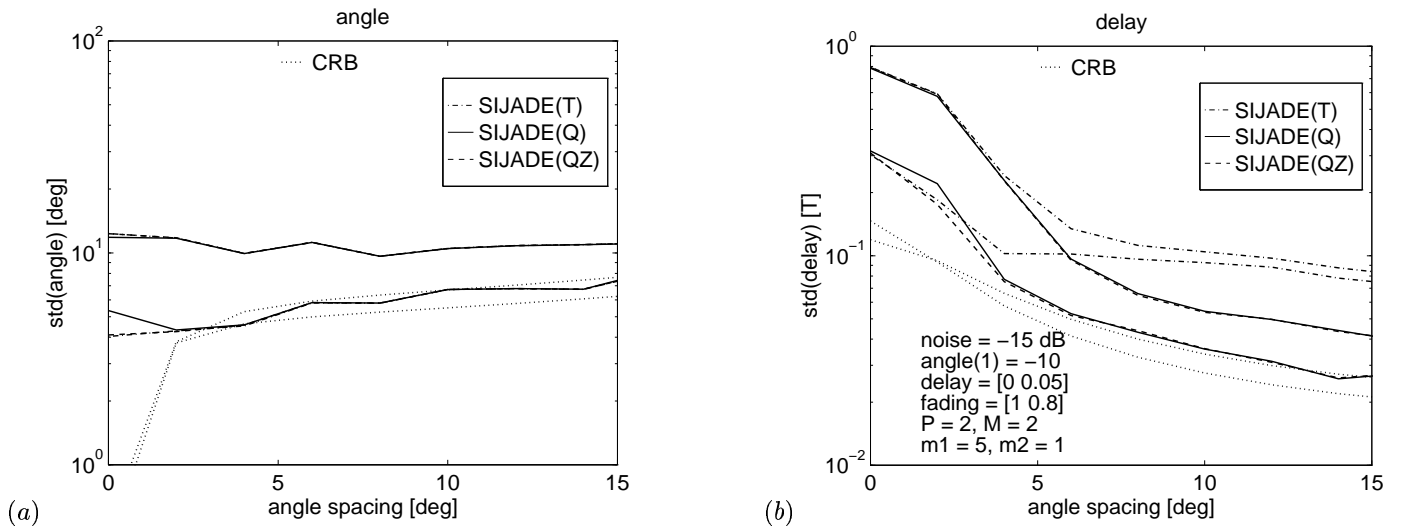


Fig. 9. Impact of diagonalization method (critical case)

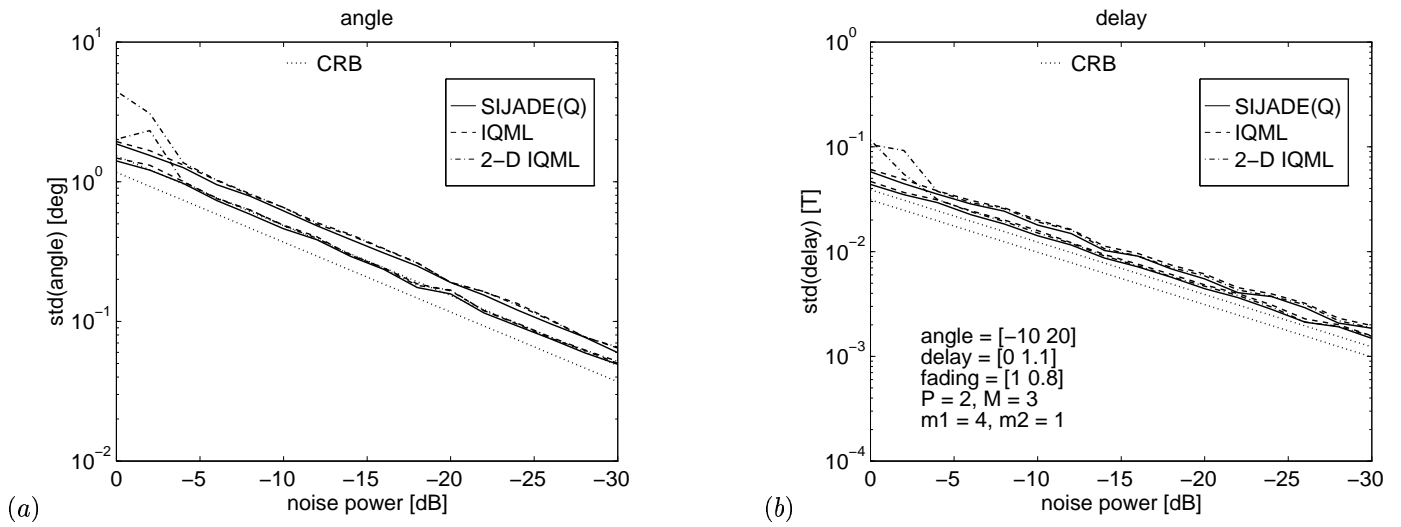


Fig. 10. Comparison to other joint angle-delay methods

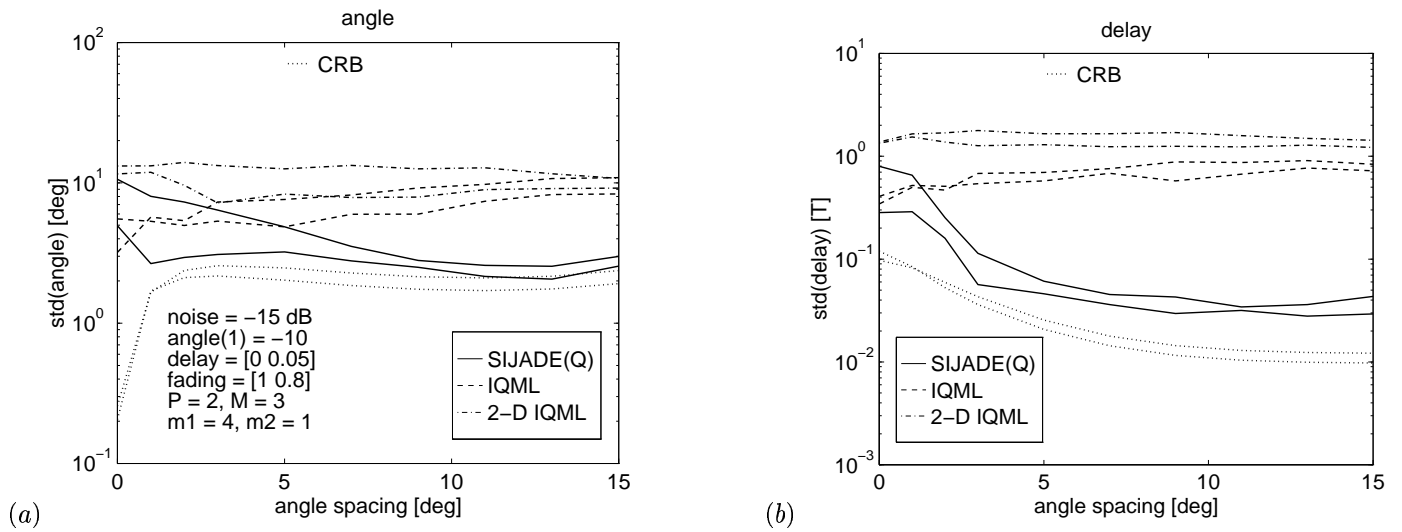


Fig. 11. Comparison to other joint angle-delay methods (critical case)