16 Direction-of-arrival estimation

Mats Viberg

16.1. Signal models and problem formulation

The purpose of this chapter of the book is to review the area of direction-of-arrival (DOA) estimation, with special focus on wireless communication systems. Such information is useful for understanding the propagation environment, and gives empirical evidence for theoretical propagation models. Both detailed modeling including resolution of all propagation paths, and statistical models of scatter clusters can be of interest. Both these cases will be considered herein. The chapter is organized as follows: first the basic signal models are motivated and the estimation problems are introduced in Section 16.1. Then, in Section 16.2 a brief introduction to the standard DOA estimation problem is given. Section 16.3 considers estimation in case of spatially spread sources (i.e., unresolved scatterers). Finally, Section 16.4 applies the DOA estimation methods to parametric modeling of the wireless channel.

16.1.1. Spatial signal modeling

A traditional wireless communication system has one transmit and one receive antenna. To enable estimation of the direction to the transmitter, the received signal must be sampled at several spatial locations. The different elements of such an antenna array will all measure the same signal, but with different time delays depending on the DOA. In a realistic scenario, these time delays are much smaller than the reciprocal of the signal bandwidth, and can therefore be considered as phase shifts (see [1] for more details regarding this narrowband assumption). Let $s(t)$ be the complex baseband equivalent of the electromagnetic field at a reference position, due to a far-field transmitter in the direction θ. The vector-valued received signal at the positions of m antenna elements can then be expressed as

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_m(t) \end{bmatrix} = \begin{bmatrix} e^{-j\omega_c \tau_1(\theta)} \\ e^{-j\omega_c \tau_2(\theta)} \\ \vdots \\ e^{-j\omega_c \tau_m(\theta)} \end{bmatrix} s(t) = \mathbf{a}(\theta) s(t), \tag{16.1}$$

where ω_c is the carrier frequency and $\tau_k(\theta)$ denotes the propagation delay from the reference to the kth element. The vector $\mathbf{a}(\theta)$ is termed the steering vector of the array. Assuming linear receivers with flat frequency responses over the signal bandwidth, the down-converted antenna outputs are proportional to $\mathbf{x}(t)$. Due to mutual coupling and other effects, the individual antenna elements may have θ-dependent gain and phase characteristics, that add to the phase-shifts due to geometry present in (16.1). As long as the receiver dynamics can be ignored, the antenna outputs can still be expressed as $\mathbf{x}(t) = \mathbf{a}(\theta)s(t)$, but the functional form of the steering vector would have to be adjusted accordingly. A structure of special interest arises if the antenna elements are nondirectional and uniformly spaced along a straight line. The steering vector for such a uniform linear array (ULA) has the form

$$\mathbf{a}_{\mathrm{ULA}}(\theta) = \left[1, e^{j\phi}, \ldots, e^{j(m-1)\phi}\right]^T, \tag{16.2}$$

where $\phi = kl\sin\theta$ is called the *electrical angle*, $k = \omega_c/c$ is the wave number, ω_c is the carrier frequency, c denotes the speed of propagation, and l is the element separation. The DOA θ is measured relative to the array normal. To uniquely determine θ, it is clear that $|\phi| \leq \pi$ must hold for all θ. In case the whole field of view is of interest, this leads to $l \leq \pi/k = \lambda/2$, where λ is the wavelength. A ULA with maximum element separation $l = \lambda/2$ is often termed a *standard* ULA, and is often used for comparison.

Due to linearity, the extension of (16.1) to the case of multiple signals is straightforward. Thus, assume there are d far-field emitters, transmitting the baseband signal waveforms $s_k(t)$, $k = 1, 2, \ldots, d$. The array output is then modeled by the familiar equation

$$\mathbf{x}(t) = \sum_{k=1}^{d} \mathbf{a}(\theta_k)s_k(t) + \mathbf{n}(t), \tag{16.3}$$

where we have also included an additive noise term $\mathbf{n}(t)$. In matrix form, the above reads

$$\mathbf{x}(t) = \mathbf{A}(\boldsymbol{\theta})\mathbf{s}(t) + \mathbf{n}(t), \tag{16.4}$$

where $\mathbf{A}(\boldsymbol{\theta}) = [\mathbf{a}(\theta_1), \ldots, \mathbf{a}(\theta_d)]$ is the steering matrix, $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_d]^T$ contains the signal parameters, and $\mathbf{s}(t) = [s_1(t), \ldots, s_d(t)]^T$ is composed of the signal waveforms. In the "standard" DOA estimation literature, no information regarding $\mathbf{s}(t)$ is assumed to be available. The two prevalent models for the signal waveforms are that they are either regarded as deterministic parameters to be estimated, or they are drawn from a d-variate complex Gaussian distribution with zero mean and covariance matrix $\mathbf{P} = \mathrm{E}[\mathbf{s}(t)\mathbf{s}^*(t)]$. To enable reliable estimation of the DOA parameters, information about the spatial covariance matrix of the noise is necessary. For simplicity, it is assumed that a prewhitening has been performed, so that

$\mathrm{E}[\mathbf{n}(t)\mathbf{n}^*(t)] = \sigma^2\mathbf{I}$ holds. Assuming stochastic signals, the spatial covariance of the array output reads

$$\mathbf{R} = \mathrm{E}\left[\mathbf{x}(t)\mathbf{x}^*(t)\right] = \mathbf{A}(\boldsymbol{\theta})\mathbf{P}\mathbf{A}^*(\boldsymbol{\theta}) + \sigma^2\mathbf{I}. \tag{16.5}$$

The array output is assumed to be sampled at N discrete time instants, say $t_1, t_2, \ldots, t_N$. The standard DOA estimation problem is then to determine the DOA parameters θ_k, $k = 1, 2, \ldots, d$ based on the measurements $\mathbf{x}(t_n)$, $n = 1, 2, \ldots, N$. Included in this problem is to estimate the number of (significant) signal components, d. Most DOA estimation methods use the measurements only to form the sample covariance matrix

$$\hat{\mathbf{R}} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{x}(t_n)\mathbf{x}^*(t_n). \tag{16.6}$$

Under fairly mild conditions on the signals and noise, it holds that $\hat{\mathbf{R}} \to \mathbf{R}$ with probability 1 as $N \to \infty$. Thus, provided the DOA estimation method gives correct estimates when $\mathbf{R}$ is used, we can expect good performance as long as N is "large enough."

The above model can be extended in various ways, for example, by taking both azimuth and elevation into account, considering near-field sources, or including effects of polarization. Of most importance in a mobile communication system, however, is the concept of multipath propagation.

16.1.2. Multipath propagation

In microwave communication, it is well known that a significant proportion of the signal energy is scattered via several propagation paths. Indeed, in many cases there is no line-of-sight (LOS) between the mobile and base station, implying that all of the energy is due to scattering. In the Jakes model [2], the scatterers are spread out evenly on a circle surrounding the mobile, but also other models have been proposed. For closely spaced scatterers, the propagation delays are similar, and the received signal components differ only in a possibly time-varying, complex scaling. The wideband case, where time delays need to be accounted for, is considered in Section 16.4. Adopting the spatial signal model (16.3), we have $s_k(t) = g_k(t)s(t)$, $k = 1, \ldots, d$, where $g_k(t)$ are termed the reflection (or scattering) coefficients. The array output is thus modeled as

$$\mathbf{x}(t) = \left(\sum_{k=1}^{d} g_k(t)\mathbf{a}\big(\theta_k(t)\big)\right)s(t) + \mathbf{n}(t) = \mathbf{v}(t)s(t) + \mathbf{n}(t). \tag{16.7}$$

The complex m-vector $\mathbf{v}(t)$ is usually referred to as the *spatial signature* of the transmitter. Since the scatterers are close, it is not practical to resolve all individual propagation paths. We therefore refer to this as a *spatially spread* source, as opposed to the standard case of point sources. In a time-varying scenario, the scatter

coefficients $g_k(t)$ are usually time-varying at a significantly higher rate than the DOAs $\theta_k(t)$ [3], and the latter may therefore be considered stationary over a short data collection interval. Both the scatterers and DOAs are modeled as random variables, with $\mathrm{E}[g_k(t)] = 0$, $\mathrm{E}[g_k(t)g_l^*(t)] = \sigma_g^2\delta_{k,l}$, $\mathrm{E}[g_k(t)g_l(t)] = 0$, $\mathrm{E}[\theta_k] = \theta_0$, and $\mathrm{E}[(\theta_k - \theta_0)^2] = \sigma_\theta^2$. The source is thus characterized by the mean (or nominal DOA) θ_0 and the standard deviation (or spreading factor) σ_θ. The scatter coefficients are usually normalized so that $\sigma_g^2 = 1/d$, by absorbing the power into $s(t)$. Under the stated assumptions, the spatial array covariance matrix takes the form

$$\mathbf{R}_s = \mathrm{E}\left[\mathbf{x}(t)\mathbf{x}^*(t)\right] = P\int_\theta \mathbf{a}(\theta)\mathbf{a}^*(\theta)p_\theta(\theta)d\theta + \sigma^2\mathbf{I}, \tag{16.8}$$

where $p_\theta(\theta)$ is the probability density function (PDF) of the DOAs. The number of incoming rays, d, is usually assumed large, implying that $\mathbf{x}(t)$ is well modeled as a zero-mean Gaussian random vector. Thus, ignoring the temporal correlation structure, $\mathbf{R}_s$ captures all information regarding the incoming energy. To get a reliable estimate of $\mathbf{R}_s$, one needs to see several realizations of the random propagation paths, all taken from the same underlying distribution. Thus, in this case the data $\mathbf{x}(t)$ should be observed slowly enough to enable each $g_k(t)$ and $\theta_k(t)$ to vary significantly between samples. In the literature, the scatter coefficients are often modeled as temporally white. Given data $\mathbf{x}(t_n)$, $n = 1,\ldots,N$, the usual sample covariance matrix

$$\hat{\mathbf{R}} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{x}(t_n)\mathbf{x}^*(t_n) \tag{16.9}$$

is taken as the estimate of $\mathbf{R}_s$. The task is now to infer the characteristics of the scattering environment based on $\hat{\mathbf{R}}$. Usually, it is sufficient to determine θ_0 and σ_θ.

16.2. DOA estimation for point sources

This section presents methods for the case where the transmitters are resolvable, that is, they are modeled as point sources. This is the "standard" DOA estimation problem, which by now can be considered a mature research area. The following outlines some of the more influential methods. More details are available in, for example, [4, 5].

16.2.1. Beamforming techniques

The duality between uniform sampling in space (ULA) and time have lead researchers to apply methods from one domain to the other. The spatial Fourier transform of a single data snapshot $\mathbf{x}(t_n)$ is given by

$$\tilde{\mathbf{x}}(t_n,\phi) = \sum_{k=0}^{m-1} x_k(t_n)e^{-jk\phi}. \tag{16.10}$$

With $\mathbf{a}_{\mathrm{ULA}}(\theta)$ given by (16.2) we can express this as the inner product

$$\tilde{\mathbf{x}}(t_n, \theta) = \mathbf{a}^*_{\mathrm{ULA}}(\theta)\mathbf{x}(t_n). \tag{16.11}$$

The spatial periodogram of $\mathbf{x}(t_n)$ is then simply $|\tilde{\mathbf{x}}(t_n, \theta)|^2$. If no temporal correlation is assumed, it is natural to average the instantaneous spatial spectra to form the final estimate. When applying this to a general array, for which $\|\mathbf{a}(\theta)\|$ may depend on θ, it is useful to introduce a normalization of (16.11). The resulting spectral estimate, termed *conventional beamforming*, takes the form

$$P_{\mathrm{BF}}(\theta) = \frac{1}{N}\sum_{n=1}^{N} \frac{|\,\mathbf{a}^*(\theta)\mathbf{x}(t_n)\,|^2}{\mathbf{a}^*(\theta)\mathbf{a}(\theta)}. \tag{16.12}$$

Expanding the square $|\mathbf{a}^*(\theta)\mathbf{x}(t_n)|^2 = \mathbf{a}^*(\theta)\mathbf{x}(t_n)\mathbf{x}^*(t_n)\mathbf{a}(\theta)$ and inserting (16.6) leads to

$$P_{\mathrm{BF}}(\theta) = \frac{\mathbf{a}^*(\theta)\hat{\mathbf{R}}\mathbf{a}(\theta)}{\mathbf{a}^*(\theta)\mathbf{a}(\theta)}. \tag{16.13}$$

The locations $\hat{\theta}_k$ of the d highest peaks of $P_{\mathrm{BF}}(\theta)$ are taken as the beamforming DOA estimates. The averaging implies that $P_{\mathrm{BF}}(\theta)$ has a much reduced variance as compared to the classical periodogram for large N, but similar to the latter it has a limited resolution. In case of a ULA, the Rayleigh resolution expressed in electrical angle is $\Delta\phi = 2\pi/m$, which for large m translates to $\Delta\theta \approx 2\pi/(klm)$.

To improve the resolution of the conventional beamformer, Capon [6] advocated the use of adaptive beamforming. Although the original derivation is different, the following interpretation has become popular:

$$\min_{\mathbf{w}} \mathbf{w}^*\hat{\mathbf{R}}\mathbf{w} \quad \text{subject to } \mathbf{w}^*\mathbf{a}(\theta) = 1. \tag{16.14}$$

The energy $\mathbf{w}^*\hat{\mathbf{R}}\mathbf{w}$ of the beamformer output $\mathbf{w}^*\mathbf{x}(t_n)$ is to be minimized, while keeping a unit gain in the "look direction" θ. The optimizing beamforming weights are easily shown to be

$$\mathbf{w}_{\mathrm{CAP}} = \frac{\hat{\mathbf{R}}^{-1}\mathbf{a}(\theta)}{\mathbf{a}^*(\theta)\hat{\mathbf{R}}^{-1}\mathbf{a}(\theta)}, \tag{16.15}$$

which when inserted into $\mathbf{w}^*\hat{\mathbf{R}}\mathbf{w}$, leads to the Capon spectral estimate

$$P_{\mathrm{CAP}}(\theta) = \frac{1}{\mathbf{a}^*(\theta)\hat{\mathbf{R}}^{-1}\mathbf{a}(\theta)}. \tag{16.16}$$

In contrast to the conventional beamformer, the resolution of (16.16) improves with the SNR (signal-to-noise ratio) [7]. It is therefore preferable, at least in high-SNR scenarios. However, the resolution does not improve for increasing N. The Capon DOA estimates, which are the peak locations of $P_{\mathrm{CAP}}(\theta)$, therefore still fail to take full advantage of data model (16.4).

16.2.2. Subspace methods

In the late seventies, a new class of spectral-based estimators were introduced. These have their roots in principal component analysis, and are based on geometrical properties of the array covariance matrix. It is clear from (16.5) that any vector that is orthogonal to $\mathbf{A}(\boldsymbol{\theta})$ is an eigenvector of $\mathbf{R}$ with corresponding eigenvalue σ^2. The remaining eigenvectors are all in the range space of $\mathbf{A}(\boldsymbol{\theta})$ (provided $m > d$ and $\mathbf{P}$ is full rank), and are therefore termed *signal eigenvectors*. The eigendecomposition of $\mathbf{R}$ (16.5) is partitioned into a signal and a noise subspace as

$$\mathbf{R} = \sum_{k=1}^{m} \lambda_k \mathbf{e}_k \mathbf{e}_k^* = \mathbf{E}_s \boldsymbol{\Lambda}_s \mathbf{E}_s^* + \mathbf{E}_n \boldsymbol{\Lambda}_n \mathbf{E}_n^*, \tag{16.17}$$

where $\lambda_1 \geq \cdots \geq \lambda_d > \lambda_{d+1} = \cdots = \lambda_m = \sigma^2$, $\mathbf{E}_s = [\mathbf{e}_1, \ldots, \mathbf{e}_d]$, $\mathbf{E}_n = [\mathbf{e}_{d+1}, \ldots, \mathbf{e}_m]$, and $\boldsymbol{\Lambda}_n = \sigma^2 \mathbf{I}$. The signal eigenvectors in $\mathbf{E}_s$ span the range space of $\mathbf{A}(\boldsymbol{\theta})$, which is termed the *signal subspace*. For the noise eigenvectors we have instead $\mathbf{E}_n \perp \mathbf{A}(\boldsymbol{\theta})$. These relations constitute the fundament for subspace methods for DOA estimation. In passing, we note that the number of signals d can easily be determined from (estimates of) the eigenvalues, either as the number of "significant" eigenvalues or by determining the multiplicity of the minimum eigenvalue [8, 9]. The MUSIC (multiple signal classification) algorithm [10, 11] exploits the orthogonality relation $\mathbf{a}^*(\theta_k)\mathbf{E}_n = 0$, for $k = 1, \ldots, d$. Provided the array is free of ambiguities, there are no false solutions to this equation [12, 13]. The noise subspace matrix is estimated from the eigendecomposition of the sample covariance

$$\hat{\mathbf{R}} = \hat{\mathbf{E}}_s \hat{\boldsymbol{\Lambda}}_s \hat{\mathbf{E}}_s^* + \hat{\mathbf{E}}_n \hat{\boldsymbol{\Lambda}}_n \hat{\mathbf{E}}_n^*. \tag{16.18}$$

Using the estimated noise subspace, the so-called MUSIC pseudospectrum is then defined as

$$P_{\mathrm{MU}}(\theta) = \frac{\mathbf{a}^*(\theta)\mathbf{a}(\theta)}{\mathbf{a}^*(\theta)\hat{\mathbf{E}}_n \hat{\mathbf{E}}_n^* \mathbf{a}(\theta)}. \tag{16.19}$$

This is not a spectrum in the usual sense, since it is in fact dimensionless. Yet, for large enough N and/or SNR, $P_{\mathrm{MU}}(\theta)$ exhibits high peaks near the true DOAs $\theta_1, \ldots, \theta_d$. The MUSIC algorithm calculates the DOA estimates by computing $P_{\mathrm{MU}}(\theta)$ at a fine grid (using FFT with zero-padding in the case of a ULA), and then locating the d largest local maxima. If desired, the estimates can be refined by using a local search.

To illustrate the performance of the different spectral-based estimators, $N = 500$ snapshots of $\mathbf{x}(t_n)$ are generated according to the model (16.4). The array is a standard ULA of $m = 6$ sensors, and $d = 3$ emitters are located at $\boldsymbol{\theta} = [0^\circ, 5^\circ, 20^\circ]^T$. The waveforms are assumed uncorrelated and of equal power and the SNR is 10 dB. Both the signal waveforms and the noise are realizations of white Gaussian random processes. The conventional and Capon spatial spectra are

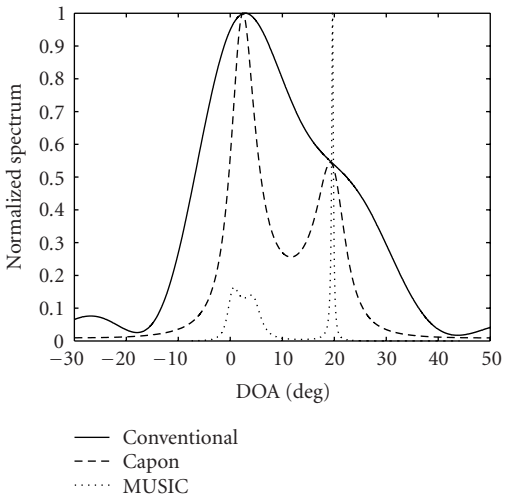

FIGURE 16.1. Spatial spectrum estimates using conventional beamforming, Capon's beamformer, and MUSIC. Standard ULA of $m = 6$ elements. True DOAs $\{0°, 5°, 20°\}$, SNR $= 10$ dB, $N = 500$.

displayed in Figure 16.1, along with the MUSIC pseudospectrum, all normalized to have maximum value one. The conventional beamformer shows just one peak around 3.4°, whereas the Capon spectrum exhibits two peaks at 2.4° and 19.2°. In contrast, the MUSIC pseudospectrum resolves all three signals with peaks at −0.5°, 5.9° and 19.5°. This is of course just one realization using random signals, but the general behavior in Figure 16.1 is representative.

Although the MUSIC algorithm can improve the resolution significantly over conventional spectral-based estimators, it requires a high SNR and/or N for this to happen. Several approaches have been proposed to increase resolution, including beamspace processing and the min-norm algorithm. For a useful comparison of these and other versions, see [14]. In case of a ULA, the threshold performance can be significantly enhanced by using a root-finding procedure instead of searching the pseudospectrum for peaks [15]. The idea is to express the spectrum as a polynomial in the complex variable $z = e^{j\phi}$. ULA steering vector (16.2) is written as

$$\mathbf{a}(z) = \left[1, z, \ldots, z^{m-1}\right]^T, \tag{16.20}$$

and the inverse of the MUSIC pseudospectrum (sometimes termed the "null spectrum") then becomes

$$P(z) = \mathbf{a}^T(z^{-1})\hat{\mathbf{E}}_n\hat{\mathbf{E}}_n^*\mathbf{a}(z). \tag{16.21}$$

Regarding $P(z)$ (or rather $z^{m-1}P(z)$) as a polynomial in z, a standard numerical routine can be used to compute the $2m - 1$ roots. Out of these, the so-called root-MUSIC algorithm uses the d roots, say $\{z_k\}_{k=1}^d$, that are closest to the unit

circle. Since $z = e^{j\phi}$ where $\phi = kl\sin\theta$, the root-MUSIC DOA estimates are then computed as $\widehat{\theta}_k = \arcsin[\text{angle}(z_k)/kl]$, $k = 1,\ldots,d$. It is known that the root-MUSIC DOA estimates have the same asymptotic properties as the standard spectral MUSIC. However, the threshold behavior is significantly improved, which according to [16] is explained by the fact that the errors in root locations have a large radial component. A further improvement is obtained by using forward-backward averaging [16].

The ESPRIT algorithm [17] is related to root-MUSIC, but requires computing only d roots. It is based on the assumption that the array is composed of two identical, spatially separated, subarrays that have the same orientation. For a ULA, there are several ways to construct subarrays. A popular choice is to let elements $1 : m-1$ constitute the first array and $2 : m$ the second. Let $\mathbf{A}_1$ denote the steering matrix for the first subarray, that is, rows 1 through $m-1$ of $\mathbf{A}(\boldsymbol{\theta})$. Similarly, $\mathbf{A}_2$ contains the $m-1$ last rows. Using (16.20), these matrices are related by

$$\mathbf{A}_2 = \mathbf{A}_1 \,\text{diag}\,[z_1,\ldots,z_d], \tag{16.22}$$

where $z_k = e^{j\phi_k}$ is the root corresponding to the kth source. This relation could be used to determine the roots if $\mathbf{A}(\boldsymbol{\theta})$ was known. In general, all that is available is an estimate of the spatial array covariance matrix $\mathbf{R}$. From the eigendecomposition (16.17), we collect the signal eigenvectors into $\mathbf{E}_s$, which is not identical to $\mathbf{A}$ but it spans the same range space. Therefore, $\mathbf{E}_s = \mathbf{AT}$ for some full-rank $d \times d$ matrix $\mathbf{T}$, and it follows that

$$\mathbf{E}_2 = \mathbf{E}_1\mathbf{T}\,\text{diag}\,[z_1,\ldots,z_d]\mathbf{T}^{-1}, \tag{16.23}$$

where $\mathbf{E}_s$ has been partitioned conformably with $\mathbf{A}$. Using the estimated $\widehat{\mathbf{E}}_s$ in (16.23), the submatrices $\widehat{\mathbf{E}}_1$ and $\widehat{\mathbf{E}}_2$ will, in general, not span the same range space, and (16.23) has no exact solution. Instead, a least-squares or total-least-squares solution $\widehat{\Psi}$ to the relation

$$\widehat{\mathbf{E}}_2 \approx \widehat{\mathbf{E}}_1\widehat{\Psi} \tag{16.24}$$

is computed. The poles $\{z_k\}_{k=1}^{d}$ are then given as the eigenvalues of $\widehat{\Psi}$, utilizing that Ψ and $\text{diag}[z_1,\ldots,z_d]$ are related by a similarity transformation and therefore share the same eigenvalues. Finally, the ESPRIT DOA estimates are computed as $\widehat{\theta}_k = \arcsin[\text{angle}(z_k)/kl]$, $k = 1,\ldots,d$. The method is computationally very efficient, but exhibits slightly worse performance than root-MUSIC. Similar to the latter, ESPRIT benefits from forward-backward averaging. A clever implementation using only real-valued operations (and therefore reduced computational burden) was proposed in [18].

16.2.3. Nonlinear least-squares estimation

The subspace-based techniques mentioned above generally provide accurate DOA estimates at an affordable computational cost. However, the performance degrades when the signal waveforms are correlated ($\mathbf{P}$ nondiagonal), and fail to operate in the presence of coherent sources. Although forward-backward averaging and spatial smoothing techniques can mitigate the drawbacks somewhat (e.g., [19]), these techniques remain suboptimal. Given the model (16.4), a straightforward approach is to use a nonlinear least-squares (NLLS) [20] fit:

$$\{\hat{\boldsymbol{\theta}}, \hat{\mathbf{s}}(t_n)\} = \arg\min_{\boldsymbol{\theta}, \mathbf{s}(t_n)} \sum_{n=1}^{N} ||\mathbf{x}(t_n) - \mathbf{A}(\boldsymbol{\theta})\mathbf{s}(t_n)||^2. \tag{16.25}$$

This is a separable least-squares problem, and for fixed (but unknown) $\boldsymbol{\theta}$, the solution with respect to the linear parameter $\mathbf{s}(t_n)$ is

$$\hat{\mathbf{s}}(t_n) = (\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}^*\mathbf{x}(t_n). \tag{16.26}$$

Substituting (16.26) into (16.25) leads to the concentrated NLLS formulation

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \operatorname{Tr}\{\boldsymbol{\Pi}_{\mathbf{A}}^{\perp}\hat{\mathbf{R}}\}, \tag{16.27}$$

where $\boldsymbol{\Pi}_{\mathbf{A}}^{\perp} = \mathbf{I} - \boldsymbol{\Pi}_{\mathbf{A}} = \mathbf{I} - \mathbf{A}(\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}^*$ is the orthogonal projection onto the nullspace of $\mathbf{A}^*$. The concentrated form (16.27) has a nice interpretation; $\operatorname{Tr}\{\boldsymbol{\Pi}_{\mathbf{A}}^{\perp}\hat{\mathbf{R}}\}$ is a measure of the remaining output power after removing all the energy stemming from the hypothetical DOA parameters $\boldsymbol{\theta}$. Clearly, this should reach its minimum when $\boldsymbol{\theta}$ is close to $\boldsymbol{\theta}_0$, because this removes all signal components from $\hat{\mathbf{R}}$. It is easy to see that the NLLS approach coincides with maximum likelihood (ML) if the signal waveforms $\mathbf{s}(t_n)$ are modeled as deterministic parameters and the noise is assumed to be Gaussian. In many applications, also $\mathbf{s}(t_n)$ can be regarded as Gaussian, which leads to a different ML estimator [21], often termed stochastic ML (SML). Regardless of the actual distribution of the signal waveforms, both the NLLS and the SML methods provide highly accurate DOA estimates [22]. However, they both require solution of a d-dimensional nonlinear optimization problem like (16.27), which is far from trivial. A practical approach at a reasonable cost is to employ a relaxed optimization procedure, where DOA parameters are adjusted one at a time. Such a method is usually able to rapidly yield preliminary estimates in the neighborhood of the true minimizers of the criterion function, but the local convergence rate is often unacceptably slow. Therefore, once sufficiently good initial estimates are available, it is preferable to switch to a Newton-type local optimization (see, e.g., [22]). In [23, 24], a modified NLLS criterion is derived, based on subspace decomposition (16.17). In its concentrated form, the *signal subspace fitting (SSF) criterion* takes the form

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \operatorname{Tr}\{\boldsymbol{\Pi}_{\mathbf{A}}^{\perp}\hat{\mathbf{E}}_s\widehat{\mathbf{W}}\hat{\mathbf{E}}_s^*\}, \tag{16.28}$$

where $\widehat{\mathbf{W}}$ is a positive definite weighting matrix, which should reflect the reliability of the different eigenvectors appearing in $\hat{\mathbf{E}}_s$. The optimal choice of weighting, in terms of minimizing the variance of the DOA estimates, is the diagonal matrix

$$\widehat{\mathbf{W}} = \left(\hat{\Lambda}_s - \hat{\sigma}^2\mathbf{I}\right)^2 \hat{\Lambda}_s^{-1}, \tag{16.29}$$

where the noise variance estimate $\hat{\sigma}^2$ can be taken as the average of the $m - d$ smallest eigenvalues of $\hat{\mathbf{R}}$. The SSF formulation has certain advantages over the data-domain NLLS, in particular when $d \ll m$. It is then significantly cheaper to compute (16.28) than (16.27).

A few techniques have been proposed in the literature [25, 26, 27, 28] for implementing a relaxed optimization of NLLS-type criteria. We will describe the so-called RELAX procedure due to [26], which simultaneously updates θ_k and $s_k(t_n)$, $n = 1,\ldots,N$ while keeping the other parameters fixed. The method is closely related to the SAGE (space-alternating generalized expected maximization) algorithm of [25], although the motivation and interpretation is simpler for RELAX. We present the technique for the data-domain NLLS, noting that the SSF-RELAX algorithm is obtained by replacing the data matrix $\mathbf{X}$ with $\hat{\mathbf{E}}_s\widehat{\mathbf{W}}^{1/2}$. We express NLLS criterion (16.25) in matrix form as

$$V(\boldsymbol{\theta},\mathbf{S}) = \left|\left|\mathbf{X} - \mathbf{A}(\boldsymbol{\theta})\mathbf{S}\right|\right|^2 = \left|\left|\mathbf{X} - \sum_{k=1}^{d} \mathbf{a}(\theta_k)\mathbf{s}_k\right|\right|^2, \tag{16.30}$$

where $\|\cdot\|$ denotes the Frobenius matrix norm, $\mathbf{X} = [\mathbf{x}(t_1),\ldots,\mathbf{x}(t_N)]$, $\mathbf{S} = [\mathbf{s}(t_1),\ldots,\mathbf{s}(t_N)]$, and $\mathbf{s}_k$ is the kth row of $\mathbf{S}$. When searching for the parameters of the kth emitter, preliminary estimates of the other signal parameters are assumed available. Using these, the "cleaned" observation matrix

$$\mathbf{X}_l = \mathbf{X} - \sum_{k\neq l} \mathbf{a}(\hat{\theta}_k)\hat{\mathbf{s}}_k \tag{16.31}$$

is formed. The relaxed criterion

$$V_l(\theta_l,\mathbf{s}_l) = \left|\left|\mathbf{X}_l - \mathbf{a}(\theta_l)\mathbf{s}_l\right|\right|^2 \tag{16.32}$$

is again a separable NLLS criterion, which is minimized by

$$\hat{\theta}_l = \arg\max_{\theta} \frac{\left|\mathbf{a}^*(\theta)\mathbf{X}_l\right|^2}{\left|\left|\mathbf{a}(\theta)\right|\right|^2}, \tag{16.33}$$

$$\mathbf{s}_l = \frac{\mathbf{a}^*(\hat{\theta}_l)\mathbf{X}_l}{\left|\left|\mathbf{a}(\hat{\theta}_l)\right|\right|^2}. \tag{16.34}$$

Maximizing (16.33) requires a nonlinear optimization over one parameter only. This is easily accomplished by a course grid search (using FFT in the case of a

ULA), followed by a local Newton-type optimization. An iteration of the RELAX algorithm is now to sequentially update θ_l and $\mathbf{s}_l$ for $l = 1, \ldots, d$. At the first iteration, no initial estimates of θ_k and $\mathbf{s}_k$ are available for $k > l$, so these are usually ignored by letting $\mathbf{s}_k = 0$, $k > l$. The iterations continue until no significant change of the DOA parameters is observed, or until one chooses to switch to a Newton-type iterative search. In case the number of signals is unknown, a parametric enumeration technique (e.g., [29, 30]) is well suited to work with the iterative RELAX procedure.

16.3. Spread sources modeling

The DOA estimation problem for spread sources is quite different from the more commonly considered point source case. Since the individual rays are not resolvable, the task is to find a statistical characterization of the incoming energy. Parametric estimation methods are based on the availability of a parameterized PDF of θ. Popular choices are a Gaussian or a uniform PDF. For physics-based statistical models, consult, for example, [31, 32]. Nonparametric techniques make no such assumptions, but can on the other hand only deliver a limited description of the ray statistics, such as the nominal DOA θ_0 and the spread σ_θ.

16.3.1. Nonparametric beamforming techniques

Assume that the signal reaches the receiving antenna array via a single scatter cluster. Provided the distribution of scatterers (DOAs) has a symmetric PDF, a natural estimate of the nominal DOA θ_0 is the location of the peak of beamforming spectrum (16.13). However, for large spreading factors σ_θ, the variance of this estimate is unacceptably high [33]. The situation is even worse when using a high-resolution method [34]. A useful remedy is to use the center of gravity rather than the peak location of the spatial spectrum [35]. Provided the resolution of the latter is sufficient, it can be shown that $p_\theta(\theta)$ is approximately proportional to the spatial spectrum. For high SNR, the Capon spectrum has superior resolution, and may therefore be the preferred choice in this application. Using the definitions of the mean and variance of a distribution, the estimates are computed according to [35]

$$\begin{aligned} \widehat{\theta}_0 &= \frac{\int_{\theta\in\Omega} \theta P(\theta)\, d\theta}{\int_{\theta\in\Omega} P(\theta)\, d\theta}, \\ \widehat{\sigma}_\theta^2 &= \frac{\int_{\theta\in\Omega} (\theta - \widehat{\theta}_0)^2 P(\theta)\, d\theta}{\int_{\theta\in\Omega} P(\theta)\, d\theta}, \end{aligned} \tag{16.35}$$

where $P(\theta)$ is either $P_{\mathrm{BF}}(\theta)$ (16.13) or $P_{\mathrm{CAP}}(\theta)$ (16.16) and Ω refers to the support of the distribution, that is, the extent of the DOA cluster. The parameterization of $P(\theta)$ in terms of the physical DOA deserves some comments. For a linear array it is perhaps more natural to use the electrical angle ϕ, which is proportional to $\sin\theta$. However, for reasonably small DOA clusters the difference between the

parameterizations is negligible. The estimate $\hat{\theta}_0$ has in general very good performance, while $\hat{\sigma}_\theta$ is more sensitive to the choice of Ω and the resolution of $P(\theta)$. These parameters are particularly critical in the presence of several DOA clusters that are (nearly) overlapping.

16.3.2. Parametric estimation

A parametric estimator exploits the structure (16.8) of $\mathbf{R}_s$ in more detail. In general, this requires that a parameterized form of the DOA PDF is known, say $p_\theta(\theta;\boldsymbol{\phi})$. Here, $\boldsymbol{\phi}$ is a vector of unknown parameters, for example $\boldsymbol{\phi} = [\theta_0, \sigma_\theta]^T$ in the Gaussian case. Basically, these parameters are determined by matching the model (16.8) to the sample covariance $\hat{\mathbf{R}}$. An interesting "semiparametric" method was recently proposed in [36]. It is based on a generalization of the Capon formulation (16.14). Instead of requiring a unit gain in the look direction $\mathbf{a}(\theta)$, the authors of [36] proposed to keep a unit average signal power for a given set of parameters $\boldsymbol{\phi}$. The optimization problem is thus

$$\min_{\mathbf{w}} \mathbf{w}^* \hat{\mathbf{R}} \mathbf{w} \quad \text{subject to } \mathbf{w}^* \bar{\mathbf{R}}_s(\boldsymbol{\phi}) \mathbf{w} = 1, \tag{16.36}$$

where

$$\bar{\mathbf{R}}_s(\boldsymbol{\phi}) = \int_\theta \mathbf{a}(\theta)\mathbf{a}^*(\theta) p_\theta(\theta;\boldsymbol{\phi}) d\theta. \tag{16.37}$$

The resulting generalized Capon (GC) spectrum is obtained as

$$P_{\mathrm{GC}}(\boldsymbol{\phi}) = \frac{1}{\lambda_{\max}\{\hat{\mathbf{R}}^{-1}\bar{\mathbf{R}}_s(\boldsymbol{\phi})\}}, \tag{16.38}$$

where $\lambda_{\max}\{\cdot\}$ refers to the maximum eigenvalue of a matrix. The GC estimates are now taken as the locations of the highest peaks of $P_{\mathrm{GC}}(\boldsymbol{\phi})$. In this way, the parameters of several clusters can be determined using a search over only one set of parameters $\boldsymbol{\phi}$. The resolution in [36] is found to be superior to that of previously proposed computationally efficient estimators [37, 38, 39], although theoretical support for this claim is yet to be seen. The computational cost for the GC estimator is quite substantial, as it requires computing the maximum eigenvalue of an $m \times m$ matrix for each criterion function evaluation, besides solving the integral in (16.37). Approximate formulae for the latter is available in special cases [37, 40]. For a ULA and Gaussian-distributed DOAs, the approximate covariance matrix is given as

$$\bar{\mathbf{R}}_s(\theta_0, \sigma_\theta) \approx [\mathbf{a}(\theta_0)\mathbf{a}^*(\theta_0)] \odot \mathbf{B}, \tag{16.39}$$

where $\odot$ means elementwise multiplication, and the ijth element of the matrix $\mathbf{B}$ is given as

$$B_{ij} = e^{-2[\pi l(j-i)]^2 \sigma_\theta^2 \sin^2\theta_0}. \tag{16.40}$$

A general approach to parametric estimation is the principle of maximum likelihood (ML). If $\mathbf{x}(t)$ is modeled as zero-mean temporally white Gaussian, the ML estimate is found by solving

$$\{\hat{\boldsymbol{\phi}}, \hat{P}, \hat{\sigma}^2\} = \arg\min_{\boldsymbol{\phi}, P, \sigma^2} \log|\mathbf{R}_s| + \mathrm{Tr}\{\mathbf{R}_s^{-1}\hat{\mathbf{R}}\}. \tag{16.41}$$

Unfortunately, this is a multidimensional (4 in the Gaussian case) nonlinear optimization problem, even in the case of a single DOA cluster. In case of several clusters, the contributions from each will have to be added to $\mathbf{R}_s$, implying an even increased number of unknown parameters. The method is therefore often deemed impractical for the current application. A possible way to decrease the complexity is to use a covariance matching estimator (COMET), exploiting the fact that $\mathbf{R}_s$ is linear in the signal and noise powers as is clear from (16.8). The optimally weighted COMET estimator of [40] is formulated as the following optimization problem:

$$\{\hat{\boldsymbol{\phi}}, \hat{P}, \hat{\sigma}^2\} = \arg\min_{\boldsymbol{\phi}, P, \sigma^2} ||(\mathbf{R}_s - \hat{\mathbf{R}})\hat{\mathbf{R}}^{-1}||^2. \tag{16.42}$$

Since $\mathbf{R}_s$ is linear in P and σ^2, the above is in the form of a separable LS problem, and the minimum with respect to these parameters can easily be found for a fixed $\boldsymbol{\phi}$. The result is that only a search over $\boldsymbol{\phi}$ is necessary, similar to GC method (16.38). However, in the presence of several clusters, (16.42) requires a simultaneous search over all parameters. This is obviously more costly, but the benefit is an increased accuracy in large samples, since (16.42) can be shown to yield asymptotically the same performance as ML. For the special case of Gaussian-distributed DOA and a ULA, [41] presents a further simplification of (16.42). An approximate, but still asymptotically equivalent, technique is derived, where the search over $\boldsymbol{\phi} = [\theta_0, \sigma_\theta]^T$ is decoupled, so that only a 1D search is necessary. See [41] for details. A drawback of the covariance-matching-based methods is that they inherently assume a large sample size, and are thus less suited to scenarios where only a small number of observations are available.

16.4. Parametric channel modeling

Parametric channel modeling refers to the problem of finding a parsimonious representation of the wireless channel. The techniques described here use parameterizations in terms of physically meaningful quantities, but it is also possible to apply pure "black-box" methods. The data is assumed to be sensor outputs collected using a known transmitted waveform (probing signal). The task is to resolve all significant propagation paths and determine their DOAs, time delays, and strengths.

16.4.1. SIMO channels

The simplest case has a single transmitter and multiple receivers. This is referred to as a SIMO (single-input multiple-output) system. For narrowband signals, in

the sense that waveforms received via different propagation paths are all coherent, this is similar to the standard DOA estimation problem. However, since equipment for channel measurements is typically wideband, we consider the case where time delays are significant with respect to the inverse bandwidth. Referring to (16.7), the received signal is then modeled by

$$\mathbf{x}(t) = \sum_{k=1}^{d} g_k \mathbf{a}(\theta_k) s(t - \tau_k) + \mathbf{n}(t). \tag{16.43}$$

Here, $s(t)$ represents the transmitted continuous-time waveform, which for a digital communication system may be given in the form

$$s(t) = \sum_{l=-\infty}^{\infty} b_l p(t - lT), \tag{16.44}$$

where b_l are the (known) information symbols and $p(t)$ the pulse shaping waveform with support on $0 \le t < T$. The noise-free part of (16.43) can be thought of as the convolution of the transmitted signal $s(t)$ and the SIMO channel

$$\mathbf{h}(t) = \sum_{k=1}^{d} g_k \mathbf{a}(\theta_k) \delta(t - \tau_k), \tag{16.45}$$

where $\delta(\cdot)$ is the Dirac delta function. The high bandwidth and multiantenna receiver allows paths to be resolved in both space and time. We therefore assume that one wishes to obtain signal parameter estimates for the individual components in (16.43), in contrast to the modeling of spatially extended sources considered in Section 16.3. Given samples $\mathbf{x}(t_n)$, $n = 1, \ldots, N$ from (16.43), where the shape of $s(t)$ is assumed known, the task is to jointly determine θ_k, τ_k, and g_k. Often, the channel-sounder equipment does not give access to the raw data, but only nonparametric channel estimates. This case will be considered later. In essence, we are facing a 2D estimation problem. To reduce the computational complexity, such problems are often solved by treating one dimension at a time in some fashion. If the noise is Gaussian and spatially white, the optimal ML estimator employs the 2D NLLS criterion

$$\begin{aligned} V(\boldsymbol{\theta}, \boldsymbol{\tau}, \mathbf{g}) &= \sum_{k=1}^{N} \left\| \mathbf{x}(t_k) - \sum_{l=1}^{d} g_l \mathbf{a}(\theta_l) s(t_k - \tau_l) \right\|^2 \\ &= \left\| \mathbf{X} - \sum_{l=1}^{d} g_l \mathbf{a}(\theta_l) \mathbf{s}(\tau_l) \right\|_F^2, \end{aligned} \tag{16.46}$$

where $\mathbf{X} = [\mathbf{x}(t_1), \ldots, \mathbf{x}(t_N)]$ as before, and

$$\mathbf{s}(\tau_k) = [s(t_1 - \tau_k), \ldots, s(t_N - \tau_k)] \tag{16.47}$$

denotes the time-delayed signal vector. For fixed $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_d]$ and $\boldsymbol{\tau} = [\tau_1, \ldots, \tau_d]$, it is easy to minimize (16.46) explicitly with respect to $\mathbf{g} = [g_1, \ldots, g_d]$. Substituting the so-obtained $\mathbf{g}(\boldsymbol{\theta}, \boldsymbol{\tau})$ back into (16.46) results in a nonlinear function of $2d$ parameters. To find the global minimum of this concentrated NLLS criterion is by no means a simple task. A natural approach is to apply the RELAX idea presented in Section 16.2.3, see also [42] for an application to time-delay estimation. This is an iterative procedure, where the joint estimation problem is in each iteration broken down into computationally simple beamforming-like (i.e., matched filter) operations. The technique has strong similarities with the SAGE algorithm, which was successfully applied in [43] to a scenario where also the Doppler-shifts due to the platform motion were taken into account. Similar to (16.31), (16.32), and (16.33), the RELAX algorithm for joint DOA and delay estimation consists of the following steps. First the data is "cleaned" from contributions from all signal paths except one:

$$\mathbf{X}_l = \mathbf{X} - \sum_{k \neq l} \hat{g}_k \mathbf{a}(\hat{\theta}_k)\mathbf{s}(\hat{\tau}_k), \tag{16.48}$$

where $\hat{(\cdot)}$ denotes the most recent estimates of the signal parameters corresponding to the other paths. The relaxed criterion is now

$$V_l(\theta_l, \tau_l, g_l) = ||\mathbf{X}_l - g_l \mathbf{a}(\theta_l)\mathbf{s}(\tau_l)||^2. \tag{16.49}$$

Minimizing explicitly with respect to g_l results after some simple algebra in the concentrated criterion

$$V_l(\theta_l, \tau_l) = \frac{|\,\mathbf{a}^*(\theta_l)\mathbf{X}_l\mathbf{s}^*(\tau_l)\,|^2}{||\mathbf{a}(\theta_l)||^2||\mathbf{s}(\tau_l)||^2}, \tag{16.50}$$

which is to be *maximized* with respect to θ_l and τ_l. The above allows a nice interpretation. The premultiplication by $\mathbf{a}^*(\theta_l)$ is a spatially matched filter, and the postmultiplication by $\mathbf{s}^*(\tau_l)$ acts as a temporally matched filter. The maximum of this normalized *space-time correlation* yields the estimates for the lth propagation path. Once updated estimates of θ_l and τ_l are found, the new g_l will be

$$\hat{g}_l = \frac{\mathbf{a}^*(\hat{\theta}_l)\mathbf{X}_l\mathbf{s}^*(\hat{\tau}_l)}{||\mathbf{a}(\hat{\theta}_l)||^2||\mathbf{s}(\hat{\tau}_l)||^2}. \tag{16.51}$$

One iteration of the RELAX algorithm is now to sequentially update the triple $\{\theta_l, \tau_l, g_l\}$ for $l = 1, \ldots, d$ as outlined above. At the first iteration one simply takes $g_k = 0$, $k > l$, unless some initial estimates of the parameters of these propagation paths are available.

The iterative procedure described above is simple in each step. But in high-resolution scenarios, several iterations may be necessary, implying a high total computational burden. Simpler subspace-based alternatives have been proposed

in, for example, [44]. The computationally most efficient techniques employ a Fourier transform of a nonparametric channel estimate along the delay dimension. Such a vector-valued impulse response estimate is obtained using a matched filter, which in matrix form is expressed as

$$\widehat{\mathbf{h}}(t) = \frac{\mathbf{X}\mathbf{s}^*(t)}{||\mathbf{s}(t)||^2}. \tag{16.52}$$

In fact, the available channel-sounder data is often given in the form of estimated impulse responses like the above. A natural approach is now to apply spatial beamforming to each available sample of the impulse response, say $\widehat{\mathbf{h}}(t_k)$, $k = 1,\ldots,M$, where $t_k = (k-1)T$ for uniform sampling. In case of a uniform linear array, each $\widehat{\mathbf{h}}(t_k)$ can be split into several subvectors, thus effectively creating several snapshots. Using this *spatial smoothing* technique, it is also possible to apply a high-resolution DOA estimation method, thus enabling closely spaced scatterers with (nearly) the same propagation delay to be resolved. Taking the discrete Fourier transform of $\widehat{\mathbf{h}}(t_k)$ and using (16.45), the nonparametric SIMO transfer function estimate is modeled by

$$\widehat{\mathbf{h}}(\omega_k) \approx \sum_{l=1}^{d} g_l \mathbf{a}(\theta_l) e^{-j\omega_k \tau_l}, \tag{16.53}$$

where $\omega_k = 2\pi(k-1)/M$, $k = 1,\ldots,M$ are the length-M DFT frequencies. The approximation above acknowledges the errors in (16.52), which are due to both noise and to finite temporal resolution. In effect, the Fourier transform has turned the delay estimation into one of frequency estimation, where τ_k plays the role of frequencies. If the array is a ULA, a computationally efficient technique for 2D frequency estimation (e.g., [45]) can now be employed to determine the signal parameters, as in [44].

16.4.2. MIMO channels

During the past several years there has been a growing interest in systems that employ multiple antennas both at the transmitter and the receiver, that is, MIMO systems. The main reason for this is a potential for dramatically increased capacity. From a propagation modeling point of view, MIMO transmission opens up an interesting new possibility. Besides finding the direction of the incoming rays, it is also possible to determine the directions-of-departure (DOD) of these rays [46], thus providing a more complete description of the wireless channel. Assuming a point source model, the received signal is now modeled as

$$\mathbf{x}(t) = \sum_{k=1}^{d} g_k \mathbf{a}_{rx}(\theta_k) \mathbf{a}_{tx}^T(\eta_k) \mathbf{s}(t-\tau_k) + \mathbf{n}(t). \tag{16.54}$$

Here, $\mathbf{a}_{tx}(\eta)$ denotes the $m_{tx} \times 1$ steering vector for the transmitter array, where η is the DOD, whereas $\mathbf{a}_{rx}(\theta)$ is the $m_{rx} \times 1$ steering vector for the receiving array with θ the DOA. Further, $\mathbf{s}(t) = [s_1(t), \ldots, s_{m_{tx}}(t)]^T$ contains the transmitted signal waveforms from each of the transmitting antennas (which are potentially different). As before, g_k is the complex gain of the kth path and τ_k is the time delay. Given $\mathbf{x}(t_n)$, $n = 1, \ldots, N$ from (16.54), where $\mathbf{s}(t)$ needs to be known to allow time-delay estimation, one desires to estimate all unknown parameters θ_k, η_k, τ_k, and g_k.

In matrix form, the NLLS criterion now takes the form

$$V(\boldsymbol{\theta}, \boldsymbol{\eta}, \boldsymbol{\tau}, \mathbf{g}) = \left\| \mathbf{X} - \sum_{l=1}^{d} g_l \mathbf{a}_{rx}(\theta_l) \mathbf{a}_{tx}^T(\eta_l) \mathbf{S}(\tau_l) \right\|^2, \tag{16.55}$$

where

$$\mathbf{S}(\tau_l) = [\mathbf{s}(t_1 - \tau_l), \ldots, \mathbf{s}(t_n - \tau_l)] \tag{16.56}$$

is the matrix of delayed transmit signals. The above is a 3D estimation problem. There is no conceptual difference to the 1D case, but the computational complexity associated with finding the signal parameters increases of course dramatically as new dimensions are opened up. The RELAX approach applied to (16.55) results in the following steps. First, a data cleaning is performed by

$$\mathbf{X}_l = \mathbf{X} - \sum_{k \neq l} g_k \mathbf{a}_{rx}(\theta_k) \mathbf{a}_{tx}^T(\eta_k) \mathbf{s}(t - \tau_k). \tag{16.57}$$

The concentrated criterion for the lth signal path is then

$$V_l(\theta_l, \eta_l, \tau_l) = \frac{|\, \mathbf{a}_{rx}^*(\theta_l) \mathbf{X}_l \mathbf{S}^*(\tau_l) \bar{\mathbf{a}}_{tx}(\eta_l) \,|^2}{||\mathbf{a}_{rx}(\theta_l)||^2 ||\mathbf{S}^*(\tau_l) \bar{\mathbf{a}}_{tx}(\eta_l)||^2}. \tag{16.58}$$

Again, we have the interpretation of space-time beamforming, but now in two spatial dimensions. The $\mathbf{a}_{rx}^*(\theta_l)$ does spatial beamforming along the rows of $\mathbf{X}_l$ (the receivers), whereas $\mathbf{S}^*(\tau_l)$ is a bank of temporal match filters, where each column (row of $\mathbf{S}(\tau_l)$) corresponds to one transmitter. The outputs of these filters are then weighted together by the transmit beamformer $\bar{\mathbf{a}}_{tx}(\eta_l)$, where $(\bar{\cdot})$ denotes complex conjugate. The description of the 3D RELAX algorithm is completed by the update of the amplitude parameter:

$$\hat{g}_l = \frac{\mathbf{a}_{rx}^*(\hat{\theta}_l) \mathbf{X}_l \mathbf{S}^*(\hat{\tau}_l) \bar{\mathbf{a}}_{tx}(\hat{\eta}_l)}{||\mathbf{a}_{rx}(\hat{\theta}_l)||^2 ||\mathbf{S}^*(\hat{\tau}_l) \bar{\mathbf{a}}_{tx}(\hat{\eta}_l)||^2}. \tag{16.59}$$

Running the above steps for $l = 1, \ldots, d$ and iterating until convergence, results ultimately in an approximate solution to the original NLLS problem (16.55).

The above assumes that the time-domain array data are available. Alternatively, the matrix-valued impulse response can be estimated by

$$\hat{\mathbf{H}}(t) = \mathbf{X}\mathbf{S}^*(t)\left(\mathbf{S}(t)\mathbf{S}^*(t)\right)^{-1}, \tag{16.60}$$

or it may be given directly from the measurement equipment. Here, the ijth element of $\hat{\mathbf{H}}(t)$ contains the estimated impulse response from transmitter i to receiver j. Similar to (16.53), the Fourier transform of $\hat{\mathbf{H}}(t)$ is modeled by

$$\hat{\mathbf{H}}(\omega_k) \approx \sum_{l=1}^{d} g_l \mathbf{a}_{rx}(\theta_l)\mathbf{a}_{tx}^T(\eta_l)e^{-j\omega_k\tau_l}, \quad k = 1,\ldots,M. \tag{16.61}$$

Either $\hat{\mathbf{H}}(t)$ or $\hat{\mathbf{H}}(\omega_k)$ can now be considered as measurement data, the task again being to determine the signal parameters g_l, θ_l, η_l, and τ_l for $l = 1,\ldots,d$. Several suboptimal techniques have been proposed to attack this problem. For example, in [47], the 2D ESPRIT method with spatial smoothing is applied to each sample $\hat{\mathbf{H}}(t_k)$ (the application in [47] is azimuth-elevation estimation, but the same approach can be applied to the DOD-DOA model). In [46], the delays are first determined using the 1D ESPRIT technique. A 2D RELAX-like approach is then applied to the so-obtained channel samples $\hat{\mathbf{H}}(\hat{\tau}_l)$. In [48], a 3D subspace method is proposed, assuming several observations of the matrix-valued transfer function $\hat{\mathbf{H}}(\omega_k)$ to be available. Finally, special methods for the case were only beamformed transmit and receive data are available are presented in [49, 50].

16.5. Summary

Direction-of-arrival (DOA) estimation is to a great extent a mature research area. We have in this chapter made an attempt to summarize the most influential methods to determine the signal parameters. For the standard DOA estimation problem, the nonlinear least-squares (NLLS) approach provides the most accurate estimates of the methods presented herein. An iterative method for computing the estimates was outlined, based on the RELAX idea presented in [26]. Subspace-based methods are computationally more attractive, especially the ESPRIT-type techniques, whenever applicable. These methods can give high resolution at moderate cost, provided the SNR and/or the number of available samples is sufficiently high. The conventional beamforming approach is applicable only when the array size is sufficiently large, whereas adaptive beamforming (Capon's method) can give increased resolution at high enough SNR.

Besides the standard DOA estimation problem, we have also considered the case where sources have a significant spatial extent, as compared to the array resolution. In this case, the incoming radiation is characterized in statistical terms, rather than as a point source only. Most commonly, the mean and the standard deviation of the DOA parameter are sought. In this case, beamforming-type methods are more likely to be useful than in the point-source case. However, the center of gravity of the beamforming spectrum should be used for estimation of the

mean DOA, rather than its peak location. Among high-resolution methods, a recent generalization of Capon's method due to [36] is found to be very promising. Finally, the case of parametric modeling of the communication channel, including propagation delay was also addressed. This leads to difficult multidimensional estimation problems, where easy solutions are likely to fail and optimal techniques are difficult to implement in practice. Iterative RELAX-type methods were presented for the various cases. Also subspace methods are available in the literature. These, and hybrids between subspace and NLLS have been found to give satisfactory results with real data, see, for example, [46, 47, 48].

Abbreviations

COMET	Covariance matching estimator
DFT	Discrete Fourier transform
DOA	Direction-of-arrival
DOD	Direction-of-departure
ESPRIT	Estimation of signal parameters via rotational invariance techniques
FFT	Fast Fourier transform
GC	Generalized Capon
LOS	Line-of-sight
LS	Least-squares
MIMO	Multiple-input multiple-output
ML	Maximum likelihood
MUSIC	Multiple signal classification
NLLS	Nonlinear least-squares
PDF	Probability density function
SAGE	Space-alternating generalized expected maximization
SIMO	Single-input multiple-output
SML	Stochastic maximum likelihood
SNR	Signal-to-noise ratio
SSF	Signal subspace fitting
ULA	Uniform linear array

Bibliography

[1] P. Stoica and R. Moses, *Introduction to Spectral Analysis*, Prentice Hall, Upper Saddle River, NJ, USA, 1997.

[2] W. Jakes, Ed., *Microwave Mobile Communications*, Wiley-Interscience, New York, NY, USA, 1974.

[3] A. Kavak, W. Yang, G. Xu, and W. J. Vogel, "Characteristics of vector propagation channels in dynamic mobile scenarios," *IEEE Trans. Antennas Propagat.*, vol. 49, no. 12, pp. 1695–1702, 2001.

[4] H. Krim and M. Viberg, "Two decades of array signal processing research: the parametric approach," *IEEE Signal Processing Mag.*, vol. 13, no. 4, pp. 67–94, 1996.

[5] H. L. Van Trees, *Optimum Array Processing*, John Wiley & Sons, Canada, 2002.

[6] J. Capon, "High resolution frequency wave number spectrum analysis," *Proc. IEEE*, vol. 57, pp. 1408–1418, 1969.

[7] G. V. Serebryakov, "Direction-of-arrival estimation of correlated sources by adaptive beamforming," *IEEE Trans. Signal Processing*, vol. 43, no. 11, pp. 2782–2787, 1995.

[8] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 33, no. 2, pp. 387–392, 1985.

[9] L. C. Zhao, P. R. Krishnaiah, and Z. D. Bai, "On detection of the number of signals in presence of white noise," *J. Multivariate Analysis*, vol. 20, no. 1, pp. 1–25, 1986.

[10] G. Bienvenu and L. Kopp, "Principe de la goniométrie passive adaptative," in *Proc. 7'eme Colloque sur le Traitement du Signal et ses Applications (GRETSI '79)*, pp. 106/1–106/10, Nice, France, 1979.

[11] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," in *Proc. RADC Spectrum Estimation Workshop*, pp. 243–258, Rome, NY, USA, October 1979.

[12] R. O. Schmidt, *A signal subspace approach to multiple emitter location and spectral estimation*, Ph.D. thesis, Stanford University, Stanford, Calif, USA, November 1981.

[13] M. Wax and I. Ziskind, "On unique localization of multiple sources by passive sensor arrays," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 37, no. 7, pp. 996–1000, 1989.

[14] W. Xu and M. Kaveh, "Comparative study of the biases of MUSIC-like estimators," *Signal Processing*, vol. 50, no. 1-2, pp. 39–55, 1996.

[15] A. J. Barabell, "Improving the resolution performance of eigenstructure-based direction-finding algorithms," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP '83)*, vol. 8, pp. 336–339, Cambridge, Mass, USA, April 1983.

[16] B. D. Rao and K. V. S. Hari, "Performance analysis of root-music," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 37, no. 12, pp. 1939–1949, 1989.

[17] R. Roy and T. Kailath, "ESPRIT—estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 37, no. 7, pp. 984–995, 1989.

[18] M. Haardt and J. Nossek, "Unitary ESPRIT: how to obtain increased estimation accuracy with a reduced computational burden," *IEEE Trans. Signal Processing*, vol. 43, no. 5, pp. 1232–1242, 1995.

[19] B. Friedlander and A. Weiss, "Direction finding using spatial smoothing with interpolated arrays," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 28, no. 2, pp. 574–587, 1992.

[20] J. F. Böhme, "Estimation of source parameters by maximum likelihood and nonlinear regression," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP '84)*, vol. 9, pp. 271–274, San Diego, Calif, USA, March 1984.

[21] J. F. Böhme, "Separated estimation of wave parameters and spectral parameters by maximum likelihood," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP '86)*, vol. 11, pp. 2818–2822, Tokyo, Japan, April 1986.

[22] B. Ottersten, M. Viberg, P. Stoica, and A. Nehorai, "Exact and large sample ML techniques for parameter estimation and detection in array processing," in *Radar Array Processing*, S. S. Haykin, J. Litva, and T. J. Shepherd, Eds., pp. 99–151, Springer, Berlin, Germany, 1993.

[23] P. Stoica and K. Sharman, "Maximum likelihood methods for direction-of-arrival estimation," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 38, no. July, pp. 1132–1143, 1990.

[24] M. Viberg and B. Ottersten, "Sensor array processing based on subspace fitting," *IEEE Trans. Signal Processing*, vol. 39, no. 5, pp. 1110–1121, 1991.

[25] J. A. Fessler and A. O. Hero, "Space-alternating generalized expectation-maximization algorithm," *IEEE Trans. Signal Processing*, vol. 42, no. 10, pp. 2664–2677, 1994.

[26] J. Li, D. Zheng, and P. Stoica, "Angle and waveform estimation via RELAXE," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 33, no. 3, pp. 1077–1087, 1997.

[27] P. Pelin, "A fast minimization technique for subspace fitting with arbitrary array manifolds," *IEEE Trans. Signal Processing*, vol. 49, no. 12, pp. 2935–2939, 2001.

[28] I. Ziskind and M. Wax, "Maximum likelihood localization of multiple sources by alternating projection," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 36, no. October, pp. 1553–1560, 1988.

[29] M. Viberg, B. Ottersten, and T. Kailath, "Detection and estimation in sensor arrays using weighted subspace fitting," *IEEE Trans. Signal Processing*, vol. 39, no. 11, pp. 2436–2449, 1991.

[30] M. Wax, "Detection and localization of multiple sources via the stochastic signals model," *IEEE Trans. Signal Processing*, vol. 39, no. 11, pp. 2450–2456, 1991.

[31] J. B. Andersen and K. I. Pedersen, "Angle-of-arrival statistics for low resolution antennas," *IEEE Trans. Antennas Propagat.*, vol. 50, no. 3, pp. 391–395, 2002.

[32] R. B. Ertel, P. Cardieri, K. W. Sowerby, T. S. Rappaport, and J. H. Reed, "Overview of spatial channel models for antenna array communication systems," *IEEE Pers. Commun.*, vol. 5, no. 1, pp. 10–22, 1998.

[33] R. Raich, J. Goldberg, and H. Messer, "Bearing estimation for a distributed source via the conventional beamformer," in *Proc. 9th IEEE Signal Processing Statistical Signal and Array Processing*, pp. 5–8, Portland, Ore, USA, September 1998.

[34] D. Astély and B. Ottersten, "The effects of local scattering on direction of arrival estimation with MUSIC," *IEEE Trans. Signal Processing*, vol. 47, no. 12, pp. 3220–3234, 1999.

[35] M. Tapio, "On the use of beamforming for estimation of spatially distributed signals," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP '03)*, vol. 3, pp. 3005–3008, Hong Kong, May 2003.

[36] A. Hassanien, S. Shahbazpanahi, and A. Gershman, "A generalized Capon estimator for multiple spread sources," *IEEE Trans. Signal Processing*, vol. 52, no. 1, pp. 280–283, 2004.

[37] M. Bengtsson and B. Ottersten, "Low-complexity estimators for distributed sources," *IEEE Trans. Signal Processing*, vol. 48, no. 8, pp. 2185–2194, 2000.

[38] Y. Meng, P. Stoica, and K. M. Wong, "Estimation of the direction of arrival of spatially dispersed sources," *IEE Proc. Radar, Sonar and Navigation*, vol. 143, no. 1, pp. 1–9, 1996.

[39] S. Valaee, B. Champagne, and P. Kabal, "Parametric localization of distributed sources," *IEEE Trans. Signal Processing*, vol. 43, no. 7, pp. 2144–2153, 1995.

[40] T. Trump and B. Ottersten, "Estimation of nominal direction of arrival and angular spread using and array of sensors," *Signal Processing*, vol. 50, no. 1-2, pp. 57–70, 1996.

[41] O. Besson and P. Stoica, "Decoupled estimation of DOA and angular spread for a spatially distributed source," *IEEE Trans. Signal Processing*, vol. 48, no. 7, pp. 1872–1882, 1983.

[42] R. Wu, J. Li, and Z.-S. Liu, "Super resolution time delay estimation via MODE-WRELAX," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 35, no. 1, pp. 294–307, 1999.

[43] K. I. Pedersen, B. Fleury, and P. E. Mogensen, "High resolution of electromagnetic waves in time-varying radio channels," in *Proc. 8th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '97)*, vol. 2, pp. 650–654, Helsinki, Finland, September 1997.

[44] M. Vanderveen, A.-J. van der Veen, and A. Paulraj, "Estimation of multipath parameters in wireless communications," *IEEE Trans. Signal Processing*, vol. 46, no. 3, pp. 682–690, 1998.

[45] S. Rouquette and M. Najim, "Estimation of frequencies and damping factors by two-dimensional ESPRIT type methods," *IEEE Trans. Signal Processing*, vol. 49, no. 1, pp. 237–245, 2001.

[46] M. Steinbauer, A. F. Molisch, and E. Bonek, "The double-directional radio channel," *IEEE Antennas Propagat. Mag.*, vol. 43, no. 4, pp. 51–63, 2001.

[47] J. Fuhl, J.-P. Rossi, and E. Bonek, "High-resolution 3-D direction-of-arrival determination for urban mobile radio," *IEEE Trans. Antennas Propagat.*, vol. 45, no. 4, pp. 672–682, 1997.

[48] M. Pesavento, C. F. Mecklenbräuker, and J. F. Böhme, "Double-directional radio channel estimation using MD-RARE," in *Proc. 36th Asilomar Conference on Signals, Systems and Computers*, vol. 1, pp. 594–598, Pacific Grove, Calif, USA, November 2002.

[49] B. D. Jeffs, E. Pyper, and B. Hunter, "A wireless MIMO channel probing approach for arbitrary antenna arrays," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP '01)*, vol. 4, pp. 2493–2496, Salt Lake City, Utah, USA, May 2001.

[50] T. Zwick, D. Hampicke, J. Maurer, et al., "Results of double-directional channel sounding measurements," in *Proc. IEEE Vehicular Technology Conference (VTC '00)*, vol. 3, pp. 2497–2501, Tokyo, Japan, May 2000.

Mats Viberg: Department of Signals and Systems, Chalmers University of Technology, 412 96 Göteborg, Sweden

Email: viberg@chalmers.se