# Algebraic Methods for Deterministic Blind Beamforming

ALLE-JAN VAN DER VEEN, MEMBER, IEEE

*Invited Paper*

*Deterministic blind beamforming algorithms try to separate superpositions of source signals impinging on a phased antenna array by using deterministic properties of the signals or the channels such as their constant modulus or directions-of-arrival. Progress in this area has been abundant over the past ten years and has resulted in several powerful algorithms. Unlike optimal or adaptive methods, the algebraic methods discussed in this review act on a fixed block of data and give closed-form expressions for beamformers by focusing on algebraic structures. This typically leads to subspace estimation and generalized eigenvalue problems. After introducing a simple and widely used multipath channel model, the paper provides an anthology of properties that are available, as well as generic algorithms that exploit them.*

***Keywords*—** *Array signal processing, blind source separation, constant modulus algorithm, delay estimation, direction of arrival estimation, frequency estimation, multipath channels, sequence estimation, singular value decomposition, space division multiplexing.*

## I. INTRODUCTION

In the context of array signal processing, beamforming is concerned with the reconstruction of source signals from the outputs of a sensor array. This can be done either by coherently adding the contributions of the desired source or by nulling out the interfering sources. The latter is an instance of the more general problem of source separation.

Classically, beamforming requires knowledge of a look direction, which is the direction of the desired source. Blind beamforming tries to recover source signals without this information, relying instead on various structural properties of the problem.

The first blind beamforming techniques proposed were based on direction finding. The direction of each incoming wavefront is estimated, at the same time producing a beamformer to recover the signal from that direction. This requires that at least that the antenna array is calibrated. If a source comes in via several directions (coherent multipath), then direction finding is more complicated. Depending on the situation, we also need to consider delay spread. Thus, the applicability of these techniques is very much dependent on the channel conditions and in general requires a small number of well-defined propagation paths per source.

More recently, new types of blind beamformers have been proposed that are not based on specific channel models, but instead exploit properties of the signals. A striking example is the constant modulus algorithm (CMA), which separates sources based on the fact that their baseband representation has a constant amplitude, such as is the case for FM or phase modulated signals. A prime advantage is that these beamformers are not dependent on channel properties or array calibration. For man-made signals, such as those encountered in wireless communications, signal properties are often well known and accurate, leading to robust algorithms. Several other properties are available, for example, cyclostationarity caused by the bauded nature of digital communication signals or introduced by small differences in carrier frequencies. Ultimately, sources can be separated based on their statistical independence, which is a somewhat weaker, but generally valid property.

*1) Deterministic Blind Beamforming:* In view of the above, it is clear that blind beamforming is a wide field, even if we limit ourselves to source separation. To restrict ourselves further, we will not consider stochastic techniques here at all (cf. the paper by Cardoso in this issue), and address cyclostationarity properties only marginally. This leaves a field that can be called "deterministic blind beamforming," which makes strong structural assumptions on the scenario, but in exchange requires only a modest number of samples. In particular, deterministic methods do not exploit the source statistics, but the can can provide exact results based on only a finite amount of data, at least under noise-free conditions. They are usually derived by first looking at how a source separation problem could be solved in the absence of noise, and then making sure that the algorithm still behaves robustly when noise is added. This often leads to elegant algorithms that have good performance, albeit suboptimal from a statistical point of view.

Manuscript received May 13, 1997; revised February 15, 1998. This work was supported by ENST, Paris, France.

The author is with the Delft University of Technology, Department of Electrical Engineering/DIMES, 2628 CD Delft, The Netherlands.

Publisher Item Identifier S 0018-9219(98)06971-0.

A second distinction is that we will only look at algebraic techniques acting on a block of data, as opposed to adaptive (updating) techniques useful for tracking. The latter are often gradient descent techniques based on cost-function minimization, where the cost derives from forcing one of the structural properties or from a maximum-likelihood (ML) criterion. Updating algorithms generally have a lower computational complexity and can track a nonstationary channel, but they place a larger demand on the number of samples and need time to converge so that their relevance depends on the requirements of the application. Issues are unpredictable convergence speed, possible convergence to suboptimal solutions, and initialization of the iteration. A considerable problem in the context of source separation with adaptive techniques is that of recovering *all* independent signals. In contrast, the algebraic techniques considered here typically find all separating beamformers jointly as the collection of eigenvectors of an associated eigenvalue problem. This makes them more reliable, but at a computational cost. Also, a model order selection is essential but often not trivial. The simplicity of the adaptive techniques have made these the only algorithms that have been implemented in actual current-day systems (cf. the paper by Treichler, Larimore, and Harp, this issue). With the advent of powerful DSP's and more finite-data burst oriented problems, this may change in the future.

Similar problems with local minima and initialization hold for optimal ML techniques, which act on a block of data and try to optimize an often highly nonlinear cost function at great computational expenses. Algebraic techniques can provide a good initial point in the search for the optimal solution. For small sample sizes, the benefit of the optimization step is not necessarily worth the effort.

A related topic is that of blind identification or equalization of convolutive channels, which is very similar except that more structure is available and only one signal is to be recovered (the others being echoes). Blind equalization is discussed in depth in the paper by Tong and Perreau in this issue (see also [50]). The main distinctive point in blind beamforming considered here is the interest in recovering all impinging signals.

The paper is thus centered around algebraic techniques for deterministic blind beamforming. We consider two classes of algorithms: those that are based on channel properties and others based on signal properties. Despite the fact that these properties are widely differing, the resulting algorithms show a remarkable homogeneity. All are subspace-based techniques and end with a generalized eigenvalue problem: the beamformers are found as the eigenvectors of a simultaneous diagonalization problem in which several matrices can be diagonalized by the same (eigenvector) matrix. The message of the paper is that joint diagonalization is *the* fundamental problem for source separation.

*2) Application Example:* By nature of this class of algorithms (i.e., they act on short data blocks with very specific structures) we will be mostly interested in applications for wireless communications. An example of a blind beamforming application in this area is the separation of aircraft transponder signals. Civil air traffic control uses a "secondary surveillance radar" (SSR) to identify and track aircraft [121], [122]. After interrogation by a ground radar station, the aircraft responds with a short data burst, providing information on its call number, airspeed, and altitude. In the newly developed SSR Mode-S, aircraft are individually addressable, but implementation of this standard has been slow. The system as it is currently used has a single carrier frequency at 1090 MHz for all return signals. It frequently occurs that several aircraft are triggered by an interrogation beam, so that ground stations receive a superposition of several data bursts, partially overlapping in time and frequency. Data bursts are short (56 or 112 bits) and do not contain training symbols. Thus, it would be very interesting to separate two or three of such messages using blind beamforming techniques. Besides direction finding, there are several opportunities for this, since signals are stochastically independent, carriers are not exactly the same (there is a tolerance of 3 MHz), and the data modulation is simple (pulse-amplitude modulation with alphabet $\{0, 1\}$) [102], [123].

*3) Outline:* The paper first introduces a compact data model by which multipath propagation channels can be described (Sections II and III). We distinguish between instantaneous and convolutive models. This is followed by an overview of properties that are available in this context (Section IV) which forms the center of the paper. The second part is a more detailed anthology of example algorithms (Sections VI and VII), which, starting at a moderate level, requires an increasing proficiency in linear algebra techniques on the part of the reader.

## II. Physical Channel Model

The propagation of signals through a radio channel is fairly complicated to model. A correct treatment would require a complete description of the physical environment, which is not very suitable for the design of signal processing algorithms. To arrive at a more useful parametric model we have to make simplifying assumptions regarding the wave propagation. Provided this model is reasonably valid, we can, in a second stage, try to derive statistical models for the parameters to obtain agreement with measurements. The purpose of this section is to discuss a simple channel model that can be used for array signal processing.

### A. Delays of Narrow-Band Signals

Let us start with a well-known but important property of narrow-band signals which says that a short time delay translates to a phase shift. In signal processing, narrow-band signals are usually represented by their lowpass equivalents [6]. A real-valued bandpass signal with center frequency $f_c$, such as received by an antenna, can be written as

$$z(t) = \text{real}\{s(t)e^{j2\pi f_c t}\}$$

where the baseband signal $s(t)$ is the complex envelope of the received signal $z(t)$. It is obtained from $z(t)$ by demod-

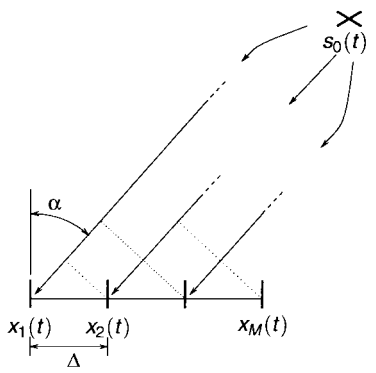



**Fig. 1.** A linear array receiving a far field point source.

ulation: multiplying the received signal with $\cos(2\pi f_c t)$ and $\sin(2\pi f_c t)$ followed by low-pass filtering.

In array signal processing, we are interested in the effect of small delays on the narrow-band baseband signal $s(t)$. A delay $\tau$ on $z(t)$ results in

$$z_\tau(t) := z(t-\tau) = \text{real}\{s(t-\tau)e^{-j2\pi f_c\tau}e^{j2\pi f_c t}\}$$

so that the complex envelope of the delayed signal is $s_\tau(t) = s(t-\tau)e^{-j2\pi f_c\tau}$. If the bandwidth $W$ of $s(t)$ is sufficiently small so that $\exp(j2\pi W\tau) \approx 1$, then standard Fourier analysis yields $s(t-\tau) \approx s(t)$, so that

$$s_\tau(t) \approx s(t)e^{-j2\pi f_c\tau} \qquad \text{for } W\tau \ll 1.$$

The well-known conclusion is that, for narrow-band signals, time delays shorter than the inverse bandwidth amount to phase shifts of the baseband signal. This is fundamental in phased array signal processing.

### *B. Antenna Array Response*

Let us consider a simple linear array consisting of $M$ identical antenna elements, as in Fig. 1. A narrow-band point source $s_0(t)$ is present in the far field and is modulated at carrier frequency $f_c$. If the distance between the array and the source is large in comparison to the extent of the array, the wave incident on the array is approximately planar. The angle $\alpha$ to the normal is the direction of arrival (DOA) of the plane wave.

Let $a(t;\,\alpha)$ be the response of a single antenna element to a signal from direction $\alpha$. We usually assume that the frequency response of the antenna is flat over the band of interest: $a(t;\,\alpha) = a(\alpha)\delta(t)$, where $a(\alpha)$ is the antenna gain pattern. If the antennas are omnidirectional, then $a(\alpha) = a$, a constant scalar.

The baseband signal at the location of the first (reference) antenna element is called $s(t)$; it differs from $s_0(t)$ by a delay and a complex attenuation (the path loss). The signal received by an antenna at a distance of $\Delta$ wavelengths from the reference location experiences an additional delay $\tau$. If $\tau$ is small compared to the inverse bandwidth of $s(t)$, we may set $s_\tau(t) = s(t)e^{-j2\pi f_c\tau}$, where the phase shift can be related to the angle of arrival $\alpha$ by

$$\theta = e^{-j2\pi f_c\tau} = e^{j2\pi\Delta\,\sin(\alpha)}.$$

An antenna array with elements at locations $\Delta_i$ receives signals $x_i(t) = a(\alpha)s_{\tau_i}(t) = a(\alpha)e^{j2\pi\Delta_i\,\sin(\alpha)}s(t)$. Collecting the signals received by the individual elements into a vector $\mathbf{x}(t)$, we obtain

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ \vdots \\ x_M(t) \end{bmatrix} = \begin{bmatrix} e^{j2\pi\Delta_1\,\sin(\alpha)} \\ \vdots \\ e^{j2\pi\Delta_M\,\sin(\alpha)} \end{bmatrix} a(\alpha)s(t)$$
$$=: \mathbf{a}(\alpha)s(t)$$

where the array response vector $\mathbf{a}(\alpha)$ is the response of the array to a planar wave with direction $\alpha$. The array manifold is the curve traced out by the vector $\mathbf{a}(\alpha)$ when $\alpha$ is varied

$$\mathcal{A} = \{\mathbf{a}(\alpha)\colon 0 \le \alpha < 2\pi\}.$$

If the curve does not intersect itself, then knowledge of the array manifold allows $\alpha$ to be determined from $\mathbf{x}$, i.e., direction finding. The common factor $a(\alpha)$ does not play a major role in this and is often omitted or lumped into the complex attenuation factor of the channel between the transmitter and receiver.

A uniform linear array (ULA) has elements equally spaced at $\Delta$. All delays between two consecutive array elements are the same, so that

$$\mathbf{x}(t) = \begin{bmatrix} 1 \\ \theta \\ \vdots \\ \theta^{M-1} \end{bmatrix} a(\alpha)s(t), \qquad \theta = e^{j2\pi\Delta\,\sin(\alpha)}. \tag{1}$$

Antenna responses are usually expressed in terms of $\theta$ rather than $\alpha$ since this is what is actually measured by the array. If $\Delta \le \frac{1}{2}$ wavelengths, there is a one-to-one relation between $\theta$ and $\alpha$. The specific structure of the array manifold of a ULA admits convenient estimation of $\theta$ and subsequently $\alpha$ from $\mathbf{x}(t)$ using algebraic techniques.

### *C. Parametric Multipath Propagation Model*

A commonly used parametric channel model for radio propagation is a multiray scattering model, also known as Jakes' model (after [1], see also [2], [3], [9], and [10]). In this model, the signal follows a number of distinct paths on its way from the source to the receiver, referred to as multipath rays. These arise from scattering, reflection, or diffraction of the radiated energy due to objects that lie in the environment. Apart from attenuation (fading), multipath propagation can also cause spreading of the signal in time, frequency and space, with significant effects on the received signal.

The scattering of the signal in the environment can be specialized into three stages: scattering local to the source at surrounding objects, reflections on distant objects of the few dominant rays that emerge out of the local clutter, and scattering local to the receiver (see Fig. 2). Let us ignore the latter for the moment and assume that there are $r$ rays bouncing off remote objects such as hills or tall buildings. The received parametric signal model is then usually written

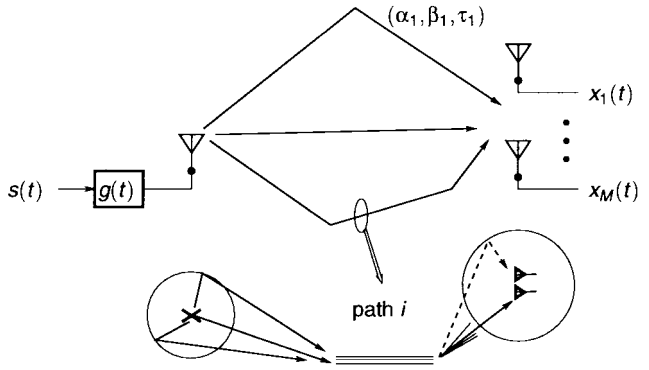



**Fig. 2.** Multipath propagation channel model.

as the convolution

$$\mathbf{x}(t) = \left[\sum_{1}^{r} \mathbf{a}(\alpha_i)\beta_i g(t-\tau_i)\right] * s(t) \qquad (2)$$

where $\mathbf{x}(t)$ is a vector consisting of the $M$ antenna outputs, $\mathbf{a}(\alpha)$ is the array response vector, and the impulse response $g(t)$ collects all temporal aspects, such as pulse shaping, and transmit and receive filtering. The model parameters of each ray are its (mean) angle-of-incidence $\alpha_i$, (mean) path delay $\tau_i$, and path loss $\beta_i$. The latter parameter lumps the overall attenuation, all phase shifts, and possibly the antenna response $a(\alpha)$ as well.

Each of the rays is itself composed of a large number of "mini-rays" due to scattering close to the source, all with roughly equal angles and delays, but arbitrary phases. This can be described by extending the model with additional parameters such as the standard deviations from the mean angle $\alpha_i$ and mean delay $\tau_i$, which depend on the radius (aspect ratio) of the scattering region and its distance to the remote scattering object [8], [29]. For macroscopic models, the standard deviations are generally small (less than a few degrees, and a fraction of $\tau_i$) and are usually, but not always, ignored.

The local scattering, however, has a major effect on the statistics and stationarity of $\beta_i$. For example, if all local rays have equal amplitude, then $\beta_i$ is the sum of a large number of arbitrary complex numbers, each with equal modulus but random phase, which gives $\beta_i$ a complex Gaussian distribution. Consequently, its amplitude has a Rayleigh distribution (hence the name Rayleigh fading). A second effect is that $\beta_i = \beta_i(t)$ is really (slowly) time varying: if the source is in motion, then the Doppler shifts and the varying location change the phase differences among the rays so that the sum can be totally different from one time instant to the next. The maximal Doppler shift $f_D$ is given by the speed of the source (in m/s) divided by the wavelength of the carrier. The *coherence time* of the channel is inversely proportional to $f_D$, roughly by a factor of 0.2; $\beta_i(t)$ can be considered approximately constant for time intervals smaller than this time [2], [3], [11]. Angles and delays are generally assumed to be stationary over much longer periods.

**Table 1** Typical Delay and Doppler Spreads in Cellular Applications at 900 MHZ

| Environment | delay spread | Doppler spread |
|---|---|---|
| Open rural | $< 0.2\mu$s | 100 Hz |
| Urban | 3 $\mu$s | 60 Hz |
| Hilly | 10–25 $\mu$s | 100 Hz |
| Indoor | $< 0.1\mu$s | 3 Hz |

Finally, scattering local to the receiver eventually results in the reception of a number of rays with roughly equal delays, but largely differing DOA's. The corresponding fading parameters have more or less equal amplitudes but different phases. This type of scattering is not present if the receiver is clear from local obstacles, e.g., on a mast, but may prevail otherwise.

### *D. Typical Channel Parameters*

Angle spread, delay spread, and Doppler spread are important characterizations of a radio channel, as it determines not only the amount of equalization that is required, but also the amount of diversity that can be obtained. In the context of mobile cellular telephony, typical channel delays and Doppler spreads that can occur at 900 MHz are provided in Table 1 [2], [3] (see also references in [9]).

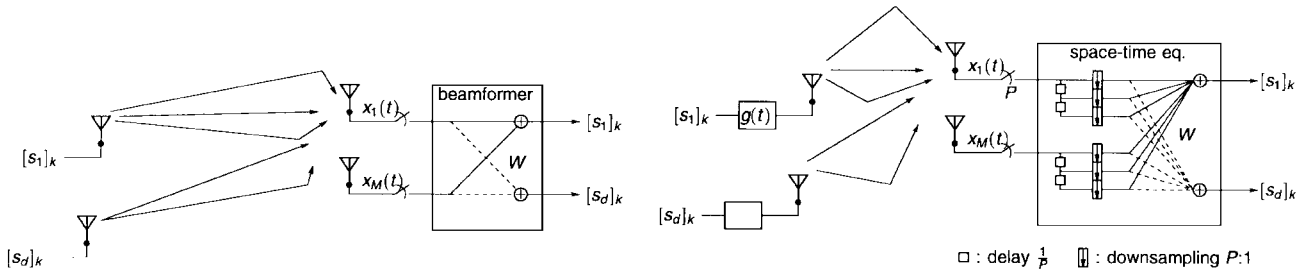
The delay spread determines the maximal symbol rate for which no equalization is required. The inverse of the delay spread is proportional to the *coherence bandwidth* [1], [2], [11]. Narrow-band signals with a bandwidth sufficiently smaller than the inverse of the delay spread experience a flat channel (in the frequency domain) that does not require equalization; $g(t)$ is essentially a scalar and can be lumped with $\beta_i$.

As noted before, the inverse of the Doppler frequency determines the coherence time, and thus the maximal temporal window in block processing algorithms, or the required speed of adaptation in adaptive algorithms.

The inverse of the angle spread (in radians) determines the *coherence distance* in wavelengths, which gives an indication of the minimal distance by which two antennas have to be spaced to enable separation of two disparate rays within this spread by (classical) spatial separation techniques. Rays without much angle spread have essentially the same $\mathbf{a}(\theta)$-vector.

Angle spreads are strongly dependent on the geometry of the environment and have not yet been studied as thoroughly as delay spreads. Current research suggests that most outdoor channels can be modeled adequately by a small number of dominant rays and that in open or suburban terrain most energy is often concentrated in a single ray in the direction of the mobile [12], with relatively small angle and delay spreads. Moreover, multiple rays usually have widely separated angles.

The first-generation American analog cellular AMPS system (FDMA) and the more recent digital IS-54 system (TDMA) have narrow-band signals at 25–30 kHz, with carrier frequencies in the 900 MHz band [3]. The symbol period for IS-54 is 41.6 $\mu$s. With delay spreads as in Table 1, it is seen that the symbol period is (much) larger in




(a) (b)

**Fig. 3.** (a) Spatial beamformer with an I-MIMO channel and (b) space-time linear equalizer with an FIR-MIMO channel.

all cases, so that the channel is usually instantaneous, not convolutive (except perhaps in hilly terrains, where some equalization may be needed). A data block in IS-54 spans 6.67 ms (162 bits). With a Doppler spread of 100 Hz, data is stationary over a fraction of 10 ms so that beamforming must be adaptive over the time slot.

The GSM system (TDMA) has signals with a bandwidth of 200 kHz in blocks (time slots) of 577 $\mu$s [3]. The data transmission rate is 270 kb/s, giving a symbol period of 3.7 $\mu$s. Thus, in hilly terrains the delay spread spans maximally five symbol periods, and equalization is necessary. The delay spread is less than one symbol period in most urban settings, and only minor equalization is required in this case. In other cases, the reception is more like an instantaneous mixture. The fading is stationary within the data block even for high Doppler shifts. Data blocks belonging to the same source are spaced at 5 ms, so fading is not stationary in going from one block to the next, although delays and angles might be the same.

In summary, knowledge of the delay spread and Doppler spread allows us to decide, *grosso modo*, if an instantaneous or a convolutive channel model is appropriate, and whether it is time invariant or time varying over the data block.

## III. Data Model for Signal Processing

In Section II, we have looked at a channel model based on physical properties of the radio channel. Though useful for generating simulated data, a detailed model is not always suitable for identification purposes, e.g., if the number of parameters is large, if the angle spreads within a cluster are large so that parameterization in terms of directions is not possible, or if there is a large and fuzzy delay spread. In these situations, it is more appropriate to work with an unstructured model, where the channel impulse responses are posed simply as arbitrary multichannel finite impulse response (FIR) filters. It is a generalization of the physical channel model considered earlier, in the sense that at a later stage we can still specify the structure of the coefficients.

### A. I-MIMO Model

Assume that $d$ source signals $s_1(t), \cdots, s_d(t)$ are transmitted from $d$ independent sources at different locations. If the delay spread is small, then what we receive at the antenna array will be a simple linear combination of these signals

$$\mathbf{x}(t) = \mathbf{a}_1 s_1(t) + \cdots + \mathbf{a}_d s_d(t)$$

where, as before, $\mathbf{x}(t)$ is a stack of the output of the $M$ antennas. We will usually write this in matrix form

$$\mathbf{x}(t) = A\mathbf{s}(t) \qquad A = [\mathbf{a}_1 \ \cdots \ \mathbf{a}_d] \qquad \mathbf{s}(t) = \begin{bmatrix} s_1(t) \\ \vdots \\ s_d(t) \end{bmatrix}.$$

Suppose we sample with a period $T$, normalized to $T = 1$, and collect a batch of $N$ samples into a matrix $X$, then

$$X = AS$$

where $X = [\mathbf{x}(0) \cdots \mathbf{x}(N-1)]$ and $S = [\mathbf{s}(0) \cdots \mathbf{s}(N-1)]$. The resulting $[X = AS]$ model is called an instantaneous multi-input multi-output model, or I-MIMO for short. It is a generic linear model for source separation, valid when the delay spread of the dominant rays is much smaller than the inverse bandwidth of the signals, e.g., for narrow-band signals, in line-of-sight situations or in scenarios where there is only local scattering. Even though this appears to limit its applicability, it is important to study it in its own right, since more complicated convolutive models can often be reduced by blind equalization techniques to $X = AS$.

The objective of beamforming for source separation is to construct a left-inverse $W$ of $A$, such that $WA = I$, hence $WX = S$ [see Fig. 3(a)]. This will recover the source signals from the observed mixture. It immediately follows that in this scenario it is necessary to have $d \leq M$ to ensure interference-free reception, i.e., not more sources than sensors. If we already know (part of) $S$, e.g., because of training, then $W = SX^{\dagger} = SX^{*}(XX^{*})^{-1}$, where $X^{\dagger}$ denotes the Moore–Penrose pseudoinverse of $X$ [105]–[107], here equal to its right inverse, and * denotes a complex conjugate transpose. Blind beamforming is to find $W$ with knowledge only of $X$.

If we adopt the multipath propagation model, then $A$ is endowed with a parametric structure: every column $\mathbf{a}_i$ is a sum of direction vectors $\mathbf{a}(\theta_{ij})$ with different fadings $\beta_{ij}$. If the $i$th source is received through $r_i$ rays, then

$$\mathbf{a}_i = \sum_{j=1}^{r_i} \mathbf{a}(\theta_{ij})\beta_{ij} = [\mathbf{a}(\theta_{i1}) \cdots \mathbf{a}(\theta_{i,r_i})] \begin{bmatrix} \beta_{i1} \\ \vdots \\ \beta_{i,r_i} \end{bmatrix}$$
$$(i = 1, \cdots, d).$$

If each source has only a single ray to the receiver array (a line-of-sight situation), then each $\mathbf{a}_i$ is a vector on the array manifold, and identification will be relatively straightforward. The more general case amounts to decomposing a given $\mathbf{a}$-vector into a sum of vectors on the manifold, which makes identification much harder.

To summarize the parametric structure in a compact way, we usually collect all $\mathbf{a}(\theta_{ij})$-vectors and path attenuation coefficients $\beta_{ij}$ of all rays of all sources in single matrices $A_\theta$ and $B$

$$A_\theta = [\mathbf{a}(\theta_{11}) \cdots \mathbf{a}(\theta_{d,r_d})] \qquad B = \text{diag}[\beta_{11} \cdots \beta_{d,r_d}].$$

To sum the rays belonging to each source into the single $\mathbf{a}_i$-vector of that source, we define a selection matrix

$$J = \begin{bmatrix} \mathbf{1}_{r_1} & & 0 \\ & \ddots & \\ 0 & & \mathbf{1}_{r_d} \end{bmatrix} : \quad r \times d \tag{3}$$

where $r = \sum_1^d r_i$ and $\mathbf{1}_m$ denotes an $m \times 1$ vector consisting of 1's. Together, this allows us to write the full (noise-free) I-MIMO data model as

$$X = AS, \qquad A = A_\theta B J. \tag{4}$$

### *B. FIR-MIMO Model*

Assume again that $d$ source signals $s_1(t), \cdots, s_d(t)$ are transmitted from $d$ independent sources, but moreover that they are now received through a convolutive channel. To limit ourselves to a practical and interesting case, let us assume that the signals are digital with a common pulse period, so that they can be described by a sequence of dirac pulses

$$s_j(t) = \sum_{k=-\infty}^{\infty} s_{j,k}\delta(t - kT).$$

For convenience, we normalize the symbol period to $T = 1$. The signal emitted by a source is a convolution of $s_j(t)$ by the pulse shape function $g(t)$, e.g., a raised cosine (generalized sinc function), which gives

$$u_j(t) = s_j(t) * g(t) = \sum_k g(t-k)\, s_{j,k}.$$

After propagation through the channel, the signal is received by an array of $M$ sensors, with outputs $x_1(t), \cdots, x_M(t)$. The impulse response of the channel from source $j$ to the $i$th sensor, $h_{ij}(t)$, is a convolution of the pulse shaping filter $g(t)$ and the actual channel response from $u_j(t)$ to $x_i(t)$. We can include any propagation delays and delays due to unsynchronized sources in $h_{ij}(t)$ as well. The data model is written compactly as the convolution

$$\mathbf{x}(t) = H(t) * \mathbf{s}(t)$$

where

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ \vdots \\ x_M(t) \end{bmatrix}, \qquad H(t) = \begin{bmatrix} h_{11}(t) & \cdots & h_{1d}(t) \\ \vdots & & \vdots \\ h_{M1}(t) & \cdots & h_{Md}(t) \end{bmatrix},$$

$$\mathbf{s}(t) = \begin{bmatrix} s_1(t) \\ \vdots \\ s_d(t) \end{bmatrix}.$$

At this point, we make the assumption that the $M$ channels $h_{ij}(t)$ associated to each source $j$ are FIR filters of (integer) length at most $L_j$, i.e., $h_{ij}(t) = 0$ for $t \notin [0, L_j)$. The maximal channel length among all sources is denoted by $L$. An immediate consequence of the FIR assumption is that, at any given moment, at most $L_j$ consecutive symbols of signal $j$ play a role in $\mathbf{x}(t)$.

Suppose that we sample each $x_i(t)$ at a rate of $P$ times the symbol rate, and collect samples during $N$ symbol periods. Then we can construct a data matrix $X$ containing all samples as

$$\begin{aligned} X &= [\mathbf{x}_0 \quad \cdots \quad \mathbf{x}_{N-1}] \\ &:= \begin{bmatrix} \mathbf{x}(0) & \mathbf{x}(1) & \cdots & \mathbf{x}(N-1) \\ \mathbf{x}\left(\frac{1}{P}\right) & \mathbf{x}\left(1+\frac{1}{P}\right) & & \cdot \\ \vdots & & & \vdots \\ \mathbf{x}\left(\frac{P-1}{P}\right) & \cdot & \cdots & \mathbf{x}\left(N-1+\frac{P-1}{P}\right) \end{bmatrix}. \end{aligned} \tag{5}$$

$X$ has size $MP \times N$; its $k$th column $\mathbf{x}_k$ contains the $MP$ spatial and temporal samples taken during the $k$th interval. Based on the FIR assumption, it follows that $X$ has a factorization

$$X = H\mathcal{S}_L \tag{6}$$

where

$$H = \begin{bmatrix} H(0) & H(1) & \cdots & H(L-1) \\ H\left(\frac{1}{P}\right) & \cdot & & \cdot \\ \vdots & & & \vdots \\ H\left(\frac{P-1}{P}\right) & \cdot & \cdots & H\left(L-\frac{1}{P}\right) \end{bmatrix} : MP \times dL$$

$$\mathcal{S}_L = \begin{bmatrix} \mathbf{s}_0 & \mathbf{s}_1 & \ddots & \mathbf{s}_{N-2} & \mathbf{s}_{N-1} \\ \mathbf{s}_{-1} & \mathbf{s}_0 & \ddots & \ddots & \mathbf{s}_{N-2} \\ \ddots & \ddots & \ddots & \ddots & \ddots \\ \mathbf{s}_{-L+1} & \ddots & \ddots & \ddots & \mathbf{s}_{N-L} \end{bmatrix} : dL \times N \tag{7}$$

and in this context $\mathbf{s}_i = \mathbf{s}(i)$, a $d$-dimensional vector. The matrix $H$ represents the unknown space-time channel, whereas $\mathcal{S} = \mathcal{S}_L$ contains the transmitted symbols. $\mathcal{S}$ has a block-Toeplitz structure: it is constant along the diagonals. This structure is a consequence of the time-invariance of the channel. Note that if the channels do not all have the same length $L$, then certain columns of $H$ are equal to zero.

A linear equalizer in this context can be written as a vector $\mathbf{w}$ which combines the rows of $X$ to generate an output $\mathbf{y} = \mathbf{w}^* X$. In the model so far, we can only equalize among the antenna outputs (simple beamforming) and

among the $P$ samples within one sample period (polyphase combining). More generally, we would want to filter over multiple sample periods, leading to a space-time equalizer. For a linear equalizer with a length of $m$ symbol periods, we have to augment $X$ with $m-1$ horizontally shifted copies of itself

$$\mathcal{X}_m = \begin{bmatrix} \mathbf{x}_0 & \mathbf{x}_1 & \iddots & \mathbf{x}_{N-m} \\ \mathbf{x}_1 & \mathbf{x}_2 & \iddots & \iddots \\ \iddots & \iddots & \iddots & \mathbf{x}_{N-2} \\ \mathbf{x}_{m-1} & \iddots & \mathbf{x}_{N-2} & \mathbf{x}_{N-1} \end{bmatrix} : mMP \times (N-m+1).$$

Each column of $\mathcal{X}_m$ is a regression vector—the memory of the filter. Using $\mathcal{X}_m$, a general space-time linear equalizer can be written as $\mathbf{y} = \mathbf{w}^* \mathcal{X}_m$, which combines $mP$ snapshots of $M$ antennas [see Fig. 3(b)]. The augmented data matrix $\mathcal{X}_m$ has a factorization

$$\begin{aligned} \mathcal{X}_m &= \mathcal{H}_m \mathcal{S}_{L+m-1} \\ &= \begin{bmatrix} \mathbf{0} & \boxed{H} & & \\ & \ddots & \ddots & \\ & \boxed{H} & & \\ \boxed{H} & & & \mathbf{0} \end{bmatrix} \\ &\quad \cdot \begin{bmatrix} \mathbf{s}_{m-1} & \ddots & \mathbf{s}_{N-2} & \mathbf{s}_{N-1} \\ \ddots & \ddots & \ddots & \mathbf{s}_{N-2} \\ \mathbf{s}_{-L+2} & \mathbf{s}_{-L+3} & \ddots & \ddots \\ \mathbf{s}_{-L+1} & \mathbf{s}_{-L+2} & \ddots & \mathbf{s}_{N-L-m+1} \end{bmatrix} \end{aligned} \tag{8}$$

where $\mathcal{H} = \mathcal{H}_m$ has size $mMP \times d(L+m-1)$ and the $m$ shifts of $H$ to the left are each over $d$ positions. $\mathcal{H}$ has a block-Hankel structure, i.e., it is constant along antidiagonals. $\mathcal{S}_{L+m-1}$ has the same structure as $\mathcal{S}_L$. A necessary condition for space-time equalization (the output $\mathbf{y}$ is equal to a row of $\mathcal{S}$) is that $\mathcal{H}$ is tall, which gives minimal conditions on $m$ in terms of $M, P, d, L$ [57]. Unlike spatial beamforming, it will not be necessary to find $\mathcal{H}^\dagger$; it suffices to reconstruct a single block row of $\mathcal{S}$, which can be done with $d$ space-time equalizers $\mathbf{w}_i$. Nonlinear equalizer structures are possible, e.g., by using feedback, but they are not discussed here.

### C. Connection to the Parametric Multipath Model

For a single source, recall the multipath propagation model (2), valid for specular multipath with small cluster angle spread

$$\mathbf{h}(t) = \sum_{i=1}^{r} \mathbf{a}(\theta_i) \beta_i g(t-\tau_i) \tag{9}$$

where $g(t)$ is the pulse shape function by which the signals are modulated, $\mathbf{a}(\theta)$ is the array response vector function, and $\beta_i$ is the complex path attenuation.

Suppose as before that $\mathbf{h}(t)$ has finite duration and is zero outside an interval $[0, L)$. Consequently, $g(t-\tau_i)$ has the same support for all $\tau_i$. At this point, we can define a parametric "time manifold" vector function $\mathbf{g}(\tau)$, collecting $LP$ samples of $g(t-\tau)$

$$\mathbf{g}(\tau) = \begin{bmatrix} g(0-\tau) \\ g\left(\frac{1}{P}-\tau\right) \\ \vdots \\ g\left(L-\frac{1}{P}-\tau\right) \end{bmatrix}, \qquad 0 \le \tau \le \max \tau_i.$$

If we also construct a vector $\mathbf{h}$ with samples of $\mathbf{h}(t)$

$$\mathbf{h} = \begin{bmatrix} \mathbf{h}(0) \\ \mathbf{h}\left(\frac{1}{P}\right) \\ \vdots \\ \mathbf{h}\left(L-\frac{1}{P}\right) \end{bmatrix}$$

then it is straightforward to verify that (9) gives

$$\mathbf{h} = \sum_{i=1}^{r} (\mathbf{g}_i \otimes \mathbf{a}_i)\beta_i = [\mathbf{g}_1 \otimes \mathbf{a}_1, \cdots, \mathbf{g}_r \otimes \mathbf{a}_r] \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_r \end{bmatrix}$$
$$\mathbf{g}_i = \mathbf{g}(\tau_i), \quad \mathbf{a}_i = \mathbf{a}(\theta_i)$$

where $\otimes$ denotes a Kronecker product, defined for vectors $\mathbf{a}$ and $\mathbf{b}$ as

$$\mathbf{a} \otimes \mathbf{b} = \begin{bmatrix} a_1 \mathbf{b} \\ \vdots \\ a_m \mathbf{b} \end{bmatrix}.$$

Thus, the multiray channel vector is a weighted sum of vectors on the space-time manifold $\mathbf{g}(\tau) \otimes \mathbf{a}(\theta)$. Because of the Kronecker product, this is a vector in an $LPM$-dimensional space, with more distinctive characteristics than the $M$-dimensional $\mathbf{a}(\theta)$-vector in a scenario without delay spread. The connection of $\mathbf{h}$ with $H$ as in (7) is that $\mathbf{h} = \text{vec}(H)$, i.e., $\mathbf{h}$ is a stacking of all columns of $H$ in a single vector.

We can define, much as before, parametric matrix functions

$$A_\theta = [\mathbf{a}(\theta_1) \cdots \mathbf{a}(\theta_r)], \quad G_\tau = [\mathbf{g}(\tau_1) \cdots \mathbf{g}(\tau_r)]$$
$$B = \text{diag}[\beta_1 \cdots \beta_r]$$

and let $(G_\tau \diamond A_\theta)$ denote a columnwise Kronecker product (Khatri–Rao product). This gives $\mathbf{h} = (G_\tau \diamond A_\theta) B \mathbf{1}_r$. Extending now to $d$ sources, we see that the $MP \times dL$-sized matrix $H$ in (7) can be rearranged into an $MPL \times d$ matrix

$$H' = [\mathbf{h}_1 \cdots \mathbf{h}_d] = (G_\tau \diamond A_\theta) B J$$

where $J$ is the selection matrix defined in (3) that sums the rays into channel vectors. $(G_\tau \diamond A_\theta)$ now plays the same role as $A_\theta$ in the previous section. Each of its columns is a vector on the space-time manifold.

**Table 2** Signal-Channel Structural Properties

| Properties | $A$ or $\mathcal{H}$ | $S$ or $\mathcal{S}$ |
|---|---|---|
| *I/FIR-MIMO* | | |
| matrix | $\mathcal{H}$ block Hankel | $\mathcal{S}$ block Toeplitz |
| | | known training segment |
| subspace | $\text{col}(A) = \text{col}(X)$ | $\text{row}(S) = \text{row}(X)$ |
| | $\text{col}(\mathcal{H}) = \text{col}(\mathcal{X})$ | $\text{row}(\mathcal{S}) = \text{row}(\mathcal{X})$ |
| modulation | | FA, CM, independence, $\cdots$ |
| temporal | | CDMA codes, temp. nonwhite |
| spectral | | cyclostationarity, res. carriers |
| *parametric* | | |
| spatial | known $\mathbf{a}(\theta)$ | |
| temporal | known $\mathbf{g}(\tau)$ | |
| spectral | residual carriers | |

## IV. Principles of Blind Beamforming

A summary of the noise-free data models developed so far is

$$\begin{aligned} &\text{I-MIMO:} \quad X = AS, \quad A = A_\theta BJ \\ &\text{FIR-MIMO:} \quad \mathcal{X} = \mathcal{H}\mathcal{S}, \quad \mathcal{H} \leftrightarrow H' = (G_\tau \diamond A_\theta)BJ. \end{aligned} \tag{10}$$

The first part of these model equations is generally valid for linear time invariant channels, whereas the second part is a consequence of the adopted multiray model.

Based on this model, the received data matrix $X$ or $\mathcal{X}$ has several structural properties. In several combinations, these are often strong enough to allow to find the factors $A$ (or $H$) and $S$ (or $\mathcal{S}$) from knowledge of $X$ or $\mathcal{X}$ alone. Very often, this will be in the form of a collection of beamformers (or space-time equalizers) $\{\mathbf{w}_i\}_1^d$ such that each beamformed output $\mathbf{w}_i^* X = \mathbf{s}_i$ is equal to one of the source signals, so that it must have the properties of that signal. Properties are listed in Table 2 and discussed below.

### A. Matrix Structure

*1) Toeplitz Structure:* The fixed baud rate of communication signals, along with time invariance, result in the fact that $\mathcal{X}$ has a factorization in which $\mathcal{H}$ is block Hankel and $\mathcal{S}$ is block Toeplitz. This is a strong property and allows, for example, the blind equalization of unknown channels carrying unknown digital signals with equal baud rates. It cannot be used for source separation, but it is very useful for reducing the FIR-MIMO problem $\mathcal{X} = \mathcal{H}\mathcal{S}$ to the instantaneous $X = AS$ problem [56], [57].

Several techniques are available nowadays: the original methods, which are phrased in a stochastic context and use the asymptotic diagonality of the source covariance matrix [51], closely related linear prediction (LP) methods [53], [54], and "deterministic" subspace-based methods working directly on $\mathcal{X}$ and exploiting either the Hankel structure of $\mathcal{H}$ [52], [53], [55] or the Toeplitz structure of $\mathcal{S}$ [56], [57]. Closely related to these are the cross-relation method [58] and the mutually referenced equalizer method [59]. It is possible to incorporate partial knowledge of the channel into some of the methods, in particular the fact that the pulse shape function is usually known [60]–[62]. This puts an additional linear constraint on the channel impulse response vector $\mathbf{h}$ and may lead to important improvements in accuracy.

The subspace-based methods exploit the linear nature of the underlying problem and work well if the channel length is known and well defined but might fail otherwise. The LP methods are robust against channel-length overestimation but rely on longer data sequences and a sufficiently large first channel coefficient. (The latter problem is overcome by a "multistep approach" [63].) See [50] and the paper by Tong and Perreau in this issue for a more complete overview.

*2) Training Sequences:* If training symbols are present in the signal, then a number of columns of $S$ or $\mathcal{S}$ are known. This number should be such that this known submatrix $S_t$ is a wide matrix, in which case it generally has a right inverse $S_t^\dagger$. This directly allows estimation of $A$ or $H$ as $X_t S_t^\dagger$, where $X_t$ is the corresponding window of the data matrix. With $A$ or $H$ known, there are a large number of suitable space-time equalizers (e.g., zero-forcing, minimum mean-square error, decision-feedback), differing in performance, complexity, and symbols/noise assumptions. Techniques are standard, and the literature is abundant.

A topic of increasing interest is that of *semiblind* techniques, where it is assumed that some training symbols are available, but perhaps not sufficient for channel estimation. Also, it is felt that use of additional structures such as the Toeplitz structure can significantly improve the channel estimates obtained from the use of training symbols only [64]. Only a few algorithms are known at this point, e.g., [65]–[67].

*3) Low Rank Factors:* An important property used by many algorithms is that $\mathcal{X} = \mathcal{H}\mathcal{S}$ is a low rank factorization: if $m$, $M$, $P$ are large enough, then $\mathcal{H}$ is a tall matrix and $\mathcal{S}$ is a wide matrix. This has several implications, most notably

$$\begin{aligned} \mathcal{H} \text{ full column rank} \quad &\Rightarrow \quad \text{row}(\mathcal{X}) = \text{row}(\mathcal{S}) \\ \mathcal{S} \text{ full row rank} \quad &\Rightarrow \quad \text{col}(\mathcal{X}) = \text{col}(\mathcal{H}) \end{aligned}$$

where $\text{row}(\cdot)$ and $\text{col}(\cdot)$ stand for the row span and column span of the matrix argument, respectively. Almost any blind separation/equalization method is (implicitly) based on this low-rank property: knowing $\mathcal{X}$ and assuming full rank factors, we have a basis for the row span of $\mathcal{S}$ and the column span of $\mathcal{H}$, and it remains to find the (hopefully unique) matrix in this row or column span that has the required structural properties, such as the Toeplitz structure, or any of the structures to follow. Also, as mentioned earlier, the low rank property is necessary in general even if $\mathcal{H}$ is known, since any space-time equalizer is a row of a left inverse of $\mathcal{H}$. For this, it is required that $\mathcal{H}$ is tall.

### B. Signal Modulation Structure

The signal modulation structure includes the instantaneous amplitude and phase of the modulated signal, and also the symbol constellation. Some typical modulation structures are listed below.

*1) Constant Modulus:* In many wireless applications, the transmitted waveform has a constant modulus (CM). This occurs, e.g., in FM modulation, or in phase modulation, as in GSM. So-called CMA's can separate arbitrary linear superpositions of such signals by finding out which linear combinations of the antenna outputs $\mathbf{w}_i^* X$ give back signals that have the CM property. Solutions are generally unique up to an arbitrary phase offset. The CM property is extremely robust and can be used for blind equalization, as well as source separation [88]–[93]. Most algorithms are based on iterative cost-function minimization, and a lot of recent research effort has been on proving global convergence and on initialization issues (see the paper by Johnson *et al.* in this issue for an overview and references). An algebraic technique is given in Section VII-A.

*2) Finite Alphabet:* Another important structure in digital communication signals is their finite alphabet (FA). The modulated signal is a linear or nonlinear map of an underlying finite alphabet, e.g., $\{+1, -1\}$ for signals with a binary phase-shift keyed (BPSK) constellation. As with the CM property, it is possible to separate arbitrary linear combinations of FA signals in a more or less unique way, given a minimal amount of samples [96]–[101]. For small constellation sizes [BPSK or quadrature PSK (QPSK)], this works very well. For high constellation sizes, only iterative algorithms are known and their performance is quite dependent on an initialization close to the solution. Often, the CMA's can be used to provide an initial point, even if the constellation does not exactly have a CM property.

*3) Distributional Properties and Independence:* More generally, if the source distribution is known and not Gaussian, separation is possible by restoring the distribution functions at the output of the beamformer, e.g., by using ML techniques. Even if the distributions are not known, we can restore distributional properties expressing the independence of sources. This is a vast area of research with many directions (cf. [75] and the paper by Cardoso in this issue). Algebraic methods are possible by using higher order stochastic moments and functions thereof, such as cumulants (e.g., see [76]–[84]). Source independence is generally applicable and very useful for audio and seismic applications, such as the separation of several speakers using multiple microphones. Because it is a stochastic property, the number of samples that are required is typically an order of magnitude larger than in the case where we can use deterministic CM or FA properties to pose conditions on every individual sample.

### C. Temporal and Spectral Structure

The temporal structure relates to $s(t)$ as well, but now with regard to its temporal properties. These can include knowledge of its pulse shape function and, in the case of CDMA signals, knowledge of the source codes, but also certain statistical properties for sources that are temporally nonwhite.

*1) CDMA Codes:* In direct-sequence CDMA, the emitted "chip symbols" $s_k$ are in fact modulations of low-rate source symbols $d_n$ by known code vectors $\mathbf{c}$ of length $L_c$

$$[s_k \cdots s_{k+L_c-1}] = d_n \mathbf{c} = d_n [c_1 \cdots c_{L_c}], \qquad n = \lfloor k/L_c \rfloor.$$

(The code vectors are different for each source.) Because the only unknowns are the $d_n$, this reduces the number of unknowns in $S$ by a factor $L_c$. The source symbols can be recovered, e.g., by row span template matching techniques [43]–[45], which are essentially straightforward least squares (LS) algorithms.

*2) Temporally Nonwhite and Independence:* If the sources are independent and temporally nonwhite, separation is possible by using the fact that the cross-covariance and cross-cumulants of the signals at the output of the beamformer should be zero for all time lags. For example

$$E(s_i(t) s_j^*(t-\tau)) \quad \begin{cases} =0, & i \neq j \\ \neq 0, & i = j,\ \tau \in \langle \text{range} \rangle. \end{cases}$$

This allows the separation of sources, but in this form it cannot be used to equalize them. Often, the second-order conditions are sufficient to find the beamformer; examples of algebraic techniques for this are in [85]–[87]. Some details are provided in Section VII-D.

*3) Cyclostationarity:* Many signals exhibit cyclostationary properties, i.e., their cyclic autocorrelation function $R_x^\alpha(\tau) = E(x(t)x(t-\tau)^* e^{-j2\pi\alpha t})$ is wide-sense stationary and has spectral lines at selective lags $\tau$ and frequencies $\alpha$ [68]. This reflects that the signal is correlated with frequency-shifted versions of itself and is typically caused by periodicities such as the symbol rate in bauded communication signals, or residual carrier frequencies after demodulation. If two sources have spectral peaks for different $(\alpha, \tau)$, then they can be separated based on this [68]–[70]. It is usually required that these parameters are known, although they can be estimated in specific cases. Recent research focuses on the explicit introduction of cyclostationarity at the transmitter, to facilitate separation at the receiver. An elementary scheme for this is simply to repeat the block or part of it [71], [72], or to deliberately introduce small carrier offsets by additional modulations with a periodic sequence [73].[1] Channel identification based on cyclostationarity properties is possible as well (e.g., see [74]). As with high-order statistics methods, these methods may in general require a considerable amount of data to yield reliable results, as convergence may be slow.

For digital communication signals, a straightforward way in which the cyclostationarity property can be expressed is by oversampling the antenna outputs, at Nyquist rate rather than the symbol rate. The multiple samples obtained during one symbol period presumably give independent linear combinations of the same transmitted bits, just as antennas give independent linear combinations from sampling in space. This fact was noted first in [51] and has generated a lot of interest (e.g., see [52]–[58]). It is the underlying reason why we could factor $X$ in (6) as $X = H\mathcal{S}_L$, where

[1] An example algorithm that separates binary sources based on small differences in carrier frequencies is given in Section VII-C.

$\mathcal{S}_L$ becomes a Toeplitz structure, and this structure then induces the more general $\mathcal{X} = \mathcal{H}\mathcal{S}$ in (8), where $\mathcal{H}$ has a similar block-Hankel structure. Although this was initially called a second-order technique, the Toeplitz structure is a deterministic rather than a stochastic property, i.e., valid for any data size and independent of source correlation properties.

### D. Parametric Structures

Parametric structures are induced by the parametric multipath model (and extensions of it) that we have derived in Section II. We use the fact that the columns of $A$ or $\mathcal{H}$ do not take just any value, but have the specific forms $A = A_\theta BJ$ and $\mathcal{H} \leftrightarrow H' = (G_\tau \diamond A_\theta)BJ$, where the parametric structure of $A_\theta$ and $G_\tau$ is known and the parameters can be estimated. It makes sense to use such models if the number of parameters is much smaller than, e.g., the number of coefficients in an unstructured FIR model.

*1) The Spatial Manifold:* In the I-MIMO model in (10), each column of $A$ is a linear combination of array response vectors $\{\mathbf{a}(\theta_i)\}$, each of which is on the array manifold. If the array manifold is known, e.g., by calibration or from structural considerations, then we can try to fit the column span of $X$ (hence $A$) to the appropriate linear combinations. This will work if the number of rays is not large and if the calibration data is reliable. For this purpose, a large number of direction finding techniques have been proposed (see the recent overview in [18]). Among the high-resolution algorithms, the MUSIC algorithm [19] is still very popular, although it is now encompassed by the more general WSF and MODE techniques [20]–[23], which provide asymptotically ML-optimal performance. These are iterative optimization algorithms that need a starting point of sufficient accuracy. Attractive closed-form algebraic techniques are possible if the geometry of the array has a shift-invariance structure, as exhibited for example by a uniform linear array (ULA), and this has led to the well-known ESPRIT algorithm and variants thereof [24]–[26].[2] The ESPRIT algorithm is discussed in Section VI-A. It is readily extended to two-dimensional direction finding of both azimuth and elevation [31]–[35]. Most DOA models assume point sources. However, the array manifold model can be generalized to include the effects of small angle spreads [29], [30].

*2) The Temporal Manifold:* Similarly, in (10), each column of $H'$ is a linear combination of vectors of the form $\{\mathbf{g}(\tau_i) \otimes \mathbf{a}(\theta_i)\}$, where $\mathbf{g}(\tau)$ is the temporal manifold function, the sampled response to an incoming pulse $g(t - \tau)$. If the specular multipath model holds true and the number of rays is not large, all received signals are constructed from several delays of $g(t)$, hence they can be viewed as superpositions of a number of vectors $\mathbf{g}(\tau)$. The temporal manifold is usually known to a good accuracy since it depends only on the pulse-shaping function and the receiver filters, both of which are under tight control. If the spatial manifold is unknown or deemed unreliable, or if the angular spread is complicated and diffuse, we can still fit to the temporal manifold and leave the spatial domain unconstrained.

Otherwise, with knowledge of both the spatial and temporal manifold, we can attempt to do a joint estimation of all angles and delays by fitting to the space-time manifold [36]–[40] (see Section VI-C).

*3) Residual Carrier Frequencies:* Independent narrowband sources modulated at high frequencies rarely have exactly the same carrier frequency. Consequently, after demodulation, the cochannel sources have unequal residual carrier frequencies, with only partially overlapping spectra. If the spectral properties of the sources are known or if we sample sufficiently fast so that we can use stationarity properties of the sources, the residual carrier frequencies can be estimated and the sources can be separated, even if the array manifold is unknown. This can be regarded as a special case of cyclostationarity. An example is given in Section VI-D.

## V. Preprocessing

In the previous section, we have listed a number of properties that are available for blind source separation and equalization. The corresponding algorithms can be broadly classified into row span and column span methods. A row span method is a method that still works even if we premultiply $X$ or $\mathcal{X}$ with an arbitrary full rank matrix on the left; this changes the mixing matrix but leaves the row span invariant. Similarly, column span methods are invariant to multiplication at the right. Algorithms that use only properties of $A$ and $H$ are column span methods: all information is contained in a basis of the column span of $X$. This reflects the fact that no constraints are placed on $S$. Methods based on properties of the signals are usually row span methods. In special cases, it is possible to translate row span information into column span information by stacking the data into block Hankel matrices. This occurs for example for the residual carrier property.

In this second part of the paper, Sections VI and VII give detailed examples of algebraic column span and row span methods to illustrate a few of the deterministic properties listed before. All algorithms can work with a basis of either the row span or the column span of the data matrix. The construction of this basis is a common and elementary preprocessing step, and is the topic of this section. In algebraic methods, it is often the main computational bottleneck as well.

### A. Subspace Estimation, SVD

Consider again the noise-free data model $X = AS$, where $A$ has size $M \times d$, $S$ has size $d \times N$, and $d \leq M \leq N$. $X$ has $M$ rows but the rank of $X$ is generically equal to $d$: each antenna output (row of $X$) is a linear combination of only $d$ source signals (rows of $S$). If we know $X$, then with linear algebra techniques we can find a basis

[2] For a ULA, the MODE algorithm can be made closed-form as well [22].

for this row span, i.e., a matrix $\hat{V}$ with $d$ rows such that $\text{row}(\hat{V}) = \text{row}(X) = \text{row}(S)$. At the same time, $X$ has only $d$ independent columns, and not $M$, and we can find a basis for it, i.e., a matrix $\hat{U}$ with $d$ columns such that $\text{col}(\hat{U}) = \text{col}(X) = \text{col}(A)$.

The numerically preferred way to obtain these bases is to compute a singular value decomposition (SVD) [105]–[107], which is a decomposition of $X$ as

$$X = \hat{U}\hat{\Sigma}\hat{V}$$

where $\hat{U}$ and $\hat{V}$ contain the orthogonal bases $\hat{U}^*\hat{U} = I_d$, $\hat{V}\hat{V}^* = I_d$, and $\hat{\Sigma}$ is a $d \times d$ diagonal matrix with positive real numbers—the nonzero singular values. These are usually sorted in nondecreasing order. The columns of $\hat{U}$ and rows of $\hat{V}$ are called the singular vectors. There is a well-known connection to eigenvalue problems: since $XX^* = \hat{U}\hat{\Sigma}^2\hat{U}^*$, it is seen that $\hat{U}$ contains the same eigenvectors as the empirical data covariance matrix, and $\hat{\Sigma}^2$ are the corresponding eigenvalues. The singular values give important information on the conditioning of the problem: signals with low power or two signals with similar $\mathbf{a}$-vectors (e.g., close directions-of-arrival) give rise to small singular values. $\hat{V}$ can be interpreted as a whitened data matrix, since $\hat{V}\hat{V}^* = I$. It can be written as $\hat{V} = (\hat{\Sigma}^{-1}\hat{U}^*)X$: its columns are obtained from those of $X$ by a filtering operation. This whitening operation is sometimes called a Mahalanobis transformation.

With noise present, the data model becomes

$$X = AS + E$$

where $E$ is the additive noise term. $X$ is no longer rank deficient but has full rank $M$. It is here that the SVD becomes useful: the SVD of $X$ can be written as

$$X = U\Sigma V = \begin{bmatrix}\hat{U} & \hat{U}_n\end{bmatrix}\begin{bmatrix}\hat{\Sigma} & 0\\ 0 & \hat{\Sigma}_n\end{bmatrix}\begin{bmatrix}\hat{V}\\ \hat{V}_n\end{bmatrix} = \hat{U}\hat{\Sigma}\hat{V} + \hat{U}_n\hat{\Sigma}_n\hat{V}_n \tag{11}$$

where $U$ and $\Sigma$ are square $M \times M$, $U^*U = UU^* = I_M$, $VV^* = I_M$, and $\Sigma$ is diagonal and partitioned into $d$ "large" and $M-d$ "small" singular values. The same decomposition holds in the noise-free case, but then with $\hat{\Sigma}_n = 0$. Under mild conditions, one can show that the new basis $\hat{U}$ is a good approximation to the noise-free basis $\hat{U}$ (and asymptotically equal to it), provided that the noise singular values, the entries of $\hat{\Sigma}_n$, are substantially smaller than the signal singular values, the entries of $\hat{\Sigma}$. Alternatively, we have to assume a sufficiently large number of samples and spatially/temporally white noise so that the noise covariance is a multiple of the identity matrix. The signal singular values depend on the signal + noise power, the number of samples, and the separation between the sources [cf. Fig. 4(a) and (b)]. The noise singular values depend on the noise power and the number of samples $N$ [cf. Fig. 4(a) and (c)]. The new $\hat{\Sigma}$ is equal to the old $\hat{\Sigma}$, but augmented with some noise power. The row space spanned by the new $\hat{V}$ can be viewed as an LS estimate of the subspace spanned by the noise-free $\hat{V}$. Thus, a rank-$d$ approximation of $X$ is $X \approx \hat{U}\hat{\Sigma}\hat{V}$, which is known as taking the truncated SVD.

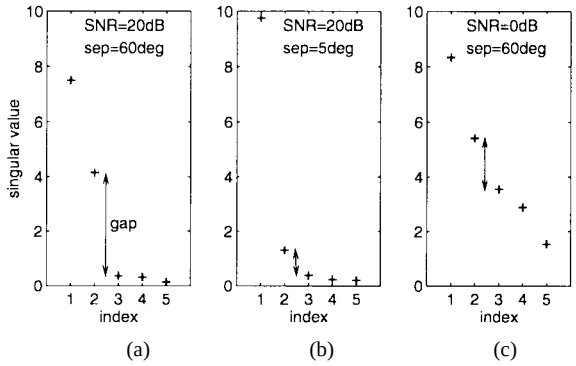



**Fig. 4.** Singular values for $d = 2$ sources, $M = 5$ antennas, $N = 10$ samples. (a) Well-separated case: large gap between signal and noise singular values. (b) Signals from close directions results in a small signal singular value. (c) Increased noise level increases noise singular values.

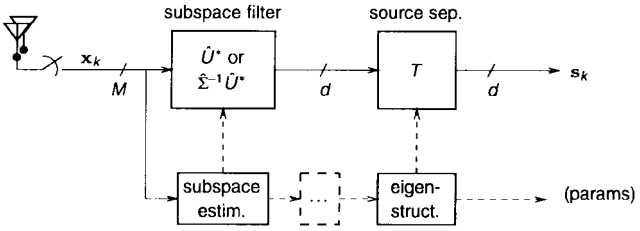



**Fig. 5.** Conceptual beamformer structure.

We will use the SVD of $X$ and subsequent truncation to rank $d$ as a first step in almost all our processing. This is useful for several reasons: 1) if the rank of $X$ without noise is much smaller than $M$, then prefiltering by $\hat{U}^*$ or $\hat{\Sigma}^{-1}\hat{U}^*$ will remove an equal ratio of noise; 2) parameter estimation is much easier from a minimal basis than from a full matrix; and 3) after truncation, a stabilized inverse of $X$ as needed in certain (MMSE-type) receivers is $X^\dagger = \hat{V}^*\hat{\Sigma}^{-1}\hat{U}^*$. An untruncated full rank inverse can lead to severe noise enhancement due to the inversion of small singular values.

Since we hardly ever use the diagonal property of $\hat{\Sigma}$, except perhaps to estimate the rank of $X$, simpler subspace estimation methods have been proposed to estimate a basis $\hat{U}$ of the principal column span. These schemes are also suitable for adaptive algorithms that update the estimate as more data columns $\mathbf{x}_k$ are observed, and either start from knowledge of the noise power, providing a level at which to truncate the rank [109]–[111], from knowledge of the rank of $X$, e.g., if the number of sources is known [112], [113], or converge to the SVD under stationary conditions [114]. Automatic detection of the rank without knowledge of the noise power or the number of sources/rays is a considerable problem which deserves additional research.

### *B. Beamformer Structure*

Let us assume that $A$ has full column rank (independent directions) and $S$ has full row rank (independent signals). Introduce the truncated SVD

$$X = \hat{U}\hat{\Sigma}\hat{V}. \tag{12}$$

Then $A$ and $\hat{U}$ span the same subspace, so that there is a $d \times d$ invertible matrix $T$ such that

$$\hat{U} = AT. \tag{13}$$

Substitution gives $S = T\hat{U}^* X$, so that

$$W = T\hat{U}^* \tag{14}$$

is a beamformer which will recover $S$ from $X$. Hence, the main problem in blind beamforming is to construct the matrix $T$ based on properties of $A$ or $S$ (or both). Note that it is sufficient to construct *any* (orthogonal) basis of the column span of $X$; we do not need the singular vectors, only the subspace they span.

Since $X = AS = \mathbf{a}_1\mathbf{s}_1 + \cdots + \mathbf{a}_d\mathbf{s}_d$ (where the $\mathbf{a}_i$ are the columns of $A$ and the $\mathbf{s}_i$ are the rows of $S$), it is clear that we cannot expect to recover the ordering of signals. Usually, we also have to permit the exchange of a phase factor between $\mathbf{a}_i$ and $\mathbf{s}_i$, or even any scalar factor if the power of the signals or the norm of vectors $\mathbf{a}_i$ is not specified.

## VI. Examples of Column Span Methods

The next two sections will elaborate on the properties listed in Section IV by demonstrating examples of how these properties can be turned into algebraic algorithms to find $T$ in (13). We first look at column span methods, which work on properties of $A$ or $\mathcal{H}$. Section VII will then go into row span methods that exploit properties of $S$. For ease of description, we will always pretend a noiseless case where $X$ is rank deficient. In the presence of noise, the first computational step is an SVD or subspace estimation, followed by a rank truncation which reduces $X = U\Sigma V$ to the quasi-noiseless case $X \approx \hat{U}\hat{\Sigma}\hat{V}$, in the notation of (11). The subsequent steps of the algorithms will remain unchanged. Of course, a correct treatment of the noise is very important—this makes the difference between a good and a bad algorithm. But looking at the noiseless case is sufficient to understand the functioning of most deterministic algorithms.

### A. No Multipath

We start with a simple scenario, in which there is no multipath and sources have only one ray toward the receiving antenna array. Since no delays are involved, all measurements are simply instantaneous linear combinations of the source signals, i.e., $X = AS$. Each source has only one ray, so that the data model is refined to $X = A_\theta BS$, where $A_\theta = [\mathbf{a}(\theta_1), \cdots, \mathbf{a}(\theta_d)]$ are the array response vectors, $B = \text{diag}[\beta_1 \cdots \beta_d]$ are the fading parameters, and the rows of $S$ contain the signals.

Computationally attractive ways to compute $\{\theta_i\}$ and hence $A_\theta$ are possible for certain regular antenna array configurations for which $\mathbf{a}(\theta)$ becomes a shift-invariant or similar recursive structure. One well-studied example of such a structure is that obtained from a ULA. For such an array, with interelement spacing $\Delta$ wavelengths, we have seen that

$$\mathbf{a}(\theta) = \begin{bmatrix} 1 \\ \theta \\ \vdots \\ \theta^{M-1} \end{bmatrix}, \qquad \theta = e^{j2\pi\Delta \sin(\alpha)} \tag{15}$$

where $\alpha$ is the direction-of-arrival.

The ESPRIT algorithm [24] is a well-known and elegant technique to find the factorization $X = A_\theta BS$ by using shift-invariance properties of $A_\theta$

$$A_\theta = [\mathbf{a}(\theta_1) \cdots \mathbf{a}(\theta_d)] = \begin{bmatrix} 1 & \cdots & 1 \\ \theta_1 & \cdots & \theta_d \\ \vdots & & \vdots \\ \theta_1^{M-1} & \cdots & \theta_d^{M-1} \end{bmatrix}. \tag{16}$$

Let us define

$$\Theta = \begin{bmatrix} \theta_1 & & 0 \\ & \ddots & \\ 0 & & \theta_d \end{bmatrix}$$

as a diagonal matrix of parameters, and selection matrices

$$J_x = [I_{M-1} \;\; \mathbf{0}_1], \qquad J_y = [\mathbf{0}_1 \;\; I_{M-1}]$$

which will select the first and last $M-1$ rows of $A_\theta$, respectively. The Vandermonde structure of $A_\theta$ ensures that

$$(J_y A_\theta) = (J_x A_\theta)\Theta$$

which is a direct expression of the shift-invariance of the array. To use this property for estimating $A_\theta$ and $S$ from the data $X$, we first compute an SVD

$$X = \hat{U}\hat{\Sigma}\hat{V}$$

where $\hat{U}$ has $d$ columns which together span the column space of $X$. Since the same space is spanned by the columns of $A_\theta$, there must exist a $d \times d$ invertible matrix $T$ such that

$$\hat{U} = A_\theta BT, \qquad S = \left(T\hat{U}^*\right)X.$$

Let us define

$$\hat{U}_x = J_x\hat{U}, \qquad \hat{U}_y = J_y\hat{U}.$$

Then the shift-invariance of $A_\theta$ implies that

$$\begin{cases} \hat{U}_x = A'_\theta BT \\ \hat{U}_y = A'_\theta \Theta BT \end{cases}$$

where $A'_\theta = J_x A_\theta$ consists of the top $M-1$ rows of $A_\theta$. Since $\Theta$ and $B$ are diagonal matrices and commute, we have $\hat{U}_y = A'_\theta B \cdot TT^{-1} \cdot \Theta T = \hat{U}_x T^{-1}\Theta T$. For $M \geq d$, $\hat{U}_x$ is "tall" and has a left-inverse $\hat{U}_x^\dagger$, so that

$$\hat{U}_x^\dagger \hat{U}_y = T^{-1}\Theta T.$$

Since $\Theta$ is a diagonal matrix, this is an eigenvalue equation: $T^{-1}$ contains the eigenvectors of $\hat{U}_x^\dagger\hat{U}_y$ (scaled arbitrarily to unit norm), and the entries of $\Theta$ on the diagonal are the eigenvalues. The blind beamformer is given by $W = T\hat{U}^*$. Thus, source separation in this case is essentially

an eigenvalue problem. (This turns out to be the case for many algebraic algorithms.) If the antennas are spaced by, at most, half a wavelength, then the DOA's are directly recovered from $\Theta$, otherwise they are ambiguous. Because the rows of $T$ are determined only up to a scaling, the fading parameters $B$ cannot be recovered unless we know the average power of each signal. This is of course inherent in the problem definition.

There are many important refinements and extensions to this algorithm. We can use the fact that all $\theta_i$ are on the unit circle along with the centro-symmetric structure of the array to augment the data matrix to $X_e = [X, \Pi X^{*T}]$, where $\Pi$ is the reverse-identity matrix which flips the rows of $X^{*T}$; this will not increase the rank but double the number of observations [25]. Using this structure, it is also possible to transform $X_e$ to a real-valued matrix by simple linear operations on its rows and columns [25], [34]. As mentioned in Section IV-D, there are many other direction finding algorithms that are applicable, in particular MODE [22]. Although ESPRIT is statistically suboptimal, its performance is usually quite adequate. Its interest to us here is its straightforward generalization to more complicated estimation problems in which shift-invariance structure is present.

### *B. Coherent Multipath*

In the above, we assumed that there was no multipath; each source had only one path to the antenna array. However, the $X = AS$ model is also valid if sources have multiple rays toward the array, as long as the delay differences are small compared to the signal bandwidth so that they can be represented by phase shifts. This is known as coherent multipath. Let $d$ be the number of sources, $r_i$ the number of rays belonging to source $i$, and $r = \sum_1^d r_i$ the total number of rays (assumed to be distinct). In that case, a more detailed model is

$$X = (A_\theta B J) S \tag{17}$$

where $A_\theta$: $M \times r$ is the Vandermonde matrix associated with the DOA's of the rays, as in (16), and $J$: $r \times d$ is a selection matrix which adds groups of rays to source signals, for example

$$J = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}$$

in case of two sources, each with two rays. $B$ is a diagonal scaling matrix representing the different amplitudes (fadings) of each ray, including phase offsets. Because the rank of $X$ is still $d$, the SVD of $X$ can retrieve only a $d$-dimensional subspace $\hat{U}$, so that

$$\hat{U} = (A_\theta B J) T, \qquad S = T \hat{U}^* X.$$

It is clear that blind beamforming is more challenging now; we try to find $T$ such that each column of $\hat{U}$ is represented by a sum of $r$ Vandermonde vectors, rather than only $d$ vectors, and $r$ is not known.

To solve this problem algebraically using ESPRIT-type techniques,[3] we first try to restore the rank to $r$. This is possible if the number of antennas $M$ is sufficiently large, in fact $M \geq r + \max(r_i)$. In that case, we can form a block-Hankel matrix out of $\hat{U}$ by taking vertical shifts of it

$$\mathcal{U}_m := \left[\hat{U}^{(1)} \;\; \hat{U}^{(2)} \;\; \cdots \;\; \hat{U}^{(m)}\right] : \;\; (M - m + 1) \times md. \tag{18}$$

Here, $\hat{U}^{(i)}$ is a submatrix of $\hat{U}$ consisting of its $i$th till $M-m+i$th row, and $m$ is known as the spatial smoothing factor [27], [28]. With the above model, we have that $\mathcal{U}_m$ satisfies the factorization

$$\mathcal{U}_m = A'_\theta B \left[JT \;\; \Theta JT \;\; \cdots \;\; \Theta^{m-1} JT\right] =: A'_\theta B \mathcal{T} \tag{19}$$

where $A'_\theta$ consists of the top $M - m + 1$ rows of $A_\theta$. If $M - m + 1 \geq r$ and $m \geq \max(r_i)$, the factors in the above factorization can be shown to have full rank $r$, so that $\mathcal{U}_m$ has rank $r$.

At this point, the structure of $\mathcal{U}_m$ in (19) shows that we have reduced the problem to an $[X = AS]$-type problem without multipath, which can be solved using the ESPRIT algorithm in Section VI-A. Thus we compute an SVD of $\mathcal{U}_m$

$$\mathcal{U}_m = \hat{U}_u \hat{\Sigma}_u \hat{V}_u$$

where $\hat{U}_u$ contains the dominant $r$ singular vectors of $\mathcal{U}_m$. From (19) it follows that there is an invertible $r \times r$ matrix $R$ such that

$$\hat{U}_u = A'_\theta B R, \quad \mathcal{T} = \left(R \hat{U}_u^*\right) \mathcal{U}_m.$$

We continue in the same way as before to compute $R$: with

$$\hat{U}_x = J_x \hat{U}_u, \quad \hat{U}_y = J_y \hat{U}_u$$

the data model satisfies the eigenvalue equation

$$\hat{U}_x^\dagger \hat{U}_y = R^{-1} \Theta R \tag{20}$$

which gives both $\Theta$ and $R$, up to scaling of its rows. At this point, we have recovered $\mathcal{T} = (R \hat{U}_u^*) \mathcal{U}_m$, up to multiplication at the left by an arbitrary diagonal matrix. The next objective is to estimate $T$ from the structure of $\mathcal{T}$ in (19). This is now a much simpler task: we have available $m$ matrices of size $r \times d$, after correction by suitable powers of $\Theta^{-1}$ all equal to $JT$. The structure of $J$ ensures that this matrix has only $d$ distinct rows, which are the $d$ rows of $T$. Hence, it suffices to estimate these $d$ unique rows, which is a simple clustering problem if the rows of $T$ are sufficiently different. This determines both $T$ and $J$, i.e., the assignment of rays to sources. With $T$ in hand, we have our blind beamformer as before: $W = T \hat{U}^*$.

[3] Other techniques, such as MODE, are directly applicable to the coherent case without modifications.

### C. Incoherent Multipath with Small Delay Spread

An extension of the previous would be to consider a true multipath scenario, where each source is received via a superposition of rays, each with its own angle $\theta_i$, delay $\tau_i$, and fading $\beta_i$. The question then becomes how to estimate these parameters, and how to construct a space-time beamformer to recover the sources. The problem is known as joint angle-delay estimation [36]–[40]. In general, this is a challenging task to perform blindly in column space, without making further assumptions on the sources.

Let us here consider a scenario which allows a simple extension of the previous and which has applications in blind CDMA beamformers. Consider $d$ sources as before. Assume that these are digital sources, i.e., discrete-time sources with a common pulse shape function $g(t)$ and a common pulse period $T$, normalized to $T = 1$. We make the following important restrictions leading to a simplified version of the $\mathcal{X} = \mathcal{HS}$ model:

1) $g(t)$ is zero outside an interval $[0, L_g)$, with $L_g < 1$;
2) the delay spread is so small that $L_g + \max(\tau_{ij}) < 1$.

The implication is that every sample of the received signal is a combination of $d$ source symbols, and not more than $d$. These assumptions are approximately valid in a CDMA receiver, after synchronization and matched filtering with the desired user code [46]. (In this case, $d = 1$ since in principle only one signal matches the code, but the interference is strong.)

The received signal at the antennas can be written as $\mathbf{x}(t) = \sum_{i=1}^{d} \sum_{j=1}^{r_i} \mathbf{a}(\theta_{ij})\beta_{ij} g(t-\tau_{ij}) s_{i,k}$, where $k = \lfloor t \rfloor$. We sample at a rate $P$ during $N$ symbol periods and collect all data samples in a matrix $X$ of size $MP \times N$

$$X = \begin{bmatrix} \mathbf{x}(0) & \mathbf{x}(1) & \cdots & \mathbf{x}(N-1) \\ \mathbf{x}\left(\frac{1}{P}\right) & \mathbf{x}\left(1+\frac{1}{P}\right) & \cdots & \\ \vdots & \vdots & & \vdots \\ \mathbf{x}\left(1-\frac{1}{P}\right) & \mathbf{x}\left(2-\frac{1}{P}\right) & \cdots & \mathbf{x}\left(N-\frac{1}{P}\right) \end{bmatrix}. \quad (21)$$

Define matrices $A_\theta = [\mathbf{a}(\theta_1), \cdots, \mathbf{a}(\theta_r)]$ and $G_\tau = [\mathbf{g}(\tau_1), \cdots, \mathbf{g}(\tau_r)]$, where $\mathbf{g}(\tau)$ is the parametric "time manifold" vector function

$$\mathbf{g}(\tau) = \begin{bmatrix} g(0-\tau) \\ g\left(\frac{1}{P}-\tau\right) \\ \vdots \\ g\left(1-\frac{1}{P}-\tau\right) \end{bmatrix}.$$

With $B$ the diagonal matrix containing the fading parameters, and $J$ the $r \times d$ selection matrix which assigns each ray to one of the sources, we find that $X$ satisfies the model

$$X = (G_\tau \diamond A_\theta) BJS$$

where $\diamond$ denotes the column-wise Kronecker product. $(G_\tau \diamond A_\theta)$ now plays the same role as $A_\theta$ in the previous section. Each of its columns is on the space-time manifold $\mathbf{g}(\tau) \otimes \mathbf{a}(\theta)$. Because of the Kronecker product, this is a vector in a high-dimensional space, which improves resolution and allows to identify more rays than sensors.

To identify the rays and derive a beamformer using similar techniques as before, we need $(G_\tau \diamond A_\theta)$ to satisfy shift-invariance properties. With a uniform linear array, $A_\theta$ already has such a property, and if the number of antennas is larger than the number of rays we can proceed as before. Otherwise, we can do a transformation such that $G_\tau$ takes a Vandermonde structure. To this end, we use a well-known property of the Fourier transformation: delays are transformed into certain phase progressions. In particular, collect the samples of the known waveform $g(t)$ into a vector $\mathbf{g}_0 = \mathbf{g}(0)$, and let $\tilde{\mathbf{g}}_0 = \mathcal{F}\mathbf{g}_0$ where $\mathcal{F}$ denotes the DFT matrix of size $P \times P$

$$\mathcal{F} := \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & \phi & \cdots & \phi^{P-1} \\ \vdots & \vdots & & \vdots \\ 1 & \phi^{P-1} & \cdots & \phi^{(P-1)^2} \end{bmatrix}, \qquad \phi = e^{-j(2\pi/P)}.$$

If $\tau$ is an integer multiple of $1/P$, then it is apparent that the Fourier transform $\tilde{\mathbf{g}}_\tau$ of $\mathbf{g}(\tau)$ is given by

$$\tilde{\mathbf{g}}_\tau = \tilde{\mathbf{g}}_0 \odot \begin{bmatrix} 1 \\ \phi^\tau \\ \vdots \\ (\phi^\tau)^{P-1} \end{bmatrix} = \mathrm{diag}(\tilde{\mathbf{g}}_0) \begin{bmatrix} 1 \\ \phi^\tau \\ \vdots \\ (\phi^\tau)^{P-1} \end{bmatrix}$$

($\odot$ is a Schur–Hadamard product: an entry-wise multiplication of two vectors or matrices). The same is to a very good approximation true if $g(t)$ is bandlimited and sampled at or above the Nyquist rate. Thus, $\mathcal{F}G_\tau = \mathrm{diag}(\tilde{\mathbf{g}}_0)F_\phi$, where

$$F_\phi = \begin{bmatrix} 1 & \cdots & 1 \\ \phi_1 & \cdots & \phi_r \\ \vdots & & \vdots \\ \phi_1^{P-1} & \cdots & \phi_r^{P-1} \end{bmatrix}, \qquad \phi_i := \phi^{\tau_i} = e^{-j(2\pi/P)\tau_i}. \quad (22)$$

It follows that if we take the Fourier transform of each oversampled antenna output over a single symbol period, we can write the resulting data model $\tilde{X} := (\mathcal{F} \otimes I_M)X$ as

$$\tilde{X} = (\mathrm{diag}(\tilde{\mathbf{g}}_0)F_\phi \diamond A_\theta) BJS. \quad (23)$$

Since $\tilde{\mathbf{g}}_0$ is known beforehand, we can divide it out of (23), which amounts to a deconvolution. Obviously, this can be done only on intervals where $\tilde{\mathbf{g}}_0$ is nonzero. The details of this are in [40] and omitted here. The result is that we can obtain a matrix $\overline{X}$ which satisfies the model

$$\overline{X} = (F_\phi \diamond A_\theta) BJS \quad (24)$$

where, because of the selection of nonzero frequency intervals, the number of rows of $F_\phi$ is typically somewhat smaller than in (22).

At this point, we have obtained a model with much the same structure as in (17), but with $A_\theta$ replaced by $F_\phi \diamond A_\theta$ where both $F_\phi$ and $A_\theta$ have a Vandermonde structure. The construction of the beamformer can now follow the same strategy as well. First note that the rank of $\overline{X}$ is only $d$,

since this is the number of rows of $S$. Thus we compute the SVD of $\overline{X}$, i.e., $\overline{X} =: \hat{U}\hat{\Sigma}\hat{V}$ where $\hat{U}$ has $d$ columns. As before, we have

$$\hat{U} = (F_\phi \diamond A_\theta)BJT, \quad S = \left(T\hat{U}^*\right)\overline{X}$$

and our objective is to compute the $d \times d$ matrix $T$. Next, construct a matrix $\mathcal{U}_m$ by composing shifted copies of $\hat{U}$

$$\mathcal{U}_m := \left[\hat{U}^{(1)} \cdots \hat{U}^{(m)}\right]$$

where now each shift is over blocks of $M$ rows rather than one. This matrix has model

$$\begin{aligned}\mathcal{U}_m &= (F_\phi \diamond A_\theta)B[JT \;\; \Phi JT \;\; \cdots \;\; \Phi^{m-1}JT] \\ &=: (F_\phi \diamond A_\theta)B\mathcal{T} \end{aligned} \tag{25}$$

where

$$\Phi = \begin{bmatrix}\phi_1 & & 0 \\ & \ddots & \\ 0 & & \phi_r\end{bmatrix}.$$

If $m$ is large enough and all rays belonging to the same source have distinct delays, then the rank of $\mathcal{U}_m$ is $r$. [Otherwise, we must also take shifts in the spatial domain to restore the rank, i.e., "spatial smoothing," as in (18).]

Let $\mathcal{U}_m = \hat{U}_u\hat{\Sigma}_u\hat{V}_u$ be the SVD of $\mathcal{U}_m$, and suppose that $\mathcal{U}_m$ has $nM$ rows. It has the model $\hat{U}_u = (F_\phi \diamond A_\theta)BR$, $\mathcal{T} = (R\hat{U}_u^*)\mathcal{U}_m$. To estimate $R$, and hence $\mathcal{T}$, we can now form *two* types of selection matrices: a pair to select submatrices of $F_\phi$, and a pair to select from $A_\theta$

$$\begin{aligned} J_{x\phi} &:= [I_{n-1} \;\; \mathbf{0}_1] \otimes I_M, \quad J_{x\theta} := I_n \otimes [I_{M-1} \;\; \mathbf{0}_1], \\ J_{y\phi} &:= [\mathbf{0}_1 \;\; I_{n-1}] \otimes I_M, \quad J_{y\theta} := I_n \otimes [\mathbf{0}_1 \;\; I_{M-1}]. \end{aligned}$$

To estimate $\Phi$, we take submatrices consisting of the first and respectively last $M(n-1)$ rows of $\hat{U}_u$

$$\hat{U}_{x\phi} = J_{x\phi}\hat{U}_u, \qquad \hat{U}_{y\phi} = J_{y\phi}\hat{U}_u$$

whereas to estimate $\Theta$ we stack, for all $n$ blocks, its first and respectively last $M-1$ rows

$$\hat{U}_{x\theta} = J_{x\theta}\hat{U}_u, \quad \hat{U}_{y\theta} = J_{y\theta}\hat{U}_u.$$

These data matrices have the structure

$$\begin{cases} \hat{U}_{x\phi} = A'BR \\ \hat{U}_{y\phi} = A'\Phi BR \end{cases} \quad \begin{cases} \hat{U}_{x\theta} = A''BR \\ \hat{U}_{y\theta} = A''\Theta BR. \end{cases} \tag{26}$$

If dimensions are such that these are low-rank factorizations, then

$$\begin{aligned} \hat{U}_{x\phi}^\dagger \hat{U}_{y\phi} &= R^{-1}\Phi R \\ \hat{U}_{x\theta}^\dagger \hat{U}_{y\theta} &= R^{-1}\Theta R. \end{aligned} \tag{27}$$

Compare this equation to (20). Instead of a single eigenvalue equation, we now have two; the same matrix $R$ can diagonalize both data matrices. As before, once we have obtained $R$, we can immediately reconstruct $\mathcal{T}$ as $\mathcal{T} = (R\hat{U}_u^*)\mathcal{U}_m$, which provides a beamformer to extract each individual ray. After that, we need to assign the rays to source signals (i.e., identify $J$ and $T$ from $\mathcal{T}$ as in Section VI-B) and combine them in any viable way to end up with a beamformer that receives the individual source signals, at the symbol rate. If we like, we can retrieve the delays and angles of each ray from the eigenvalue matrices $\Phi$ and $\Theta$, respectively. The correct pairing of angles to delays follows simply from the fact that they share the same eigenvectors.

Joint diagonalization problems such as the above are overdetermined; one matrix already gives $R$, provided that the eigenvalues are distinct. For example, we could work only with the first matrix (since we already assumed once that the delays are distinct), and in this case we do not have to make any assumptions on the structure of the antenna array, i.e., we do not use its shift-invariance. We can also form any linear combination of the two matrices and try to ensure that the combination has distinct eigenvalues (such an approach was taken in [34]). Several Jacobi-type algorithms have been proposed as well, although some of these assume that $R$ is a unitary matrix [31], [32], [35], [77], [81], [83], [87], [93], [115]–[120].

Although these algorithms usually yield good performance, the problem of joint diagonalization with nonhermitian matrices has not yet been optimally solved. It is very relevant to study such overdetermined eigenvalue problems. Indeed, a third matrix arises if we use a two-dimensional uniform antenna array, by which we can measure both azimuth and elevation, or any other array with multiple independent baselines. We will see several other examples of joint eigenvalue problems later in this paper.

### *D. Space-Frequency Beamforming; Residual Carriers*

A somewhat different scenario than what we considered before, which, however, leads to the same type of data models (and thus the same beamforming algorithms), is the following. Suppose that we observe a frequency band of interest and want to separate all sources that are present. Assume that the sources are narrowband, typically with different carrier frequencies, but that the spectra might be partly overlapping. The objective is to construct a beamformer to separate the sources based on differences in angles or carrier frequencies. This is a problem of joint angle-frequency estimation [48], [49]. We will assume that the sample rates in this application are much higher than the data rates of each source and that there is only coherent multipath, although generalizations are possible.

Suppose that the narrow-band signals have a bandwidth of less than $1/T$, so that they can be sampled with a period $T$ to satisfy the Nyquist rate. We normalize to $T = 1$. Also assume that the bandwidth of the band to be scanned is $P$ times larger; after demodulation to IF we have to sample at rate $P$. Without multipath, the data model of the modulated sources at the receiver is

$$\mathbf{x}(t) = \sum_1^d \mathbf{a}(\theta_i)\beta_i e^{j(2\pi/P)f_i t}s_i(t)$$

where $f_i$ is the residual modulation frequency of the $i$th source ($-(P/2) \le f_i < P/2$). In matrix form this is

written as

$$\mathbf{x}(t) = A_\theta B \Phi^t \mathbf{s}(t) \tag{28}$$

where

$$\Phi = \begin{bmatrix} \phi_1 & & 0 \\ & \ddots & \\ 0 & & \phi_d \end{bmatrix}, \qquad \phi_i = e^{j(2\pi/P)f_i}.$$

Since $P$ can be quite large (order 100, say), it would be very expensive to construct a full data matrix of all samples. In fact, it is sufficient to subsample: collect $m$ subsequent samples at rate $P$, then wait till the next period before sampling again, resulting in a data matrix $X$ of size $mM \times N$

$$X = \begin{bmatrix} \mathbf{x}(0) & \mathbf{x}(1) & \cdots & \mathbf{x}(N-1) \\ \mathbf{x}\left(\frac{1}{P}\right) & \mathbf{x}\left(1+\frac{1}{P}\right) & \cdots & \mathbf{x}\left(N-1+\frac{1}{P}\right) \\ \vdots & \vdots & & \vdots \\ \mathbf{x}\left(\frac{m-1}{P}\right) & \mathbf{x}\left(1+\frac{m-1}{P}\right) & \cdots & \mathbf{x}\left(N-1+\frac{m-1}{P}\right) \end{bmatrix}. \tag{28a}$$

With the model of $\mathbf{x}(t)$ in (28), we find that $X$ has a factorization

$$X = \begin{bmatrix} A_\theta B\mathbf{s}(0) & A_\theta B\Phi^P \mathbf{s}(1) & \cdots \\ A_\theta B\Phi \mathbf{s}\left(\frac{1}{P}\right) & A_\theta B\Phi^{P+1}\mathbf{s}\left(1+\frac{1}{P}\right) & \cdots \\ \vdots & \vdots & \\ A_\theta B\Phi^{m-1}\mathbf{s}\left(\frac{m-1}{P}\right) & A_\theta B\Phi^{P+m-1}\mathbf{s}\left(1+\frac{m-1}{P}\right) & \cdots \end{bmatrix}. \tag{28b}$$

Let us assume at this point that $P \gg m$. In that case, $s(t)$ is relatively bandlimited with respect to the observed band, which allows to make the crucial assumption that

$$\mathbf{s}(t) \approx \mathbf{s}\left(t+\frac{1}{P}\right) \approx \cdots \approx \mathbf{s}\left(t+\frac{m-1}{P}\right)$$

so that the model of $X$ simplifies to

$$\begin{aligned} X &\approx \begin{bmatrix} A_\theta \\ A_\theta\Phi \\ \vdots \\ A_\theta\Phi^{m-1} \end{bmatrix} B[\mathbf{s}_0 \;\; \Phi^P\mathbf{s}_1 \;\; \cdots \;\; \Phi^{(N-1)P}\mathbf{s}_{N-1}] \\ &= (F_\phi \diamond A_\theta) B (F_P \odot S). \end{aligned}$$

$F_\phi$ is as in (22), only it has a different interpretation: $\phi$ is now related to the carrier frequency. $F_P$ is similar to $F_\phi$ except for a transpose and different powers, and the pointwise multiplication represents the modulation on the signals. Obviously, beamforming will not remove this modulation, but after estimating $\Phi$ we can easily correct for it.

If we do consider coherent multipath, the data model becomes

$$X = (F_\phi \diamond A_\theta) B J (F_P \odot S). \tag{29}$$

The column span of this model has precisely the same structure as $\overline{X}$ in (24) before, and hence we can use the same algorithm to find the beamformer.

If sources are assumed not to have equal carrier frequencies and $m > d$, we can separate them based on the structure of $F_\phi$ only. In this case we do not need the array structure and an arbitrary array can be used, but we do not recover the DOA's. If frequencies can be close, however, we will have to separate the signals based on differences in angles as well. It is then also necessary to restore the rank of $X$ to $r$ by spatial smoothing.

## VII. Examples of Row Span Methods

Column span methods require rather sophisticated assumptions on the channel, and their accuracy largely depends on the validity of these assumptions. In contrast, row span methods only pose $X = AS$ or $X = H\mathcal{S}$ and put all conditions on $S$. For communication signals with significant structures, this leads to powerful and robust blind beamforming algorithms. We will be mainly concerned with I-MIMO scenarios here, although extensions to general FIR-MIMO models have been derived; e.g., in [56] and [57] the Toeplitz structure of $\mathcal{S}$ is exploited to reduce $\mathcal{X} = \mathcal{H}\mathcal{S}$ to $X = AS$. In fact, both problems are the same if we do not use the Toeplitz structure of $\mathcal{S}$.

As always, the first step of row span methods is to reduce dimensions to that of row$(X)$. Via an SVD, an orthogonal basis for this is obtained as $\hat{V}$.

### A. Constant Modulus

For a signal (row vector) $\mathbf{s} = [s_1 \cdots s_N]$, the CM property can be written as

$$|s_k| = 1, \qquad (k = 1, \cdots, N). \tag{30}$$

The property holds for phase or frequency modulated signals, or any single-level digital constellation. Our objective is, for a given $X$, to find a factorization $X = AS$ where all rows of $S$ have this CM property. Let us assume that we have computed an SVD $X = \hat{U}\hat{\Sigma}\hat{V}$ and have done the subspace filtering by $\hat{\Sigma}^{-1}\hat{U}^*$, so that at this point we have a $d \times N$ matrix $\hat{V} = (\hat{\Sigma}^{-1}\hat{U}^*)X$. It remains to identify the $d \times d$ matrix $T$ such that $T\hat{V} = S$ is a CM matrix.

Let $\hat{V} = [\mathbf{v}_1 \cdots \mathbf{v}_N]$ and $T = [\mathbf{t}_1 \cdots \mathbf{t}_d]^*$. We are looking for all beamforming vectors $\mathbf{t}$ such that $\mathbf{t}^*\hat{V} = \mathbf{s}$ is a CM signal. One can prove that, generically and for $N > 2d$, solutions are unique so that any CM signal that is recovered this way is bound to be one of the original source signals, up to a phase factor [93]. Substituting $\mathbf{t}^*\mathbf{v}_k = s_k$ in (30), i.e., $s_k s_k^* = 1$, shows that $\mathbf{t}$ satisfies the property

$$\mathbf{t}^*\mathbf{v}_k\mathbf{v}_k^*\mathbf{t} = 1, \qquad (k = 1, \cdots, N). \tag{31}$$

The CM problem is to find all independent vectors $\mathbf{t}$ that satisfy this equation. An alternative way to write this equation is by using the Kronecker product. By expanding (31) into a sum of terms and rearranging, it follows that $\mathbf{t}^*\mathbf{v}_k\mathbf{v}_k^*\mathbf{t} = [\overline{\mathbf{v}}_k \otimes \mathbf{v}_k]^T(\mathbf{t} \otimes \overline{\mathbf{t}}) = 1$, so that

$$P\mathbf{y} = \mathbf{1}, \qquad \mathbf{y} = \mathbf{t} \otimes \overline{\mathbf{t}} \tag{32}$$

where $\bar{\mathbf{t}}$ is the complex conjugate of $\mathbf{t}$, $\mathbf{1}$ is a vector of 1's, and $P$ is an $N \times d^2$ matrix whose rows are given by $[\bar{\mathbf{v}}_k \otimes \mathbf{v}_k]^T$. Hence, we have turned the overdetermined system of quadratic equations (31) into a linear system of equations, subject to a quadratic constraint. Any solution of the linear system in (32) can be written as

$$\mathbf{y} = \alpha_1 \mathbf{y}_1 + \alpha_2 \mathbf{y}_2 + \cdots + \alpha_\delta \mathbf{y}_\delta, \qquad (\alpha_1 = 1) \tag{33}$$

where $\mathbf{y}_1$ is a particular solution of the system ($P\mathbf{y}_1 = \mathbf{1}$), the other $\mathbf{y}_k$ are a basis of the null space of $P$, and $\delta - 1$ is the dimension of this space. An important result is that, generically, $\delta = d$ once $N > d^2$ [93]. The remaining problem is to find out which linear combinations of the $\{\mathbf{y}_k\}$ lead to a vector $\mathbf{y}$ that can be written as $\mathbf{y} = \mathbf{t} \otimes \bar{\mathbf{t}}$.

A convenient way to rephrase the latter problem is to work with a matrix $Y = \mathbf{t}\mathbf{t}^*$. For any matrix, we can form a vector by simply stacking its columns. Similarly, we can "unvec" vectors into matrices. A notable property is that $\mathbf{t}\mathbf{t}^* \leftrightarrow \mathbf{t} \otimes \bar{\mathbf{t}}$. Thus, applying the unvec operation to every $\mathbf{y}_k$ in (33) leads to the equivalent

$$\mathbf{t}\mathbf{t}^* = \alpha_1 Y_1 + \alpha_2 Y_2 + \cdots + \alpha_\delta Y_\delta$$

where each $Y_k$ has size $d \times d$. Hence, the problem is to form linear combinations of known matrices $Y_k$ such that the result is rank-1 hermitian, so that it has a factorization as $\mathbf{t}\mathbf{t}^*$. In the present case, there are $\delta = d$ matrices, and we are looking for all $d$ solutions $\mathbf{t}_k$, $k = 1, \cdots, d$ to the problem; we are in fact trying to rewrite the given arbitrary basis $\{Y_k\}$ as a rank-one basis $\{\mathbf{t}_k \mathbf{t}_k^*\}$. The $\mathbf{t}_k^*$ are the rows of the beamforming matrix $T$.

Conversely, this means we can write each matrix $Y_k$ as some linear combination of the rank-one basis

$$Y_k = \lambda_{k1} \mathbf{t}_1 \mathbf{t}_1^* + \lambda_{k2} \mathbf{t}_2 \mathbf{t}_2^* + \cdots + \lambda_{kd} \mathbf{t}_d \mathbf{t}_d^*, \qquad k = 1, \cdots, d.$$

After collecting the coefficients into diagonal matrices, $\Lambda_k = \text{diag}[\lambda_{k1}, \cdots, \lambda_{kd}]$, it follows that the $Y_k$ satisfy the equations

$$\begin{aligned} Y_1 &= T^* \Lambda_1 T \\ Y_2 &= T^* \Lambda_2 T \\ &\vdots \\ Y_d &= T^* \Lambda_d T \end{aligned} \tag{34}$$

where all $\Lambda_k$'s are diagonal. This is a very fundamental set of equations. The collection $\{Y_k\}$ are similar by congruence to diagonal matrices, and can be jointly diagonalized. In fact, this is again a generalized eigenvalue problem. This is perhaps more clearly seen by looking at ratios, e.g., $Y_2^{-1} Y_1$ (assuming invertibility) has a factorization as $T^{-1}(\Lambda_2^{-1}\Lambda_1)T$, and $T$ can in principle be found as the eigenvectors of $Y_2^{-1} Y_1$. Of course, given the full set of equations and the presence of noise, we would rather try to find a single $T$ that optimally satisfies all equations rather than only one, which leads to the problem of simultaneous diagonalization. Similar algorithms as mentioned for the joint angle-delay estimation problem are available, e.g., based on Jacobi iterations [77], [81], [93], [115]–[118]. Since we have a good starting point from the eigenvalue problem of a pair of matrices, such iterations generally converge extremely fast, i.e., in two or three iterations.

The CM problem is well studied as an adaptive blind equalization technique. The difference with source separation is that, in equalization, we are interested in only one signal since the others are shifted copies (echos) of it. Applied to the problem of source separation, the adaptive techniques have major problems in making sure that *all* independent sources are recovered, and special attention has to be paid to this [92]. The algorithm described above was called algebraic constant modulus algorithm (ACMA) [93] and is very robust in this respect. Its complexity, on the other hand, is reasonable only for a small number of sources (say $d \leq 6$), which limits applications for equalization.

*Experiment:* We can illustrate the performance of ACMA by an experiment with measured data.[4] The sources are six FM-modulated analog speech/music signals, occupying the same bandwidth of 25 kHz in the 900 MHz band. Since the signals are narrowband, this is an I-MIMO scenario where no equalization is necessary, even though multipath may be present. The uncalibrated antenna array consists of $M = 6$ omnidirectional antennas, arranged nonuniformly roughly on a line, with a maximal baseline of 2.5 m. The signal-to-noise (SNR) ratio is around 17 dB per antenna per source, and the sources have roughly equal powers. Their DOA's are, respectively, $-1.5$, 7, 0, 42, 100, and 42° (nearfield). With all sources present, the condition number of $A$ is around 20, which is medium conditioned.[5]

Fig. 6 shows the worst signal-to-interference ratio (SIR) among the signals after beamforming, as a function of the number of samples that have been used, and for a varying number of sources. The reference "true" $A$-matrix has been estimated from 500 samples. It is seen that only a small number of samples are required (order $2d^2$ or so) to give a good separation of all $d$ sources, even though some of the sources are only spaced by 1.5°.

### B. Binary Symbol Constellation

Another property based on which digital sources can be separated is their finite alphabet. Frequently, a BPSK constellation is used, i.e., the transmitted signals are vectors $\mathbf{s}$ with all entries $s_k \in \{+1, -1\}$. The signals are of course modulated by some pulse shape function, but this can be absorbed in the $A$-matrix, or perhaps the $H$-matrix, which will result in the same problem since we do not use its structure. Hence we arrive at the following problem: given $X$, determine the factorization $X = AS$, where $S_{ij} \in \{+1, -1\}$. This can be viewed as a specialization of the CM problem if we restrict the signals to be real as well.

[4]The experimental data was kindly provided by F. McCarthy, ARGOSystems, Sunnyvale, CA, June 1994, and is described in [93].

[5]The condition number of a matrix is the ratio of the largest to the smallest singular value and can be interpreted as the maximal noise amplification of a zero-forcing beamformer $W = A^\dagger$, here a factor 20, or 26 dB. With only two sources present, the amplification is $\text{cond}(A) = 1.2$, or only 1.5 dB.

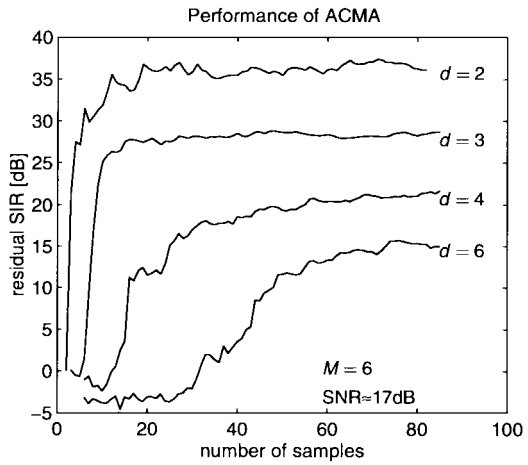



**Fig. 6.** Residual SIR after blind beamforming of a mixture of $d$ signals using the constant-modulus property.

Since $S$ is real-valued, it is advantageous to write

$$X = AS \Leftrightarrow \begin{bmatrix} \text{real}(X) \\ \text{imag}(X) \end{bmatrix} = \begin{bmatrix} \text{real}(A) \\ \text{imag}(A) \end{bmatrix} S \Leftrightarrow X_R = A_R S \tag{35}$$

with obvious definitions of $X_R \in \mathbb{R}^{2M \times N}$ and $A_R \in \mathbb{R}^{2M \times d}$. If we work with $X_R$ and take care to stay in the real domain, then this forces our estimate of $S$ to be real. At the same time, $A_R$ is usually much better conditioned than $A$.

Our first step is again an SVD of $X_R$, which will reduce $X_R$ to $\hat{V}$ and the dimension from $2M$ to $d$. Thus, the problem is equivalent to finding all independent vectors $\mathbf{t} \in \mathbb{R}^d$ such that

$$\mathbf{t}^T \hat{V} = \mathbf{s}, \qquad s_k \in \{\pm 1\}. \tag{36}$$

The alphabet condition is written as

$$s \in \{\pm 1\} \Leftrightarrow (s-1)(s+1) = 0 \Leftrightarrow s^2 = 1. \tag{37}$$

Denoting the $k$th column of $\hat{V}$ by $\mathbf{v}_k$, substitution of (36) into (37) leads to

$$\mathbf{t}^T \mathbf{v}_k \mathbf{v}_k^T \mathbf{t} = 1, \qquad k = 1, \cdots, N. \tag{38}$$

Hence, we get the same type of problem as in the CM case, and we arrive at the same joint diagonalization problem (34) as before [100]. Other discrete symbol alphabets can in principle be treated by extensions of (37), but for higher order constellations the complexity of the algorithm quickly gets out of hand.

### *C. Binary Symbols with Residual Carrier Frequencies*

In the case of digital signals from independent sources, it is very reasonable to assume that the carrier frequencies are slightly different. For example, if the sources are modulated to 900 MHz then even if the carrier frequencies are the same up to five to six orders of magnitude, after demodulation to baseband using the nominal carrier frequency each of the sources will have a residual carrier of up to roughly 5 kHz. If the sources have a bandwidth of 20 kHz, then we can expect phase shifts in the order of 90° per symbol. Hence, the BPSK model $X = AS$ with $S_{ij} \in \{\pm 1\}$ is too naive in this case. We can either revert to the CM model to separate the sources, or we can try to separate them based on these small differences in residual carrier frequencies.

Modern-day communication systems use a common reference signal, so that the residual carrier is typically much smaller, reportedly around 500 Hz or less. The residual carrier method discussed below already works once the phase shift between the first and last symbol in the data batch is more than ±180°. For sources with a bandwidth of 20 kHz and a difference in carriers of 500 Hz, this amounts to a data batch of 20 samples. We can envision systems where cochannel users are deliberately shifted by such small amounts in order to facilitate separation. This can be regarded as a special instance of separation by "coding-induced cyclo-stationarity," and such schemes have been proposed, e.g., in [68], [71], and [73]. The main difference is that here the frequency offsets are regarded as unknown parameters. One of the few algorithms that considers this case can be found in [103]; it is a two-step iterative approach.

A more accurate source model for BPSK sources is

$$s_k \in \left\{\pm e^{j2\pi f k}\right\}, \qquad (k = 1, \cdots, N).$$

where $f$ is the unknown residual carrier frequency of the source. If we now look at $s_k^2$, we have

$$s_k^2 = \phi^k, \qquad \phi = e^{j4\pi f}.$$

Similarly as before, substitute $s_k = \mathbf{t}^* \mathbf{v}_k = \overline{\mathbf{t}}^T \mathbf{v}_k$ (note that we have to work with the complex data model). In terms of Kronecker products, we can rewrite the equation as $[\mathbf{v}_k \otimes \mathbf{v}_k]^T (\overline{\mathbf{t}} \otimes \overline{\mathbf{t}}) = \phi^k$. If we collect the row vectors $[\mathbf{v}_k \otimes \mathbf{v}_k]^T$ in a matrix $P$ as before, we obtain

$$P\mathbf{y} = \begin{bmatrix} 1 \\ \phi \\ \vdots \\ \phi^{N-1} \end{bmatrix}, \qquad \mathbf{y} = \overline{\mathbf{t}} \otimes \overline{\mathbf{t}}, \ |\phi| = 1. \tag{39}$$

The difference with the CM problem we had before is that $\phi$ is unknown. However, it can readily be estimated using the same shift-invariance ideas as before; with $P^{(1)} = J_x P$ and $P^{(2)} = J_y P$ denoting the shifted matrices, we obtain

$$P^{(1)}\mathbf{y} = \begin{bmatrix} 1 \\ \phi \\ \vdots \\ \phi^{N-2} \end{bmatrix}, \qquad P^{(2)}\mathbf{y} = \begin{bmatrix} \phi \\ \phi^2 \\ \vdots \\ \phi^{N-1} \end{bmatrix} \Rightarrow P^{(2)}\mathbf{y} = \phi P^{(1)}\mathbf{y}. \tag{40}$$

Hence, $\phi$ is a generalized eigenvalue of the matrix pair $(P^{(1)}, P^{(2)})$, and $\mathbf{y}$ is its corresponding eigenvector. If the residual frequencies are not the same, then this problem can be solved using standard linear algebra techniques [104]. The resulting $\{\phi_i\}$ are then distinct and the corresponding eigenvectors $\{\mathbf{y}_i\}$ are unique up to scaling and can directly be factored as $\mathbf{y}_i = \overline{\mathbf{t}}_i \otimes \overline{\mathbf{t}}_i$. This gives the collection of beamforming vectors.

Of course, we could also separate the sources based on their CM properties. As can be guessed from the equations, the difference in accuracy turns out to be only marginal. The main benefit in solving (40) is a somewhat reduced computational complexity.

### D. Source Independence

Algorithms similar to ACMA and residual carrier recovery have been derived in a more stochastic context and are applicable to signals that are statistically independent. The property that is used is that, for independent signals, the rows of the $S$-matrix are asymptotically orthogonal to each other.

For example, as in [87], suppose we have source signals which are uncorrelated, but have a certain temporal autocorrelation

$$R_{ss}(m) = E[\mathbf{s}_k \mathbf{s}_{k+m}^*] = D_m$$

where $D_m$ is a diagonal matrix which should be nonzero for the chosen value of the lag. If the noise is temporally uncorrelated and spatially white, the correlation matrix for the received data has the form

$$R_{xx}(0) = AD_0A^* + \sigma^2 I$$
$$R_{xx}(m) = AD_mA^*.$$

Thus, the blind beamformer $W = A^\dagger$ follows by solving for the eigenvectors of $R_{xx}(0)^{-1}R_{xx}(m)$, or equivalently, by jointly diagonalizing both $R_{xx}(0)$ and $R_{xx}(m)$. The condition for this to work is that the eigenvalues are distinct, which implies that the signals should have different spectral signatures. Even for identically distributed sources this can be assumed if the source signals have been received through different channels, or, e.g., if they have different residual carrier frequencies.

If we take multiple values for $m$, then we obtain a joint diagonalization problem, much as in (34). In essence, $W$ is computed such that the beamforming outputs look "independent," in this case with respect to their second-order statistics at selected time lags.

Depending on the signals, there might not be much distinctive information in the temporal correlations. More general techniques lift this requirement by working on higher order statistics, such as fourth-order cumulants. For the $X = AS$ problem, this has led to an algorithm called JADE that is very similar in structure to ACMA, except that it arrives at a collection of $d^2$ fourth-order cumulant matrices $C_{ij}$, all of which can be simultaneously diagonalized by $W$ [77]. An extra processing step reduces this to a more compact set of $d$ matrices to be diagonalized.

The main limitation of stochastic techniques for source separation is that they require data matrices to be fairly large so that the experimental source covariance matrices are sufficiently diagonal. This means that typically in an order of magnitude more samples are required than in the ACMA algorithms, which can make use of stronger assumptions on the data to arrive at properties that are pointwise satisfied in the absence of noise.

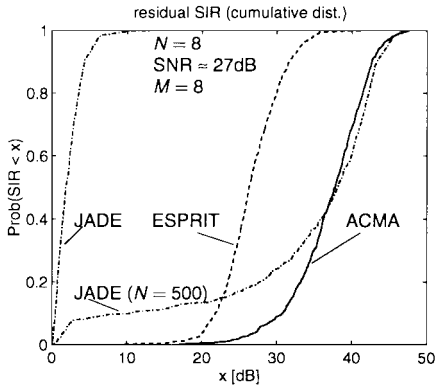



**Fig. 7.** Experimental comparison between ESPRIT, ACMA, and JADE [94]. Shown are the empirical cumulative distributions of the residual SIR, after blind beamforming based on $N = 8$ samples.

### E. Experimental Comparison

Having seen many different approaches for computing basically the same factorization, the question of choice arises. This is still a dark area, and any preference is clearly dependent on the availability and reliability of the properties, among other considerations. Nonetheless, a preliminary experimental performance comparison between ESPRIT, ACMA, JADE, and a few other methods has been reported in [94]. In one of these experiments, data was collected from a ULA with $M = 8$ elements, spaced at slightly less than half wavelength in the 900 MHz band. Two sources moving in residential traffic in a suburban setting were present at a distance of 2–3 km from the array, transmitting analog FM-modulated 1 kHz tones at slightly different carrier frequencies (15 kHz separation). The average SNR was around 27 $\pm$ 6 dB, and the sources were at least 10° spaced in azimuth. Since the sources are narrowband, the I-MIMO model $X = AS$ is appropriate.[6]

The results of this experiment are shown in Fig. 7, taken from [94]. In this case, separation based on the CM property offers some 10 dB additional SIR improvement over the direction-finding method (coherent ESPRIT). The JADE algorithm is based on restoring statistical independence properties, which requires many more samples to become effective. Eventually, it surpasses both ESPRIT and ACMA once $N > 600$ or so. Although one should be careful in drawing general conclusions from a single experiment, it seems fair to say that row span methods are quite promising for blind source separation.

## VIII. Conclusion

This paper has described algebraic methods for deterministic blind beamforming. Even within this limited framework, many properties are available and can be used to blindly separate sources and equalize channels. Column span methods are mostly parametric and try to fit a multiray channel model to the observed data. These methods are applicable if this model is valid to a reasonable accuracy, with a small number of specular rays. The requirement of

[6]The results reported here were provided courtesy of A. L. Swindlehurst and are based on measurement data shared by ArrayComm, Inc.

a model order estimation and the sensitivity to model order mismatch can be considered their Achilles heel. On the other hand, potentially useful side information is obtained, such as delays and angles of multipath rays, which enables source localization. Uncalibrated antenna arrays can be employed if there is sufficient resolution in the delays or residual carrier frequencies.

Row span methods use properties of the signals such as a CM. If these properties are present, they are very powerful and robust and not dependent on the validity of the channel model or array calibration. The strength, and at the same time the limitation, of deterministic row span methods is that they almost always require the signals to be man made. More generally applicable signal separation methods are based on stochastic properties, and, e.g., force the independence of the outputs of the beamformer, or reconstruct their distributions. Depending on the signal distributions and the *a priori* knowledge, stochastic techniques can be far superior but may require many more samples.

A major challenge in this area is to combine several signal and channel properties at the same time, since this would result in a superior and more robust beamformer. A cue to this is the observation that all algebraic methods considered in this paper lead to generalized eigenvalue problems, where the beamforming coefficients are given by the eigenvectors—the same for each method! Hence, joint diagonalization algorithms are envisioned to play an increasingly important role.

## ACKNOWLEDGMENT

The author wishes to gratefully acknowledge the several years of pleasant and fruitful collaboration with Prof. A. Paulraj and members of the Smart Antenna Research Group at Stanford University, which yielded many of the visions and ideas expressed in this paper. The author also thanks Prof. J.-F. Cardoso and Prof. P. Duhamel for their feedback and hospitality during the preparation of this work.

## REFERENCES

[1] W. C. Jakes, Ed., *Microwave Mobile Communications.* New York: Wiley, 1974.
[2] W. C. Y. Lee, *Mobile Communications Design Fundamentals.* New York: Wiley, 1993.
[3] T. S. Rappaport, *Wireless Communications: Principles and Practice.* Englewood Cliffs, NJ: Prentice-Hall, 1996.
[4] E. Lee and D. Messerschmitt, *Digital Communication.* Boston: Kluwer, 1994.
[5] R. Steele, Ed., *Mobile Radio Communications.* New York: IEEE Press, 1994.
[6] J. G. Proakis, *Digital Communications.* New York: McGraw-Hill, 1983.
[7] W. C. Y. Lee, "Overview of cellular CDMA," *IEEE Trans. Veh. Technol.,* vol. 40, pp. 291–302, May 1991.
[8] B. Ottersten, "Array processing for wireless communications," in *Proc. IEEE Workshop Statistics, Signal, Array Processing,* Corfu, Greece, June 1996, pp. 466–473.
[9] B. Sklar, "Rayleigh fading channels in mobile digital communication systems, Part I: Characterization," *IEEE Commun. Mag.,* vol. 35, pp. 90–100, July 1997.
[10] A. J. Paulraj and C. B. Papadias, "Space-time processing for wireless communications," *IEEE Signal Processing Mag.,* vol. 14, pp. 49–83, Nov. 1997.
[11] K. Pahlavan and A. H. Levesque, "Wireless data communications," *Proc. IEEE,* vol. 82, pp. 1398–1430, Sept. 1994.
[12] A. Klein, W. Mohr, R. Thomas, P. Weber, and B. Wirth, "Direction of arrival of partial waves in wideband mobile radio channels for intelligent antenna concepts," in *Proc. IEEE Vehicular Technology Conf.,* Nov. 1996, pp. 849–853.
[13] R. A. Monzingo and T. W. Miller, *Introduction to Adaptive Arrays.* New York: Wiley, 1980.
[14] B. D. van Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE Audio Speech Signal Processing Mag.,* vol. 5, pp. 4–24, Apr. 1988.
[15] D. H. Johnson and D. E. Dudgeon, *Array Signal Processing.* Englewood Cliffs, NJ: Prentice-Hall, 1993.
[16] S. Haykin, *Array Signal Processing.* Englewood Cliffs, NJ: Prentice-Hall, 1984.
[17] S. U. Pillai, *Array Signal Processing.* New York: Springer, 1989.
[18] H. Krim and M. Viberg, "Two decades of array signal processing research, the parametric approach," *IEEE Signal Processing Mag.,* vol. 13, pp. 67–94, July 1996.
[19] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propagat,* vol. 34, pp. 276–280, Mar. 1986.
[20] M. Viberg and B. Ottersten, "Sensor array processing based on subspace fitting," *IEEE Trans. Signal Processing,* vol. 39, pp. 1110–1121, May 1991.
[21] M. Viberg, B. Ottersten, and T. Kailath, "Detection and estimation in sensor arrays using weighted subspace fitting," *IEEE Trans. Signal Processing,* vol. 39, pp. 2436–2449, Nov. 1991.
[22] P. Stoica and K. C. Sharman, "Maximum likelihood methods for direction-of-arrival estimation," *IEEE Trans. Acoustics, Speech, Signal Processing,* vol. 38, pp. 1132–1143, July 1990.
[23] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Cramér–Rao bound: Further results and comparisons," *IEEE Trans. Acoustics, Speech, Signal Processing,* vol. 38, pp. 2140–2150, Dec. 1990.
[24] R. Roy and T. Kailath, "ESPRIT—Estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoustics, Speech, Signal Processing,* vol. 37, pp. 984–995, July 1989.
[25] 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.
[26] B. Ottersten, M. Viberg, and T. Kailath, "Performance analysis of the total least squares ESPRIT algorithm," *IEEE Trans. Signal Processing,* vol. 39, pp. 1122–1135, May 1991.
[27] T. J. Shan, M. Wax, and T. Kailath, "On spatial smoothing for direction-of-arrival estimation of coherent signals," *IEEE Trans. Acoustics, Speech, Signal Processing,* vol. 33, pp. 806–811, Apr. 1985.
[28] U. K. Pillai and B. H. Kwon, "Forward/backward spatial smoothing techniques for coherent signal identification," *IEEE Trans. Acoustics, Speech, Signal Processing,* vol. 37, pp. 8–15, Jan. 1989.
[29] T. Trump and B. Ottersten, "Estimation of nominal direction of arrival and angular spread using an array of sensors," *Signal Processing,* vol. 50, nos. 1–2, pp. 57–69, Apr. 1996.
[30] D. Asztély, B. Ottersten, and A. L. Swindlehurst, "A generalized array manifold model for local scattering in wireless communications," in *Proc. IEEE ICASSP,* Munich, Germany, 1997, pp. 4021–4024.
[31] 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.
[32] Y. Hua, "Estimating two-dimensional frequencies by matrix enhancement and matrix pencil," *IEEE Trans. Signal Processing,* vol. 40, pp. 2267–2280, Sept. 1992.
[33] C. P. Mathews and M. D. Zoltowski, "Eigenstructure techniques for 2-D estimation with uniform circular arrays," *IEEE Trans. Signal Processing,* vol. 42, pp. 2395–2407, Sept. 1994.
[34] 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.
[35] M. Haardt, "Efficient one-, two-, and multidimensional high-resolution array signal processing," Ph.D. dissertation, TU München, Munich, Germany, 1997.
[36] Y. Ogawa, N. Hamaguchi, K. Ohshima, and K. Itoh, "High-resolution analysis of indoor multipath propagation structure,"

*IEICE Trans. Commun.*, vol. E78-B, pp. 1450–1457, Nov. 1995.
[37] J. Gunther and A. L. Swindlehurst, "Algorithms for blind equalization with multiple antennas based on frequency domain subspaces," in *Proc. IEEE ICASSP*, Atlanta, GA, 1996, pp. 2421–2424.
[38] M. Wax and A. Leshem, "Joint estimation of directions-of-arrival and time-delays of multiple reflections of known signal," *IEEE Trans. Signal Processing*, vol. 45, pp. 2477–2484, Oct. 1997.
[39] M. C. Vanderveen, C. B. Papadias, and A. Paulraj, "Joint angle and delay estimation (JADE) for multipath signals arriving at an antenna array," *IEEE Commun. Lett.*, vol. 1, pp. 12–14, Jan. 1997.
[40] A. J. van der Veen, M. C. Vanderveen, and A. Paulraj, "Joint angle and delay estimation using shift-invariance techniques," *IEEE Trans. Signal Processing*, vol. 46, no. 2, pp. 405–418, Feb. 1998.
[41] B. H. Khalaj, A. Paulraj, and T. Kailath, "2D RAKE receivers for CDMA cellular systems," in *Proc. Globecom*, San Francisco, CA, Nov. 1994, vol. 1, pp. 400–404.
[42] M. Honig, U. Madhow, and S. Verdu, "Blind adaptive multiuser detection," *IEEE Trans. Inform. Theory*, vol. 41, pp. 944–960, July 1995.
[43] H. Liu and G. Xu, "A subspace method for signature waveform estimation in synchronous CDMA systems," *IEEE Trans. Commun.*, vol. 44, pp. 1346–1354, Oct. 1996.
[44] H. Liu and M. D. Zoltowski, "Blind equalization in antenna array CDMA systems," *IEEE Trans. Signal Processing*, vol. 45, pp. 161–172, Jan. 1997.
[45] M. Torlak and G. Xu, "Blind multiuser channel estimation in asynchronous CDMA systems," *IEEE Trans. Signal Processing*, vol. 45, pp. 137–147, Jan. 1997.
[46] J. Ramos and M. D. Zoltowski, "Reduced complexity blind 2D RAKE receiver for CDMA," in *Proc. IEEE SP Workshop on Statistics Signal Array Processing*, Corfu, Greece, 1996, pp. 502–505.
[47] J. Ramos, M. D. Zoltowski, and H. Liu, "A low-complexity space-time RAKE receiver for DS-CDMA communications," *IEEE Signal Processing Lett.*, vol. 4, pp. 262–265, Sept. 1997.
[48] M. D. Zoltowski and C. P. Mathews, "Real-time frequency and 2-D angle estimation with sub-Nyquist spatio-temporal sampling," *IEEE Trans. Signal Processing*, vol. 42, pp. 2781–2794, Oct. 1994.
[49] K.-B. Yu, "Recursive super-resolution algorithm for low-elevation target angle tracking in multipath," *Proc. Inst. Elect. Eng.—Radar, Sonar and Navigation*, vol. 141, no. 4, pp. 223–229, Aug. 1994.
[50] K. Abed-Meraim, W. Qui, and Y. Hua, "Blind system identification," *Proc. IEEE*, vol. 85, no. 8, pp. 1310–1322, Aug. 1997.
[51] L. Tong, G. Xu, and T. Kailath, "Blind identification and equalization based on second-order statistics: A time domain approach," *IEEE Trans. Inform. Theory*, vol. 40, pp. 340–349, Mar. 1994.
[52] E. Moulines, P. Duhamel, J.-F. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Trans. Signal Processing*, vol. 43, pp. 516–525, Feb. 1995.
[53] D. Slock, "Blind fractionally-spaced equalization, perfect-reconstruction filter banks and multichannel linear prediction," in *Proc. IEEE ICASSP*, 1994, pp. IV:585–588.
[54] K. Abed-Meraim, E. Moulines, and P. Loubaton, "Prediction error method for second-order blind identification," *IEEE Trans. Signal Processing*, vol. 45, pp. 694–705, Mar. 1997.
[55] Y. Hua, K. Abed-Meraim, and M. Wax, "Blind system identification using minimum noise subspace," *IEEE Trans. Signal Processing*, vol. 45, pp. 770–773, Mar. 1997.
[56] H. Liu and G. Xu, "Closed-form blind symbol estimation in digital communications," *IEEE Trans. Signal Processing*, vol. 43, pp. 2714–2723, Nov. 1995.
[57] A. J. van der Veen, S. Talwar, and A. Paulraj, "A subspace approach to blind space-time signal processing for wireless communication systems," *IEEE Trans. Signal Processing*, vol. 45, pp. 173–190, Jan. 1997.
[58] G. Xu, H. Liu, L. Tong, and T. Kailath, "A least-squares approach to blind channel identification," *IEEE Trans. Signal Processing*, vol. 43, pp. 2982–2993, Dec. 1995.
[59] D. Gesbert, P. Duhamel, and S. Mayrargue, "On-line blind multichannel equalization based on mutually referenced filters," *IEEE Trans. Signal Processing*, vol. 45, pp. 2307–2317, Sept. 1997.
[60] S. V. Schell and D. L. Smith, "Improved performance of blind equalization using prior knowledge of transmitter filter," in *Proc. IEEE MILCOM*, 1994, pp. 128–132 vol. 1.
[61] Z. Ding and Z. Mao, "Knowledge-based identification of fractionally sampled channels," in *Proc. IEEE ICASSP*, Detroit, MI, May 1995, pp. 1996–1999.
[62] M. Cedervall, B. C. Ng, and A. Paulraj, "Structured methods for blind multi-channel identification," in *Proc. 13th Int. Conf. Digital Signal Processing*, Santorini, Greece, July 1997, pp. 387–390.
[63] D. Gesbert and P. Duhamel, "Robust blind channel identification and equalization based on multi-step predictors," in *Proc. IEEE ICASSP*, Munich, Germany, 1997, pp. 3621–3624.
[64] E. De Carvalho and D. T. M. Slock, "Cramér–Rao bounds for semi-blind, blind and training sequence based channel estimation," in *IEEE Workshop Signal Processing Advances in Wireless Communications*, Paris, Apr. 1997, pp. 129–132.
[65] D. Pal, "Fractionally spaced equalization of multipath channels: A semi-blind approach," in *Proc. IEEE ICASSP*, 1993, pp. 9–12 vol. 3.
[66] A. Gorokhov and P. Loubaton, "Semi-blind second order identification of convolutive channels," in *Proc. IEEE ICASSP*, 1997, pp. 3905–3908.
[67] E. de Carvalho and D. T. M. Slock, "Maximum-likelihood blind FIR multi-channel estimation with Gaussian prior for the symbols," in *Proc. IEEE ICASSP*, 1997, pp. 3593–3596.
[68] B. G. Agee, S. V. Schell, and W. A. Gardner, "Spectral self-coherence restoral: A new approach to blind adaptive signal extraction using antenna arrays," *Proc. IEEE*, vol. 78, pp. 753–767, Apr. 1990.
[69] S. V. Schell, "An overview of sensor array processing for cyclostationary signals," in *Cyclostationarity in Communications and Signal Processing*, W. A. Gardner, Ed. Piscataway, NJ: IEEE Press, 1994, pp. 168–239.
[70] A. Chevreuil and P. Loubaton, "On the use of conjugate cyclo-stationarity: A blind second-order multi-user equalization method," in *Proc. IEEE ICASSP*, Atlanta, GA, 1996, pp. 2439–2442.
[71] M. K. Tsatsanis and G. B. Giannakis, "Coding induced cyclostationarity for blind channel equalization," in *Proc. 29th Conf. Information Sciences Systems*, Baltimore, MA, 1995, pp. 685–690.
[72] M. K. Tsatsanis and G. B. Giannakis, "Transmitter induced cyclostationarity for blind channel equalization," *IEEE Trans. Signal Processing*, vol. 45, pp. 1785–1794, July 1997.
[73] A. Chevreuil and Ph. Loubaton, "Blind second-order identification of FIR channels: Forced cyclostationarity and structured subspace method," *IEEE Signal Processing Lett.*, vol. 4, pp. 204–206, July 1997.
[74] W. A. Gardner, "A new method of channel identification," *IEEE Trans. Commun.*, vol. 39, pp. 813–817, June 1991.
[75] K. J. Pope and R. E. Bogner, "Blind signal separation. I. Linear, instantaneous combinations," *Digital Signal Processing*, vol. 6, no. 1, pp. 5–16, Jan. 1996.
[76] C. Jutten and J. Herault, "Blind separation of sources. I. An adaptive algorithm based on neuromimetic architecture," *Signal Processing*, vol. 24, no. 1, pp. 1–10, July 1991.
[77] J. F. Cardoso and A. Souloumiac, "Blind beamforming for non-Gaussian signals," *Proc. Inst. Elect. Eng.*, vol. 140, pt. F, no. 6, pp. 362–370, Dec. 1993.
[78] L. Tong, Y. Inouye, and R.-W. Liu, "Waveform-preserving blind estimation of multiple independent sources," *IEEE Trans. Signal Processing*, vol. 41, pp. 2461–2470, July 1993.
[79] P. Comon, "Independent component analysis, a new concept?," *Signal Processing*, vol. 36, no. 3, pp. 287–314, Apr. 1994.
[80] J.-F. Cardoso and B. H. Laheld, "Equivariant adaptive source separation," *IEEE Trans. Signal Processing*, vol. 44, pp. 3017–3030, Dec. 1996.
[81] L. De Lathauwer, B. De Moor, and J. Vandewalle, "Independent component analysis based on higher-order statistics only," in *Proc. IEEE Signal Processing Workshop on Statistical Signal Array Processing*, Corfu, Greece, 1996, pp. 356–359.

[82] A. Mansour and C. Jutten, "A direct solution for blind separation of sources," *IEEE Trans. Signal Processing*, vol. 44, no. 3, pp. 746–748, Mar. 1996.
[83] L. De Lathauwer, "Signal processing based on multilinear algebra," Ph.D. dissertation, KU Leuven, Leuven, Belgium, 1997.
[84] E. Gönen and J. M. Mendel, "Applications of cumulants to array processing—Part III: Blind beamforming for coherent signals," *IEEE Trans. Signal Processing*, vol. 45, pp. 2252–2264, Sept. 1997.
[85] S. Van Gerven and D. Van Compernolle, "On the use of decorrelation in scalar signal separation," in *Proc. IEEE ICASSP*, 1994, pp. III.57–60.
[86] U. Lindgren and A. J. van der Veen, "Source separation based on second order statistics—An algebraic approach," in *Proc. IEEE Signal Processing Workshop Statistical Signal Array Procssing*, Corfu, Greece, June 1996, pp. 324–327.
[87] A. Belouchrani, K. A. Meraim, J.-F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE Trans. Signal Processing*, vol. 45, pp. 434–444, Feb. 1997.
[88] Y. Sato, "A method of self-recovering equalization for multilevel amplitude-modulation systems," *IEEE Trans. Commun.*, vol. 23, pp. 679–682, June 1975.
[89] D. N. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," *IEEE Trans. Commun.*, vol. 28, pp. 1867–1875, Nov. 1980.
[90] J. R. Treichler and B. G. Agee, "A new approach to multipath correction of constant modulus signals," *IEEE Trans. Acoustics, Speech, Signal Processing*, vol. 31, pp. 459–471, Apr. 1983.
[91] J. R. Treichler and M. G. Larimore, "New processing techniques based on constant modulus adaptive algorithm," *IEEE Trans. Acoustics, Speech, Signal Processing*, vol. 33, pp. 420–431, Apr. 1985.
[92] J. J. Shynk and R. P. Gooch, "The constant modulus array for cochannel signal copy and direction finding," *IEEE Trans. Signal Processing*, vol. 44, pp. 652–660, Mar. 1996.
[93] A. J. van der Veen and A. Paulraj, "An analytical constant modulus algorithm," *IEEE Trans. Signal Processing*, vol. 44, pp. 1136–1155, May 1996.
[94] A. L. Swindlehurst, M. J. Goris, and B. Ottersten, "Some experiments with array data collected in actual urban and suburban environments," in *IEEE Workshop Signal Processing Advances Wireless Communications*, Paris, France, Apr. 1997, pp. 301–304.
[95] C. B. Papadias and A. J. Paulraj, "A constant modulus algorithm for multiuser signal separation in presence of delay spread using antenna arrays," *IEEE Signal Processing Lett.*, vol. 4, pp. 178–181, June 1997.
[96] D. Yellin and B. Porat, "Blind identification of FIR systems excited by discrete-alphabet inputs," *IEEE Trans. Signal Processing*, vol. 41, pp. 1331–1339, Mar. 1993.
[97] S. Talwar, M. Viberg, and A. Paulraj, "Blind estimation of synchronous co-channel digital signals using an antenna array. Part I: Algorithms," *IEEE Trans. Signal Processing*, vol. 44, pp. 1184–1197, May 1996.
[98] K. Anand, G. Mathew, and V. U. Reddy, "Blind separation of multiple co-channel BPSK signals arriving at an antenna array," *IEEE Signal Processing Lett.*, vol. 2, pp. 176–178, Sept. 1995.
[99] A. Swindlehurst, S. Daas, and J. Yang, "Analysis of a decision directed beamformer," *IEEE Trans. Signal Processing*, vol. 43, pp. 2920–2927, Dec. 1995.
[100] A. J. van der Veen, "Analytical method for blind binary signal separation," *IEEE Trans. Signal Processing*, vol. 45, pp. 1078–1082, Apr. 1997.
[101] L. K. Hansen and G. Xu, "A hyperplance-based algorithm for the digital co-channel communications problem," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1536–1548, Sept. 1997.
[102] A. J. van der Veen and J. Tol, "Separation of zero/constant modulus signals," in *Proc. IEEE ICASSP*, Munich, Germany, Apr. 1997, pp. 3445–3448.
[103] L. Tong, "Joint blind signal detection and carrier recovery over fading channels," in *Proc. IEEE ICASSP*, Detroit, MI, 1995, pp. 1205–1208.
[104] A. J. van der Veen, "Blind separation of BPSK sources with residual carriers," submitted for publication.
[105] P. Lancaster and M. Tismenetsky, *The Theory of Matrices: With Applications*, 2nd ed. Orlando, FL: Academic, 1985.
[106] R. A. Horn and C. R. Johnson, *Matrix Analysis.* Cambridge, U.K.: Cambridge, 1985.
[107] G. Golub and C. F. Van Loan, *Matrix Computations.* Baltimore, MD: Johns Hopkins Press, 1989.
[108] G. W. Stewart and J. G. Sun, *Matrix Perturbation Theory*. Boston, MA: Academic, 1990.
[109] S. Chandrasekaran and I. C. F. Ipsen, "On rank-revealing factorizations," *SIAM J. Matrix Anal. Applicat.*, vol. 15, no. 2, pp. 592–622, Apr. 1994.
[110] G. W. Stewart, "An updating algorithm for subspace tracking," *IEEE Trans. Signal Processing*, vol. 40, pp. 1535–1541, June 1992.
[111] A. J. van der Veen, "A Schur method for low-rank matrix approximation," *SIAM J. Matrix Anal. Applicat.*, vol. 17, no. 1, pp. 139–160, Jan. 1996.
[112] P. Comon and G. H. Golub, "Tracking a few extreme singular values and vectors in signal processing," *Proc. IEEE*, vol. 78, pp. 1327–1343, Aug. 1990.
[113] B. Yang, "Projection approximation subspace tracking," *IEEE Trans. Signal Processing*, vol. 43, pp. 95–107, Jan. 1995.
[114] M. Moonen, P. Van Dooren, and F. Vanpoucke, "On the QR algorithm and updating the SVD and URV decomposition in parallel," *Linear Algebra Applicat.*, vol. 188/189, pp. 549–568, July 1993.
[115] P. Binding, "Simultaneous diagonalization of several Hermitian matrices," *SIAM J. Matrix Anal. Applicat.*, vol. 4, no. 11, pp. 531–536, 1990.
[116] M. T. Chu, "A continuous Jacobi-like approach to the simultaneous reduction of real matrices," *Lin. Algebra Applicat*, vol. 147, pp. 75–96, 1991.
[117] A. Bunse-Gerstner, R. Byers, and V. Mehrmann, "Numerical methods for simultaneous diagonalization," *SIAM J. Matrix Anal. Applicat.*, vol. 4, pp. 927–949, 1993.
[118] B. D. Flury and B. E. Neuenschwander, "Simultaneous diagonalization algorithms with applications in multivariate statistics," in *Approximation and Computation*, R. V. M. Zahar, Ed. Basel, Germany: Birkhäuser, 1995, pp. 179–205.
[119] J.-F. Cardoso and A. Souloumiac, "Jacobi angles for simultaneous diagonalization," *SIAM J. Matrix Anal. Applicat.*, vol. 17, no. 1, pp. 161–164, 1996.
[120] M. Wax and J. Sheinvald, "A least-squares approach to joint diagonalization," *IEEE Signal Processing Lett.*, vol. 4, pp. 52–53, Feb. 1997.
[121] R. M. Trim, "Mode S: An introduction and overview," *Electron. Commun. Eng. J.*, vol. 2, pp. 53–59, Apr. 1990.
[122] "International standards and recommended practices, aeronautical telecommunications: Annex 10," in *Convention International Civil Aviation*, International Civil Aviation Organization, 1985.
[123] E. Chaumette, P. Comon, and D. Muller, "ICA-based technique for radiating sources estimation: Application to airport surveillance," *Proc. Inst. Elect. Eng.* vol. 140, pt. F, no. 6, pp. 395–401, Dec. 1993.

**Alle-Jan van der Veen** (Member, IEEE) was born in The Netherlands in 1966. He received the M.Sc. degree (cum laude) in electrical engineering in 1988 and the Ph.D. degree (cum laude) in 1993, both from Delft University of Technology, Delft, The Netherlands..

Throughout 1994 he was a Postdoctoral Scholar at Stanford University, Stanford, CA, in the Scientific Computing/Computational Mathematics Group and in the Information Systems Lab. Presently he is Assistant Professor in the Signal Processing group of DIMES, Delft University of Technology. His research interests are in the general area of system theory applied to signal processing, and, in particular, algebraic methods for array signal processing.

Dr. van der Veen is the recipient of the 1994 and 1999 IEEE Signal Processing paper awards.