A Subspace Approach to Blind Space-Time Signal Processing for Wireless Communication Systems

Alle-Jan van der Veen, *Member, IEEE*, Shilpa Talwar, and Arogyaswami Paulraj, *Fellow, IEEE*

***Abstract*—The two key limiting factors facing wireless systems today are multipath interference and multiuser interference. In this context, a challenging signal processing problem is the joint space-time equalization of multiple digital signals transmitted over multipath channels. We propose a blind approach that does not use training sets to estimate the transmitted signals and the space-time channel. Instead, this approach takes advantage of spatial and temporal oversampling techniques and the finite alphabet property of digital signals to determine the user symbol sequences. The problem of channels with largely differing and ill-defined delay spreads is discussed. The proposed approach is tested on actual channel data.**

I. Introduction

A challenging problem in signal processing is the blind joint space-time equalization of multiple digital signals transmitted over multipath channels. An important area where such a problem arises is wireless (mobile) communications. Consider a scenario where several users are trying to talk to a central base station, which has several antennas (viz., Fig. 1). A space-time equalizer at the base station combines two signal processing aspects: *equalization* (or echo canceling) to combat the intersymbol interference due to large-delay multipath and *source separation* to combat cochannel interference (CCI). The CCI might be interfering signals at the same frequency from neighboring communication cells, or we might intentionally allow multiple users at the same frequency in order to increase the system capacity. The latter is known as space division multiple access (SDMA) because it essentially separates users based on differences in location.

Current communication systems such as IS-54 and GSM require some amount of equalization (up to five symbol periods in GSM and up to one symbol period in IS-54) but are not designed to handle cochannel users. To assist "classical" single-user channel identification algorithms, a fair number of training symbols are incorporated in the data packets.

Manuscript received December 1, 1995; revised August 21, 1996. This work was supported by the Department of the Army, Army Research Office, under Grant DAAH04-95-1-0436. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Army Research Office or the U.S. Government. Additional support was provided by Ericsson Inc., Bell Northern Research Inc., Qualcomm Inc., and Hughes Network Systems Inc.

A. J. van der Veen is with the Department of Electrical Engineering/DIMES, Delft University of Technology, Delft, The Netherlands.

S. Talwar was with the Scientific Computing Program, Stanford University, Stanford, CA 94305 USA. She is now with Stanford Telecom, Sunnyvale, CA USA.

A. Paulraj is with the Information Systems Laboratory, Stanford University, Stanford, CA 94305 USA.

Publisher Item Identifier S 1053-587X(97)00529-1.

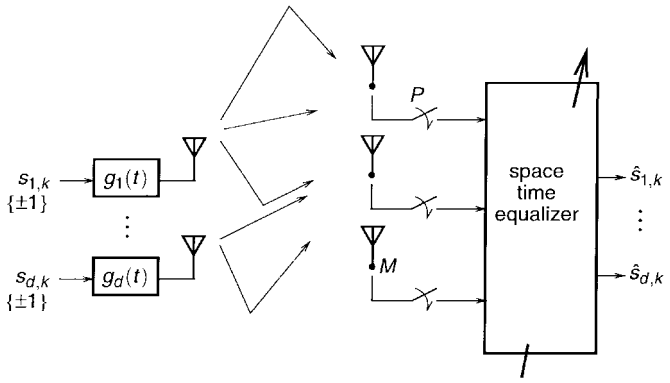


Fig. 1. Multiray scenario in wireless communications.

However, in recent years, it became gradually known that digital signals can also be separated and equalized *blindly*, i.e., without the aid of training sequences, by exploiting the underlying structure of the signals. Although the use of training sequences is an inherently more robust way to estimate the channel, there are several reasons for studying blind algorithms, aside from the obvious academic and military motivations. Most notably, adding unnecessary training bits is a direct waste of the available bandwidth. In addition, training is not efficient in rapidly time-varying channels or in protocols with very small data packages, such as the uplink of wireless teletypes in PCS or in distributed networks. Training requires synchronization, which is not always available or feasible in multiuser scenarios. Finally, the insights gained are also applicable to other systems such as CDMA, where it might be used to improve the near-far resistance.

Blind algorithms have now become a very active research area, in particular in the context of digital communication signals, where there are several leverages for solving the blind finite impulse response, multiple inputs, multiple outputs (FIR-MIMO) identification problem considered in this paper. For example, the *fixed symbol rate* of digital signals in combination with linear channels, multiple antennas, and oversampling allows us to blindly synchronize and equalize (but not separate) such signals. Statistically, oversampling digital signals gives rise to cyclostationarity of the spectrum [3]. Tong *et al.* were the first to realize that cyclostationarity allows the identification of nonminimum phase FIR-SISO channels from second-order statistics [4], [5]. In a deterministic discrete-time setting, the property leads to structured (Toeplitz) matrices and has inspired several subspace-based algorithms [6]–[9]. A second useful property is the *finite alphabet* (FA) structure

of digital signals. For equalization, this has been exploited in decision-directed adaptive algorithms [10]–[13] as well as in joint channel estimation and sequence detection [14], [15]. Several iterative algorithms for the separation of instantaneous superpositions of multiple finite alphabet signals (I-MIMO) were originally proposed by Talwar *et al.* [16]–[18]; an algorithm based on expectation maximization appeared in [19]. The two properties are in fact readily combined into one algorithm to solve the FIR-MIMO problem, as was discovered independently by Liu and Xu [20], [21] and the present authors [1], [2]. Related work on blind FIR-MIMO identification was carried out in parallel by Abed-Meraim *et al.* as well [22].

Many other signal properties can also be used for blind estimation, for example, source independence and high-order statistical properties and constant modulus properties. In addition, the spatial properties of the receiving antenna array might be known, which allows signal separation based on differences in directions of arrival, provided the number of antennas is large enough [23]. Assuming a multiray propagation scenario, knowledge of both the pulse-shape function and the array manifold allows a joint delay and angle estimation of all propagation paths (viz. [24], [25]).

A. Contributions

In this paper, we consider the FIR-MIMO source separation and equalization problem. We assume all sources transmit digital communication signals with the same symbol rate and alphabet, both of which are known *a priori*. The measured data is obtained from a cluster of fractionally sampled antennas. We do not assume knowledge of the antenna array mainly because we do not attempt to resolve the individual directions of the incoming rays.

The proposed methods are subspace-based block-algorithms and attempt to provide a structured factorization of the data matrix into a channel matrix times a symbol matrix. The constant symbol rate translates to a block-Toeplitz structure of the symbol matrix, which is sufficient for equalization. We rely strongly on the finite alphabet property for the separation of the individual signals and to some extent also for the equalization.

The outline of the paper is as follows. The data model for the FIR-MIMO scenario is presented in Section II. Maximum-likelihood techniques for blind sequence estimation are not computationally feasible. As an alternative, a subspace-based approach is presented in Section III. A proof of identifiability and some insight into the underlying subspace intersection method is given in Section IV, as well as a comparison among a few alternative methods via computer simulation. In Section V, we indicate problems that occur with subspace intersection techniques if the channel lengths are not well-defined and suggest a possible solution. Finally, in Section VI, the proposed algorithm is tested on simulated data based on actual wireless indoor channels.

The paper encompasses preliminary short versions [1], [2]. During review, the full version of Liu and Xu's approach [20], [21] appeared in [26]. Although the original methods used are quite identical, the present paper extends beyond [26] by providing a significantly more efficient implementation, comparing the proposed approach to a similar technique in which the channel is identified first and subsequently inverted (cf. [7]) and addressing the case with differing and ill-defined channel lengths.

B. Notation

Lower-case bold, as in $\mathbf{x}$, denotes vectors. For a matrix A, A^T is the transpose, A^* is the complex conjugate transpose, $A^\dagger$ is the Moore–Penrose pseudo-inverse, col(A) and row(A) denote the column span and row span of A, and $\| A \|_F$ is the Frobenius norm of A. vec (A) is the "vectoring" operation that stacks all columns of A in a single vector, and $\otimes$ is the Kronecker product.

II. Data Model

An array of M sensors, with outputs $x_1(t), \cdots, x_M(t)$, receives d digital signals $s_1(t), \cdots, s_d(t)$, each of which is described as a sequence of dirac pulses $s_j(t) = \sum_{k=-\infty}^{\infty} s_{j,k}\delta(t - kT)$. For convenience, we assume the symbol rate T is normalized to $T = 1$, and the digital symbols $s_{j,n}$ belong to a known finite alphabet $\Omega = \{\pm 1, \pm 3, \ldots, \pm(N_\Omega - 1)\}$ for real signals, or $\Omega = \{\pm 1, \pm 3, \ldots, \pm(N_\Omega - 1)\} \oplus \{\pm j, \pm j3, \ldots, \pm j(N_\Omega - 1)\}$ for complex signals. The waveform received at the array consists of multiple paths per signal, with echos arriving from different angles, with different delays and attenuations. The impulse response of the channel from the jth source to the ith sensor $h_{ij}(t)$ is a convolution of the pulse shaping filter $g_j(t)$ and the actual channel from $s_j(t)$ to $x_i(t)$. We include any propagation delays and delays due to asynchronous signals in $h_{ij}(t)$. The data model is written compactly as the convolution $\mathbf{x}(t) = H(t) * \mathbf{s}(t)$, where

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ \vdots \\ x_M(t) \end{bmatrix}, \; H(t) = \begin{bmatrix} h_{11}(t) & \cdots & h_{1d}(t) \\ \vdots & & \vdots \\ h_{M1}(t) & \cdots & h_{Md}(t) \end{bmatrix},$$
$$\mathbf{s}(t) = \begin{bmatrix} s_1(t) \\ \vdots \\ s_d(t) \end{bmatrix}.$$

It is common to assume at this point that all M channels $h_{ij}(t)$ are FIR filters of length at most $L_j \in \mathbb{N}$:

$$h_{ij}(t) = 0, \quad t \notin [0, L_j), \qquad i = 1, \cdots, M; \; j = 1, \cdots, d.$$

The maximal channel length among all sources is denoted by $L = \max_j L_j$. 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 $\mathbf{x}(t)$. Indeed, for $t = n + \tau$, where $n \in \mathbb{Z}$ and $0 \leq \tau < 1$, the convolution $x_i(t) = \sum_j h_{ij} * s_j(t)$ can be expressed as

$$x_i(n+\tau) = \sum_{k=0}^{L_1 - 1} h_{i1}(k+\tau) s_{1,n-k} + \cdots + \sum_{k=0}^{L_d - 1} h_{id}(k+\tau) s_{d,n-k}. \quad (1)$$

For simplicity of the exposition, we initially assume that all channels have the same length L and generalize later on.

Suppose we sample each $x_i(t)$ at a rate $P \in \mathbb{N}$, where P is the oversampling factor, and collect samples during N symbol

periods; then, we can construct a data matrix X as

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

The kth column $\mathbf{x}_k$ of X contains the MP spatial and temporal samples taken during the kth interval. Based on the model of $x_i(t)$ in (1), it follows that X has a factorization

$$\begin{aligned} X &= HS_L \\ H &= \begin{bmatrix} H(0) & H(1) & \cdots & H(L-1) \\ H(\frac{1}{P}) & \cdot & & \cdot \\ \vdots & & & \vdots \\ H(\frac{P-1}{P}) & \cdot & \cdots & H(L-\frac{1}{P}) \end{bmatrix} \\ S_L &= \begin{bmatrix} \mathbf{s}_0 & \mathbf{s}_1 & \ddots & \mathbf{s}_{N-2} & \mathbf{s}_{N-1} \\ \ddots & \ddots & \ddots & \ddots & \mathbf{s}_{N-2} \\ \mathbf{s}_{-L+2} & \mathbf{s}_{-L+3} & \ddots & \ddots & \ddots \\ \mathbf{s}_{-L+1} & \mathbf{s}_{-L+2} & \ddots & \ddots & \mathbf{s}_{N-L} \end{bmatrix} \end{aligned} \qquad (2)$$

$$H : MP \times dL, \quad S_L : dL \times N.$$

The matrix H represents the unknown space-time channel, and the block-Toeplitz matrix $S = S_L$ contains the transmitted symbols.[1] For generality, we have assumed that the measured block of data starts while the transmission of each of the signals was already in progress, i.e., $\mathbf{x}_0$ is determined by previous symbols $\mathbf{s}_{-L+1}, \cdots, \mathbf{s}_{-1}$ as well as $\mathbf{s}_0$. A similar assumption is made on $\mathbf{x}_{N-1}$. Note that if the channels do not all have the same length L, then certain columns of H are equal to zero. Given X, our goal in blind estimation is to find H and S such that S is a block-Toeplitz matrix, and the symbols in S satisfy the finite alphabet property.

If the source alphabet is real, then it is customary to work with a real-valued data model by redefining (with some abuse of notation)

$$X := \begin{bmatrix} \text{real}(X) \\ \text{imag}(X) \end{bmatrix}, \quad H := \begin{bmatrix} \text{real}(H) \\ \text{imag}(H) \end{bmatrix}. \qquad (3)$$

This effectively doubles the number of observables MP while halving the noise power on each entry.

The algorithms we consider in this paper rely on the existence of a "filtering matrix" W such that $WX = S$. This implies that the row span of S is equal to (or contained in) the row span of X. For this to be true, it is necessary that H has full column rank, which implies that $MP \geq dL$. This may put undue requirements on the number of antennas or oversampling rate. However, it is possible to ease this condition by making use of time-invariance and the structure of S. Extending X to a block-Hankel matrix by left shifting and stacking m times (as discussed later, m can be viewed as an equalizer length measured in symbol periods), we obtain

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

This augmented data matrix $\mathcal{X}_m$ has a factorization

$$\mathcal{X}_m = \mathcal{H}_m \mathcal{S}_{L+m-1}$$
$$= \begin{bmatrix} \mathbf{0} & H & & \\ & \iddots & \iddots & \\ & & H & \\ H & & & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{s}_{m-1} & \ddots & \mathbf{s}_{N-2} & \mathbf{s}_{N-1} \\ \ddots & \ddots & \ddots & \mathbf{s}_{N-2} \\ \mathbf{s}_{-L+2} & \mathbf{s}_{-L+3} & \ddots & \ddots \\ \mathbf{s}_{-L+1} & \mathbf{s}_{-L+2} & \ddots & \mathbf{s}_{N-L-m+1} \end{bmatrix}$$
$$\mathcal{H}_m : mMP \times d(L+m-1),$$
$$\mathcal{S}_{L+m-1} : d(L+m-1) \times (N-m+1) \qquad (4)$$

(the m shifts of H to the left are each over d positions) and the objective becomes, for given $\mathcal{X}$, to determine factors $\mathcal{H}$ and $\mathcal{S}$ of the indicated structure such that the entries of $\mathcal{S}$ belong to the finite alphabet. As we show in the sequel, identification is possible if this is a minimal-rank factorization. *Necessary* conditions for $\mathcal{X}$ to have a unique factorization $\mathcal{X} = \mathcal{H}\mathcal{S}$ are that $\mathcal{H}$ is a "tall" matrix and that $\mathcal{S}$ is a "wide" matrix, which for $L > 1$ leads to

$$\begin{aligned} MP &> d \\ m &\geq \frac{dL-d}{MP-d} \\ N &> dL + (d+1)(m-1). \end{aligned} \qquad (5)$$

Given sufficient data, only $MP > d$ poses a fundamental identifiability restriction.

Note that these conditions are *not sufficient* for $\mathcal{H}$ and $\mathcal{S}$ to have full rank. One case where $\mathcal{H}$ does not have full rank is when the channels do not have equal lengths, in which case the rank of $\mathcal{X}$ is at most $\sum L_j + d(m-1)$. Ill-conditioned cases might occur when the channels are bandwidth limited so that sampling faster than the Nyquist rate does not provide independent linear combinations of the same symbols. In principle, the maximal effective P is given by the ratio of the Nyquist rate and the symbol rate [27]. (There may be other practical reasons to select a larger P, e.g., to correct for errors in carrier recovery. This is not considered here.)

For SISO models, the condition that $\mathcal{H}_m$ is of full rank is usually formulated in terms of "common zeros"; if the z-transforms $h_i(z)$ of the rows of H do not have a root in common, then $\mathcal{H}_m$ has full column rank for at least all $m \geq L$ (viz., e.g., [7], [26]). For arbitrary channels, this technical condition holds almost surely. In the FIR-MIMO case, the corresponding requirement is that $H(z)$ is "irreducible and column reduced" (viz. [22]).

III. Subspace-based Approaches

According to the previous section, the basic problem in solving the blind FIR-MIMO problem is, for a given matrix

[1] The subscript L denotes the number of block rows in S. We usually omit the subscript if this does not lead to confusion.

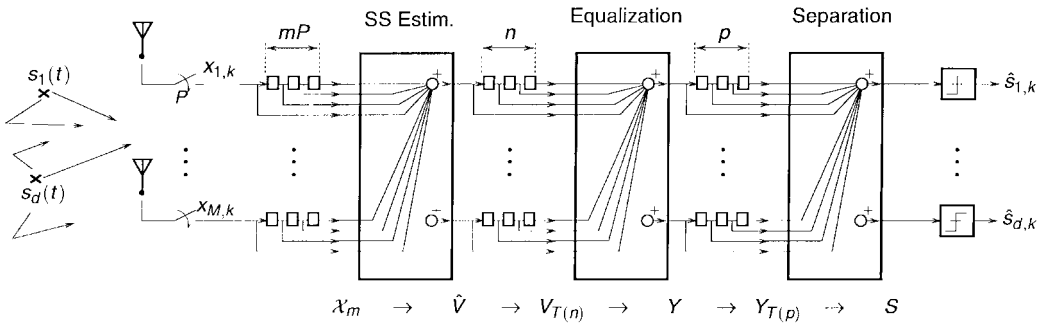


Fig. 2. Multistage equalization/separation filter.

X, to find a factorization $X = H\mathcal{S}$, where $\mathcal{S} \equiv \mathcal{S}_L$ is block-Toeplitz with entries $(\mathcal{S})_{ij} \in \Omega$. If we assume that the data matrix X is corrupted by additive white Gaussian noise, then the maximum likelihood criterion yields the nonlinear least squares minimization problem

$$\min_{H,\ \mathcal{S}\in\Omega:\ \text{block-Toeplitz}} \|X - H\mathcal{S}\|_F^2. \tag{6}$$

To find an exact solution of this nonlinear optimization problem is computationally formidable. It is possible to approach the optimum via iterative techniques that alternatingly estimate H and $\mathcal{S}$, starting from some initial estimate for H [28]. This approach is still computationally expensive due to the repeated enumeration of all possible sequences of length L using the Viterbi algorithm. In addition, the initial point has to be quite accurate in order to converge to the global minimum, rather than one of the numerous local minima.

The subspace-based approaches derived in this section simplify the problem by breaking it up into two subproblems. Suppose that the channels have equal lengths L and that the conditions (5) are satisfied. Then,

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

To factor $\mathcal{X}$ into $\mathcal{X} = \mathcal{H}\mathcal{S}$, the strategy is to find either $\mathcal{S}$, which is a block-Toeplitz matrix with a specified row span, or $\mathcal{H}$, which is a block-Hankel matrix with a specified column span. In the scalar case ($d = 1$ signal), a number of algorithms have been proposed for doing the latter, in particular, [7] and [8], and it is straightforward to extend these algorithms to the vector case ($d > 1$), presuming the channel lengths are all equal. However, for $d > 1$, subspace information alone leads to an ambiguity: $\mathcal{X} = (\mathcal{H}D^{-1})(D\mathcal{S})$ is a factorization with the same subspaces for $D = \text{diag}[A, \cdots, A]$ and A any invertible $d \times d$ matrix. This ambiguity is resolved in a second step by taking advantage of the finite-alphabet property of the signals.

We outline three approaches: one that directly estimates $\mathcal{S}$ from its row span, as was originally proposed in [20], [1], and [2], then an entirely equivalent but computationally more attractive version, and finally, an approach in which $\mathcal{H}$ is estimated first. In the absence of noise, all approaches give exact results. Note that none of these approaches provides a factorization $\mathcal{X} = \mathcal{H}\mathcal{S}$ in which *both* factors are forced to have the required Toeplitz or Hankel structure so that they are suboptimal in that respect.

A. Step 1: Estimating the Row Span of $\mathcal{S}$

We work with the extended matrix $\mathcal{X} = \mathcal{X}_m$ in (4). The first step in the direct algorithms is the estimation of an (orthonormal) basis of the row span of $\mathcal{S} = \mathcal{S}_{L+m-1}$ from the row span of $\mathcal{X}$. Suppose, as before, that the channels have equal lengths, (5) holds, and $\mathcal{H}$ has full column rank. Then, row$(\mathcal{X})$ = row$(\mathcal{S})$ so that we can determine the row span of $\mathcal{S}$ from that of $\mathcal{X}$. This requires the computation of an SVD of $\mathcal{X}$, $\mathcal{X} = U\Sigma V$ where U, V are unitary matrices, and Σ is a diagonal matrix containing the singular values in nonincreasing order [29]. Without noise, the rank $d_{\mathcal{X}}$ of $\mathcal{X}$ is equal to the number of nonzero singular values, and we can write $\mathcal{X} = \hat{U}\hat{\Sigma}\hat{V}$, where $\hat{U}$ consists of the first $d_{\mathcal{X}}$ columns of U, $\hat{\Sigma}$ is a diagonal $d_{\mathcal{X}} \times d_{\mathcal{X}}$ matrix consisting of the nonzero singular values, and $\hat{V}$ is the first $d_{\mathcal{X}}$ rows of V, forming an orthonormal basis for row$(\mathcal{X})$. For well-conditioned problems with equal channel lengths L, we expect $d_{\mathcal{X}} = d(L+m-1)$. If $\mathcal{X}$ is corrupted by noise, then the numerical rank $\hat{d}_{\mathcal{X}}$ of $\mathcal{X}$ is estimated by deciding how many singular values of $\mathcal{X}$ are above the noise level. The estimated row span $\hat{V}$ is given by the first $\hat{d}_{\mathcal{X}}$ rows of V.

If the noise on $\mathcal{X}$ is white and i.i.d., with covariance matrix $\sigma^2 I$, then $\hat{U}$ asymptotically converges to a basis for the column span of the noise-free data $\mathcal{H}\mathcal{S}$. For i.i.d. signals, $\frac{1}{N}\mathcal{S}\mathcal{S}^* \rightarrow I$ (or a multiple thereof) so that $\frac{1}{N}\hat{\Sigma}^2$ asymptotically converges to $\Sigma_{\mathcal{H}}^2 + \sigma^2 I$, where $\Sigma_{\mathcal{H}}$ contains the singular values of $\mathcal{H}$. Note that $\hat{V}$ does not converge to its noise-free value since its dimension grows along with N. However, we can write $\hat{V} = [\hat{\Sigma}^{-1}\hat{U}^*]\mathcal{X}$ so that each column of $\hat{V}$ is determined by a linear combination of the corresponding column of $\mathcal{X}$. Because of the Hankel structure of $\mathcal{X}$, this column contains samples from m consecutive symbol periods. Hence, the matrix multiplication can be viewed as an FIR filter, where m is the equalizer length, and the rows of $\hat{V}$ can be viewed as a new, filtered data set. This is depicted in Fig. 2, where we have just covered the first stage (of three). The filter coefficients are given by the entries of $[\hat{\Sigma}^{-1}\hat{U}^*]$. The main purpose of such a subspace filter is dimensionality reduction, although we will use the orthonormality of $\hat{V}$ as well.

Computing the SVD of an $mMP \times (N-m)$ matrix requires about $3(mMP)^2(N-m) + 10(mMP)^3$ operations [29]. It is possible to replace the SVD by computationally more efficient adaptive subspace tracking algorithms, which update the filter coefficients as increasingly more columns of $\mathcal{X}$ are taken into

account. Several updating algorithms are available; see, for example, [30]–[32].

B. Step 2: Forcing the Toeplitz Property of $\mathcal{S}$

The next step in computing the structured factorization $\mathcal{X} = \mathcal{H}\mathcal{S}$ is to find a description of all possible matrices $\mathcal{S} = \mathcal{S}_{L+m-1}$ that have a block-Toeplitz structure with $L+m-1$ block rows and are such that $\text{row}(\mathcal{S}) = \text{row}(\mathcal{X})$. The latter condition can be true only if each row of $\mathcal{S}$ is in the row span of $\mathcal{X}$:

$$\begin{array}{llll} [\mathbf{s}_{m-1} & \mathbf{s}_m & \cdots & \mathbf{s}_{N-1}] \in \text{row}(\mathcal{X}) \\ [\mathbf{s}_{m-2} & \mathbf{s}_{m-1} & \cdots & \mathbf{s}_{N-2}] \in \text{row}(\mathcal{X}) \\ & & & \vdots \\ [\mathbf{s}_{-L+1} & \mathbf{s}_{-L+2} & \cdots & \mathbf{s}_{N-(L+m-1)}] \in \text{row}(\mathcal{X}). \end{array} \tag{7}$$

These $L+m-1$ conditions can be aligned to apply to a single block vector in several ways. We choose to work with

$$S := \mathcal{S}_1 = [\mathbf{s}_{-L+1} \quad \mathbf{s}_{-L+2} \quad \cdots \quad \mathbf{s}_{N-1}],$$

which is the generator of the Toeplitz matrix $\mathcal{S}$. Hence, S is in the *intersection* of the row span of $\mathcal{X}$ and shifts of this row span (suitably embedded with zeros). Alternatively, we can say that S is orthogonal to the *union* of the complements of these row spans. This leads to a standard procedure to enforce the Toeplitz property of $\mathcal{S}$ and was originally used in [20] and [1]. We will briefly describe the method for reference and then show how row span intersections are computationally more efficient in producing exactly the same result.

1) Null Space Union: Let G be a matrix whose columns constitute a basis for $\ker(\mathcal{X})$, i.e., G is the complement of $\hat{V}$ and can be determined from the SVD of $\mathcal{X}$. If $\mathcal{H}$ has full column rank, then G has dimensions $(N-m+1)\times(N-m+1-d(L+m-1)) =: m_G \times N_G$. Moreover, $\mathcal{X}G = 0 \Rightarrow \mathcal{S}G = 0$. Using the fact that $\mathcal{S}$ is block-Toeplitz, we obtain

$$\mathcal{S}G = 0 \quad \Leftrightarrow \quad SG_{T(L)} = 0,$$

$$G_{T(\ell)} := \begin{bmatrix} G & & \mathbf{0} \\ & G & \\ & & \ddots \\ & & \ddots \\ \mathbf{0} & & \ddots \\ & & & G \end{bmatrix} : \quad (N+\ell-1) \times N_G(\ell+m-1). \tag{8}$$

The number of block columns of $G_{T(\ell)}$ is equal to $\ell+m-1$, where ℓ is a parameter chosen equal to the channel length L (or maybe smaller, as we will propose later). The blocks are each shifted down over one position.

If $G_{T(L)}$ is a wide matrix (this gives additional conditions on m and N), then $\ker(G^*_{T(L)})$ determines S, but only up to a left invertible $d \times d$ matrix A, because $Y = AS$ also satisfies $YG_{T(L)} = 0$. Given $G_{T(L)}$, we take Y to be a matrix whose rows form a basis for $\ker(G^*_{T(L)})$. Hence, the Toeplitz matrix $\mathcal{S}$ is determined uniquely up to multiplication at the right by $D = \text{diag}[T, \cdots, T]$. Now, to identify S, we have to find the factorization $Y = AS$, which, in the case of finite alphabet signals, can be done using a suitable I-MIMO signal separation algorithm, as outlined in Section III-C.

The computation of $\ker(G^*_{T(L)})$ calls for an SVD of $G_{T(L)}$, which is a matrix with dimensions of order $N \times N(L+m)$. Hence, this approach requires order $N^3(L+m)$ operations, which is not feasible for $N > 50$ or so. It is possible to alleviate the computational requirements as we need only the d basis vectors in the null space, which does not require a full SVD. For example, a "spherical subspace" updating algorithm, if applicable, would yield a complexity of roughly $dN^2(L+m)$.

Row Span Intersections: We again consider (7) and let $\hat{V}$ be a basis for $\text{row}(\mathcal{X})$ as determined in the first step. Define

$$\hat{V}^{(k)} := \begin{bmatrix} \mathbf{0} & \hat{V} & \mathbf{0} \\ I_{k-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & I_{n-k} \end{bmatrix} \tag{9}$$

where we take $n = L+m-1$ for now, although we will consider other values for n later. The conditions in (7) can be realigned into

$$\begin{array}{lll} S \in \text{row}\hat{V}^{(1)}, & \hat{V}^{(1)} = \begin{bmatrix} \hat{V} & \mathbf{0} \\ \mathbf{0} & I_{L+m-2} \end{bmatrix}, \\ S \in \text{row}\hat{V}^{(2)}, & \hat{V}^{(2)} = \begin{bmatrix} \mathbf{0} & \hat{V} & \mathbf{0} \\ 1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & I_{L+m-3} \end{bmatrix}, \\ \quad \vdots \\ S \in \text{row}\hat{V}^{(L+m-1)}, & \hat{V}^{(L+m-1)} = \begin{bmatrix} \mathbf{0} & \hat{V} \\ I_{L+m-2} & \mathbf{0} \end{bmatrix}. \end{array} \tag{10}$$

Indeed, the identity matrices in each $\hat{V}^{(k)}$ reflect the fact that, at that point, there are no range conditions on the corresponding columns of S. Thus, S is in the intersection of the row spans of $\hat{V}^{(1)}$ until $\hat{V}^{(L+m-1)}$, and the problem is one of determining a basis for the intersection of a set of given subspaces. One approach, as we saw in the previous subsection, is to compute the union of the complements of the subspaces and take the complement again. However, it is possible to compute subspace intersections without forming complements. To this end, we use the fact that for *orthonormal* bases $\hat{V}^{(k)}$ in (9), precisely the same subspace intersection is obtained by computing the right singular vectors of a matrix formed by stacking the basis vectors (see Appendix A), i.e., by computing an SVD of

$$\begin{bmatrix} \hat{V}^{(1)} \\ \vdots \\ \hat{V}^{(L+m-1)} \end{bmatrix}. \tag{11}$$

More conveniently, (for n intersections, $n = L + m - 1$), we can compute the SVD of where the n copies of $\hat{V}$ are each

$$V_{T(n)} := \begin{bmatrix} \hat{V} & & \mathbf{0} \\ & \hat{V} & \\ \mathbf{0} & \ddots & \ddots \\ & \ddots & \hat{V} \\ J_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & J_2 \end{bmatrix} \Big\} n \tag{12}$$

shifted over one entry and

$$J_1 = \begin{bmatrix} \sqrt{n-1} & & \mathbf{0} \\ & \ddots & \\ & \sqrt{2} & \\ \mathbf{0} & & 1 \end{bmatrix},$$

$$J_2 = \begin{bmatrix} 1 & & \mathbf{0} \\ & \sqrt{2} & \\ & \ddots & \\ \mathbf{0} & & \sqrt{n-1} \end{bmatrix}.$$

The matrices J_1, J_2 summarize the identity matrices present in $\{\hat{V}^{(k)}\}$, which is possible because we are only interested in the singular values and right singular vectors of $V_{T(n)}$, and these do not change by replacing the stack of identity matrices by J_1, J_2. This is immediately seen by looking at $V^*_{T(n)}V_{T(n)}$ and observing that it is the same as the square of (11).

The estimated basis for the intersection Y is given by the right singular vectors of $V_{T(n)}$ that correspond to the *largest* singular values of $V_{T(n)}$:by Appendix A, those that are equal to $\sqrt{n}$ (if there is no noise). As we will motivate later in Section IV-D, the next largest singular values are close to $\sqrt{n-1}$. Thus, the ISI filtering process is based on distinguishing singular values between $\sqrt{n}$ and $\sqrt{n-1}$. It is clear that for large n, this becomes a delicate matter. This motivates us to keep $n = L + m - 1$ small, i.e., not to make the stacking parameter m larger than necessary.

The relation between V_T and G_T in (8) is (cf. Appendix A)

$$V^*_{T(n)}V_{T(n)} + G_{T(\ell)}G^*_{T(\ell)} = nI \qquad (n = \ell + m - 1). \tag{13}$$

Hence, the right singular vectors of $V_{T(\ell+m-1)}$ and $G^*_{T(\ell)}$ are pairwise identical, except for a reversal in ordering. In addition, their squared singular values pairwise add up to n. This is independent of any noise influence and is entirely caused by the fact that we took $\hat{V}$ and G^* to be *orthonormal* bases of complementary subspaces. Hence, the null span union method is just as delicate: The two methods give exactly the same results and have the same robustness and sensitivity to noise.

We let $\hat{d}_S$ denote the dimension of Y, which is the estimated basis of the intersection. Under the (noise-free) conditions specified in Section IV-A, with $n = L + m - 1$ intersections, we obtain $\hat{d}_S = d$, i.e., the intersecting subspace is precisely d-dimensional. This implies that $S = TY$ for some invertible $d \times d$ matrix T.

With noise, we follow the same procedure. We would like to solve

$$\mathcal{S} = \arg \min_{\mathcal{S}:\text{block Toeplitz}} \text{dist}^2(\text{row}(\mathcal{S}), \text{row}(\mathcal{X}))$$

where "dist" denotes the distance between two subspaces [29]. Directly finding a solution to this optimization problem is not feasible. Instead, we find a generator S for the Toeplitz matrix by solving

$$S = \arg \min_{S} \sum_{k=1}^{L+m-1} \text{dist}^2(\text{row}(S), \text{row}(\hat{V}^{(k)})) \tag{14}$$

which determines a matrix S such that the row span of each segment of S is as close to the row span of $\hat{V}$ as possible. The two optimization problems are not precisely the same. The solution of (14) is given in terms of the SVD of $V_{T(n)}$ and is equal to the right singular vectors corresponding to the largest $\hat{d}_S$ singular values of this matrix: those that are close to $\sqrt{n}$ and larger than $\sqrt{n-1}$. Thus, the proposed intersection algorithm solves the second optimization problem.

In Fig. 2, the second stage indicates how $V_{T(n)}$ is formed from $\hat{V}$ and $n-1$ delays of $\hat{V}$ and that the basis Y is obtained by linear combinations of the rows of $V_{T(n)}$. The coefficients of the filter are obtained from the U- and Σ-matrix of the SVD of V_T, which is similar to the first stage. The matrices J_1, J_2 are ignored in the figure as they only play a role in the first and last few columns of a block of data and not during the filtering process itself.

If we take $n = L + m - 1$, then $V_{T(n)}$ has dimensions $(d(L+m-1)^2+2(L+m-2)) \times (N+L-1)$. Using an SVD, this gives the subspace intersection algorithm a complexity of $\mathcal{O}(d^2(L+m)^4N)$, which is linear in N. Similar to stage 1, we can consider an updating implementation of this stage as well, although the required orthonormality of the input signals to this stage gives rise to some interesting complications. A spherical subspace tracker, if applicable, would yield a complexity of order $d^2(L+m)^2N$. An investigation of the details is beyond the scope of the paper.

If the complexity of the intersection step is too large, it may be interesting to consider a multistage intersection approach. Instead of computing the joint intersection of $L + m - 1$ subspaces, which requires a stack of $n = L + m - 1$ shifts of $\hat{V}$, we may place, e.g., two intersection stages in cascade, each consisting of a joint intersection of $n = \frac{1}{2}(L + m - 1)$ subspaces. Likewise, it is ultimately possible to have $L+m-1$ stages, each consisting of one pairwise intersection. This reduces the complexity to $\mathcal{O}(d^2(L + m)^3N)$, and without noise, the result is precisely the same independent of the scheme. Numerically and with noise, however, the result is suboptimal because it is sensitive to the order in which the intersections are performed.

C. Step 3: Forcing the Finite Alphabet Property

At this point, we have only obtained a *basis* Y of a $\hat{d}_S$-dimensional subspace that contains $S = [\mathbf{s}_{-L+1} \cdots \mathbf{s}_{N-1}]$.

To find S, we have to determine which linear combinations of the rows of Y give a finite alphabet structure. This problem is a structured factorization of the form $Y = AS,\ S_{ij} \in \Omega$, which is interpreted as separating an instantaneous linear mixture of finite alphabet signals. Several algorithms have been proposed to solve such problems. In particular, a maximum-likelihood (ML) formulation of the problem leads to

$$\min_{A,S;S_{ij}\in\Omega} \| Y - AS \| \tag{15}$$

which is precisely the problem studied in [16] and [17]. In that paper, two iterative block algorithms are introduced—ILSE and ILSP—which are summarized below. Starting from an initial estimate $A^{(0)}$, the algorithms proceed as follows:

ILSE
for $k = 1, 2, \cdots$
a) $\mathbf{s}_i^{(k)} = \arg\min_{\mathbf{s}_i\in\Omega} \|\mathbf{y}_i - A^{(k)}\mathbf{s}_i\|, \quad \forall i$
b) $A^{(k+1)} = YS^{(k)\dagger}$

ILSP
for $k = 1, 2, \cdots$
a) $S^{(k)} = \text{Proj}_\Omega\left[A^{(k)\dagger}Y\right]$
b) $A^{(k+1)} = YS^{(k)\dagger}$.

The operator Proj_Ω denotes element-wise projection on to the alphabet Ω. The ILSE algorithm converges to the ML estimate of the pair (A, S), provided the initial estimate for A is close to the true value. Solving step a in ILSE involves enumeration over all possible combinations of symbols. The ILSP algorithm avoids the enumeration by replacing it with a least-squares solve for S followed by a projection onto the alphabet. This is computationally cheaper but suboptimal. Unless the alphabet is BPSK and the number of rows of Y is small, it is important to have a reasonably accurate initial estimate of A. Good initial points are obtained by the recently introduced "analytical constant modulus algorithm" (ACMA) [33], which is readily specialized to give closed-form eigenvalue-based solutions for simple finite alphabets such as BPSK and MSK [34]. Depending on the matrix dimensions and noise level, ILSE and ILSP usually converge to a fixed point in less than 5–10 iterations [16]. As mentioned in the introduction, several other I-MIMO source separation algorithms are described in the literature, some based on different properties such as source independence or constant modulus [18], [19], [35]–[38].

Alternatively, we can minimize the MMSE criterion

$$\min_{T \text{ full rank}, S\in\Omega} \| S - TY \|^2 \tag{16}$$

which essentially fits the subspace Y to a FA matrix S. An iterative algorithm to solve this problem is called iterative least squares with subspace fitting (ILSF) and is listed in Table I. It is very similar to ILSP but has the advantage that the pseudoinverse of the kth iterate $S^{(k)}$ is avoided and replaced by a pseudoinverse of Y, which is constant. Since Y is an orthonormal basis, this inverse is simply equal to the complex conjugate transpose $Y^\dagger = Y^*$. For a small number of sources, each iteration requires $2Nd^2$ flops.

One aspect of the problem (16) that is different from (6) is that we explicitly require T to be full rank in order to guarantee independent rows of S. Indeed, T should be a well-conditioned

TABLE I
ILSF ALGORITHM

In: Y, **out:** S s.t. $S \in \Omega$, $S \approx TY$
Choose $T^{(0)}$
for $k = 1, 2, \cdots$
a. $S^{(k)} = \text{proj}_\Omega[T^{(k-1)}Y]$
b. $T^{(k)} = S^{(k)}Y^\dagger$
c. Ensure $T^{(k)}$ is full rank (see text)
until $T^{(k)} - T^{(k-1)} = 0$.

matrix since the rows of S become orthogonal for large N (as the signals are uncorrelated). In addition, since Y is orthogonal as well, T is close to unitary (up to a scaling). It rotates one orthogonal basis into another. Hence, T is the solution to an *orthogonal Procrustes* problem [29]. Forcing T to be close to unitary provides one way of enforcing independent signals in S. Note that for a unitary matrix T, the criteria (16) and (15) are the same; therefore, the performance of ILSF is quite similar to ILSP.

The ILSF step is the last stage of the filter in Fig. 2, with $p = 1$ ($p > 1$ is considered in Section V-C). The coefficients of the filter in this stage are the entries of T. Similar to ILSP in [17], it is straightforward to replace ILSF by an updating version, which operates in a decision directed feedback mode.

D. Alternative: Computation of $\mathcal{H}$ First

Instead of estimating $\mathcal{S}$ directly, we can also first estimate $\mathcal{H}$ and invert the resulting channel to estimate S. This is potentially interesting since the dimensions of $\mathcal{H}$ do not grow with N; therefore, it can be estimated consistently. We briefly describe the procedure, which is basically an extension of [7] to multiple signals.

Let G' be a basis of the left null space of $\mathcal{X}_m$. Assuming $\mathcal{S}$ to be of full rank, we have $G'\mathcal{X} = 0 \ \Rightarrow\ G'\mathcal{H}_m = 0$. Write

$$\begin{array}{lll} H &=: & [H_0\ \cdots\ H_{L-1}], \quad H_i : MP \times d \\ G' &=: & [G'_1\ \cdots\ G'_m], \quad G'_i : (mMP - d_{\mathcal{X}}) \times MP. \end{array}$$

Then $G'\mathcal{H}_m = 0 \quad \Leftrightarrow$

$$\begin{bmatrix} G'_m & & \mathbf{0} \\ \vdots & \ddots & \\ \vdots & & G'_m \\ G'_1 & & \vdots \\ & \ddots & \vdots \\ \mathbf{0} & & G'_1 \end{bmatrix} \begin{bmatrix} H_0 \\ \vdots \\ H_{L-1} \end{bmatrix} = 0.$$

If the matrix on the left is "tall" (this gives minimal conditions on m), then generically its right null space specifies H up to a right block-diagonal factor $\text{diag}[A, \cdots, A]$. For any solution H, the basis $Y = [Y_{-L+1} \cdots Y_{N-1}]$ is found from an inverse

filter associated with H as

$$\begin{bmatrix} Y_{N-1} \\ \vdots \\ Y_{-L+1} \end{bmatrix} = \mathcal{H}_N^{\dagger}\mathbf{x}$$

where $\mathbf{x} = \text{vec}(X) \equiv \mathcal{X}_N$ is a stack of all input data. At this point, we are back at the model $Y = AS$, and the ILSE/P/F algorithm is employed to remove the ambiguity that A represents.

For the estimation of $\mathcal{H}$, it is only required that $\mathcal{S}$ be of full row rank, which is a mild condition. In particular, it is not necessary that all channels have equal length, although certain modifications are in order (see [22], which also contains some identifiability results).

It is unclear whether a direct estimation of S is to be preferred over an indirect estimation via $\mathcal{H}$. The former initially forces only the structure of $\mathcal{S}$, neglecting that of $\mathcal{H}$, whereas the latter does the opposite. In general, estimating $\mathcal{H}$ is computationally easier for large N and can be done consistently. Our experience with simulations, however, is that estimating S directly might be more accurate in the presence of model mismatch (see Section IV-E). In addition, if the channel lengths are not well defined (i.e., the FIR assumption is only approximately true), row span methods can potentially obtain a better model fit. This is because they do not force zeros in the lower right block of $\mathcal{H}$ but have the freedom to insert the actual (nonzero) coefficients instead. Finally, without going into details, we mention that the row span methods are almost immediately applicable to more general ARMA (rational) channel models, in which a state space model is assumed.

IV. Aspects of the Algorithm

A. Identifiability

Does the intersection/FA algorithm provide a unique estimate of $\mathcal{S}$? This identifiability issue is the subject of the following theorem. Similar results for $d = 1$ were presented in [7] but from the point of view of estimating $\mathcal{H}$ from its Hankel structure. An alternative proof appears in [26].

Theorem 1: Consider the FIR-MIMO scenario with d sources and channels of equal length $L_j = L$. Suppose that the dimension conditions (5) are satisfied for some m and that the $\text{rank}(\mathcal{X}_m) = d(L+m-1)$ and $\text{rank}(\mathcal{X}_{m+1}) = d(L+m)$.

Let $\mathcal{X}_m = \mathcal{H}_m\mathcal{S}_{L+m-1}$ be a structured factorization of $\mathcal{X}_m$. Taking only the Toeplitz structure into account, $\mathcal{S}_{L+m-1}$ is uniquely specified by the condition $\text{row}(\mathcal{X}_m) = \text{row}(\mathcal{S}_{L+m-1})$ up to a left block-diagonal factor $D = \text{diag}[T, \cdots, T]$, where T is an invertible $d \times d$ matrix.

Taking also the FA property into account, under conditions of [17, Theorem 3.2],[2] $\mathcal{S}_{L+m-1}$ is unique up to $D = \text{diag}[T, \cdots, T]$, where T can take the form of a permutation and a diagonal scaling by $\pm 1, \pm j$.

We first derive the following lemma, where n can be any number of intersections between 1 and $L+m-1$.

[2]This theorem basically requires that S contains all possible d-dimensional columns that can be generated by the finite alphabet. This is a sufficient but pessimistically large condition on N.

Lemma 1: For $1 \le n \le L+m-1$, let $\hat{V}$ be an orthonormal basis of $\text{row}(\mathcal{X}_m)$, and define $\hat{V}^{(k)}$ as in (9). Under the conditions of Theorem 1, $\text{row}(\hat{V}^{(1)}) \cap \cdots \cap \text{row}(\hat{V}^{(n)})$ is a subspace of dimension $d(L+m-n)$ and contains $\mathcal{S}_{L+m-n}$ $(1 \le n \le L+m-1)$.

Proof of the Lemma: The rank condition on $\mathcal{X}_{m+1}$ implies that $\mathcal{S}_{L+m}$ has full row rank. In turn, this implies that $\mathcal{S}_k$ has full row rank equal to dk for $1 \le k \le L+m$ (since any subset of the rows of $\mathcal{S}_{L+m}$ has full row rank as well).

Suppose $n = 2$. In investigating $\text{row}(\hat{V}^{(1)}) \cap \text{row}(\hat{V}^{(2)})$, we may as well look at $\mathcal{S}_{L+m-1}$ instead of $\hat{V}$ since they span the same space. Consider

$$\begin{bmatrix} \mathcal{S}_{L+m-1} & * \\ * & \mathcal{S}_{L+m-1} \end{bmatrix} =$$

$$\begin{bmatrix}
\mathbf{s}_{m-1} & \mathbf{s}_m & \ddots & \mathbf{s}_{N-2} & \mathbf{s}_{N-1} & * \\
\mathbf{s}_{m-2} & \mathbf{s}_{m-1} & \ddots & \mathbf{s}_{N-3} & \mathbf{s}_{N-2} & * \\
\ddots & \ddots & \ddots & \ddots & \ddots & * \\
\mathbf{s}_{-L+2} & \mathbf{s}_{-L+3} & \ddots & \ddots & \ddots & * \\
\mathbf{s}_{-L+1} & \mathbf{s}_{-L+2} & \ddots & \ddots & \mathbf{s}_{N-L-m+1} & * \\
* & \mathbf{s}_{m-1} & \ddots & \ddots & \mathbf{s}_{N-2} & \mathbf{s}_{N-1} \\
* & \ddots & \ddots & \ddots & \ddots & \ddots \\
* & \mathbf{s}_{-L+2} & \mathbf{s}_{-L+3} & \ddots & \mathbf{s}_{N-L-m+1} & \mathbf{s}_{N-L-m+2} \\
* & \mathbf{s}_{-L+1} & \mathbf{s}_{-L+2} & \ddots & \mathbf{s}_{N-L-m} & \mathbf{s}_{N-L-m+1}
\end{bmatrix}$$

where "$*$" stands for an arbitrary extension as enabled by the identity matrices in the $\{\hat{V}^{(k)}\}$. The intersection removes all rows that are not linearly dependent on the rows of the opposite block. With suitable extensions, this means only the first and last rows are candidates for removal. Note that they cannot be linearly dependent on the other rows because the submatrix of the above matrix obtained by removing the first and last column has the same set of rows as $\mathcal{S}_{L+m}$, which has full row rank. Hence, both rows are removed, and the result of the intersection is a space with precisely d less rows and is generated by the rows of $\mathcal{S}_{L+m-2}$ (since it is of full row rank). The result for larger n is obtained by repeating the same argument.[3]

Proof of the Theorem: Setting $n = L+m-1$ in the above lemma gives an intersection subspace of dimension d, which is spanned by the d rows of $S = \mathcal{S}_1$. Hence, $\mathcal{S}_1$ is unique up to left multiplication by some invertible $d \times d$ matrix T; consequently, $\mathcal{S}_{L+m-1}$ is unique up to left multiplication by $D = \text{diag}[T, \cdots, T]$.

Taking the FA property into account as well, [17, theorem 3.2] claims that for sufficiently diverse symbols, $\mathcal{S}_1$ is unique up to permutation and scaling by $\pm 1, \pm j$ □.

[3]The rank condition on $\mathcal{X}_{m+1}$ is necessary to avoid pathological cases: Consider, e.g., a periodic symbol matrix

$$\mathcal{S}_3 = \begin{bmatrix} \mathbf{s}_2 & \mathbf{s}_0 & \mathbf{s}_1 & \mathbf{s}_2 & \mathbf{s}_0 & \mathbf{s}_1 \\ \mathbf{s}_1 & \mathbf{s}_2 & \mathbf{s}_0 & \mathbf{s}_1 & \mathbf{s}_2 & \mathbf{s}_0 \\ \mathbf{s}_0 & \mathbf{s}_1 & \mathbf{s}_2 & \mathbf{s}_0 & \mathbf{s}_1 & \mathbf{s}_2 \end{bmatrix}.$$

In this case the intersections do not remove any row. Note that $\mathcal{S}_4$ has rank $\le 3d$; therefore, it is not of full rank.

B. Detection of d and L

If $\mathcal{H}$ and $\mathcal{S}$ have full column rank and row rank, respectively, then the rank of $\mathcal{X}_m$ is $d_\mathcal{X} = d(L+m-1)$. The number of signals d can be estimated by increasing m by one and looking at the increase in rank of $\mathcal{X}_m$. This property provides a very effective detection mechanism even if the noise level is quite high since it is independent of the actual (observable) channel length. Furthermore, it still holds if all channels do not have equal lengths (see Section IV-C). In case they do, then L can be determined from the estimated rank of $\mathcal{X}$, $\hat{d}_\mathcal{X}$ and the estimated number of signals $\hat{d}$ by $\hat{L} = \hat{d}_\mathcal{X}/\hat{d} - m + 1$.

It is interesting to note that there is an efficient updating algorithm for estimating the rank of $\mathcal{X}_k$ and the corresponding column span for all k from 1 to m at once, without requiring SVD's and using only the full-size $\mathcal{X}_m$. The SSE-1 subspace estimator derived in [32] is a technique for computing the number of singular values of a matrix X that are larger than a given threshold γ, and a basis for a subspace that is γ-close to the column span of the matrix in some norm. The algorithm is such that at the same time, this information is produced on all principal submatrices of X as well. Applied to $\mathcal{X}_m$, it produces the ranks of all $\mathcal{X}_k$, $k \le m$, with respect to a given threshold at the complexity of a QR factorization.

C. Unequal Channel Lengths

For simplicity of presentation, we have only considered channels with equal length up to now: $L_j = L, j = 1, \cdots, d$. In general, however, the lengths L_j may be different. In that case, it is perhaps more natural to write the factorization $\mathcal{X}_m = \mathcal{H}_m \mathcal{S}_{m+L-1}$ with a rank-deficient $\mathcal{H}_m$ as

$$\mathcal{X}_m = \mathcal{H}_m^{(1)} \mathcal{S}_{m+L_1-1}^{(1)} + \cdots + \mathcal{H}_m^{(d)} \mathcal{S}_{m+L_d-1}^{(d)} \qquad (17)$$

where each $\mathcal{H}_m^{(k)}$ and $\mathcal{S}^{(k)}$ correspond to the channel and symbol matrix of source k only. Generically, these factors are of full rank $L_j + m - 1$. The rank of $\mathcal{X}_m$ is thus expected to be

$$\operatorname{rank}(\mathcal{X}_m) = L_{\text{tot}} + d(m-1), \qquad L_{\text{tot}} = \sum_1^d L_j$$

assuming the terms in (17) are linearly independent. To obtain $\operatorname{row}(\mathcal{X}_m)$ equal to the linear envelope $\operatorname{row}(\mathcal{S}^{(1)}) \dotplus \cdots \dotplus \operatorname{row}(\mathcal{S}^{(d)})$, it is necessary that $mMP \ge L_{\text{tot}} + d(m-1)$, i.e.,

$$MP > d, \qquad m \ge \frac{L_{\text{tot}} - d}{MP - d}.$$

To describe the result of the subspace intersections, we need to define a "rank profile"[4]

$$r(\ell) := \#\{j : L_j \ge \ell\}$$

i.e., $r(\ell)$ is equal to the number of sources with a channel length $L_j \ge \ell$. Thus, $r(\ell)$ is monotonically decreasing from d ($\ell = 1$) to 0 ($\ell > L$), and $\sum_1^L r(\ell) = L_{\text{tot}}$.

If we perform intersections step by step, then the first intersection removes the top row of every $\mathcal{S}_{m+L_k-1}^{(k)}$, and the rank of the intersecting subspace is d less than the rank of $\mathcal{X}_m$. The next intersection removes the second row and drops the rank again by d, etc. This continues until the rank of one or more of $\mathcal{S}^{(k)}$ is exhausted, in which case, the drop in rank per intersection is now less. The latter starts to happen once more than m intersections are taken, and the drop in rank follows the rank profile $r(\ell)$. In general, after $n \ge m$ intersections, the rank of the resulting intersecting subspace is

$$\begin{aligned} &\operatorname{rank}\left[\cap_1^n \operatorname{row}(\hat{V}^{(k)})\right] \\ &\quad = \left[L_{\text{tot}} + d(m-1)\right] - \left[d(m-1) + \textstyle\sum_1^{n-m} r(\ell)\right] \\ &\quad = L_{\text{tot}} - \textstyle\sum_1^{n-m} r(\ell). \end{aligned}$$

In principle, this allows one to determine the rank profile $r(\ell)$ and, hence, the individual channel lengths.

In an approach outlined by Liu and Xu in [21], a technique for estimating source signals with unequal channel lengths is presented. Essentially, the idea is to compute all intersecting subspaces for $n = m$ to $n = m + L - 1$. Starting with the smallest dimensional subspace (i.e., $n = m + L - 1$), first, all the multiple signals that are in this subspace are separated (by ILSF), which are precisely the $r(L)$ signals with channel length $L_j = L$. With these signals known, the next higher dimensional intersection (smaller n) is computed, and the signals in it are separated, using the signals that were already found (and their shifts) as partial initial conditions for T in the ILSF algorithm. In this way, it is in theory possible to unwind the separation problem.

If, instead of the SVD, we apply the SSE-1 subspace estimator [32] to the full-size $V_{T(L+m-1)}$, we obtain rank and subspace information of all principal submatrices of this matrix as well. Since these principal submatrices are equal to the smaller size $V_{T(n)}$, $n = 1, \cdots, L+m-1$ (ignoring the effect of $J_{1,2}$), this gives sufficient information to find the complete rank profile at once, as well as a way to reconstruct all intersections.

D. Singular Value Model of Intersections

Under noise-free conditions, we already know (by Appendix A) that the largest d singular values of $V_{T(L+m-1)}$ are precisely equal to $\sqrt{L+m-1}$. What is the magnitude of other singular values? It is straightforward to give an answer for $N \to \infty$.

Since $\hat{V}$ is a basis of $\operatorname{row}(\mathcal{S})$, we have $\hat{V} = Q\mathcal{S}$ for some square matrix Q. Hence $V_{T(n)}$ can be factored as (18), shown at the bottom of the next page, where "$*$" denotes entries that are not of interest. For large N and i.i.d. signals, the rows of $\mathcal{S}$ are approximately orthogonal to each other, that is, $\frac{1}{N}\mathcal{S}\mathcal{S}^* \to I$, which implies that $\sqrt{N}Q$ is close to a unitary matrix. In that case, it follows that the columns of $Q_{T(n)}$ are asymptotically orthogonal to each other. Ignoring the second term in the factorization (18) for the moment, the factorization of the first term directly translates into the SVD of V_T. In particular, the singular values of V_T are the norms of the columns of $\sqrt{N}Q_T$ and, thus, are equal to $1, \cdots, \sqrt{n-1}, \sqrt{n}, \sqrt{n-1}, \cdots, 1$ each repeated d times. The left singular vectors are just normalizations of the columns of Q_T, and the right singular vectors are normalizations of

[4] For a set E, $\#(E)$ denotes the number of elements in E.

the rows of $\mathcal{S}_{ext}$. The latter normalization is in the order of $\frac{1}{\sqrt{N}}$. For large N, it is clear that rows of the second term (containing $J_{1,2}$) become orthogonal to $\frac{1}{\sqrt{N}}\mathcal{S}_{ext}$ since the inner product is proportional to $\frac{1}{\sqrt{N}}$. Obviously, the columns of this term are orthogonal to Q_T. Hence, the second term contributes additional singular values $1, \cdots, \sqrt{n-1}$ each repeated two times. Altogether, asymptotically and under noise-free conditions, $V_{T(n)}$ for $n = L+m-1$ has d singular values equal to $\sqrt{n}$ and groups of $2d+2$ singular values equal to $\sqrt{n-1}, \cdots, 1$. If we take $n < L+m-1$, then similarly, we can show that there are $d_S := d(L+m-1) - d(n-1)$ singular values equal to $\sqrt{n}$, followed by the groups of $2d+2$ singular values equal to $\sqrt{n-1}, \cdots, 1$. The right singular vectors corresponding to the d_S largest singular values are a basis for $\mathcal{S}_{L+m-1-(n-1)}$; the n intersections have removed $n-1$ echos of each signal.

If N is not large and if there is noise, then obviously, the singular values start to deviate from these asymptotic values, and in particular, the gap between the singular values around $\sqrt{n}$ and $\sqrt{n-1}$ closes. The assessment of these deviations is subject to future research. Such an analysis would give pointers to suitable minimal values for N (in relation to the noise level) such that there still can be a gap.

E. Comparison by Simulation

To assess and compare the performance of the proposed algorithms, we consider a simple but unrealistic scenario in which all assumptions on the model are satisfied. A more challenging test case is deferred to Section VI. We took $d = 2$ real-valued BPSK sources and a randomly selected complex channel matrix with $MP = 4$ observables and equal channel lengths $L = 3$. (M and P are equivalent in this example because there is no modulation function and no multiray model.) We added complex white Gaussian i.i.d. noise with variance σ^2. The number of snapshots was $N = 50$. The signal-to-noise ratio (SNR) is defined as $\|H\mathcal{S}_L\|_F^2/(dMPN\sigma^2)$, which is the average SNR per signal per observable. The relative power of both sources was set equal.

The singular values of $\mathcal{X}_m$ are displayed in Fig. 3. Without noise, the rank of $\mathcal{X}_m$ is expected to be $d_{\mathcal{X}} = d(L+m-1)$, which turns out to be the case. The number of sources d can be identified from the graph by looking at the increase in rank of $\mathcal{X}_m$ as m increases. In addition, L can be estimated, assuming

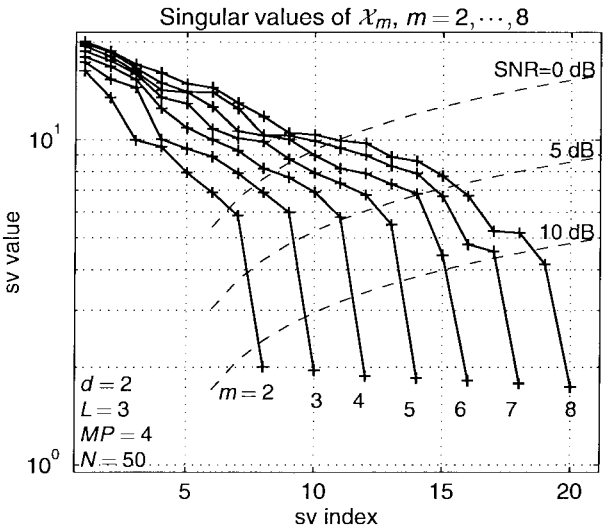
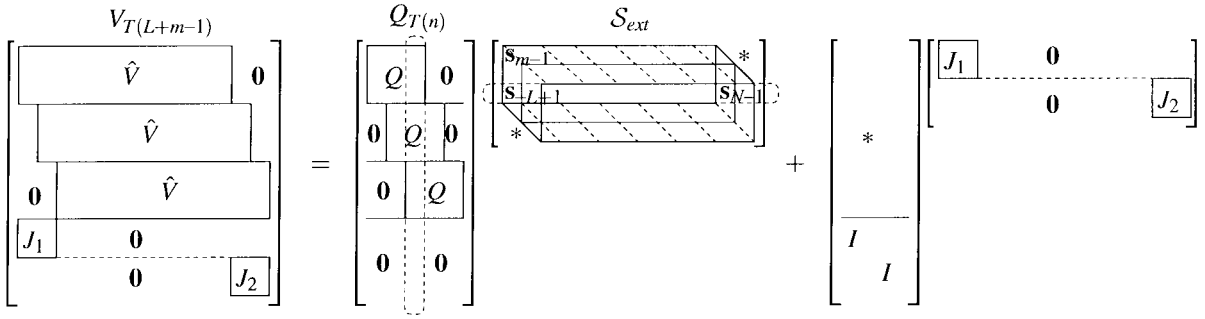


Fig. 3. Singular values of $\mathcal{X}_m$ (noise-free) for a range of m. The dashed lines indicate which singular values will be masked by the noise.

equal channel lengths. The dashed lines in the graph indicate at which noise level the small singular values of $\mathcal{X}_m$ will be obscured. This level increases with $\sqrt{m}$, but the singular values of $\mathcal{X}_m$ do not; therefore, it is advantageous to keep m small. Below 10 dB, the true rank of $\mathcal{X}_m$ is no longer visible, and in practice, we would estimate the rank of $\mathcal{X}_m$ too low.

The singular values of $V_{T(n)}$ are considered next (Fig. 4). For convenience, they are converted into the singular values of G_T by computing $(nI - \text{sv}(V_{T(n)})^2)^{1/2}$. (Recall from (13) that the singular values of V_T and G_T squared, add up to n by construction.) After transformation, we expect for full intersections ($n = L+m-1$) a total of $d = 2$ zero singular values corresponding to the sources and groups of $2d+2 = 6$ singular values around $1, \sqrt{2}, \cdots, \sqrt{n}$ (as indicated by the dotted lines in the figure). For SNR's of 10 dB or more, this is indeed the case, but for SNR $= 5$ dB, the second source is no longer present after intersections (Fig. 4(b)). If we take $m = 2$ and truncate the rank of $\mathcal{X}$ at 7, which is its observable rank, then the rank equation produces an estimated channel length $\hat{L} = 3\frac{1}{2}$. Setting $\hat{L} = 2$ instead, we can only take $n = 2$ intersections, and we are left with $\hat{d}_S = \hat{d}_{\mathcal{X}} - d(\hat{L}+m-1)+d = 3$ signals/echos. As seen in Fig. 4(c), this number of remaining signals is still well visible, even in the SNR $= 5$ dB case. This indicates that for the row span intersection method, there is an advantage in underestimating L and $d_{\mathcal{X}}$.



(18)

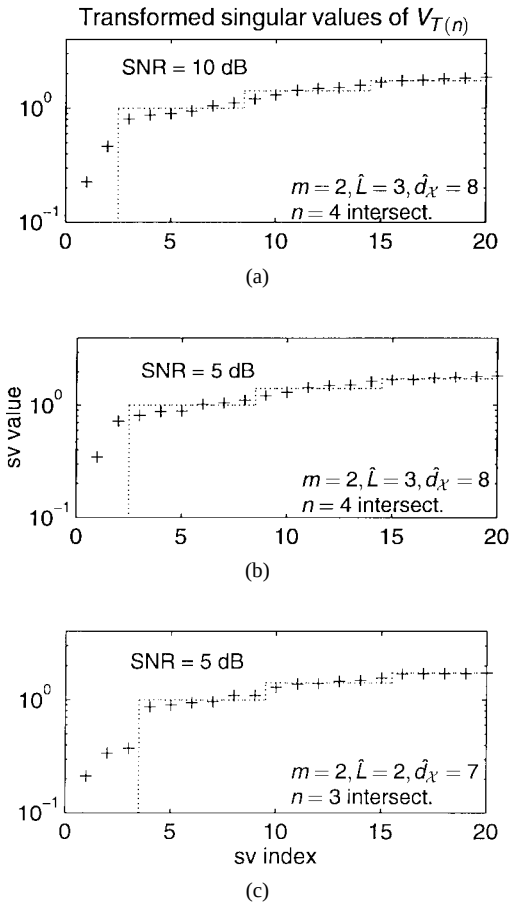


Fig. 4. Transformed singular values of $V_{T(n)}$, namely, $(n - \text{sv}\ (V_{T(n)})^2)^{1/2}$. Small values indicate the number of remaining signals after intersections. (a) SNR = 10 dB, full intersections. (b) SNR = 5 dB, full intersections (second source not resolved). (c) SNR = 5 dB, underestimating $d_{\mathcal{X}}$ and taking less intersections. After intersections, $\hat{d}_S = 3$ signals remain.

This is confirmed by Fig. 5, which shows the bit error rates (BER) for varying SNR for various choices of the parameters $\hat{L}$ and $\hat{d}_{\mathcal{X}}$. Here, we compare the directly estimating-$\mathcal{S}$ method (row span intersection) with the estimating-$\mathcal{H}$-first method (column nullspace union). If the exact parameters are used in the identification, the performance of both methods is approximately the same.

In [26], a Cramer–Rao lower bound (CRLB) is derived for the blind equalization of one source if only the Toeplitz/Hankel structure of $\mathcal{S}$ and $\mathcal{H}$ is taken into account but not the finite-alphabet property. The result is readily generalized to $d > 1$. However, a correction to [26] by about a factor 2 is in order, which is discussed in Appendix B.

As seen in Fig. 5(a), the methods do not reach the approximate blind CRLB (22) because they only force one of the factors to have structure (either $\mathcal{S}$ or $\mathcal{H}$) but not both. For comparison, we also show the CRLB for estimation of S if H is known (the performance for a zero-forcing equalizer), which has a better performance, especially for the second signal. The "ILSE" curves are obtained by running the ILSE algorithm on X, initialized by the exact H so that it gives the ML estimate of the factor $\mathcal{S}_L$ if its Toeplitz structure is ignored. As its BER is well above that of the blind CRLB, this indicates that using the Toeplitz structure is relevant.

Fig. 5(b) shows the case where the rank of $\mathcal{X}$ is underestimated, which would happen in practice below 10 dB. In that case, underestimating L as well (hence, taking less intersections $n = 3$) leads to $\hat{d}_S = 3$ remaining signals after intersections, which are separated by ILSE. As shown by the dotted lines, this greatly improves the performance. We even go below the blind CRLB for the second signal, which is possible because the estimators are not necessarily unbiased and because the FA structure is used more strongly now but is not considered in the bound. This holds for the row span method. Using similar techniques, we were not able to to improve the performance of column span method. Instead, it collapsed on rank-truncated data.

The conclusions of the simulation can be summarized as follows:

- The current and proposed blind equalization/separation methods force only one structural property out of three: the Hankel structure of $\mathcal{H}$, the Toeplitz structure of $\mathcal{S}$, and its finite alphabet structure. For the assumed model, each of these properties by itself is approximately equally strong. As shown by the theoretical bounds, significant gains can still be obtained by simultaneously forcing more than one property.
- The performance of the row span method can be significantly improved by truncating the rank of $\mathcal{X}$ at the noise level, underestimating the channel length L, and separating the remaining signals plus echos based on the finite alphabet property. The column span method is apparently not robust on truncated data.

V. Ill-Defined Channel Lengths

In reality, channels do not have well-defined channel lengths. Multipath echos with a long delay generally have a smaller amplitude; therefore, the channel responses trail down to zero rather than filling out a sharply defined interval in time. In such cases, $\mathcal{H}$ is ill conditioned, and subspace intersections cannot be used to precisely cancel all the echos. Ill-conditioned channel matrices are also expected for bandlimited signals [27].

A. Effect on Intersections

To illustrate the effect of ill-conditioned channels on the computation of the intersecting subspace, consider the impulse response $h(t)$ shown in Fig. 6(a). This is the convolution of an actual line-of-sight indoor channel at 2.4 GHz with a raised cosine pulse ($T = 10$ ns, modulation index $\beta = 0.5$, oversampling rate $P = 5$). The main peak has a width of about two symbols, but there are several smaller peaks as well. For this example, we consider the data obtained from $M = 2$ antennas, with $d = 1$ signal present, which is already sufficient to make our point. The singular values of $\mathcal{H}_{10}$ are shown in Fig. 6(b). The rank of $\mathcal{H}_{10}$ is not clear; it is certainly not of low rank in a mathematical sense, and the numerical rank depends on the truncation level we choose. To avoid an excessively large inverse of $\mathcal{H}$, we would in this case decide a rank of $\hat{d}_{\mathcal{X}} = 12$ or so, corresponding to an estimated channel length of $\hat{L} = 3$.

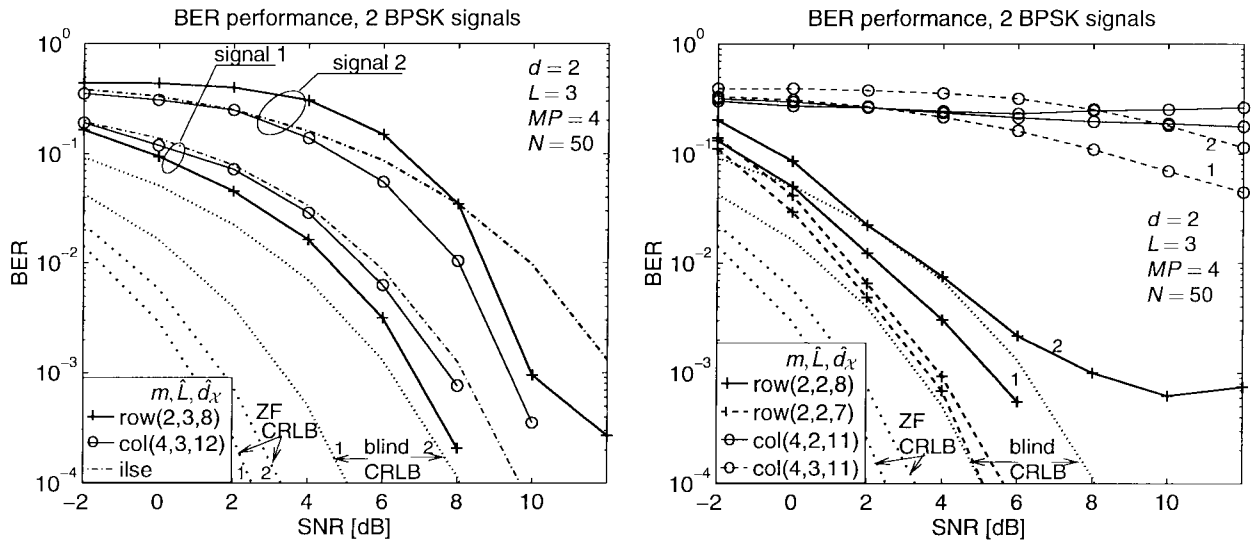


Fig. 5. BER performance for $d = 2$ BPSK signals. (a) Using exact values $\hat{L} = L$, $\hat{d}_{\mathcal{X}} = d_{\mathcal{X}} = d(L + m - 1)$. (b) Using approximate values. For comparison, the CRLB for a zero-forcing equalizer ($\mathcal{H}_N^{\dagger}$) is indicated, assuming perfect knowledge of H, the CRLB for the blind scenario (not using the FA property), and the performance of ILSE initialized with the exact H.

For large N, the signals are approximately orthogonal to their shifts, and in that case, the singular values of $\mathcal{H}$ are equal to the singular values of $\frac{1}{\sqrt{N}}\mathcal{X}$. In fact, let $\mathcal{H} = U_{\mathcal{H}}\Sigma_{\mathcal{H}}Q$ be an SVD of $\mathcal{H}$. Then

$$\mathcal{X} = \mathcal{H}\mathcal{S} = [U_{\mathcal{H}}\Sigma_{\mathcal{H}}Q]\mathcal{S} = U_{\mathcal{X}}\Sigma_{\mathcal{X}}V$$

so that, for orthogonal $\mathcal{S}$, $\Sigma_{\mathcal{X}} = \sqrt{N}\Sigma_{\mathcal{H}}$ and $V = \frac{1}{\sqrt{N}}Q\mathcal{S}$. If we approximate $\mathcal{X}$ by truncating its SVD to some rank $\hat{d}_{\mathcal{X}}$, then

$$\hat{\mathcal{X}} = [\hat{U}_{\mathcal{H}}\hat{\Sigma}_{\mathcal{H}}\hat{Q}]\mathcal{S} = \hat{U}_{\mathcal{X}}\hat{\Sigma}_{\mathcal{X}}\hat{V}.$$

Ideally, $\mathcal{H}$ is full rank, and Q is square, but for ill-defined channel lengths, $\hat{Q}$ has size $\hat{d}_{\mathcal{X}} \times (L+m-1)$, where L is the "actual" (large and fuzzy) channel length. Fig. 6(c) shows the magnitude of the entries of $\hat{Q}$ (up to the first 24 rows of $\hat{Q}$). The first 10 rows of $\hat{Q}$ have 11 large entries; thus, the first 10 rows of $\hat{V}$ are a linear combination of 11 rows of $\mathcal{S}$, plus some weaker ISI from other rows. The reason for this is that $h(t)$ contains a sharp peak, which is smeared by the $m = 10$ shifts over 11 symbols. The next few rows of $\hat{Q}$ show the influence of the smaller peaks in $h(t)$: An increasing number of rows of $\mathcal{S}$ get involved.

For large N, we may write, as in (18),

$$V_{T(n)} \quad \approx \quad Q_{T(n)}\mathcal{S}_{ext}$$

neglecting the edge effects caused by $J_{1,2}$. Since the rows of $\mathcal{S}_{ext}$ are close to orthogonal, the SVD of V_T can be written as the SVD of Q_T times $\mathcal{S}_{ext}$. Fig. 6(d) shows the singular values of $Q_{T(n)}$ when $n = \hat{d}_{\mathcal{X}} = 12$. There is one singular value close to $\sqrt{12}$ and two around $\sqrt{11}, \sqrt{10}, \cdots$, as expected. Hence, there is one vector in the intersection. This vector is given by the product of the corresponding right singular vector of Q_T times $\mathcal{S}_{ext}$. The right singular vectors are shown in Fig. 6(e). Since we expect the result to be basically one symbol sequence out of $\mathcal{S}_{ext}$, the top row should have only one large entry. However, it is seen that the top row has at least eight large entries; therefore, the vector in the intersection is still a linear combination of at least eight symbols. Thus, the intersection did not produce the desired effect of removing all ISI. The "Λ" structure of this figure is very characteristic and shows how the intersections work. Indeed, small singular values of V_T (or Q_T) correspond to the top and bottom rows of $\mathcal{S}_{ext}$ since these are repeated only a few times in V_T. The large singular values correspond to rows in the middle of $\mathcal{S}_{ext}$, which are repeated up to n times. The width of the legs of "Λ" is nearly constant. For well-defined channel lengths, the width of the legs is expected to be 1 because the right singular vectors corresponding to a singular value are specific echos (rows of $\mathcal{S}_{ext}$; cf. (18)). The widening of the second leg of the "Λ" in our example shows the influence of the structured noise that is introduced by truncating the rank of $\mathcal{H}$ at 12. Qualitatively, it can be attributed to the second peak in $h(t)$, which is partly (but not entirely) eliminated by the truncation of $\mathcal{H}$ to rank 12. The truncated data matrix still contains one or a few linear combinations of echos, but since there are fewer combinations than symbols that play a role after truncation, the echos cannot be removed by intersections.

The conclusion drawn from this experiment is that for actual channels the SVD-based intersection scheme may not remove all the ISI if the rank of $\mathcal{H}$ is ill-defined.

B. Effect of Taking Fewer Intersections

What happens if we take less than $L+m-1$ intersections? We provide an intuitive analysis. Let us say that $\hat{d}_{\mathcal{X}}$ is the true rank of $\mathcal{X}$ and that the resulting approximation error is lumped into the noise term. Since $\hat{V} = [\hat{\Sigma}^{-1}\hat{U}^*]\mathcal{X}$, it is seen that the noise on the rows of $\hat{V}$ is not uniform: $\hat{\Sigma}^{-1}$ amplifies the noise at the top rows of $\hat{U}^*\mathcal{X}$ less than at the later rows. Consider a simple example where $d = 1$ and $\hat{Q} = I$. If we take $n < \hat{d}_{\mathcal{X}}$ intersections, then the basis Y of singular vectors of

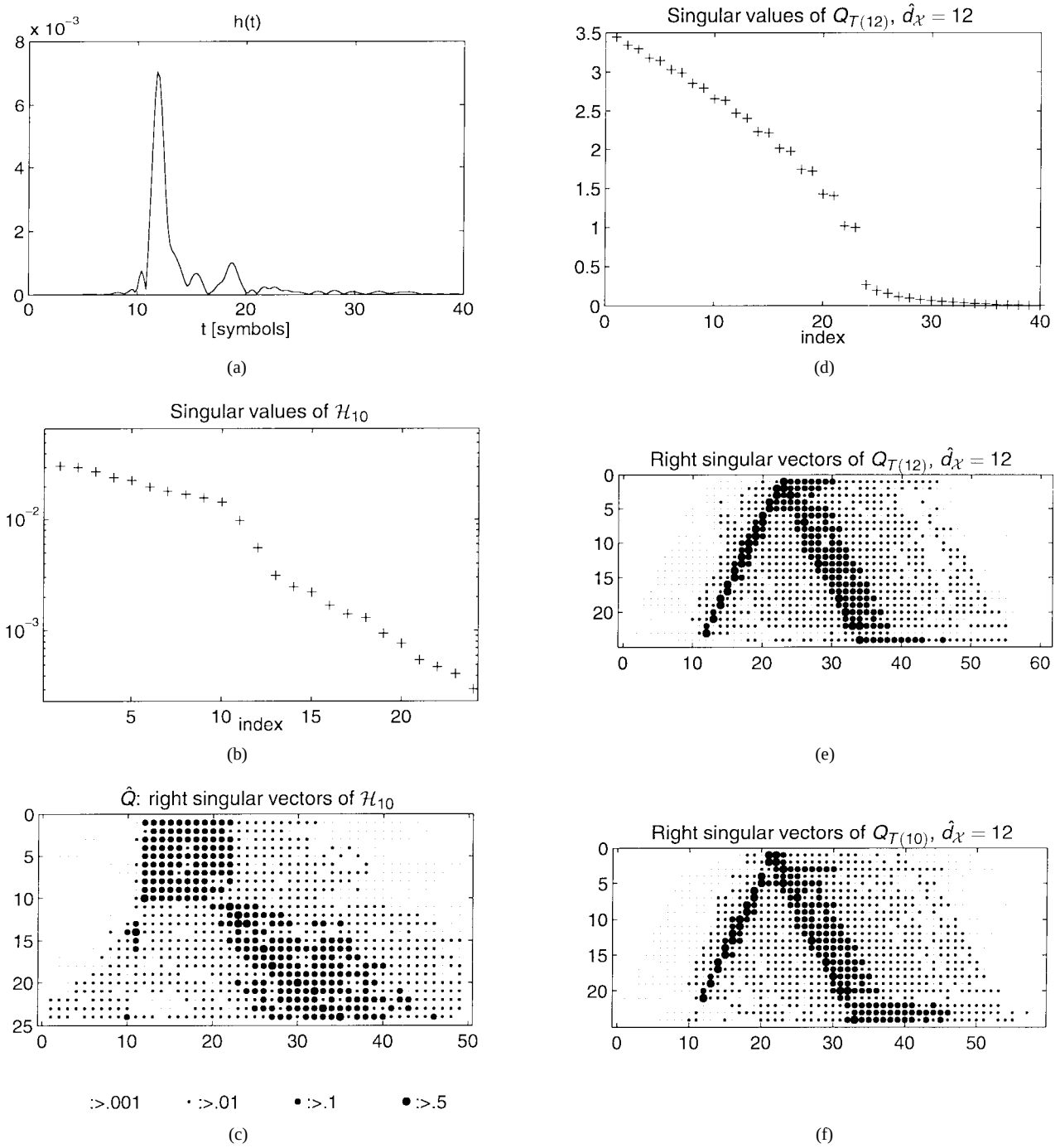


Fig. 6. (a) Channel impulse response $h(t)$. (b) Singular values of $\mathcal{H}_{10}$: the numerical rank of $\mathcal{H}_{10}$ is about 12. (c) $\hat{Q}$, i.e., right singular vectors of $\mathcal{H}_{10}$ (magnitude of entries). (d) Singular values of $Q_{T(12)}$, with $\hat{d}_{\mathcal{X}} = 12$. (e) Right singular vectors of $Q_{T(12)}$ and (f) of $Q_{T(10)}$.

$V_{T(n)}$ corresponding to singular values close to n still contains $d_S = \hat{d}_{\mathcal{X}} - (n-1) > 1$ echos of each signal. A straightforward generalization of the singular value model (18) in Section IV-D to $n < L + m - 1$ shows that, with $\hat{Q} = I$, each row of Y is an average of n out of $\hat{d}_{\mathcal{X}}$ rows of $\hat{V}$. Rows of Y that are formed by combining the top n rows of $\hat{V}$ contain less noise than others.

In general, $\hat{V}$ is some other combination of the symbols ($Q \neq I$ and $d > 1$). However, the effect that, with fewer intersections, some rows of Y are less contaminated by noise than others is still often observed. This is illustrated in Fig. 6(f), where we have taken $n = m = 10$ intersections rather than 12. The first two singular vectors are each a linear combination of only three symbols, rather than eight, as we had before with full row span intersections. Y has $d_S = 3$ rows, and the third singular vector is indeed noisy; it is seen to be a linear combination of nine symbols.

C. Multistage Intersections

Motivated by the preceding subsection, we propose a multistage intersection scheme. The first intersection stage only

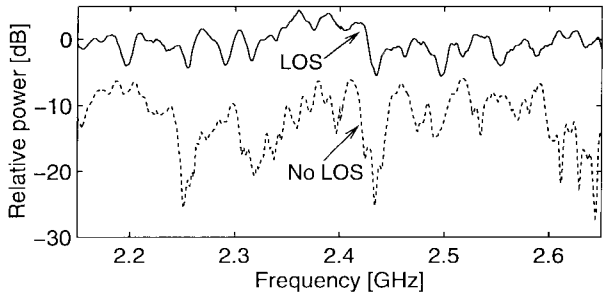
TABLE II
BLIND FIR-MIMO IDENTIFICATION ALGORITHM

In: $\mathcal{X}$, **out:** generator of $\mathcal{S}$ s.t. $\mathcal{X} = \mathcal{H}\mathcal{S}$, w. $\mathcal{S}$ Toeplitz+FA

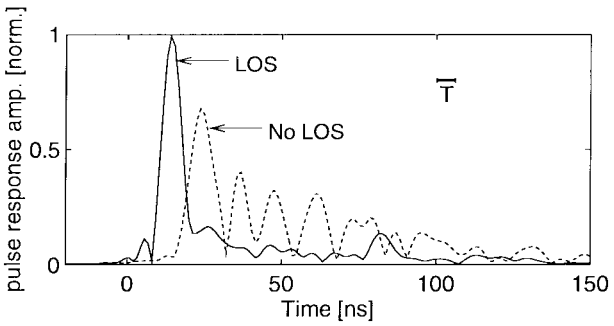
1. *Estimate row*($\mathcal{X}$):
   a. Compute SVD($\mathcal{X}$): $\mathcal{X} =: U\Sigma V$
   b. Estimate $\hat{d}_{\mathcal{X}} = \text{rank}(\mathcal{X})$ from Σ
   c. $\hat{V}$ = first $\hat{d}_{\mathcal{X}}$ rows of V
   (d. Estimate d from rank($\mathcal{X}_{m+1}$))
2. *Partial time-equalization: do n subspace intersections*:
   a. Set $n = m + \hat{L} - 1$, with $\hat{L} = 1$ or $\hat{L} = \min(L_j)$
   b. Construct $V_{T(n)}$ in equation (12)
   c. Compute SVD($V_{T(n)}$)
   d. Set $\hat{d}_S = \hat{d}_{\mathcal{X}} - d(n-1)$
   e. Y:= largest $\hat{d}_S$ right singular vectors
3. *Separate signals based on FA property*:
   a. Select $p \leq \lceil \frac{\hat{d}_S}{d} \rceil$
   b. Do ILSF on Toeplitz matrix $Y_{T(p)}$ from Y
   c. Detect echos by symbol sequence comparison, keep d independent signals with lowest variance

takes the well-defined intersections: At most, $n = m + \min_j(L_j) - 1$, where L_j is an underestimate of the channel length of signal j, but without prior knowledge of channel lengths, perhaps even only $n = m$. This produces a basis Y which is too large. In fact, it contains $\hat{d}_S = \hat{d}_{\mathcal{X}} - d(n-1)$ rows and is a basis for $\mathcal{S}_{L+m-n}$. The second intersection stage has to remove the remaining ISI. Instead of using an SVD, we combine this stage with the separation stage (ILSF or some other I-MIMO algorithm), i.e., the finite alphabet property is used to do the remaining equalization and the signal separation as well.

In principle, we can apply ILSF directly on Y. We could recover all rows of $\mathcal{S}_{L+m-n}$ and select those d rows that are not shifts of each other and that have the smallest deviation from the alphabet. However, since the rows of Y are a basis for the Toeplitz matrix $\mathcal{S}_{\hat{d}_S}$, it is more general to prepare for a subspace intersection step, i.e., augment Y to a Toeplitz matrix $Y_{T(p)}$, where p is some small number. Similar to the construction of V_T from $\hat{V}$, we construct $Y_{T(p)}$ by stacking p shifted copies of Y (omitting J_1,J_2). However, instead of applying an SVD to Y_T, we apply ILSF so that signals and echos are separated based on finite alphabet properties. The resulting variance on the symbol estimates should be lower since ILSF has the same degrees of freedom as the SVD but is not blind to symbol variance. The value for p could vary between 1 and $\hat{d}_{\mathcal{X}}/d - n$. A larger p will always result in symbol estimates with lower variance. In the latter extreme case, we act on the same data that a secondary SVD-based subspace intersection stage would use. However, p cannot be too large because the complexity and the reliability of convergence of ILSF to the global minimum deteriorates with growing dimensions.



(a)



(b)

Fig. 7. (a) Relative power and (b) response to a raised-cosine pulse ($T = 6$ ns, $\beta = 0.5$) of two measured indoor channels.

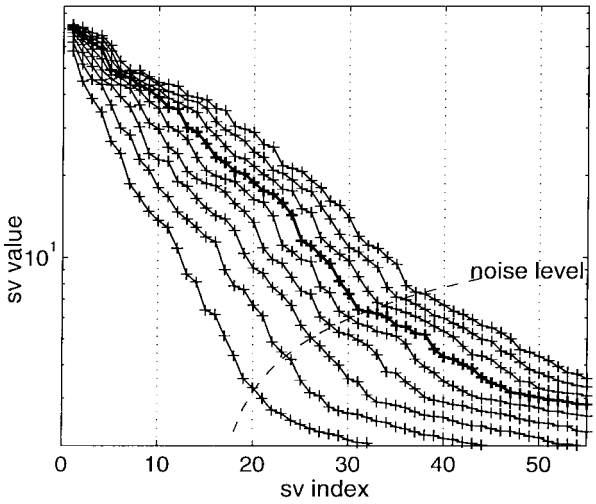


Fig. 8. Singular values of $\mathcal{X}_m$ for $m = 2, \cdots, 10$.

The resulting algorithm has the general structure of Fig. 2 and is listed in Table II. The significance of taking $p > 1$ will be clear from the simulation results in Section VI.

VI. SIMULATION RESULTS

In this section, we report on a test of the algorithm in an off-line experiment, in which we simulate the reception of a number of BPSK signals through an indoor wireless channel at 2.4 GHz. The channel impulse responses are derived from experimental data measured in an office at FEL-TNO, which is in The Hague, The Netherlands, in 1992 [39].[5]

[5]We are grateful to G. J. M. Janssen (now at TU Delft) for sharing his measurement data.

The office has dimensions 5.6 m $\times$ 5.0 m and height 3.5 m. The actual measurement set-up had a transmit antenna in the center of the room at a height of 3.0 m and a receiving antenna cluster located at varying positions at a height of 1.5 m. The cluster consisted of six wideband antennas spaced $\lambda/2$ in a circular array.

Assuming reciprocity (not quite true), we can pretend to simulate a central basestation antenna array of up to six elements, receiving a superposition of signals from a number of user locations. We have used data from two such locations: one with a direct line of sight (RMS delay spread $=$ 7.3 ns) and one without LOS (RMS delay spread $=$ 16.7 ns). The relative powers in the frequency domain are plotted in Fig. 7(a). Fig. 7(b) shows the amplitude of the impulse responses to a raised-cosine pulse ($T = 6$ ns, $\beta = 0.5$ demodulated to baseband from a carrier frequency of 2.4 GHz), each normalized to unit power. We do not have any application in mind with these numbers; they are chosen to provide an ambitious test case that uses all of the measured bandwidth.

In the experiment, we took $d = 2$ BPSK sources, transmitted over the above channels, $M = 3$ antennas, $P = 3$ times oversampling, and $N = 300$ samples. The received power of both signals was scaled to be equal, and we added complex white Gaussian noise with variance σ^2 such that the signal-to-noise ratio SNR $= \|X\|_F^2/(dMPN\sigma^2) = 15$ dB per antenna per sample per signal. The singular values of $\mathcal{X}$ are plotted in Fig. 8 for a range of values of m. It is seen that the numerical rank of $\mathcal{X}$ ($= \hat{d}_{\mathcal{X}}$) cannot be estimated very well, but clearly, $d = 2$, as deduced from the horizontal shifts for increasing m. For $m = 7$, it seems reasonable to set $\hat{d}_{\mathcal{X}}$ in the range 20–30, which makes the "observable channel length" $\hat{L}$ equal to 4–9, if the channels had equal lengths. As in the single-user case (Fig. 6(a) and (b)), the actual channel lengths cannot really be deduced from the data.

Fig. 9 shows the standard deviations of the symbol estimates (before classification as ± 1) for a range of parameter settings: estimated rank $\hat{d}_{\mathcal{X}}$, number of intersections n, and secondary equalizer p. ILSF initialized with $T^{(0)} = I$ was used as finite alphabet separation algorithm. The values of these parameters have a deliberate impact on the performance, but precisely how to find the best settings *a priori* is an open problem. As a general observation, it is possible to underestimate $\hat{d}_{\mathcal{X}}$, but in that case, it is essential that n is taken small ($n \approx m$) and that ILSF is used as an equalizer as well ($p > 1$). However, p should not be taken too large because then, the matrices on which ILSF acts become too big, leading to an abundance of local minima. To put the graphs into perspective, note that at this noise level, the standard deviations of the symbol estimates in an ISI-free scenario, where $L = 1$ and the antennas and oversampling produce MP independent observations per symbol would be $\sigma/\sqrt{2MP} = 0.04$. (The factor 2 is due to the transformation of X to a real matrix, as in (3).)

VII. Conclusion

We have presented an algorithm for blind sequence estimation of multiple digital sources in a general multipath environment. The algorithm uses information from multiple sensors, oversampling to exploit the constant symbol period, and the finite alphabet property of digital signals. It is set in a deterministic framework and uses subspace properties of the underlying structured matrix factorization problem. This approach is effective in situations where the channel lengths are well determined. We have indicated some problems that may arise in subspace intersections algorithms when the channel lengths are not well defined and suggested a modification that should give improvements for channels with well-defined peaks.

Appendix A
Intersection of Subspaces

Let $\mathcal{H}_1, \mathcal{H}_2$ be subspaces in $\mathbb{C}^N$ with orthogonal complements $\mathcal{K}_1 = \mathcal{H}_1^\perp$, $\mathcal{K}_2 = \mathcal{H}_2^\perp$. Then,

$$\mathcal{H}_1 \cap \mathcal{H}_2 = (\mathcal{K}_1 + \mathcal{K}_2)^\perp.$$

The computation of the intersection via this equation requires the formation of three orthogonal complements. With K_1, K_2 matrices whose columns form orthonormal bases for $\mathcal{K}_1$, $\mathcal{K}_2$, we can obtain a basis for the intersection of $\mathcal{H}_1$ and $\mathcal{H}_2$ from the kernel of $[K_1 \quad K_2]$. With noisy data, this requires the computation of an SVD of $[K_1 \quad K_2]$: A basis of the estimated kernel is given by the singular vectors that correspond to small singular values.

In our application, the dimension of the $\mathcal{H}_i$ is independent of N so that the dimension of the complements grows with N. This means that for large N, it is not attractive to compute the intersection in this manner. We show in the following proposition that precisely the same information may be gleaned from the large singular values and corresponding singular vectors of a matrix $[H_1 \quad H_2]$, where H_1, H_2 are orthonormal bases of $\mathcal{H}_1, \mathcal{H}_2$.

Proposition 1: Let $\mathcal{H}_1, \mathcal{H}_2$ be subspaces in $\mathbb{C}^N$ with orthonormal bases H_1, H_2, and let $[H_1 \quad H_2] = U_H \Sigma_H V_H^*$ be an SVD. Suppose that K_1, K_2 are orthonormal bases for $\mathcal{H}_1^\perp$, $\mathcal{H}_2^\perp$. Then, $[K_1 \quad K_2]$ has an SVD $[K_1 \quad K_2] = U_H(2I - \Sigma_H^2)^{1/2} V_K^*$ for some unitary matrix V_K.

Proof: Since $H_1 H_1^* + K_1 K_1^* = I$, $H_2 H_2^* + K_2 K_2^* = I$, we have

$$[H_1 \quad H_2][H_1 \quad H_2]^* + [K_1 \quad K_2][K_1 \quad K_2]^* = 2I.$$

Substituting $[H_1 \quad H_2] = U_H \Sigma_H V_H^*$ and multiplying the above equation with $U_H^*(\cdots)U_H$, we obtain

$$\Sigma_H^2 + U_H^*[K_1 \quad K_2][K_1 \quad K_2]^* U_H = 2I.$$

Since both Σ_H^2 and $2I$ are diagonal, this implies that there is a unitary matrix V_K such that $U_H^*[K_1 \quad K_2]V_K$ is diagonal. This, however, constitutes precisely an SVD for $[K_1 \quad K_2]$. □

This result is readily generalized to the joint intersection of n subspaces $\mathcal{H}_i$, $i = 1, \cdots, n$. Likewise, we compute an SVD of $[H_1 \quad \cdots \quad H_n]$ but now look for singular values that are close to $\sqrt{n}$.

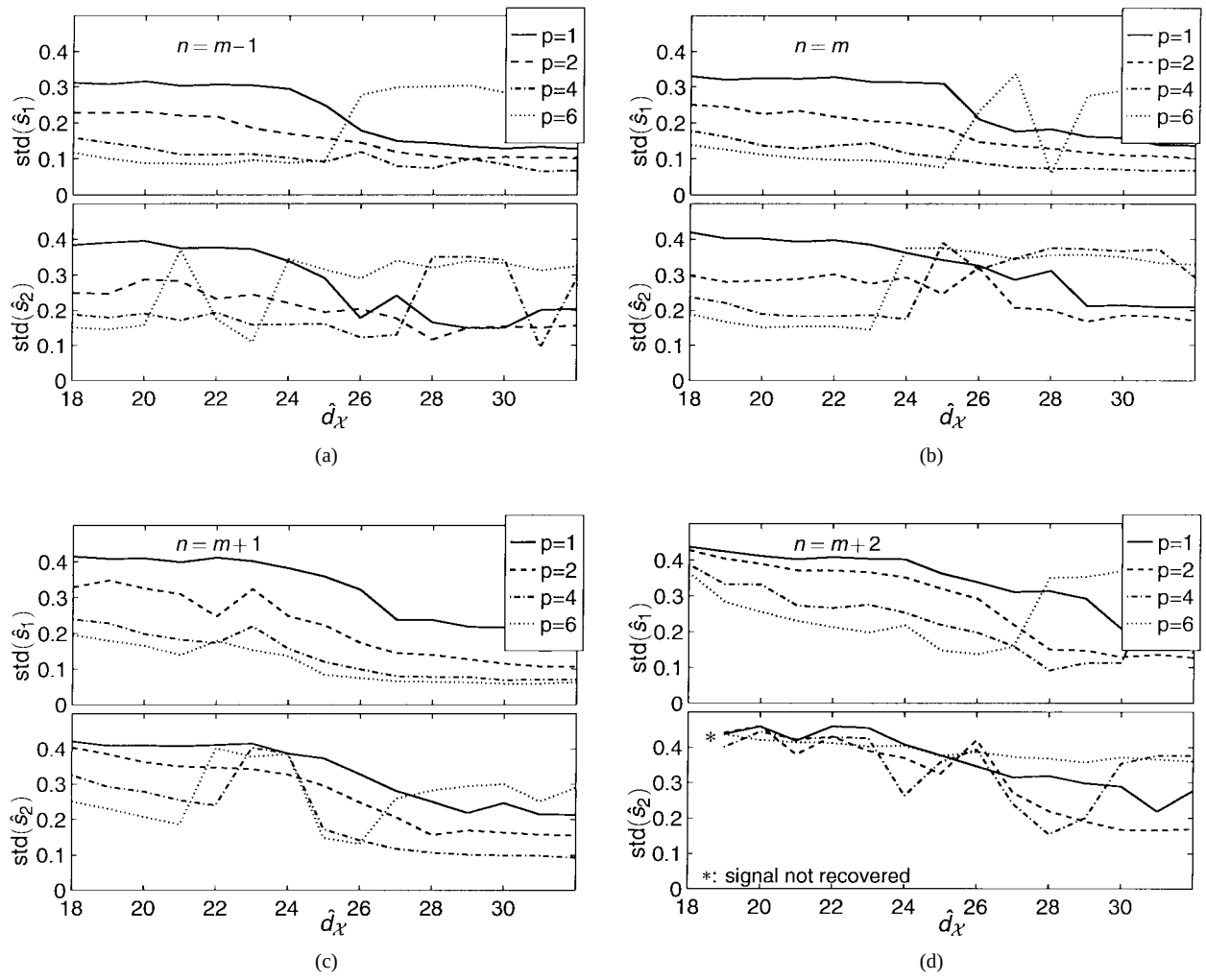


Fig. 9. Standard deviations of signal estimates for $m = 7$ and varying settings of $\hat{d}_{\mathcal{X}}$, n, p.

APPENDIX B
APPROXIMATE CRAMER–RAO BOUNDS

Suppose $X = H\mathcal{S}_L + E$, where E is a white i.i.d. complex Gaussian noise process with covariance matrix $\sigma^2 I$. For simplicity of future notation, let us specialize to the case of real signals (e.g., BPSK signals $\Omega = \{\pm 1\}$). Define vectors of unknown parameters $\mathbf{h} := \text{vec}\,(H)$ and $\mathbf{s} := [\mathbf{s}_{N-1}^T \ \cdots \ \mathbf{s}_{-L+1}^T]^T$. We assume that the number of sources d is known and that the channels have equal known channel length L. If we do not take into account that the entries of $\mathbf{s}$ belong to a finite alphabet, then the concentrated Fisher information matrix for $\theta = [\mathbf{s}^T, \mathbf{h}^T]^T$ is derived in [26] as

$$\Phi = \frac{1}{\sigma^2}\begin{bmatrix} \mathcal{H}_N^T\mathcal{H}_N & \mathcal{H}_N^T\mathcal{C} \\ \mathcal{C}^T\mathcal{H}_N & \mathcal{C}^T\mathcal{C} \end{bmatrix}, \qquad \mathcal{C} := \mathcal{S}_L^T \otimes I_{2MP}$$

(originally for a single signal, but the results are readily generalized for $d > 1$ and adapted here for a real data model). The CRLB that describes the lower bound on the covariance of any unbiased estimator for $\mathbf{s}$ and $\mathbf{h}$ is obtained by inverting Φ. However, as noted in [26], this matrix turns out to be singular because there is ambiguity in the parameter values: Without forcing the FA property, we can only identify H and S up to an invertible $d \times d$ matrix A. Indeed, the dimension of the null space of Φ is observed to be d^2 in generic examples. To fix A, one has to assume that certain symbols are known.

For $d = 1$, knowing the value of one symbol of S suffices, and the variance of the remaining estimates is obtained by deleting the corresponding column of $\mathcal{H}_N$. Let $\mathcal{H}'_N$ be equal to $\mathcal{H}_N$ with the column corresponding to the known symbol taken out, and define $\mathbf{s}'$ and Φ' accordingly. Then, the CRLB on the covariance of $\mathbf{s}'$ is

$$C' = (\Phi'^{-1})_{1,1} = \sigma^2\{\mathcal{H}_N'^T\mathcal{H}'_N - \mathcal{H}_N'^T\mathcal{C}(\mathcal{C}^T\mathcal{C})^{-1}\mathcal{C}^T\mathcal{H}'_N\}^{-1} \tag{19}$$

(the subscript '1, 1' denotes the (1,1) block of Φ'^{-1} following the partitioning of Φ') so that, in particular,

$$\text{var}(\mathbf{s}') \geq \text{diag}(C') \qquad \text{[CRLB with "training," no FA]}. \tag{20}$$

This is basically the result in [26], where it is also noted that although the bound is dependent on both H and S, its dependence on S is only weak in practice. However, a number of remarks that go beyond [26] are in order.

1) It makes a difference *which* symbol is assumed to be known. Not surprisingly, knowing one of the center symbols gives significantly lower variances than knowing

one of the first or last $L-1$ symbols because these play a less significant role in $\mathcal{S}_L$. Additionally, the variances of the symbols in the range $0, \cdots, N-L$ are usually approximately equal to each other, but the first and last $L-1$ symbols have a significantly larger variance. In the computation of the expected bit error rates, we have taken these tail symbols out of consideration.

2) The result (20) strictly speaking applies to a scenario in which we have a "training sequence" of length 1. It is readily generalized for training sequences longer than 1 by leaving out more columns of $\mathcal{H}_N$.
3) The above remark implies that in the *actual* blind algorithm, the above lower bound on the variance is too large by about a factor 2. Indeed, (20) is valid for estimates where the variance of one symbol, $\mathbf{s}_r$ say, is made zero. This is conceptually done by estimating any sequence and then normalizing the rth entry by dividing the estimated sequence by the estimated value of $\mathbf{s}_r$ and multiplication by its desired value. Assuming relatively small variances, the division causes the variance of all other symbol estimates to be enlarged by the variance of the estimate of $\mathbf{s}_r$.[6] In the actual blind scheme, we do not normalize on a single symbol $\mathbf{s}_r$ but normalize $\|S\|_F$. In that case, the lower bound (20) is too high and formally not applicable. To attempt to correct for this, we have to estimate the variance of $\mathbf{s}_r$, e.g., as $\frac{1}{2}$median diag(C') and subtract to get

$$\operatorname{var}(\mathbf{s}'_{\text{blind}}) \stackrel{\geq}{\approx} \operatorname{diag}(C') - \tfrac{1}{2}\text{median diag}(C') \quad \text{[approx. blind CRLB, no FA, } d=1\text{]}. \tag{21}$$

(We take the median instead of the mean to avoid the influence of outliers at the tails of the sequence.) This is the (approximate) CRLB for a blind scheme that relies on a structured factorization $X = H\mathcal{S}_L$, not taking the finite alphabet into account other than for removing the ambiguity. If all estimates have approximately equal variance, the originally derived bound (20) is about a factor 2 too high.

For $d > 1$, roughly the same derivation holds, except that we have to pretend that more symbols are known because the ambiguity factor A now has d^2 degrees of freedom. Hence, we have to fix d symbols of d signals, i.e., a $d \times d$ submatrix S_r of S somewhere in the center of S. An extra complication is that this submatrix S_r has to be full rank or else some ambiguity in A remains. Hence, in computing the bound, we have to select d independent columns of S, which are located somewhere in the center, and delete the corresponding columns of $\mathcal{H}_N$ to obtain $\mathcal{H}'_N$. After this, the bound (20) is derived as before. The correction for the unnatural normalization as assumed in that bound is somewhat more intricate now. Indeed, before normalization, let us say we have symbol estimates

$$S_{\text{blind}} = [S_r \quad S'] + [E_r \quad E'].$$

S_r and S' are the exact symbols, and E_r and E' represent the noise on the estimates. Normalization to arrive at an estimate in which the known symbols have zero variance gives the modified estimates $\tilde{S}'$ for which the CRLB (20) holds as

$$\begin{aligned} \tilde{S}' &= S_r(S_r + E_r)^{-1}(S' + E') \\ &\approx S_r S_r^{-1}(I - E_r S_r^{-1})(S' + E') \\ &\approx S' - E_r S_r^{-1} S' + E'. \end{aligned}$$

Note that S_r^{-1} can actually amplify the noise contribution by E_r. In estimating the correction on the bound, assume (not entirely correctly) that the columns of $E = [E_r \;\; E']$ are independent, zero mean, and have equal distribution $\mathrm{E}(\mathbf{e}_i \mathbf{e}_i^T) =: R_{\mathbf{e}}$. Let $\mathbf{s}_i$ be the ith column of S', and then

$$\mathrm{E}((\tilde{\mathbf{s}}_i - \mathbf{s}_i)(\tilde{\mathbf{s}}_i - \mathbf{s}_i)^T) \approx R_{\mathbf{e}} \|S_r^{-1}\mathbf{s}_i\|^2 + R_{\mathbf{e}} = R_{\mathbf{e}}(1 + \|S_r^{-1}\mathbf{s}_i\|^2).$$

The left-hand side of this expression is given by the uncorrected CRLB, namely, C'_i, which is the $d \times d$ submatrix of C' in (20) corresponding to $\mathbf{s}_i$. It follows that an estimate of $R_{\mathbf{e}}$ and an approximate lower bound on $\operatorname{var}(\mathbf{s}_{i,\text{blind}})$ can be obtained as

$$\begin{aligned} R_{\mathbf{e}} &\approx \text{median}_i \left\{ \frac{C'_i}{1 + \|S_r^{-1}\mathbf{s}_i\|^2} \right\} \\ \operatorname{var}(\mathbf{s}_{i,\text{blind}}) &\stackrel{\geq}{\approx} C'_i - R_{\mathbf{e}} \|S_r^{-1}\mathbf{s}_i\|^2. \end{aligned} \tag{22}$$

For $d = 2$ BPSK signals, $\|S_r^{-1}\mathbf{s}_i\| = 1$ always, and the above expression reduces to (21).

[6] Here, the first-order approximation $s_r(s_r + e_r)^{-1}(s+e) \approx s - e_r s_r^{-1} s + e$ is used, as well as the BPSK assumption $|s_i| = 1$.

REFERENCES

[1] A. J. van der Veen, S. Talwar, and A. Paulraj, "Blind estimation of multiple digital signals transmitted over FIR channels," *IEEE Signal Processing Lett.*, vol. 2, pp. 99–102, May 1995.

[2] ——, "Blind identification of FIR channels carrying multiple finite alphabet signals," in *Proc. IEEE ICASSP*, 1995, pp. 1213–1216.

[3] W. A. Gardner, "A new method of channel identification," *IEEE Trans. Commun.*, vol. 39, pp. 813–817, June 1991.

[4] 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.

[5] L. Tong, G. Xu, B. Hassibi, and T. Kailath, "Blind channel identification based on second-order statistics: a frequency-domain approach," *IEEE Trans. Inform. Theory*, vol. 41, pp. 329–334, Jan. 1995.

[6] L. Tong, G. Xu, and T. Kailath, "Fast blind equalization via antenna arrays," in *Proc. IEEE ICASSP*, 1993, pp. IV:272–274.

[7] E. Moulines, P. Duhamel, J. 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.

[8] D. Slock, "Blind fractionally-spaced equalization, perfect-reconstruction filter banks and multichannel linear prediction," in *Proc. IEEE ICASSP*, vol. IV, 1994, pp. 585–588.

[9] G. Xu and H. Liu, "A deterministic approach to blind identification of multi-channel FIR systems," in *Proc. IEEE ICASSP*, 1994, vol. IV, pp. 581–584,

[10] F. R. Magee and J. G. Proakis, "Adaptive maximum-likelihood sequence estimation for signaling in the presence of intersymbol interference," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 120–124, Jan. 1973.

[11] G. Ungerboeck, "Adaptive maximum-likelihood receiver for carrier-modulated data transmission systems," *IEEE Trans. Commun.*, vol. COM-22, pp. 624–636, May 1974.

[12] G. Picchi and G. Prati, "Blind equalization and carrier recovery using a 'stop-and-go' decision-directed algorithm," *IEEE Trans. Commun.*, vol. COM-35, pp. 877–887, Sept. 1987.

[13] Z. Ding, "Blind equalization based on joint minimum MSE criterion," *IEEE Trans. Commun.*, vol. 42, pp. 648–654, Feb. 1994.

[14] N. Seshadri, "Joint data and channel estimation using fast blind trellis search techniques," in *Proc. Globecom*, 1991, pp. 1659–1653.

[15] 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.

[16] S. Talwar, M. Viberg, and A. Paulraj, "Blind estimation of multiple co-channel digital signals using an antenna array," *IEEE Signal Processing Lett.*, vol. 1, pp. 29–31, Feb. 1994.

[17] ——, "Blind separation of synchronous co-channel digital signals using an antenna array. Part I: Algorithms," *IEEE Trans. Signal Processing*, vol. 44, pp. 1184–1197, May 1996.
[18] A. Swindlehurst, S. Daas, and J.Yang, "Analysis of a decision directed beamformer," *IEEE Trans. Signal Processing*, vol. 43, pp. 2920–2927, Dec. 1995.
[19] A. Belouchrani and J.-F. Cardoso, "Maximum likelihood source separation for discrete sources," in *Proc. Eusipco*, 1994, pp. 768–771.
[20] H. Liu and G. Xu, "A deterministic approach to blind symbol estimation," *IEEE Signal Processing Lett.*, vol. 1, pp. 205–207, Dec. 1994.
[21] H. Liu and G. Xu, "Multiuser blind channel estimation and spatial channel pre-equalization," in *Proc. IEEE ICASSP*, 1995, pp. 1756–1759.
[22] K. Abed-Meraim, P. Loubaton, and E. Moulines, "A subspace algorithm for certain blind identification problems," *IEEE Trans. Inform. Theory*, vol. 42, July 1996.
[23] S. Andersson, M. Millnert, M. Viberg, and B. Wahlberg, "An adaptive array for mobile communication systems," *IEEE Trans. Vehic. Technol.*, vol. 40, no. 1, pp. 230–236, 1991.
[24] 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, vol. 5, pp. 2421–2424.
[25] A. J. van der Veen, M. C. Vanderveen, and A. Paulraj, "Joint angle and delay estimation using shift-invariance properties," submitted to *IEEE Signal Processing Lett.*
[26] H. Liu and G. Xu, "Closed-form blind symbol estimation in digital communications," *IEEE Trans. Signal Processing*, vol. 43, pp. 2714–2723, Nov. 1995.
[27] A. J. van der Veen, "Resolution limits of blind multi-user multi-channel identification schemes - the bandlimited case," in *Proc. IEEE ICASSP*, Atlanta, GA, May 1996, pp. 2722–2725.
[28] S. N. Diggavi, Y. Cho, and A. Paulraj, "Blind estimation of multiple co-channel digital signals in vector FIR channels," in *Proc. IEEE Globecom*, 1995.
[29] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Baltimore, MD: John Hopkins Univ. Press, 1984.
[30] P. Comon and G. H. Golub, "Tracking a few extreme singular values and vectors in signal procesing," *Proc. IEEE*, vol. 78, pp. 1327–1343, Aug. 1990.
[31] B. Yang, "Projection approximation subspace tracking," *IEEE Trans. Signal Processing*, vol. 43, pp. 95–107, Jan. 1995.
[32] A. J. van der Veen, "A Schur method for low-rank matrix approximation," *SIAM J. Matrix Anal. Appl.*, vol. 17, pp. 139–160, Jan. 1996.
[33] A. J. van der Veen and A. Paulraj, "An analytical constant modulus algorithm," *IEEE Trans. Signal Processing*, vol. 44, pp. 1136–1155, May 1996.
[34] A. J. van der Veen, "Analytical method for blind binary signal separation," accepted for publication in *IEEE Trans. Signal Processing*, Oct. 1996.
[35] J.-F. Cardoso, "Super-symmetric decomposition of the fourth-order cumulant tensor. Blind identification of more sources than sensors," in *Proc. IEEE ICASSP*, Toronto, Ont., Canada, 1991, vol. 5, pp. 3109–3112.
[36] ——, "Iterative techniques for blind source separation using only fourth-order cumulants," in *Signal Processing VI: Proc. EUSIPCO-92*, Brussels, J. Vandewalle, Ed. New York: Elsevier, 1992, vol. 2, pp. 739–742.
[37] L. Tong, Y. Inouye, and R. Liu, "Waveform-preserving blind estimation of multiple independent sources," *IEEE Trans. Signal Processing*, vol. 41, pp. 2461–2470, July 1993.
[38] 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.
[39] G. J. M. Janssen and R. Prasad, "Propagation measurements in an indoor radio environment at 2.4 GHz, 4.75 GHz and 11.5 GHz," in *Proc. 42nd VTS Conf.*, Denver, vol. 2, 1992, pp. 617–620.

Alle-Jan van der Veen (S'87–M'94) was born in The Netherlands in 1966. He graduated (cum laude) from the Department of Electrical Engineering, Delft University of Technology, Delft, The Netherlands, in 1988, and received the Ph.D. degree (cum laude) from the same institute in 1993.

Throughout 1994, he was a postdoctoral scholar at Stanford University, Stanford, CA, in the Scientific Computing/Computational Mathematics Group and in the Information Systems Laboratory. At present, he is a researcher 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, in particular, system identification, time-varying system theory, and in numerical methods and parallel algorithms for linear algebra problems.

Dr. van der Veen is the recipient of a 1994 IEEE SP paper award.

Shilpa Talwar received the M.S. degree in electrical engineering and the Ph.D. degree in scientific computing and computational mathematics from Stanford University, Stanford, CA, in 1996.

She is currently employed by Stanford Telecom, Sunnyvale, CA. Her research interests include wireless communications, array signal processing, and numerical linear algebra.

Arogyaswami Paulraj (F'91) was educated at the Naval Engineering College, India, and at the Indian Institute of Technology, New Delhi, where he received the Ph.D. degree in 1973.

A large part of his career to date has been spent in research laboratories in India, where he supervised the development of several electronic systems. His contributions include a sonar receiver in 1973–1974, a surface ship sonar in 1976–1983, a parallel computer in 1988–1991, and telecommunications systems. He has held visiting appointments at several universities: the Indian Institute of Technology, Delhi, from 1973 to 1974, Loughborough University of Technology, UK, from 1974 to 1975, and Stanford University, Stanford, CA, from 1983 to 1986. His research has spanned several disciplines, emphasizing estimation theory, sensor signal processing, antenna array processing, parallel computer architectures/algorithms and communication systems. He is currently a Professor of Electrical Engineering at Stanford University, working in the area of mobile communications. He is the author of about 90 research papers and holds several patents.

Dr. Paulraj has won a number of national awards in India for his contributions to technology development.