Intrinsic Subspace Convergence in TDD MIMO Communication

Tobias Dahl, Silvana Silva Pereira, Nils Christophersen, and David Gesbert

Abstract—In numerical linear algebra, students encounter early the iterative power method, which finds eigenvectors of a matrix from an arbitrary starting point through repeated normalization and multiplications by the matrix itself. In practice, more sophisticated methods are used nowadays, threatening to make the power method a historical and pedagogic footnote. However, in the context of communication over a time-division duplex (TDD) multiple-input multiple-output (MIMO) channel, the power method takes a special position. It can be viewed as an intrinsic part of the uplink and downlink communication switching, enabling estimation of the eigenmodes of the channel without extra overhead. Generalizing the method to vector subspaces, communication in the subspaces with the best receive and transmit signal-to-noise ratio (SNR) is made possible. In exploring this *intrinsic subspace convergence* (ISC), we show that several published and new schemes can be cast into a common framework where all members benefit from the ISC.

Index Terms—Channel identification, eigenmodes, multiple-input multiple-output (MIMO) systems, singular modes, singular value decomposition (SVD).

I. INTRODUCTION

OVER the last few years, a number of papers have been published on communication in time-division duplex (TDD) multiple-input multiple-output (MIMO) systems using the top singular modes of the channel [4], [5], [15], [21], [11]. These techniques are potentially very useful—they are blind and computationally tractable and offer maximum signal-to-noise ratio (SNR) for independent data streams when successful. Combined with an optimization of the number of streams and water-filling algorithms, singular-vector-based methods can realize an interesting tradeoff between rate and diversity maximization.

However, these methods may also enjoy another feature that has been little highlighted in the literature: the *intrinsic subspace convergence* (ISC) of the MIMO channel. As a first example, Andersen [1] noted that transmitting a vector repeat-

T. Dahl and N. Christophersen are with the Department of Informatics, University of Oslo, N-0316 Blindern, Norway (e-mail: tobias@ifi.uio.no; nilsch@ifi.uio.no).

S. Silva Pereira is with the Universitat Politècnica de Catalunya, Catalunya, Campus Nord C. Jordi Girona 1-3, ES-08034 Barcelona, Spain (e-mail: silvana @gps.tsc.upc.edu).

D. Gesbert is with the Mobile Communication Department, Eurecom, BP 193, F-06904, France (e-mail: gesbert@eurecom.fr).
Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2007.896901

edly forth and back through a MIMO channel led to convergence towards the top singular mode (i.e., with largest singular value). This procedure is almost identical to the algebraic power method, an iterative method for finding eigenvalues and eigenvectors of a matrix [9]. It also has strong connections with time-reversal mirrors [8], now gaining popularity in communication [12], [6]. The power method is based on the observation that for a Hermitian matrix $\mathbf{H} \in C^{N \times N}$ and an arbitrary starting point vector $\mathbf{x} \in C^N$, the following sequence:

$$\mathbf{y}(n) = \underbrace{\mathbf{H}\mathbf{H}\cdots\mathbf{H}}_{n}\mathbf{x} = \mathbf{H}^n\mathbf{x} \qquad (1)$$

has the property that $\mathbf{y}/\|\mathbf{y}\|$ converges towards the unit-norm eigenvector of $\mathbf{H}$ with the largest associated eigenvalue λ_1 of $\mathbf{H}$, assuming λ_1 to be strictly larger than the other eigenvalues. The power method can be modified to become a method for finding left and right *singular vectors* of a non-Hermitian matrix $\mathbf{H} \in C^{N \times M}$, through the following iterations:

$$\mathbf{y}(i) = \mathbf{H}\mathbf{x}(i)$$
$$\mathbf{x}(i+1) = \mathbf{H}^H\mathbf{y}(i). \qquad (2)$$

Here, $\mathbf{x}(i) \in C^M$ and $\mathbf{y}(i) \in C^N$ converge towards scalar multiples of the top right and left singular vectors of $\mathbf{H}$, respectively, and subsequent normalizations at each step yield unit-norm singular vector estimates. Successive singular modes can be estimated by a QR-like procedure [9], which will be described below. $\mathbf{H}$ will denote the *channel matrix* used in TDD MIMO communication.

In [4] and [5], we proposed a technique [Blind Iterative MIMO Algorithm (BIMA), see below] based on this principle, where feedback through transmission and retransmission of the channel is exploited: A set of top singular modes is estimated directly, without estimating the channel matrix itself, and while simultaneously using them for communication. Expanding on this idea, we now present a more general framework for utilizing the ISC effect. Independently of our work, Poon *et al.* [15] and Wen *et al.* [21] suggested related blind methods for directly estimating and utilizing the channel eigenstructure. Their approach is based on subspace estimation through decomposition of the spatial correlation matrix combined with pre- and postfiltering using the singular vectors. However, these authors do not take into account the ISC, which is necessary to understand the effects of feedback in the channel.

Blind methods [7], [16], [18], [19] are known to be notorious with respect to convergence problems and local minima. Furthermore, they often rely on higher order statistics and are thus computationally expensive and may also be prone to error

Fig. 1. Central steps in the ISC Framework. $\mathbf{H}$ is the channel matrix; UL—uplink and DL—downlink communication.

propagation. Proving global convergence, even for the simplest ideas, is often cumbersome if not impossible.

There is an ongoing debate about the usefulness of blind and semiblind methods in the signal processing community. In light of this debate, it should be noted that blind communication over TDD MIMO channels has several important advantages over one-way schemes and frequency-divided systems. The ability to have important channel parameters estimated without the need for pilot sequences, using subspace techniques, is but one.

In the paper at hand, we show convergence of an entire class of MIMO TDD methods and provide theoretical results as well as practical examples showing the usefulness of the power method in TDD MIMO communication. The results are not limited to blind methods, they should appeal to designers of semiblind and training-based systems equally well. To be specific, we show that the methods suggested by Dahl *et al.* [4], [5], Wen *et al.* [21], and Liu and Host-Madsen [11] all belong to a wider class of methods that benefit from the ISC. We also show an example where a training-based method utilizes the ISC.

The layout of the ISC scheme is illustrated in Fig. 1.

II. MATHEMATICAL BACKGROUND

A. Channel Model

We assume two-way communication through a $N(\text{receive}) \times M(\text{transmit})$ flat-fading MIMO channel matrix $\mathbf{H} \in C^{N \times M}$ at time slot/index i.

$$\mathbf{Z}(i) = \mathbf{H}\mathbf{X}(i) \text{ (uplink)} \tag{3}$$
$$\mathbf{W}(i+1) = \mathbf{H}^T\mathbf{Y}(i) \text{ (downlink)} \tag{4}$$

where $\mathbf{X}(i) \in C^{M \times n}$ and $\mathbf{Y}(i) \in C^{N \times n}$ are the transmitted data blocks of length n uplink (UL) and downlink (DL), respectively. A model such as the one above describes a TDD system, provided the ping-pong time—the time between the beginning of a DL frame and the beginning of the next UL frame—is small compared with the channel coherence time. For algorithm derivation purposes, we initially assume no channel noise, but subsequently consider the practical noisy case as well. It is also

convenient to work within a framework where the transpose channel $\mathbf{H}^T$ in (4) is effectively replaced by the complex conjugate transpose $\mathbf{H}^H$. This is achieved by letting the transmit data $\mathbf{Y}(i)$ be complex conjugated prior to transmission, and the received data block $\mathbf{W}(i)$ be complex conjugated prior to any further processing. In this way, (4) may effectively be replaced by

$$\mathbf{W}(i+1) = \mathbf{H}^H\mathbf{Y}(i) \text{ (downlink).} \tag{5}$$

The rank of $\mathbf{H}$ is denoted by $K_0 \leq \min(N, M)$, and its singular value decomposition (SVD) is $\mathbf{H} = \mathbf{U}\mathbf{S}\mathbf{V}^H$. $\mathbf{S}$ is the diagonal matrix of singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{K_0} > 0$, and

$$\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_{K_0}] \in C^{N \times K_0} \tag{6}$$
$$\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_{K_0}] \in C^{M \times K_0} \tag{7}$$

are unitary matrices whose columns can be used as receive and transmit vectors $\{\mathbf{u}_i\}$ and $\{\mathbf{v}_i\}$, respectively. Clearly, one can select a number $K(K \leq K_0)$ of transmit–receive vector pairs for communication through orthogonal singular modes.

B. Classical MIMO Eigenmode Transmission

Assume, initially, that the singular vectors are known: One party [e.g., the base station (BTS)] knows the top K left singular vectors $\{\mathbf{u}_j\}, j = 1, \dots, K$ of $\mathbf{H}$ and the other side [subscriber unit (SU)] knows the corresponding subset of right singular vectors $\{\mathbf{v}_j\}, j = 1, \dots, K$. The top singular values $\{\sigma_1, \sigma_2, \dots, \sigma_K\}$ are assumed to be known by both parties; however, if the symbol alphabet has unit modulo, this is not strictly necessary. Let

$$\mathbf{U}_K = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_K] \tag{8}$$
$$\mathbf{V}_K = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_K] \tag{9}$$
$$\mathbf{S}_K = \text{diag}\{\sigma_1, \sigma_2, \dots, \sigma_K\} \tag{10}$$

denote these subsets arranged into matrices, and let $\mathbf{C_U}(i), \mathbf{C_D}(i) \in C^{K \times n}$ $(n \geq K)$ be the *symbol matrices* comprising the UL and DL symbol blocks, respectively. Each row of a symbol matrix represents an individual data stream.

The elements of these matrices are symbols from a modulation constellation (e.g., PSK and QAM). Using these (known) singular vectors and values and neglecting noise, the UL transmit data block $\mathbf{X}(i) = \mathbf{V}_K \mathbf{C_U}(i)$ would be received as

$$\mathbf{Z}(i) = \mathbf{H}\mathbf{X}(i) = \mathbf{U}\mathbf{S}\mathbf{V}^H \mathbf{V}_K \mathbf{C_U} = \mathbf{U}_K \mathbf{S}_K \mathbf{C_U}(i). \quad (11)$$

The last equality is obtained by noting that $\mathbf{V}^H \mathbf{V}_K = \mathbf{I}_{K_0 \times K}$, which has "1"s in the K main diagonal elements and zeros elsewhere. From this, it follows that $\mathbf{U}\mathbf{S}\mathbf{I}_{K_0 \times K} = \mathbf{U}_K \mathbf{S}_K$. Decoding is then simply performed using $\mathbf{S}_K$ and $\mathbf{U}_K$, noting that

$$\mathbf{C_U}(i) = \mathbf{S}_K^{-1} \mathbf{U}_K^H \mathbf{Z}(i). \quad (12)$$

The DL transmission block is set up as $\mathbf{Y}(i) = \mathbf{U}_K \mathbf{C_D}(i)$, and decoding is carried out at the subscriber unit correspondingly. Both parties require $\mathbf{S}_K$, but the key remark is that no party requires complete knowledge of $\mathbf{H}$. In fact, the base station both receives and transmits using only $\mathbf{U}_K$, while the subscriber unit only needs $\mathbf{V}_K$. In addition, if only phase modulation is used (e.g., QPSK), the singular values are not even necessary for decoding.

C. ISC Framework

Consider the situation where we more generally want to communicate on K transmit/receive vectors. These could be singular vector estimates held in $M \times K$ matrices $\mathbf{T_U}$ (uplink) and $N \times K$ matrices $\mathbf{T_D}$ (downlink), but other possibilities exist. $K \times n$ symbol blocks $\mathbf{C_U}, \mathbf{C_D}$ are transmitted after premultiplying them with $\mathbf{T_U}$ and $\mathbf{T_D}$, which serve as UL and DL "prefilters," respectively. Note that the product matrices $\mathbf{T_U}\mathbf{C_U}$ and $\mathbf{T_D}\mathbf{C_D}$, which are the actual prefilter-and-symbol blocks transmitted over the channel, have rank equal to $K \leq K_0$.

As an informal analogy of what we are going to describe, consider the following story of a communications engineer who "got lucky." In constructing a communication algorithm for a two-way TDD MIMO system, she decides to start with a random set of orthogonal prefilter vectors, whose number K is less than the rank K_0 of the channel. Using those prefilters for coding the independent K data streams to construct a rank-reduced transmit data block, that block is transmitted over the channel in one direction. On the other side, some blind algorithm is used to decipher the symbols just transmitted. Then, the column vectors of the received data block (deciphered or not) are linearly combined to construct another set of K orthogonal prefilters. As we shall see, there are many ways to obtain such prefilters.

The prefilter vector set is used to send another block of K data streams in the other direction, where once again, symbol deciphering and column vector combinations are used to construct orthogonal prefilters. The procedure is carried out over and over again. The engineer notices that the blind symbol decision starts getting better as the iterations go on. Why? Seemingly, she is just receiving data blocks and combining them more or less at random to obtain new sets of prefilter vectors. Why should blind decoding of the symbols gradually get easier? Because, "magically," the prefilter vectors will tend towards lying in the top left and right singular subspaces of the channel, the subspaces

maximizing the effective SNR associated with the communication process. The main contribution of the present paper is to explain why and to what extent this *intrinsic subspace convergence* occurs.

Formalizing the above story, the rules of operating the ISC scheme are the following.
A) Pick an initial arbitrary set of orthonormal transmit vectors.
B) When a data block has been received, construct a new, orthonormal set of transmit vectors as a linear combination of the columns of the receive data block, according to the rules explained below.

Addressing the issue of additive channel noise, we will add a rule or principle.
C) Try to obtain knowledge about your received symbol block *up to an orthogonal rotation*.

Using this information in constructing your next transmit block, one will minimize the effect of the channel noise in the task of obtaining a better set of transmit–receive vectors, as will be explained in detail below.

We consider first the situation without channel noise, but return to this issue later. In the following, $R\{\mathbf{A}\}$ denotes the column subspace associated with a matrix $\mathbf{A}$.

ISC Algorithm:
1. Let $\mathbf{C_U}(0) \in C^{K \times n}, K \leq K_0$ be the first uplink symbol block and let $\mathbf{T_U}(0) \in C^{M \times K}$ be the first orthonormal guess of the UL transmit vectors. Set $i = 0$.
2. Transmit uplink through the channel $\mathbf{H}$

$$\begin{aligned} \text{SU}: \quad &\text{transmit} \quad \mathbf{X}(i) = \mathbf{T_U}(i)\mathbf{C_U}(i) \\ \text{BTS}: \quad &\text{receive} \quad \mathbf{Z}(i) = \mathbf{H}\mathbf{X}(i). \end{aligned}$$

3. BTS: From the block $\mathbf{Z}(i)$, decide the symbols $\hat{\mathbf{C}}_U(i)$.
4. BTS: Create a new set of K downlink orthonormal transmit vectors held as column vectors in $\mathbf{T}_D(i)$ by linearly combining the columns of $\mathbf{Z}(i)$ such that

$$R\{\mathbf{T_D}(i)\} = R\{\mathbf{Z}(i)\}. \quad (13)$$

5. BTS: Transmit downlink through the complex conjugate channel $\mathbf{H}^H$

$$\begin{aligned} \text{BTS}: \quad &\text{transmit} \quad \mathbf{Y}(i) := \mathbf{T_D}(i)\mathbf{C_D}(i). \\ \text{SU}: \quad &\text{receive} \quad \mathbf{W}(i+1) = \mathbf{H}^H \mathbf{Y}(i). \end{aligned}$$

6. SU: From the block $\mathbf{W}(i+1)$, decide the symbols $\hat{\mathbf{C}}_D(i)$.
7. SU: Create a new set of K orthonormal uplink transmit vectors held as column vectors in $\mathbf{T_U}(i+1)$ by linearly combining the columns of $\mathbf{W}(i+1)$ such that

$$R\{\mathbf{T_U}(i+1)\} = R\{\mathbf{W}(i+1))\}. \quad (14)$$

8. Increase i, and repeat from 2.

The results given in the following paragraphs show the convergence of the transmit vectors towards bases for the leading left and right subspaces of $\mathbf{H}$ under typical circumstances. To this end, consider first a simpler algorithm, not dealing with actual symbol transmission.

Two-Way QR Power Iterations:
1) Set $\mathbf{T}_U(0) \in C^{M \times K} =$ random orthogonal, $\quad i = 0$.

2) $\mathbf{Y}(i) = \mathbf{H}\mathbf{T}_U(i)$.
3) $[\mathbf{T}_D(i), \mathbf{R}] = qr(\mathbf{Y}(i))$.
4) $\mathbf{X}(i+1) = \mathbf{H}^H\mathbf{T}_D(i)$.
5) $[\mathbf{T}_U(i+1), \mathbf{R}] = qr(\mathbf{X}(i+1))$.
6) Increase i, and repeat from 2.

Here, qr denotes a QR orthonormalization step [9]. Both the unique positive diagonal-term version of the QR decomposition or a nonunique version without this requirement could be used.

We should make the extra requirement for both the algorithms above and for others following, that the columns of the initial matrix $\mathbf{T}_U(0)$ have some component in the leading K-dimensional row subspace of $\mathbf{H}$, e.g., that $\mathbf{V}_K^H\mathbf{T}_U(0) \neq 0$. This is to avoid the possibility that the basis search is conducted outside the subspace of interest, as explained for power iterations more generally in [9]. However, in practice, roundoff errors, and later on, added channel noise, will make this scenario implausible. When channel noise is added, the theoretical possibility that $\mathbf{V}_K^H\mathbf{T}_U(i) = 0$ for $i > 0$ also arises; however, the same variability in the noise subsequently will normally compensate for this deficit in the next iteration. An equivalent argument holds for $\mathbf{T}_D(i)$ relative to $\mathbf{U}_K$. Unless explicitly stated, we shall assume the transmit vectors to have some components in the subspaces of interest. In the following, we shall only prove results for estimating one set of singular vectors, observing that equivalent results hold for estimation of the other.

Theorem 1: Convergence of the two-way power method. If the K and the $K+1$'th singular values of $\mathbf{H}$ are distinct, then in the two-way power iterations, $\mathbf{T}_U(i) \rightarrow \mathbf{V}_K \cdot \Lambda_U$ and $\mathbf{T}_D(i) \rightarrow \mathbf{U}_K \cdot \Lambda_D$, where Λ_U, Λ_D are diagonal matrices in $C^{K \times K}$ with unit-modulo elements.

This result was shown in [5] and could also be elaborated via the *natural power method* described in [10]. It states that the column vectors held in the matrices $\mathbf{T}_U(i)$ and $\mathbf{T}_D(i)$ will converge towards the leading right and left singular vectors respectively, up to a multiplication by a constant number of unit modulo, which is due to the scalar ambiguity inherent in the eigenestimation. This produces an ambiguity in the symbol stream decoding which is typical for blind methods, and which is circumvented using differential decoding schemes. The matrix $\mathbf{R}$ in steps 3 and 5 is the upper triangular matrix of the QR decomposition, and is of no further importance in this procedure. The above algorithm is a straightforward generalization of the QR-based power iterations for symmetric matrices [9] and is also know as the NIPALS algorithm in the field of chemometrics [22]. A new and important generalization of Theorem 1 is given below. In the subsequent theorems, we shall assume that the K and $K+1$th singular values are distinct unless explicitly stated.

Theorem 2: If the QR decomposition in steps 3 and 5 of the two-way power method is replaced by some other orthogonal decomposition of $\mathbf{Y}(i)$ and $\mathbf{X}(i+1)$, respectively, $R\{\mathbf{T}_U(i)\} \rightarrow R\{\mathbf{V}_K\}$ and $R\{\mathbf{T}_D(i)\} \rightarrow R\{\mathbf{U}_K\}$ as $i \rightarrow \infty$.

The proof is given in the Appendix. The immediate consequence of this theorem is that we do not rely on using the QR method as our decomposition method if we are content with convergence in a *subspace sense*. Any other orthogonal method, such as eigenvector/spectral decomposition or the polar decomposition [9], will do equally well. As we shall see, this opens for

a number of possibilities in designing SNR-improving communication schemes, as the above results can be extended to cover actual symbol transmission situations.

Theorem 3: The ISC algorithm is converging in the sense that $R\{\mathbf{T}_U(i)\} \rightarrow R\{\mathbf{U}_K\}$ and $R\{\mathbf{T}_D(i)\} \rightarrow R\{\mathbf{V}_K\}$ as $i \rightarrow \infty$.

The proof is given in the Appendix.

This theorem addresses subspaces spanned by leading singular vectors rather than the vectors themselves. Relating this to the classic MIMO eigenmode transmission of Section II-B, one may ask, "Why is this subspace important, when it is the *vectors* that are *de facto* used for communication?" The simplicity of eigenmode TDD communications seems to lie in the fact that one party only needs to know one set of singular vectors each. As explained in Section II-B, a right singular *vector* $\mathbf{u}_i$ is translated into a scaled left singular *vector* $\sigma_i\mathbf{v}_i$ upon multiplication by the channel matrix $\mathbf{H}$. When only subspace information is available, and not the singular vectors themselves, this property is lost. If the matrix $\mathbf{T}_U(i)$ is not to $\mathbf{V}_K$, but say $\mathbf{T}_U(i) = \mathbf{V}_K\mathbf{Q}$, where $\mathbf{Q} \in C^{K \times K}$ is an orthogonal matrix, then the received symbol block will appear rotated prior to decoding. In turn, this means that some blind method must be used to find the missing rotation so that the symbols can be correctly decided in steps 3 or 6 in the ISC algorithm.

Value of Subspace Information: An important advantage of MIMO eigenmode communication, however, remains intact even when the singular transmit/receive vectors are not perfectly known themselves: Communication using *linear combinations* of the leading singular vectors has an SNR-improving effect relative to using random vectors, or e.g., vectors constituting an inverse channel matrix. Furthermore, as we shall see, even "erroneous" symbol block estimation, i.e., estimation up to a orthogonal factor, has important benefits for the increasing convergence speeds and reducing the impact of the noise in the subsequent blind symbol decision step, as will be shown.

Fig. 2 illustrates the usefulness of knowing the leading singular left and right subspaces even if the singular *vectors* themselves are not known. Consider the situation where the parties communicate over a random 6×6 MIMO channel. Two random sets of three orthogonal, six-dimensional complex vectors are held in the matrices $\mathbf{T_U}(i)$ and $\mathbf{T_D}(i)$, with i denoting the iteration index. White additive noise is included in the model, and the matrices $\mathbf{T_U}(i)$ and $\mathbf{T_D}(i)$ are used for prefiltering as well as postfiltering of the received data block, in the same way as $\mathbf{V}_K$ and $\mathbf{U}_K$ were used both for prefiltering and postfiltering in the classic MIMO eigenmode communications. Upon convergence, however, $\mathbf{T_U}(i)$ and $\mathbf{T_D}(i)$ will only estimate $\mathbf{V}_K$ and $\mathbf{U}_K$ up to orthogonal transformations, and the symbol block will appear rotated at the decoder. To resolve this, the JADE [2] algorithm was used for symbol stream separation. Furthermore, since the introduction of noise increases the effective dimensionality of $\mathbf{Z}(i)$ and $\mathbf{W}(i)$, the computation of their subspace bases in steps 3 and 6 is replaced by a basis approximation. This was done by performing a QR decomposition of the autocorrelation matrix at each station, using the K vectors associated with the largest diagonal element of the upper triangular matrix as the prefilter to use, letting them serve as (approximate) basis vectors for $R\{\mathbf{Z}(i)\}$. In Section III-A-4), we explain why this QR-based method works so well. The figure shows the bit error
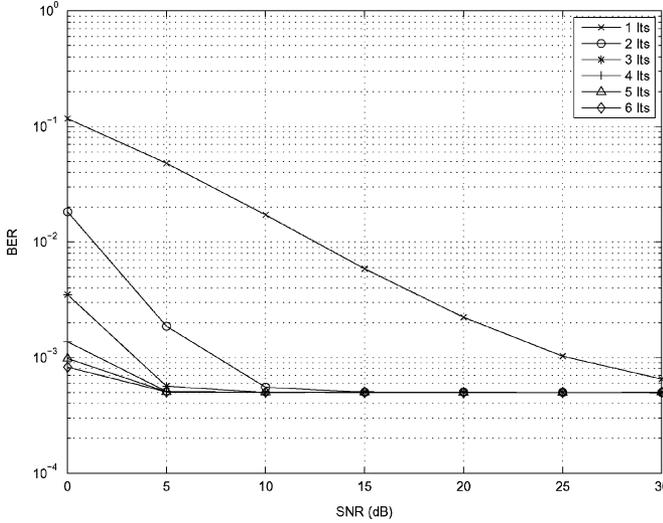
Fig. 2. Average bit error rates at the various iteration steps of the ISC framework. The autocorrelation matrix of the receive data block was used to extract K orthogonal prefilters via a QR decomposition. Each curve in the graph corresponds to the average error after a given number of iterations.

rates at the various iteration steps of the algorithm at various SNRs. As the iterations advance, the average BER drops. The results are based on 10 000 Monte Carlo runs.

A consequence of all this is that the ISC scheme can be seen as a way *to prepare* the transmission prefilters before classic (blind) estimation methods for symbol detection are put to use.

D. Channel Noise Effect

Having shown convergence of the ISC scheme for the deterministic case in the proofs of Theorems 2 and 3, we now examine its convergence properties in the presence of noise. In particular, we want to study the deviation between the deterministic process and the corresponding stochastic process throughout the iterations. Mathematically, we modify the steps 2 and 5 to become

$2'$. BTS : receive $\mathbf{Z}(i)' = \mathbf{HX}(i) + \mathbf{N}_U(i)$

$5'$. SU : receive $\mathbf{W}(i+1)' = \mathbf{H}^H \mathbf{Y}(i) + \mathbf{N}_D(i)$

where $\mathbf{N}_U(i)$ and $\mathbf{N}_D(i)$ are white noise matrices in $C^{N \times n}$ and $C^{M \times n}$, respectively. A particularly important issue arises in steps 4 and 7 of the ISC scheme. In the deterministic case, the task "to find a set of K orthonormal basis vectors" spanning the column spaces $R\{\mathbf{Y}(i)\}''$ (or $R\{\mathbf{W}(i+1)\}$) was straightforward. However, it is not as well defined for the stochastic case. In the deterministic case, the rank of the column spaces of the matrices $\mathbf{Y}(i)$ and $\mathbf{W}(i+1)$ was exactly equal to K (assuming $\mathbf{C}_U$ and $\mathbf{C}_D$ to be of full rank). Now, we have to compute K-dimensional bases for the K_0-dimensional column spaces instead. This naturally leads to the question of approximation of data in a subspace and how different basis-approximation methods perform. This question is beyond the scope of the present paper. On a general basis, however, computing basis vectors for the column space of the noise-corrupted receive block $\mathbf{Z}(i)'$ by linearly combining its columns can always be expressed as

$$\mathbf{T}_D(i) = \mathbf{Z}(i)' \Psi^H \qquad (15)$$

where Ψ is some matrix in $C^{K \times n}$ such that $\mathbf{T}_D(i)$ becomes orthogonal. The trick in establishing subsequent results is to observe how the ISC algorithm reduces to the two-way power method. We write out the above product as

$$
\begin{aligned}
\mathbf{Z}(i)' \Psi^H &= [\mathbf{HX}(i) + \mathbf{N}_U(i)]\Psi^H \\
&= [\mathbf{HT}_{\mathbf{U}}(i)\mathbf{C}_{\mathbf{U}}(i) + \mathbf{N}_U(i)]\Psi^H \\
&= \mathbf{HT}_{\mathbf{U}}(i)\mathbf{C}_{\mathbf{U}}(i)\Psi^H + \mathbf{N}_U(i)\Psi^H \\
&= \mathbf{HT}_{\mathbf{U}}(i)\mathbf{P} + \mathbf{N}_2(i) \qquad (16)
\end{aligned}
$$

where $\mathbf{N}_2(i) = \mathbf{N}_U(i)\Psi^H$ and $\mathbf{P} = \mathbf{C}_{\mathbf{U}}(i)\Psi^H$ is assumed invertible for reasons that will be discussed. We see that the last line is equivalent to step 2 in the two-way power method, with the exception of the presence of the matrices $\mathbf{P}$ and $\mathbf{N}_2(i)$. Note first, that if there was no channel noise, e.g., $\mathbf{N}_2(i) = 0$, then the matrix $\mathbf{P}$ has no influence on the subspace convergence (as long as it is of full rank). What we now have to study is the *combined* effects of the presences of $\mathbf{P}$ and $\mathbf{N}_2(i)$ in the equations. The two main issues we will deal with is 1) how to minimize the effect of the noise term and 2) showing that the error in the subspace estimates is limited by the norm of the (effective) noise matrix. The former is an important issue for the engineer, seeking to maximize the effect of the subspace convergence which is the subject of the paper. The latter issue is important both theoretically and practically: Assuming the central claim of this paper to be true—that there *is* such a thing as a subspace convergence gradually driving the transmit vectors towards a subspace with desirable signal-to-noise properties—this would be of little value if the channel noise associated with a single cycle could result in a complete distortion of that convergence. To simplify the elaborations, we can put the "distortion terms" $\mathbf{P}$ and $\mathbf{N}_2(i)$ into a common additive noise factor. To see this, note that

$$\mathbf{HT}_{\mathbf{U}}(i)\mathbf{P} + \mathbf{N}_2(i) = [\mathbf{HT}_{\mathbf{U}}(i) + \mathbf{N}_3(i)]\mathbf{P} \qquad (17)$$

where $\mathbf{N}_3(i) = \mathbf{N}_2(i)\mathbf{P}^{-1}$. Now, it follows from the invertibility of $\mathbf{P}$ that

$$R\{[\mathbf{HT}_{\mathbf{U}}(i) + \mathbf{N}_3(i)]\mathbf{P}\} = R\{[\mathbf{HT}_{\mathbf{U}}(i) + \mathbf{N}_3(i)\}. \qquad (18)$$

Hence, in deriving properties for the column subspace $R\{\mathbf{HT}_{\mathbf{U}}(i)\mathbf{P} + \mathbf{N}_2(i)\}$, we can without loss of generality derive those properties for the simpler form $R\{[\mathbf{HT}_{\mathbf{U}}(i) + \mathbf{N}_3(i)\}$. Most subspace distortion results, and indeed those we will include, depend on some norm of the error term, in this case $\mathbf{N}_3(i)$. Approaching our core distortion argument, we consider the following theorem.

Theorem 4: Let

$$\mathbf{N}_3(i) = \mathbf{N}_2(i)\mathbf{P}^{-1} = \mathbf{N}_U(i)\Psi^H[\mathbf{C}_{\mathbf{U}}(i)\Psi^H]^{-1}$$

be the effective noise term in estimating an orthonormal subspace basis for $R\{[\mathbf{HT}_{\mathbf{U}}(i) + \mathbf{N}_3(i)\} = R\{\mathbf{Z}(i)'\Psi^H\}$, assume $[\mathbf{C}_{\mathbf{U}}(i)\Psi^H]$ to be invertible, and the rows of $\mathbf{C}_{\mathbf{U}}(i)$ to be mutually orthogonal and of the same L_2 norm l, $\mathbf{C}_{\mathbf{U}}(i)\mathbf{C}_{\mathbf{U}}(i)^H = l^2 \cdot \mathbf{I}$. The expected Frobenious norm $E\{\|\mathbf{N}_3(i)\|_F\}$ is minimized over all choices of Ψ given $\mathbf{C}_{\mathbf{U}}(i)$ if and only if $R\{\Psi^H\} = R\{\mathbf{C}_{\mathbf{U}}(i)^H\}$.

The proof is given in the Appendix.

Let us consider what this theorem states. First, the assumption that $\mathbf{C_U}(i)$ has orthogonal rows seems realistic, given that the symbol blocks are sufficiently long and the transmitted data streams' independence. Next, any orthonormal basis for the column space of $\mathbf{Z}(i)' \in C^{M \times n}$ can be found by post-multiplying it with a suitable matrix $\Psi \in C^{K \times n}$ complex conjugate transposed. Finally, the invertibility of $\mathbf{C_U}(i)\Psi^H$ relates to maintaining full dimensionality in the column spaces $\mathbf{T_D}(i)$, and $\mathbf{T_U}(i)$. If invertibility was not the case, the implication would have to be that some component of the row spaces of Ψ and $\mathbf{C_U}(i)$ were perpendicular. This, in turn, would lead to a rank reduction in (16), representing a situation where one of the communication channels is effectively "lost." This case will not be considered in further detail here but could imply either that the noise has completely corrupted the signal or that the assumption about the transmit vectors having components in the important singular subspaces is broken. What is stated in the theorem, then, is that if the matrix Ψ, or rather, *its row subspace*, has a particular relation with the *row subspace* of the symbol block $\mathbf{C_U}(i)$, the distortion is at its minimum.

Relations With Classic Subspace Analysis: The above has an interesting consequence for improving the success of blind symbol detection algorithms. In constructing a suitable (and optimal) basis-generating matrix Ψ, it is sufficient to have knowledge of the *row subspace* of the symbol block $\mathbf{C_U}(i)$, which can be estimated straightforwardly using e.g., a subspace analysis of the received data block. This is a somewhat unusual result for a blind method: In e.g., decision-directed (DD) estimation, convergence of the channel parameters towards the correct optimum will fail unless the symbols are decided correctly, and *vice versa* for the symbols with respect to the channel estimate. Here, we are in a situation where there is some convergence, if only towards a desired *subspace*, even if the symbol block is not perfectly estimated. That subspace, in turn, is associated with high SNR, increasing the chance of correct (blind) symbol detection. Contrasting our method with other blind schemes, the value of performing a subspace analysis of the received data block differs. Subspace analysis is normally used as a means of obtaining *partial channel information*, such as, e.g., one set of singular vectors and singular values, as in [11] or [21]. Obtaining the remaining set of singular vectors is typically achieved using statistics (HOSs), which becomes an extra computational burden. In the context of the ISC framework, the situation is slightly different. If, e.g., Ψ is chosen such that

$$\Psi^H = \underbrace{\mathbf{C_U}(i)^H \Theta}_{\text{Rotated block}} \cdot \underbrace{\mathbf{R}}_{\text{Orthogonalizer}} \qquad (19)$$

where the term $\mathbf{C_U}(i)^H \Theta$, for an orthogonal $\Theta \in C^{K \times K}$, is not the symbol block itself but a *rotated version of it*, then the error-minimizing property still holds. Hence, the introduction of a subspace analysis, e.g., through an SVD, has important benefits without invoking the additional costs associated with higher order statistics.

Below, we shall see that a number of previously published schemes indeed inherit such error-minimizing properties, al-

though the designers might not have been aware of it. Having established the form of the noise/distortion term and how to minimize it, we recapture some subspace perturbation results from Wedin for quantifying these errors. We effectively address issue (2) above in showing that the effect of the distortion term on the subspace estimation process is bounded. Dealing with distances between subspace, the following definition from [20] is useful.

Definition 1: Let L and M be two subspaces in the complex field, and let $\mathbf{P}_L, \mathbf{P}_M$ be projection operator matrices onto those spaces. Then, a measure of the distance between those subspaces are

$$d(L, M) = \|\sin \theta(L, M)\| = \|(I - \mathbf{P}_M)\mathbf{P}_L\| \qquad (20)$$

for any unitary invariant norm.

Theorem 5: Assume that $\mathbf{N}_3(i) = \mathbf{N}_U(i)\Psi^H[\mathbf{C_U}(i)\Psi^H]^{-1}$ is the effective distortion term in estimating an orthogonal basis for $\mathbf{Z}(i)'$ Then

$$d(R\{\mathbf{HT_U}(i)\}, R\{\mathbf{HT_U}(i) + \mathbf{N}_3\}) \leq \frac{\|\mathbf{N}_3\|_F}{\delta} \qquad (21)$$

where $\delta = \sigma_{\min}(\mathbf{HT_U}(i) + \mathbf{N}_3)$.

The proof is given in the Appendix.

We claim that the subspace estimation error *at each iteration step* is bounded by the Frobenious norm of the effective noise term $\|\mathbf{N}_3\|_F$. The denominator term δ is a measure of smallest singular value in the matrix $\mathbf{HT_U}(i)$ after noise has been added to it. However, since the effect of additive white noise is usually an *increase* in the singular values, this is not likely to lead to division problems. In total, the smaller the error term $\mathbf{N}_3(i)$ is, and the larger the singular values in $\mathbf{HT_U}(i)$ are, the lower is the subspace estimation error. Clearly, the singular values (with or without noise added) are at their highest when the columns of $\mathbf{T_U}(i)$ lie in the leading right singular subspace of $\mathbf{H}$. The consequence, not surprisingly, is that an optimum choice of $\mathbf{T_U}(i)$, e.g., one where the desired convergence has occurred is a relatively stable condition: When convergence towards the optimum subspace has occurred, subsequent (re-)estimates of this subspace have a lower associated error than was the case prior to convergence.

Convergence of Primitive Schemes: Note finally, that even if the matrix Ψ does not conform with the symbol block subspace principle as stated in Theorem 4, the subspace estimation error will still be bounded by the error term $\mathbf{N}_3(i)$, only it will not be at its minimum. This means that even "naive" ways of constructing orthogonal basis vectors may succeed in converging towards basis vectors for the optimal subspace. Consider, as one example, the choice $\Psi = [\mathbf{e}_1, \mathbf{e}_2, 0, \ldots \ldots, 0]\mathbf{R}$, where $\mathbf{e}_1$ and $\mathbf{e}_2$ are the first and second elementary vectors, and where the postmultiplication matrix $\mathbf{R}$ has the effect of orthonormalizing the product $\mathbf{Z}(i)'[\mathbf{e}_1, \mathbf{e}_2, 0, \ldots \ldots, 0]$. Choosing Ψ this way corresponds simply to extracting the two first column vectors of the receive block $\mathbf{Z}(i)'$ and orthonormalizing them to get a new basis estimate $\mathbf{T_D}(i)$. Even with this simple strategy, there will be a "pull" towards the leading singular subspaces, albeit with a larger effective noise/distortion term than is accomplished when Theorem 4 is fulfilled.

III. RESULTS

A. Fitting Techniques Into the ISC Framework

With the framework in place and its convergence established, we now show how some techniques fit into it. A detailed description of each of the algorithms is outside the scope of the present papers, and for this we refer to the original works. However, the main ideas that they use should be recognizable, as should their place in the suggested ISC framework be. In the following, $\mathbf{P}$ is used to denote various specific matrix operations, and Π denotes variations of the spatial correlation matrix.

1) BIMA: The BIMA [5] works in two steps. First, it tries to estimate the received symbols using the previous set of estimated singular vectors. Then, assuming these estimates to be sufficiently good, it combines them with a QR decomposition of a certain matrix product to obtain a new set of singular vectors. BIMA can be cast into step 4 of the ISC framework as follows: Let $\hat{\mathbf{C}}_U(i)$ be an estimate of $\mathbf{C}_U(i)$. BIMA works by postmultiplying the receive block $\mathbf{Z}(i)'$ by $\hat{\mathbf{C}}_U(i)^\dagger$, $(.)^\dagger$ denoting the Moore–Penrose inverse, which gives

$$
\begin{aligned}
\mathbf{Z}(i)'\hat{\mathbf{C}}_U(i)^\dagger &= [\mathbf{HT}_{\mathbf{U}}(i)\mathbf{C}_{\mathbf{U}}(i) + \mathbf{N}_U(i)]\hat{\mathbf{C}}_U(i)^\dagger \\
&= \mathbf{HT}_{\mathbf{U}}(i)\mathbf{C}_{\mathbf{U}}\hat{\mathbf{C}}_U(i)^\dagger + \mathbf{N}_U(i)\hat{\mathbf{C}}_U(i)^\dagger \\
&= \mathbf{HT}_{\mathbf{U}}(i) + \mathbf{N}_U(i)\hat{\mathbf{C}}_U(i)^\dagger
\end{aligned} \tag{22}
$$

assuming the estimate $\hat{\mathbf{C}}_U(i)$ to be correct. The idea is that if the effective noise term $\mathbf{N}_U(i)\hat{\mathbf{C}}_U(i)^\dagger$ is low, then performing a QR decomposition of $\mathbf{Z}(i)'\hat{\mathbf{C}}_U(i)^\dagger$ is sufficient to obtain a K-dimensional orthogonal basis for $R\{\mathbf{HT}_{\mathbf{U}}(i)\}$, which again represents a cycle in the two-way power method. In order to speed up the convergence, the columns of $\mathbf{Z}(i)'\hat{\mathbf{C}}_U(i)^\dagger$ are sorted in order of decreasing L_2 norm. This sorting can be formalized as postmultiplication by an orthogonal matrix $\mathbf{P}$ consisting of elementary vector columns. Performing a QR decomposition of the column-sorted matrix can be expressed by further postmultiplying by a matrix $\mathbf{R}^{-1}$, so that

$$
\mathbf{T}_{\mathbf{D}}(i) = \mathbf{Z}(i)'\hat{\mathbf{C}}_U(i)^\dagger\mathbf{P}\mathbf{R}^{-1} \tag{23}
$$

where $\mathbf{QR} = \mathbf{Z}(i)'\hat{\mathbf{C}}_U(i)^\dagger\mathbf{P}$ is an economy-version QR decomposition of $\mathbf{Z}(i)'\hat{\mathbf{C}}_U(i)^\dagger\mathbf{P}$. Now consider how this strategy relates to the derived error-minimizing properties of the ISC framework. Finding $\mathbf{T}_{\mathbf{D}}(i)$ within the ISC framework is formalized by setting

$$
\Psi^H = \hat{\mathbf{C}}_U(i)^\dagger\mathbf{P}\mathbf{R}^{-1}. \tag{24}
$$

Provided that $\hat{\mathbf{C}}_U(i)$ is correctly estimated up to premultiplication by any square full-rank matrix, the error-minimizing effect described in Theorem 4 sets in. This is because $R\{\Psi^H\} = R\{\hat{\mathbf{C}}_U(i)^\dagger\} = R\{\hat{\mathbf{C}}_U(i)^H\}$, the latter following from the definition of the Moore–Penrose inverse.

2) PROTEUS: Wen *et al.* [21] proposes to use the PROTEUS algorithm [3] for subspace tracking of the spatial correlation matrix. Extracting the eigenvectors from the spatial correlation matrices at each side is essentially the same as performing an SVD on the received data blocks. PROTEUS explicitly tracks

the eigenvectors of a time-varying covariance matrix by incorporating a forgetting factor. Ignoring the forgetting factor momentarily, the process can be fitted into the ISC framework as follows. Let $\mathbf{Z}(i)' = \Phi\Lambda\Omega^H$ be the economy-version SVD of $\mathbf{Z}(i)'$, and let

$$
\Psi^H = \mathbf{Z}(i)'^H \cdot \mathbf{R}_1 \tag{25}
$$

where $\mathbf{R}_1 = \Phi_K\Lambda_K^{-1}$, with Φ_K containing the first K columns of Φ and Λ_K the K first columns and row of Λ. Using Ψ to extract an orthogonal matrix $\mathbf{T}_{\mathbf{D}}(i)$ amounts to computing the K leading column singular vectors of $\mathbf{Z}(i)'$, which is the same as performing a (column) principal component analysis (PCA) of $\mathbf{Z}(i)'$. But note also that

$$
\begin{aligned}
\Psi^H &= \mathbf{Z}(i)'^H \cdot \mathbf{R}_1 \\
&= [\mathbf{HT}_{\mathbf{U}}(i)\mathbf{C}_{\mathbf{U}}(i) + \mathbf{N}_U(i)]^H\mathbf{R}_1 \\
&= \mathbf{C}_{\mathbf{U}}(i)^H\underbrace{\mathbf{T}_{\mathbf{U}}(i)^H\mathbf{H}^H\mathbf{R}_1}_{\mathbf{P}} + \underbrace{\mathbf{N}_U(i)^H\mathbf{R}_1}_{\mathbf{N}}.
\end{aligned} \tag{26}
$$

Assuming the noise term $\mathbf{N}$ to be small and $\mathbf{P} = \mathbf{T}_{\mathbf{U}}(i)^H\mathbf{H}^H\mathbf{R}_1$ to be full rank, it follows that $R\{\Psi^H\} \approx R\{\mathbf{C}_{\mathbf{U}}(i)^H\}$, e.g., the distortion-minimizing criterion holds approximately. The assumption that $\mathbf{P}$ should be of full rank is plausible as long as the columns of $\mathbf{T}_{\mathbf{U}}(i)$ are not perpendicular to the columns of $\mathbf{H}$, $\mathbf{T}_{\mathbf{U}}(i)^H\mathbf{H}^H$ is nondegenerate, which was previously assumed. The leading principal components extracted through postmultiplication with $\mathbf{R}_1$ will be dominated by the signal part $\mathbf{C}_{\mathbf{U}}(i)^H\mathbf{T}_{\mathbf{U}}(i)^H\mathbf{H}^H$ of Ψ^H, provided the noise is sufficiently small. Concluding, the PROTEUS algorithm, like the BIMA, tends towards distortion minimization in the subspace estimation process. As for the forgetting factor, it makes the above claim a bit less certain—its use implies that the latest basis vector estimates not only are linear combinations of the received data block, but also contain contributions from the column space of previous data blocks. However, for relatively slowly changing channels and/or forgetting factors implying a short memory, both of which are typical situations, the proposition is justified. The convergence occurs not only in a subspace sense, but also on a vector-by-vector basis. In fact, we can show such convergence, even for situations where only a subset of $K \leq K_0$ singular vector pairs are used for communication.

Convergence of the PROTEUS Algorithm: It is relatively straightforward to see that if all $K = K_0$ singular modes are used for communication in PROTEUS, then the estimated basis vectors held in $\mathbf{T}_{\mathbf{U}}, \mathbf{T}_{\mathbf{D}}$ will be close to the singular vectors of some square channel matrix $\mathbf{H}$. This is because the spatial correlation matrix at, say, the BTS, can be computed in the noise-free case as

$$
\begin{aligned}
\frac{1}{n-1}\mathbf{Z}(i)\mathbf{Z}(i)^H &= \mathbf{HT}_{\mathbf{U}}(i)\mathbf{C}_{\mathbf{U}}(i)\mathbf{C}_{\mathbf{U}}(i)^H\mathbf{T}_{\mathbf{U}}(i)^H\mathbf{H}^H \\
&\approx c \cdot \mathbf{HT}_{\mathbf{U}}(i)\mathbf{T}_{\mathbf{U}}(i)^H\mathbf{H}^H \\
&= c \cdot \mathbf{HH}^H
\end{aligned} \tag{27}
$$

where we have used the assumption that independent data streams held as rows in the symbol matrix yield $\mathbf{C}_{\mathbf{U}}(i)\mathbf{C}_{\mathbf{U}}(i)^H \approx c \cdot \mathbf{I}$ where c is some constant, and that

since $\mathbf{T_U}(i)$ is square, the relation $\mathbf{T_U}(i)\mathbf{T_U}(i)^H = \mathbf{I}$ holds true. The left singular vectors of $\mathbf{H}$ are the eigenvectors of $\mathbf{HH}^H$, and hence the singular vectors of $\mathbf{H}$ can be estimated directly from $\mathbf{Z}(i)$. It is also known from the literature that white additive noise does not affect the expectation of the estimated singular/eigenvectors, only their eigenvalues. Hence, the estimation procedure is unbiased with respect to white noise. The next natural question now is, "What happens if only a *a subset* of the singular vectors are used for transmission?" In this case, $\mathbf{T_U}(i)\mathbf{T_U}(i)^H \neq \mathbf{I}$, and a central condition is broken. However, let us assume that the columns of $\mathbf{T_U}(i)$ are indeed in the leading right K-dimensional singular subspace of $\mathbf{H}$ or that $R\{\mathbf{T_U}(i)\} = R\{\mathbf{V}_K\}$. We can then apply the following theorem.

Theorem 6: Assume the rows of $\mathbf{C_U}(i)$ to be orthogonal and of equal L_2-norm l, and $R\{\mathbf{T_U}(i)\} = R\{\mathbf{V}_K\}$. Then, the leading K left singular vectors of the matrix $\mathbf{Z}(i)$ are equal to the leading K right singular vectors of the matrix $\mathbf{H}$ up to multiplications by a scalar.

The proof is given in the Appendix.

The theorem states is that if the transmit vectors are already in the right subspace, the correct singular vector estimates will be obtained by extracting the K first eigenvectors of the spatial correlation matrix. But the "if" part is precisely what the ISC scheme, wherein PROTEUS is a natural member, provides. Clearly, the effects of noise and the question about the orthogonality of $\mathbf{C_U}(i)$ will affect the precision of the claim, but the tendencies for convergence of the PROTEUS algorithm should be explainable by the above arguments.

3) Liu and Host-Madsen: In the paper [11], a method is presented for communicating both on a single-input multiple-output (SIMO) and on a MIMO TDD channel. Initially, a full block of symbols is transmitted from the base station to the mobile subscriber, to enable the computation of the spatial correlation matrix. At the SU, the top eigenvector of this matrix, corresponding to the top right singular vector $\mathbf{v}_1$ is estimated via an eigenvector-eigenvalue decomposition (EVD) of this matrix and is used for transmitting a binary-phase-shift-keying (BPSK) symbol uplink. It is then shown that for a constant $\mathbf{H}$ matrix, the receive spatial correlation matrix at the BTS is

$$\Pi = \mathbf{u}_1\mathbf{u}_1^H\lambda_1^2 + \sigma^2\mathbf{I} \tag{28}$$

where $(\mathbf{u}_1, \lambda_1)$ is the maximum right singular vector/singular value pair of the channel. In turn, this means that $\mathbf{u}_1$ can be estimated again by an EVD. At the BTS, $\mathbf{u}_1$ is used again for downlink communication, and the spatial correlation matrix at the SU becomes

$$\Pi = \mathbf{v}_1\mathbf{v}_1^H\lambda_1^2 + \sigma^2\mathbf{I} \tag{29}$$

enabling the re-estimation of $\mathbf{v}_1$ in the same fashion. These steps are repeated, and it is argued that the singular vectors will be tracked also for time-varying channels. However, no comment is made as to *why* these singular vectors are traceable by such a simple scheme. Two things can be noted a) that after the initialization, it is practically identical with BIMA using one top

singular mode only, and b) the reason *why* this technique succeeds in *tracking* the channel parameters $\mathbf{v}_1$ and $\mathbf{u}_1$ when $\mathbf{H}$ changes with time is once again down to the ISC.

Until now, we have only investigated published methods in light of the ISC framework. However, the principle also allows construction of new and other simple methods. Two first examples are given below.

4) A QR-Based Method: Returning to the example of Fig. 2, consider the following algorithm. At the BTS, the matrix $\mathbf{Z}(i)'$ is received, and the spatial autocorrelation matrix computed as

$$\Pi = \frac{1}{n-1}\mathbf{Z}(i)'\mathbf{Z}(i)'^H. \tag{30}$$

Next, we sort the columns of Π in decreasing order according to their L_2 norm and perform a reduced QR decomposition where only the K first columns of the orthogonal matrix are retained. These two steps can be expressed as postmultiplication by $\mathbf{P}$, a square column-perturbation matrix, and subsequent multiplication by a matrix $\mathbf{J}$ in $C^{N \times K}$. In total, this yields

$$\mathbf{T_D}(i) = \Pi\mathbf{PJ} = \mathbf{Z}(i)'\underbrace{\left[\frac{1}{n-1}\mathbf{Z}(i)'^H\mathbf{PJ}\right]}_{\Psi^H} \tag{31}$$

which is again on the form of the ISC framework. Note furthermore that

$$\begin{aligned}
\Psi^H &= \frac{1}{n-1}\mathbf{Z}(i)'^H\mathbf{PJ} \\
&= \frac{1}{n-1}[\mathbf{HT_U}(i)\mathbf{C_U}(i) + \mathbf{N}_U(i)]^H\mathbf{PJ} \\
&= \frac{1}{n-1}\mathbf{C_U}(i)^H\mathbf{T_U}(i)^H\mathbf{H}^H\mathbf{PJ} + \mathbf{N}_U(i)^H\mathbf{PJ} \\
&= \frac{1}{n-1}\mathbf{C_U}(i)^H\mathbf{Q}(i) + \mathbf{N}_{UU}(i) \tag{32}
\end{aligned}$$

where $\mathbf{Q}(i) = \mathbf{T_U}(i)^H\mathbf{H}^H\mathbf{PJ}$ and $\mathbf{N}_{UU}(i) = \mathbf{N}_U(i)\mathbf{PJ}$. Assuming the additive noise factor to be small, we can make the approximation $R\{\Psi^H\} \approx R\{\mathbf{C_U}(i)^H\}$, as was the case with PROTEUS. The effect of sorting the columns prior to the QR decomposition is to favor vectors with large norms as candidates for basis vectors in $\mathbf{T_D}(i)$. The rationale behind this is that vectors with large norms are likely to have been "*scaled large*" by previous iterations and are likely already to be in the leading singular subspace. This method still results in transmit–receive filter matrices $\mathbf{T_D}(i), \mathbf{T_U}(i)$, which are rotated relative to $\mathbf{U}_K, \mathbf{V}_K$.

5) A Training-Based Method: So far, the methods we have considered are blind. A natural question is whether the ISC effect has any consequences for training-based methods, or whether training-based schemes could be constructed to take advantage of the ISC. To shed some light on this, consider the following training-based scheme. First, a set of known training symbols held in the matrix $\mathbf{C_U}(i)'$ are transmitted uplink using the prefilters held in $\mathbf{T_U}(i)$, so that

$$\mathbf{X}(i)' = \mathbf{T_U}(i)\mathbf{C_U}(i)' \tag{33}$$

which is received at the BTS as

$$\mathbf{Z}(i)' = \mathbf{H}\mathbf{X}(i) + \mathbf{N}(i)$$
$$= \mathbf{H}\mathbf{T_U}(i)\mathbf{C_U}(i)' + \mathbf{N}(i). \qquad (34)$$

Now, since the training symbols held in $\mathbf{C_U}(i)'$ are known by both parties, we can use the pseudoinverse $\mathbf{C_U}(i)'^\dagger$ to compute

$$\widehat{HT_U(i)} = \mathbf{Z}(i)'\mathbf{C_U}(i)'^\dagger$$
$$= [\mathbf{H}\mathbf{T_U}(i)\mathbf{C_U}(i)']\mathbf{C_U}(i)'^\dagger + \mathbf{N}(i)\mathbf{C_U}(i)'^\dagger$$
$$= \mathbf{H}\mathbf{T_U}(i) + \mathbf{N}(i)\mathbf{C_U}(i)'^\dagger. \qquad (35)$$

Assuming the noise to be sufficiently small, and computing an orthonormal basis for the result $\mathbf{Z}(i)'\mathbf{C_U}(i)'^\dagger \in C^{N \times K}$ by using a QR decomposition—or indeed by using any other orthogonal basis for $\mathbf{Z}(i)'\mathbf{C_U}(i)'^\dagger$, we are once again invoking the convergence of the two-way power method. The same procedure is repeated in the DL direction.

B. Simulation Results

Evaluating the latter training-based method, a two-way MIMO communication scheme is constructed, where a pilot sequence is sent and the channel is estimated at the receiver side prior to decoding of the payload data. Both the training sequence and the actual data blocks have less rows/independent (K) streams than the maximum number (K_0). Each symbol block then has to be premultiplied by a $N \times K$ transmit vector block, $\mathbf{T_D}(i)$ or $\mathbf{T_U}(i)$ on the base stations and subscriber units, respectively. The effective uplink and downlink channels then become $\mathbf{H}\mathbf{T_U}(i)$ and $\mathbf{H}^H\mathbf{T_D}(i)$. We now compare two different settings, as follows:

  a) the one where $\mathbf{T_U}(i)$ and $\mathbf{T_D}(i)$ are chosen randomly anew in each iteration;
  b) the one where $\mathbf{T_U}(i)$ and $\mathbf{T_D}(i)$ are constructed as (arbitrary) orthonormalized linear combinations of the received data blocks.

The latter is in accordance with rule B) we put up for successful convergence of the vectors towards the top singular subspaces. As it turns out, version b) has the interesting property of mimicking some characteristics of a water-filling algorithm. The idea of water-filling is to use only the singular modes with a sufficiently high receive SNR for communication, and skip subsequent modes. However, to do this, one must perform an SVD of the channel at the transmitter. Here, the estimation of the top eigenmodes (or at least a basis for them) occurs "by itself" without the need for an actual SVD. The exact linear combinations for obtaining the precise singular vectors may not be at hand, but using a *linear combination* of the best singular vectors is a good second best to using the vectors themselves. Version b) of the scheme could be properly described as a "poor man's water-filling method," taking a limited advantage of its main characteristics (SNR maximization), while avoiding its highest costs.

Fig. 3 shows the results, using a 6×6 MIMO channel, transmitting on $K = 3$ streams, with a block length of 200 symbols, and a training rate at 20% of the channel capacity over 500 Monte Carlo runs, each with 20 transmitted blocks. The upper curve shows performance without the subspace scheme,
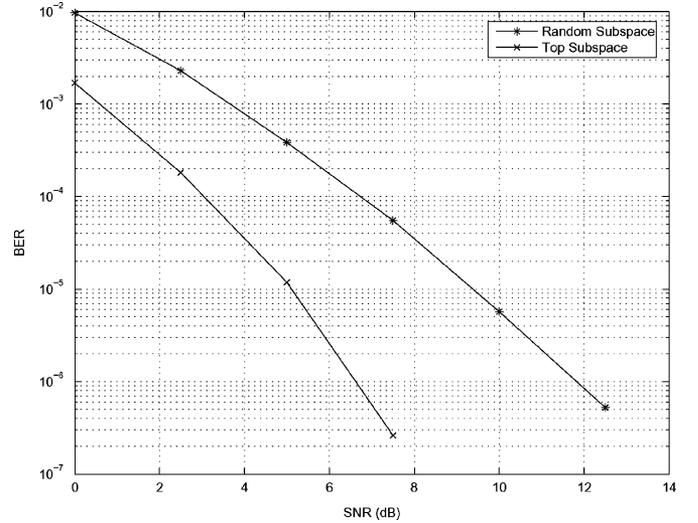
Fig. 3. Bit error rates for a 6×6 MIMO TDD scheme, communicating on $K = 3$ modes, with the use of subspace power iterations and without.

the lower using it. The initial five blocks were considered an acquisition period for the subspace basis vectors to converge and were, hence, kept out of the computations. In the presence of noise, the top singular subspaces are the ones that provide the highest SNRs. Hence, it is not surprising that the method using "top subspaces" outperforms the "random subspace" method.

IV. DISCUSSION

We have presented a framework for utilizing the impact of the power method for communicating in a subspace with maximum SNR. This was accomplished by constructing the transmit vector blocks in accordance with some simple rules enabling the *intrinsic subspace convergence* to start operating. It was shown that several published methods rely on or could benefit from the ISC effect. The stability of the framework and robustness against noise was demonstrated both by mathematical analysis and by simulations. Although the theory of the framework was developed for constant channels $\mathbf{H}$, channels that are slowly or intermediately time varying can also be tracked: At each iteration step, the estimates of the basis vectors are realized from a process containing the latest channel. If the time variation is not too fast, e.g., if the convergence rate of the algorithm is faster than the speed of the channel time variation, tracking will be successful, as shown in previous literature [5].

A typical way of assessing a communication scheme is by its symbol or bit error rates. Two typical sources of symbol decision errors are noise and faulty channel estimation. Furthermore, many adaptive schemes contain an inherent dilemma: "If the symbol decision is poor, so will the channel parameter estimates be, and vice versa." The ISC has the potential of obtaining good channel parameter estimates without good symbol decisions. It improves transmission schemes by allowing communication in the "best subspace," the one maximizing the receive SNR. The convergence towards these top singular modes is robust, in the sense that as long as the transmit–receive vector design is made linear, a *convergence pull* occurs regardless of other factors, such as success rate of the symbol decision. This

is in contrast with, e.g., decision-directed estimation, where erroneous symbol decision might lead to erroneous channel estimation *ad infinitum*.

A. Further Research

A large number of methods for blind, semiblind, or training-based communication could be designed within the ISC framework. Subspace-boosted DD estimation for channel tracking is but one obvious example.

More work needs to be done and more results derived to assess the success of the eigenestimation for various algorithms within the framework, and also for various channel settings. We have assumed the symbol blocks to have independent streams, but one could ask what would happen if they were not. Furthermore, colored channel noise will affect the estimated basis vector in each step. The overall performance effect of this is not clear, but deficiency bounds could possibly be derived using results such as Wedin's theorem. Simulations and previous methods implemented on the basis of the eigenestimation effects do show good promise.

APPENDIX

Proof of Theorem 2: Consider step 3 of the two-way power method. In replacing the QR method with another orthogonal decomposition, the term $[\mathbf{T}_D(i), \mathbf{R}]$ is effectively replaced with some other term $[\mathbf{T}_D(i)', \mathbf{R}']$, where $\mathbf{R}$ is not necessarily upper triangular. However, since both $\mathbf{T}_D(i)$ and $\mathbf{T}_D(i)'$ must be orthogonal matrices spanning $R\{\mathbf{Y}(i)\}$, there must exist some matrix $\mathbf{Q} \in C^{K \times K}$ such that $\mathbf{T}_D(i)' = \mathbf{T}_D(i)\mathbf{Q}$. In step 4 then, $\mathbf{X}(i+1) = \mathbf{H}^H \mathbf{T}_D(i)' = \mathbf{H}^H \mathbf{T}_D(i)\mathbf{Q}$. In step 5, finding an orthogonal basis for $R\{\mathbf{X}(i+1)\}$ by QR or some other method, the presence of the matrix $\mathbf{Q}$ becomes irrelevant *in the subspace sense*. Finding an orthogonal basis for $R\{\mathbf{H}^H \mathbf{T}_D(i)\mathbf{Q}\}$, or finding it for $R\{\mathbf{H}^H \mathbf{T}_D(i)\}$ is equivalent, since the subspaces are the same. Since this relation holds in every step of the algorithm in both directions, it also holds upon convergence.

Proof of Theorem 3: Consider step 4 in the ISC framework. Computing an orthonormal basis for the columns of $\mathbf{Z}(i)$ can be done by postmultiplying by some matrix $\Psi \in C^{K \times n}$, as previously argued. Writing out, we have

$$\mathbf{T}_D(i) := \mathbf{Z}(i)\Psi^H = \mathbf{H}\mathbf{T}_U(i)\mathbf{C}_U(i)\Psi^H$$
$$= \mathbf{H}\mathbf{T}_U(i)\mathbf{P}(i) \tag{36}$$

where $\mathbf{P}(i) = \mathbf{C}_U(i)\Psi^H$ is in $C^{K \times K}$, and according to the definitions of the ISC framework, Ψ is chosen so that $\mathbf{T}_D(i)$ is orthogonal. Compare the last line in (36) with steps 2 and 3 in the two-way QR power Iterations. Comparing with step 2, the matrix $\mathbf{Y}(i)$ is on identical form, with the exception of the multiplication by the matrix $\mathbf{P}$. Following the same reasoning as in the proof of Theorem 2, the matrix $\mathbf{P}$ does not affect convergence in the subspace sense, and the theorem follows.

Proposition 1: Let $\mathbf{A}, \mathbf{B}$ in $C^{N \times M}$ be orthogonal matrices, $\mathbf{A}^H \mathbf{A} = \mathbf{B}^H \mathbf{B} = \mathbf{I}$, where $N \geq M$. Then, $R\{\mathbf{A}\} = R\{\mathbf{B}\} \Leftrightarrow \mathbf{A} = \mathbf{B}\mathbf{Q}$ for some orthogonal matrix $\mathbf{Q} \in C^{M \times M}$.

Proof: The *if* part is straightforward: for the *only if* part, note that $R\{\mathbf{A}\} = R\{\mathbf{B}\} \Rightarrow \mathbf{A} = \mathbf{B}\mathbf{X}$ for some invertible matrix $\mathbf{X} \in C^{M \times M}$. But $\mathbf{A}$ is orthogonal, and hence $\mathbf{I} = \mathbf{A}^H \mathbf{A} = \mathbf{X}^H \mathbf{B}^H \mathbf{B}\mathbf{X} = \mathbf{X}^H \mathbf{X}$, which means that $\mathbf{X}$ is orthogonal.

Definition 2: For a fixed orthogonal matrix $\mathbf{B} \in C^{N \times M}$ and $N \geq M$, let S_B be the set of orthogonal matrices in $C^{N \times M}$, such that $\mathbf{A} \in S_B$ if and only if $\mathbf{A}^H \mathbf{B}\mathbf{B}^H \mathbf{A}$ is invertible.

Proposition 2: Define $f(\mathbf{A}|\mathbf{B}) = \text{tr}\{(\mathbf{A}^H \mathbf{B}\mathbf{B}^H \mathbf{A})^{-1}\}$ for fixed orthogonal $\mathbf{B} \in C^{N \times M}, N \geq M$. The minimum value of f over $\mathbf{A} \in S_B$ is M.

Proof: We can rewrite $\mathbf{f}(\mathbf{A}|\mathbf{B}) = \sum_{i=1}^{M} \lambda_i(\mathbf{A}|\mathbf{B})^{-1}$, where $\lambda_i(\mathbf{A}|\mathbf{B})$ are the eigenvalues of $\mathbf{A}^H \mathbf{B}\mathbf{B}^H \mathbf{A}$. The largest eigenvalue of $\mathbf{A}^H \mathbf{B}\mathbf{B}^H \mathbf{A}$ is

$$\max_{\|x\|=1} \mathbf{x}^H \mathbf{A}^H \mathbf{B}\mathbf{B}^H \mathbf{A}\mathbf{x} = \|\mathbf{B}^H \mathbf{A}\mathbf{x}\|_2^2$$
$$\leq \|\mathbf{B}^H\|_2^2 \|\mathbf{A}\|_2^2 \|\mathbf{x}\|_2^2 = 1. \tag{37}$$

Hence, $f(\mathbf{A}|\mathbf{B}) \geq M$. To show that there exist $\mathbf{A}$ for which $f(\mathbf{A}|\mathbf{B}) = M$, take $\mathbf{A} = \mathbf{B}\mathbf{Q}$, where $\mathbf{Q}$ is orthogonal, and observe that $\mathbf{A}^H \mathbf{B}\mathbf{B}^H \mathbf{A} = \mathbf{Q}^H \mathbf{B}^H \mathbf{B}\mathbf{B}^H \mathbf{B}\mathbf{Q} = \mathbf{I}$, implying all eigenvalues are 1 and $f(\mathbf{A}|\mathbf{B}) = M$.

Proposition 3: For fixed orthogonal $\mathbf{B} \in C^{N \times M}, N \geq M$, $f(\mathbf{A}|\mathbf{B}) = \text{tr}\{(\mathbf{A}^H \mathbf{B}\mathbf{B}^H \mathbf{A})^{-1}\}$ is minimized over $\mathbf{A} \in S_B$ $\Leftrightarrow R\{\mathbf{A}\} = R\{\mathbf{B}\}$.

Proof: The *if* part follows from the proof of Proposition 2 via Proposition 1. For the *only if* part, note that $f(\mathbf{A}|\mathbf{B})$ is minimum when another function

$$g(\mathbf{A}|\mathbf{B}) = \text{tr}\mathbf{A}^H \mathbf{B}\mathbf{B}^H \mathbf{A} \tag{38}$$

is maximum. This is because $g(\mathbf{A}|\mathbf{B}) = \sum_{i=1}^{M} \lambda_i(\mathbf{A}|\mathbf{B})$ whose maximum is attained when all $\lambda_i(\mathbf{A}|\mathbf{B})$ are maximized and equal to 1 (from Proposition 2). Then also, and only then $\mathbf{f}(\mathbf{A}|\mathbf{B})$ minimum. But we can rewrite $g(\mathbf{A}|\mathbf{B}) = \|\mathbf{B}^H \mathbf{A}\|_F^2 = \sum_{i=1}^{M} \|\mathbf{B}^H \mathbf{a}\|_i^2$, where $\{\mathbf{a}_i\}$ are the columns of $\mathbf{A}$. If $\mathbf{a}_i \in R\{\mathbf{B}\}$ for all $i \leq 1 \leq M$, the term $\|\mathbf{B}^H \mathbf{a}\|_i$ is 1 for all $i \in [1, M]$ and g is maximum. If some $\mathbf{a}_i$ has a component outside $R\{\mathbf{B}\}$, the term $\|\mathbf{B}^H \mathbf{a}\|_i < 1$ and $g(\mathbf{A}|\mathbf{B})$ is not maximum. Thus, $\mathbf{a}_i \in R\{\mathbf{B}\}$ must hold for all i. This implies $R\{\mathbf{A}\} \subset R\{\mathbf{B}\}$. Since $\mathbf{A}$ and $\mathbf{B}$ are full rank and $\dim(\mathbf{A}) = \dim(\mathbf{B})$, it follows that $R\{\mathbf{A}\} = R\{\mathbf{B}\}$.

Proof of Theorem 4: Let $\Psi = \mathbf{U}\mathbf{S}\mathbf{V}^H$ be a singular value decomposition of $\Psi \in C^{N,m}$ with $\mathbf{U} \in C^{N \times N}, \mathbf{S} \in C^{N \times N}$ and $\mathbf{V} \in C^{n \times N}$, giving

$$\mathbf{N}_3(i) = \mathbf{N}_U(i)\Psi^H[\mathbf{C}_U(i)\Psi^H]^{-1}$$
$$= \mathbf{N}_U(i)\mathbf{V}\mathbf{S}\mathbf{U}^H[\mathbf{C}_U(i)\mathbf{V}\mathbf{S}\mathbf{U}^H]^{-1}$$
$$= \mathbf{N}_U(i)\mathbf{V}\mathbf{S}\mathbf{U}^H[\mathbf{W}\mathbf{S}\mathbf{U}^H]^{-1} \tag{39}$$

where $\mathbf{W} = \mathbf{C}_U(i)\mathbf{V}$ is in $C^{N \times N}$. Hence

$$\mathbf{N}_3(i) = \mathbf{N}_U(i)\mathbf{V}\mathbf{S}\mathbf{U}^H \mathbf{U}\mathbf{S}^{-1}\mathbf{W}^{-1}$$
$$= \mathbf{N}_U(i)\mathbf{V}\mathbf{W}^{-1}$$
$$= \mathbf{N}_{UU}(i)\mathbf{W}^{-1} \tag{40}$$

assuming $\mathbf{W}$ to be invertible, a direct consequence of the theorem's definition, and where $\mathbf{N}_{UU}(i) = \mathbf{N}_U(i)\mathbf{V}$ is now a noise term that will have the same statistical properties as

$\mathbf{N}_U(i)$, since it is merely a rotation of the latter. Now, consider the expected Frobenious norm of the error

$$
\begin{aligned}
E\{\|\mathbf{N}_3(i)\|_F^2\} &= E\{\operatorname{tr}\mathbf{N}_3(i)^H\mathbf{N}_3(i)\} \\
&= E\{\operatorname{tr}\mathbf{W}^{-H}\mathbf{N}_{UU}(i)^H\mathbf{N}_{UU}(i)\mathbf{W}^{-1}\} \\
&= \operatorname{tr}\mathbf{W}^{-H}E\{\mathbf{N}_{UU}(i)^H\mathbf{N}_{UU}(i)\}\mathbf{W}^{-1} \\
&= nN\sigma^2 \cdot \operatorname{tr}\mathbf{W}^{-H}\mathbf{W}^{-1} \\
&= nN\sigma^2 \cdot \operatorname{tr}(\mathbf{W}\mathbf{W}^H)^{-1} \\
&= nN\sigma^2 \cdot \operatorname{tr}(\mathbf{W}^H\mathbf{W})^{-1}
\end{aligned}
\tag{41}
$$

where σ^2 is the noise variance. Assuming $\mathbf{C}_{\mathbf{U}}(i)$ to have orthogonal rows (as stated)

$$
\operatorname{tr}(\mathbf{W}^H\mathbf{W})^{-1} = \operatorname{tr}(\mathbf{V}^H\mathbf{C}_{\mathbf{U}}(i)^H\mathbf{C}_{\mathbf{U}}(i)\mathbf{V})^{-1}.
\tag{42}
$$

Letting $\mathbf{B} = (1/l)\mathbf{C}_{\mathbf{U}}(i)^H$ and $\mathbf{A} = \mathbf{V}$ in Proposition 3 above, the theorem follows by noting that $R\{\Psi^H\} = R\{\mathbf{V}\}$.

Theorem [20]: Let $\mathbf{A}, \mathbf{B}$, and $\mathbf{T}$ be matrices in $C^{M \times K}$, where $K < M$. We define $\mathbf{B} = \mathbf{A} + \mathbf{T}$, where $\mathbf{T}$ represents a perturbation of $\mathbf{A}$. Let $\mathbf{Y}$ and $\mathbf{X}$ be matrices containing orthogonal basis vectors spanning $R(\mathbf{B})$ and $R(\mathbf{B}^H)$ respectively. Let furthermore $\mathbf{R}_1 = -\mathbf{T}\mathbf{X}$ and $\mathbf{R}_2 = -\mathbf{T}^H\mathbf{Y}$. Let $\delta = \sigma_{\min}(\mathbf{B})$ be the smallest singular value of the matrix $\mathbf{B}$, and $\epsilon = \max(\|\mathbf{R}_1\|, \|\mathbf{R}_2\|)$. Then, for every unitary invariant norm

$$
d(R(\mathbf{A}), R(\mathbf{B})) \leq \frac{\epsilon}{\delta}.
\tag{43}
$$

Corollary: Since the matrices $\mathbf{X}$ and $\mathbf{Y}$ merely represent projections of the error term $\mathbf{T}$ onto the row and column subspaces $R\{\mathbf{B}\}$ and $R\{\mathbf{B}^H\}$, it follows that $\|\mathbf{T}\|_F \geq \|\mathbf{R}_1\|_F$ and $\|\mathbf{T}\|_F \geq \|\mathbf{R}_2\|_F$. Hence, $\epsilon \leq \|\mathbf{T}\|_F$, and

$$
d(R(\mathbf{A}), R(\mathbf{B}))_F \leq \frac{\|\mathbf{T}\|_F}{\delta}
\tag{44}
$$

where $d(\cdot, \cdot)_F$ denotes the distance according to Definition 1, and the Frobenious norm is the chosen norm.

Proof of Theorem 5: Using the corollary above, setting $\mathbf{T} = \mathbf{N}_3(i)$, $\mathbf{A} = \mathbf{H}\mathbf{T}_U(i)$ and $\mathbf{B} = \mathbf{A} + \mathbf{T}$, the theorem follows.

Proof of Theorem 6: Taking $\mathbf{H} = \mathbf{U}\mathbf{S}\mathbf{V}^H$ as an SVD of $\mathbf{H}$, one can write

$$
\begin{aligned}
\mathbf{Z}(i)\mathbf{Z}(i)^H &= \mathbf{H}\mathbf{T}_{\mathbf{U}}(i)\mathbf{C}_U(i)\mathbf{C}_U(i)^H\mathbf{T}_{\mathbf{U}}(i)^H\mathbf{H}^H \\
&= \mathbf{U}\mathbf{S}\mathbf{V}^H\mathbf{T}_{\mathbf{U}}(i)\mathbf{C}_U(i)\mathbf{C}_U(i)^H\mathbf{T}_{\mathbf{U}}(i)^H\mathbf{V}\mathbf{S}\mathbf{U}^H \\
&= l \cdot \mathbf{U}\mathbf{S}\underbrace{\mathbf{V}^H\mathbf{T}_{\mathbf{U}}(i)}_{\mathbf{Q}(i)}\underbrace{\mathbf{T}_{\mathbf{U}}(i)^H\mathbf{V}}_{\mathbf{Q}(i)^H}\mathbf{S}\mathbf{U}^H \\
&= l \cdot \mathbf{U}\mathbf{S}\mathbf{Q}(i)\mathbf{Q}(i)^H\mathbf{S}\mathbf{U} \\
&= l \cdot \mathbf{U}\mathbf{S}\mathbf{I}_{N,K}\mathbf{S}\mathbf{U}^H \\
&= l \cdot \mathbf{U}_K\mathbf{S}_K^2\mathbf{U}_K^H.
\end{aligned}
\tag{45}
$$

We have used the relation $\mathbf{Q}(i)\mathbf{Q}(i)^H = \mathbf{I}_{N,K}$, where $\mathbf{I}_{N,K}$ is an $N \times N$ matrix with ones on the first K diagonal elements and zeros elsewhere. This follows from the fact that if $\mathbf{T}_{\mathbf{U}}(i)$ is in the subspace spanned by the first columns of $\mathbf{V}$, then

$$
\mathbf{Q}(i) = \mathbf{V}^H\mathbf{T}_{\mathbf{U}}(i) = \begin{pmatrix} \mathbf{W}(i) \\ 0 \end{pmatrix}
\tag{46}
$$

where $\mathbf{W}(i)$ is an orthogonal matrix. The leading singular vectors of $\mathbf{H}$ can then be extracted as the K leading eigenvectors of the autocorrelation matrix, and the theorem follows.

REFERENCES

[1] J. B. Andersen, "Array gain and capacity for known random channels with multiple element arrays at both ends," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 11, pp. 2172–2178, Nov. 2000.

[2] J.-F. Cardoso and A. Souloumiac, "Blind beamforming for non-Gaussian signals," *Proc. Inst. Elect. Eng. F, Radar Signal Processing*, vol. 140, no. 6, pp. 362–370, Dec. 1993.

[3] B. Champagne and Q. G. Liu, "Plane rotation-based EVD updating schemes for efficient subspace tracking," *IEEE Trans. Signal Process.*, vol. 46, no. 7, pp. 1886–1900, Jul. 1998.

[4] T. Dahl, N. Christophersen, and D. Gesbert, "BIMA—Blind iterative MIMO algorithm," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing*, 2002, vol. 3, pp. 2365–2368.

[5] T. Dahl, N. Christophersen, and D. Gesbert, "Blind MIMO eigenmode transmission based on the algebraic power method," *IEEE Trans. Signal Process.*, vol. 52, no. 9, pp. 2424–2431, Sep. 2004.

[6] T. Dahl and J. E. Kirkebø, "Blind communication in a frequency divided MISO system based on time reversal mirrors," presented at the Workshop Signal Process. Adv. Wireless Commun., New York, Jun. 2005.

[7] Z. Ding, "Matrix outer-product decomposition method for blind multiple channel identification," *IEEE Trans. Signal Process.*, vol. 45, no. 12, pp. 3053–3061, Dec. 1997.

[8] M. Fink, "Time-reversed acoustics," *Phys. Today*, vol. 50, pp. 34–40, 1997.

[9] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: The Johns Hopkins Univ. Press, 1996.

[10] B. Hua, Y. Xiang, T. Chen, K. Abeb-Merian, and Y. Miao, "A new look at the power method for fast subspace tracking," *Digit. Signal Process.*, vol. 9, no. 4, pp. 297–314, Oct. 1999.

[11] J. Liu and A. Host-Madsen, "Novel communication schemes with blind channel estimation in TDD MIMO system," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Montreal, Canada, May 2004, pp. II-33–II-36.

[12] H. T. Nguyen, J. B. Andersen, and G. F. Pedersen, "The potential use of time reversal techniques in multiple element antenna systems," *IEEE Commun. Lett.*, vol. 9, no. 1, pp. 40–42, Jan. 2005.

[13] C. B. Papadias, "A multi-user kurtosis algorithm for blind source separation," presented at the Int. Conf. Acoustics, Speech, Signal Processing (ICASSP), Istanbul, Turkey, Jun. 5, 2000.

[14] G. Picchi and G. Prati, "Blind equalization and carrier recovery using a 'stop-and-go' decision-directed algorithm," *IEEE Trans. Commun.*, vol. COM-35, no. 9, pp. 877–887, Sep. 1987.

[15] A. S. Y. Poon, D. N. C. Tse, and R. W. Brodersen, "An adaptive multiantenna transceiver for slowly flat fading channels," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1820–1827, Nov. 2003.

[16] A. L. Swindlehurst and G. Leus, "Blind and semi-blind equalization for generalized space-time block codes," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2489–2498, Oct. 2002.

[17] M. Tanter, J.-L. Thomas, and M. Fink, "Time reversal and the inverse filter," *J. Acoust. Soc. Amer.*, vol. 108, no. 1, pp. 223–233, 2000.

[18] A. Touzni, I. Fijalkow, M. G. Larimore, and J. R. Treichler, "A globally convergent approach for blind MIMO adaptive deconvolution," *IEEE Trans. Signal Process.*, vol. 49, no. 6, pp. 1166–1178, Jun. 2001.

[19] J. K. Tugnait, "Identification and deconvolution of multichannel nongaussian processes using higher order statistics and inverse filter criteria," *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 658–672, Mar. 1997.

[20] P.-Â. Wedin, "Perturbation bounds in connection with singular value decomposition," *BIT*, vol. 12, pp. 99–111, 1972.

[21] C.-K. Wen, Y.-C. Wang, and J.-T. Chen, "An adaptive spatio-temporal coding scheme for indoor wireless communication," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 2, pp. 161–170, Feb. 2003.

[22] H. Wold, "Partial least squares," in *Encyclopedia of Statistical Sciences*, S. Kotz and N. L. Johnson, Eds. New York: Wiley, 1985, vol. 6, pp. 581–591.

Tobias Dahl received the M.Sc. degree in applied mathematics and the Ph.D. degree from the University of Oslo, Norway, in 1998 and 2002, respectively.

He worked as a Postdoctoral Research Fellow at Imperial College, U.K. in 2002–2003 and is currently working part-time at the University of Oslo. His current research interests include blind source separation, positioning and detection, acoustic imaging, and multivariate analysis.

Dr. Dahl received the 2005 IEEE Best Young Paper Award for a paper coauthored with Gesbert and Christophersen, for his work in blind multiple-input multiple-output channel estimation.

Silvana Silva Pereira received the M.Sc. degree from the Department of Informatics, University of Oslo, Norway, in 2003. She is currently working towards the Ph.D. degree at the Signal Theory and Communications Department at the Technical University of Catalonia (UPC), Spain, with a grant from the Spanish Ministry of Education and Science.

In 2005, she joined the Signal Theory and Communications Department at UPC, Spain, as a Research Assistant for the R+D project ECO (TEC2004-04526) financed by the Spanish government. Her current research interests include array signal processing for wireless communications, MIMO wireless communications, and sensor networks.

Nils Christophersen received the M.Sc. degree in cybernetics and the Ph.D. degree from the University of Oslo, Norway, in 1972 and 1982, respectively.

He worked at the Norwegian Defence Research Establishment and The Foundation for Scientific and Industrial Research (SINTEF) before joining the Department of Informatics at the University of Oslo in 1991. He spent a year as a Visiting Scientist at the Centre for Ecology and Hydrology, Wallingford, U.K., from 1988 to 1989 and was a visiting scholar at the Department of Statistics at Stanford University,

Stanford, CA, from 1997 to 1998. His current research interests include array signal processing algorithms within acoustic imaging and instrumentation with emphasis on the synergy between sonar, seismics, and ultrasound.

Dr. Christophersen is on the Institute for Scientific Information (ISI) list of highly cited researchers.

David Gesbert received the Ph.D. degree from Ecole Nationale Superieure des Telecommunications, Paris, France, in 1997.

From 1993 to 1997, he was with France Telecom Research, Paris. From April 1997 to October 1998, he was a Research Fellow at the Information Systems Laboratory, Stanford University, Stanford, CA. He took part in the founding team of Iospan Wireless, Inc., San Jose, CA, a startup company pioneering MIMO-OFDM. Starting in 2001, he has been with the University of Oslo as an Adjunct Professor. He is currently a Professor at the Eurecom Institute, France. He has published about 100 papers and several patents all in the area of signal processing and communications.

Dr. Gesbert co-edited several special issues for the IEEE JOURNAL OF SELECTED AREAS IN COMMUNICATIONS (JSAC) (2003), the EURASIP *Journal on Advances in Signal Processing* (2004), and the *IEEE Communications Magazine* (2006). He is an elected member of the IEEE Signal Processing for Communications Technical Committee. He authored or coauthored papers winning the 2004 IEEE Best Tutorial Paper Award (Communications Society) for a 2003 JSAC paper on MIMO systems, the 2005 Best Paper (Young Author) Award for Signal Processing Society journals, and the Best Paper Award for the 2004 ACM International Symposium on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM) workshop. He was co-organizer, with Prof. D. Slock, of the IEEE Workshop on Signal Processing Advances in Wireless Communications, Cannes, France, 2006.