Blind Channel Approximation: Effective Channel Order Determination

Athanasios P. Liavas, *Member, IEEE*, Phillip A. Regalia, *Senior Member, IEEE*, and Jean-Pierre Delmas

***Abstract*—A common assumption of blind channel identification methods is that the order of the true channel is known. This information is not available in practice, and we are obliged to estimate the channel order by applying a rank detection procedure to an "overmodeled" data covariance matrix. Information theoretic criteria have been widely suggested approaches for this task. We check the quality of their estimates in the context of order estimation of measured microwave radio channels and confirm that they are very sensitive to variations in the SNR and the number of data samples. This fact has prohibited their successful application for channel order estimation and has created some confusion concerning the classification into under- and over-modeled cases. Recently, it has been shown that blind channel approximation methods should attempt to model only the significant part of the channel composed of the "large" impulse response terms because efforts toward modeling "small" leading and/or trailing terms lead to effective overmodeling, which is generically ill-conditioned and, thus, should be avoided. This can be achieved by applying blind identification methods with model order equal to the order of the significant part of the true channel called the effective channel order. Toward developing an efficient approach for the detection of the effective channel order, we use numerical analysis arguments. The derived criterion provides a "maximally stable" decomposition of the range space of an "overmodeled" data covariance matrix into signal and noise subspaces. It is shown to be robust to variations in the SNR and the number of data samples. Furthermore, it provides useful effective channel order estimates, leading to sufficiently good blind approximation/equalization of measured real-world microwave radio channels.**

I. INTRODUCTION

FOLLOWING the work of Tong *et al.* [1], many methods have been proposed recently that claim blind single-input/multi-output channel identification under the so-called length and zero conditions [2]–[4]. A common assumption in all these works is that the order of the true channel is known. Of course, such information is not available in practice, and we are thus obliged to estimate the channel order by applying a rank detection procedure to an "overmodeled" data covariance matrix. The use of information theoretic criteria, as proposed in [5], has become the standard first step of many methods that treat the blind channel identification problem.

Manuscript received June 23, 1998; revised June 1, 1999. A. P. Liavas was supported by the Training and Mobility of Researchers (T.M.R.) Program of the European Commission under Contract ERBFMBICT960659. The associate editor coordinating the review of this paper and approving it for publication was Dr. Mahmood R. Azimi-Sadjadi.

A. P. Liavas was with the Department SIM, Institut National des Télécommunications, Evry, France. He is now with the Department of Computer Science, University of Ioannina, Ioannina, Greece.

P. A. Regalia and J.-P. Delmas are with the Department SIM, Institut National des Télécommunications, Evry, France.

Publisher Item Identifier S 1053-587X(99)09208-9.

The development of information theoretic criteria is based on the assumptions that successive data vectors are i.i.d. zero-mean Gaussian random vectors and that the noise is white Gaussian and uncorrelated from the signal. These assumptions seem realistic in some applications, such as directions of arrival, but in other applications, such as blind channel identification, they do not seem the most appropriate. First, in blind channel identification, the data covariance matrix is built from vectors that exhibit the so-called "shift property;" thus, successive data vectors are *not* statistically independent. Furthermore, existence of "colored noise" due to the influence of long tails of "small" leading and/or trailing impulse response terms, is practically inevitable. These terms should *not* be modeled because the quality of our estimate may degrade dramatically [6], [7], and thus, we consider their influence on the data covariance matrix as "colored noise."

Hence, the assumptions on which information theoretic criteria are based do *not* hold true in the blind channel identification context. Thus, a natural question arises: "*Do information theoretic criteria provide* reliable *effective channel order estimates?*" We check the quality of their estimates in the context of order estimation of measured microwave radio channels. We observe that they are very sensitive to variations in the SNR and the number of data samples. This fact has prohibited their successful application for effective channel order estimation and has created, arguably, some confusion concerning the classification into under- and over-modeled cases. This, in turn, has created confusion regarding the robustness and applicability of blind channel identification methods under realistic conditions.

In order to overcome the shortcomings of information theoretic criteria, we propose a new approach based on numerical analysis arguments. Using the concept of canonical angles between subspaces and invariant subspace perturbation results, we develop a criterion that provides a "maximally stable" decomposition of the range space of an "overmodeled" data covariance matrix into signal and noise subspaces. When used for the effective channel order determination of measured microwave radio channels, the proposed criterion is shown to be insensitive to variations in the SNR and the number of data samples. Furthermore, it permits a classification into stable or well-conditioned and unstable or ill-conditioned cases. In the stable cases, it provides useful effective channel order estimates, leading to sufficiently good blind channel approximation/equalization; this is *not* always the case with information theoretic criteria. In the unstable cases, sufficiently good blind channel approximation/equalization seems difficult.

II. Effective Channel Order Determination Using Information Theoretic Criteria

In this section, we recall the information theoretic criteria [5], [9], [10], and we apply them to the effective channel order determination problem.

A. Information Theoretic Criteria

Let us assume that we measure a sequence of p-dimensional data vectors $\{\mathbf{x}(n)\}_{n=1}^{N}$, which obey the model

$$\mathbf{x}(n) = \mathbf{A}\,\mathbf{s}(n) + \mathbf{w}(n).$$

$\mathbf{A}$ is a $p \times q$ full-column rank matrix $(q<p)$. $\{\mathbf{s}(n)\}$ is a sequence of zero-mean stationary ergodic circular complex Gaussian q-dimensional random vectors with nonsingular covariance matrix

$$\mathcal{S} \triangleq E\{\mathbf{s}(n)\,\mathbf{s}^H(n)\}$$

where superscript ^{H} denotes Hermitian transpose; $\{\mathbf{w}(n)\}$ is a sequence of zero-mean stationary ergodic circular complex Gaussian p-dimensional random vectors with covariance matrix

$$\mathcal{W} \triangleq E\{\mathbf{w}(n)\,\mathbf{w}^H(n)\} = \sigma^2\mathbf{I}$$

where $\mathbf{I}$ denotes the identity matrix, (its dimension becomes clear from the context); furthermore, $\{\mathbf{s}(n)\}$ and $\{\mathbf{w}(n)\}$ are uncorrelated. Under these assumptions, the covariance matrix of $\mathbf{x}(n)$ is

$$\mathcal{R} \triangleq E\{\mathbf{x}(n)\,\mathbf{x}^H(n)\} = \mathbf{A}\,\mathcal{S}\,\mathbf{A}^H + \sigma^2\mathbf{I}$$

where $\mathbf{A}\,\mathcal{S}\,\mathbf{A}^H$ is a rank-q matrix. The q-dimensional subspace spanned by the columns of $\mathbf{A}$ is usually called *signal* subspace, whereas its orthogonal complement is called *noise* subspace.

A very important problem arising in many application areas is the determination of the dimension of the signal subspace. Denoting the eigenvalues of $\mathcal{R}$ as $\lambda_1(\mathcal{R}) \geq \lambda_2(\mathcal{R}) \geq \cdots \geq \lambda_p(\mathcal{R})$, we obtain that the $(p-q)$ smallest eigenvalues of $\mathcal{R}$ are equal to σ^2, i.e.,

$$\lambda_{q+1}(\mathcal{R}) = \lambda_{q+2}(\mathcal{R}) = \cdots = \lambda_p(\mathcal{R}) = \sigma^2.$$

Hence, in theory, we can determine the dimension of the signal subspace from the multiplicity of the smallest eigenvalue of $\mathcal{R}$. However, in practice, we do not have access to the true data covariance matrix but to its finite data sample estimate

$$\mathbf{R} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{x}(n)\,\mathbf{x}^H(n). \quad (1)$$

In this case, the smallest eigenvalues of $\mathbf{R}$ are all different with high probability, thus complicating the determination of the dimension of the signal subspace.

Akaike's information theoretic criterion (AIC) [9] selects the model that minimizes

$$\text{AIC} = -2\log f(\mathbf{x}(1),\cdots,\mathbf{x}(N)|\hat{\Theta}) + 2k$$

where $f(\mathbf{x}(1),\cdots,\mathbf{x}(N)|\Theta)$ is a parametrized family of probability densities, $\hat{\Theta}$ is the maximum likelihood estimate of a parameter vector Θ, and k is the number of free adjusted parameters in Θ.

The minimum description length (MDL) criterion [10] selects the model that instead minimizes

$$\text{MDL} = -\log f(\mathbf{x}(1),\cdots,\mathbf{x}(N)|\hat{\Theta}) + \tfrac{1}{2}k\log N.$$

In the estimated data covariance matrix case described by (1), assuming that the observed vectors $\{\mathbf{x}(n)\}_{n=1}^{N}$ are zero-mean i.i.d Gaussian random vectors, one may show that [5]

$$\text{AIC}(k) = -2\,\log\left[\frac{\prod_{i=k+1}^{p}\lambda_i^{1/(p-k)}}{\frac{1}{p-k}\sum_{i=k+1}^{p}\lambda_i}\right]^{(p-k)N} + 2k(2p-k)$$

and

$$\text{MDL}(k) = -\,\log\left[\frac{\prod_{i=k+1}^{p}\lambda_i^{1/(p-k)}}{\frac{1}{p-k}\sum_{i=k+1}^{p}\lambda_i}\right]^{(p-k)N} + \frac{1}{2}k(2p-k)\,\log N$$

where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$ denote the eigenvalues of $\mathbf{R}$. The dimension of the signal subspace is taken to be the value of $k \in \{0, 1\cdots, p-1\}$ for which either AIC(k) or MDL(k) is minimized. The MDL criterion is shown to be asymptotically consistent, whereas the AIC tends to overestimate the dimension of the signal subspace [5]. Taking into account the sensitivity of blind channel identification methods, with respect to effective channel overmodeling, the MDL criterion has often been favored over the AIC.

B. Application of Information Theoretic Criteria for Effective Channel Order Determination

A very important application area requiring the determination of a subspace dimension is blind channel identification. In Fig. 1, we present the one-input/two-output channel setting that is derived either by channel oversampling by a factor of 2 or by using two sensors at the receiver. Although, in the sequel, we present the one-input/two-output case, the results can be trivially extended to the one-input/p-output case, with $p>2$.

If the true channel order is M, and the channel impulse response is denoted by $\mathbf{h}_M \triangleq [\mathbf{h}_M^{1\,T}\mathbf{h}_M^{2\,T}]^T$, where superscript ^{T} denotes transpose, then the data vector $\mathbf{x}_L(n)$ composed of the $(L+1)$ most recent samples of each subchannel, i.e., $\mathbf{x}_L(n) \triangleq [x_n^{(1)}\cdots x_{n-L}^{(1)}\,x_n^{(2)}\cdots x_{n-L}^{(2)}]^T$, can be expressed as

$$\mathbf{x}_L(n) = \mathbf{y}_L(n) + \mathbf{w}_L(n) = \mathcal{H}_L(\mathbf{h}_M)\mathbf{s}_{L+M}(n) + \mathbf{w}_L(n)$$

where

$$\mathbf{y}_L(n) \triangleq [y_n^{(1)}\cdots y_{n-L}^{(1)}\,y_n^{(2)}\,\cdots\,y_{n-L}^{(2)}]^T$$
$$\mathbf{w}_L(n) \triangleq [w_n^{(1)}\cdots w_{n-L}^{(1)}\,w_n^{(2)}\cdots w_{n-L}^{(2)}]^T$$
$$\mathbf{s}_{L+M}(n) \triangleq [s_n\cdots s_{n-L-M}]^T.$$

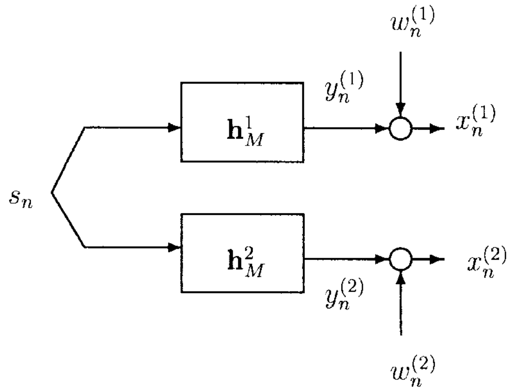


Fig. 1. One-input/two-output channel setting.

The convolution matrix $\mathcal{H}_L(\mathbf{h}_M)$ is defined as

$$\mathcal{H}_L(\mathbf{h}_M) \triangleq \begin{bmatrix} \mathcal{F}_L(\mathbf{h}_M^1) \\ \mathcal{F}_L(\mathbf{h}_M^2) \end{bmatrix}$$

with the $(L+1) \times (L+M+1)$ matrix $\mathcal{F}_L(\mathbf{h}_M^i)$ given by

$$\mathcal{F}_L(\mathbf{h}_M^i) \triangleq \begin{bmatrix} h_0^{(i)} & \cdots & \cdots & h_M^{(i)} & & \\ & \ddots & & & \ddots & \\ & & h_0^{(i)} & \cdots & \cdots & h_M^{(i)} \end{bmatrix}.$$

The input $\{s_n\}$ is assumed zero-mean unit-variance white noise, whereas the additive channel noise is assumed temporally and spatially white, i.e.,

$$E\{\mathbf{w}_L(i)\,\mathbf{w}_L^H(j)\} = \delta_{i,j}\,\sigma^2\,\mathbf{I}.$$

Furthermore, the input and the additive channel noise are assumed to be uncorrelated.

If $L \geq M-1$ and subchannels $\mathbf{h}_M^1$, $\mathbf{h}_M^2$ do not share common zeros, then $\mathcal{H}_L(\mathbf{h}_M)$ is of full-column rank, i.e.,

$$\text{rank}(\mathcal{H}_L(\mathbf{h}_M)) = L + M + 1.$$

Thus, the data covariance matrix

$$\begin{aligned}\mathcal{R}_L &\triangleq E\{\mathbf{x}_L(n)\mathbf{x}_L^H(n)\} = \mathcal{H}_L(\mathbf{h}_M)\,\mathcal{H}_L^H(\mathbf{h}_M) + \sigma^2\mathbf{I} \\ &= \tilde{\mathcal{R}}_L + \sigma^2\mathbf{I}\end{aligned}$$

is the sum of a rank-$(L+M+1)$ matrix and a multiple of the identity. By determining the rank of $\tilde{\mathcal{R}}_L$, we can estimate the order of the channel M as

$$M = \text{rank}(\tilde{\mathcal{R}}_L) - L - 1.$$

However, in reality, the situation is somewhat different. The true impulse response $\mathbf{h}_M$ is often very long [11], that is, usually $M \gg L$, and it can be partitioned into the significant part and the tails. By *significant part*, we mean the part that is usually found near the middle of the true impulse response and contains the "large" terms; it may contain some intermediate "small" terms as well. Its order, called *effective channel order*, is denoted by (the unknown integer) m. By *tails* we mean the part of the true channel that is complementary to the significant part; it is composed of "small" leading and/or trailing terms. Notationally, this partitioning can be expressed for $0 \leq m_1 < m_2 \triangleq m_1 + m \leq M$ as [6]–[8]

$$\mathbf{h}_M = \mathbf{h}_{m_1,m_2}^{\mathrm{z}} + \mathbf{d}_{m_1,m_2}^{\mathrm{z}}$$

where superscript ^{z} means "appropriately zero-padded" and

$$\mathbf{h}_{m_1,m_2}^{\mathrm{z}} \triangleq \begin{bmatrix} \mathbf{h}_{m_1,m_2}^{\mathrm{z}1} \\ \mathbf{h}_{m_1,m_2}^{\mathrm{z}2} \end{bmatrix}, \quad \mathbf{d}_{m_1,m_2}^{\mathrm{z}} \triangleq \begin{bmatrix} \mathbf{d}_{m_1,m_2}^{\mathrm{z}1} \\ \mathbf{d}_{m_1,m_2}^{\mathrm{z}2} \end{bmatrix}$$

with

$$\mathbf{h}_{m_1,m_2}^{\mathrm{z}j} \triangleq [\underbrace{0\cdots0}_{m_1}\underbrace{h_{m_1}^{(j)}\cdots h_{m_2}^{(j)}}_{m+1}\underbrace{0\cdots0}_{M-m_2}]^T, \qquad j = 1,2$$

$$\mathbf{d}_{m_1,m_2}^{\mathrm{z}j} \triangleq [\underbrace{h_0^{(j)}\cdots h_{m_1-1}^{(j)}}_{m_1}\underbrace{0\cdots0}_{m+1}\underbrace{h_{m_2+1}^{(j)}\cdots h_M^{(j)}}_{M-m_2}]^T$$

$$j = 1,2$$

and

$$\|\mathbf{d}_{m_1,m_2}^{\mathrm{z}}\|_2 \ll \|\mathbf{h}_M\|_2. \tag{2}$$

With $\mathbf{h}_{m_1,m_2}$, we denote the truncated significant part of the channel

$$\mathbf{h}_{m_1,m_2} \triangleq \begin{bmatrix} \mathbf{h}_{m_1,m_2}^1 \\ \mathbf{h}_{m_1,m_2}^2 \end{bmatrix}$$

$$\mathbf{h}_{m_1,m_2}^j \triangleq [h_{m_1}^{(j)}\cdots h_{m_2}^{(j)}]^T, \qquad j = 1,2.$$

In this case, the data covariance matrix is

$$\begin{aligned}\mathcal{R}_L &= \mathcal{H}_L(\mathbf{h}_M)\mathcal{H}_L^H(\mathbf{h}_M) + \sigma^2\mathbf{I} \\ &= \mathcal{H}_L(\mathbf{h}_{m_1,m_2}^{\mathrm{z}} + \mathbf{d}_{m_1,m_2}^{\mathrm{z}})\mathcal{H}_L^H(\mathbf{h}_{m_1,m_2}^{\mathrm{z}} + \mathbf{d}_{m_1,m_2}^{\mathrm{z}}) \\ &\quad + \sigma^2\mathbf{I} \\ &= \{\mathcal{H}_L(\mathbf{h}_{m_1,m_2}^{\mathrm{z}}) + \mathcal{H}_L(\mathbf{d}_{m_1,m_2}^{\mathrm{z}})\} \\ &\quad \cdot \{\mathcal{H}_L^H(\mathbf{h}_{m_1,m_2}^{\mathrm{z}}) + \mathcal{H}_L^H(\mathbf{d}_{m_1,m_2}^{\mathrm{z}})\} + \sigma^2\mathbf{I} \\ &= \mathcal{H}_L(\mathbf{h}_{m_1,m_2}^{\mathrm{z}})\,\mathcal{H}_L^H(\mathbf{h}_{m_1,m_2}^{\mathrm{z}}) + \mathbf{E}_L + \sigma^2\mathbf{I} \\ &= \tilde{\mathcal{R}}_L + \mathbf{E}_L + \sigma^2\mathbf{I}\end{aligned}$$

where $\tilde{\mathcal{R}}_L$ denotes the covariance matrix associated with the significant part of the channel, and $\mathbf{E}_L$ expresses the influence of the tails. Due to (2), $\mathbf{E}_L$ is assumed to be "small" with respect to $\tilde{\mathcal{R}}_L$; σ^2 is also assumed to be "small."

As shown in [6] and [7], blind channel identification methods should attempt to model only the significant part of the channel because efforts toward modeling "small" leading and/or trailing terms lead to effective overmodeling, which is generically ill-conditioned and thus should be avoided. This can be achieved by applying blind identification methods with model order equal to the effective channel order. Thus, the development of efficient approaches for the determination of the effective channel order is of great importance.

Since the significant part of the channel $\mathbf{h}_{m_1,m_2}$ has order m, if $L \geq m-1$ and subchannels $\mathbf{h}_{m_1,m_2}^1$, $\mathbf{h}_{m_1,m_2}^2$ do not share common zeros, then $\mathcal{H}_L(\mathbf{h}_{m_1,m_2})$ is of full-column rank, i.e.,

$$\text{rank}(\mathcal{H}_L(\mathbf{h}_{m_1,m_2})) = L + m + 1.$$

It can be verified easily that

$$\mathcal{H}_L(\mathbf{h}_{m_1,m_2}^{\mathrm{z}})\,\mathcal{H}_L^H(\mathbf{h}_{m_1,m_2}^{\mathrm{z}}) = \mathcal{H}_L(\mathbf{h}_{m_1,m_2})\,\mathcal{H}_L^H(\mathbf{h}_{m_1,m_2})$$

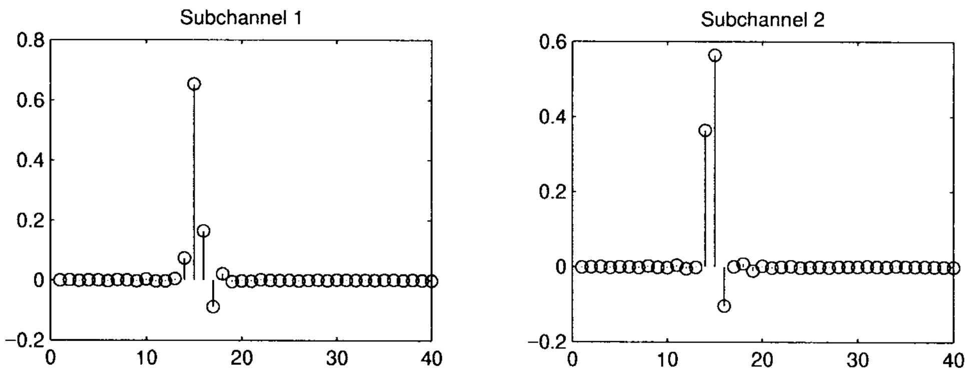


Fig. 2. Portion of the real part of subchannels.

yielding

$$\text{rank}(\tilde{\mathcal{R}}_L) = \text{rank}(\mathcal{H}_L(\mathbf{h}_{m_1,m_2})) = L+m+1. \quad (3)$$

This means that the estimated $2(L+1)\times 2(L+1)$ covariance matrix $\mathcal{R}_L$ is "close" to the rank-$(L+m+1)$ matrix $\tilde{\mathcal{R}}_L$.

By computing the rank of $\tilde{\mathcal{R}}_L$, we can deduce the effective channel order m. Now, however, our problem is more demanding than that attacked in the previous subsection because the perturbation onto the "ideal" matrix $\tilde{\mathcal{R}}_L$ is no longer a multiple of the identity. Despite this difference, the only methods that have been suggested for channel order estimation are the information theoretic criteria, as proposed in [5]. Therefore, their behavior in realistic cases is of great importance.

In Fig. 2, we plot a portion of the real part of the two subchannels constructed by the complex-valued oversampled, by a factor of 2, FIR microwave radio channel *chan10.mat*, which is found at http://spib.rice.edu/spib/microwave.html. The partitioning into the significant part and the tails is clear. Intuitively satisfying effective channel order estimates are two or three, that is, three or four taps, for each subchannel.

In order to estimate the effective channel order, we perform the following experiment: We put as input to the channel $N = 200$ independent samples from a 4-QAM constellation. At the channel output, we add spatially and temporally white noise, with SNR = 90, 70, 50, and 30 dB. The SNR is defined as

$$\text{SNR} = 10\log_{10}\frac{\sum_{j=1}^{2} E\{|y_n^{(j)}|^2\}}{2\sigma^2}$$

where σ^2 is the variance of the circular complex additive white subchannel noise. Then, we compute the "overmodeled" covariance matrix of the noisy channel output $\mathbf{R}_{20}$ (with dimensions 42×42). In Fig. 3, we plot the Akaike criterion, that is, AIC(k) versus k, for the various SNR's. The estimates of the rank of $\mathbf{R}_{20}$, i.e., the k's for which AIC(k) is minimized, are 40, 41, 30, and 23, respectively. Using (3), we compute the corresponding effective channel order estimates as 19, 20, 9, and 2. The application of the MDL criterion is illustrated in Fig. 4; the effective channel order estimates are 19, 19, 9, and 2.

The most striking observation in the cases presented in Figs. 3 and 4, and in extensive studies using all the measured channels found at this site, is that the estimates of information theoretic criteria are very sensitive to variations in the SNR and the number of data samples. For high SNR (SNR > 30 dB) and/or many data samples ($N > 300$), they usually lead to effective overmodeling. Such estimates are practically useless [6], [7]. For low SNR and few data samples, they may provide useful estimates, leading to sufficiently good blind channel approximation/equalization. However, their high sensitivity is clearly unsatisfactory and has arguably created some confusion concerning the correct classification into under- and overmodeled cases. This, in turn, has created confusion regarding the robustness and applicability of blind channel identification methods in realistic cases.

In the sequel, we provide an entirely different approach, based on numerical analysis arguments. The derived criterion appears to be much more robust than information theoretic criteria and its estimates much more useful, as validated in many simulations.

III. A New Rank Detection Criterion

Let us consider the $2(L+1)$-dimensional estimated data covariance matrix $\mathbf{R}_L$, which is assumed to be the sum of the unknown "ideal" rank-q matrix $\tilde{\mathcal{R}}_L$ and the unknown "perturbation" matrix $\mathbf{E}_L$, i.e.,

$$\mathbf{R}_L = \tilde{\mathcal{R}}_L + \mathbf{E}_L$$

where $q = L+m+1$, with m being the assumed effective channel order; q and, thus, m remain to be determined. The "ideal" matrix $\tilde{\mathcal{R}}_L$ denotes the exact statistics covariance matrix associated with the significant part of the channel, i.e.,

$$\tilde{\mathcal{R}}_L \triangleq \mathcal{H}_L(\mathbf{h}_{m_1,m_2})\mathcal{H}_L^H(\mathbf{h}_{m_1,m_2}).$$

The "perturbation" matrix $\mathbf{E}_L$ incorporates the influence of the tails, the influence of the additive, not necessarily white, channel noise, and the influence of the estimated, inexact statistics; $\mathbf{E}_L$ is assumed to be "small" with respect to $\tilde{\mathcal{R}}_L$.

Let us denote the eigenvalues of $\tilde{\mathcal{R}}_L$ as

$$\lambda_1(\tilde{\mathcal{R}}_L) \geq \cdots \geq \lambda_q(\tilde{\mathcal{R}}_L) > \lambda_{q+1}(\tilde{\mathcal{R}}_L) = \cdots = \lambda_{2(L+1)}(\tilde{\mathcal{R}}_L) = 0.$$

The smallest nonzero eigenvalue of $\tilde{\mathcal{R}}_L$, $\lambda_q(\tilde{\mathcal{R}}_L)$, being the distance, in the matrix 2-norm of $\tilde{\mathcal{R}}_L$ from the matrices with rank $(q-1)$ measures "how well" fulfilled our assumption

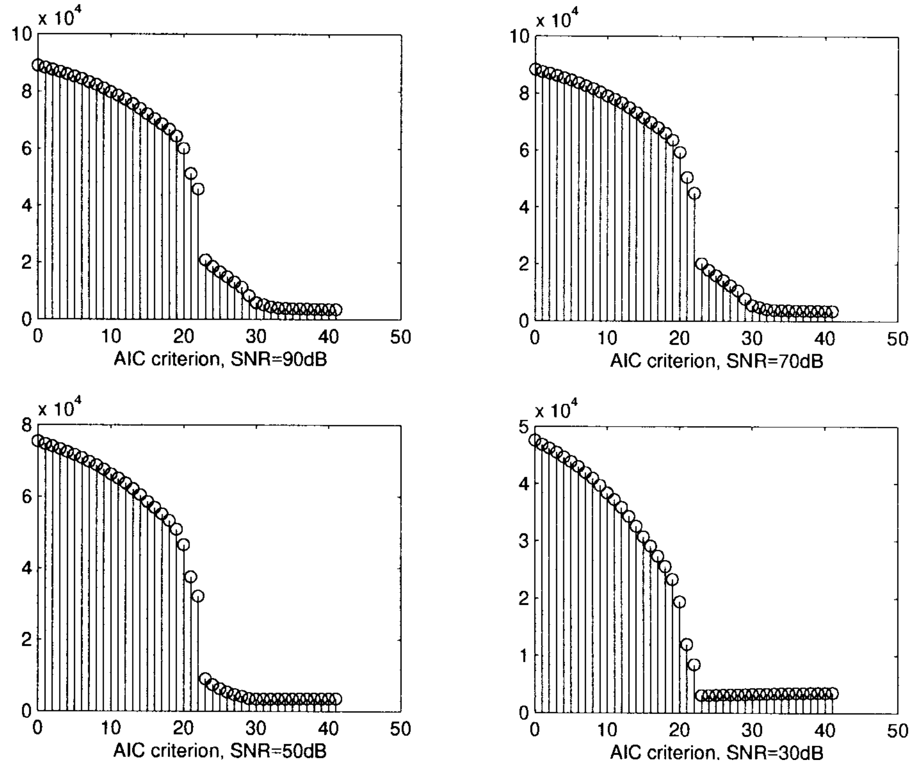


Fig. 3. AIC(k) versus k, for SNR = 90, 70, 50, and 30 dB.

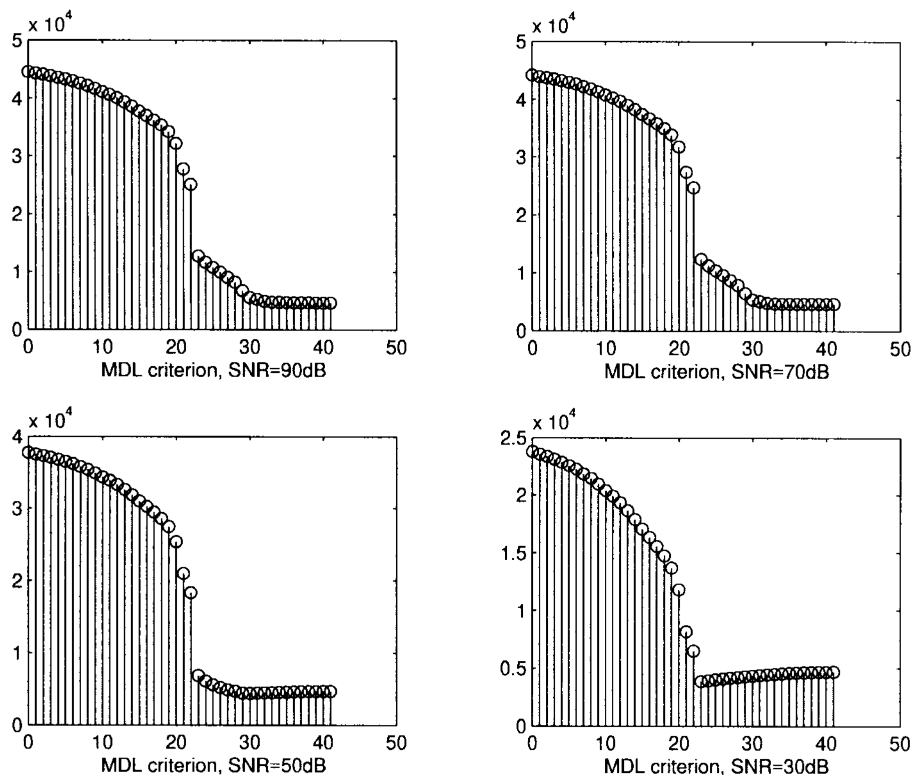


Fig. 4. MDL(k) versus k, for SNR = 90, 70, 50, and 30 dB.

concerning rank$(\tilde{\mathcal{R}}_L)$ is. Since

$$\lambda_q(\tilde{\mathcal{R}}_L) = \sigma_q^2(\mathcal{H}_L(\mathbf{h}_{m_1,m_2})) \tag{4}$$

with $\sigma_i(\mathcal{A})$ being the ith singular value of matrix $\mathcal{A}$, we may consider $\lambda_q(\tilde{\mathcal{R}}_L)$ as a measure of *diversity* of channel $\mathbf{h}_{m_1,m_2}$.

Let the spectral decomposition of $\mathbf{R}_L$ be

$$\mathbf{R}_L = \sum_{i=1}^{2(L+1)} \lambda_i \, \mathbf{u}_i \, \mathbf{u}_i^H.$$

The "ideal" signal subspace is the *unknown* q-dimensional subspace spanned by the columns of $\tilde{\mathcal{R}}_L$, denoted $R(\tilde{\mathcal{R}}_L)$. However, in practice, we consider *incorrectly* as signal subspace the "perturbed" subspace $R(\mathbf{S}_q)$, where $\mathbf{S}_q$ is the matrix associated with the q largest eigenpairs of $\mathbf{R}_L$, i.e.,

$$\mathbf{S}_q \triangleq \sum_{i=1}^{q} \lambda_i \mathbf{u}_i \mathbf{u}_i^H$$

and noise subspace as its orthogonal complement, which is spanned by the columns of $\mathbf{N}_q$ defined as

$$\mathbf{N}_q \triangleq \sum_{i=q+1}^{2(L+1)} \lambda_i \mathbf{u}_i \mathbf{u}_i^H. \tag{5}$$

The assumption that $\tilde{\mathcal{R}}_L$ is of rank q implies that

$$\|\mathbf{E}_L\|_2 \geq \lambda_{q+1} \tag{6}$$

which means that the unknown "ideal" and the estimated "perturbed" subspaces $R(\tilde{\mathcal{R}}_L)$ and $R(\mathbf{S}_q)$, respectively, are related through a perturbation whose size, as measured by the matrix 2-norm, is greater than or equal to λ_{q+1}. Hence, assuming that rank$(\tilde{\mathcal{R}}_L) = q$, (6) is the *only* information we may deduce for the perturbation $\mathbf{E}_L$.

Since $\mathbf{E}_L$ is assumed to be "small" with respect to $\tilde{\mathcal{R}}_L$, we would like $R(\mathbf{S}_q)$ to be "close" to $R(\tilde{\mathcal{R}}_L)$. However, (6) is insufficient to calculate the distance between $R(\tilde{\mathcal{R}}_L)$ and $R(\mathbf{S}_q)$. We can, however, examine the sensitivity of $R(\mathbf{S}_q)$ with respect to "small" perturbations. To this end, we shall compute how far $R(\mathbf{S}_q)$ may be moved by a perturbation, which is denoted $\mathcal{E}_L$, whose 2-norm is the smallest that the actual perturbation $\mathbf{E}_L$ may have, that is

$$\|\mathcal{E}_L\|_2 = \lambda_{q+1}. \tag{7}$$

If $R(\mathbf{S}_q)$ is insensitive to $\mathcal{E}_L$, then we have reason to believe that $R(\tilde{\mathcal{R}}_L)$ and $R(\mathbf{S}_q)$ are close each other. If, on the other hand, $R(\mathbf{S}_q)$ is sensitive to $\mathcal{E}_L$, then $R(\tilde{\mathcal{R}}_L)$ and $R(\mathbf{S}_q)$ may be far from each other.

We shall then take our rank estimate as the index q, for which $R(\mathbf{S}_q)$ is the least-sensitive, with respect to all perturbations $\mathcal{E}_L$ with $\|\mathcal{E}_L\|_2 = \lambda_{q+1}$, over all q.

Thus, let us consider

$$\hat{\mathbf{R}}_L = \mathbf{R}_L + \mathcal{E}_L$$

with spectral decomposition

$$\hat{\mathbf{R}}_L = \sum_{i=1}^{2(L+1)} \hat{\lambda}_i \hat{\mathbf{u}}_i \hat{\mathbf{u}}_i^H.$$

The eigenpartitionings defined analogously to $\mathbf{S}_q$ and $\mathbf{N}_q$ become

$$\hat{\mathbf{S}}_q \triangleq \sum_{i=1}^{q} \hat{\lambda}_i \hat{\mathbf{u}}_i \hat{\mathbf{u}}_i^H, \quad \hat{\mathbf{N}}_q \triangleq \sum_{i=q+1}^{2(L+1)} \hat{\lambda}_i \hat{\mathbf{u}}_i \hat{\mathbf{u}}_i^H. \tag{8}$$

A distance measure between two linear subspaces, commonly employed in numerical analysis, is the sine of their *canonical angles*, [12, p. 43] [13]

$$\begin{aligned} \|\sin\angle(R(\hat{\mathbf{S}}_q), R(\mathbf{S}_q))\|_2 &= \|\mathbf{P}_{R(\hat{\mathbf{S}}_q)} - \mathbf{P}_{R(\mathbf{S}_q)}\|_2 \\ &= \|\mathbf{P}_{R(\mathbf{S}_q)^\perp} \mathbf{P}_{R(\hat{\mathbf{S}}_q)}\|_2 \end{aligned} \tag{9}$$

where $\mathbf{P}_{R(\mathcal{A})}$ is the orthogonal projector onto $R(\mathcal{A})$, and $R(\mathcal{A})^\perp$ is the orthogonal complement of $R(\mathcal{A})$. Since $R(\hat{\mathbf{S}}_q)$ and $R(\mathbf{S}_q)$ are related through the perturbation $\mathcal{E}_L$, we can compute an upper bound for this distance using invariant subspace perturbation results. It turns out [13] that the distance between $R(\hat{\mathbf{S}}_q)$ and $R(\mathbf{S}_q)$ is determined by the size of $\mathcal{E}_L$ and the *separation* between the eigenvalues associated with $R(\hat{\mathbf{S}}_q)$ and $R(\mathbf{N}_q)$, which is defined as

$$\delta \triangleq \hat{\lambda}_q - \lambda_{q+1}. \tag{10}$$

As we are going to see shortly, we use δ only in the cases in which its positivity is guaranteed.

Denoting by $\hat{\mathbf{S}}_q^{-}$ the Moore–Penrose generalized inverse of $\hat{\mathbf{S}}_q$, we obtain

$$\mathbf{P}_{R(\hat{\mathbf{S}}_q)} = \hat{\mathbf{S}}_q \hat{\mathbf{S}}_q^{-}.$$

Then, (9) yields [12, p. 268], [13]

$$\begin{aligned} &\|\sin\angle(R(\hat{\mathbf{S}}_q), R(\mathbf{S}_q))\|_2 \\ &\quad = \|\mathbf{P}_{R(\mathbf{S}_q)^\perp} \mathbf{P}_{R(\hat{\mathbf{S}}_q)}\|_2 = \|\mathbf{P}_{R(\mathbf{S}_q)^\perp} \hat{\mathbf{S}}_q \hat{\mathbf{S}}_q^{-}\|_2 \\ &\quad = \|\mathbf{P}_{R(\mathbf{S}_q)^\perp} (\mathbf{S}_q + \mathbf{N}_q + \mathcal{E}_L - \hat{\mathbf{N}}_q) \hat{\mathbf{S}}_q^{-}\|_2 \\ &\quad = \|(\mathbf{P}_{R(\mathbf{S}_q)^\perp} \mathcal{E}_L \mathbf{P}_{R(\hat{\mathbf{S}}_q)} + \mathbf{N}_q \mathbf{P}_{R(\mathbf{S}_q)^\perp} \mathbf{P}_{R(\hat{\mathbf{S}}_q)}) \hat{\mathbf{S}}_q^{-}\|_2. \end{aligned} \tag{11}$$

Defining $\mathcal{T} \triangleq \|\sin\angle(R(\hat{\mathbf{S}}_q), R(\mathbf{S}_q))\|_2$ and using (5), (8) and (10), we obtain from (11)

$$\begin{aligned} \mathcal{T} &\leq (\|\mathcal{E}_L\|_2 + \|\mathbf{N}_q\|_2 \mathcal{T}) \|\hat{\mathbf{S}}_q^{-}\|_2 \\ &= \frac{\|\mathcal{E}_L\|_2 + \lambda_{q+1}\mathcal{T}}{\hat{\lambda}_q} = \frac{\lambda_{q+1} + \lambda_{q+1}\mathcal{T}}{\lambda_{q+1} + \delta} \end{aligned}$$

which simplifies to

$$\mathcal{T} \leq \frac{\lambda_{q+1}}{\delta} = \frac{\lambda_{q+1}}{\hat{\lambda}_q - \lambda_{q+1}}.$$

Using standard eigenvalue perturbation results [15, p. 411], we have further that

$$\hat{\lambda}_q \geq \lambda_q - \|\mathcal{E}_L\|_2 = \lambda_q - \lambda_{q+1}$$

giving that if $\lambda_{q+1} \leq (\lambda_q/3)$, then $\delta > 0$ and

$$\mathcal{T} \leq \frac{\lambda_{q+1}}{\lambda_q - 2\lambda_{q+1}}.$$

Otherwise, our upper bound is equal to 1. Thus, we have

$$\mathcal{T} \leq r(q) \triangleq \begin{cases} \dfrac{\lambda_{q+1}}{\lambda_q - 2\lambda_{q+1}}, & \text{if } \lambda_{q+1} \leq \dfrac{\lambda_q}{3} \\ 1, & \text{otherwise.} \end{cases} \tag{12}$$

Relation (12) reveals that the sensitivity of $R(\mathbf{S}_q)$, with respect to perturbations satisfying (7), is governed by the separation of the eigenvalues λ_q and λ_{q+1}.

If $r(q) \ll 1$, meaning that $\lambda_{q+1} \ll \lambda_q$, then the "estimated" signal subspace $R(\mathbf{S}_q)$ is insensitive to perturbations with size λ_{q+1}, lending credence to its proximity to the unknown "ideal" signal subspace $R(\tilde{R}_L)$. Furthermore, from an eigenvalue point of view, since $\lambda_{q+1} \ll \lambda_q$, it does not seem plausible that the "perturbed" eigenvalues λ_q and λ_{q+1}, which have been associated with different subspaces, come from the same multiple "ideal" eigenvalue $\lambda_*(\tilde{\mathcal{R}}_L)$.

If, on the other hand, $r(q) \approx 1$, meaning that $\lambda_q \approx \lambda_{q+1}$, then the "estimated" signal subspace $R(\mathbf{S}_q)$ may be very sensitive to perturbations with size λ_{q+1}, casting serious doubt as to its proximity to the unknown "ideal" signal subspace $R(\tilde{R}_L)$. Furthermore, since $\lambda_q \approx \lambda_{q+1}$, it seems plausible that λ_q and λ_{q+1}, which have been associated with different subspaces, come from the same multiple "ideal" eigenvalue, for example, $\lambda_*(\tilde{\mathcal{R}}_L) = 0$.

Our rank estimate for $\tilde{\mathcal{R}}_L$ will be taken as the integer $\tilde{q}$, leading to the effective channel order estimation $\tilde{m}$, which minimizes $r(q)$.

Thus, our criterion becomes

$$\tilde{q} = \arg\min_q r(q). \tag{13}$$

If $r(\tilde{q}) \ll 1$, then there is a *gap* between $\lambda_{\tilde{q}}$ and $\lambda_{\tilde{q}+1}$. Subspaces $R(\mathbf{S}_{\tilde{q}})$ and $R(\mathbf{N}_{\tilde{q}})$ are insensitive to perturbations with size $\lambda_{\tilde{q}+1}$. This fact makes us consider the problem of decomposition into signal and noise subspaces, namely, $R(\mathbf{S}_{\tilde{q}})$ and $R(\mathbf{N}_{\tilde{q}})$, *stable* or *well-conditioned*. On the other hand, absence of such a gap, i.e., $r(\tilde{q}) \approx 1$, means that there does *not* exist a clear-cut separation between the signal and the noise, making us consider the problem of decomposition into signal and noise subspaces *unstable* or *ill-conditioned*.

A. Connections with the Blind Channel Approximation Problem

In order to give a physical interpretation to our results, we may say that if the diversity of the $\tilde{m}$th-order significant part of the true channel, which is denoted by $\mathbf{h}_{\tilde{m}_1,\tilde{m}_2}$, is sufficiently large with respect to the size of the "noise"—here, "noise" is a generic term that incorporates the influence of the tails, the additive, not necessarily white, channel noise and the estimated, inexact statistics—expressed as [recall (4)]

$$\lambda_{\tilde{q}}(\tilde{\mathcal{R}}_L) \gg \|\mathbf{E}_L\|_2$$

then

$$\lambda_{\tilde{q}} \geq \lambda_{\tilde{q}}(\tilde{\mathcal{R}}_L) - \|\mathbf{E}_L\|_2 \gg \|\mathbf{E}_L\|_2 \geq \lambda_{\tilde{q}+1}$$

which means that there will exist a gap between two consecutive eigenvalues of the estimated data covariance matrix, making us consider the decomposition into signal and noise subspaces stable.

As shown in the blind channel identification context [6], if the diversity of the $\tilde{m}$th-order significant part of the channel is sufficiently large with respect to the size of the tails, then the $\tilde{m}$th-order subspace method can approximate the unknown channel sufficiently well.

The measures of diversity in these cases are not identical. In the subspace decomposition problem, the measure is $\sigma_{L+\tilde{m}+1}(\mathcal{H}_L(\mathbf{h}_{\tilde{m}_1,\tilde{m}_2}))$, whereas in the blind channel approximation problem, it is $\sigma_{2\tilde{m}+1}(\mathcal{H}_{\tilde{m}}(\mathbf{h}_{\tilde{m}_1,\tilde{m}_2}))$. These quantities are not orderable, in general, that is, one is not always larger than the other. However, since they measure the distance of $\mathcal{H}_L(\mathbf{h}_{\tilde{m}_1,\tilde{m}_2})$ and $\mathcal{H}_{\tilde{m}}(\mathbf{h}_{\tilde{m}_1,\tilde{m}_2})$, respectively, from the matrices with rank one less than the assumed rank, we expect that if one measure is "large" (resp. "small") then the other will be "large" (resp. "small"), as well. Extensive simulations have revealed that they are reasonably close. Thus, we should expect a close relationship between the stability of the decomposition of the data covariance matrix into signal and noise subspaces and the stability of the approximation of unknown channels by blind channel identification methods.

B. Determination of the Effective Channel Order with the Proposed Criterion

In Fig. 5, we plot the inverse of the rank detection criterion (13), i.e., $1/r(q)$ versus q, for the data set used for the computation of AIC(k) and MDL(k), shown in Figs. 3 and 4. We observe that the proposed criterion is insensitive to variations in the SNR. In all cases, the minimum of $r(q)$ appears at the position 23, giving 2 as the effective channel order estimate, that is, three taps for each subchannel. In all cases, there exists a gap between two consecutive eigenvalues of the estimated data covariance matrix, namely, λ_{23} and λ_{24}, making us consider the signal-noise subspace decomposition problem stable. Furthermore, in all cases, the first-order "zero-forcing" or Wiener equalizers, which are computed by the impulse responses estimated by the second-order subspace method, can open the eye. In Fig. 6, we plot the output of the first-order Wiener equalizer for the SNR $=$ 30 dB case.

In extensive simulations, we have observed that the proposed criterion is insensitive to variations in the number of data samples.

Two quite dissapointing facts concerning the behavior of information theoretic criteria are that, as shown in Figs. 3–5, in many cases

1) they *cannot* detect a gap of about two orders of magnitude between two consecutive data covariance matrix eigenvalues;
2) they associate, *erroneously*, "close" eigenvalues with different subspaces.

This means that their estimates may be poor even in cases considered stable. This is a characteristic of unstable numerical procedures, and clearly, it is unsatisfactory.

In some cases, there does *not* exist a big gap between two consecutive eigenvalues of $\mathbf{R}_L$. This is the case, for example, of *chan3.mat*, found at the same website. We compute $r(q)$ using noiseless data obtained at the output of this channel. In Fig. 7, we plot $1/r(q)$ versus q, where it seems that there is *no* clear-cut separation between the signal and the noise, making us consider the signal–noise subspace decomposition problem unstable. This fact implies that there does *not* exist an m such that the diversity of the mth-order significant part of this channel is sufficiently large, with

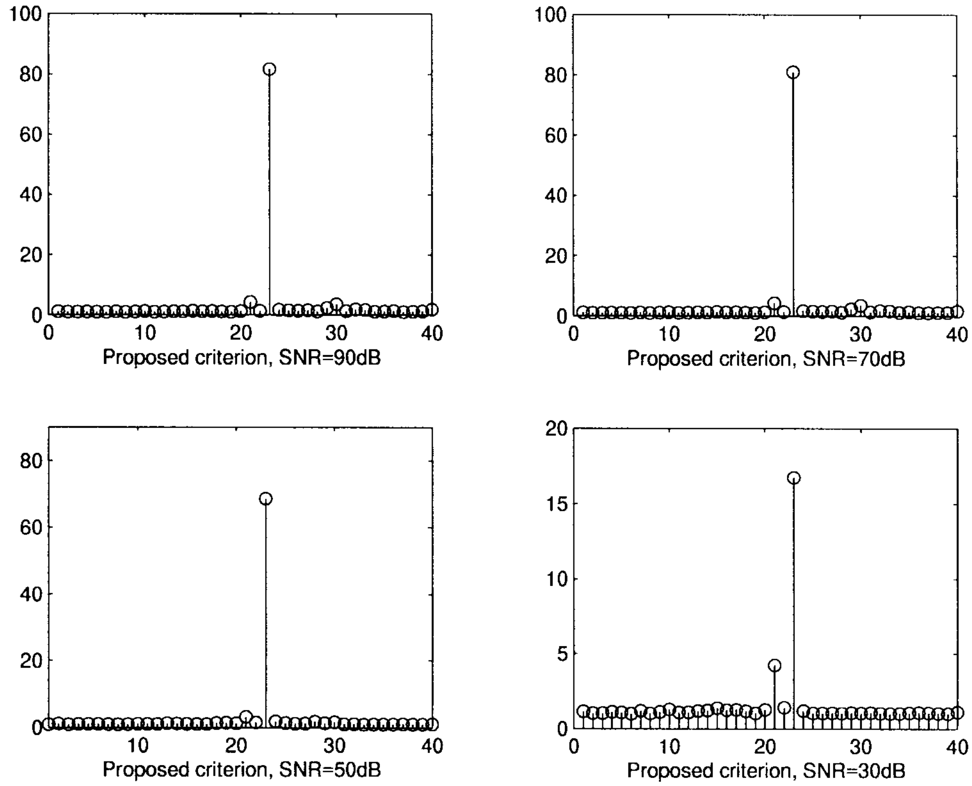


Fig. 5. Inverse of the proposed criterion, i.e., $1/r(q)$ versus q for SNR $=$ 90, 70, 50, and 30 dB.

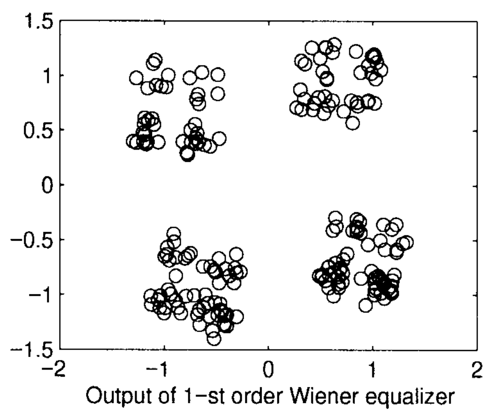


Fig. 6. Best case output of first-order Wiener equalizer (delay= 2) computed using the impulse response "identified" by the second-order subspace method (SNR $=$ 30 dB).

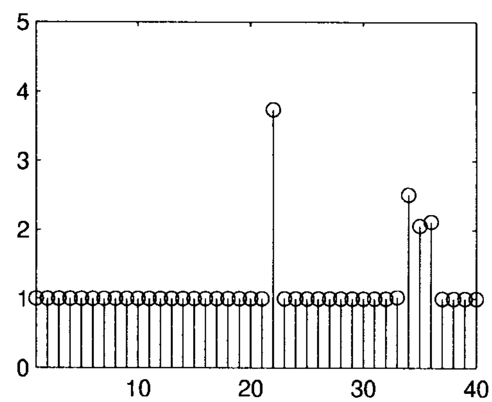


Fig. 7. Inverse of the proposed criterion, i.e., $1/r(q)$ versus q for noiseless data obtained by *chan3.mat*.

respect to the "noise." Consequently, we may anticipate that it is difficult to approximate and subsequently equalize this channel sufficiently well. In some realizations of high SNR $(\mathrm{SNR} > 40$ dB), it is possible to open the eye using the Wiener equalizers corresponding to the estimates of the subspace method with order the estimated by the proposed criterion effective channel order, i.e., 1. However, for $\mathrm{SNR} < 40$ dB, it seems difficult to approximate *chan3.mat* sufficiently well, using blind channel identification methods. Thus, the value of $r(\tilde{q})$ may provide useful information not only for the stability of the decomposition of the data covariance matrix into signal and noise subspaces but also for the stability of the blind channel approximation problem.

IV. Conclusions

Effective channel order determination is critical to the successful application of blind channel identification procedures. We considered the performance of information theoretic criteria for this task. It turns out that they are very sensitive to variations in the SNR and the number of data samples. More specificaly, for high SNR and/or many data samples, they usually lead to effective overmodeling. Such estimates are practically useless. For low SNR and few data samples, they may lead to useful effective channel order estimates. However, their high sensitivity is unsatisfactory and has impeded their successfull application to the channel order determination problem. Furthermore, it has contributed to the creation of some confusion concerning the classification into under- and over-modeled cases and the applicability of blind channel identification methods in realistic cases.

In order to avoid these shortcomings, we proposed a new criterion based on numerical analysis arguments. Using the concept of canonical angles between subspaces and invariant subspace perturbation results, we provided a "maximally stable" decomposition of the range space of the data covariance matrix into signal and noise subspaces. Simulations with realistic data have shown this criterion to be insensitive to variations in the SNR and the number of data samples.

Existence of a gap between two consecutive eigenvalues of the estimated data covariance matrix makes us consider the subspace decomposition problem stable and gives reason to believe that the blind channel approximation problem is stable as well. On the other hand, absence of such a gap makes us consider both problems unstable. In the stable cases, our criterion provides useful effective channel order estimates, leading to sufficiently good blind channel approximation/equalization; this is *not* always the case with information theoretic criteria. In the unstable cases, sufficiently good blind channel approximation seems difficult.

A very important question concerning the widespread applicability of blind channel identification methods is whether the majority of real-life microwave radio impulse responses leads to stable or unstable signal–noise subspace decompositions. In order to answer this question, extensive experimentation with measured data is needed. Thus, the development of an extensive database with measured data is of great importance.

REFERENCES

[1] L. Tong, G. Xu, and T. Kailath, "Blind identification based on second-order statistics: A time domain approach," *IEEE Trans. Inform. Theory*, vol. 40, pp. 340–349, 1994.

[2] E. Moulines, P. Duhamel, J. F. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Trans. Signal Processing*, vol. 43, pp. 516–525, Feb. 1995.

[3] G. Xu, H. Liu, L. Tong, and T. Kailath, "A least-squares approach to blind channel identification," *IEEE Trans. Signal Processing*, vol. 43, pp. 2982–2993, Dec. 1995.

[4] D. Slock, "Blind fractionally-spaced equalization, perfect reconstruction filterbanks, and multilinear prediction," in *Proc. ICASSP*, Adelaide, Australia, Apr. 1994.

[5] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 387–392, Apr. 1985.

[6] A. P. Liavas, P. A. Regalia, and J. P. Delmas, "Robustness of least-squares and subspace methods for blind channel identification/equalization with respect to effective channel undermodeling/overmodeling," *IEEE Trans. Signal Processing*, vol. 47, pp. 1636–1645, June 1999.

[7] ——, "Robustness of the linear prediction method for blind channel identification/equalization with respect to effective channel undermodeling/overmodeling," submitted for publication.

[8] ——, "Blind channel approximation: Effective channel length determination," in *Proc. 32nd Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Nov. 1998, pp. 1153–1157.

[9] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Automat. Contr.*, vol. AC-19, pp. 716–723, Dec. 1974.

[10] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, pp. 465–471, 1978.

[11] J. R. Treichler, I. Fijalkow, and C. R. Johnson, Jr., "Fractionally spaced equalizers. How long should they really be?," *IEEE Signal Processing Mag.*, pp. 65–81, May 1996.

[12] G. Stewart and J. Sun. *Matrix Perturbation Theory*. New York: Academic, 1990.

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

[14] G. W. Stewart, "Error and perturbation bounds for subspaces associated with certain eigenvalue problems," *SIAM Rev.*, vol. 15, no. 4, pp. 727–764, Oct. 1973.

[15] G. Golub and C. V. Loan, *Matrix Computations*, 2nd ed. Baltimore, MD: Johns Hopkins Univ. Press, 1991.

Athanasios P. Liavas (M'94) was born in Pyrgos, Greece, in 1966. He received the diploma and the Ph.D. degrees in computer engineering from the University of Patras, Patras, Greece, in 1989 and 1993, respectively.

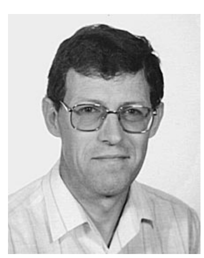
From 1996 to 1998, he was a research fellow at the Institut National des Télécommunications, Evry, France, under the framework of the Training and Mobility of Researchers (T.M.R.) program of the European Commission. He is currently with the Department of Computer Science, University of Ioannina, Ioannina, Greece. His research interests include adaptive signal processing algorithms, blind system identification, and biomedical signal processing.

Dr. Liavas is a member of the Technical Chamber of Greece.

Phillip A. Regalia (SM'96) was born in Walnut Creek, CA, in 1962. He received the B.Sc. degree with Highest Honors and the M.Sc. and Ph.D. degrees in electrical and computer engineering from the University of California, Santa Barbara, in 1985, 1987, and 1988, respectively, and the Habilitation à Diriger des Recherches degree from the University of Paris, Orsay, France, in 1994.

He is presently a Professor at the Institut National des Télécommunications, Evry, France, with research interests focused in adaptive signal processing.

Dr. Regalia serves as an associate editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and as an editor for the *International Journal of Adaptive Control and Signal Processing*.



Jean-Pierre Delmas was born in France in 1950. He received the engineering degree from the Ecole Centrale de Lyon, Lyon, France, in 1973 and the Certificat d' études supérieures from the Ecole Nationale Supérieure des Télécommunications, Paris, France, in 1982.

Since 1980, he has been with the Institut National des Télécommunications, Evry, France, where he is currently Maitre de Conférences. His research interests are in statistical signal processing.