Subspace methods for blind symbol estimation: The unknown channel order case

Athanasios P. Liavas

30 September 1999

Abstract

We consider subspace-based methods for blind symbol estimation when the "known channel order" assumption is violated; the estimated channel length is l, while the true channel length is L, with $L > l$. Using the concepts of the *length-l significant part* and the *unmodeled tails*, we show that if the size of the unmodeled tails is small with respect to the diversity of the length-l significant part, then the algorithm performs well; otherwise, it may perform poorly. The generically ill-conditioned case of effective overmodeling results when our channel model attempts to model "small" leading and/or trailing channel terms.

EDICS Category: 3-CEQU: Channel Modeling, Estimation and Equalization (Correspondence).

The author is with the Department of Computer Science, University of Ioannina, 45110 Ioannina, Greece. E-mail: `liavas@cs.uoi.gr`.

This work was supported by the EPETII Program of the G.S.R.T.

1 Introduction

Many methods have appeared recently, claiming perfect blind signal recovery in the single-input/multi-output channel context, under the so-called zero-forcing conditions [1]–[4]. However, the validity of the zero-forcing conditions is very difficult to guarantee in a "real-life" scenario. In particular, the assumption that the true channel be FIR of known order is rarely true. Taking into account that most of the algorithms so derived are very sensitive to the true channel order, a theoretical analysis of the behavior of the algorithms under realistic conditions seems of great importance.

We consider the so-called row-span methods for blind symbol estimation [3], [4], when the "known channel order" assumption is violated; the estimated channel length is l, while the true channel length is L, with $L > l$. Using the concepts of the length-l significant part and the unmodeled tails [5], we show that if the size of the unmodeled tails is small with respect to the diversity of the length-l significant part, as measured by the smallest nonzero singular value of a certain filtering matrix, then the algorithm performs well; otherwise, it may perform poorly. The generically ill-conditioned case of effective overmodeling appears when our channel model attempts to model "small" leading and/or trailing channel terms. These results show a striking similarity to those concerning subspace-based blind channel identification methods [5].

2 Row-span methods

We consider a single-input/two-output baseband FIR channel model; extension to the single-input/p-output case, with $p > 2$, is trivial. If the true channel length is L, then the 2-dimensional output $\mathbf{x}_n$ is given by the convolution $\mathbf{x}_n = \sum_{k=0}^{L-1} \mathbf{h}_k \, s_{n-k}$, where $\{\mathbf{h}_k\}_{k=0}^{L-1}$ is the channel impulse response and $\{s_n\}$ is the scalar-valued input sequence. We denote the entire channel parameter vector as $\mathbf{h} \triangleq [\, \mathbf{h}_0^T \; \cdots \; \mathbf{h}_{L-1}^T \,]^T$, where superscript ^{T} denotes transpose.

The subspace-based methods of [3] and [4] exploit the structure of the $2m \times (N - m + 1)$

data matrix

$$\mathcal{X}(\mathbf{h}) \triangleq \begin{bmatrix} \mathbf{x}_0 & \mathbf{x}_1 & \cdots & \mathbf{x}_{N-m} \\ \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_{N-m+1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_{m-1} & \mathbf{x}_m & \cdots & \mathbf{x}_{N-1} \end{bmatrix}$$

which can be expressed as $\mathcal{X}(\mathbf{h}) = \mathcal{H}(\mathbf{h})\ \mathcal{S}_{N-1}^{-L+1}$, where

$$\mathcal{H}(\mathbf{h}) \triangleq \begin{bmatrix} \mathbf{h}_{L-1} & \cdots & \mathbf{h}_0 & & \\ & \ddots & & \ddots & \\ & & \mathbf{h}_{L-1} & \cdots & \mathbf{h}_0 \end{bmatrix} \quad \text{and} \quad \mathcal{S}_{N-1}^{-L+1} \triangleq \begin{bmatrix} s_{-L+1} & s_{-L+2} & \cdots & s_{N-m-L+1} \\ s_{-L+2} & s_{-L+3} & \cdots & s_{N-m-L+2} \\ \vdots & \vdots & \ddots & \vdots \\ s_{m-1} & s_m & \cdots & s_{N-1} \end{bmatrix}$$

with dimensions $2m \times (L+m-1)$ and $(L+m-1) \times (N-m+1)$, respectively. If the subchannels of $\mathbf{h}$ do not share common zeros and $m \geq L-1$, then $\mathcal{H}(\mathbf{h})$ is a tall full-column rank matrix. Consequently, $\text{row}(\mathcal{X}(\mathbf{h})) = \text{row}\left(\mathcal{S}_{N-1}^{-L+1}\right)$, where $\text{row}(\mathcal{A})$ denotes the row space of matrix $\mathcal{A}$, enabling the identification of the input sequence $\mathbf{s}_{N-1}^{-L+1} \triangleq [s_{-L+1} \ \cdots \ s_{N-1}]^T$ [3], [4].

In the sequel, we focus on the subspace intersection method of [4]. However, our results apply to the null space union method of [3], as well, because the two methods are equivalent, even in the presence of noise [4]. At first, one performs an SVD of $\mathcal{X}(\mathbf{h})$, $\mathcal{X}(\mathbf{h}) = U\Sigma V$, and obtains V, whose rows form an orthonormal basis for $\text{row}(\mathcal{X}(\mathbf{h}))$. Under sufficient richness conditions on the input, $\text{row}(\mathcal{X}(\mathbf{h}))$ is an $(L+m-1)$-dimensional subspace. Setting $n = L+m-1$, $\mathbf{V}_{T(n)}$ is constructed as

$$\mathbf{V}_{T(n)} \triangleq \begin{bmatrix} V^{(1)} \\ \vdots \\ V^{(n)} \end{bmatrix} \quad \text{with} \quad V^{(k)} \triangleq \begin{bmatrix} \mathbf{0} & V & \mathbf{0} \\ I_{k-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & I_{n-k} \end{bmatrix} \tag{1}$$

and the input sequence $\mathbf{s}_{N-1}^{-L+1}$ is identified, to within a scaling factor, as the right singular vector of $\mathbf{V}_{T(n)}$ associated with its largest singular value.

However, in practice, the true channels are usually long, i.e., $L > m$, and they are composed of a *significant part* and "small" leading and trailing *tails*. In such a case, $\mathcal{H}(\mathbf{h})$ is not a tall full-column rank matrix and $\mathrm{row}(\mathcal{X}(\mathbf{h})) \neq \mathrm{row}(\mathcal{S}_{N-1}^{-L+1})$. Thus, is it possible, in such cases, to use $\mathrm{row}(\mathcal{X}(\mathbf{h}))$ to identify or, at least, estimate the input sequence?

To answer this question, we shall describe a more realistic implementation of the method. In practice, one first estimates the *effective rank* r ($r \leq 2m$) of data matrix $\mathcal{X}(\mathbf{h})$, by adopting, e.g., the criterion developed in [6]. The truncated rank-r data matrix $\widehat{\mathcal{X}}(\mathbf{h})$ and matrix $\widehat{V}$, whose rows form an orthonormal basis for $\widehat{\mathcal{V}} = \mathrm{row}(\widehat{\mathcal{X}}(\mathbf{h}))$, are related through $\widehat{\mathcal{X}}(\mathbf{h}) = \widehat{U}\,\widehat{\Sigma}\,\widehat{V}$, where the variables with the hats are associated with the r largest singular values of $\mathcal{X}(\mathbf{h})$. The *effective channel length* l is estimated as $l = r - m + 1$. Then, we may write

$$\mathcal{X}(\mathbf{h}) = \underbrace{\mathcal{H}(\mathbf{h}^{\mathrm{zs}})\, \mathcal{S}_{N-1}^{-L+1}}_{\mathcal{X}(\mathbf{h}^{\mathrm{zs}})} + \mathcal{E}, \quad \text{with} \quad \mathcal{E} \triangleq \mathcal{H}(\mathbf{h} - \mathbf{h}^{\mathrm{zs}})\mathcal{S}_{N-1}^{-L+1},$$

where $\mathbf{h}^{\mathrm{zs}}$ is the *zero-padded* length-l significant part of the true channel, with the nonzero terms lying between positions $m1$ and $m2$ (see [5, eqs. (3)–(6)]), and $\mathcal{X}(\mathbf{h}^{\mathrm{zs}})$ is the associated rank-r data matrix. Since $m \geq l - 1$, the $2m \times (l + m - 1)$ filtering matrix $\mathcal{H}(\mathbf{h}^{\mathrm{s}})$, associated with the *truncated* length-l significant part of the channel $\mathbf{h}^{\mathrm{s}}$ (see [5, eq. 7]), is a tall matrix. If the subchannels of $\mathbf{h}^{\mathrm{s}}$ do not share common zeros, then $\mathcal{H}(\mathbf{h}^{\mathrm{s}})$ is of full-column rank. As a result,

$$\mathcal{V}^{\mathrm{s}} \triangleq \mathrm{row}(\mathcal{X}(\mathbf{h}^{\mathrm{zs}})) = \mathrm{row}\left(\mathcal{H}(\mathbf{h}^{\mathrm{zs}})\mathcal{S}_{N-1}^{-L+1}\right) = \mathrm{row}\left(\mathcal{H}(\mathbf{h}^{\mathrm{s}})\mathcal{S}_{N-m1-1}^{-m2}\right) = \mathrm{row}\left(\mathcal{S}_{N-m1-1}^{-m2}\right).$$

Thus, if $\mathrm{row}(\mathcal{X}(\mathbf{h}^{\mathrm{zs}}))$ were known, the subspace intersection method of [4] would permit the identification of the input sequence $\mathbf{s}_{N-m1-1}^{-m2}$, through the sequence of computations:

$$\mathcal{X}(\mathbf{h}^{\mathrm{zs}}) \rightarrow \mathcal{V}^{\mathrm{s}} \rightarrow V^{\mathrm{s}} \rightarrow \mathbf{V}_{T(r)}^{\mathrm{s}} \rightarrow \mathbf{s}_{N-m1-1}^{-m2} \equiv \mathbf{s},$$

where the rows of matrix V^{s} form an orthonormal basis for $\mathcal{V}^{\mathrm{s}}$ and $\mathbf{V}_{T(r)}^{\mathrm{s}}$ is built from V^{s}, as in (1). Note though that $\mathrm{row}(\mathcal{X}(\mathbf{h}^{\mathrm{zs}}))$, which has been introduced just for analysis purposes, is unknown. However, by continuity arguments, one would expect that if $\mathcal{E}$ is "sufficiently" small, then knowledge of $\mathrm{row}(\widehat{\mathcal{X}}(\mathbf{h}))$ may permit an approximation to $\mathbf{s}_{N-m1-1}^{-m2}$, through

$$\widehat{\mathcal{X}}(\mathbf{h}) \rightarrow \widehat{\mathcal{V}} \rightarrow \widehat{V} \rightarrow \widehat{\mathbf{V}}_{T(r)} \rightarrow \widehat{\mathbf{s}}_{N-m1-1}^{-m2} \equiv \widehat{\mathbf{s}}.$$

In the sequel, we provide an upper bound for the sine of the angle between the unit 2-norm vectors $\mathbf{s}$ and $\widehat{\mathbf{s}}$, revealing, in this way, the factors that govern the behavior of the subspace intersection method.

Recall that $\mathcal{V}^{\mathrm{s}} = \mathrm{row}(\mathcal{X}(\mathbf{h}^{\mathrm{zs}}))$ and $\widehat{\mathcal{V}} = \mathrm{row}(\widehat{\mathcal{X}}(\mathbf{h}))$ are r-dimensional subspaces associated with the r largest singular values of matrices $\mathcal{X}(\mathbf{h}^{\mathrm{zs}})$ and $\mathcal{X}(\mathbf{h})$, respectively. These matrices are related through perturbation $\mathcal{E}$. Thus, as shown in [8] (see also [7, p. 261])

$$\|\sin\angle(\widehat{\mathcal{V}}, \mathcal{V}^{\mathrm{s}})\|_F \leq \frac{\|\mathcal{E}\|_F}{\sigma_{\min}(\mathcal{X}(\mathbf{h}^{\mathrm{zs}})) - \|\mathcal{E}\|_2} \equiv \mathcal{B}_1 \tag{2}$$

where $\sin\angle(\widehat{\mathcal{V}}, \mathcal{V}^{\mathrm{s}})$ is the diagonal matrix with elements the sines of the canonical angles θ_i between subspaces $\widehat{\mathcal{V}}$ and $\mathcal{V}^{\mathrm{s}}$, $\|\cdot\|_F$ and $\|\cdot\|_2$ denote the matrix Frobenious and 2-norm, respectively, and $\sigma_{\min}(\mathcal{A})$ denotes the smallest nonzero singular value of matrix $\mathcal{A}$.

If the rows of matrix V^{s} form an orthonormal basis for $\mathcal{V}^{\mathrm{s}}$, then we can find matrix $\widehat{V}$, whose rows form an orthonormal basis for $\widehat{\mathcal{V}}$, such that [7, p. 95]

$$\|\widehat{V} - V^{\mathrm{s}}\|_F \leq \sqrt{2\sum_{i=1}^{r}(1-\cos\theta_i)} \leq \sqrt{2}\,\mathcal{B}_1 \equiv \mathcal{B}_2. \tag{3}$$

In this case, $\widehat{V}$ and V^{s} are as close as possible, with respect to the metric associated with the matrix Frobenious norm. Now, the structure of $\mathbf{V}^{\mathrm{s}}_{T(r)}$ and (3) imply

$$\|\widehat{\mathbf{V}}_{T(r)} - \mathbf{V}^{\mathrm{s}}_{T(r)}\|_F = \sqrt{r}\,\|\widehat{V} - V^{\mathrm{s}}\|_F \leq \sqrt{r}\,\mathcal{B}_2 \equiv \mathcal{B}_3. \tag{4}$$

Finally, since $\mathbf{s}$ and $\widehat{\mathbf{s}}$ are the largest right singular vectors of $\mathbf{V}^{\mathrm{s}}_{T(r)}$ and $\widehat{\mathbf{V}}_{T(r)}$, respectively, we obtain [8]:

$$\sin\angle(\widehat{\mathbf{s}}, \mathbf{s}) \leq \frac{\|\widehat{\mathbf{V}}_{T(r)} - \mathbf{V}^{\mathrm{s}}_{T(r)}\|_2}{\sigma_1\left(\mathbf{V}^{\mathrm{s}}_{T(r)}\right) - \sigma_2\left(\mathbf{V}^{\mathrm{s}}_{T(r)}\right) - \|\widehat{\mathbf{V}}_{T(r)} - \mathbf{V}^{\mathrm{s}}_{T(r)}\|_2} \leq \frac{\mathcal{B}_3}{\sigma_1\left(\mathbf{V}^{\mathrm{s}}_{T(r)}\right) - \sigma_2\left(\mathbf{V}^{\mathrm{s}}_{T(r)}\right) - \mathcal{B}_3}. \tag{5}$$

We observe that the perturbation on $\mathbf{V}^{\mathrm{s}}_{T(r)}$ and the gap between the two largest singular values of $\mathbf{V}^{\mathrm{s}}_{T(r)}$ govern the behavior of the subspace intersection algorithm. If the perturbation is small with respect to the gap, then the algorithm performs well; otherwise, it may perform poorly.

Result (5) is difficult to interpet in terms of channel characteristics and estimated effective channel length, unless we perform an asymptotic (as $N \to \infty$) analysis. In this case, and for zero-mean, unit-variance, i.i.d. input sequence $\{s_n\}$, we obtain $\frac{1}{N} S\, S^T \to I$ [4]. Consequently,

$$\sigma_i\left(\mathcal{E}\right) = \sigma_i(\mathcal{H}(\mathbf{h} - \mathbf{h}^{\text{zs}})\, S) \to \sqrt{N}\, \sigma_i\left(\mathcal{H}(\mathbf{h} - \mathbf{h}^{\text{zs}})\right),$$

$$\sigma_{\min}\left(\mathcal{X}(\mathbf{h}^{\text{zs}})\right) = \sigma_{\min}\left(\mathcal{H}(\mathbf{h}^{\text{zs}})\, S\right) \to \sqrt{N}\, \sigma_{\min}\left(\mathcal{H}(\mathbf{h}^{\text{s}})\right),$$

where $\sigma_i(\mathcal{A})$ denotes the i–th singular value of matrix $\mathcal{A}$. Bound (2) becomes:

$$\|\sin\angle(\widehat{\mathcal{V}}, \mathcal{V})\|_F \leq \frac{\|\mathcal{H}(\mathbf{h} - \mathbf{h}^{\text{zs}})\|_F}{\sigma_{\min}(\mathcal{H}(\mathbf{h}^{\text{s}})) - \|\mathcal{H}(\mathbf{h} - \mathbf{h}^{\text{zs}})\|_2} \leq \frac{\sqrt{m}\, \|\mathbf{h} - \mathbf{h}^{\text{zs}}\|_2}{\sigma_{\min}(\mathcal{H}(\mathbf{h}^{\text{s}})) - \sqrt{m}\, \|\mathbf{h} - \mathbf{h}^{\text{zs}}\|_2}. \quad (6)$$

Thus, if the size of the unmodeled part $\|\mathbf{h} - \mathbf{h}^{\text{zs}}\|_2$ is small with respect to the smallest nonzero singular value of filtering matrix $\mathcal{H}(\mathbf{h}^{\text{s}})$, then subspaces $\widehat{\mathcal{V}}$ and $\mathcal{V}^{\text{s}}$ and, consequently, matrices $\widehat{\mathbf{V}}_{T(r)}$ and $\mathbf{V}^{\text{s}}_{T(r)}$ are close each other; term $\sigma_{\min}(\mathcal{H}(\mathbf{h}^{\text{zs}}))$, relative to the largest singular value of $\mathcal{H}(\mathbf{h}^{\text{s}})$, may be interpreted as a measure of diversity of channel $\mathbf{h}^{\text{s}}$ (see also [5]). Furthermore, for large N, one may show [4] that $\sigma_1\left(\mathbf{V}^{\text{s}}_{T(r)}\right) = \sqrt{r}$ and $\sigma_2\left(\mathbf{V}^{\text{s}}_{T(r)}\right) \to \sqrt{r-1}$. Thus, finally, if $\|\widehat{\mathbf{V}}_{T(r)} - \mathbf{V}^{\text{s}}_{T(r)}\|_F$ is small with respect to $(\sqrt{r} - \sqrt{r-1})$, then estimate $\widehat{\mathbf{s}}$ will be "close" to the true input $\mathbf{s}$.

The above analysis shows great similarity to the one developed in [5], concerning blind channel identification using subspace techniques. In both approaches, the diversity of the significant part of the channel and the size of the unmodeled part govern the performance of the methods. The derived bounds are not tight, in general. However, 1) they provide an indication for the performance of the method, and 2) making use of theorem 2 of [5], they imply that the effective channel length or, equivalently, the effective rank of data matrix $\mathcal{X}(\mathbf{h})$, should be chosen in such a way so that "small" leading and trailing terms be excluded from our channel model; otherwise, an effective overmodeling case results, leading to potentially poor performance of the method.

3 Simulation

We illustrate the above concepts by passing a BPSK data sequence $\mathbf{s}_{99}^{0}$ through the real part of the oversampled, by a factor of 2, *chan4* found at *http://spib.rice.edu/spib/microwave.html*; the data before time instant $n = 0$ are zero. In Fig. 1, we plot a portion of the subchannels of *chan4*; intuitively satisfying estimates of the significant subchannel length are 2 or 3. We want to recover the input sequence using equalizers of length $m = 8$. In Fig. 2, we plot the effective rank detection criterion [6] $r(q) = \sigma_q(\mathcal{X}(\mathbf{h}))/\sigma_{q+1}(\mathcal{X}(\mathbf{h}))$, for $q = 1, \ldots, 15$. The estimated effective rank is

$$r = \arg\max_{q}\ r(q) = 10,$$

leading to the effective channel length estimate $l = 3$. We perform the subspace intersection method of [4] for channel lengths $l = 1, \ldots, 9$. In Fig. 3, we plot the cosine between the original input sequence $\mathbf{s}_{99}^{0}$ and the estimated ones, for the various channel lengths. We observe that we obtain the best performance for channel length $l = 3$. For channel length $l < 3$, the estimation is poor due to undermodeling error, while for $l > 3$ it is poor due to lack of diversity. The criterion developed in [6] proves to be useful in practice.

4 Conclusion

We performed a theoretical analysis of the performance of subspace methods for blind symbol estimation, when the true channel length is L, while the estimated channel length is l, with $L > l$. By partitioning the true channel impulse response into the length-l significant part and the unmodeled tails we showed that if the size of the unmodeled tails is small with respect to the diversity of the length-l significant part, then the algorithm performs well; otherwise, it may perform poorly. Effective overmodeling results when the length-l significant part includes small leading and/or trailing terms. Such cases are generically ill-conditioned (see also ([5]) and should be avoided.

References

[1] L. Tong, G. Xu and T. Kailath, "Blind identification based on second-order statistics: A time domain approach," *IEEE Trans. Information Theory*, vol. 40, no. 2, 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, February 1995.

[3] H. Liu and G. Xu, "Closed–form blind symbol estimation in digital communications," *IEEE Trans. Signal Processing*, vol. 43, no. 11, pp. 2714–2723, November 1995.

[4] A.-J. van der Veen, S. Talwar and A. Paulraj, "A subspace approach to blind space–time signal processing for wireless communication systems," *IEEE Trans. Signal Processing*, vol. 45. no. 1, pp. 173–190, January 1997.

[5] 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, no. 6, pp. 1636–1645, June 1999.

[6] A. P. Liavas, P. A. Regalia and J.-P. Delmas, "Blind channel approximation: Effective channel order determination," *IEEE Trans. Signal Processing*, vol. 47, no. 12, December 1999 (to appear).

[7] G. Stewart and J. Sun. *Matrix perturbation theory*. Academic Press, 1990.

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

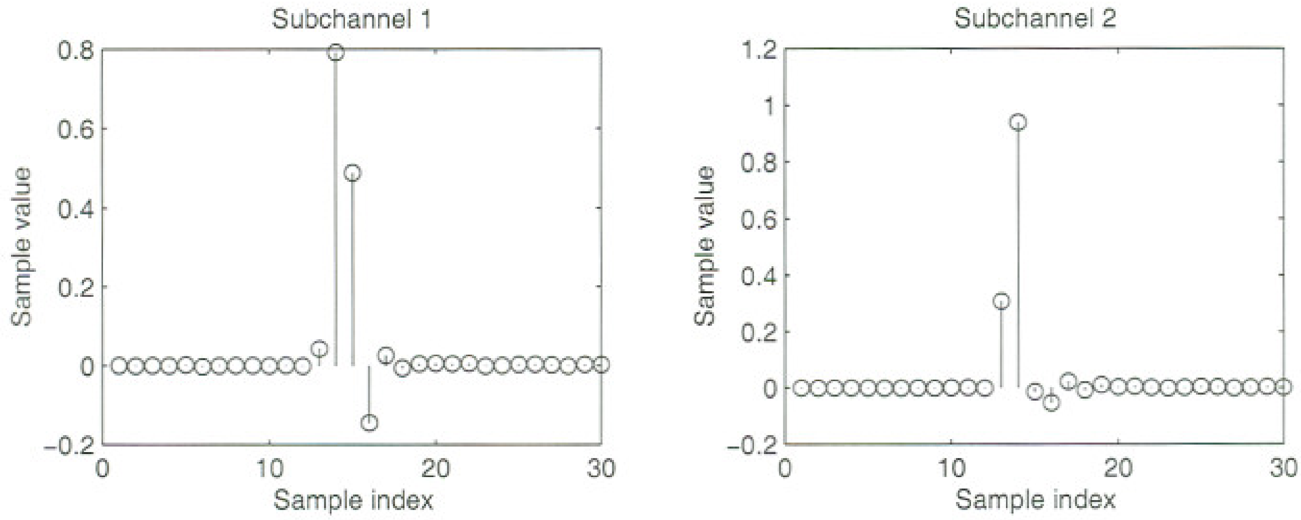

Fig. 1. Portion of the real part of the subchannels of *chan4.mat*.

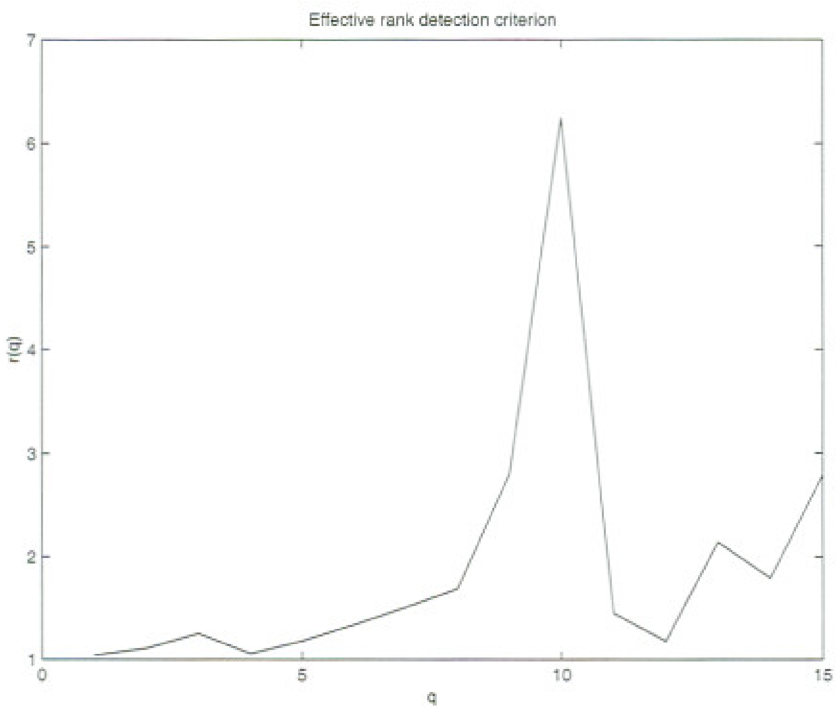

Fig. 2. Plot of $r(q) = \frac{\sigma_q(\mathcal{X}(\mathbf{h}))}{\sigma_{q+1}(\mathcal{X}(\mathbf{h}))}$ versus q

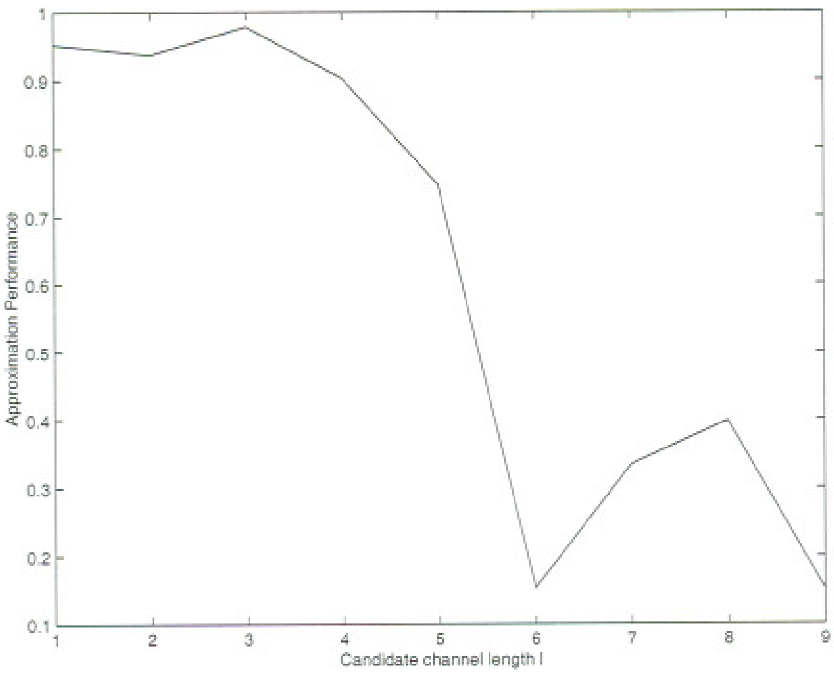

Fig. 3. Cosine of the angle between $\mathbf{s}_{99}^{0}$ and $\widehat{\mathbf{s}}_{99}^{0}$ versus the candidate channel length.