Asymptotic Eigenvalue Distribution of Block Toeplitz Matrices and Application to Blind SIMO Channel Identification

Houcem Gazzah, *Student Member, IEEE*, Phillip A. Regalia, *Fellow, IEEE*, and Jean-Pierre Delmas, *Member, IEEE*

***Abstract*—Szegö's theorem states that the asymptotic behavior of the eigenvalues of a Hermitian Toeplitz matrix is linked to the Fourier transform of its entries. This result was later extended to block Toeplitz matrices, i.e., covariance matrices of multivariate stationary processes. The present work gives a new proof of Szegö's theorem applied to block Toeplitz matrices. We focus on a particular class of Toeplitz matrices, those corresponding to covariance matrices of single-input multiple-output (SIMO) channels. They satisfy some factorization properties that lead to a simpler form of Szegö's theorem and allow one to deduce results on the asymptotic behavior of the lowest nonzero eigenvalue for which an upper bound is developed and expressed in terms of the subchannels frequency responses. This bound is interpreted in the context of blind channel identification using second-order algorithms, and more particularly in the case of band-limited channels.**

***Index Terms*—Asymptotic eigenvalue distribution, band-limited channels, blind identification, block Toeplitz matrices, multivariate processes, second-order statistics algorithms.**

I. Introduction

In a celebrated result appearing in [1], Szegö states that the eigenvalues of a sequence of Hermitian Toeplitz matrices are asymptotically distributed like the samples of the Fourier transform of its entries. The lowest/highest eigenvalues are decreasing/increasing and converge to the minimum/maximum of this Fourier transform. The application of this result to covariance matrices of scalar stationary[1] processes is straightforward. Several extensions have since been made (see [2]). The most important extends Szegö's theorem to block Toeplitz matrices with non-Toeplitz blocks where the number of blocks tends to infinity [3], [4]. However, the proof made therein relies on sophisticated mathematics. In this correspondence, we suggest a simpler proof than that in [3], [4] of the extension of the Szegö theorem to block Toeplitz structured matrices. We use the asymptotic equivalence of matrix sequences and more particularly the result established by Gray in [5] on asymptotic equivalence of Toeplitz matrix sequences and circulant matrix sequences. We focus then on a special class of block Toeplitz matrices, frequently encountered in signal processing, to give a simpler form of the Szegö theorem and deduce results about the lowest nonzero eigenvalue, which expresses the conditioning with respect to inversion of such matrices.

We target in particular second-order statistics based blind identification algorithms of single-input multiple-output (SIMO) channels where channel output covariance matrices are manipulated in such a way that the performance of the algorithms depends heavily on how well-conditioned the matrix is [6], [7]. Therefore, the interest in eigenvalues (and more particularly the lowest nonzero eigenvalue) of block Toeplitz matrices is highly justified and constitutes the subject of this correspondence.

This correspondence is organized as follows. In Section II, results on asymptotic equivalence of Toeplitz matrix sequences as well as Szegö's theorem are reviewed for convenience of the reader and in order to fix notations. In Section III, we propose a new proof of Szegö's theorem extended to block Toeplitz matrices with non-Toeplitz blocks where the number of blocks tends to infinity. We address then a specific class of block Toeplitz matrices, that of SIMO channel covariance matrices. In Section IV, implications for blind channel identification are discussed and the case of band-limited channels is particularly addressed.

II. Notation and Previous Results

Let $\{t_k\}_{k=\ldots,-1,0,1,\ldots}$ be an absolutely summable infinite complex sequence (i.e., $\sum_k |t_k| < \infty$) so that the associated 2π-periodic Fourier transform $t(w) \hat{=} \sum_k t_k e^{-ikw}$ is well defined. We define the infinite matrix sequence $\{\boldsymbol{T}_n(t)\}_{n\geq 1}$ where $\boldsymbol{T}_n(t)$ is the $n \times n$ Toeplitz matrix given by

$$\boldsymbol{T}_n(t) \hat{=} \begin{bmatrix} t_0 & t_{-1} & \cdots & t_{-(n-1)} \\ t_1 & \ddots & \ddots & t_{-(n-2)} \\ \vdots & \ddots & & \vdots \\ t_{n-1} & t_{n-2} & \cdots & t_0 \end{bmatrix}.$$

Consider a sequence of $n \times n$ matrices $\boldsymbol{A}_n$. To study the asymptotic equivalence of sequences of such matrices, two norms have been introduced [5]. The strong norm $\|\boldsymbol{A}_n\|$ and the weak norm $|\boldsymbol{A}_n|$ are defined, respectively, as the spectral norm

$$\|\boldsymbol{A}_n\|^2 \hat{=} \max_{\|\boldsymbol{x}\|=1} \boldsymbol{x}^H \boldsymbol{A}_n^H \boldsymbol{A}_n \boldsymbol{x}$$

and as the normalized Frobenius norm

$$|\boldsymbol{A}_n|^2 \hat{=} \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} |a_{i,j}|^2.$$

$\sigma_k(\boldsymbol{A})$ (resp., $\lambda_k(\boldsymbol{A})$) refers to the kth largest singular value (resp., eigenvalue) of the matrix (resp., the square matrix) $\boldsymbol{A}$. $\boldsymbol{K}_{r,s}$ represents the vec-permutation matrix [8] such that $\mathrm{Vec}(\boldsymbol{A}) = \boldsymbol{K}_{r,s}\mathrm{Vec}(\boldsymbol{A}^T)$ for all $r \times s$ matrices $\boldsymbol{A}$. It satisfies $\boldsymbol{K}_{r,s} = \boldsymbol{K}_{s,r}^{-1} = \boldsymbol{K}_{s,r}^T$.

The following definitions and results, necessary for Section III, are recalled.

Definition 1: Asymptotic equivalence [5].

Two matrix sequences $\{\boldsymbol{A}_n\}$ and $\{\boldsymbol{B}_n\}$, $n = 1, 2, \ldots$ are said to be asymptotically equivalent and noted $\{\boldsymbol{A}_n\} \sim \{\boldsymbol{B}_n\}$ if

$$\exists M < \infty \text{ such that } \forall n,\ \|\boldsymbol{A}_n\| \leq M \text{ and } \|\boldsymbol{B}_n\| \leq M \tag{1}$$

$$\lim_{n\to\infty} |\boldsymbol{A}_n - \boldsymbol{B}_n| = 0. \tag{2}$$

Lemma 1 [5]: If $\{\boldsymbol{A}_n\} \sim \{\boldsymbol{B}_n\}$ and if

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \lambda_k^s(\boldsymbol{A}_n)$$

Manuscript received November 26, 1999; revised July 14, 2000.

The authors are with the Département Signal et Image, Institut National des Télécommunications (INT), 91011 Evry Cedex, France (e-mail: gazzah@sim.int-evry.fr; Phillip.Regalia@int-evry.fr; Jean-Pierre.Delmas@int-evry.fr).

Communicated by J. A. O'Sullivan, Associate Editor for Detection and Estimation.

Publisher Item Identifier S 0018-9448(01)00585-5.

[1]Throughout the correspondence, stationary processes denote second-order stationary processes.

exists and is finite for any positive integer s, then

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \lambda_k^s(\boldsymbol{A}_n) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \lambda_k^s(\boldsymbol{B}_n.$$

Lemma 2 [5]: For all absolutely summable sequences $\{t_k\}_{k=\ldots,-1,0,1,\ldots}$, there exists a sequence of circulant matrices $\{\boldsymbol{C}_n(t)\}$ asymptotically equivalent to $\{\boldsymbol{T}_n(t)\}$ and given by $\boldsymbol{C}_n(t) = \boldsymbol{U}_n^H \boldsymbol{D}_n(t) \boldsymbol{U}_n$ where $\boldsymbol{D}_n(t)$ is a diagonal matrix with its kth entry given by

$$(\boldsymbol{D}_n(t))_{k,k} = t\left(\frac{2\pi(k-1)}{n}\right)$$

and $\boldsymbol{U}_n$ is the unitary discrete Fourier transform (DFT) matrix

$$(\boldsymbol{U}_n)_{k,l} = \frac{1}{\sqrt{n}} e^{-i2\pi \frac{(k-1)(l-1)}{n}}.$$

We will make use of the fact that the eigenvectors of $\boldsymbol{C}_n(t)$ are independent of the sequence $\{t_k\}$ and that its eigenvalues are equally spaced samples of the Fourier transform $t(w)$. We finally recall the Szegö result for a Toeplitz matrix that we will extend in Section III to the block Toeplitz matrices:

Theorem 1—Szegö's Theorem [1]: For all absolutely summable sequences $\{t_k\}_{k=\ldots,-1,0,1,\ldots}$, if $\boldsymbol{T}_n(t)$ is Hermitian, then for all functions F continuous on $[\min_w t(w), \max_w t(w)]$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} F(\lambda_k(\boldsymbol{T}_n(t))) = \frac{1}{2\pi} \int_{-\pi}^{\pi} F(t(w))\, dw.$$

Theorem 2 [1], [9]: For all absolutely summable sequences $\{t_k\}_{k=\ldots,-1,0,1,\ldots}$, if $\boldsymbol{T}_n(t)$ is Hermitian, then, for any l, the lowest (resp., largest) l eigenvalues of $\boldsymbol{T}_n(t)$ are decreasing (resp., increasing) with n and converge to $\min_w t(w)$ (resp., $\max_w t(w)$).

III. Block Toeplitz Matrices

A. Definitions

To extend the preceding results to block Toeplitz matrices, we define the block Toeplitz matrix

$$\mathcal{T}_n'(\{t^{u,v}\}) = \begin{bmatrix} \boldsymbol{T}_0 & \boldsymbol{T}_{-1} & \cdots & \boldsymbol{T}_{-(n-1)} \\ \boldsymbol{T}_1 & \ddots & \ddots & \boldsymbol{T}_{-(n-2)} \\ \vdots & \ddots & & \vdots \\ \boldsymbol{T}_{n-1} & \boldsymbol{T}_{n-2} & \cdots & \boldsymbol{T}_0 \end{bmatrix} \tag{3}$$

where $(\boldsymbol{T}_k)_{k=-(n-1),\ldots,n-1}$ are $c \times c$ matrices (not necessarily Toeplitz) of entries $t_k^{u,v} \hat{=} (\boldsymbol{T}_k)_{u,v}$, $u, v = 1, \ldots, c$. We consider the associated matrix

$$\mathcal{T}_n(\{t^{u,v}\}) \hat{=} \begin{bmatrix} \boldsymbol{T}_n(t^{1,1}) & \boldsymbol{T}_n(t^{1,2}) & \cdots & \boldsymbol{T}_n(t^{1,c}) \\ \boldsymbol{T}_n(t^{2,1}) & \boldsymbol{T}_n(t^{2,2}) & \cdots & \boldsymbol{T}_n(t^{2,c}) \\ \vdots & & & \vdots \\ \boldsymbol{T}_n(t^{c,1}) & \boldsymbol{T}_n(t^{c,2}) & \cdots & \boldsymbol{T}_n(t^{c,c}) \end{bmatrix} \tag{4}$$

where $\boldsymbol{T}_n(t^{u,v})$, $u, v = 1, \ldots, c$ are the $n \times n$ the Toeplitz matrices

$$\begin{bmatrix} t_0^{u,v} & t_{-1}^{u,v} & \cdots & t_{-(n-1)}^{u,v} \\ t_1^{u,v} & \ddots & \ddots & t_{-(n-2)}^{u,v} \\ \vdots & \ddots & & \vdots \\ t_{n-1}^{u,v} & t_{n-2}^{u,v} & \cdots & t_0^{u,v} \end{bmatrix}.$$

We suppose $\{t_k^{u,v}\}$, $u, v = 1, \ldots, c$ to be a finite set of absolutely summable infinite sequences.[2] So, the Fourier transform

$$t^{u,v}(w) \hat{=} \sum_k t_k^{u,v} e^{-ikw}$$

can be associated with each sequence. Using the vec-permutation matrix $\boldsymbol{K}_{r,s}$ [8], we have

$$\mathcal{T}_n'(\{t^{u,v}\} = \boldsymbol{K}_{n,c} \mathcal{T}_n(\{t^{u,v}\} \boldsymbol{K}_{c,n} = \boldsymbol{K}_{n,c} \mathcal{T}_n(\{t^{u,v}\} \boldsymbol{K}_{n,c}^{-1}$$

so that the matrices $\mathcal{T}_n'(\{t^{u,v}\})$ and $\mathcal{T}_n(\{t^{u,v}\})$ are similar and hence, equivalent from an eigenvalue point of view. However, the formulation in (4) is preferred as it allows one to handle Toeplitz blocks for which results recalled in Section II can be used. Notice that $\mathcal{T}_n(\{t^{u,v}\})$ is Hermitian if and only if

$$\boldsymbol{T}_n(t^{v,u}) = \boldsymbol{T}_n^H(t^{u,v}), \qquad u, v = 1, \ldots, c$$

or equivalently

$$t^{v,u}(w) = (t^{u,v}(w))^*, \qquad u, v = 1, \ldots, c.$$

B. Asymptotic Distribution of Eigenvalues

Lemma 2 extends straightforwardly to the block Toeplitz matrices in the following.

Lemma 3: For all absolutely summable sequences

$$\{t_k^{u,v}\}_{k=\ldots,-1,0,1,\ldots}$$

there exists a sequence of matrices $\{\mathcal{C}_n(\{t^{u,v}\})\}$ asymptotically equivalent to $\{\mathcal{T}_n(\{t^{u,v}\})\}$ and given by

$$\mathcal{C}_n(\{t^{u,v}\} = \mathcal{U}_n^H \mathcal{D}_n(\{t^{u,v}\}) \mathcal{U}_n$$

where $\mathcal{U}_n$ is an $nc \times nc$ unitary matrix independent of $\mathcal{T}_n(\{t^{u,v}\})$ and where $\mathcal{D}_n(\{t^{u,v}\})$ is the following matrix:

$$\mathcal{D}_n(\{t^{u,v}\}) \hat{=} \begin{bmatrix} \boldsymbol{D}_n(t^{1,1}) & \boldsymbol{D}_n(t^{1,2}) & \cdots & \boldsymbol{D}_n(t^{1,c}) \\ \boldsymbol{D}_n(t^{2,1}) & \boldsymbol{D}_n(t^{2,2}) & \cdots & \boldsymbol{D}_n(t^{2,c}) \\ \vdots & & & \vdots \\ \boldsymbol{D}_n(t^{c,1}) & \boldsymbol{D}_n(t^{c,2}) & \cdots & \boldsymbol{D}_n(t^{c,c}) \end{bmatrix} \tag{5}$$

where $\boldsymbol{D}_n(t^{u,v})$ is a diagonal matrix defined as in Lemma 2.

Notice that $\mathcal{C}_n(\{t^{u,v}\})$ is no longer a circulant matrix, nor is $\mathcal{D}_n(\{t^{u,v}\})$ diagonal.

We next prove a result on the asymptotic eigenvalue moments of block Toeplitz matrices.

Lemma 4: For all integers $s \geq 1$

$$\begin{aligned} &\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{cn} \lambda_k^s(\mathcal{T}_n(\{t^{u,v}\})) \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \sum_{1 \leq k_1,\ldots,k_s \leq c} t^{k_1,k_2}(w) t^{k_2,k_3}(w) \ldots t^{k_s,k_1}(w)\, dw \end{aligned} \tag{6}$$

Proof: Because $\mathcal{T}_n(\{t^{u,v}\})$ and $\mathcal{C}_n(\{t^{u,v}\})$ are asymptotically equivalent, thanks to Lemma 1, we have

$$\lim_{n\to\infty} \frac{1}{cn} \sum_{k=1}^{cn} \lambda_k^s(\mathcal{T}_n(\{t^{u,v}\})) = \lim_{n\to\infty} \frac{1}{cn} \sum_{k=1}^{cn} \lambda_k^s(\mathcal{C}_n(\{t^{u,v}\})).$$

[2]The Szegö theorem [1], as well as the extensions in [4], [3], [10], were proved under weaker hypotheses on the entries of the Toeplitz matrices. The associated sequences are there supposed to be only square summable. In this case, the Fourier transform is defined differently and the Szegö theorem and its extension are more complicated to prove.

Because $\lambda_k^s(\mathcal{C}_n(\{t^{u,v}\}))$ are all the eigenvalues of $\mathcal{C}_n^s(\{t^{u,v}\})$, the preceding summation equals the trace of $\mathcal{C}_n^s(\{t^{u,v}\})$, i.e., that of $\mathcal{D}_n^s(\{t^{u,v}\})$. This can be easily proven to be equal to

$$\sum_{1\le k_1,\ldots,k_s\le c}\sum_{k=0}^{n-1} t^{k_1,k_2}\left(\frac{2\pi k}{n}\right)t^{k_2,k_3}\left(\frac{2\pi k}{n}\right)\cdots t^{k_s,k_1}\left(\frac{2\pi k}{n}\right)$$

and so

$$\begin{aligned}&\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{cn}\lambda_k^s(\mathcal{T}_n(\{t^{u,v}\}))\\ &=\lim_{n\to\infty}\frac{1}{n}\sum_{1\le k_1,\ldots,k_s\le c}\sum_{k=0}^{n-1} t^{k_1,k_2}\left(\frac{2\pi k}{n}\right)\\ &\quad\cdot t^{k_2,k_3}\left(\frac{2\pi k}{n}\right)\cdots t^{k_s,k_1}\left(\frac{2\pi k}{n}\right)\\ &=\sum_{1\le k_1,\ldots,k_s\le c}\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1} t^{k_1,k_2}\left(\frac{2\pi k}{n}\right)\\ &\quad\cdot t^{k_2,k_3}\left(\frac{2\pi k}{n}\right)\cdots t^{k_s,k_1}\left(\frac{2\pi k}{n}\right).\end{aligned}$$

Thanks to the definition of the Riemann integral where the continuity of the 2π-periodic Fourier transform guarantees the existence, the proof is complete. □

If we let

$$\boldsymbol{T}(w)\hat{=}\begin{bmatrix} t^{1,1}(w) & \cdots & t^{1,c}(w)\\ \vdots & & \vdots\\ t^{c,1}(w) & \cdots & t^{c,c}(w)\end{bmatrix}$$

(Hermitian for all w if $\mathcal{T}_n(\{t^{u,v}\})$ is Hermitian), the summation in the right-hand side of (6) is nothing but the trace of $(\boldsymbol{T}(w))^s$ whose eigenvalues are those of $\boldsymbol{T}(w)$ to the power of s. Consequently, (6) is equivalent to

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{cn}\lambda_k^s(\mathcal{T}_n(\{t^{u,v}\}))=\frac{1}{2\pi}\int_{-\pi}^{\pi}\sum_{u=1}^{c}\lambda_u^s(\boldsymbol{T}(w))\,dw.$$

Hence, for any polynomial P, we have

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{cn}P(\lambda_k(\mathcal{T}_n(\{t^{u,v}\})))=\frac{1}{2\pi}\int_{-\pi}^{\pi}\sum_{u=1}^{c}P[\lambda_u(\boldsymbol{T}(w))]\,dw.$$

Invoking the Stone–Weierstrass approximation theorem (recalled in [5]), when $\boldsymbol{T}(w)$ is Hermitian for all w, this relation extends to all functions F continuous on $[\min_w \lambda_c(\boldsymbol{T}(w)), \max_w \lambda_1(\boldsymbol{T}(w))]$. Thus, the following result extends Szegö's theorem to block Toeplitz matrices.

Theorem 3: Assume that $\mathcal{T}_n(\{t^{u,v}\})$ is Hermitian; then for all continuous functions F

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{cn}F(\lambda_k(\mathcal{T}_n(\{t^{u,v}\})))=\frac{1}{2\pi}\int_{-\pi}^{\pi}\sum_{u=1}^{c}F[\lambda_u(\boldsymbol{T}(w))]\,dw.$$

Added to the fact that, for all n, the eigenvalues of $\mathcal{T}_n(\{t^{u,v}\})$ lie in $[\min_w \lambda_c(\boldsymbol{T}(w)), \max_w \lambda_1(\boldsymbol{T}(w))]$ [4, Theorem 3.1], Theorem 3 implies that (see [4], [11]) for any integer l, the lowest (resp., largest) l eigenvalues are convergent in n and

$$\lim_{n\to\infty}\lambda_{cn-l+1}(\mathcal{T}_n(\{t^{u,v}\}))=\min_w \lambda_c(\boldsymbol{T}(w)) \quad (7)$$

$$\lim_{n\to\infty}\lambda_{l}(\mathcal{T}_n(\{t^{u,v}\}))=\max_w \lambda_1(\boldsymbol{T}(w)). \quad (8)$$

C. A Class of Block Toeplitz Matrices

We investigate the following special case of block Toeplitz matrices.

Hypothesis H1: $\boldsymbol{T}(w)$ has rank 1, for all w.

This is equivalent to having for all w

$$\boldsymbol{T}(w)=[t^1(w),\ldots,t^c(w)]^T|\mathrm{Tr}(\boldsymbol{T}(w))|[t'^1(w),\ldots,t'^c(w)]$$

where $[t^1(w),\ldots,t^c(w)]^T$ and $[t'^1(w),\ldots,t'^c(w)]^T$ are unit-norm left and right singular vectors, respectively, associated with the unique nonzero singular value of $\boldsymbol{T}(w)$. As this singular value is of multiplicity one, it is, as well as the associated singular vectors, a continuous function of $\boldsymbol{T}(w)$ [12, p. 14, Theorem 1.2.8][3] which, in turn, is a continuous function of w by construction. Hence, by redefining $t^u(w)$ and $t'^u(w)$ to be, respectively,

$$\sqrt{|\mathrm{Tr}(\boldsymbol{T}(w))|}\,t^u(w) \quad\text{and}\quad \sqrt{|\mathrm{Tr}(\boldsymbol{T}(w))|}\,t'^u(w).$$

H1 is equivalent to the following: there exist continuous functions $t^u(w)$ and $t'^u(w)$, with

$$[t^1(w),\ldots,t^c(w)]^T\neq\mathbf{0}$$

and

$$[t'^1(w),\ldots,t'^c(w)]^T\neq\mathbf{0}$$

for all w, such that $t^{u,v}(w)=t^u(w)t'^v(w)$, for all $u, v=1,\ldots,c$.

As $t^u(w)$ and $t'^u(w)$ are continuous, the infinite sequences $\{t_k^u\}_{k=\ldots,-1,0,1,\ldots}$ and $\{t'^u_k\}_{k=\ldots,-1,0,1,\ldots}$, obtained as the inverse Fourier transforms of $t^u(w)$ and $t'^u(w)$, respectively, $u=1,\ldots,c$, are square summable and not identically zero. *H1* is equivalent to having

$$t_i^{u,v}=t_i^u * t'^v_i\,\hat{=}\sum_k t_k^u t'^v_{i-k}, \qquad i=\ldots,-1,0,1,\ldots$$

i.e.,

$$t_i^{u,v}=[\ldots,t_{-1}^u,t_0^u,t_1^u,\ldots][\ldots,t'^v_{i+1},t'^v_i,t'^v_{i-1},\ldots]^T.$$

Consequently, the hypothesis *H1* is equivalent to

$$\mathcal{T}_n(\{t^{u,v}\})=\begin{bmatrix}\boldsymbol{T}_{(n)}(t^1)\\ \vdots\\ \boldsymbol{T}_{(n)}(t^c)\end{bmatrix}\begin{bmatrix}\boldsymbol{T}_{(n)}(t'^1)\\ \vdots\\ \boldsymbol{T}_{(n)}(t'^c)\end{bmatrix}^T \quad (9)$$

where $\boldsymbol{T}_{(n)}(t^u)$ (resp., $\boldsymbol{T}_{(n)}(t'^u)$), $u=1,\ldots,c$, denotes the n rows Toeplitz matrix of first row $[\ldots,t_{-1}^u,t_0^u,t_1^u,\ldots]$ (resp., $[\ldots,t'^u_{-1},t'^u_0,t'^u_1,\ldots]$). Furthermore, if $\mathcal{T}_n(\{t^{u,v}\})$ is Hermitian and positive semidefinite, *H1* is fulfilled iff $t'^u_k=(t^u_{-k})^*$, $u=1,\ldots,c$ or equivalently $\boldsymbol{T}_{(n)}(t'^u)=\boldsymbol{T}^*_{(n)}(t^u)$, i.e.,

$$\mathcal{T}_n(\{t^{u,v}\})=\begin{bmatrix}\boldsymbol{T}_{(n)}(t^1)\\ \vdots\\ \boldsymbol{T}_{(n)}(t^c)\end{bmatrix}\begin{bmatrix}\boldsymbol{T}_{(n)}(t^1)\\ \vdots\\ \boldsymbol{T}_{(n)}(t^c)\end{bmatrix}^H. \quad (10)$$

This preceding condition is frequently encountered in signal processing applications because (10) represents the covariance matrix of a c-variate stationary process obtained by filtering a white scalar stationary process. However, we note that this factorization and thus *H1* is not satisfied for covariance matrices of more general c-variate stationary processes. In the same way, (9) represents the cross-covariance matrix of two c-variate stationary processes obtained by filtering

[3]The uniqueness of the left singular vector is guaranteed by limiting the domain of the associated Givens parameterization. So, the application of [12, p. 14, Theorem 1.2.8] implies the continuity of the individual components of $[t^1(w),\ldots,t^c(w)]^T$. The same holds for $[t'^1(w),\ldots,t'^c(w)]^T$.

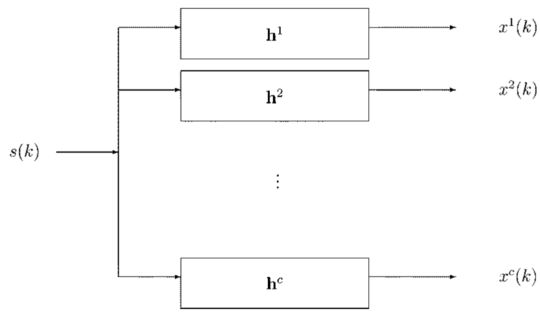


Fig. 1. Single-input multiple-output (SIMO) channel.

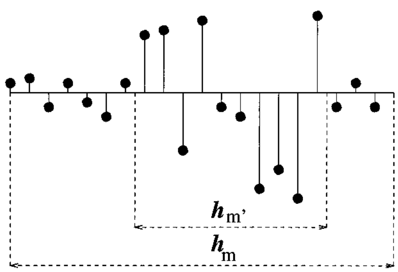


Fig. 2. Channel response with small heading and trailing terms.

a white complex-valued scalar stationary process. With the hypothesis *H1*, the following theorem is straightforwardly proved.

Theorem 4: Assume that $\mathcal{T}_n(\{t^{u,v}\})$ is Hermitian and fulfills *H1*, then for all continuous functions F

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{cn} F(\lambda_k(\mathcal{T}_n(\{t^{u,v}\}))) = (c-1)F(0) + \frac{1}{2\pi}\int_{-\pi}^{\pi} F\left(\sum_{u=1}^{c} t^{u,u}(w)\right) dw.$$

IV. Application to SIMO Channel Identification

A. Results for the SIMO Channel Filtering Matrix

An m-order SIMO channel, as depicted in Fig. 1, is a set of c filters $\boldsymbol{h}^i \hat{=} [h_0^i, \ldots, h_m^i]^T$, $i = 1, \ldots, c$, driven by a common scalar input $s(k)$, related to the c-dimensional vector output $\boldsymbol{x}(k)$ by

$$\boldsymbol{x}(k) \hat{=} [x^1(k), \ldots, x^c(k)]^T = \boldsymbol{G}(\boldsymbol{h})\boldsymbol{s}_{m+1}(k)$$

with $\boldsymbol{s}_{m+1}(k) \hat{=} [s(k), \ldots, s(k-m)]^T$ and $\boldsymbol{G}(\boldsymbol{h}) \hat{=} [\boldsymbol{h}(0) \ldots \boldsymbol{h}(m)]$ where $\boldsymbol{h}(k) \hat{=} [h_k^1 \ldots h_k^c]^T$. This setting corresponds to a multisensor reception or a polyphase representation of an oversampled signal, or a possibly hybrid situation. The SIMO channel order m is defined as the maximum order among those of the different filters $\boldsymbol{h}^1 \ldots \boldsymbol{h}^c$. n successive output observations are stacked, time by time, into

$$\boldsymbol{x}'_n(k) \hat{=} [\boldsymbol{x}^T(k) \ldots \boldsymbol{x}^T(k-(n-1))]^T$$

and the covariance matrix is defined as $\boldsymbol{R}'_n \hat{=} E[\boldsymbol{x}'_n(k)\boldsymbol{x}'^H_n(k)]$. If input $s(k)$ is zero-mean and white with variance σ_s^2 then $\boldsymbol{R}'_n = \sigma_s^2 \boldsymbol{G}_n(\boldsymbol{h})\boldsymbol{G}_n^H(\boldsymbol{h})$, where

$$\boldsymbol{G}_n(\boldsymbol{h}) \hat{=} \begin{bmatrix} \boldsymbol{G}(\boldsymbol{h}) & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{G}(\boldsymbol{h}) & \cdots & \boldsymbol{0} \\ & & \vdots & \\ \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{G}(\boldsymbol{h}) \end{bmatrix}$$

is the $cn \times (n+m)$ *filtering matrix* and $\boldsymbol{0}$ is the c-dimensional null vector. Alternatively, the data set can be arranged space/phase by space/phase in the vector $\boldsymbol{x}_n(k) \hat{=} \boldsymbol{K}_{c,n}\boldsymbol{x}'_n(k)$. It is straightforwardly proved that $\boldsymbol{x}_n(k) = \boldsymbol{H}_n(\boldsymbol{h})\boldsymbol{s}_{m+1}(k)$ where

$$\boldsymbol{H}_n(\boldsymbol{h}) \hat{=} \begin{bmatrix} \boldsymbol{H}_n(\boldsymbol{h}^1) \\ \vdots \\ \boldsymbol{H}_n(\boldsymbol{h}^c) \end{bmatrix}$$

and

$$\boldsymbol{H}_n(\boldsymbol{h}^u) \hat{=} \begin{bmatrix} \boldsymbol{h}^{u^T} & 0 & \cdots & 0 \\ 0 & \boldsymbol{h}^{u^T} & \cdots & 0 \\ & & \vdots & \\ 0 & \cdots & 0 & \boldsymbol{h}^{u^T} \end{bmatrix}$$

is the $n \times (n+m)$ filtering matrix associated with the uth filter, $u = 1, \ldots, c$.

$$\boldsymbol{R}_n \hat{=} E[\boldsymbol{x}_n(k)\boldsymbol{x}_n^H(k)] = \boldsymbol{K}_{c,n}\boldsymbol{R}'_n\boldsymbol{K}_{n,c} = \sigma_s^2 \boldsymbol{H}_n(\boldsymbol{h})\boldsymbol{H}_n^H(\boldsymbol{h})$$

is hence a block Toeplitz structured matrix that fulfills *H1* and can be written, with respect to notation of Section III, as $\boldsymbol{R}_n = \sigma_s^2 \mathcal{T}_n(\{h^u h^{v^*}\})$. The application of theorem 4 to $\boldsymbol{R}_n$ implies that for all continuous functions F, we have

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{cn} F(\lambda_k(\boldsymbol{R}_n)) = (c-1)F(0) + \frac{1}{2\pi}\int_{-\pi}^{\pi} F\left(\sigma_s^2 \sum_{u=1}^{c} |h^u(w)|^2\right) dw. \quad (11)$$

If $n+m \le cn$, we let $\sigma_k^{(n)}$, $k = 1, \ldots, n+m$, be the kth largest singular value of $\boldsymbol{G}_n(\boldsymbol{h})$ (and of $\boldsymbol{H}_n(\boldsymbol{h})$), so that $\lambda_k(\boldsymbol{R}_n) = (\sigma_k^{(n)})^2$ for $k = 1, \ldots, n+m$ and $\lambda_k(\boldsymbol{R}_n) = 0$ for $k > n+m$. Then (11) gives

$$(c-1)F(0) + \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n+m} F\left(\sigma_k^{(n)^2}\right) = (c-1)F(0) + \frac{1}{2\pi}\int_{-\pi}^{\pi} F\left(\sigma_s^2 \sum_{u=1}^{c} |h^u(w)|^2\right) dw.$$

So, the following theorem is proved.

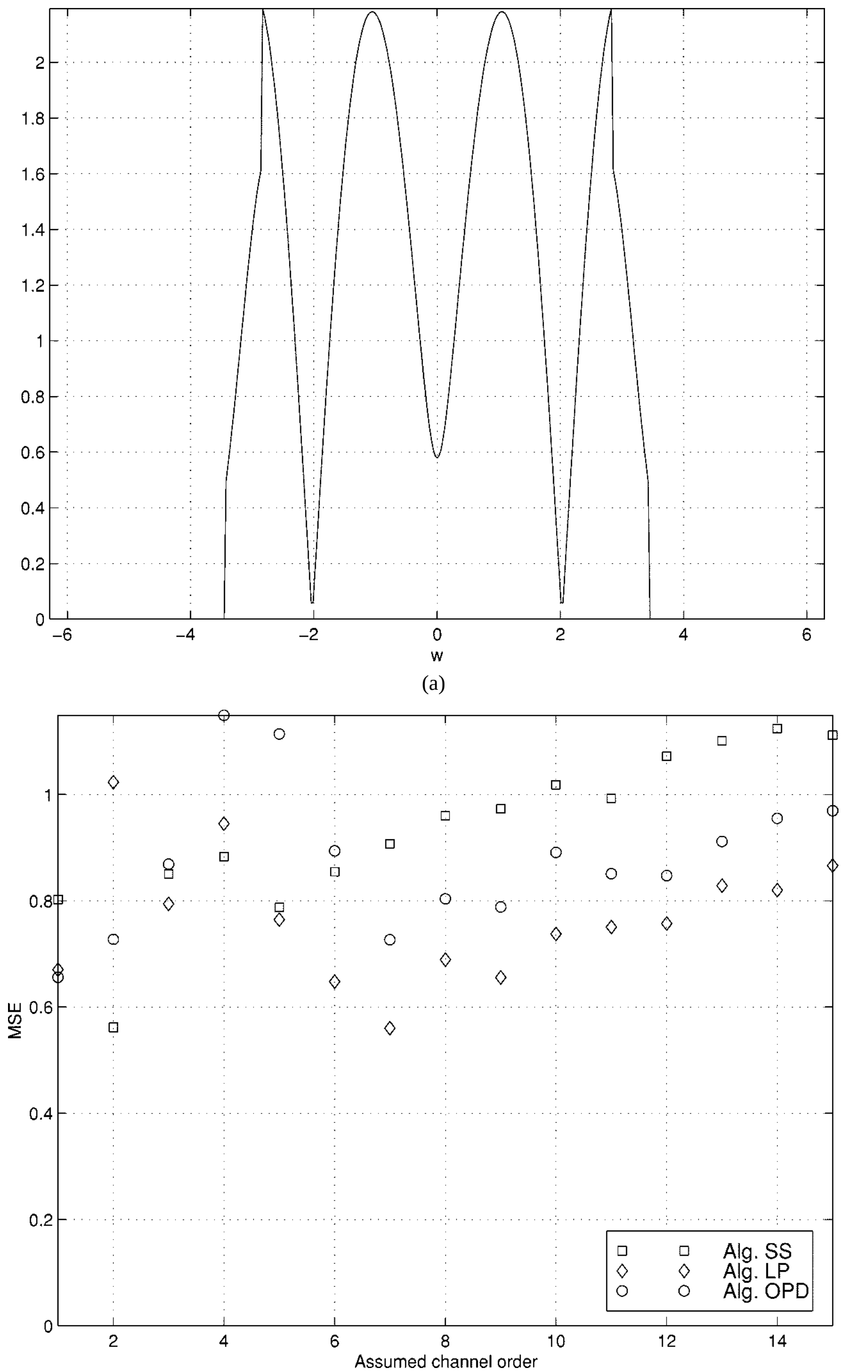
Theorem 5: For all continuous functions F

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n+m} F\left(\sigma_k^{(n)}\right) = \frac{1}{2\pi}\int_{-\pi}^{\pi} F\left(\sigma_s \sqrt{\sum_{u=1}^{c} |h^u(w)|^2}\right) dw. \quad (12)$$

It is interesting to study the asymptotic behavior of the smallest singular value $\sigma_{n+m}^{(n)}$. However, it cannot be written as $\lambda_l(\boldsymbol{R}_n)$ or $\lambda_{nc-l+1}(\boldsymbol{R}_n)$, for some fixed l and hence we can apply neither (7) nor (8). Only the following is proved.

Theorem 6: If $\min_k \sigma_k^{(n)}$ converges in n, then

$$\lim_{n\to\infty}\left(\min_k \sigma_k^{(n)}\right) \le \sigma_s \min_w \left(\sqrt{\sum_{u=1}^{c} |h^u(w)|^2}\right). \quad (13)$$



(b)

Fig. 3. Respectively, path delays $0, 1.0341T, 1.1826T, 1.3605T, 1.8780T, 1.9039T, 2.1689T, 2.2460T, 3.3337T, 3.4301T, 3.7009T$, and $4T$ and attenuations $1, 0.9868, -0.3123$, -0.3661, -0.6635, 0.6918, -0.4045, 0.5188, -0.4929, -0.7333, -0.0459, and 0.6. The upper bound in (13) equals 0.0851.

Proof: The proof is inspired by that of [4, Corollary 3.9]. We consider the real function

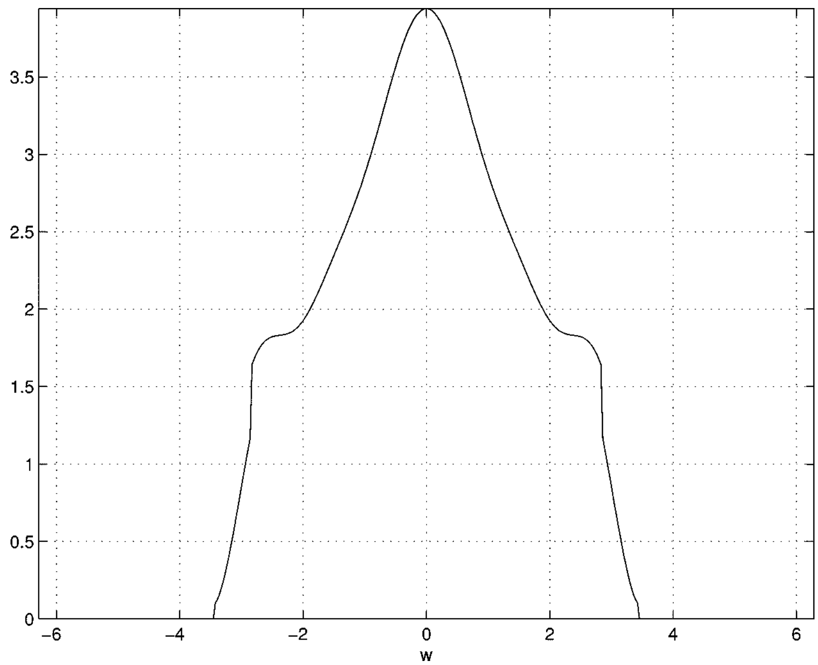
$$t(w) \hat{=} \sigma_s \sqrt{\sum_{u=1}^{c} |h^u(w)|^2}$$

and $m_t \hat{=} \min_w t(w)$. We assume $\min_k \sigma_k^{(n)}$ to be convergent to the limit $\mathcal{L}$ when $n \to \infty$. Suppose that $\mathcal{L} > m_t$. There exist a and b such that $m_t < a < b < \mathcal{L}$. We define the function $F(x) = 1$ if $x \leq a$, $F(x) = 0$ if $x \geq b$. For $x \in [a, b]$, $F(x)$ is chosen so that F is continuous and positive.

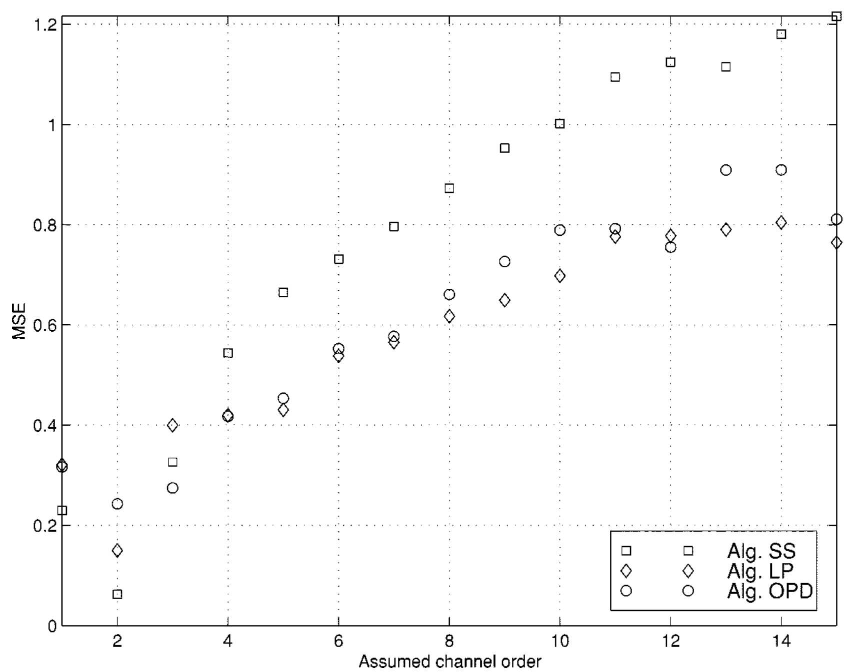
There exists an integer N such that for all $n > N$ and all k, $\sigma_k^{(n)} \geq \min_k \sigma_k^{(n)} > b$ and hence, $F(\sigma_k^{(n)}) = 0$. Consequently, the left-hand side of (12) equals 0. The right-hand side, however, equals

$$\begin{aligned}\frac{1}{2\pi}\int_{-\pi}^{\pi} F(t(w)) &= \frac{1}{2\pi}\int_{w\in[-\pi,+\pi] \text{ and } t(w)\leq b} F(t(w))\,dw \\ &> \frac{1}{2\pi}\int_{w\in[-\pi,+\pi] \text{ and } t(w)\leq a} F(t(w))\,dw \\ &= \frac{1}{2\pi}\int_{w\in[-\pi,+\pi] \text{ and } t(w)\leq a} dw > 0.\end{aligned}$$

Consequently, we must have $\mathcal{L} \leq m_t$. □



(a)



(b)

Fig. 4. Frequency response $h(w)$. Respectively, path delays $0, 0.0470T, 0.5461T, 1.0093T, 1.0858T, 2.9492T, 3.5756,$ and $4T$ and attenuations $1, 0.8381, 0.4418, 0.3689, 0.8672, 0.6775, -0.6518,$ and 0.4. The upper bound in (13) equals 1.0690.

B. Implications for Blind SIMO Channel Identification

The covariance matrix $\boldsymbol{R}_n$ contains channel phase information and is used to deduce the channel coefficients $\boldsymbol{h}(k)$, the so-called identification problem, for which a variety of second-order algorithms (among them the subspace (SS) [13], the linear prediction (LP) [14], and the outer product decomposition (OPD) [15] algorithms) has been developed. They all implicitly or explicitly need inversion of the channel output covariance matrix and hence their performance depends largely on how well conditioned the matrix is [6], [7]. Hence, its smallest nonzero eigenvalue is critical to the performance of the blind identification algorithm.

We point out that in this context (blind SIMO channel identification), the herein proved result (Theorem 3) can be considered more appropriate than that of [16]–[19]. The asymptotic results proved therein are established for block Toeplitz matrices with Toeplitz blocks (BTTB) where both the size and number of blocks tend to infinity; while in this correspondence, only the size of the blocks n tends to infinity. This is more relevant for stationary processes where n refers to the observation time and c refers to the size of the antenna array and/or the amount

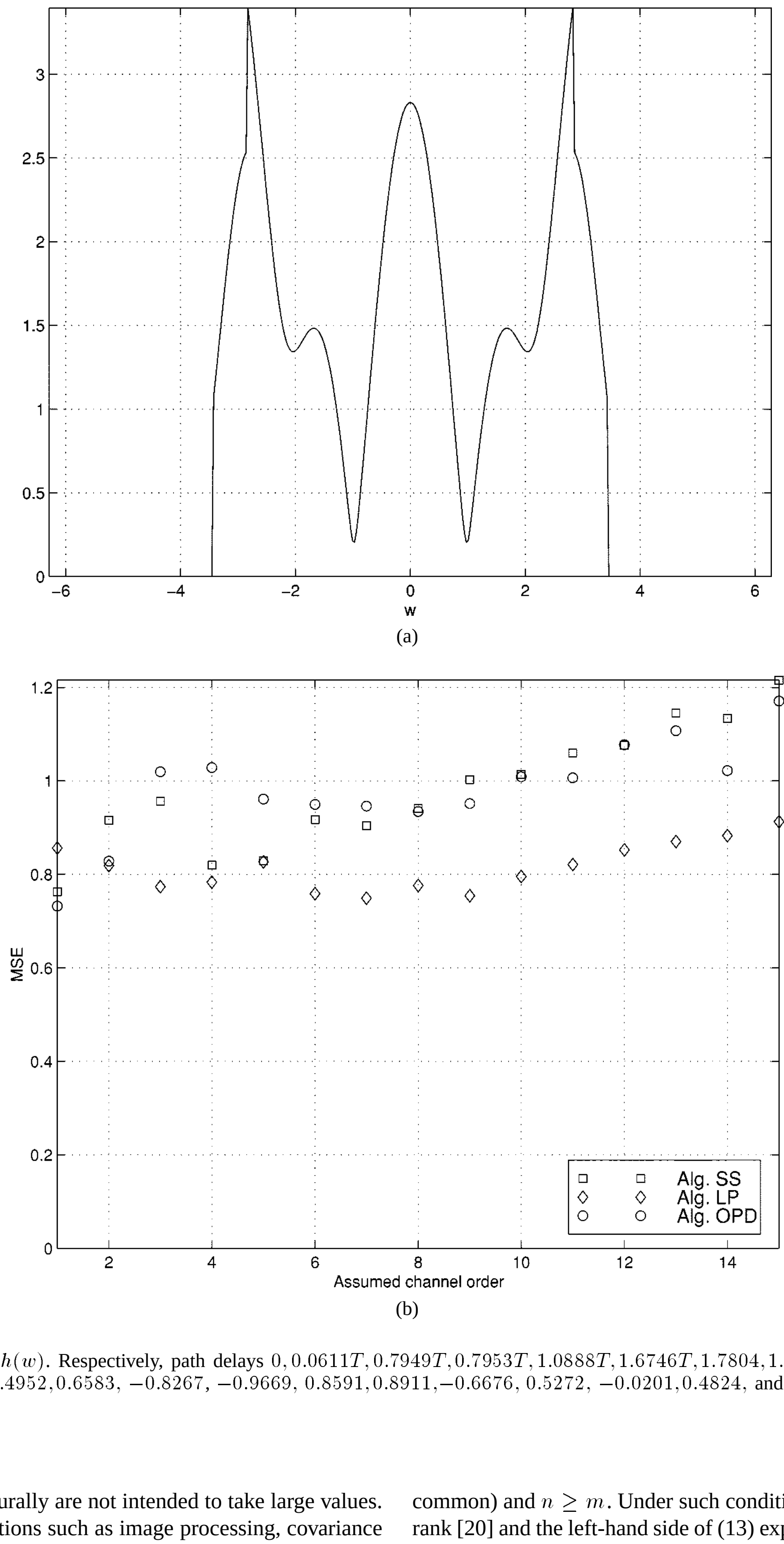


Fig. 5. Frequency response $h(w)$. Respectively, path delays $0, 0.0611T, 0.7949T, 0.7953T, 1.0888T, 1.6746T, 1.7804, 1.8640, 2.4152, 2.9871, 3.7273,$ and $4T$ and attenuations $1, 0.4952, 0.6583,\ -0.8267,\ -0.9669,\ 0.8591, 0.8911, -0.6676,\ 0.5272,\ -0.0201, 0.4824,$ and 0.4. The upper bound in (13) equals 0.3033.

of oversampling which naturally are not intended to take large values. However, for other applications such as image processing, covariance matrices of the involved two-dimensional (2-D) stationary processes are BTTB. The number and size of blocks refer to the spatial samples of the process, possibly large. In this context, results in [16]–[19] appear better adapted.

Channel blind identifiability from its second-order statistics (i.e., $\boldsymbol{R}_n$) requires the SIMO channel to be zero-coprime (i.e., the Z-transforms of the sequences $\{h_k^u,\ k = 1,\ \ldots,\ m\}$ do not have any zero in common) and $n \geq m$. Under such conditions, $\boldsymbol{G}_n(\boldsymbol{h})$ has full column rank [20] and the left-hand side of (13) expresses the square root of the asymptotic lowest nonzero eigenvalue of $\boldsymbol{R}_n$. When observed over finite time intervals and in the presence of noise, the above is insufficient and the channel needs to exhibit enough *diversity* to allow for accurate response estimation. Channel diversity has often been described as the closeness in the Z plane of the zeros of the subchannels transfer functions [21]. This definition is rather subjective and counterexamples can be found where a channel has closer zeros while its covariance matrix is better conditioned. We, therefore, suggest the left-hand side of (13) as an algorithm-independent *measure* of the channel diversity. Indeed, it approximates well the square root of the lowest nonzero eigenvalue of $\boldsymbol{R}_n$ for practical values of n.

The upper bound in (13) is better suited to assess channel blind (un)identifiability under practical observation conditions. In fact, in cases where the right-hand side in (13) is small, the channel output covariance matrix is poorly conditioned and blind algorithms are expected to fail to identify the channel if its output is observed over a limited time duration. This bound has also the advantage of giving a spectral interpretation of channel diversity.

This bound has a further interpretation in the practical case when the channel response includes small heading and/or trailing terms (Fig. 2). The whole m-order channel response $\boldsymbol{h}$ can be written as the sum of an m'-order effective response $\boldsymbol{h}_{m'}$, $m' < m$, and a perturbation vector due to the small trailing terms [22]. If we let $h^u_{m'}(w)$ be the Fourier transform associated with the subchannel $u = 1, \ldots, c$ of $\boldsymbol{h}_{m'}$, then

$$\sum_{u=1}^{c} |h^u(w)|^2 \simeq \sum_{u=1}^{c} |h^u_{m'}(w)|^2$$

i.e., the bound in (13) is approximately the same when evaluated for $\boldsymbol{h}$ or $\boldsymbol{h}_{m'}$. When this bound is weak, it implies poor diversity of the whole response as well as the effective response. In such a case, the channel will not be identifiable whatever the assumed channel order. When assumed to be $>m'$, it leads to a badly conditioned covariance matrix because of the small trailing terms. When $<m'$, the identification procedure will fail because some significant terms were ignored. When equal to m', blind identification is still not possible because of the bound (and hence channel diversity) being weak. Hence, while generally not tight as verified through computations, the upper bound in (13), when low, indicates *absolute nonidentifiability* of the channel, i.e., neither the channel nor a part of it can be identified from a finite observation set. Examples are given in the practical case of fractionally received band-limited channels.

C. Fractionally Spaced Band-Limited Channels

We now focus on fractionally spaced band-limited channels. If subchannels $h^k(w)$, $k = 1, \ldots, c$ are issued from the oversampling of a waveform $h(t)$, then

$$h^k(w) = \sum_l h(w - 2l\pi)e^{-j(w-2l\pi)\frac{k-1}{c}}$$

where

$$h(w) \hat{=} \int h(t)e^{-jwt}\, dt.$$

When $h(t)$ is band-limited (to $[-\frac{1}{T}, \frac{1}{T}]$), then

$$h^k(w) = h(w)e^{-jw\frac{k-1}{c}} + h(w - 2\pi)e^{-j(w-2\pi)\frac{k-1}{c}}$$

for $w \in [0, 2\pi]$, and it can be proved that (13) simplifies to

$$\lim_{n\to\infty} \left(\min_k \sigma_k^{(n)}\right) \le \sigma_s\sqrt{c}\, \min_w(\sqrt{(|h(w)|^2 + |h(w-2\pi)|^2)})$$

More commonly, $h(w)$ is a band-limited shaping filter response (a raised cosine waveform most often) propagating through a frequency-selective multipath channel. Because of severe selectivity, some frequency components can be significantly attenuated leading to the upper bound above to be weak. This justifies the poor performance of blind algorithms in identifying communication channels using fractional receivers, and concurs with remarks in [23].[4]

A series of simulations was conducted with a raised cosine waveform[5] with rolloff 0.3, propagating through randomly selected multipath channels with a 4 symbol period delay spread.[6] Channels for which the upper bound of (13) was weak (≤ 0.1), such as in Fig. 3, were systematically *absolutely nonidentifiable*[7] in the sense given in Section IV-B. On the contrary, however, when the upper bound of (13) was not weak, no conclusion could be made. An order with which reliable identification can be performed may exist (Fig. 4) or not (Fig. 5).

[4]The analysis there is, however, algorithm-dependent (subspace algorithm) and uses sophisticated mathematics (spheroidal wave sequences).

[5]The waveform response was truncated over 40 symbol periods.

[6]The direct path is not delayed and not attenuated while the number, delays and attenuations of the weaker and delayed paths are randomly chosen. The channel response was normalized so that $\|\boldsymbol{h}\| = 1$.

[7]Identification was tried using the subspace algorithm. The channel was observed over 300 symbol periods with a signal-to-noise ratio (SNR) of 20 dB and was $T/2$ sampled. The channel was declared nonidentified when the mean-square error exceeded 0.1.

V. Conclusion

We have given a new and simpler proof, inspired by that in [5], of Szegö's theorem extension to block Toeplitz matrices [2]. Block Toeplitz matrices are encountered in signal processing as covariance matrices that always verify some factorization properties. We exploited these properties to get a simpler form of Szegö's theorem extension and derive results about the asymptotic behavior of their lowest nonzero eigenvalue. Application to SIMO channels can help justifying cases where the channel covariance matrix is poorly conditioned resulting in poor performance of the blind identification algorithms; as is shown to be practically the case of fractionally spaced band-limited channels.

References

[1] U. Grenander and G. Szegö, *Toeplitz Forms and Their Applications*. New York: Chelsea, 1984.

[2] P. Tilli, "Asymptotic spectral distribution of Toeplitz-related matrices," in *Fast Reliable Algorithms for Matrices with Structure*, T. Kailath and A. H. Sayed, Eds. Philadelphia, PA: SIAM, 1999, pp. 153–187.

[3] ——, "Singular values and eigenvalues of non-Hermitian block Toeplitz matrices," *Linear Algebra Its Appl.*, vol. 272, pp. 59–89, 1998.

[4] M. Miranda and P. Tilli, "Asymptotic spectra of Hermitian block Toeplitz matrices and preconditioning results," *SIAM J. Matrix Anal. Applic.*, vol. 21, no. 3, pp. 867–881, 2000.

[5] R. M. Gray, "On the asymptotic eigenvalue distribution of Toeplitz matrices," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 725–730, Nov. 1972.

[6] A.-J. van der Veen, "Resolution limits of blind multi-user multi-channel identification schemes—The band-limited case," in *Proc. IEEE ICASSP*, vol. 5, 1996, pp. 2722–2727.

[7] 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," *IEEE Trans. Signal Processing*, vol. 47, pp. 1636–1645, June 1999.

[8] H. V. Henderson and S. R. Searle, "The vec-permutation matrix, the vec operator and Kronecker products: A review," *Lin. Multilin. Algebra*, vol. 9, pp. 271–288, 1981.

[9] R. M. Gray. (2000, Mar.) Toeplitz and circulant matrices: A review. Information Theory Laboratory, Stanford Univ., Stanford, CA. [Online]. Available: http://www-isl.stanford.edu/ gray/compression.html

[10] E. E. Tyrtyshnikov, "A unifying approach to some old and new theorems on distribution and clustering," *Linear Algebra Its Applic.*, vol. 232, pp. 1–43, 1996.

[11] M. Miranda and P. Tilli, *Block Toeplitz Matrices and Preconditioning*: Calcolo, 1996, vol. 33, pp. 79–86.

[12] Å. Bjørck, *Numerical Methods for Least Squares Problems*. Philadelphia, PA: SIAM, 1996.

[13] 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.

[14] K. Abed-Meraim, E. Moulines, and P. Loubaton, "Prediction error method for second-order blind identification," *IEEE Trans. Signal Processing*, vol. 45, pp. 694–705, Mar. 1997.

[15] Z. Ding, "Matrix outer product decomposition method for blind multiple channel identification," *IEEE Trans. Signal Processing*, vol. 45, pp. 3053–3061, Dec. 1997.

[16] P. A. Voois, "A theorem on the asymptotic eigenvalue distribution of Toeplitz-block-Toeplitz matrices," *IEEE Trans. Signal Processing*, vol. 44, pp. 1837–1841, July 1996.

[17] N. K. Bose and K. J. Boo, "Asymptotic eigenvalue distribution of block-Toeplitz matrices," *IEEE Trans. Inform. Theory*, vol. 44, pp. 858–861, Mar. 1998.

[18] S. Serra, "On the extreme eigenvalues of Hermitian (block) Toeplitz matrices," *Linear Algebra Its Applic.*, vol. 270, pp. 109–129, 1998.

[19] P. Tilli, "On the asymptotic spectrum of Hermitian block Toeplitz matrices with Toeplitz blocks," *Math. Comput.*, vol. 6, no. 219, pp. 1147–1159, 1997.

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

[21] I. Fijalkow, A. Touzni, and J. R. Treichler, "Fractionally spaced equalization using CMA: Robustness to channel noise and lack of disparity," *IEEE Trans. Signal Processing*, vol. 45, pp. 56–66, Jan. 1997.

[22] A. P. Liavas, P. A. Regalia, and J. P. Delmas, "Blind channel approximation: Effective channel order determination," *IEEE Trans. Signal Processing*, vol. 47, pp. 3336–3344, Dec. 1999.

[23] P. Ciblat and P. Loubaton, "Second order blind equalization: The band-limited case," in *Proc. IEEE ICASSP*, 1998.

On Preprocessing for Mismatched Classification of Gaussian Signals

Yariv Ephraim, *Fellow, IEEE*, and William J. J. Roberts, *Member, IEEE*

***Abstract*—The optimal linear preprocessor for classifying two zero-mean Gaussian discrete-time signals which have been corrupted by additive zero-mean Gaussian noise is studied. Conditions for existence of the optimal linear preprocessor that achieves the performance of the likelihood ratio test for the noisy signals are given and the preprocessor is explicitly derived.**

***Index Terms*—Hypothesis testing, mismatched classification, preprocessing.**

I. Introduction

We study a binary hypothesis testing problem in which a classifier was designed for clean signals but the observed signals are noisy. We assume zero-mean Gaussian discrete-time signals and an additive statistically independent zero-mean Gaussian discrete-time noise process. These mismatched conditions may result in significant performance degradation especially at low signal-to-noise ratio (SNR).

The optimal classifier for the clean signals compares the likelihood ratio of these signals to a threshold that depends on the optimality criterion. When the observed signals are noisy, the classifier must use the likelihood ratio for the noisy signals and a different threshold may be required. Under the Gaussian regime considered here, the likelihood ratio for the noisy signals can be obtained from the original likelihood ratio by replacing the covariance matrices of the clean signals by the covariance matrices of the noisy signals. In some applications, however, it is not desirable to modify the original classifier. Instead, the original likelihood ratio test is supplemented by a preprocessor that provides an estimate of the clean signal from the observed noisy signal. This approach may be chosen, for example, in designing smart antennas where the preprocessor and the classifier may be in separate locations.

In this correspondence, we study the hookup of a linear preprocessor with the original likelihood ratio test which provides the desired likelihood ratio test for the noisy signals. Optimal classifications in the sense of minimum probability of error and in the Neyman–Pearson sense [20] are considered. We provide conditions for the linear preprocessing approach to be optimal and give the explicit form of the optimal preprocessor. When the Gaussian signals and noise have circulant covariance matrices [7], the optimal preprocessor is proportional to the geometric mean of the Wiener filters for the two hypotheses. For independent and identically distributed (i.i.d.) signals and noise, we calculate the probability of error and compare the optimal preprocessor for classification with the optimal linear preprocessor for estimation of the signal in the minimum mean-square error (MMSE) sense. This is the Wiener estimator for the mixture covariance of the signals under the two hypotheses. It is demonstrated that the optimal linear preprocessor for classification can substantially outperform the optimal linear preprocessor for estimation especially at low SNRs. Other signal estimation preprocessors that adaptively estimate the clean signals from the noisy signals are often used but will not be considered here [17].

The optimal preprocessor for classification derived here under Gaussian assumptions may also be useful in approximating the optimal linear preprocessor when the signals and noise are not strictly Gaussian and compensation of the likelihood ratio for the clean signals is not trivial. In such situations, only second-order statistics of the signals and noise are required for designing the linear preprocessor for classification.

The preprocessing approach is motivated by the infamous Kailath–Duncan theorem [10], [11]. This theorem draws an analogy between the likelihood ratio functions for two binary detection problems. In one problem, the signal to be detected is deterministically known while in the other problem, it is a sample function of a finite-energy random process that is not necessarily Gaussian. In both cases, the signals and noise are continuous-time processes, and the noise is assumed to be a zero-mean Gaussian white process. The signal and noise are assumed statistically independent. The theorem shows that the likelihood ratio for detecting the random signal has a similar form as that of the likelihood ratio for detecting the deterministic signal. The former likelihood ratio can formally be obtained from the latter by replacing the deterministic signal with the MMSE causal estimator of the signal random process, and by interpreting the correlator integral as an Ito integral. Thus, this theorem shows that the optimal detector for the signal random process is an estimator–correlator receiver in which the signal is first estimated from the observed process and then the estimated signal is applied to the correlation detector as if it were the known deterministic signal. An excellent review of this and related results can be found in [13]. Weaker conditions for the theorem are also given in [13].

The Kailath–Duncan theorem deals with a detection problem that is different from the mismatched classification problem we study here. Nevertheless, it is often cited as the rationale for replacing unavailable clean signals by their estimates in detection and classifications problems. Furthermore, the Kailath–Duncan theorem applies to continuous-time signals only. No analogous theorem for discrete-time sig-

Manuscript received July 1, 1999; revised September 7, 2000.

Y. Ephraim is with the Department of Electrical and Computer Engineering, George Mason University, Fairfax, VA 22030 USA (e-mail: yephraim@gmu.edu).

W. J. J. Roberts was with the School of Information Technology and Engineering, George Mason University, Fairfax, VA 22030 USA. He is now with the Information Technology Division, Defense Science Technology Organisation, Salisbury, SA 5108, Australia (e-mail: william.roberts@dsto.defence.gov.au).

Communicated by S. R. Kulkarni, Associate Editor for Nonparametric Estimation, Classification, and Neural Networks.

Publisher Item Identifier S 0018-9448(01)01347-5.