Asymptotically equivalent sequences of matrices and Hermitian block Toeplitz matrices with continuous symbols:
Applications to MIMO systems

Jesús Gutiérrez-Gutiérrez and Pedro M. Crespo, Senior Member, IEEE

Abstract

For the engineering community, Gray’s tutorial monograph on Toeplitz and circulant matrices has been, and remains, the best elementary introduction to the Szegő theory on large Toeplitz matrices. In the present paper the most important results of the cited monograph are generalized to block Toeplitz matrices by maintaining the same mathematical tools used by Gray, that is, by using asymptotically equivalent sequences of matrices. As applications of these results, the geometric MMSE for both an infinite-length multivariate linear predictor and an infinite-length DFE for MIMO channels, are obtained as a limit of the corresponding finite-length cases. Similarly, a short derivation of the well known capacity of a time-invariant MIMO Gaussian channel with ISI and fixed input covariance matrix is also presented.

Index Terms

Block circulant matrices; block Toeplitz matrices; channel capacity; Frobenius norm; functions of matrices; Hermitian matrices; MIMO MMSE-DFE; MMSE linear predictor; multivariate stationary processes; spectral norm.

I. INTRODUCTION

The most famous theorem on large Toeplitz matrices is the Szegő theorem [1]. This result deals with the asymptotic behavior of the eigenvalues of an $n \times n$ Hermitian Toeplitz matrix $T_n = (t_{j-k})_{j,k=1}^n$, where the complex numbers $t_k$ with $k \in \mathbb{Z}$, are the Fourier coefficients of a bounded function called the symbol of the sequence $\{T_n\}$. Unfortunately, the level of mathematical sophistication for understanding [1] is often beyond the level provided

This work was partially supported by the Spanish Ministry of Education and Science, by the European Social Fund and by the European Regional Development Fund through the Torres-Quevedo program and the MultiMIMO project (no. TEC2007-68020-C04-03).

Both authors are with CEIT and Tecnun (University of Navarra), Manuel de Lardizábal 15, 20018, San Sebastián, Spain. Tel.: +34 943212800; fax: +34 943213076. Email addresses: jgutierrez@ceit.es (J. Gutiérrez-Gutiérrez), pcrespo@ceit.es (P. M. Crespo)

September 1, 2008 DRAFT
by a typical engineering background. For this reason, Gray simplified the Szegő theory in [2], by assuming more restrictive conditions, namely, Toeplitz matrices with symbols in the Wiener algebra (i.e., $\sum_{k \in \mathbb{Z}} |t_k| < \infty$). Since his approach is purely matrix-analytic (mainly based on his concept of asymptotically equivalent sequences of matrices), he succeeded in conveying the main ideas of the Szegő theory to a wider audience.

In the paper of Gazzah, Regalia and Delmas [3], the Szegő theorem for Toeplitz matrices with symbols in the Wiener algebra is generalized to block Toeplitz matrices by using asymptotically equivalent sequences of matrices (i.e., Gray’s tools) and the concept of vec-permutation matrix.

In the present paper we go a step further. By using again the concept of asymptotically equivalent sequences of matrices, we extend the most important results (the Szegő theorem among others) of the Gray tutorial [2] to block Toeplitz matrices with continuous symbols (i.e., the $t_k$ with $k \in \mathbb{Z}$, are given by the Fourier coefficients of a continuous $N \times N$ matrix function), which is a less restrictive condition than being in the Wiener algebra.

In [4], Bose and Boo also derived asymptotic results of block Toeplitz matrices based on Gray’s tools. However, they look at a completely different problem to the one considered in our paper. Here, $T_n$ is an $n \times n$ block Toeplitz (BT) matrix with $N \times N$ blocks, and we look, among other things, at the asymptotic behavior of the eigenvalues of $T_n$ as $n$ goes to infinity while $N$ remains fixed (specifically, we study the Szegő theorem for continuous matrix-valued symbols). On the contrary, in [4], $T_n$ is an $n \times n$ block Toeplitz matrix with Toeplitz blocks (BTTB) of order $N \times N$, and they look at the asymptotic behavior of the eigenvalues of $T_n$ when both $n$ and $N$ go to infinity (specifically, they study the Szegő theorem for certain 2-variate symbols). It should be pointed out that their derivation can not be particularized to the case where $N$ is kept fixed.

Finally, it should be mentioned that, based on matrix analysis but without using the Gray concept of asymptotically equivalent sequences of matrices, Tilli [5] and Serra-Capizzano [6] derive results of Szegő type for Lebesgue integrable multivariate matrix-valued symbols, that are stronger than the results given by Bose and Boo [4] and the ones presented here. Such generality is based on introducing another concept of approximation of matrix sequences, namely the notion of approximating class of sequences (a.c.s.), which is more general than that of asymptotically equivalent sequences. In the Appendix we report some analysis based on this new approach, since it has the potential of becoming a further tool available in the engineering community.

Therefore, the main contribution of our paper is to present a plain generalization to block Toeplitz matrices of Gray’s famous tutorial [2] by maintaining the Gray tools. The theory on block Toeplitz matrices is in general useful in signal processing, communications and information theory, since this kind of matrices arise, for instance, as covariance matrices of multivariate stationary processes or as matrix representations of linear time-invariant discrete-time multivariate filters. As applications of this theory to MIMO systems, the geometric MMSE for both an infinite-length multivariate linear predictor and an infinite-length DFE for MIMO channels, are obtained as a limit of the corresponding finite-length cases. Similarly, a short derivation of the well known capacity of a time-invariant MIMO Gaussian channel with ISI and fixed input covariance matrix is also presented.

The remainder of the paper is organized as follows. The next section states preliminary mathematical considerations. Section III derives a new result on matrix sequences that will be used in Section IV. Section IV is the
main part of the paper and derives several fundamental theorems on large Hermitian block Toeplitz matrices. We conclude with Section V where the furnished theoretical results are applied to three problems in the area of signal processing and communications.

II. PRELIMINARY MATHEMATICAL DEFINITIONS AND RESULTS

A. Asymptotically equivalent sequences of matrices

If $A$ is an $n \times n$ diagonalizable matrix then we will indicate by $\lambda_k(A)$, $1 \leq k \leq n$, all the eigenvalues of $A$ counted with their multiplicities. The following well known matrix norms (see e.g. [7] or [8]) will be used:

Definition 1: The Frobenius norm $\|A\|_F$ and the spectral norm $\|A\|_2$ of an $n \times n$ matrix $A = (a_{j,k})_{j,k=1}^n$ are defined as

$$\|A\|_F = \sqrt{\text{tr}(A^*A)} = \left(\sum_{j=1}^n \sum_{k=1}^n |a_{j,k}|^2\right)^{\frac{1}{2}}$$

$$\|A\|_2 = \max_{x \neq 0} \left(\frac{x^*A^*Ax}{x^*x}\right)^{\frac{1}{2}} = \left(\max_{1 \leq k \leq n} \lambda_k(A^*A)\right)^{\frac{1}{2}}$$

where $\text{tr}$ denotes trace, $*$ denotes conjugate transpose and $x$ is a column vector.

The Frobenius norm and the spectral norm are special instances of unitarily invariant norms (see e.g. [7] or [8]). We recall that a norm $\| \cdot \|$ is unitarily invariant if $\|UAV\| = \|A\|$ for every $n \times n$ matrix $A$ and for every $n \times n$ unitary matrices $U$ and $V$.

The next definition is due to Gray (see [2] or [9]).

Definition 2: Let $A_n$ and $B_n$ be $n \times n$ matrices for all $n \in \mathbb{N}$. We say that the sequences $\{A_n\}$ and $\{B_n\}$ are asymptotically equivalent, and write $\{A_n\} \sim \{B_n\}$, if

$$\exists M \geq 0; \quad \|A_n\|_2, \|B_n\|_2 \leq M \quad \forall n \in \mathbb{N}$$  \hspace{1cm} (1)

and

$$\lim_{n \to \infty} \frac{\|A_n - B_n\|_F}{\sqrt{n}} = 0.$$  \hspace{1cm} (2)

In the following lemma the basic properties of this kind of sequences are reviewed, most of them are directly from Gray’s monograph [2].

Lemma 1: 1) If $\{A_n\} \sim \{B_n\}$ then $\{B_n\} \sim \{A_n\}$.

2) If $\{A_n\} \sim \{B_n\}$ and $\{B_n\} \sim \{C_n\}$, then $\{A_n\} \sim \{C_n\}$.

3) If $\{A_n\} \sim \{B_n\}$ then $\{\alpha A_n\} \sim \{\alpha B_n\}$ for every $\alpha \in \mathbb{C}$.

4) If $\{A_n\} \sim \{B_n\}$ and $\{C_n\} \sim \{D_n\}$, then $\{A_n + C_n\} \sim \{B_n + D_n\}$ and $\{A_nC_n\} \sim \{B_nD_n\}$.

Notice that $\sim$ is not reflexive.

B. Functions of matrices

We first review the concept of polynomial function of a matrix.
**Definition 3:** If \( p(x) = \alpha_q x^q + \ldots + \alpha_1 x + \alpha_0 \) is a given polynomial with coefficients in \( \mathbb{C} \), then for every \( n \times n \) matrix \( A \) we define \( p(A) = \alpha_q A^q + \ldots + \alpha_1 A + \alpha_0 I_n \), where \( I_n \) denotes the \( n \times n \) identity matrix.

We now introduce the concept of function of a matrix. For a thorough treatment we refer the reader to [10] or [11]. We also recommend the text [12] for a treatment of functions of matrices, developed in a signal processing context.

**Definition 4:** We consider an \( n \times n \) diagonalizable matrix \( A = U \text{diag}(\lambda_1(A), \ldots, \lambda_n(A)) U^{-1} \). If \( g \) is a complex function on the set \( \{\lambda_1(A), \ldots, \lambda_n(A)\} \), we define the \( n \times n \) matrix \( g(A) = U \text{diag}(g(\lambda_1(A)), \ldots, g(\lambda_n(A))) U^{-1} \).

Observe that the last definition agrees with Definition 3. Furthermore, if \( g \) is a complex function on the set \( \{\lambda_1(A), \ldots, \lambda_n(A)\} \) of the eigenvalues of an \( n \times n \) diagonalizable matrix \( A \), there exists a polynomial of degree at most \( n - 1 \) such that \( g(A) = p(A) \). This is a direct consequence of the fact that the interpolation problem of determining a polynomial \( p(x) \) of degree at most \( n - 1 \) satisfying \( p(\lambda_k(A)) = g(\lambda_k(A)) \), \( 1 \leq k \leq n \), has solution (see e.g. [8]). Therefore, Definition 4 is independent of the chosen eigenvalue decomposition of the matrix \( A \).

**C. Block circulant matrices**

We now review a special type of block Toeplitz matrices.

**Definition 5:** An \( n \times n \) block circulant matrix with \( N \times N \) blocks is one having the form

\[
C = \begin{pmatrix}
C_0 & C_1 & C_2 & \cdots & C_{n-1} \\
C_{n-1} & C_0 & C_1 & \ddots & \\
C_{n-2} & C_{n-1} & C_0 & \ddots & \\
& \ddots & \ddots & \ddots & C_1 \\
C_1 & \cdots & C_{n-2} & C_{n-1} & C_0
\end{pmatrix}
\]

where \( C_k \in \mathbb{C}^{N \times N} \), \( 0 \leq k \leq n - 1 \), and \( \mathbb{C}^{N \times N} \) is the set of all \( N \times N \) complex matrices.

The next well known lemma characterizes block circulant matrices.

**Lemma 2:** \( C = \left( C_{j,k} \right)_{j,k=1}^n \), with \( C_{j,k} \in \mathbb{C}^{N \times N} \), is an \( n \times n \) block circulant matrix with \( N \times N \) blocks if and only if

\[
C = (V_n \otimes I_N) \text{diag}(M_1, \ldots, M_n)(V_n \otimes I_N)^*
\]

where \( \otimes \) is the Kronecker product, \( V_n = (v_{j,k}^{(n)})_{j,k=1}^n \) is the \( n \times n \) Fourier unitary matrix:

\[
g_{j,k}^{(n)} = \frac{1}{\sqrt{n}} e^{-\frac{2\pi i (j-1)(k-1)}{n}},
\]

and \( \text{diag}(M_1, \ldots, M_n) = (\delta_{j,k} M_j)_{j,k=1}^n \) with \( \delta_{j,k} \) being the Kronecker delta and \( M_j \in \mathbb{C}^{N \times N} \) satisfying

\[
\begin{pmatrix}
M_1 \\
M_2 \\
\vdots \\
M_n
\end{pmatrix} = \sqrt{n} (V_n \otimes I_N)^* \begin{pmatrix}
C_{1,1} \\
C_{2,1} \\
\vdots \\
C_{n,1}
\end{pmatrix}.
\]

If \( C = \left( C_{j,k} \right)_{j,k=1}^n \) is an \( n \times n \) block circulant matrix, it is customary to write \( C = \text{circ}(C_{1,1}, \ldots, C_{n,1}) \).
Finally, we review two well known results regarding the determinant and the inverse of a partitioned matrix (see e.g. [8]), which are based on block two-sided Gaussian elimination steps.

**Lemma 3:** Let $A$ be a square matrix partitioned as

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

where $A_{11}$ and $A_{22}$ are also square. If $A_{11}$ and $A_{22}$ are nonsingular then

$$\det(A) = \det(A_{11}) \det(A_{22} - A_{21}A_{11}^{-1}A_{12}) = \det(A_{11} - A_{12}A_{22}^{-1}A_{21}).$$

**Lemma 4:** Let $A$ be a square matrix partitioned as in Lemma 3. If $A$, $A_{11}$ and $A_{22}$ are nonsingular then the corresponding partitioned presentation of $A^{-1}$ is

$$A^{-1} = \begin{pmatrix} (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1} & A_{11}^{-1}A_{12}(A_{21}A_{11}^{-1}A_{12} - A_{22})^{-1} \\ (A_{21}A_{11}^{-1}A_{12} - A_{22})^{-1}A_{21}A_{11}^{-1} & (A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1} \end{pmatrix}.$$

### III. A NEW RESULT ON ASYMPTOTICALLY EQUIVALENT SEQUENCES OF MATRICES

**Theorem 1:** Consider two asymptotically equivalent sequences of Hermitian matrices $\{A_n\}$ and $\{B_n\}$. Let $[a, b]$ be a closed interval containing the eigenvalues of the matrices of these sequences. Then

$$\{g(A_n)\} \sim \{g(B_n)\}, \quad \forall g \in C[a, b],$$

where $C[a, b]$ denotes the set of all continuous complex functions on $[a, b]$.

**Proof:** Since $\{A_n\}$ and $\{B_n\}$ are sequences of Hermitian matrices, (1) implies that there is $M \geq 0$ such that $-M \leq \lambda_k(A_n), \lambda_k(B_n) \leq M$, $n \in \mathbb{N}$, $1 \leq k \leq n$. Therefore, there exists such an interval $[a, b]$, for instance $[-M, M]$. From Lemma 1 the theorem holds for any polynomial $g$.

Now, let $g$ be an arbitrary continuous function on $[a, b]$. Since the spectral norm is unitarily invariant, we deduce that

$$\|g(A_n)\|_2 = \|\text{diag}(g(\lambda_1(A_n)), \ldots, g(\lambda_n(A_n)))\|_2 = \max_{1 \leq k \leq n} |g(\lambda_k(A_n))| \leq K$$

for all $n \in \mathbb{N}$, where $K = \max_{a \leq x \leq b} |g(x)|$. Analogously for the case of $B_n$.

On the other hand, by the Stone-Weierstrass approximation theorem (see e.g. [13]), given $\epsilon > 0$ there is a polynomial $p$ such that $|p(x) - g(x)| < \epsilon$, $x \in [a, b]$. Furthermore, from the fact that $\{p(A_n)\} \sim \{p(B_n)\}$, there exists $n_0 \geq 1$ such that

$$\frac{\|p(A_n) - p(B_n)\|_F}{\sqrt{n}} < \epsilon \quad \forall n \geq n_0.$$
Consequently, by using the triangle inequality and the fact that the Frobenius norm is unitarily invariant, we have
\[
\|g(A_n) - g(B_n)\|_F \leq \sqrt{\frac{1}{n}} \sum_{k=1}^{n} |(g - p)(\lambda_k(A_n))|^2 + \sqrt{\frac{1}{n}} \sum_{k=1}^{n} |(p - g)(\lambda_k(B_n))|^2
\]
for every \( n \geq n_0 \). This ends the proof.

This result will be very useful as shown in the following section. It should be mentioned that Theorem 1 can be stated in a more general way (see Appendix).

IV. LARGE HERMITIAN BLOCK TOEPLITZ MATRICES

In the present section we will derive several fundamental theorems on \( n \times n \) Hermitian block Toeplitz matrices with \( N \times N \) blocks, for large \( n \).

Fix \( N \in \mathbb{N} \), let \( A_n \) and \( B_n \) be \( nN \times nN \) matrices for all \( n \in \mathbb{N} \). We say that the sequences \( \{A_n\} \) and \( \{B_n\} \) are asymptotically equivalent, and write \( \{A_n\} \sim \{B_n\} \), if they satisfy the expressions (1) and (2). Note that Lemma 1 and Theorem 1 still hold in this more general context.

In the sequel let \( F : \mathbb{R} \rightarrow \mathbb{C}^{N \times N} \) be an \( N \times N \) matrix function, which is continuous and \( 2\pi \)-periodic. We denote by \( \{F_k\}_{k=-\infty}^{\infty} \) the sequence of Fourier coefficients of \( F \):
\[
F_k = \frac{1}{2\pi} \int_{0}^{2\pi} F(\omega) e^{-ik\omega} d\omega,
\]
and by \( \{F_m\}_{m=0}^{\infty} \) the sequence of partial sums of the Fourier series of \( F \):
\[
F_m(\omega) = \sum_{k=-m}^{m} F_k e^{ik\omega}.
\]

Therefore from the Parseval theorem (see e.g [13]) we deduce that
\[
\lim_{m \rightarrow \infty} \frac{1}{2\pi} \int_{0}^{2\pi} \|F_m(\omega) - F(\omega)\|_F^2 d\omega = 0. \tag{3}
\]

Let us consider the \( n \times n \) block Toeplitz (BT) matrix \( T_n(F) = (F_{j-k})_{j,k=1}^{n} \), with \( N \times N \) blocks, that is,
\[
T_n(F) = \begin{pmatrix}
F_0 & F_{-1} & F_{-2} & \cdots & F_{1-n} \\
F_1 & F_0 & F_{-1} & \cdots & \vdots \\
F_2 & F_1 & F_0 & \cdots & F_{-2} \\
& \vdots & \ddots & \ddots & \ddots \\
& & & F_{1-n} & \cdots & F_2 & F_1 & F_0
\end{pmatrix}
\]
The function \( F \) is called the symbol of the \( T_n(F) \) block Toeplitz matrices. Since
\[
\|T_n(F)\|_F^2 = n\|F_0\|_F^2 + \sum_{k=1}^{n-1} (n - k) (\|F_k\|_F^2 + \|F_{-k}\|_F^2)
\]
by Parseval’s theorem we have,
\[
\frac{\|T_n(F)\|_F^2}{n} \leq \frac{1}{2\pi} \int_{0}^{2\pi} \|F(\omega)\|_F^2 d\omega.
\]  
(4)

We also associate with the function \(F\) a sequence of block circulant matrices \(\{C_n(F)\}\) given by
\[
\{(V_n \otimes I_N) \text{diag} (F(0), F(2\pi/n), \ldots, F(2\pi(n-1)/n)) (V_n \otimes I_N)^*\}.
\]

Our definition of \(C_n(F)\) is the natural generalization of the one given by Gray for the scalar case [2, p. 48]. Notice that when the \(T_n(F)\) are banded, that is, \(F = F_m\) for certain \(m\), it is deduced from Lemma 2 that
\[
C_n(F) = C_n(F_m) = \text{circ}(F_0, F_1, \ldots, F_m, 0_N, \ldots, 0_N, F_{-m}, \ldots, F_{-1})
\]
(5)
for all \(n > 2m\), where \(0_N\) denotes the \(N \times N\) zero matrix.

Suppose moreover that \(F(\omega)\) is Hermitian for all \(\omega \in \mathbb{R}\). Consequently, the \(T_n(F)\) and \(C_n(F)\) matrices will also be Hermitian. By reasoning as in [2] for the scalar case, it is not difficult to show (see e.g. [14, p. 871]) that the eigenvalues of the \(T_n(F)\) matrices are contained in the closed interval with finite endpoints
\[
\inf F = \inf_{1 \leq k \leq N, \omega \in [0, 2\pi]} \lambda_k(F(\omega))
\]
and
\[
\sup F = \sup_{1 \leq k \leq N, \omega \in [0, 2\pi]} \lambda_k(F(\omega)).
\]

Observe that the eigenvalues of the \(C_n(F)\) matrices also lie inside \([\inf F, \sup F]\), since they are given by
\[
\lambda_k \left( F \left( \frac{2\pi(j-1)}{n} \right) \right), \quad 1 \leq k \leq N, \quad 1 \leq j \leq n.
\]

We are now in a position to give an interesting example of asymptotically equivalent sequences of matrices.

**Lemma 5 (BT matrices):** Let \(F : \mathbb{R} \rightarrow \mathbb{C}^{N \times N}\) be a Hermitian matrix function, which is continuous and \(2\pi\)-periodic. Then \(\{T_n(F)\} \sim \{C_n(F)\}\).

**Proof:** For every \(n \in \mathbb{N}\), since \([\inf F, \sup F]\) contains the eigenvalues of the Hermitian matrices \(T_n(F)\) and \(C_n(F)\), we have \(\|T_n(F)\|_2, \|C_n(F)\|_2 \leq \max(|\inf F|, |\sup F|)\).

On the other hand from (3), given \(\epsilon > 0\) there exists \(m\) such that
\[
\frac{1}{2\pi} \int_{0}^{2\pi} \|F_m(\omega) - F(\omega)\|_F^2 d\omega < \epsilon^2.
\]  
(6)

Thus, by the triangle inequality
\[
\frac{\|T_n(F) - C_n(F)\|_F}{\sqrt{n}} < \frac{\|T_n(F) - T_n(F_m)\|_F}{\sqrt{n}} + \frac{\|T_n(F_m) - C_n(F_m)\|_F}{\sqrt{n}} + \frac{\|C_n(F_m) - C_n(F)\|_F}{\sqrt{n}} - \left( \frac{1}{2\pi} \int_{0}^{2\pi} \|F_m(\omega) - F(\omega)\|_F^2 d\omega \right)^{\frac{1}{2}} + \epsilon.
\]  
(7)

From (4) and (6) we obtain
\[
\frac{\|T_n(F) - T_n(F_m)\|_F}{\sqrt{n}} = \frac{\|T_n(F - F_m)\|_F}{\sqrt{n}} < \epsilon.
\]  
(8)
On account of (5) we have
\[
\frac{\|T_n(F_m) - C_n(F_m)\|_F^2}{n} = \sum_{k=1}^n k\left(\|F_k\|_F^2 + \|F_{-k}\|_F^2\right).
\]
Therefore, there exists \(n_0 \in \mathbb{N}\) such that
\[
\frac{\|T_n(F_m) - C_n(F_m)\|_F}{\sqrt{n}} < \epsilon \quad \forall n \geq n_0.
\] (9)

Furthermore, since the Frobenius norm is unitarily invariant we deduce that
\[
\lim_{n \to \infty} \frac{\|C_n(F_m) - C_n(F)\|_F^2}{n} = \lim_{n \to \infty} \frac{\|C_n(F_m - F)\|_F^2}{n} = \frac{1}{2\pi} \int_0^{2\pi} \|F_m(\omega) - F(\omega)\|_F^2 d\omega
\]
where the continuity of \(F - F_m\) on \([0, 2\pi]\) guarantees the existence of the above limit as a Riemann integral. Consequently, there is \(n_1 \in \mathbb{N}\) such that
\[
\frac{\|C_n(F_m) - C_n(F)\|_F}{\sqrt{n}} - \left(\frac{1}{2\pi} \int_0^{2\pi} \|F_m(\omega) - F(\omega)\|_F^2 d\omega\right)^{\frac{1}{2}} < \epsilon
\] (10)
for all \(n \geq n_1\). Finally, from (7), (8), (9) and (10) we conclude that
\[
\frac{\|T_n(F) - C_n(F)\|_F}{\sqrt{n}} < 4\epsilon \quad \forall n \geq \max(n_0, n_1).
\]

The product of two block Toeplitz matrices \(T_n(F)\) and \(T_n(G)\), is not in general block Toeplitz. However, the next theorem shows that this product is asymptotically equivalent to the block Toeplitz matrix of the product of the two symbols, i.e., \(T_n(FG)\).

**Theorem 2 (Product of BT matrices):** Let \(F, G : \mathbb{R} \to \mathbb{C}^{N \times N}\) be Hermitian matrix functions, which are continuous and \(2\pi\)-periodic. If \(FG\) is also Hermitian then \(\{T_n(F)T_n(G)\} \sim \{T_n(FG)\}\).

**Proof:** By combining Lemmas 1 and 5 we have \(\{T_n(F)T_n(G)\} \sim \{C_n(F)C_n(G)\} = \{C_n(FG)\} \sim \{T_n(FG)\}\).

It should be mentioned that Widom, by using advanced tools from functional analysis, proved a stronger and more general result on the product of two block Toeplitz matrices that can be found for instance in [6, Section 3.3.1], [15, p. 192] or [16].

From Theorem 2, \(\{g(T_n(F))\} \sim \{T_n(g(F))\}\) with \(g(x) = x^q, q \in \mathbb{N}\), i.e., \(\{(T_n(F))^q\} \sim \{T_n(F^q)\}\). Based on Theorem 1 and Lemma 5 we will show that this is also true for more general \(g\) functions.

**Theorem 3 (Functions of BT matrices):** If \(F\) is as in Lemma 5, then \(\{g(T_n(F))\} \sim \{T_n(g(F))\}\) \(\forall g \in C[\inf F, \sup F]\).

**Proof:** Fix \(g \in C[\inf F, \sup F]\). We begin by showing that \(g(F(\omega)) = [(g(F(\omega))]_{j,k}\) \(j,k=1\) is continuous on \(\mathbb{R}\). We will denote by \(Q(\omega)\text{diag}(\lambda_1(F(\omega)), \ldots, \lambda_N(F(\omega)))(Q(\omega))^*\) an eigenvalue decomposition of \(F(\omega)\) for all
\( \omega \in \mathbb{R} \). By the Stone-Weierstrass approximation theorem there exists a sequence of polynomials \( \{p_n\} \) that converges uniformly to \( g \) on \([\inf F, \sup F] \). Therefore, given \( \epsilon > 0 \) we can find \( n_0 \in \mathbb{N} \) such that
\[
|p_n(x) - g(x)| < \epsilon \quad x \in [\inf F, \sup F], \quad n \geq n_0.
\]
By using the Cauchy-Schwarz inequality and the fact that \( Q(\omega) = ([Q(\omega)]_{j,k})_{j,k=1}^{N} \) is a unitary matrix, we see that
\[
||p_n(F(\omega))_{j,k} - [g(F(\omega))]_{j,k}|| = \left| \sum_{l=1}^{N} (p_n - g)(\lambda_l(F(\omega))[Q(\omega)]_{j,l})[Q(\omega)]_{k,l} \right| \leq \epsilon \sum_{l=1}^{N} ||[Q(\omega)]_{j,l}||[Q(\omega)]_{k,l} \leq \epsilon.
\]
for every \( n \geq n_0 \). Thus \( \{[p_n(F(\omega))]_{j,k}\} \) converges uniformly to \([g(F(\omega))]_{j,k}\) on \( \mathbb{R} \). Since the \([p_n(F(\omega))]_{j,k}\) are continuous on \( \mathbb{R} \), we conclude that \([g(F(\omega))]_{j,k}\) is also continuous on \( \mathbb{R} \).

We are now ready to prove the assertion of the theorem. We denote the real part and the imaginary part of \( F \) by \( \text{Re}[g] \) and \( \text{Im}[g] \), respectively. Obviously, \( \text{Re}[g](F(\omega)) \) and \( \text{Im}[g](F(\omega)) \) are Hermitian for all \( \omega \in \mathbb{R} \). Therefore, from Theorem 1, Lemma 1 and Lemma 5, we deduce that
\[
\{g(T_n(F))\} \sim \{g(C_n(F))\} = \{C_n(g(F))\} = \{C_n(\text{Re}[g](F)) + iC_n(\text{Im}[g](F))\}
\]
\[
\sim \{T_n(\text{Re}[g](F)) + iT_n(\text{Im}[g](F))\} = \{T_n(g(F))\}.
\]
As in the earlier case, based on advanced tools from functional analysis, a stronger result linking \( g(T_n(F)) \) to \( T_n(g(F)) \) can be obtained (see Section 2.7 of [15]).

By particularizing \( g \) in Theorem 3 to: \( g(x) = \exp x \), \( g(x) = x^{-1} \) and \( g(x) = \sqrt{x} \), we obtain the next result:

**Theorem 4 (Exponential, inverse and square root of BT matrices):** Let \( F \) be as in Lemma 5.

1) Then \( \{\exp(T_n(F))\} \sim \{T_n(\exp F)\} \).
2) If \( 0 \not\in [\inf F, \sup F] \) then \( \{(T_n(F))^{-1}\} \sim \{T_n(F^{-1})\} \).
3) If \( \inf F \geq 0 \) then \( \{\sqrt{T_n(F)}\} \sim \{T_n(\sqrt{F})\} \).

We now derive the Szegö theorem for block Toeplitz matrix sequences with continuous symbols.

**Theorem 5 (Trace of functions of BT matrices):** Let \( F \) be as in Lemma 5. Then
\[
\lim_{n \to \infty} \frac{1}{nN} \sum_{k=1}^{N} g(\lambda_k(T_n(F))) = \frac{1}{2\pi} \int_{0}^{2\pi} \frac{1}{N} \sum_{k=1}^{N} g(\lambda_k(F(\omega)))d\omega
\]
for all \( g \in C[\inf F, \sup F] \).

**Proof:** Since \( |\text{tr}(A)| \leq \sqrt{n} ||A||_F \) for all \( n \times n \) matrix \( A \), from Theorem 3 we have
\[
\lim_{n \to \infty} \frac{1}{nN} \text{tr}(g(T_n(F))) = \lim_{n \to \infty} \frac{1}{nN} \text{tr}(T_n(g(F))) = \frac{1}{N} \text{tr} \left( \frac{1}{2\pi} \int_{0}^{2\pi} g(F(\omega))d\omega \right).
\]

The extension of Theorem 5 to Lebesgue integrable symbols \( F \in L^1 \) can be found in [5] and [6]. It should be mentioned that when \( F \) is bounded \( (F \in L^\infty) \) the assumption of bounded support for the test function \( g \) is not necessary. Moreover for \( F \in L^p \), \( 1 \leq p < \infty \), the assumption of bounded support can be replaced by an
assumption of mild growth at infinity. More precisely, it is required that the test functions $g(z)$ are continuous and $g(z)/(1+|z|^p)$ is bounded (see [17]).

The following theorem is a straightforward consequence of Theorem 5, and deals with determinants instead of traces.

**Theorem 6 (Determinant of functions of BT matrices):** Let $F$ be as in Lemma 5. Then

$$
\lim_{n \to \infty} \left( \det(g(T_n(F))) \right)^{1/nN} = \exp \left( \frac{1}{2\pi} \int_0^{2\pi} \frac{1}{N} \sum_{k=1}^N \ln(g(\lambda_k(F(\omega)))) d\omega \right)
$$

for all positive function $g \in C[\inf F, \sup F]$.

**Proof:** We have

$$
\left( \det(g(T_n(F))) \right)^{1/nN} = \left( \prod_{k=1}^{nN} g(\lambda_k(T_n(F))) \right)^{1/nN} = \exp \left( \frac{1}{nN} \sum_{k=1}^{nN} \ln(g(\lambda_k(T_n(F)))) \right).
$$

The result now follows from Theorem 5 and the continuity of the exponential function. □

By particularizing $g$ in Theorem 6 to $g(x) = x$, we obtain the next result on limits of block Toeplitz determinants.

**Theorem 7 (BT determinants):** Let $F$ be as in Lemma 5. If $\inf F > 0$ then

$$
\lim_{n \to \infty} \left( \det(T_n(F)) \right)^{1/nN} = \exp \left( \frac{1}{2\pi} \int_0^{2\pi} \frac{1}{N} \sum_{k=1}^N \ln(\lambda_k(F(\omega))) d\omega \right).
$$

As we have seen, Theorem 5 deals with the asymptotic behavior of the arithmetic average (scalar) of the entries belonging to the main diagonal of $g(T_n(F))$, i.e., $\text{tr}(g(T_n(F)))/(nN)$. The last result of the paper is stronger than Theorem 5, since it deals with the arithmetic average ($N \times N$ matrix) of the block entries belonging to an arbitrary diagonal of $g(T_n(F))$, for large $n$.

**Theorem 8:** Let $F$ be as in Lemma 5. If $r$ is a nonnegative integer then for every $g \in C[\inf F, \sup F]$,

$$
\lim_{n \to \infty} \frac{1}{n-r} \sum_{j=1}^{n-r} [g(T_n(F))]_{j,j+r} = \frac{1}{2\pi} \int_0^{2\pi} g(F(\omega)) e^{r\omega i} d\omega
$$

and

$$
\lim_{n \to \infty} \frac{1}{n-r} \sum_{j=1}^{n-r} [g(T_n(F))]_{j+r,j} = \frac{1}{2\pi} \int_0^{2\pi} g(F(\omega)) e^{-r\omega i} d\omega
$$

where $[g(T_n(F))]_{j,k} \in \mathbb{C}^{N \times N}$ is the $j,k$ block entry of $g(T_n(F))$.

**Proof:** Since in the finite dimensional vector space $\mathbb{C}^{N \times N}$ all norms are equivalent, we will prove the assertion
of the theorem for the Frobenius norm. Using the Cauchy-Schwarz inequality we have

\[
\left\| \frac{1}{n-r} \sum_{j=1}^{n-r} [g(T_n(F))]_{j,j+r} - \frac{1}{n-r} \sum_{j=1}^{n-r} [T_n(g(F))]_{j,j+r} \right\|_F
\]

\[
= \left\| \frac{1}{n-r} \sum_{j=1}^{n-r} [g(T_n(F))]_{j,j+r} - \frac{1}{n-r} \sum_{j=1}^{n-r} [T_n(g(F))]_{j,j+r} \right\|_F
\]

\[
\leq \frac{1}{n-r} \sum_{j=1}^{n-r} \|g(T_n(F)) - [T_n(g(F))]_{j,j+r}\|_F
\]

\[
\leq \frac{1}{n-r} \left( \sum_{j=1}^{n-r} \|g(T_n(F)) - [T_n(g(F))]_{j,j+r}\|_F^2 \right)^{1/2} \sqrt{n-r}
\]

\[
\leq \sqrt{\frac{n}{n-r}} \|g(T_n(F)) - T_n(g(F))\|_F
\]

for all \( n > r \). Notice that by virtue of Theorem 3, (11) now follows. The proof for (12) is similar.

Observe that all the results of this section still hold when the sequence of matrices \( \{T_n(F)\} \) is replaced by another sequence of Hermitian matrices \( \{A_n\} \) with \( \{A_n\} \sim \{T_n(F)\} \), and \([\inf F, \sup F]\) is replaced by a bigger closed interval containing the eigenvalues of the \( A_n \) matrices.

Finally, notice that most of the main results of this section (Lemma 5, Theorem 2, assertion 2 of Theorem 4, Theorem 5 and Theorem 7), were proved for the scalar case by Gray [2], assuming more restrictive conditions, namely, for \( F \) functions in the Wiener algebra rather than with continuous symbols. It should also be noted that as in [2], the asymptotic results presented in our paper do not concern individual entries of the matrices considered but rather, they describe an “average” behavior. For recent results on Toeplitz matrices regarding elementwise convergence with engineering applications, we refer the reader to [18]-[20].

V. APPLICATIONS

We now present three examples of problems in the area of signal processing and communications that can be solved by using the theoretical results of the previous sections.

A. Infinite-length multivariate MMSE linear predictor

In the first example we evaluate the geometric minimum mean square error (MMSE) for an infinite-length multivariate linear predictor, as a limit of the corresponding finite-length cases. The observed multivariate stationary process is given by \( y_k = s_k + n_k \). The process \( s_k \) is a moving average (MA) process obtained at the output of a FIR MIMO filter (i.e., a multivariate finite impulse response filter)

\[
s_k = \sum_{l=0}^{m} H_l x_{k-l}
\]

where \( x_k = (x_k^{(1)}, \ldots, x_k^{(N_i)})^\top \) denotes the input process and the filter taps \( H_0, H_1, \ldots, H_m \) are \( N_o \times N_i \) matrices. The variables \( x_k^{(j)} \), with \( k \in \mathbb{Z} \) and \( 1 \leq j \leq N_i \), are uncorrelated random variables with zero mean and variance...
The observation noise, \( n_k = (n_k^{(1)}, \ldots, n_k^{(N_o)})^\top \), is uncorrelated with the input, i.e., \( E[x_k, n_k^{(j)}] = 0_{N_x \times N_o} \) for all \( k_1, k_2 \in \mathbb{Z} \); here \( 0_{N_x \times N_o} \) is the \( N_x \times N_o \) zero matrix. Furthermore, the noise variables \( n_k^{(j)} \), with \( k \in \mathbb{Z} \) and \( 1 \leq j \leq N_o \), are uncorrelated random variables with zero mean and variance \( \sigma^2 > 0 \).

Let \( H_n \) be the \( n \times (n + m) \) block Toeplitz matrix with \( N_o \times N_i \) blocks defined by \( H_n = H_n H_n^* \). This matrix can be written as

\[
T = \begin{pmatrix}
H_0 & H_1 & \cdots & H_m & 0_{N_o \times N_i} & \cdots & 0_{N_o \times N_i} \\
0_{N_o \times N_i} & H_0 & H_1 & \cdots & H_m & 0_{N_o \times N_i} & \cdots & 0_{N_o \times N_i} \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\
0_{N_o \times N_i} & \cdots & 0_{N_o \times N_i} & H_0 & H_1 & \cdots & H_m
\end{pmatrix}.
\] (13)

We denote by \( T_n \) the \( n \times n \) banded Hermitian block Toeplitz matrix with \( N_o \times N_o \) blocks defined by \( T_n = H_n H_n^* \).

We linearly estimate \( y_k \) from the previous \( n - 1 \) vectors \( y_{k-1}, y_{k-2}, \ldots, y_{k-n+1} \). Let \( \hat{y}_k = \sum_{j=1}^{n-1} W_j y_{k-j} \) denote such an estimate. Figure 1 shows this forward linear predictor. The \( N_o \times N_o \) weight matrices \( W_j \) are chosen to minimize the mean square error (MSE) defined by the trace of the error auto-correlation matrix \( R_{ee}(n-1) = E[\mathbf{e}_k \mathbf{e}_k^\top] \) with \( \mathbf{e}_k = y_k - \hat{y}_k \). By reasoning as in [21] for the univariate case of the forward linear predictor, it is not difficult to prove that the resulting error auto-correlation matrix can be expressed as

\[
R_{ee,\text{min}}(n-1) = ([g(T_n)]_{1,1})^{-1}
\]

where \( g(x) = 1/(\sigma^2 + x) \) and \( [g(T_n)]_{1,1} \in \mathbb{C}^{N_o \times N_o} \) is the 1, 1 block entry of \( g(T_n) \). Thus, the corresponding geometric MMSE is now computed as

\[
\text{MMSE}_n = \det(([g(T_n)]_{1,1})^{-1}) = \frac{1}{\det([g(T_n)]_{1,1})}.
\]

The matrix \( T_n \) can be partitioned as

\[
T_n = \begin{pmatrix}
[T_n]_{1,1} & A_{12} \\
A_{21} & T_{n-1}
\end{pmatrix}
\]

and consequently

\[
\sigma^2 I_{N_o} + T_n = \begin{pmatrix}
\sigma^2 I_{N_o} + [T_n]_{1,1} & A_{12} \\
A_{21} & \sigma^2 I_{(n-1)N_o} + T_{n-1}
\end{pmatrix}.
\]

Combining Lemmas 3 and 4 we deduce that

\[
\det([g(T_n)]_{1,1}) = \det(([\sigma^2 I_{N_o} + T_n]^{-1})_{1,1}) = \frac{\det(\sigma^2 I_{(n-1)N_o} + T_{n-1})}{\det(\sigma^2 I_{N_o} + T_n)}.
\]

Since we are dealing with multi-dimensional error random processes either the trace or the determinant of \( R_{ee}(n-1) \) can be used as a mean square error measure. It turns out that by using either criterion, the obtained optimal system parameters and the associated error auto-correlation matrix, \( R_{ee,\text{min}}(n-1) \), are the same.
and therefore, \( \text{MMSE}_n = \frac{\det(g(T_{n-1}))}{\det(g(T_n))} \). Since \( \det(g(T_n)) > 0 \), using a well known property regarding limits of sequences of positive numbers, yields

\[
\lim_{n \to \infty} \text{MMSE}_n = \lim_{n \to \infty} \frac{\det(g(T_{n-1}))}{\det(g(T_n))} = 1
\]

Finally by Theorem 6, the geometric MMSE for an infinite-length multivariate forward linear predictor can be written as

\[
\text{MMSE} = \exp \left( \frac{1}{2\pi} \int_0^{2\pi} \ln \left( \sigma^2 + \lambda_k(H(\omega)H(\omega)^*) \right) d\omega \right)
\]

\[
= \exp \left( \frac{1}{2\pi} \int_0^{2\pi} \ln \det \left( \sigma^2 I_N + H(\omega)H(\omega)^* \right) d\omega \right).
\]

**B. Infinite-length MIMO MMSE-DFE**

In the second example, we evaluate the geometric MMSE for an infinite-length decision feedback equalizer (DFE) in multiple-input multiple-output (MIMO) channels, as a limit of the corresponding finite-length cases. We consider the communication system shown in Figure 2 where the MIMO channel has \( N_i \) inputs and \( N_o \) outputs. Its discrete channel impulse response is given by \( H = (H_0, \ldots, H_m) \), where the \( H_l \) are \( N_o \times N_i \) matrices. The channel output, \( y_k \), can be written as

\[
y_k = \sum_{l=0}^{m} H_l x_{k-l} + n_k
\]

where \( x_k = (x_k^{(1)}, \ldots, x_k^{(N_i)})^\top \) and \( n_k = (n_k^{(1)}, \ldots, n_k^{(N_o)})^\top \) denote the input and noise respectively. The input components symbols, \( x_k^{(j)} \), are discrete uncorrelated random variables with \( \mathbb{E}[x_k^{(j)}] = 0 \) and \( \mathbb{E}[|x_k^{(j)}|^2] = 1 \). The observation noise is uncorrelated with the input, and its samples, \( n_k^{(j)} \), are uncorrelated random variables with zero mean and variance \( \sigma^2 > 0 \).

The equalizer block is a standard DFE that comprises an \( n \) tapped delay line feedforward filter (FFF), and an infinite tapped delay line feedback filter (FBF). We linearly estimate the input vector \( x_{k-\triangle} \), where \( 0 \leq \triangle \leq n + m - 1 \) is the decision delay, from the channel outputs \( y_k, y_{k-1}, \ldots, y_{k-n+1} \) and from the inputs \( x_{k-\triangle-1}, x_{k-\triangle-2}, \ldots, x_{k-n-m+1} \) (assuming correct previous decisions). Thus, such an estimate \( \tilde{x}_{k-\triangle} \) has the
form
\[ \tilde{x}_{k-\triangle} = \sum_{j=0}^{n-1} W_j y_{k-j} + \sum_{l=1}^{n+m-\triangle-1} B_l \tilde{x}_{k-\triangle-l}, \]
with the \( W_j \) and \( B_l \) being \( N_i \times N_o \) matrices and \( N_i \times N_i \) matrices, respectively. These matrix taps \( W_j \) and \( B_l \) are chosen to minimize the MSE defined by the trace\(^2\) of the error auto-correlation matrix
\[ R_{ee}(n) = E[e_{k-\triangle} e_{k-\triangle}^*] \]
with \( e_{k-\triangle} = x_{k-\triangle} - \tilde{x}_{k-\triangle} \). From [22] it can be easily shown that the resulting error auto-correlation matrix can be expressed as
\[ R_{ee,\text{min}}(n) = [g(\tilde{A}_{\triangle+1})]_{\triangle+1,\triangle+1} \]
where \( \tilde{A}_{\triangle+1} \) is the matrix obtained by deleting the last \( (n+m-\triangle-1)N_i \) rows and columns of the matrix \( H_n^*H_n \), with \( H_n \) being the channel matrix (13); the function \( g \) is now
\[ g(x) = \frac{\sigma^2}{\sigma^2 + x}, \]
and \( [g(\tilde{A}_{\triangle+1})]_{\triangle+1,\triangle+1} \in \mathbb{C}^{N_i \times N_i} \) is the \( \triangle + 1, \triangle + 1 \) block entry of \( g(\tilde{A}_{\triangle+1}) \). Thus, the corresponding geometric MMSE can be computed as
\[ \text{MMSE}_n(\triangle + 1) = \det([g(\tilde{A}_{\triangle+1})]_{\triangle+1,\triangle+1}). \]

Let now \( F(\omega) \) be the \( N_i \times N_i \) matrix function \( H(\omega)^*H(\omega) \) where \( H(\omega) = \sum_{k=0}^{m} H_k e^{-k\omega} \) is the transfer function of the channel. Although the \( (n+m)N_i \times (n+m)N_i \) Hermitian matrix \( H_n^*H_n \) is in general not block Toeplitz, it is equal to the banded Hermitian block Toeplitz matrix \( T_{n+m}(F) \) except perhaps for the \( mN_i \times mN_i \) upper-left and lower-right corners. Furthermore, the entries of these corners do not depend on the order. Consequently, if we define \( A_n \) as the \( nN_i \times nN_i \) matrix obtained by deleting the last \( mN_i \) rows and columns of \( H_n^*H_n \), then \( A_n \) is equal to \( T_n(F) \) except perhaps for the \( mN_i \times mN_i \) upper-left corner and therefore, \( \{A_n\} \sim \{T_n(F)\} \). The matrix \( A_n \) can be partitioned as
\[ A_n = \begin{pmatrix} A_{n-1} & A_{12} \\ A_{21} & [A_n]_{n,n} \end{pmatrix}, \]

\(^2\)As in the first application, the trace or the determinant of \( R_{ee}(n) \) can be used as a mean square error measure. In [22] it was shown that the obtained optimal system parameters and the associated error auto-correlation matrix, \( R_{ee,\text{min}}(n) \), are the same by using either criterion.
Thus, 
\[ I_{nN_i} + \sigma^{-2} A_n = \begin{pmatrix} I_{(n-1)N_i} + \sigma^{-2} A_{n-1} & \sigma^{-2} A_{12} \\ \sigma^{-2} A_{21} & I_{N_i} + \sigma^{-2} [A_n]_{n,n} \end{pmatrix}. \]

Combining Lemmas 3 and 4 we deduce that
\[ \det((g(A_n))_{n,n}) = \frac{\det(I_{(n-1)N_i} + \sigma^{-2} A_{n-1})}{\det(I_{N_i} + \sigma^{-2} [A_n]_{n,n})} = \frac{\det(g(A_n))}{\det(g(A_{n-1}))}. \]  

(14)

Without any loss of generality, suppose that \( \Delta + 1 = \varphi(n + m) \) where \( \varphi : \mathbb{N} \rightarrow \mathbb{N} \) is a function such that \( 1 \leq \varphi(n) \leq n, \varphi(n) \rightarrow \infty, \) and \( m \leq \lim_{n \rightarrow \infty} n - \varphi(n) \) (for example, \( \varphi(n + m) = n \)). Consequently, for sufficiently large \( n \), the matrix \( A_{\Delta+1} \) is equal to \( A_{\varphi(n+m)} \) and by using (14) we have
\[ \lim_{n \rightarrow \infty} \text{MMSE}_n(\varphi(n + m)) = \lim_{n \rightarrow \infty} \frac{\det((g(A_{\varphi(n+m)}))_{n,n})}{\varphi(n+m), \varphi(n+m)} = \lim_{n \rightarrow \infty} \frac{\det(g(A_n))}{\text{det}(g(A_{n-1}))} = \lim_{n \rightarrow \infty} (\text{det}(g(A_n)))^{1/n}. \]

Since \( \{A_n\} \sim \{T_n(F)\} \), from Theorem 6 the geometric MMSE for an infinite-length MIMO DFE can be written as
\[ \text{MMSE} = \lim_{n \rightarrow \infty} \left( \frac{\det(g(T_n(F)))}{\det(g(A_n)))} \right)^{1/n} = \exp \left( -\frac{1}{2\pi} \int_0^{2\pi} \sum_{k=1}^{N_i} \ln \left( 1 + \sigma^{-2} \lambda_k(H(\omega)^* H(\omega)) \right) d\omega \right) \]
\[ = \exp \left( -\frac{1}{2\pi} \int_0^{2\pi} \ln \text{det} \left( I_{N_i} + \sigma^{-2} H(\omega)^* H(\omega) \right) d\omega \right). \]

C. Capacity of a time-invariant MIMO Gaussian channel with ISI and fixed input covariance matrix

Finally, we derive the capacity of a time-invariant MIMO Gaussian channel with impulse response \( \{H_j\}_{j=0}^m \) subject to an input constraint, namely, a flat input power spectral density. The input-output relation is given by
\[ y_k = \sum_{j=0}^m H_j x_{k-j} + n_k \]
where \( x_{k-m} \in \mathbb{C}^{N_i} \) and \( y_k \in \mathbb{C}^{N_o}, \) \( k \in \mathbb{N}, \) are the input and output of the channel, respectively. The channel noise is independent of the input, and the noise random vectors \( n_k \in \mathbb{C}^{N_o} \) are assumed to be i.i.d. (independent identically distributed) and circularly symmetric Gaussian with zero mean and covariance \( \sigma^2 I_{N_o}. \) For every \( n \in \mathbb{N}, \) the output of the channel up to time \( n \) can be written in matrix notation as \( y^{(n)} = H_n x^{(n+m)} + n^{(n)} \), where \( H_n \) is given by expression (13), \( y^{(n)} = (y_1^{(n)}, \ldots, y_{N_o}^{(n)})^\top \), \( x^{(n+m)} = (x_{1-m}^{(n)}, \ldots, x_N^{(n)})^\top \) and \( n^{(n)} = (n_1^{(n)}, \ldots, n_{N_o}^{(n)})^\top \).

Assuming we restrict the covariance of the input random vector, \( x^{(n+m)} \), to be \( K = \frac{P}{N_i} I_{(n+m)N_i} \), the maximum of the normalized mutual information \( \frac{1}{n} I(x^{(n+m)}; y^{(n)}) \) is \( \frac{1}{n} \log_2 \det \left( I_{N_i} + \frac{\text{SNR}}{N_i} H_n H_n^* \right) \) bits per channel use, and it is obtained when the input process is Gaussian, with zero mean [23]. By denoting \( \text{SNR} = \frac{P}{\sigma^2} \) the signal to noise ratio at the input of the MIMO channel, the channel capacity is now given by the limiting expression
\[ C = \lim_{n \rightarrow \infty} \frac{1}{n} \log_2 \det \left( I_{N_i} + \frac{\text{SNR}}{N_i} T_n(F) \right) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^{n N_i} \log_2 \left( 1 + \frac{\text{SNR} \lambda_k(T_n(F))}{N_i} \right) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^{n N_i} g(\lambda_k(T_n(F))) \]
where $F(\omega)$ is the $N_e \times N_e$ matrix function $H(\omega)^* H(\omega)$, with 
$H(\omega) = \sum_{k=0}^{N_e} H_k e^{-j\omega_k t}$, and $g(x) = \log_2 \left( 1 + \frac{\text{SNR}}{N_i} \right)$.

Now, by applying Theorem 5 we obtain the well known expression for the desired capacity [24]

$$C = \frac{1}{2 \pi} \int_0^{2 \pi} \sum_{k=1}^{N_e} \log_2 \left( 1 + \frac{\text{SNR}}{N_i} \lambda_k (H(\omega)^* H(\omega)) \right) d\omega = \frac{1}{2 \pi} \int_0^{2 \pi} \log_2 \det \left( I_{N_e} + \frac{\text{SNR}}{N_i} H(\omega)^* H(\omega) \right) d\omega.$$

**APPENDIX**

**OTHER CONCEPTS OF APPROXIMATION OF MATRIX SEQUENCES**

In this appendix we present the notions of asymptotically $p$-equivalent sequences of matrices and that of approximating class of sequences (a.c.s.). Both are concepts of approximation of matrix sequences that generalize Gray’s concept of asymptotically equivalent sequences. We present these two notions to show that Theorem 1 can be stated in a more general way, by proving that this theorem can be considered as a special instance of certain known results, namely [25, Theorems 2.1-2.2].

We begin with the definition of asymptotically $p$-equivalent sequences of matrices that can be obtained from the definition of asymptotically equivalent sequences of matrices by substituting the Frobenius norm for any other Schatten $p$-norm. We first review these well known norms (see e.g. [7] or [8]).

**Definition 6:** The Schatten $p$-norm $\| A \|_{S,p}$ of an $n \times n$ matrix $A$, $p \in [1, \infty)$, is the $l_p$ norm on $\mathbb{C}^n$ of the vector of the singular values of $A$, i.e.,

$$\| A \|_{S,p} = \begin{cases} \left( \sum_{k=1}^{n} \sigma_k(A)^p \right)^{1/p} & \text{if } p \in [1, \infty), \\ \max_{1 \leq k \leq n} \sigma_k(A) & \text{if } p = \infty, \end{cases}$$

where $\sigma_k(A)$, with $1 \leq k \leq n$, are the singular values of $A$ counted with their multiplicities.

Notice that $\| A \|_{S,\infty} = \| A \|_2$ (the spectral radius if $A$ is Hermitian) and $\| A \|_{S,2} = \| A \|_F$ (the $l_2$ norm on $\mathbb{C}^n$ of the vector containing the eigenvalues of $A$ when $A$ is Hermitian).

All the Schatten norms are special instances of unitarily invariant norms (see e.g. [7] or [8]).

Using these norms, the definition of asymptotically $p$-equivalent sequences of matrices is as follows.

**Definition 7:** Consider a strictly increasing sequence of natural numbers $\{d_n\}$. Let $A_n$ and $B_n$ be $d_n \times d_n$ matrices for all $n \in \mathbb{N}$. We say that the sequences $\{A_n\}$ and $\{B_n\}$ are asymptotically $p$-equivalent, $p \in [1, \infty)$, and write $\{A_n\} \sim_p \{B_n\}$, if

$$\exists M \geq 0; \quad \| A_n \|_2, \| B_n \|_2 \leq M \quad \forall n \in \mathbb{N} \quad (15)$$

and

$$\lim_{n \to \infty} \frac{\| A_n - B_n \|_{S,p}}{d_n^{1/p}} = 0. \quad (16)$$

The notion of asymptotically $p$-equivalent sequences is a concept of approximation in norm of matrix sequences. Observe that for $p = 2$ and $d_n = n$, $\{A_n\} \sim_p \{B_n\}$ is the same as $\{A_n\} \sim \{B_n\}$ (the standard notion introduced by Gray).

The second and last concept is the notion of a.c.s. due to Serra-Capizzano [26]. It deals with approximation in norm and in rank.
Definition 8: Consider a strictly increasing sequence of natural numbers \( \{d_n\} \). Let \( A_n \) be a \( d_n \times d_n \) matrix for all \( n \in \mathbb{N} \). We say that \( \{\{B_{n,m}\}_n\}_m \) is an a.c.s. for \( \{A_n\} \) if, for all sufficiently large \( m \in \mathbb{N} \), the following splitting holds:

\[
A_n = B_{n,m} + R_{n,m} + N_{n,m} \quad \forall n > n_m,
\]

with

\[
\text{rank}(R_{n,m}) \leq d_n c(m), \quad \|N_{n,m}\|_2 \leq \omega(m),
\]

where \( n_m, c(m) \) and \( \omega(m) \) depend only on \( m \) and, moreover,

\[
\lim_{m \to \infty} c(m) = \lim_{m \to \infty} \omega(m) = 0.
\]

The a.c.s notion is more general than that of asymptotically \( p \)-equivalent sequences. In fact, we show next that if \( \{B_n\} \sim_p \{A_n\} \) then \( \{\{B_n\} : m\} \) is an a.c.s for \( \{A_n\} \). Moreover, the requirement of uniform boundedness in (15) is not necessary.

Lemma 6: Consider \( p \in [1, \infty) \) and a strictly increasing sequence of natural numbers \( \{d_n\} \). Let \( A_n \) and \( B_n \) be \( d_n \times d_n \) matrices for all \( n \in \mathbb{N} \). If expression (16) holds then \( \{\{B_n\} : m\} \) is an a.c.s for \( \{A_n\} \).

Proof: Let \( V_n \text{diag}(\sigma_{1,n}, \ldots, \sigma_{d_{n,n}})W_n^* \) be a singular value decomposition of the matrix \( A_n - B_n \) with \( \sigma_{1,n} \geq \sigma_{2,n} \geq \cdots \geq \sigma_{d_{n,n}} \). We define

\[
\phi(n, m) = \min \left\{ j \in \{1, \ldots, d_n\} : \sigma_{j,n} < \frac{1}{m} \right\},
\]

\[
R_{n,m} = V_n \text{diag}(\sigma_{1,n}, \ldots, \sigma_{\phi(n,m)-1,n}, 0, \ldots, 0)W_n^*,
\]

\[
N_{n,m} = V_n \text{diag}(0, \ldots, 0, \sigma_{\phi(n,m),n}, \ldots, \sigma_{d_{n,n}})W_n^*.
\]

Obviously \( A_n = B_n + R_{n,m} + N_{n,m} \), since \( R_{n,m} + N_{n,m} = A_n - B_n \). On the one hand,

\[
\|N_{n,m}\|_2 = \sigma_{\phi(n,m),n} < \frac{1}{m} = \omega(m).
\]

On the other hand, since (16) holds there exists \( n_m \in \mathbb{N} \) such that

\[
\frac{\|A_n - B_n\|_{S,p}}{d_n^{1/p}} < \frac{1}{m^2}, \quad \forall n > n_m,
\]

and therefore

\[
\frac{\text{rank}(R_{n,m})}{d_n} = \frac{\text{rank}(\text{diag}(\sigma_{1,n}, \ldots, \sigma_{\phi(n,m)-1,n}, 0, \ldots, 0))}{d_n} = \frac{\phi(n, m) - 1}{d_n}
\]

\[
\leq \frac{1}{d_n} \sum_{k=1}^{\phi(n,m)-1} \left( \frac{\sigma_{k,n}}{\sigma_{\phi(n,m)-1,n}} \right)^p \leq \frac{1}{d_n} \sum_{k=1}^{d_n} \left( \frac{\sigma_{k,n}}{\sigma_{\phi(n,m)-1,n}} \right)^p
\]

\[
= \frac{1}{d_n \sigma_{\phi(n,m)-1,n}} \|A_n - B_n\|_{S,p}^p \leq \frac{m^p}{d_n} \|A_n - B_n\|_{S,p}^p
\]

\[
= \left( m \frac{\|A_n - B_n\|_{S,p}}{d_n^{1/p}} \right)^p < \frac{1}{m^p} = c(m)
\]

for all \( n > n_m \).
Finally, it should be mentioned that by using the previous lemma (with \( p = 2 \) and \( d_n = n \)), Theorem 1 can be obtained from [25, Theorems 2.1-2.2]. In fact, Theorem 1 can be stated in a more general way:

i) If \( \{ A_n \} \sim_p \{ B_n \} \), with some \( p \in [1, \infty) \), then \( \{ g(A_n) \} \sim_p \{ g(B_n) \} \);

ii) If \( \{ \{ B_n \} : m \} \) a.c.s for \( \{ A_n \} \), then \( \{ \{ g(B_n) \} : m \} \) a.c.s for \( \{ g(A_n) \} \),

for all continuous function \( g \) with bounded support. Theorem 2 also holds if \( \sim \) is replaced by \( \sim_p \) for every \( p \in [1, \infty) \) (see [26], [27]), even for bounded symbols. If the sequence \( \{ C_n(F) \} \) is substituted by \( \{ \hat{C}_n(F) \} \), where the matrix \( \hat{C}_n(F) \) is obtained by minimizing \( \| T_n(F) - X \|_F \) over all the block circulant matrices \( X \) (see [28, Definition 4.1]), then Lemma 5 also holds when \( \sim \) is replaced by \( \sim_p \) for all \( p \in [1, \infty) \), even for bounded symbols. In fact, it holds that (see [17], [27] and [28, Proposition 4.3])

i) \( \lim_{n \to \infty} \| T_n(F) - \hat{C}_n(F) \|_{p,\infty} = 0 \ \forall F \in L^p \),

ii) \( \| \hat{C}_n(F) \| \leq \| T_n(F) \| \) for every \( F \in L^1 \) and for every unitarily invariant norm \( \| \cdot \| \),

iii) \( \| T_n(F) \|_2 \) is uniformly bounded for all \( F \in L^\infty \).

Moreover, the matrix \( \hat{C}_n(F) \), unlike \( C_n(F) \), has a simple explicit expression (see the formulae (18) in [28]) in terms of the Fourier coefficients of \( F \).

**ACKNOWLEDGMENT**

The authors would like to thank the anonymous referees for their thorough review.

**REFERENCES**


Jesús Gutiérrez-Gutiérrez was born in Granada, Spain. He received his degree in Mathematics from the University of Granada, Spain, in 1999 and his Ph.D. degree in Electronics and Communications from the University of Navarra, Spain, in 2004. He is currently with the Electronics and Communications Department at the R&D center CEIT, San Sebastián, Spain. He is also an associate professor at the Engineering School of the University of Navarra (Tecnun). His current research interests include matrix analysis, random matrix theory, spectral theory and orthogonal polynomials applied to problems in communications.
Pedro M. Crespo (S’80-M’84-SM’91) was born in Barcelona, Spain. In 1978, he received the engineering degree in Telecommunications from Universidad Politécnica de Barcelona. He received the M.Sc. in Applied Mathematics and Ph.D. in Electrical Engineering from the University of Southern California (USC), in 1983 and 1984, respectively. From September 1984 to April 1991, he was a member of the technical staff in the Signal Processing Research group at Bell Communications Research, New Jersey, USA, where he worked in the areas of digital communication and signal processing. He actively contributed in the definition and development of the first prototypes of xDSL (Digital Subscriber Lines transceivers). From May 1991 to August 1999 he was a district manager at Telefónica Investigación y Desarrollo, Madrid, Spain. From 1999 to 2002 he was the technical director of the Spanish telecommunication operator Jazztel. At present he is the head of the Electronics and Communications Department at the R&D center CEIT, San Sebastián, Spain. He is also a full professor at the Engineering School of the University of Navarra (Tecnun).

Dr. Crespo is a Recipient of the Bell Communication Research Award of excellence. He holds seven patents in the areas of digital subscriber lines and wireless communications. His research interests include the general areas of digital communications, signal processing and information theory.