Displacement Structure and Maximum Entropy

Tiberiu Constantinescu, Ali H. Sayed, *Member, IEEE*, and Thomas Kailath, *Fellow, IEEE*

***Abstract*—The study of matrices with a displacement structure is mainly concerned with recursions for the so-called generator matrices. The recursion usually involves free parameters, which can be chosen in several ways so as to simplify the resulting algorithm. In this correspondence we present a choice for the parameters that is motivated by a maximum-entropy formulation. This choice further motivates the introduction of the so-called generalized reflection coefficients which are, in general, different from the better known Schur coefficients.**

***Index Terms*— Displacement structure, maximum entropy, reflection coefficient, Schur algorithm.**

I. INTRODUCTION

The maximum-entropy extension (or loading) problem has attracted considerable attention in the literature. The first solution by Burg [1] treated Toeplitz matrices and emphasized their parametrization in terms of the so-called reflection coefficients, also known as Schur coefficients. In this correspondence, we exploit the fact that the Toeplitz/Schur ideas can be extended to more general classes of matrices by invoking the concept of displacement structure [2], and show that a very general formulation of the maximum entropy problem is possible. In particular, we provide both global and recursive solutions to the generalized problem.

The connection between maximum-entropy extensions and structured matrices will be established in terms of the cascade or transmission-line structures, that arise naturally when the Cholesky factorizations of structured matrices are efficiently computed via a generalized Schur algorithm [2]. For a given structured matrix R, the algorithm operates recursively on its so-called generator matrix G and provides, for each step, a first-order section (or transfer function/operator). Each section is usually parametrized in terms of two free parameters: a J-unitary rotation matrix Θ_i and a complex scalar τ_i that is restricted to lie on a circle of a given radius. The details of the algorithm in the time-variant scenario are provided in [3] and [4].

A sequence of $(n+1)$ steps of the generalized Schur algorithm would lead to a cascade of n such sections, known as a transmission line and which we will denote by $\boldsymbol{T}$ (see Fig. 1). Under certain positivity and finite-dimensionality conditions [3], the cascade $\boldsymbol{T}$ is known to map, in a certain way, contractive operators K to contractive operators S, written simply as $S = \boldsymbol{T}[K]$.

Different choices for $\{\tau_i, \Theta_i\}$ lead to different expressions for the first-order sections and to different forms for the generator recursion itself. For example, one particular choice for $\{\tau_i, \Theta_i\}$, which will be discussed in Section V-A, allows the generator recursion to be written

Manuscript received January 18, 1996; revised August 1,. 1996. This work was supported in part by the Army Research Office under Contract DAAL03-89-K-0109, and by a grant from the National Science Foundation under award MIP-9409319.

T. Constantinescu is with the Programs in Mathematical Sciences, University of Texas at Dallas, Richardson, TX 75083 USA.

A. H. Sayed is with the Department of Electrical Engineering, University of California, Los Angeles, CA 90095 USA.

T. Kailath is with the Department of Electrical Engineering, Stanford University, Palo Alto, CA 94305 USA.

Publisher Item Identifier S 0018-9448(97)02318-3.

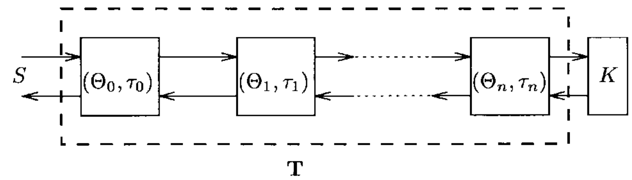

Fig. 1. A transmission-line mapping K into S.

in a simplified so-called proper form, which is often desirable from a computational point of view, e.g., in interpolation problems [3]–[5].

Other choices for $\{\tau_i, \Theta_i\}$, while leading to different forms for the first-order sections and for the generator recursion, further allow to impose other desirable properties on the cascade $\boldsymbol{T}$. The present correspondence addresses one such issue. More specifically, it shows how to construct a cascade $\boldsymbol{T}$, and in particular how to choose the above-mentioned free parameters $\{\tau_i, \Theta_i\}$, such that the resulting cascade $\boldsymbol{T}$ will map the zero load $(K = 0)$ to the maximum-entropy solution, as in the classical result [1]. We shall see that, in general, the cascade that corresponds to the proper choice for $\{\tau_i, \Theta_i\}$ does not map the zero-load to the maximum-entropy solution. Moreover, we shall be motivated to introduce a new set of contractive coefficients, one for each section of the cascade, and which will be shown, in general, to be distinct from the Schur parameters encountered in the proper case (see, e.g., [4, Sec. 5] and [3]).

A. Related Works in the Literature

Similar issues of relating the maximum-entropy solution to the central solution (corresponding to the zero load) have been addressed in the literature [6]–[8].

The work [6] deals with time-dependent entropy problems and also considers contractive extension problems. The framework of the lifting of commutants is employed in [7], while [8] employs tools of the W-transform technique studied in [9]. In particular, the work [8] poses a maximum-entropy problem in the context of linear fractional transformations that arise in time-variant discrete-time H_∞ control. The work shows how to choose a particular contractive load that maximizes a time-variant entropy measure, and provides state-space formulas and global expressions for the entropy operator.

The current work departs from earlier work in the sense that it focuses on a recursive (rather than a global) construction of the maximum-entropy solution. This is useful in situations when the available data is updated and it is desired to re-evaluate the corresponding maximum-entropy solution by exploiting the available cascade from the earlier calculations. A recursive procedure allows us to evaluate this new solution by simply appending a new section to the earlier cascade. Global expressions, on the other hand, need to be evaluated afresh whenever the data is modified, which is not convenient in recursive scenarios that arise, for example, in adaptive schemes.

We have chosen to present the results of this correspondence in an operator setting for generality of exposition. The results, however, can be easily specialized to particular situations.

B. Notation

The symbol $\mathbb{Z}$ denotes the set of integers, and for two Hilbert spaces $\mathcal{H}$ and $\mathcal{H}'$ we write $\mathcal{L}(\mathcal{H}, \mathcal{H}')$ to denote the set of bounded linear operators acting from $\mathcal{H}$ into $\mathcal{H}'$. We further consider three families $\{\mathcal{U}(t), \mathcal{V}(t), \mathcal{R}(t)\}_{t\in\mathbb{Z}}$ of Hilbert spaces depending on the parameter $t \in \mathbb{Z}$, two families of bounded linear operators

$G(t) \in \mathcal{L}(\mathcal{U}(t) \oplus \mathcal{V}(t), \mathcal{R}(t))$ and $F(t) \in \mathcal{L}(\mathcal{R}(t-1), \mathcal{R}(t))$, and we define the symmetry $J(t) = (I_{\mathcal{U}(t)} \oplus -I_{\mathcal{V}(t)})$ acting on $\mathcal{U}(t) \oplus \mathcal{V}(t)$, where $I_{\mathcal{U}(t)}$ denotes the identity operator on the space $\mathcal{U}(t)$. We partition $G(t) = [U(t) \;\; V(t)]$, where $U(t) \in \mathcal{L}(\mathcal{U}(t), \mathcal{R}(t))$ and $V(t) \in \mathcal{L}(\mathcal{V}(t), \mathcal{R}(t))$. We also use the symbol * to denote the adjoint operator and we write $F^*(t) = (F(t))^*$.

Definition 1: A family of operators $\{R(t) \in \mathcal{L}(\mathcal{R}(t))\}_{t\in\mathbb{Z}}$ is said to have a time-variant displacement structure with respect to $\{F(t), G(t)\}_{t\in\mathbb{Z}}$ if $\{R(t)\}_{t\in\mathbb{Z}}$ is uniformly bounded, viz., there exists $r > 0$ such that $\|R(t)\| \le r$ for all $t \in \mathbb{Z}$, and $R(t)$ satisfies the time-variant Lyapunov (or displacement) equation

$$R(t) - F(t)R(t-1)F^*(t) = G(t)J(t)G^*(t). \tag{1}$$

The cardinal number $r(t) = \dim \mathcal{U}(t) + \dim \mathcal{V}(t)$ is called the displacement rank of $R(t)$ in (1). We say that (1) has a Pick solution if $R(t)$ is positive-semidefinite for every $t \in \mathbb{Z}$. ■

Throughout the correspondence we assume that the following conditions hold (viz., [4, conditions (8a)–(8e)]): a) there exists a positive integer n such that $\mathcal{R}(t) = \oplus_{i=0}^{n-1} \mathcal{R}_i(t)$, for all t; b) $\dim \mathcal{R}_i(t)$ are all equal and finite; c) $\dim \mathcal{U}(t)$ and $\dim \mathcal{V}(t)$ are finite; and d) $\{F(t)\}$ is a uniformly bounded family of lower triangular operators with stable families of diagonal entries $\{f_i(t)\}_{i=0}^{n-1}$ (i.e., there exist $c_i > 0$ such that $\|f_i(t)\| \le c_i < 1$ for all t); e) $\{G(t)\}$ is a uniformly bounded family. Under these assumptions, the infinite block matrices

$$\begin{aligned} \boldsymbol{U}(t) &= [\cdots \quad F(t)F(t-1)U(t-2) \quad F(t)U(t-1) \quad U(t)] \\ \boldsymbol{V}(t) &= [\cdots \quad F(t)F(t-1)V(t-2) \quad F(t)V(t-1) \quad V(t)] \end{aligned}$$

are well-defined bounded linear operators, and the displacement equation (1) is guaranteed to have a unique uniformly bounded solution that is given by

$$R(t) = \boldsymbol{U}(t)\boldsymbol{U}^*(t) - \boldsymbol{V}(t)\boldsymbol{V}^*(t).$$

We further assume the following so-called nondegeneracy condition: f) the operator $\boldsymbol{U}(t)\boldsymbol{U}^*(t)$ is uniformly bounded from below, viz., $\exists\ \mu > 0$ such that $\boldsymbol{U}(t)\boldsymbol{U}^*(t) \ge \mu > 0$ for all $t \in \mathbb{Z}$.

Assumptions a)–f) allow us to state (see [4, Theorem 4.7]) that the time-variant displacement equation (1) has a Pick solution $R(t)$ such that $R(t) > \epsilon I > 0$ for a constant ϵ and for all $t \in \mathbb{Z}$ if, and only if, there exists an upper-triangular strict contraction $S(\|S\| < 1)$

$$S \in \mathcal{L}(\underset{t\in\mathbb{Z}}{\oplus} \mathcal{V}(t), \underset{t\in\mathbb{Z}}{\oplus} \mathcal{U}(t))$$

such that

$$\boldsymbol{V}(t) = \boldsymbol{U}(t)P_{\mathcal{U}}(t)S/\underset{j\le t}{\oplus}\mathcal{V}(j), \quad \text{for every } t \in \mathbb{Z} \tag{2}$$

where $P_{\mathcal{U}}(t)$ denotes the orthogonal projection of $\oplus_{t\in\mathbb{Z}}\mathcal{U}(t)$ onto $\oplus_{j\le t}\mathcal{U}(j)$.

Let $\mathcal{S}$ denote the set of all upper-triangular strictly-contractive operators S that satisfy (2). For every such $S \in \mathcal{S}$ it follows that $I - S^*S$ is a positive operator. Let Ψ_S denote its spectral factor (as defined in [10]–[12]). In the following, we write $D(A)$ to denote the diagonal of an upper-triangular operator A.

Problem 1: Let Ψ_S denote the spectral factor of an upper-triangular strictly-contractive operator $S \in \mathcal{S}$. The maximum-entropy problem is to solve the following optimization criterion:

$$\max_{S\in\mathcal{S}} \{D(\Psi_S)^* D(\Psi_S)\}. \tag{3}$$

■

Interpretations, motivations, and applications of problems of this kind abound in the literature. For formulations close to the above one we refer to [6]–[8], [13].

II. Solution of the Optimization Problem

Define the direct sum $\boldsymbol{J} = \oplus_{t\in\mathbb{Z}} J(t)$ and consider a bounded upper-triangular operator

$$\boldsymbol{T} \in \mathcal{L}([\underset{t\in\mathbb{Z}}{\oplus} \mathcal{U}(t)] \oplus [\underset{t\in\mathbb{Z}}{\oplus} \mathcal{V}(t)])$$

whose matrix entries $\{T_{lj}\}$ are partiotioned accordingly with $J(l)$ and $J(j)$, say

$$T_{lj} = \begin{bmatrix} T_{11}^{lj} & T_{12}^{lj} \\ T_{21}^{lj} & T_{22}^{lj} \end{bmatrix}.$$

We further construct the upper-triangular operators

$$\begin{aligned} &\boldsymbol{T}_{11} = [T_{11}^{lj}], \quad \boldsymbol{T}_{21} = [T_{21}^{lj}], \quad \boldsymbol{T}_{12} = [T_{12}^{lj}], \\ &\boldsymbol{T}_{22} = [T_{22}^{lj}], \quad \text{for } -\infty < l,\ j < \infty. \end{aligned}$$

The operator $\boldsymbol{T}$ will be said to be $\boldsymbol{J}$-inner (see, e.g., [15, Theorem 2.3]) if i) it is $\boldsymbol{J}$-unitary, i.e.,

$$\boldsymbol{TJT}^* = \boldsymbol{T}^*\boldsymbol{JT} = \boldsymbol{J}$$

and ii) $\boldsymbol{T}_{22}^{-1}$ is a bounded upper-triangular operator. In this case, it follows that $\boldsymbol{T}_{22}^{-1}\boldsymbol{T}_{21}$ is an upper-triangular strictly-contractive operator $\|\boldsymbol{T}_{22}^{-1}\boldsymbol{T}_{21}\| < 1$.

It was shown in [4, Theorem 4.8] that, starting with $\{F(t), G(t), J(t)\}$ of (1), there exists a bounded upper-triangular $\boldsymbol{J}$-inner operator $\boldsymbol{T}$ that can be determined as a function of the given $\{F(t), G(t), J(t)\}$, and such that $S \in \mathcal{S}$ if and only if there exists K such that

$$S = \boldsymbol{T}[K] = -(\boldsymbol{T}_{11}K + \boldsymbol{T}_{12})(\boldsymbol{T}_{21}K + \boldsymbol{T}_{22})^{-1} \tag{4}$$

where K is an upper-triangular strictly-contractive operator, $\|K\| < 1$. We now have the following.

Lemma 1: Consider an $S \in \mathcal{S}$ and let K be the associated operator, $S = \boldsymbol{T}[K]$. Then its spectral factor Ψ_S can be chosen according to the formula $\Psi_S = \Psi_K(\boldsymbol{T}_{21}K + \boldsymbol{T}_{22})^{-1}$.

Proof: It follows from the $\boldsymbol{J}$-innerness of $\boldsymbol{T}$ that

$$I - S^*S = (K^*\boldsymbol{T}_{21}^* + \boldsymbol{T}_{22}^*)^{-1}(I - K^*K)(\boldsymbol{T}_{21}K + \boldsymbol{T}_{22})^{-1}. \tag{5}$$

Let $\Psi_K \in \mathcal{L}(\oplus_{t\in\mathbb{Z}}\mathcal{V}(t), \oplus_{t\in\mathbb{Z}}\mathcal{V}'(t))$ be the spectral factor of K, and define $\Psi = \Psi_K(\boldsymbol{T}_{21}K + \boldsymbol{T}_{22})^{-1}$. We thus have that Ψ is an upper-triangular operator that obeys the condition that the space $\Psi[\oplus_{j\le t}\mathcal{V}(j)]$ is dense in $\oplus_{j\le t}\mathcal{V}'(j)$ for all $t \in \mathbb{Z}$. Moreover, the inequality $\Psi_K^*\Psi_K \le I - K^*K$ allows us to conclude, in conjunction with (5), that $\Psi^*\Psi \le I - S^*S$. Now consider any other upper-triangular contraction $Z \in \mathcal{L}(\oplus_{t\in\mathbb{Z}}\mathcal{V}(t), \oplus_{t\in\mathbb{Z}}\mathcal{V}(t))$ such that $Z^*Z \le I - S^*S$. It follows from (5) that

$$(K^*\boldsymbol{T}_{21}^* + \boldsymbol{T}_{22}^*)Z^*Z(\boldsymbol{T}_{21}K + \boldsymbol{T}_{22}) \le I - K^*K$$

and using the properties of the spectral factors, we must certainly have

$$(K^*\boldsymbol{T}_{21}^* + \boldsymbol{T}_{22}^*)Z^*Z(\boldsymbol{T}_{21}K + \boldsymbol{T}_{22}) \le \Psi_K^*\Psi_K.$$

This implies that $Z^*Z \le \Psi^*\Psi$ and, consequently, $\Psi = \Psi_K(\boldsymbol{T}_{21}K + \boldsymbol{T}_{22})^{-1}$ can be chosen as the spectral factor of $S = \boldsymbol{T}[K]$. ■

We are now in a position to state the solution of problem (3) (see also [7] and [8] for alternative arguments). For this purpose, and for notational convenience, we denote the upper-triangular operators $\boldsymbol{T}_{22}^{-1}\boldsymbol{T}_{21}$ and $\boldsymbol{T}_{22}^{-1}$ by χ and φ, respectively.

Lemma 2: Assume conditions a)–f) hold and let $\boldsymbol{T}$ be the $\boldsymbol{J}$-inner upper-triangular operator described above. Then

$$\begin{aligned}\max_{S\in\mathcal{S}} \{D(\Psi_S)^* D(\Psi_S\} &= [D(\boldsymbol{T}_{22})D(\boldsymbol{T}_{22})^* - D(\boldsymbol{T}_{21})D(\boldsymbol{T}_{21})^*]^{-1}, \\ &= [D(\boldsymbol{T}_{22})^*]^{-1}[I - D(\chi)D(\chi)^*]^{-1}[D(\boldsymbol{T}_{22})]^{-1}.\end{aligned}$$

Moreover, the maximum is attained if, and only if, $S = S_0 = \boldsymbol{T}[D(\chi)^*]$. In particular, if $D(\chi) = 0$ or, equivalently, $D(\boldsymbol{T}_{21}) = 0$, then

$$\max_{S\in\mathcal{S}}\{D(\Psi_S)^* D(\Psi_S\} = [D(\boldsymbol{T}_{22})^*]^{-1}[D(\boldsymbol{T}_{22})]^{-1}$$

and the maximum is attained for $S = S_0 = \boldsymbol{T}[0] = -\boldsymbol{T}_{12}\boldsymbol{T}_{22}^{-1}$.

Proof: The argument uses Lemma 1 and follows closely the proof of the main result in [13] (which is in Russian). (The monograph [14, especially, ch. 11] contains a number of examples of maximum entropy problems that might be more accessible to an English reader). ■

The unique

$$S_0 = \boldsymbol{T}[D(\chi)^*] = \boldsymbol{T}[D(\boldsymbol{T}_{22}^{-1}\boldsymbol{T}_{21})^*]$$

is called the maximum-entropy solution of (3). The unique $\boldsymbol{T}[0] = -\boldsymbol{T}_{12}\boldsymbol{T}_{22}^{-1}$ is called the central solution since it corresponds to choosing $K = 0$.

As mentioned earlier in Section I, the above statement provides a global characterization of the maximum entropy solution (and has also been studied in [6]–[8]). In particular, note that the expression for the required load is given in terms of the (block) entries of the entire cascade $\boldsymbol{T}$. The contribution of this correspondence is to exhibit a recursive construction of the maximum-entropy solution S_0 that does not require prior knowledge of the global expression for $\boldsymbol{T}$. The details are presented in the remaining sections.

III. A Recursive Solution

The recursive procedure will follow from an algorithm derived in [3], [4] for the triangular factorization of time-variant matrices with displacement structure.

To clarify this, consider block matrices $R(t) = [r_{lj}(t)]_{l,j=0}^{n-1}$ and let $R_i(t)$ denote the Schur complement of the leading $i \times i$ block submatrix of $R(t)$. If $l_i(t)$ and $d_i(t)$ stand for the first block column and the $(0, 0)$ block entry of $R_i(t)$, respectively, then the successive Schur complements of $R(t)$ are recursively related as follows:

$$R_i(t) - l_i(t)d_i^{-1}(t)l_i^*(t) = \begin{bmatrix} 0 & 0 \\ 0 & R_{i+1}(t)\end{bmatrix} \qquad R_0(t) = R(t).$$

We further note that the positive-definiteness of $R(t)$ guarantees $d_i(t) > 0$ for all i. Also, the notation $d^{-1}(t)$ stands for $d(t)^{-1}$.

After n consecutive Schur complement steps we obtain the block-triangular factorization of $R(t)$, viz.,

$$\begin{aligned}R(t) &= l_0(t)d_0^{-1}(t)l_0^*(t) + \begin{bmatrix} 0 \\ l_1(t)\end{bmatrix} d_1^{-1}(t)\begin{bmatrix} 0 \\ l_1(t)\end{bmatrix}^* + \cdots \\ &= L(t)D^{-1}(t)L^*(t)\end{aligned}$$

where $D(t) = \mathrm{diag}\{d_0(t), \cdots, d_{n-1}(t)\}$ is a block-diagonal matrix, and the (nonzero parts of the) columns of the block lower-triangular matrix $L(t)$ are $\{l_0(t), \cdots, l_{n-1}(t)\}$. It was shown in [4], [3] that for structured matrices $R(t)$ as in (1), the triangular factor at time $t - 1$, viz., $L(t - 1)$, can be time-updated to the triangular factor at time t, $L(t)$, via a recursive procedure on the generator matrix $G(t)$ as described below:

Start with $F_0(t) = F(t), G_0(t) = G(t)$, and repeat for $i \geq 0$.

- Choose uniformly bounded sequences $\{h_i(t), k_i(t)\}_{t\in\mathbb{Z}}$ that satisfy the following time-variant embedding relation:

$$\begin{bmatrix} f_i(t) & g_i(t) \\ h_i(t) & k_i(t)\end{bmatrix}\begin{bmatrix} d_i(t-1) & 0 \\ 0 & J(t)\end{bmatrix}\begin{bmatrix} f_i(t) & g_i(t) \\ h_i(t) & k_i(t)\end{bmatrix}^* = \begin{bmatrix} d_i(t) & 0 \\ 0 & J(t)\end{bmatrix} \tag{6}$$

where $g_i(t)$ denotes the top block row of $G_i(t)$.

- Apply the recursion

$$\begin{bmatrix} & 0 \\ l_i(t) & \\ & G_{i+1}(t)\end{bmatrix} = [F_i(t)l_i(t-1) \quad G_i(t)] \cdot \begin{bmatrix} f_i^*(t) & h_i^*(t)J(t) \\ J(t)g_i^*(t) & J(t)k_i^*(t)J(t)\end{bmatrix}. \tag{7}$$

Moreover,

$$d_i(t) = f_i(t)d_i(t-1)f_i^*(t) + g_i(t)J(t)g_i^*(t)$$

and $R_{i+1}(t)$ satisfies the time-variant displacement equation

$$R_{i+1}(t) - F_{i+1}(t)R_{i+1}(t-1)F_{i+1}^*(t) = G_{i+1}(t)J(t)G_{i+1}^*(t)$$

where $F_{i+1}(t)$ is the submatrix obtained after deleting the first row and column of $F_i(t)$. Let $\boldsymbol{T}_i = [T_{lj}^{(i)}]$ denote the upper-triangular transfer operator with time-variant Markov parameters:

$$\begin{aligned} T_{ll}^{(i)} &= J(l)k_i^*(l)J(l) \\ T_{l,l+1}^{(i)} &= J(l)g_i^*(l)h_i^*(l+1)J(l+1) \\ T_{lj}^{(i)} &= J(l)g_i^*(l)f_i^*(l+1)f_i^*(l+2)\cdots f_i^*(j-1)h_i^*(j)J(j), \\ &\qquad \text{for } j > l+1. \end{aligned} \tag{8}$$

After n recursive steps we obtain a cascade of sections $\boldsymbol{T} = \boldsymbol{T}_0\boldsymbol{T}_1\cdots\boldsymbol{T}_{n-1}$, which may be regarded as a generalized transmission line. This is the $\boldsymbol{J}$-inner operator that parametrizes all $S \in \mathcal{S}$ in (4).

The choice of $\{h_i(t), k_i(t)\}$ in (6) is nonunique and, therefore, the generator matrix $G_{i+1}(t)$ in (7) is also nonunique. Each choice for $\{h_i(t), k_i(t)\}$ would lead to a valid $G_{i+1}(t)$. There are, for instance, special choices for $\{h_i(t), k_i(t)\}$ that would lead to considerable simplifications in the computational requirements, since they lead to what are known as *proper* generators, as developed in [18] for the time-invariant case and in [3] for the time-variant case. But these choices do not generally lead to a maximum-entropy solution.

We shall show, however, that it is always possible to find $\{h_i(t), k_i(t)\}$, usually distinct from the choice in the proper case, so as to result in a cascade $\boldsymbol{T}$ whose central value, viz., $\boldsymbol{T}[0] = -\boldsymbol{T}_{12}\boldsymbol{T}_{22}^{-1}$, will correspond to the maximum-entropy solution. To achieve this, all we need to do is to exhibit uniformly bounded choices for $\{h_i(t), k_i(t)\}$ that would result in a cascade $\boldsymbol{T}$ for which

$$D(\chi) = D(\boldsymbol{T}_{22}^{-1}\boldsymbol{T}_{21}) = 0.$$

One way to guarantee this is to require that for each individual operator $\boldsymbol{T}_i$ we have

$$D(\boldsymbol{T}_{22,i}^{-1}\boldsymbol{T}_{21,i}) = 0$$

where the index i in $\boldsymbol{T}_{22,i}$ and $\boldsymbol{T}_{21,i}$ refers to the ith section.

But first let us elaborate on the nonunique choice of $\{h_i(t), k_i(t)\}_{t\in\mathbb{Z}}$ so as to satisfy the embedding relation (6).

For this purpose, we recall a result in [4, Theorem 4.1] where it was shown that the following choices for $h_i(t)$ and $k_i(t)$ satisfy (6):

$$\begin{aligned} h_i(t) &= \Theta_i^{-1}(t)J(t)g_i^*(t)[d_i^{*/2}(t) - \tau_i(t)d_i^{*/2}(t-1)f_i^*(t)]^{-1} \\ &\quad \cdot [\tau_i(t)d_i^{-(1/2)}(t-1) - d_i^{-(1/2)}(t)f_i(t)] \\ k_i(t) &= \Theta_i^{-1}(t)\{I - J(t)g_i^*(t)[d_i^{*/2}(t) \\ &\quad - \tau_i(t)d_i^{*/2}(t-1)f_i^*(t)]^{-1}d_i^{-(1/2)}(t)g_i(t)\} \end{aligned} \tag{9}$$

for an arbitrary $J(t)$-unitary operator $\Theta_i(t)$ and an arbitrary unitary operator $\tau_i(t)$, whenever the inverse of $d_i^{*/2}(t) - \tau_i(t)d_i^{*/2}(t-1) \cdot f_i^*(t)$ exists. Here, $d_i^{1/2}(t)$ denotes the operator defined by $d_i(t) = d_i^{1/2}(t)d_i^{*/2}(t)$. (The finite-dimensionality conditions guarantee that it is always possible to choose a unitary matrix $\tau_i(t)$ so as to assure the invertibility of $d_i^{*/2}(t) - \tau_i(t)d_i^{*/2}(t-1)f_i^*(t)$.)

A specific choice for $\tau_i(t)$, along with the choice $\Theta_i(t) = I$, was shown in [4] to guarantee the corresponding $\{h_i(t), k_i(t)\}_{t\in\mathbb{Z}}$, which we shall denote by $\{\overline{h}_i(t), \overline{k}_i(t)\}_{t\in\mathbb{Z}}$, to be uniformly bounded. But other choices for $(\tau_i(t), \Theta_i(t)\}$ that would guarantee the uniform boundedness of the corresponding $\{h_i(t), k_i(t)\}_{t\in\mathbb{Z}}$ are also possible. Examples to this effect, with specific values for $(\tau_i(t), \Theta_i(t)\}$, are given later (see, e.g., (13)).

With each uniformly bounded choice $\overline{k}_i(t)$, we associate a strict contraction $\overline{\rho}_i(t)$ that is defined below, and which will be referred to as a generalized reflection coefficient.

Definition 2: Let $\{\overline{k}_i(t)\}_{t\in\mathbb{Z}}$ be any uniformly bounded sequence that satisfies the embedding relation (6), and partition it accordingly with $J(t)$

$$\overline{k}_i(t) = \begin{bmatrix} \overline{k}_i^{(11)}(t) & \overline{k}_i^{(12)}(t) \\ \overline{k}_i^{(21)}(t) & \overline{k}_i^{(22)}(t) \end{bmatrix}.$$

The corresponding generalized reflection coefficient $\overline{\rho}_i(t)$ is defined by

$$\overline{\rho}_i(t) = -\overline{k}_i^{(12)}(t)(\overline{k}_i^{(22)}(t))^{-1} \in \mathcal{L}(\mathcal{V}(t), \mathcal{U}(t)). \quad ■$$

We can now state the main result of this correspondence.

Theorem 1: Assume conditions a)–f) hold and let $R(t)$ be the unique Pick solution of (1), viz., $R(t) > \epsilon I > 0$ for a constant ϵ and for all $t \in \mathbb{Z}$. Then we can always choose uniformly bounded families $\{h_i(t), k_i(t)\}_{t\in\mathbb{Z}}$, such that the associated $\boldsymbol{J}$-inner operator $\boldsymbol{T}$ has the property that $S_0 = \boldsymbol{T}[0] = -\boldsymbol{T}_{12}\boldsymbol{T}_{22}^{-1}$. That is, the central solution of the cascade coincides with the maximum-entropy solution of Problem 3.

Proof: The proof is constructive. It follows from Lemma 2 that the central solution $\boldsymbol{T}[0] = -\boldsymbol{T}_{12}\boldsymbol{T}_{22}^{-1}$ coincides with the maximum-entropy solution S_0 if, and only if, $D(\boldsymbol{T}_{21}) = 0$. We now show how to choose uniformly bounded families $\{h_i(t), k_i(t)\}_{t\in\mathbb{Z}}$ so as to guarantee $D(\boldsymbol{T}_{21,i}) = 0$ for each $i = 0, 1, \cdots, n-1$.

We have indicated above that it is always possible to find uniformly bounded families $\{\overline{h}_i(t)\}_{t\in\mathbb{Z}}, \{\overline{k}_i(t)\}_{t\in\mathbb{Z}}$ such that the embedding relation (6) holds. Let $\overline{\boldsymbol{T}}_i = [\overline{T}_{lj}^{(i)}]_{l,j}$ denote the transfer operator associated with $\{f_i(t), g_i(t), \overline{h}_i(t), \overline{k}_i(t)\}$, as in (8). We conclude from the embedding relation (6) that

$$\overline{h}_i(t)d_i(t-1)\overline{h}_i^*(t) + \overline{k}_i(t)J(t)\overline{k}_i^*(t) = J(t)$$

and, consequently,

$$J(t) - \overline{k}_i(t)J(t)\overline{k}_i^*(t) = \overline{h}_i(t)d_i(t-1)\overline{h}_i^*(t) \geq 0.$$

Since $\dim \mathcal{U}(t) < \infty$ and $\dim \mathcal{V}(t) < \infty$, we also conclude that $J(t) - \overline{k}_i^*(t)J(t)\overline{k}_i(t) \geq 0$ for all $t \in \mathbb{Z}$. If we partition $\overline{k}_i(t)$ accordingly with $J(t)$

$$\overline{k}_i(t) = \begin{bmatrix} \overline{k}_i^{(11)}(t) & \overline{k}_i^{(12)}(t) \\ \overline{k}_i^{(21)}(t) & \overline{k}_i^{(22)}(t) \end{bmatrix}$$

we then obtain

$$\overline{k}_i^{*(22)}(t)\overline{k}_i^{(22)}(t) \geq I + \overline{k}_i^{*(12)}(t)\overline{k}_i^{(12)}(t).$$

Therefore, $\overline{k}_i^{(22)}(t)$ is invertible and $\|(\overline{k}_i^{(22)}(t))^{-1}\| \leq 1$. We also know that $\|\overline{k}_i^{(22)}(t)\| \leq M$ for a certain $M > 0$. We now define the corresponding generalized reflection coefficient

$$\overline{\rho}_i(t) = -\overline{k}_i^{(12)}(t)(\overline{k}_i^{(22)}(t))^{-1} \in \mathcal{L}(\mathcal{V}(t), \mathcal{U}(t)) \tag{10}$$

which satisfies

$$I - \overline{\rho}_i^*(t)\overline{\rho}_i(t) \geq \overline{k}_i^{*(22)}(t)^{-1}\overline{k}_i^{(22)}(t)^{-1} \geq \frac{1}{M^2}.$$

Hence,

$$(I - \overline{\rho}_i^*(t)\overline{\rho}_i(t))^{-1} \leq M^2.$$

Moreover, from the identity

$$(I - \overline{\rho}_i(t)\overline{\rho}_i^*(t))^{-1} = I + \overline{\rho}_i(t)(I - \overline{\rho}_i^*(t)\overline{\rho}_i(t))^{-1}\overline{\rho}_i^*(t)$$

we obtain that

$$I - \overline{\rho}_i(t)\overline{\rho}_i^*(t)) \leq 1 + M^2.$$

We further define the family of $J(t)$-unitary matrices $\Theta_i(t) = H(\overline{\rho}_i(t))$, and remark that it is uniformly bounded. Using this choice for $\Theta_i(t)$ in (9) we conclude that the choices

$$h_i(t) = \Theta_i^{-1}(t)\overline{h}_i(t) \qquad k_i(t) = \Theta_i^{-1}(t)\overline{k}_i(t)$$

satisfy the embedding relation (6), are uniformly bounded over t, and result in $D(\boldsymbol{T}_{21,i}) = 0$ since the choice for $\Theta_i(t)$ forces $k_i(t)$ to be block-lower-triangular or, equivalently, $J(t)k_i(t)^*J(t)$ to be block-upper-triangular. ■

We should note, however, that the construction used in the previous proof is only one, among several possibilities, that would guarantee the condition $D(\boldsymbol{T}_{21}) = 0$. This is because the above construction achieves $D(\boldsymbol{T}_{21}) = 0$ by assuring that each individual section, or operator, satisfies a similar condition, $D(\boldsymbol{T}_{21,i}) = 0$; thus resulting in an overall cascade that satisfies $D(\boldsymbol{T}_{21}) = 0$. But, as we shall show in an example in the next section, it is possible to have $D(\boldsymbol{T}_{21}) = 0$ without requiring all the individual sections to satisfy $D(\boldsymbol{T}_{21,i}) = 0$.

A. Strictly Lower-Triangular $F(t)$

Let us first concentrate on the case of strictly lower-triangular matrices $F(t)$, viz., $f_i(t) = 0$ for all $t \in \mathbb{Z}$ and $i = 0, \cdots, n-1$.

We begin with the additional assumption

$$\dim \mathcal{R}_i(t) = \dim \mathcal{U}(t) \ \text{ for all } t \in \mathbb{Z}, \qquad i = 0, 1, \cdots, n-1. \tag{11}$$

The more general case can be similarly treated and we ommit the details. Let $g_i(t) = [u_i(t) \quad v_i(t)]$ denote the top block row of $G_i(t)$, and note that it follows from the displacement equation for $R_i(t)$ that

$$g_i(t)J(t)g_i^*(t) = d_i(t) > 0.$$

This implies that there exists a uniquely determined matrix $\gamma_i(t)$, $\|\gamma_i(t)\| < 1$, such that

$$v_i(t) = u_i(t)\gamma_i(t) \tag{12}$$

and we can define the $J(t)$-unitary rotation $H(\gamma_i(t))$. It reduces the top row of $G_i(t)$ to the form $g_i(t)H(\gamma_i(t)) = [\delta_i(t) \quad 0_{\mathcal{V}(t)}]$, and we say that $G_i(t)$ is reduced to *proper form*. This will allow us to further simplify the generator recursion (7) as detailed ahead. We shall refer to the $\gamma_i(t)$ as the *Schur parameters* associated with the displacement equation (1), when $F(t)$ is strictly lower-triangular. Consider further the following uniformly bounded choices (recall (9)):

$$\begin{aligned}\overline{k}_i(t) &= I - J(t)g_i^*(t)d_i^{-1}(t)g_i(t)\\ \overline{h}_i(t) &= J(t)g_i^*(t)d_i^{-1/2}(t)\tau_i(t)d_i^{-1/2}(t-1)\end{aligned} \tag{13}$$

where $\tau_i(t)$ is unitary and $\Theta_i(t) = I$. We further partition $\overline{k}_i(t)$ accordingly with $J(t)$, and introduce the generalized reflection coefficients

$$\overline{\rho}_i(t) = -\overline{k}_i^{(12)}(t)(\overline{k}_i^{(22)}(t))^{-1}. \tag{14}$$

Despite of the simple proof, the following result is quite unexpected.

Theorem 2: Consider the setting of Theorem 1 and let $R(t)$ be the unique Pick solution of (1), viz., $R(t) > \epsilon I > 0$ for a constant ϵ and for all $t \in \mathbb{Z}$. Assume further that $F(t)$ is strictly lower-triangular and $\dim \mathcal{R}_i(t) = \dim \mathcal{U}(t)$ for all $t \in \mathbb{Z}$ and $i = 0, 1, \cdots, n-1$. Then the Schur parameters $\{\gamma_i(t)\}$, defined via (12), and the generalized reflection coefficients $\{\overline{\rho}_i(t)\}$, defined via (14), coincide.

$$\overline{\rho}_i(t) = \gamma_i(t), \qquad \text{for } t \in \mathbb{Z}, i = 0, 1, \cdots, n-1.$$

Proof: Since $\dim \mathcal{R}_i(t) = \dim \mathcal{U}(t)$ for all $t \in \mathbb{Z}$, $i = 0, 1, \cdots, n-1$, and $u_i(t)u_i^*(t) \geq \epsilon + v_i(t)v_i^*(t)$ for a certain $\epsilon > 0$, we get that $u_i(t)$ are invertible matrices. Consequently,

$$\begin{aligned}\overline{\rho}_i(t) &= u_i^*(t)d_i^{-1}(t)v_i(t)(I + v_i^*(t)d_i^{-1}(t)v_i(t))^{-1}\\ &= u_i^*(t)(u_i(t)(I - \gamma_i(t)\gamma_i^*(t))u_i^*(t))^{-1}u_i(t)\gamma_i(t)\\ &\quad \times (I + \gamma_i^*(t)u_i^*(t)(u_i(t)\\ &\quad \cdot (I - \gamma_i(t)\gamma_i^*(t))u_i^*(t))^{-1}u_i(t)\gamma_i(t))^{-1}\\ &= (I - \gamma_i(t)\gamma_i^*(t))^{-1}\gamma_i(t)(I + \gamma_i^*(t)\\ &\quad \cdot (I - \gamma_i(t)\gamma_i^*(t))^{-1}\gamma_i(t))^{-1}\\ &= (I - \gamma_i(t)\gamma_i^*(t))^{-1}\gamma_i(t)(I - \gamma_i^*(t)\gamma_i(t)) = \gamma_i(t).\end{aligned}$$

This result also follows by noting that the generator recursion (7) gets simplified once we incorporate into it the special choice $\Theta_i(t) = H(\gamma_i(t))$ and use (9) to write

$$k_i(t) = H(\gamma_i(t))^{-1}\overline{k}_i(t) \qquad h_i(t) = H(\gamma_i(t))^{-1}\overline{h}_i(t)$$

where $\{\overline{h}_i(t), \overline{k}_i(t)\}$ are as in (13). We readily conclude that

$$\begin{aligned}J(t)k_i^*(t)J(t) &= H(\gamma_i(t))\begin{bmatrix}0 & 0\\ 0 & I\end{bmatrix}\\ h_i^*(t)J(t) &= d_i^{-1/2}(t-1)\tau_i^*(t)d_i^{-1/2}(t)[\delta_i(t) \quad 0].\end{aligned}$$

Because of the assumption $\dim \mathcal{R}_i(t) = \dim \mathcal{U}(t)$ for all $t \in \mathbb{Z}$, $i = 0, 1, \cdots, n-1$, and the fact that

$$\delta_i(t)\delta_i^*(t) = g_i^*(t)J(t)g_i(t) = d_i(t)$$

it follows from a simple Schur complement argument that

$$I - \delta_i^*(t)d_i^{-1}(t)\delta_i(t) = 0.$$

These facts further allow us to choose the unitary matrix $\tau_i(t)$ so as to satisfy the relation

$$\delta_i^*(t-1)d_i^{-1/2}(t-1)\tau_i^*(t) = \delta_i^*(t)d_i^{-1/2}(t)$$

and the generator recursion (7) gets simplified to the following:

$$\begin{aligned}\begin{bmatrix}0\\ G_{i+1}(t)\end{bmatrix} &= F_i(t)G_i(t-1)H(\gamma_i(t-1))\begin{bmatrix}I & 0\\ 0 & 0\end{bmatrix}\\ &\quad + G_i(t)H(\gamma_i(t))\begin{bmatrix}0 & 0\\ 0 & I\end{bmatrix}.\end{aligned} \tag{15}$$

It is thus clear that

$$\begin{aligned}J(t)k_i^*(t)J(t) &= H(\gamma_i(t))\begin{bmatrix}0 & 0\\ 0 & I\end{bmatrix}\\ &= \begin{bmatrix}0 & -\gamma_i(t)[I - \gamma_i^*(t)\gamma_i(t)]^{-1/2}\\ 0 & [I - \gamma_i^*(t)\gamma_i(t)]^{-1/2}\end{bmatrix}\end{aligned}$$

is block upper-triangular, and the entire cascade will exhibit $D(\boldsymbol{T}_{21}) = 0$.

We can also obtain an expression for the value of (3).

Theorem 3: Consider the setting of Theorem 2 and $\Delta = [\Delta_{tt}]_{t \in \mathbb{Z}}$ denote the optimal diagonal operator

$$\Delta = \max_{S \in \mathcal{S}} \{D(\Psi_S)^* D(\Psi_S\}.$$

Then

$$\begin{aligned}\Delta_{tt} &= [I - \gamma_0^*(t)\gamma_0(t)]^{1/2}[I - \gamma_1^*(t)\gamma_1(t)]^{1/2}\\ &\quad \cdots [I - \gamma_{n-1}^*(t)\gamma_{n-1}(t)] \cdots [I - \gamma_1^*(t)\gamma_1(t)]^{1/2}\\ &\quad \cdot [I - \gamma_0^*(t)\gamma_0(t)]^{1/2}.\end{aligned}$$

Proof: Let $\boldsymbol{T}_i$ denote the ith section associated with the proper generator recursion (15). We already know that the central solution of the corresponding cascade $\boldsymbol{T}$ coincides with the maximum-entropy solution and, consequently,

$$\Delta = (D(\boldsymbol{T}_{22})^*)^{-1}(D(\boldsymbol{T}_{22}))^{-1}.$$

But, for each section $\boldsymbol{T}_i$, we have

$$[D(\boldsymbol{T}_{22,i})]_{tt} = (I - \gamma_i^*(t)\gamma_i(t))^{-(1/2)}.$$

Therefore,

$$\begin{aligned}[D(\boldsymbol{T}_{22})]_{tt} &= (I - \gamma_0^*(t)\gamma_0(t))^{-(1/2)}(I - \gamma_1^*(t)\gamma_1(t))^{-(1/2)}\\ &\quad \cdots (I - \gamma_{n-1}^*(t)\gamma_{n-1}(t))^{-(1/2)}\end{aligned}$$

and the required result now follows. ■

The previous discussion can be extended even if we drop assumption (11), viz., that $\dim \mathcal{R}_i(t) = \dim \mathcal{U}(t)$ for all $t \in \mathbb{Z}$ and $i = 0, 1, \cdots, n-1$. We omit the details here.

We may add that the case of strictly lower-triangular $F(t)$ covers the band completion problems studied in [16], as well as some contractive extension problems considered in [6], [9], [19]—see [4], [3] for details. It is also connected with the so-called time-domain model validation problem (see, e.g., [4]).

B. Lower-Triangular $F(t)$

The notion of proper generators can also be extended, under additional assumptions, to the case of lower-triangular $F(t)$ (i.e., an $F(t)$ that is not necessarily strictly lower-triangular) [3]. However, as [3, formulas (24) and (26)] show, the associated proper recursion does not lead to upper-triangular terms $J(t)k_i^*(t)J(t)$ and, consequently, the individual sections $\boldsymbol{T}_i$ will not satisfy the requirement $D(\boldsymbol{T}_{21,i}) = 0$. This means that the central solution of the cascade $\boldsymbol{T}$ that is constructed via the proper recursion, and using the classical Schur parameters, will not generally correspond to the maximum-entropy solution. The best illustration of this case is the consideration of the classical Nevanlinna recursion, which maps Schur functions $s_i(z)$ (i.e., functions that are analytic and bounded by unity in the unit disc) to Schur functions $s_{i+1}(z)$ as follows:

$$\begin{aligned}s_{i+1}(z) &= \frac{1 - f_i^* z}{z - f_i}\,\frac{s_i(z) - \gamma_i}{1 - \gamma_i^* s_i(z)},\\ &\qquad \gamma_i = s_i(f_i), \quad s_0(z) = s(z), \quad i \geq 0.\end{aligned} \tag{16}$$

This relation can be linearized by expressing $s_i(z)$ as the ratio of two power series, $s_i(z) = v_i(z)/u_i(z)$. It follows from (16) that we can also write

$$(z - f_i)[u_{i+1}(z) \;\; v_{i+1}(z)] = [u_i(z) \;\; v_i(z)]H(\gamma_i)\begin{bmatrix} \frac{z-f_i}{1-f_i^* z} & 0 \\ 0 & 1 \end{bmatrix} \tag{17}$$

where $H(\gamma_i)$ is the elementary hyperbolic rotation

$$H(\gamma_i) = \frac{1}{\sqrt{1-|\gamma_i|^2}}\begin{bmatrix} 1 & -\gamma_i \\ -\gamma_i^* & 1 \end{bmatrix} \qquad \gamma_i = \lim_{z \to f_i} \frac{v_i(z)}{u_i(z)}.$$

We see that each step of (17) gives rise to a first-order J-lossless section with transfer function [5]

$$\boldsymbol{T}_i(z) = H(\gamma_i)\begin{bmatrix} \frac{z-f_i}{1-f_i^* z} & 0 \\ 0 & 1 \end{bmatrix}.$$

The resulting cascade $\boldsymbol{T}(z)$ that can be associated with n steps of the above recursion is given by (see [5] for details, where these cascades were discussed in the context of time-invariant displacement equations of the form $R - FRF^* = GJG^*$)

$$\boldsymbol{T}(z) = \boldsymbol{T}_0(z)\boldsymbol{T}_1(z)\cdots\boldsymbol{T}_{n-1}(z).$$

Let us partition $\boldsymbol{T}(z)$ accordingly with $J = (1 \oplus -1)$

$$\boldsymbol{T}(z) = \begin{bmatrix} \boldsymbol{T}_{11}(z) & \boldsymbol{T}_{12}(z) \\ \boldsymbol{T}_{21}(z) & \boldsymbol{T}_{22}(z) \end{bmatrix}$$

and consider its central solution

$$\boldsymbol{T}[0] = -\frac{\boldsymbol{T}_{12}(z)}{\boldsymbol{T}_{22}(z)}.$$

(*Remark*. The notation $\boldsymbol{T}[0]$ for the central solution should not be confused with $\boldsymbol{T}(0)$, the value of $\boldsymbol{T}(z)$ at $z = 0$).

The question of interest is whether this central solution, which corresponds to the classical Schur parameters $\{\gamma_i\}$, has the maximum-entropy property. According to Lemma 2, the central solution coincides with the maximum-entropy solution if, and only if, $\boldsymbol{T}_{21}(z)$ is a strictly proper rational matrix function or, equivalently, $\boldsymbol{T}_{21}(0) = 0$. So let us verify if this condition is always met in the Nevanlinna case. For this purpose, we focus only, and without loss of generality, on the first two sections. That is, assume we have $n = 2$. This leads to a cascade $\boldsymbol{T}(z) = \boldsymbol{T}_0(z)\boldsymbol{T}_1(z)$

$$\boldsymbol{T}(z) = H(\gamma_0)\begin{bmatrix} B_0(z) & 0 \\ 0 & 1 \end{bmatrix} H(\gamma_1)\begin{bmatrix} B_1(z) & 0 \\ 0 & 1 \end{bmatrix}$$

whose $(2, 1)$ entry is then equal to

$$\boldsymbol{T}_{21}(z) = -\frac{1}{\sqrt{1-|\gamma_0|^2}}\frac{1}{\sqrt{1-|\gamma_1|^2}}[\gamma_0^* B_0(z)B_1(z) + \gamma_1^* B_1(z)].$$

Therefore,

$$\boldsymbol{T}_{21}(0) = -\frac{1}{\sqrt{1-|\gamma_0|^2}}\frac{1}{\sqrt{1-|\gamma_1|^2}}[\gamma_0^* B_0(0)B_1(0) + \gamma_1^* B_1(0)]$$

and it is clear that, in general, we have $\boldsymbol{T}_{21}(0) \neq 0$; thus confirming our earlier claim that the central solution of the Nevanlinna cascade does not coincide, in general, with the maximum-entropy solution. It is also clear that if $f_1 = 0$ and, consequently, $B_1(z) = z$, then $\boldsymbol{T}_{21}(0) = 0$ and the central solution will coincide with the maximum-entropy solution.

We now show how to use our earlier results in order to modify the Nevanlinna recursion and obtain an algorithm that leads to a cascade whose central solution coincides with the maximum-entropy solution. To clarify this, we first elaborate on the connection of the Schur parameters $\{\gamma_i\}$ and the generalized reflection coefficients $\{\overline{\rho}_i\}$. Indeed, we choose $\Theta_i = I$ and $\tau_i = 1 + f_i/1 + f_i^*$ in (9) and write

$$\overline{k}_i = I - Jg_i^* d_i^{-1}\left(1 - \frac{1+f_i}{1+f_i^*}f_i^*\right)^{-1} g_i.$$

The generalized reflection coefficient is then related to the Schur parameter γ_i via

$$\overline{\rho}_i = \frac{1+f_i^*}{1+f_i^*|\gamma_i|^2}\gamma_i. \tag{18}$$

This leads to the choices

$$h_i = H(\overline{\rho}_i)^{-1} d_i^{-1} J g_i^* \qquad k_i = H(\overline{\rho}_i)^{-1}\overline{k}_i$$

and to the first-order sections (see [5] for details)

$$\begin{aligned} \boldsymbol{T}_{\rho,i}(z) &= \left\{ I + [B_i(z) - 1]\frac{Jg_i^* g_i}{g_i J g_i^*} \right\} H(\overline{\rho}_i) \\ B_i(z) &= \frac{z - f_i}{1 - f_i^* z}. \end{aligned} \tag{19}$$

These sections are related to the earlier $T_i(z)$ via

$$\boldsymbol{T}_{\rho,i}(z) = \boldsymbol{T}_i(z)H(\gamma_i)^{-1}H(\overline{\rho}_i).$$

The corresponding generator recursion is given by

$$\begin{bmatrix} 0 \\ G_{i+1} \end{bmatrix} = \left[G_i + (\Phi_i - I)G_i \frac{Jg_i^* g_i}{g_i J g_i^*} \right] H(\overline{\rho}_i). \tag{20}$$

A simple computation shows that

$$H(\gamma_i)^{-1}H(\overline{\rho}_i) = \frac{|1 + f_i^*|\gamma_i|^2|}{(1 - |f_i|^2|\gamma_i|^2)^{1/2}} \cdot \begin{bmatrix} \frac{1}{1+f_i|\gamma_i|^2} & -\frac{-f_i^*\gamma_i}{1+f_i^*|\gamma_i|^2} \\ -\frac{f_i\gamma_i^*}{1+f_i|\gamma_i|^2} & \frac{1}{1+f_i^*|\gamma_i|^2} \end{bmatrix}.$$

If we define

$$\zeta_i = \frac{|1 + f_i^*|\gamma_i|^2|}{1 + f_i|\gamma_i|^2} \qquad c_i = f_i^*\gamma_i$$

then the generator recursion (20) leads to a modified Nevanlinna recursion of the type

$$\frac{\zeta_i^* s_{i+1}(z) + c_i}{1 + c_i^*\zeta_i^* s_{i+1}(z)} = \frac{1 - f_i^* z}{z - f_i}\frac{s_i(z) - \gamma_i}{1 - \gamma_i^* s_i(z)}, \qquad \gamma_i = s_i(f_i), \;\; s_0 = s, \;\; i \geq 0. \tag{21}$$

The central solution of the cascade associated with this modified recursion now coincides with the maximum-entropy solution. We should mention that a detailed analysis of this type of recursions appears in [17], where it is shown that (21) facilitates the study of the Nevanlinna–Pick problem for an infinite number of data.

IV. Concluding Remarks

We have shown that the displacement structure theory allows a general formulation of the maximum entropy problem and yields both global and recursive solutions. A new set of contractive coefficients has also been shown to arise in this context, and which are different from those encountered in other applications of the displacement theory, e.g., in factorization and interpolation problems.

REFERENCES

[1] J. P. Burg, "Maximum entropy spectral analysis," in *Proc. 37th Meet. Soc. Exploration Geophysicists* (Oklahoma City, OK, Oct. 1967).

[2] T. Kailath and A. H. Sayed, "Displacement structure: Theory and applications," *SIAM Rev.*, vol. 37, no. 3, pp. 297–386, Sept. 1995.

[3] A. H. Sayed, T. Constantinescu, and T. Kailath, "Time-variant displacement structure and interpolation problems," *IEEE Trans. Automat. Contr.*, vol. 39, pp. 960–976, May 1994.

[4] T. Constantinescu, A. H. Sayed, and T. Kailath, "Displacement structure and completion problems," *SIAM J. Matrix Anal. Appl.*, vol. 16, no. 1, pp. 58–78, Jan. 1995.

[5] A. H. Sayed, T. Kailath, H. Lev-Ari, and T. Constantinescu, "Recursive solutions of rational interpolation problems via fast matrix factorization," *Integral Equations and Operator Theory*, vol. 20, pp. 84–118, Oct. 1994.

[6] I. Gohberg, M. A. Kaashoek, and H. J. Woerdeman, "A maximum entropy principle in the general framework of the band method," *J. Functional Anal.*, vol. 95, pp. 231–254, 1991.

[7] C. Foias, A. Frazho, and I. Gohberg, "Central intertwining lifting, maximum entropy and their permanence," *Integral Equations and Operator Theory*, vol. 18, pp. 166–201, 1994.

[8] P. A. Iglesias, "An entropy formula for time-varying discrete-time control problems," in *Proc. 28th Conf. on Information Sciences and Systems* (Princeton, NJ, Mar. 1994), pp. 214–219. (Also, *SIAM J. Contr. Optimiz.*, Sept. 1996.)

[9] P. Dewilde and H. Dym, "Interpolation for upper triangular pperators," in *Operator Theory: Advances and Applications*, vol. 56, I. Gohberg, Ed. Basel, Switzerland: Birkhäuser-Verlag, 1992, pp. 153–260.

[10] T. Constantinescu, "Schur analysis of positive block-matrices," in *Operator Theory: Advances and Applications*, vol. 18, I. Gohberg, Ed. Boston, MA: Birkhäuser, 1986, pp. 191–206.

[11] M. Rosenblum and J. Rovnyak, *Hardy Classes and Operator Theory*. Oxford, U.K.: Oxford Univ. Press, 1985.

[12] B. Sz. Nagy and C. Foias, *Harmonic Analysis of Operators on Hilbert Space*. Amsterdam–Budapest: North Holland, 1970.

[13] D. Z. Arov and M. G. Krein, "On the evaluation of entropy functionals and their minima in generalized extension problems," *Acta Sci. Math. (Szeged)*, vol. 45, pp. 33–50, 1983.

[14] H. Dym, *J-Contractive Matrix Functions, Reproducing Kernel Hilbert Spaces and Interpolation*, CBMS, Amer. Math. Soc., vol. 71, RI, 1989.

[15] J. A. Ball, I. Gohberg, and M. A. Kaashoek, "Nevanlinna-Pick interpolation for time-varying input-output maps: The discrete case," in *Operator Theory: Advances and Applications*, vol.. 56, I. Gohberg, Ed. Basel, Switzerland: Birkhäuser-Verlag, 1992, pp. 1–51.

[16] H. Dym and I. Gohberg, "Extensions of band matrices with band inverses," *Linear Alg. its Appl.*, vol. 36, pp. 1–24, 1981.

[17] J. B. Garnett. *Bounded Analytic Functions*. New York: Academic, 1981.

[18] H. Lev-Ari and T. Kailath, "Lattice filter parametrization and modeling of nonstationary processes," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 2–16, Jan. 1984.

[19] J. A. Ball and I. Gohberg, "A commutant lifting theorem for triangular matrices with diverse applications," *Integral Equations and Operator Theory*, vol. 8, pp. 205–267, 1985.

Information and Entropy of Continuous Random Variables

Zdeněk Fabián

***Abstract*—The mean value of the square of a generalized score function is shown to be interpretable as information associated with a continuous random variable. This information is in particular cases equal to the Fisher information of the corresponding distribution.**

***Index Terms*—Entropy, Fisher information, information function.**

I. INTRODUCTION

It is still an open question which quantity should be taken as a measure of the average amount of information associated with a continuous random variable X with density p. It is well known that Shannon's differential entropy

$$h_S(X) = E_p(-\log p) = \int -\log p(x)\ p(x)\ dx$$

cannot be a natural candidate since h_S can be negative. We show that a suitably modified Fisher information can serve as such a quantity.

Let $T \subset R$, where R denotes the real line, be an open interval with the σ-field $\mathcal{B}_T$ of its Borel subsets and let $\Theta \subset R^m$ be an open set. Consider the usual parametric model

$$\mathcal{P}_T = \{T, \mathcal{B}_T, p(u|\theta)\colon u \in T, \theta \in \Theta\}$$

with densities regular in the Cramér–Rao sense. A simple particular case is the location model $\{R, \mathcal{B}_R, p(x-\mu), x, \mu \in R\}$, where the location parameter μ represents a shift along the x-axis.

Fisher information is usually defined with respect to parameters of $\mathcal{P}_T$. Recall that the Fisher information matrix $(g_{jk}(\theta))_1^m$ is given by

$$g_{jk}(\theta) = E_p s_j s_k, \qquad j, k = 1, \cdots, m$$

where

$$s_j(u|\theta) = \frac{\partial \log p(u|\theta)}{\partial \theta_j} \tag{1}$$

is the *likelihood score* for the parameter θ_j.

The concept of the Fisher information of a distribution is much less frequent. It is defined (e.g., [2, pp. 494]) as

$$I(X) = E_p s^2 = \int_R s^2(x) p(x)\ dx \tag{2}$$

where s is the *score function* of the distribution p, given by

$$s(x) = -\frac{p'(x)}{p(x)}. \tag{3}$$

It is easy to see that in the location model we have

$$I(X) = g_{11}(\mu)|_{\mu=0}. \tag{4}$$

Consider the function $s^2\colon R \to [0, \infty)$. In the case of an unimodal distribution on $(R, \mathcal{B}_R)$, it attains its minimum value at the least informative point $x = 0$ of the distribution. By (4), the mean value $E_p s^2$ has the meaning of an information. It seems that the value

Manuscript received May 31, 1995; revised August 13, 1996. This work was supported in part by the Grant Agency CAS under Grant A107570.

The author is with the Institute of Computer Science, Academy of Sciences of the Czech Republic, 182 07 Prague 8, Czech Republic.

Publisher Item Identifier S 0018-9448(97)02324-9.