1 RECURSIVE CONSTRUCTION OF MULTI-CHANNEL TRANSMISSION LINES WITH A MAXIMUM ENTROPY PROPERTY*

A. H. Sayed[1], T. Constantinescu[2], and T. Kailath[3]

[1]Department of Electrical Engineering
University of California
Los Angeles, CA 90024 USA
email: sayed@ee.ucla.edu

[2]Programs in Mathematical Sciences
University of Texas
Richardson, TX 75083 USA
email: tiberiu@utdallas.edu

[3]Department of Electrical Engineering
Stanford University
Stanford, CA 94305 USA
email: tkailath@leland.stanford.edu

*Published in *Codes, Graphs, and Systems*, R. E. Blahut and R. Koetter, editors, pp. 259–290, Kluwer, 2002. This work was partially supported by the National Science Foundation under award CCR-9732376 and by the Army Research Office under grant DAAH-04-96-1-0176.

Abstract: J. P. Burg gave a now-famous solution of the problem of finding the maximum-entropy extension of a section $\{c_0, c_1, \ldots, c_n\}$ of a scalar (stationary) covariance sequence, with a nice interpretation in terms of the reflection coefficients of an associated discrete transmission line: namely, that the maximum-entropy extension is obtained by adding trivial sections with zero reflection coefficients. This work inspired various extensions and generalizations, mostly without any associated physical interpretation. In this paper we shall use the generalized transmission lines associated with the basic triangular matrix factorization algorithms of the Displacement Structure Theory to obtain several interesting maximum entropy extension results. One is the perhaps surprising fact that Burg's zero-reflection-coefficient extension solution may not always hold when the coefficients $\{c_i\}$ are matrix-valued.

DEDICATED TO G. DAVID FORNEY WITH ADMIRATION AND APPRECIATION.

1.1 INTRODUCTION

The breadth of Dave Forney's abilities, interest and contributions is truly remarkable. He has studied, developed, learned and applied a surprising diversity of mathematical tools to a variety of problems in information, communication and system theory. At another end of the spectrum, he has been closely involved with software and hardware development and with standards issues in the conception, development and successful deployment of pioneering communications modems.

It is not possible, for several reasons, to contribute a paper directly overlapping some of Dave's research areas. The present contribution is not too far afield, however, and in fact relates to one of the most appealing, at least to us, features of several of Dave's mathematical contributions. Namely, that the development of many of his apparently quite abstract results has been grounded in a concrete physical implementation: for example, his very important contributions to the algebraic structure of rational matrices are intimately connected to the structure and properties of convolutional codes.

In our work, such a concrete foundation for several purely mathematical results has been provided by generalized transmission line realizations of a set of fundamental matrix factorization algorithms that we have called Generalized Schur algorithms; in essence, these provide fast algorithms for a large variety of mathematical and physical problems by incorporating various kinds of Displacement Structure into the fundamental and ubiquitous Gaussian elimination algorithm (see [1], [2]).

In this paper we show how this generalized transmission line structure can be used to generalize Burg's famous result [3] on maximum entropy extensions of a scalar covariance sequence, $\{1, c_1, \ldots, c_N\}$. In the transmission line interpretation, the classical Schur algorithm (Burg used the closely related Levinson algorithm) is used to construct N sections of a lossless transmission line, each section being characterized by a scalar reflection coefficient (of course, less than unity in magnitude). All extensions of the given covariance sequence can be obtained by adding additional sections to the line; the maximum entropy extension is obtained by adding (trivial) sections with reflection coefficients that are zero. Another characterization is that all extensions are obtained by terminating the constructed N-section line with a passive load; the maximum entropy section corresponds to using a zero load.

Burg's result has been further studied and generalized in several ways, see *e.g.*, [4], [5]. This paper presents some further extensions with some surprising features, especially that in the apparently straightforward case where the c_i are matrix valued, the zero load termination may not give the maximum entropy based extension (see *Sec. 7*). An outline of the rest of the paper is as follows: in the next section we present classical material about the trigonometric moment problem and the Schur test for its solution. Then we describe the array form Schur's algorithm and its scattering interpretation. The maximum entropy extension is introduced in *Sec. 5*. The last two sections contain our results concerning the connection with structured matrices and the construction of the multichannel maximum entropy extension.

1.2 CLASSICAL BACKGROUND MATERIAL

Consider a sequence of covariance, (or auto-correlation) coefficients $\{c_k\}$ of a zero-mean stationary random process $\{y_i\}$, say

$$c_k \triangleq \mathrm{E}\, y_i y_{i-k}^* \qquad \text{for} \quad k = \ldots -2, -1, 0, 1, 2 \ldots$$

where the letter E denotes expectation. The $z-$spectrum of the process $\{y_i\}$ is defined as the z-transform of its covariance sequence,

$$S_y(z) \triangleq \sum_{k=-\infty}^{\infty} c_k z^k \,. \tag{1.1}$$

The $z-$spectrum $S_y(z)$ has two fundamental properties:

i) Para-Hermitian symmetry, which means that $S_y(z)$ satisfies the relation $S_y(z) = \left[S_y\left(\frac{1}{z^*}\right)\right]^*$. This can be easily established from the definition (1.1) and the fact that $c_{-k} = c_k^*$ for all k.

ii) Nonnegativity, which means that $S_y(e^{j\omega}) \geq 0$. This is a non-trivial property that characterizes any $z-$spectrum of a truly stochastic process.

More importantly perhaps, is that the converse statement also holds. That is, any function of z obeying the above two properties must be the $z-$transform of a covariance (also known as a moment) sequence.

Two classical problems that were extensively studied at the beginning of the last century are the so-called trigonometric moment problem and the extension problem (see, *e.g.*, [6] for an account of these earlier contributions).

Problem 1 (Trigonometric Moment Problem) *Given a sequence of complex numbers $\{\ldots c_{-2}, c_{-1}, c_0, c_1, c_2 \ldots\}$, satisfying $c_k = c_{-k}^*$, find necessary and sufficient conditions for the sequence to be a moment (or covariance) sequence.*

In other words, the above problem seeks conditions that guarantee that a given sequence $\{c_k\}$ can be regarded as the covariance sequence of a true stochastic process, *viz.*, that its entries can be recovered via a relation of the form

$$c_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-j\omega k} d\mu(\omega) \tag{1.2}$$

for some positive measure μ on $[-\pi, \pi)$ (say $d\mu(\omega) = S_y(e^{j\omega})d\omega$, for some power spectral density function $S_y(e^{j\omega})$.

Problem 2 (Extension Problem) *Given a finite sequence of complex numbers $\{c_{-n}, \ldots, c_{-1}, c_0, c_1, \ldots, c_n\}$, satisfying $c_k = c_{-k}^*$, describe all possible extensions $\{c_k, |k| > n\}$ that result in a moment sequence.*

The two problems are related and, not surprisingly, have applications in many areas of statistical signal processing (*e.g.*, in the form of maximum entropy spectral estimation), estimation theory (*e.g.*, Wiener-Hopf theory and spectral factorizations), circuit theory (*e.g.*, lossless and passive circuit design), and control theory (*e.g.*, positive-real systems and the so-called Kalman-Yakubovich-Popov characterization of such systems).

The solutions of the above two problems were given in a series of works by Toeplitz, Carathéodory, and Schur. They were later extended, and at times significantly generalized, in several other directions.

In particular, Toeplitz (1907) showed that a sequence $\{c_k,$ such that $c_k = c_{-k}^*\}_{-\infty}^{\infty}$ is a moment sequence if, and only if, the matrices (now known as *Toeplitz* matrices)

$$T_k \triangleq \begin{bmatrix} c_0 & c_{-1} & c_{-2} & \cdots & c_{-k} \\ c_1 & c_0 & c_{-1} & \cdots & c_{-k+1} \\ c_2 & c_1 & c_0 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & c_{-1} \\ c_k & c_{k-1} & \cdots & c_1 & c_0 \end{bmatrix} \quad \text{for} \quad k = \ldots, -1, 0, 1, \ldots \tag{1.3}$$

are positive semi-definite for all k, *i.e.*, $T_k \geq 0$. Checking this condition would, at first sight, involve checking the nonnegativity of large determinantal formulas. Carathéodory (1911) connected this nonnegativity condition with function theory. He defined the power series

$$c(z) \triangleq c_0 + 2\sum_{k=1}^{\infty} c_k z^k \,, \tag{1.4}$$

and showed that $\{c_k\}_{-\infty}^{\infty}$ is a moment sequence if, and only if,

$$c(z) \text{ is analytic and } \mathrm{Re}\,[c(z)] \geq 0 \text{ in } |z| < 1 \,. \tag{1.5}$$

In other words, $c(z)$ needs to be a positive-real function (now known as a *Carathéodory* function). Following this work, Schur (1917) considered the bilinear transformation

$$s(z) \triangleq \frac{c(z) - c_0}{c(z) + c_0}, \tag{1.6}$$

and noted that $c(z)$ is positive-real if, and only if, $s(z)$ is analytic and bounded by unity ($|s(z)| \leq 1$) in $|z| < 1$ (such functions are now called *Schur functions*). Moreover, and strikingly for the times, Schur presented a recursive test (rather than large determinantal expressions) for checking if a scalar function $s(z)$ is of Schur type or not.

Theorem 1 (Schur's Test) *Consider the following recursive algorithm that starts with a given initial function $s(z)$,*

$$s_{i+1}(z) = \frac{1}{z}\,\frac{s_i(z) - \gamma_i}{1 - \gamma_i^* s_i(z)}\ , \quad s_0(z) = s(z), \quad \textit{and} \quad \gamma_i = s_i(0). \qquad (1.7)$$

Then the following statements hold:

- *$s(z)$ is analytic and bounded by unity in $|z| < 1 \iff |\gamma_i| \leq 1$ for all i.*
- *$|\gamma_i| < 1$ for $0 \leq i < n$ and $|\gamma_n| = 1$ for some n if, and only if, $s(z)$ is a finite Blaschke product of degree n (i.e., a rational all-pass function). This means that $s(z)$ is analytic and bounded by unity in $|z| < 1$ and $|s(e^{j\omega})| = 1$.*
- *Starting with a Schur function $s(z)$, each function $s_i(z)$ is also of Schur type.*

The above theorem also shows that Schur functions can be of two types. One type is characterized by reflection coefficients that are always strictly bounded by one, $\{|\gamma_i| < 1\}$ for all i. The other type corresponds to all-pass functions and is characterized by a finite set of initial reflection coefficients that are strictly bounded by one, say $\{|\gamma_i| < 1,\ 0 \leq i < n\}$, followed by $|\gamma_n| = 1$. It can be shown that any Schur-type function $s(z)$ that satisfies

$$\int_{-\pi}^{\pi} \log[1 - |s(e^{j\omega})|^2] d\omega \ > \ -\infty \qquad (1.8)$$

belongs to the first type. That is, all its reflection coefficients will be strictly less than one in magnitude. Now, in view of relation (1.10), the condition (1.8) is implied by the following so-called Paley-Wiener condition on the power spectral density function of the random process $\{y_i\}$,

$$\int_{-\pi}^{\pi} \log[S_y(e^{j\omega})] d\omega > -\infty. \qquad (1.9)$$

A stationary stochastic process that satisfies this condition is said to be *purely nondeterministic* to indicate that even prediction from an *infinite* past leaves a nonzero residual error. This is because in prediction theory, the integral in (1.9) is related to the prediction error energy from an infinite-order prediction problem. More specifically, and briefly, let $\hat{y}_{i|i-1:i-m}$ denote the linear least-mean-squares estimate of y_i that is based on the past m data, and define

$$\sigma_m^2 \triangleq E\left|y_i - \hat{y}_{i|i-1:i-m}\right|^2 .$$

That is, σ_m^2 is the variance of the corresponding prediction error (its independence of i is due to the stationarity assumption on the random process $\{y_i\}$). The asymptotic behaviour of σ_m^2 as $m \to \infty$ has been studied in some detail in [7, 8]. In Ch. 2 of the first reference it is shown that

$$\lim_{m\to\infty} \sigma_m^2 = \exp\left[\frac{1}{2\pi}\int_{-\pi}^{\pi} \log[S_y(e^{j\omega})]d\omega\right] ,$$

which explains our earlier claim that for a purely nondeterministic process $\{y_i\}$, the prediction residual variance is finite even for an infinite-order prediction. In fact, it is further shown in [7] that this happens if, and only if, the sequence of reflection coefficients is square summable,

$$\sum_{i=1}^{\infty} |\gamma_i|^2 < \infty.$$

This is known as Szegö's formula and is a celebrated result in the prediction theory of stationary processes.

We therefore have a sequence of results that address the moment problem by working with the interrelated quantities $\{c(z), s(z), S_y(z)\}$, which can be seen to be related as follows:

$$1 - s(z)\left[s\left(\frac{1}{z^*}\right)\right]^* = \frac{4S_y(z)}{[c(z)+1][c^*(z^{-*})+1]}. \tag{1.10}$$

But what about the extension problem? It is well-known that the solution is highly nonunique and that, among all solutions, a so-called central solution with a maximum entropy property stands out. This central solution has received considerable attention in the literature, especially since Burg's original work on spectral analysis [3], and it can be associated with the construction of a special scattering cascade with a zero load.

The purpose of this article is to show how to construct more general scattering cascades in a multi-channel scenario under a certain maximum entropy requirement. The cascades are more general not only in the sense that they involve multi-input/multi-output ports but also in that the pure delays that are associated with classical scattering media are now replaced by first-order Blaschke functions. In addition, and in order to meet the maximum entropy property, special care is needed in constructing the scattering cascade and in choosing the load.

1.3 ARRAY FORM OF SCHUR'S ALGORITHM

It turns out that a convenient framework to pursue these extensions is by the study of matrices with so-called displacement structure (see, e.g., [1, 9] and the references therein). A generalized form of Schur's recursion plays a significant role in this development and is therefore reviewed next.

To begin with, note that Schur's recursion (1.7) is nonlinear in $s_i(z)$. A linearized form can be obtained by expressing $s_i(z)$ as the ratio of two power series, say

$$s_i(z) \triangleq \frac{y_i(z)}{x_i(z)} ,$$

so from (1.7) we must have

$$\frac{y_{i+1}(z)}{x_{i+1}(z)} = \frac{y_i(z) - \gamma_i x_i(z)}{z\left[x_i(z) - \gamma_i^* y_i(z)\right]} .$$

We can therefore employ the following updates for the numerator and denominator of $s_{i+1}(z)$:

$$y_{i+1}(z) = \alpha_i \left[y_i(z) - \gamma_i x_i(z)\right] \quad (1.11)$$

$$x_{i+1}(z) = \alpha_i z \left[x_i(z) - \gamma_i^* y_i(z)\right] \quad (1.12)$$

for some arbitrary nonzero complex number α_i. If we choose (assuming $|\gamma_i| < 1$, which will be our focus in this paper for reasons to become clear as we progress),

$$\alpha_i \triangleq \frac{1}{\sqrt{1 - |\gamma_i|^2}} ,$$

then we can write (1.11)–(1.12) in compact matrix form as follows (after extracting a z term)

$$z \begin{bmatrix} x_{i+1}(z) & y_{i+1}(z) \end{bmatrix} = \begin{bmatrix} x_i(z) & y_i(z) \end{bmatrix} \underbrace{\Theta_i \begin{bmatrix} z & 0 \\ 0 & 1 \end{bmatrix}}_{\Theta_i(z)} \quad (1.13)$$

where Θ_i is the 2×2 elementary hyperbolic rotation,

$$\Theta_i \triangleq \frac{1}{\sqrt{1 - |\gamma_i|^2}} \begin{bmatrix} 1 & -\gamma_i \\ -\gamma_i^* & 1 \end{bmatrix}, \quad \gamma_i = \lim_{z \to 0} \frac{y_i(z)}{x_i(z)} ,$$

and $\Theta_i(z)$ is a so-called 2×2 $J-$lossless elementary section,

$$\Theta_i(z) \triangleq \Theta_i \begin{bmatrix} z & 0 \\ 0 & 1 \end{bmatrix} . \quad (1.14)$$

It is $J-$lossless since it is analytic in $|z| < 1$ and is $J-$unitary on the unit circle,

$$\Theta_i(e^{j\omega}) J \left[\Theta_i(e^{j\omega})\right]^* = J , \quad J = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} .$$

[This follows from the fact that the hyperbolic rotation satisfies $\Theta_i J \Theta_i^* = J$.]

We further invoke the power series expansions of $x_i(z)$ and $y_i(z)$, say

$$\begin{aligned} x_i(z) &= x_{ii} + x_{i+1,i}z + x_{i+2,i}z^2 + \dots \\ y_i(z) &= y_{ii} + y_{i+1,i}z + y_{i+2,i}z^2 + \dots \end{aligned}$$

and collect the coefficients $\{x_{ik}, y_{ik}\}$ into an array:

$$G_i \triangleq \begin{bmatrix} x_{ii} & y_{ii} \\ x_{i+1,i} & y_{i+1,i} \\ x_{i+2,i} & y_{i+2,i} \\ \vdots & \vdots \end{bmatrix}.$$

Then recursion (1.13) can be interpreted as follows:

$$G_i = \begin{bmatrix} \times & \times \\ \times & \times \\ \times & \times \\ \vdots & \vdots \end{bmatrix} \overset{\Theta_i(\gamma_i)}{\longrightarrow} \begin{bmatrix} \times' & 0 \\ \times' & \times' \\ \times' & \times' \\ \vdots & \vdots \end{bmatrix} \overset{shift}{\longrightarrow} \begin{bmatrix} 0 & 0 \\ \times'' & \times' \\ \times'' & \times' \\ \vdots & \vdots \end{bmatrix} = \begin{bmatrix} 0 \quad 0 \\ G_{i+1} \end{bmatrix}.$$

In words,

1. We start with an array G_i and define γ_i as the ratio of the entries of its top row.

2. We apply Θ_i to G_i and thus annihilate the rightmost entry of its top row.

3. We then shift down the first column of $G_i\Theta_i$, while leaving the second column unaltered.

4. The resulting array has a top zero row. Its lower part is G_{i+1}, which contains the coefficients of the power series expansions of $\{x_{i+1}(z), y_{i+1}(z)\}$ that define $s_{i+1}(z)$.

In matrix language, (1.13) and the above array picture can be expressed as

$$\begin{bmatrix} 0 \quad 0 \\ G_{i+1} \end{bmatrix} = \mathcal{Z} G_i \Theta_i \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + G_i \Theta_i \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \tag{1.15}$$

where $\mathcal{Z}$ denotes the (semi-infinite) lower triangular shift matrix with ones on the first subdiagonal and zeros elsewhere.

We therefore described an equivalent description of Schur's algorithm in terms of a procedure that involves only matrix operations on arrays of numbers. Such matrix-based descriptions are very useful and also very powerful; for one thing, they allow us to pursue generalizations that are difficult to develop (or perhaps envision) if we persist with the function language.

We can now state Schur's solution to the trigonometric moment problem in this array form and using the bilinear representation (1.6).

Algorithm 1 (Schur's Solution to the Moment Problem) *Apply Schur's test to* $s(z) = y_0(z)/x_0(z)$ *where,*

$$y_0(z) = \frac{1}{2}[c(z) - c_0] \quad \textit{and} \quad x_0(z) = \frac{1}{2}[c(z) + c_0] \,,$$

or, equivalently, to the array

$$G_0 = \begin{bmatrix} c_0 & 0 \\ c_1 & c_1 \\ c_2 & c_2 \\ c_3 & c_3 \\ \vdots & \vdots \end{bmatrix} ,$$

and verify whether the resulting reflection coefficients $\{\gamma_i\}$ *are bounded by one,* $|\gamma_i| \leq 1$. *More specifically,*

1. *If* $|\gamma_j| > 1$ *for some* j, *then the sequence* $\{c_k\}$ *is not a moment sequence.*
2. *If* $|\gamma_j| < 1$ *for* $0 \leq j < n$ *and* $|\gamma_n| = 1$ *at some iteration* n, *then* $\{c_k\}$ *is a moment sequence.*
3. *If* $|\gamma_j| < 1$ *for all* j *then* $\{c_k\}$ *is also a moment sequence.*

1.4 THE SCATTERING FRAMEWORK

An important aspect of Schur's algorithm is that it admits a physical interpretation in terms of a cascade of elementary sections that combine together to form a layered-medium structure. Such structures play a fundamental role in modern signal processing.

Returning to the array recursion (1.15), note that it can be graphically depicted as a cascade of elementary sections as shown in Fig. 1.1, where we defined

$$\gamma_i^c \triangleq \sqrt{1 - |\gamma_i|^2} \,.$$

Each section consists of a 2×2 hyperbolic rotation Θ_i followed by a unit-time delay (or storage) element denoted by z. The cascade represents a feedforward (and pipelineable) implementation of the array algorithm, where the entries of the two columns of G_0 are available at the input lines of the first section.

We can regard each section as a 2−input 2−output system with transfer function $\Theta_i(z)$ as in (1.14). Moreover, our convention in this paper is that inputs to a transfer matrix function are applied to its left (they are row input vectors) and, correspondingly, outputs are also row vectors. Hence, we may write (see Fig. 1.2)

$$\begin{aligned} \begin{bmatrix} o_1(z) & o_2(z) \end{bmatrix} &= \begin{bmatrix} i_1(z) & i_2(z) \end{bmatrix} \Theta_i(z) \\ &= \begin{bmatrix} i_1(z) & i_2(z) \end{bmatrix} \begin{bmatrix} 1/\gamma_i^c & -\gamma_i/\gamma_i^c \\ -\gamma_i^*/\gamma_i^c & 1/\gamma_i^c \end{bmatrix} \begin{bmatrix} z & 0 \\ 0 & 1 \end{bmatrix} , \end{aligned}$$

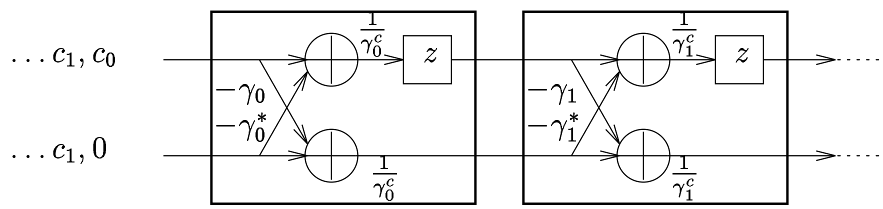

Figure 1.1 The feedforward structure associated with Schur's recursion.

to denote the input-output map of each elementary section in the z-transform domain, with $\{i_1(z), i_2(z)\}$ generically denoting the input signals and $\{o_1(z), o_2(z)\}$ denoting the output signals.

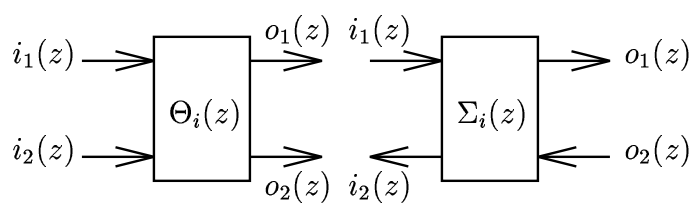

Figure 1.2 The scattering formulation for individual sections.

By reversing the direction of flow in a section we obtain the mapping

$$\begin{bmatrix} o_1(z) & i_2(z) \end{bmatrix} = \begin{bmatrix} i_1(z) & o_2(z) \end{bmatrix} \underbrace{\begin{bmatrix} \gamma_i^c & \gamma_i \\ -\gamma_i^* & \gamma_i^c \end{bmatrix} \begin{bmatrix} z & 0 \\ 0 & 1 \end{bmatrix}}_{\Sigma_i(z)}$$

where $\{i_1, o_2\}$ are now treated as input ports and $\{o_1, i_2\}$ are now treated as output ports (see Fig. 1.2). Each section $\Sigma_i(z)$ is now composed of a unitary gain matrix Σ_i ($\Sigma_i \Sigma_i^* = I$) followed by a unit-time delay element.

By combining several elementary sections $\Theta_i(z)$ we obtain a feedforward cascade whose overall 2×2 transfer matrix we shall denote by $\Theta(z)$ and is given by (*cf.* Fig. 1.1),

$$\Theta(z) = \Theta_0(z)\Theta_1(z)\Theta_2(z)\ldots$$

If we partition its entries accordingly with $J = (1 \oplus -1)$, say

$$\Theta(z) \triangleq \begin{bmatrix} \Theta_{11}(z) & \Theta_{12}(z) \\ \Theta_{21}(z) & \Theta_{22}(z) \end{bmatrix}, \tag{1.16}$$

where $\Theta_{ij}(z)$ are 1×1 transfer functions, then the entries of the scattering matrix $\Sigma(z)$ that corresponds to the combination of the corresponding scattering sections $\Sigma_i(z)$, i.e.,

$$\Sigma(z) = \Sigma_0(z) \star \Sigma_1(z) \star \Sigma_2(z) \star \ldots,$$

where $\star$ denotes the Redheffer product, are given by

$$\Sigma(z) = \begin{bmatrix} \Theta_{11}(z) - \Theta_{12}(z)\Theta_{22}^{-1}(z)\Theta_{21}(z) & -\Theta_{12}(z)\Theta_{22}^{-1}(z) \\ \Theta_{22}^{-1}(z)\Theta_{21}(z) & \Theta_{22}^{-1}(z) \end{bmatrix}. \tag{1.17}$$

The scattering cascade is depicted in Fig. 1.3, where we are denoting the entries of the sequences at its left-most ports generically by $\{x_{k0}, y_{k0}\}$. [In particular, if we use the special input sequences of Fig. 1.1, then $\{x_{k0}\} = \{c_0, c_1, \ldots\}$ and $\{y_{k0}\} = \{0, c_1, \ldots\}$. The transfer function from the left-most input to output ports in the scattering cascades is seen to be $-\Theta_{12}(z)\Theta_{22}^{-1}(z)$.

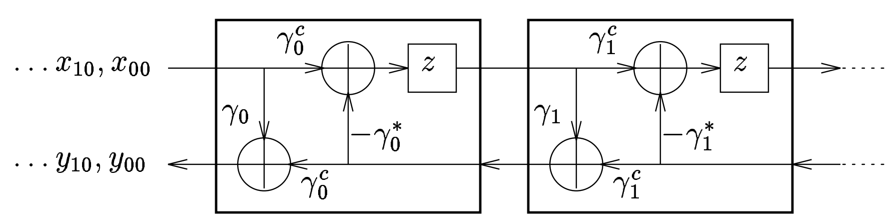

Figure 1.3 The feedback structure associated with Schur's algorithm.

We are now in a position to describe a solution to the extension problem that we posed earlier. Indeed, it is well-known that the above cascade can be used to solve the following slightly more specialized version of the extension problem, in which we require the Carathéodory function to be strictly positive-real.

Problem 3 (Specialized Extension Problem) *Given a finite sequence of complex numbers* $\{c_{-n}, \ldots, c_{-1}, c_0, c_1, \ldots, c_n\}$, *satisfying* $c_k = c_{-k}^*$, *describe all possible extensions* $\{c_k, |k| > n\}$ *that result in a moment (or covariance) sequence such that the corresponding* $c(z)$ *is strictly positive-real.*

The strict positivity requirement on $c(z)$ guarantees that the corresponding Schur function $s(z)$ that is obtained via the bilinear relation (1.6) is strictly bounded by unity in $|z| \leq 1$. We say that it is a strict Schur function and, in view of (1.8), its reflection coefficients will all be strictly bounded by one. This further means that the extended moment sequence will be the covariance sequence of a purely nondeterministic stationary stochastic process.

Now given the first $n+1$ moment values $\{c_k, 0 \leq k \leq n\}$ (and their complex conjugate values), we can determine the coefficients $\{s_k, 0 \leq k \leq n\}$ from the bilinear relation (1.6). Likewise, given $\{s_k, 0 \leq k \leq n\}$ (and c_0), we can determine the corresponding $\{c_k, 0 \leq k \leq n\}$. In order to guarantee a one-to-one correspondence between the $\{c_k\}$ and the $\{s_k\}$, we shall assume without loss of generality that the moment sequence is normalized so that $c_0 = 1$. We therefore pose the following equivalent problem.

Problem 4 (Interpolation Problem) *Given* $\{s_k, 0 \leq k \leq n\}$, *satisfying* $s_0 = 0$, *what are the possible extensions* $\{s_k, k > n\}$ *that result in a strict Schur function* $s(z)$*?*

The reason we refer to the above equivalent formulation as an interpolation problem is that the given coefficients $\{s_k\}$ can be interpreted as interpolation conditions on a function $s(z)$, *viz.*, we seek strict Schur functions $s(z)$ that satisfy

$$\frac{1}{k!}\frac{d^k}{dz^k}s(z)\bigg|_{z=0} = s_k \,. \tag{1.18}$$

This is a special analytic interpolation problem that can be solved in many different ways. For our purposes in this article, it is convenient to consider a solution that is based on the scattering cascade of Fig. 1.3 [10].

Given $\{s_k, s_0 = 0, 0 \leq k \leq n\}$, we form the matrix

$$G_0 = \begin{bmatrix} 1 & s_0 \\ 0 & s_1 \\ 0 & s_2 \\ \vdots & \vdots \\ 0 & s_n \end{bmatrix} , \tag{1.19}$$

and apply Schur's recursion (1.15) to these rows. Notice that in Sec. 1.3 we used the notation G_i (and G_0) to refer to a semi-infinite matrix. Here, we are using the same notation for a finite matrix (with $(n+1)$ rows). Referring to the array formulation (1.15) of Schur's algorithm we see that, for any k, the first k rows of the resulting array

$$\begin{bmatrix} 0 \quad 0 \\ G_{i+1} \end{bmatrix}$$

are completely determined by the first k rows of the array G_i. Hence, when we are only interested in a finite number of Schur steps, we can use a finite-dimensional version of (1.15), which we write as

$$\begin{bmatrix} 0 & 0 \\ G_{i+1} & \end{bmatrix} = ZG_i\Theta_i \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + G_i\Theta_i \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \qquad (1.20)$$

where Z now denotes a finite-dimensional lower triangular shift matrix (of appropriate dimensions) with ones on the first subdiagonal and zeros elsewhere.

By applying (1.20) to the above G_0, we identity $(n+1)$ reflection coefficients $\{\gamma_i, i = 0, \ldots, n\}$. If the coefficients $\{s_k, 0 \leq k \leq n\}$ we started with arose from numbers $\{c_k, c_0 = 1, c_k = c^*_{-k}, 0 \leq k \leq n\}$ that are part of a valid strict moment sequence, then the result of Schur's test in Thm. 1 (and the corresponding array solution in Alg. 1) indicates that all these $(n+1)$ reflection coefficients must be strictly bounded by one. If at any step k, $0 \leq k \leq n$, we find $|\gamma_k| \geq 1$, then the above interpolation problem is not solvable.

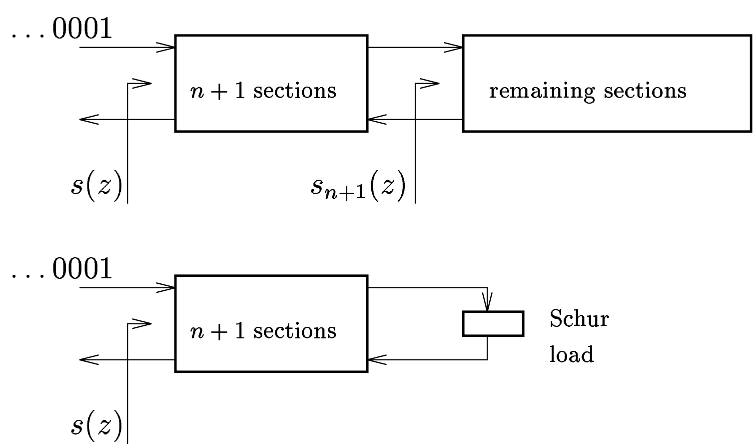

Figure 1.4 A scattering interpretation of the first $(n+1)$ steps.

Assuming that we succeed in getting $\{|\gamma_i| < 1, 0 \leq i \leq n\}$, we then obtain the $(n+1)$ sections of the feedforward cascade of Fig. 1.1, with individual transfer functions $\{\Theta_i(z), 0 \leq i \leq n\}$. The corresponding overall transfer matrix is

$$\Theta(z) = \Theta_0(z)\Theta_1(z)\ldots\Theta_n(z) .$$

It is analytic in $|z| \leq 1$ and $J-$unitary on the unit circle. This property in turn implies that if we partition the entries of $\Theta(z)$ as in (1.16) then $-\Theta_{12}(z)\Theta_{22}^{-1}(z)$ is a strict Schur function.

The corresponding scattering cascade $\Sigma(z)$ is shown schematically in Fig. 1.4. According to the partitioning (1.17), the above strict Schur function is the $(1,2)$ entry of $\Sigma(z)$ and is equal to the transfer function at the input of the scattering cascade of these $(n+1)$ sections. This transfer function is denoted by $s(z)$ in Fig. 1.4 because, besides being a strict Schur function, it also satisfies the required interpolation conditions. There are many ways to see this. Here we follow [2, Sec. 3.4]. Let $G_0(z)$ denote the $z-$transform of the initial array G_0,

$$G_0(z) \triangleq \begin{bmatrix} 1 & z & z^2 & \ldots & z^n \end{bmatrix} G_0 = \begin{bmatrix} 1 & \sum_{k=1}^{n} s_k z^k \end{bmatrix} . \qquad (1.21)$$

Then recall from the linearized function form (1.13) that the array algorithm (1.20) creates successive arrays G_i whose $z-$transforms satisfy

$$zG_{i+1}(z) = G_i(z)\Theta_i(z) .$$

This means that each step of Schur's algorithm produces a new row function $zG_{i+1}(z)$ with one more zero at the origin than the previous function $G_i(z)$. Therefore, after $(n+1)$ such steps and starting with $G_0(z)$ we must have

$$G_0(z)\Theta_0(z)\Theta_1(z)\ldots\Theta_n(z) = G_0(z)\Theta(z) = O(z^{n+1}) .$$

The notation $O(z^{n+1})$ means that the powers of z^i, $0 \leq i \leq n$, in the Taylor series expansion of $G_0(z)\Theta(z)$ around $z=0$ are all zero. Using the partitionings (1.16) and (1.21) and defining $s(z) = -\Theta_{12}(z)\Theta_{22}^{-1}(z)$, we obtain that

$$s(z) = \left[\sum_{k=1}^{n} s_k z^k\right] + O(z^{n+1}) .$$

This establishes that $s(z)$ satisfies (1.18).

In summary, we showed that the $(1,2)$ entry of the scattering cascade that is obtained from G_0 gives us one solution $s(z)$ in the form of the transfer function at the input of the cascade. But what about other solutions? It turns out that all of these can be obtained by attaching to the right-hand side of $\Sigma(z)$ any strict Schur-type function (see, e.g., [5]). This is indicated in Fig. 1.4, where $s_{n+1}(z)$ denotes the transfer function at the input terminals of the load and the block denoted "remaining sections" would correspond to a scattering implementation of this Schur load, if desired. The sequence $\{\ldots, 0001\}$ in the figure is used to indicate a unit pulse sequence and, therefore, that the impulse response of such a cascade will be $\{\boxed{s_0}, s_1, \ldots, s_n,\}$ followed by valid $\{s_k, k > n\}$ such that the corresponding $s(z)$ is a Schur function that satisfies the interpolation conditions.

In the above, we constructed the cascade $\Sigma(z)$ by using G_0 in (1.19). Alternatively, we could have also started with

$$G_0 = \begin{bmatrix} c_0 & 0 \\ c_1 & c_1 \\ c_2 & c_2 \\ \vdots & \vdots \\ c_n & c_n \end{bmatrix}. \qquad (1.22)$$

This is because the resulting reflection coefficients, which fully characterize the cascade, will be the same as can be seen by using again the bilinear transformation (1.6).

Algorithm 2 (Solution of the Interpolation Problem) *Given* $\{s_k, 0 \le k \le n\}$*, we apply* $(n+1)$ *steps of Schur's recursion (1.20) to the matrix*

$$G_0 = \begin{bmatrix} 1 & s_0 \\ 0 & s_1 \\ 0 & s_2 \\ \vdots & \vdots \\ 0 & s_n \end{bmatrix},$$

and construct the scattering cascade that corresponds to the first $(n+1)$ *sections. We then terminate the cascade with any strict Schur-type function. The transfer function at the input of the combined cascade solves the interpolation problem.*

It is well-known that this procedure can be used to solve the original extension problem (which asks for the coefficients $\{c_k\}$ rather than the $\{s_k\}$) by means of a perfect reflection experiment, as shown in Fig. 1.5 and as described in the following statement.

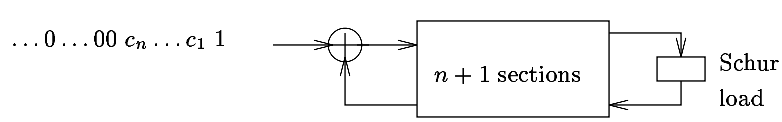

Figure 1.5 The perfect reflection experiment.

Algorithm 3 (Solution of the Extension Problem) *Given* $\{c_k, c_0 = 1, |k| < n\}$*, all valid extensions to a moment sequence can be obtained as follows:*

1. *Apply* $(n+1)$ *steps of Schur's recursion (1.15) to the matrix*

$$G_0 = \begin{bmatrix} 1 & 0 \\ c_1 & c_1 \\ c_2 & c_2 \\ \vdots & \vdots \\ c_n & c_n \end{bmatrix} ,$$

   and construct the scattering cascade that corresponds to the first $(n+1)$ *sections. We then terminate the cascade with any strict Schur-type function. The problem has a solution only if this step can be performed and results in all* γ_i *strictly less than unity in magnitude.*

2. *Connect the left-most output through an adder to the left-most input (cf. Fig. 1.5).*

3. *Excite the system with the input sequence* $\{\ldots 0 \ldots 00\, c_n \ldots c_1\, \boxed{1}\}$ *and measure the values at the left-most output for time instants* $k \geq n+1$. *These values provide the successive extension values* $\{c_j\}$.

1.5 MAXIMUM ENTROPY EXTENSIONS

The discussion in the last section shows that the solution of the extension problem, and of its equivalent interpolation problem, is highly nonunique due to the freedom in choosing the Schur load. Different loads would lead to different solutions and we therefore have a multitude of solutions.

Among all possibilities, the one that is most natural (in the sense that it is the most random or imposes the least constraints on the solution) corresponds to the so-called *maximum entropy* choice. The maximum entropy principle has been often used in the literature to choose, among the many solutions of an under-determined problem, the most natural (or the least-constrained) one — see [3, 11].

The maximum entropy criterion relies on the concept of differential entropy of a random variable. Let $f(y)$ denote the probability density function of a random variable y. Its differential entropy is defined by [11]

$$h(y) \triangleq -\int_{\text{SUP}} f(y) \ln(f(y)) dy = -E\left[\ln f(y)\right] , \tag{1.23}$$

where SUP is the support set of the random variable. When y has a normal distribution, say

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp^{-\frac{y^2}{2\sigma^2}} ,$$

its differential entropy can be shown to be $h(y) = \frac{1}{2}\ln(2\pi e\sigma^2)$ (where $\ln e = 1$). If y is an $n-$dimensional vector-valued random variable that is still normally distributed with covariance matrix R_y, say

$$f(y) = \frac{1}{\left(\sqrt{2\pi}\right)^n \sqrt{\det R_y}} \exp^{-\frac{1}{2}y^T R_y^{-1} y} ,$$

then its differential entropy reduces to $h(y) = \frac{1}{2}\ln[(2\pi e)^n \det R_y]$. In the limit, when y becomes a Gaussian random process (*viz.*, a multi-variable random vector with infinitely many components), its differential entropy reduces to

$$h(y) \;=\; \frac{1}{2\pi}\int_{-\pi}^{\pi} \ln[S_y(e^{j\omega})]d\omega \;, \tag{1.24}$$

where $S_y(e^{j\omega})$ is the power spectral density function of the process $\{y_i\}$.

This last expression motivates us to consider the following problem formulation. Given a finite-length sequence $\{c_k, c_0 = 1, 0 \le k \le n\}$, we know how to extend it to an infinitely-long valid moment sequence (that is strictly positive-real). Now since there are many solutions, one for each Schur-type load, we also obtain many possible $z-$spectra $S_y(z)$. Among all possible extensions of the moment sequence we seek that one whose $z-$spectrum is such that it has the most entropy.

Problem 5 (Maximum Entropy Problem) *Given a finite-length sequence*

$$\{c_k, c_0 = 1, 0 \le k \le n\} \,,$$

we wish to extend it, if possible, to a strictly positive-real moment sequence $\{c_k, |k| \ge 0\}$ *that meets the following optimization criterion:*

$$\max_{c_k, |k|>n} \left[\frac{1}{2\pi}\int_{-\pi}^{\pi} \ln[S_y(e^{j\omega})]d\omega\right] \,,$$

which is also equivalent to

$$\max_{\text{Schur load}} \left[\frac{1}{2\pi}\int_{-\pi}^{\pi} \ln[S_y(e^{j\omega})]d\omega\right] \,.$$

We can rewrite the cost function that appears in the above statement in a slightly different form that will be useful for our generalizations further ahead. For this purpose, we first remark that we can write

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \ln[S_y(e^{j\omega})]d\omega \;=\; 4\pi \ln\left[L(0)\right], \tag{1.25}$$

where $L(z)$ is the so-called Szegö function of $S_y(z)$. It is defined as

$$L(z) \;\triangleq\; \exp\left(\frac{1}{4\pi}\int_{-\pi}^{\pi} \frac{e^{j\omega}+z}{e^{j\omega}-z} \ln\left[S_y(e^{j\omega})\right] d\omega\right). \tag{1.26}$$

The Szegö function $L(z)$ is an outer function (*i.e.*, both $L(z)$ and its inverse are analytic in $|z| < 1$) that has the property that

$$|L(e^{j\omega})|^2 = S_y(e^{j\omega}),$$

almost everywhere. Consequently, it can be chosen as the spectral factor of $S_y(z)$.

Using the earlier relation (1.10) we also see that this spectral factor can be chosen as (recall that we are taking $c_0 = 1$)

$$L(z) = [c(z) + 1]L_s(z) \ , \qquad (1.27)$$

where $L_s(z)$ is the spectral factor of the function $1 - s(z)\left[s\left(\frac{1}{z^*}\right)\right]^*$. Equation (1.27) therefore shows that $L(0) = 2L_s(0)$, and the maximum entropy problem becomes equivalent to the following:

$$\max_{\text{Schur load}} \ [4\pi \ln L(0)] \iff \max_{\text{Schur load}} \ [8\pi \ln L_s(0)] \quad . \qquad (1.28)$$

In other words, we would like to choose the moments $\{c_k, |k| > n\}$ so as to maximize the value at 0 of either of the spectral factors, $L(z)$ or $L_s(z)$.

It is shown in [3] that the solution to the above problem is obtained when the Schur load in the scattering cascade of Fig. 1.4 is taken to be simply zero. This guarantees that there will be no direct path from the input to the output at the right-end of the scattering cascade; a condition that, as we shall explain further ahead, guarantees the maximum entropy property.

This construction that corresponds to the choice of the zero load is known as the *central solution*, and it arises in many contexts in systems, control, and estimation.

Algorithm 4 (Solution of Maximum Entropy Problem) *Simply use Algorithm 3 with the arbitrary Schur load in step 1 replaced by the zero load.*

1.6 GENERALIZED SCATTERING CASCADES

In this section we consider a generalization of the maximum entropy formulation of the previous section. The generalization is in two respects:

1. We study scattering cascades that involve multiple-input and multiple-output ports. In this case, the construction of maximum entropy solutions needs more care.

2. We replace the pure delay elements z that appear in the cascade of Fig. 1.3 by more general Blaschke factors, *viz.*, first-order rational all-pass functions of the form
$$\frac{z - f_i}{1 - f_i^* z} \ ,$$
for some $|f_i| < 1$.

It turns out that a convenient framework to pursue this extension is within the displacement structure theory (see, *e.g.,* [1, 9] for recent surveys and examples of many applications). No major background in the theory is required to follow the discussions in the sequel, especially since we shall motivate most of the arguments. Some basic background in matrix theory is enough.

1.6.1 Structured Matrices

The displacement structure concept provides a unifying framework for handling a variety of problems in signal processing, control, and mathematics, that involve some sort of matrix structure. For the moment and extension problems of the earlier sections, the matrix structure that we encountered was the Toeplitz structure in the matrices T_k in (1.3). It turns out that the Schur algorithm, when studied in the context of such structured matrices, can be extended and generalized rather significantly.

Consider again the Toeplitz matrix T_n of (1.3) — we continue to assume $c_0 = 1$ for simplicity. Let Z denote the lower triangular shift matrix (of appropriate dimensions). Then it can be verified that T_n satisfies the equation

$$T_n - ZT_nZ^* = G_0JG_0^* \ , \quad J = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} . \tag{1.29}$$

That is, the difference $T_n - ZT_nZ^*$ has low rank equal to 2 and independent of the dimension n, and its inertia is $\{1, -1, 0\}$. This is a consequence of the Toeplitz structure of T_n since it can be easily verified that

$$T_n - ZT_nZ^* = \begin{bmatrix} 1 & c_{-1} & c_{-2} & \dots & c_{-n} \\ c_1 & & & & \\ \vdots & & \bigcirc & & \\ c_n & & & & \end{bmatrix} .$$

Moreover, the above difference can be factored as $G_0JG_0^*$, where G_0 is the same matrix that we encountered before in (1.22), *viz.,*

$$G_0 = \begin{bmatrix} 1 & 0 \\ c_1 & c_1 \\ c_2 & c_2 \\ \vdots & \vdots \\ c_n & c_n \end{bmatrix} . \tag{1.30}$$

We call G_0 a generator matrix for T_n since it contains all the information needed about T_n. Note further that while T_n has n^2 entries, G_0 has $2n$ entries; an order of magnitude less than T_n. Hence, algorithms that operate on G_0 in order to extract some information about T_n will tend to be an order of magnitude faster than algorithms that operate on the entries of T_n themselves. The Schur algorithm is one such procedure. It can be shown that it not only verifies the positive-definiteness of T_n (which is the context in which it was used earlier in this paper by checking whether $|\gamma_i| < 1$), but it also provides the Cholesky factorization of T_n using only a multiple of n^2 operations.

More specifically, it can be verified that the first columns of the successive $G_i\Theta_i$ in the array form (1.20) correspond to the columns of the Cholesky factor of T_n, say $T_n = \bar{L}\bar{L}^*$, where $\bar{L}$ is lower triangular with columns

$$\bar{l}_i \;=\; G_i\Theta_i \begin{bmatrix} 1 \\ 0 \end{bmatrix} . \tag{1.31}$$

Even more important, it turns out that the Schur algorithm (1.20) can be used to (Cholesky) factor more general structured matrices. For example, it can factor any positive-definite matrix R that satisfies an equation of the form

$$R - ZRZ^* = GJG^* \ ,$$

for some 2-column matrix G and $J = (1 \oplus -1)$. That is, any matrix R for which the difference $R - ZRZ^*$ has rank 2 and inertia $\{1, -1, 0\}$. The factorization is achieved by simply applying (1.20) starting with $G_0 = G$ and using (1.31). Here, the only difference from the Toeplitz case is the initial general matrix G_0.

1.6.2 Generalized Schur Algorithm

Generalizations of both structure and Schur's work have been studied extensively in the literature (see [1, 9] and the references therein). Here we focus on one such extension.

Let F be an $n \times n$ lower triangular stable matrix with diagonal entries $\{f_i,\ |f_i| < 1\}$, and let R be an $n \times n$ positive-definite Hermitian matrix. We shall say that R has *displacement structure* if it satisfies an equation of the form (also called a displacement equation):

$$R - FRF^* = G \underbrace{\begin{bmatrix} I_p & 0 \\ 0 & -I_q \end{bmatrix}}_{J} G^* \ , \tag{1.32}$$

where J is a signature matrix that specifies the displacement inertia of R, and G is an $n \times r$ generator matrix with $r \ll n$ and $r = (p+q)$. We say that R has structure to indicate that the difference $R - FRF^*$ is low rank; its rank is called the displacement rank of R. The condition $|f_i| < 1$ is sufficient to guarantee that the displacement equation (1.32) uniquely specifies R from knowledge of $\{F, G, J\}$. Note also that the Toeplitz structure (1.29) is a special case of (1.32) by choosing $F = Z$, $G = G_0$, $p = q = 1$, and $J = (1 \oplus -1)$.

A major result concerning structured matrices R, is that the successive Schur complements of R, denoted by R_i, inherit a similar structure. That is, if R_i is the Schur complement of the leading $i \times i$ submatrix of R, then R_i also exhibits displacement structure of the form

$$R_i - F_iR_iF_i^* = G_iJG_i^* \ ,$$

where F_i is the submatrix obtained after deleting the first i rows and columns of F, and the generator G_i satisfies a recursive construction that we explain below.

Algorithm 5 (Generalized Schur Algorithm) *Generator matrices G_i for the successive Schur complements R_i of a positive-definite structured matrix R, as in (1.32), can be recursively constructed as follows. Start with $G_0 = G, F_0 = F$, and repeat for $i \geq 0$:*

1. *At step i we have F_i and G_i. Let g_i denote the top row of G_i.*
2. *Choose any $J-$unitary rotation Θ_i that reduces g_i to the form*

$$g_i\Theta_i = \begin{bmatrix} \delta_i & 0 & \dots & 0 \end{bmatrix} \tag{1.33}$$

   Such a rotation always exists in view of the positive-definiteness of R and it can be implemented in many different ways, e.g., *as a sequence of elementary unitary and hyperbolic rotations.*

3. *Apply Θ_i to G_i leading to the next generator as follows:*

$$\begin{bmatrix} \mathbf{0} \\ G_{i+1} \end{bmatrix} = G_i\Theta_i \begin{bmatrix} 0 & \mathbf{0} \\ \mathbf{0} & I \end{bmatrix} + \Phi_i G_i \Theta_i \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} , \tag{1.34}$$

   where Φ_i denotes a so-called Blaschke matrix,

$$\Phi_i = (F_i - f_i I)(I - f_i^* F_i)^{-1} .$$

4. *The columns of the Cholesky factor of $R = \bar{L}\bar{L}^*$ are given by*

$$\bar{l}_i = \sqrt{1 - |f_i|^2}\, (I - f_i^* F_i)^{-1} G_i \Theta_i \begin{bmatrix} 1 \\ \mathbf{0} \end{bmatrix} . \tag{1.35}$$

Recursion (1.34) is an extension of (1.20). Pictorially, we have the following (see Fig. 1.6):

$$G_i = \begin{bmatrix} \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \\ \vdots & & \vdots \end{bmatrix} \xrightarrow{\Theta_i} \begin{bmatrix} \delta_i & 0 & 0 \\ \times' & \times' & \times' \\ \times' & \times' & \times' \\ \vdots & & \vdots \end{bmatrix} \xrightarrow{\Phi_i} \begin{bmatrix} 0 & 0 & 0 \\ \times'' & \times' & \times' \\ \times'' & \times' & \times' \\ \vdots & & \vdots \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ G_{i+1} \end{bmatrix}$$

In words:

- Choose an $r \times r$ $J-$unitary rotation Θ_i that reduces the top row of G_i as in (1.33). We say that G_i is reduced to proper form.
- Apply Θ_i to G_i.

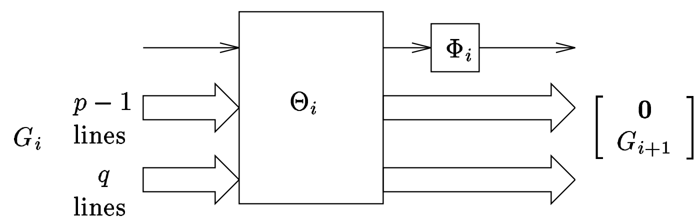

Figure 1.6 Pictorial representation of the generalized Schur algorithm.

- Multiply the first column of $G_i\Theta_i$ by Φ_i and keep all other columns unchanged.

A useful conclusion follows by combining the generator recursion (1.34) with the expression (1.35) for the i−th column of the Cholesky factor. Indeed, define

$$l_i \triangleq \sqrt{d_i}\,\bar{l}_i\,,$$

where $1/\sqrt{d_i}$ is the top entry of $\bar{l}_i$ (and equal to $|\delta_i|^2/(1-|f_i|^2)$). That is, the top entry of l_i is normalized to unity. Then it can be verified that (1.34) and (1.35) lead to the following expression

$$\begin{bmatrix} l_i & \begin{matrix} 0 \\ G_{i+1} \end{matrix} \end{bmatrix} = \begin{bmatrix} F_i l_i & G_i \end{bmatrix} \begin{bmatrix} f_i^* & \frac{\delta_i}{d_i}\begin{bmatrix} 1 & 0 \end{bmatrix} \\ \Theta_i \begin{bmatrix} \delta_i \\ 0 \end{bmatrix} & \Theta_i \begin{bmatrix} -f_i & 0 \\ 0 & I_{r-1} \end{bmatrix} \end{bmatrix} . \quad (1.36)$$

We can therefore regard the transformation

$$\begin{bmatrix} f_i^* & \frac{\delta_i}{d_i}\begin{bmatrix} 1 & 0 \end{bmatrix} \\ \Theta_i \begin{bmatrix} \delta_i \\ 0 \end{bmatrix} & \Theta_i \begin{bmatrix} -f_i & 0 \\ 0 & I_{r-1} \end{bmatrix} \end{bmatrix}$$

as the system matrix of a first-order state-space linear system; the rows of $\{G_i\}$ and $\{G_{i+1}\}$ can be regarded as inputs and outputs of this system, respectively, and the entries of $\{l_i, F_i l_i\}$ can be regarded as the corresponding current and future states. If we let $\Theta_i(z)$ denote the transfer function of the linear system (with inputs from the left), *viz.*,

$$\Theta_i(z) = \Theta_i \begin{bmatrix} -f_i & 0 \\ 0 & I_{r-1} \end{bmatrix} + \Theta_i \begin{bmatrix} \delta_i \\ 0 \end{bmatrix} (z^{-1} - f_i^*)^{-1} \frac{\delta_i}{d_i} \begin{bmatrix} 1 & 0 \end{bmatrix} ,$$

simple algebra will show that the above expression collapses to

$$\Theta_i(z) = \Theta_i \begin{bmatrix} B_i(z) & \mathbf{0} \\ \mathbf{0} & I_{r-1} \end{bmatrix}, \quad B_i(z) = \frac{z - f_i}{1 - f_i^* z}. \tag{1.37}$$

We therefore see that each step of the generalized Schur recursion gives us to a first-order section $\Theta_i(z)$ — see Fig. 1.7. [Compare this with $\Theta_i(z)$ in (1.14), which is obtained as a special case by setting $f_i = 0$ and $r = 2$.]

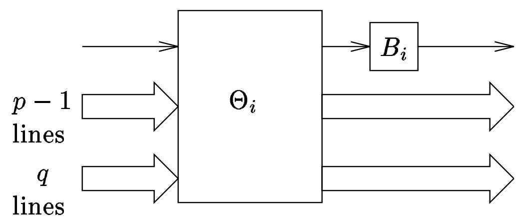

Figure 1.7 Elementary section $\Theta_i(z)$.

A succession of steps of the generalized Schur algorithm would therefore lead to a feedforward cascade of sections, say for $(n+1)$ steps,

$$\Theta(z) = \Theta_0(z)\Theta_1(z)\ldots\Theta_n(z) = \begin{bmatrix} \Theta_{11}(z) & \Theta_{12}(z) \\ \Theta_{21}(z) & \Theta_{22}(z) \end{bmatrix}, \tag{1.38}$$

which we partition accordingly with $J = (I_p \oplus -I_q)$. That is, $\Theta_{11}(z)$ is $p \times p$ $\Theta_{12}(z)$ is $p \times q$ $\Theta_{21}(z)$ is $q \times p$ $\Theta_{22}(z)$ is $q \times q$. The transformation implied by $\Theta(z)$ is depicted in Fig. 1.8. It is also a $J-$lossless transformation.

The associated scattering cascade would then be (see Fig. 1.9)

$$\Sigma(z) = \begin{bmatrix} \Theta_{11}(z) - \Theta_{12}(z)\Theta_{22}^{-1}(z)\Theta_{21}(z) & -\Theta_{12}(z)\Theta_{22}^{-1}(z) \\ \Theta_{22}^{-1}(z)\Theta_{21}(z) & \Theta_{22}^{-1}(z) \end{bmatrix}. \tag{1.39}$$

The array form that we described for the the generalized Schur algorithm is in fact a special case of a more general description that also turns out to be relevant to our discussions.

Algorithm 6 (General Generator Recursion) *Generator matrices G_i for the successive Schur complements R_i of a positive-definite structured matrix R,*

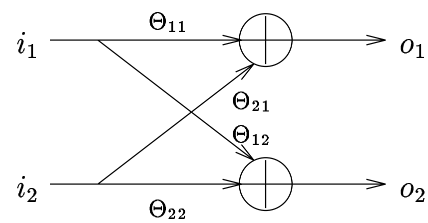

Figure 1.8 The feedforward cascade.

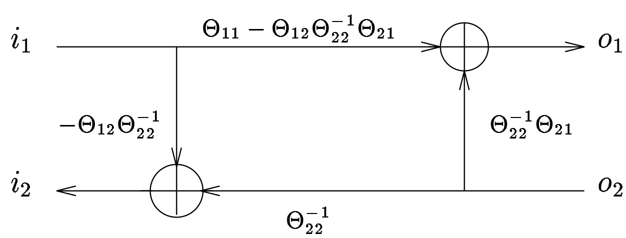

Figure 1.9 The scattering cascade.

as in (1.32), can be recursively constructed as follows. Start with $G_0 = G, F_0 = F$*, and repeat for* $i \geq 0$*:*

$$\begin{bmatrix} \mathbf{0} \\ G_{i+1} \end{bmatrix} = \left\{ G_i + (\Phi_i - I_{n-i}) G_i \frac{Jg_i^* g_i}{g_i J g_i^*} \right\} \Theta_i \ , \tag{1.40}$$

for <u>*any*</u> $J-$*unitary* Θ_i*. Moreover,*

$$\bar{l}_i = \frac{\sqrt{1 - |f_i|^2}}{\sqrt{g_i J g_i^*}} \ (I - f_i^* F_i)^{-1} G_i J g_i^* \ , \tag{1.41}$$

where the positive-definiteness of R*, and the fact that* $|f_i| < 1$*, guarantee* $g_i J g_i^* > 0$*. Also, the elementary sections* $\Theta_i(z)$ *that are generated by the algorithm are given by*

$$\Theta_i(z) = \left\{ I_r + [B_i(z) - 1] \frac{Jg_i^* g_i}{g_i J g_i^*} \right\} \Theta_i \ . \tag{1.42}$$

In the above description, Θ_i can be <u>any</u> $J-$unitary rotation matrix and we will again be led to a feedforward cascade $\Theta(z)$ and to a feedback cascade $\Sigma_i(z)$. The individual sections of $\Theta(z)$ will now be given by (1.42), while the individual sections of $\Sigma(z)$ will be obtained from (1.42) via an expression similar to (1.39). We may also note that (1.40) and (1.41) can be combined together as follows

$$\begin{bmatrix} l_i & 0 \\ & G_{i+1} \end{bmatrix} = \begin{bmatrix} F_i l_i & G_i \end{bmatrix} \begin{bmatrix} f_i^* & \frac{1}{d_i} g_i \Theta_i \\ J g_i^* & \left(I_r - (1 + f_i) \frac{J g_i^* g_i}{g_i J g_i^*} \right) \Theta_i \end{bmatrix} , \tag{1.43}$$

and that $\Theta_i(z)$ in (1.42) is now the transfer matrix of the above state-space transformation, with

$$d_i = \frac{g_i J g_i^*}{1 - |f_i|^2} \ .$$

The array form (1.34) follows as a special case of (1.40) when we take Θ_i to be that rotation that reduces g_i to the proper form (1.33). In this case, the expression (1.42) for $\Theta_i(z)$ also collapses to (1.37). Why do we consider this more general Schur recursion? The point is, as we shall see, the freedom in choosing Θ_i turns out to be useful in solving maximum entropy problems. Rather than restricting ourselves to that particular choice of Θ_i that reduces the generator to proper form, we can consider other choices that force certain (maximum entropy) properties on the cascade, as we now explain.

1.7 MULTICHANNEL MAXIMUM ENTROPY CONSTRUCTION

The generalized scattering cascade $\Sigma(z)$ of Fig. 1.9 now maps $p \times q$ Schur-type matrix-valued loads to $p \times q$ Schur-type transfer matrices $S(z)$ at the input of the cascade – this is the transfer matrix from input-to-output at the left-side of the cascade. Also, we now have a system with multiple inputs (p of them) and multiple outputs (q of them).

For any such matrix Schur function $S(z)$, we define its spectral factor as the maximal outer function $L_s(z)$ such that

$$I - S(z)\left[S\left(\frac{1}{z^*}\right)\right]^* \geq L_s(z)\left[L_s\left(\frac{1}{z^*}\right)\right]^* .$$

Using the analogy with the single channel case (1.28), we introduce the following maximum entropy problem.

Problem 6 (Multichannel Entropy Problem) *Consider a positive-definite structured matrix R satisfying (1.32), and let $\Theta(z)$ and $\Sigma(z)$ denote the feedforward and feedback cascades obtained by the generalized Schur algorithm. Every Schur-type load now leads to a Schur-type transfer function $S(z)$ at the input of the cascade. The maximum entropy problem is to choose the load that results in*

$$\max_{\text{Schur load}} \left[\ln \det L_s(0)\right] . \tag{1.44}$$

The solution to the earlier maximum entropy problem in Alg. 4 (which corresponded to the special case $F = Z$ and $r = 2$), was obtained by terminating the corresponding cascade with the zero load. This solution does not work for general F. The following result is a special case of the main theorem we established in [12].

Algorithm 7 (Solution of Multichannel Version) *Given a positive-definite matrix R satisfying (1.32), let $\Theta(z)$ and $\Sigma(z)$ denote the feedforward and feedback cascades obtained by the generalized Schur algorithm (either by applying (1.34) or (1.40)). The optimal load is obtained by terminating the scattering cascade with the following (constant and contractive) load*

$$\text{Optimal Schur load} \;=\; [\Theta_{22}^{-1}\Theta_{21}]^*(0) . \tag{1.45}$$

In other words, the result requires that we determine $\Theta_{22}^{-1}(z)\Theta_{21}(z)$ and evaluate its value at $z = 0$. Then the complex conjugate of this value should be taken as the load (see Fig. 1.10).

This construction admits the following interpretation. Recall from expression (1.39) for the scattering cascade that the transfer matrix from the input to the output at the right-hand side is equal to $\Theta_{22}^{-1}(z)\Theta_{21}(z)$. The value at zero gives us the value of the direct (*i.e.*, delayless) path that exists between these two ports. Therefore, in a sense, the load $[\Theta_{22}^{-1}\Theta_{21}]^*(0)$ compensates for this direct path. To see this more clearly, let us denote this load by A^* for

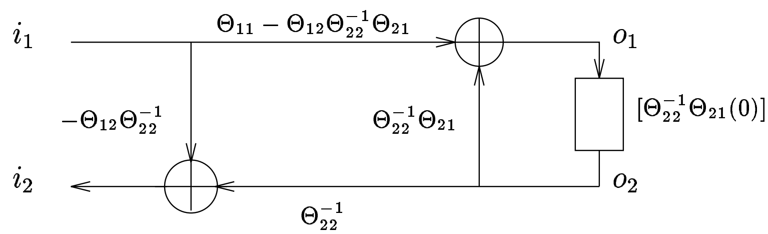

Figure 1.10 Maximum entropy load in the multichannel case.

simplicity of notation, $A^* = [\Theta_{22}^{-1}\Theta_{21}]^*(0)$. Then the gain of the direct path from o_2 to o_1 is obtained by combining the direct-path relations

$$o_2 = o_1 A^* \quad \text{and} \quad o_1 = o_2 A \ ,$$

which lead to $o_1 = o_1 A^* A$. But since A is strictly contractive, this equality holds only for $o_1 = 0$ and, consequently, $o_2 = 0$. That is, the response from o_2 to o_1 can only be strictly causal.

In the $F = Z$ case that we studied earlier (Alg. 4), the elementary sections of the scattering cascade obtained via the array form (1.20) had the special form (1.14). Hence, note that each such section has the following special form at $z = 0$,

$$\Theta_i(0) = \begin{bmatrix} 0 & \times \\ 0 & \times \end{bmatrix} .$$

If we now form $\Theta(z)$ by combining several such $\Theta_i(z)$, say $\Theta(z) = \Theta_0(z) \ldots \Theta_n(z)$, then $\Theta(z)$ at $z = 0$ will have a similar form,

$$\Theta(0) = \begin{bmatrix} 0 & \times \\ 0 & \times \end{bmatrix} .$$

Consequently, the $(2, 1)$ entry of $\Theta(z)$ will satisfy $\Theta_{21}(0) = 0$. This means that the resulting $\Theta_{21}(z)$ is strictly proper, which in turn implies that the resulting maximum-entropy load, according to the construction $[\Theta_{22}^{-1}\Theta_{21}]^*(0)$, will indeed be the zero load (as claimed earlier in the statement of Alg. 4.)

A similar conclusion does not hold for the array form (1.34) for general lower triangular F, and even for $F = Z$ with higher displacement rank (higher than 2). That is, the cascade $\Theta(z)$ that results in these cases from applying the array form (1.34) (or even the general form (1.40) without special care) will not be such that $\Theta_{21}(z)$ is strictly proper. In such cases, a nonzero load will be necessary and, consequently, the central solution will not be the maximum entropy solution.

To demonstrate this, consider the simple example of a diagonal F and displacement rank 2, *i.e.,* G in (1.32) has two columns and $J = (1 \oplus -1)$. Using the array form (1.34) of the generalized Schur algorithm we obtain a cascade of sections of the form

$$\Theta_i(z) = \Theta_i \begin{bmatrix} \frac{z-f_i}{1-f_i^* z} & 0 \\ 0 & 1 \end{bmatrix} ,$$

which implies that $\Theta_i(0)$ will in general have the form

$$\Theta_i(0) = \begin{bmatrix} \times & \times \\ \times & \times \end{bmatrix} .$$

Therefore, by combining several such sections the resulting $\Theta_{21}(z)$ will not generally be strictly proper.

In any case, the solution (1.45) to the multichannel maximum entropy problem requires that we first determine the cascade $\Theta(z)$ and then evaluate the load $[\Theta_{22}^{-1}\Theta_{21}]^*(0)$. It does not matter which form of the generalized Schur algorithm we use (the array form (1.34) or the more general form (1.40)). In either case, we should determine the optimal load via (1.45).

There is an alternative solution of the maximum entropy problem that relies instead on the general form (1.40). It employs a different choice for the arbitrary $J-$unitary rotations Θ_i in order to guarantee that the cascade $\Theta(z)$ will always satisfy $\Theta_{21}(0) = 0$ and, therefore, that the zero load will be sufficient to solve the maximum entropy problem. This alternative solution is therefore such that the central solution will always be the maximum entropy solution. It avoids the additional burden of determining the right load via (1.45) since all the calculations are embedded into the recursive construction of the cascade.

The new choice for the arbitrary rotation Θ_i is one that enforces that the $(2,1)$ entry of the elementary section $\Theta_i(z)$ in (1.42) is itself strictly proper. In so doing, when these sections are combined together to yield $\Theta(z)$, the $(2,1)$ entry $\Theta_{21}(z)$ of the overall cascade will be strictly proper and the required load will then be the zero load.

The rotation Θ_i in (1.40) is chosen as follows. Consider the general expression (1.42) for $\Theta_i(z)$ and partition g_i accordingly with J, say

$$g_i \triangleq \begin{bmatrix} u_i & v_i \end{bmatrix} ,$$

where u_i is $1 \times p$ and v_i is $1 \times q$. Let also $|\delta_i|^2 \triangleq g_i J g_i^*$. Then some straightforward algebra shows that the value of $\Theta_i(z)$ of (1.42) at $z = 0$ is given by

$$\Theta_i(0) = \begin{bmatrix} I_p - \frac{1+f_i}{|\delta_i|^2} u_i^* u_i & -\frac{1+f_i}{|\delta_i|^2} u_i^* v_i \\ \frac{1+f_i}{|\delta_i|^2} v_i^* u_i & I_q + \frac{1+f_i}{|\delta_i|^2} v_i^* v_i \end{bmatrix} \Theta_i .$$

This suggests that we could choose Θ_i in order to reduce the $r \times r$ matrix

$$\begin{bmatrix} I_p - \frac{1+f_i}{|\delta_i|^2} u_i^* u_i & -\frac{1+f_i}{|\delta_i|^2} u_i^* v_i \\ \frac{1+f_i}{|\delta_i|^2} v_i^* u_i & I_q + \frac{1+f_i}{|\delta_i|^2} v_i^* v_i \end{bmatrix}$$

into the reversed upper triangular form

$$\begin{bmatrix} \times & \times \\ 0_{q\times p} & \times \end{bmatrix}.$$

That is, we should choose a $J-$unitary Θ_i that performs the triangularization

$$\begin{bmatrix} I_p - \frac{1+f_i}{|\delta_i|^2} u_i^* u_i & -\frac{1+f_i}{|\delta_i|^2} u_i^* v_i \\ \frac{1+f_i}{|\delta_i|^2} v_i^* u_i & I_q + \frac{1+f_i}{|\delta_i|^2} v_i^* v_i \end{bmatrix} \Theta_i = \begin{bmatrix} \times & \times \\ 0_{q\times p} & \times \end{bmatrix}. \tag{1.46}$$

Once this is done for each $\Theta_i(z)$, the resulting $\Theta(z)$ will be such that $\Theta_{21}(z)$ is strictly proper. We showed in [12] that such rotations Θ_i exist by considering a family of generalized reflection coefficients.

Algorithm 8 (Recursive Solution of Multichannel Version) *Given a positive definite structured matrix R satisfying (1.32), apply the generalized Schur algorithm (1.40) where the Θ_i are chosen so as to enforce the transformation (1.46). Now terminate the resulting scattering cascade with the zero load and the Schur function mapped at the input of the cascade will be such that it solves (1.45).*

Using the construction (1.46) for the Θ_i in (1.43), we see that each section $\Theta_i(z)$ now has a state-space description of the form

$$\begin{bmatrix} l_i & \begin{matrix} 0 \\ G_{i+1} \end{matrix} \end{bmatrix} = \begin{bmatrix} F_i l_i & G_i \end{bmatrix} \begin{bmatrix} f_i^* & \frac{1}{d_i} g_i \Theta_i \\ J g_i^* & \begin{bmatrix} \times & \times \\ 0 & \times \end{bmatrix} \end{bmatrix}. \tag{1.47}$$

1.8 CONCLUDING REMARKS

Several classical moment and extension problems, and the famous solutions of Toeplitz, Caratheodory, and Schur, were used to introduce the maximum entropy problem for scalar-valued covariance sequences. A recursive construction in terms of the generalized Schur algorithm was then presented that allows us to solve the maximum entropy problem for matrix-valed covariance sequences and some more general sequences.

References

[1] T. Kailath and A. H. Sayed, Displacement structure: Theory and applications, *SIAM Review*, vol. 37, no. 3, pp. 297–386, September 1995.

[2] A. H. Sayed, *Displacement Structure in Signal Processing and Mathematics*. PhD thesis, Stanford University, Stanford, CA, August 1992.

[3] J. P. Burg, Maximum entropy spectral analysis, in *Proc. 37−th Meeting of Society of Exploration Geophysicists*, Oklahoma City, Okla., October 1967.

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

[5] H. Dym, *J-Contractive Matrix Functions, Reproducing Kernel Hilbert Spaces and Interpolation*, vol. 71, Conference Board of the Mathematical Sciences, American Mathematical Society, Providence, Rhode Island, 1989.

[6] N. I. Akhiezer, *The Classical Moment Problem and Some Related Questions in Analysis*, Hafner Publishing Company, NY, 1965 (in Russian, 1961).

[7] L. Y. Geronimus, *Orthogonal Polynomials*, Transl. Consultants Bureau, NY, 1961 (in Russian, 1958).

[8] G. Szego, *Orthogonal Polynomials*, vol. 23, American Mathematical Society Colloquium Publishing, Providence, RI, 1939.

[9] T. Kailath and A. H. Sayed, editors, *Fast Reliable Algorithms for Matrices with Structure*, SIAM, PA, 1999.

[10] 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, September 1994.

[11] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, 1991.

[12] T. Constantinescu, A. H. Sayed, and T. Kailath, Displacement structure and maximum entropy, *IEEE Trans. Information Theory*, vol. 43, no. 3, pp. 1074–1080, May 1997.