

A Maximum Entropy Solution of the Covariance Extension Problem for Reciprocal Processes

Francesca P. Carli, Augusto Ferrante, Michele Pavon, and Giorgio Picci, *Life Fellow, IEEE*

Abstract—Stationary reciprocal processes defined on a finite interval of the integer line can be seen as a special class of Markov random fields restricted to one dimension. Nonstationary reciprocal processes have been extensively studied in the past especially by Jamison *et al.* The specialization of the nonstationary theory to the stationary case, however, does not seem to have been pursued in sufficient depth in the literature. Stationary reciprocal processes (and reciprocal stochastic models) are potentially useful for describing signals which naturally live in a finite region of the time (or space) line. Estimation or identification of these models starting from observed data seems still to be an open problem which can lead to many interesting applications in signal and image processing. In this paper, we discuss a class of reciprocal processes which is the acausal analog of auto-regressive (AR) processes, familiar in control and signal processing. We show that maximum likelihood identification of these processes leads to a covariance extension problem for block-circulant covariance matrices. This generalizes the famous covariance band extension problem for stationary processes on the integer line. As in the usual stationary setting on the integer line, the covariance extension problem turns out to be a basic conceptual and practical step in solving the identification problem. We show that the maximum entropy principle leads to a complete solution of the problem.

Index Terms—Circulant matrices, covariance extension, covariance selection, maximum entropy, reciprocal processes.

I. INTRODUCTION

Reciprocal processes have been introduced at the beginning of the last century [2], [35], [36] even earlier than the idea of Markov process was formalized by Kolmogorov. The basic defining property is conditional independence given the values taken by the process at the boundary, which resembles a widely accepted definition of Markov random fields. When the “time” parameter is 1-D, reciprocal processes can in fact be seen as Markov random fields restricted to one dimension. For this reason, reciprocal processes are actually more general than Markov processes (a Markov process is reciprocal but not conversely). In fact, these processes naturally live in a finite region of the time (or space) variable and specification of *boundary values* at the extremes of the interval is an essential part of their probabilistic description. In discrete-time they are naturally defined on a finite interval of the integer line. Reciprocal

processes have been extensively studied in the past notably by Jamison, Krener, Levy and co-workers, see [15], [19]–[23], [26], [27]. However, the specialization of the nonstationary theory to the stationary case, except for a few noticeable exceptions, e.g. [19], [33], [34], does not seem to have been pursued in sufficient depth in the literature. Stationary reciprocal processes (and reciprocal stochastic models) are potentially useful for describing signals which naturally live in a finite region of the time or space line. They can be described by constant coefficient models which are a natural generalization of the Gauss-Markov state space models widely used in engineering and applied sciences. Estimation and identification of these models starting from observed data seems to be a completely open problem which can lead to many interesting applications in signal and image processing.

In this paper, after a general introduction to stationary processes defined on a finite interval (Section II), we discuss a class of reciprocal processes described by models which are the acausal analog of auto-regressive (AR) processes, familiar in control and signal processing (Section III). In Section IV we show that maximum likelihood identification of these processes leads to a covariance extension problem for block-circulant covariance matrices. This generalizes the famous covariance extension problem for stationary processes on the integer line. As in the usual stationary setting on the integer line, the covariance extension problem turns out to be a basic conceptual and practical step in solving the identification problem. The circulant covariance extension problem looks similar to a classical extension problem for positive block-Toeplitz matrices widely studied in the literature, [13], [17], which belongs to the class of *band extension* problems for positive matrices. All problems of this kind are solvable by factorization techniques. However, the banded algebra framework on which this literature relies does not apply to circulant matrices, see [5]. Circulant band extension appears to be a new kind of matrix extension problem.

In the present context, we are seeking a (reciprocal) AR extension. One may speculate that this extension should possess the analog of the so-called “maximum entropy” property, which holds for stationary processes on the line. In the literature, this property is usually presented as a final embellishment of the solution which is obtained by factorization techniques (typically computed via the Levinson–Whittle algorithm [24], [40]). In our case, where there are no factorization techniques at hand, we resort to maximum entropy as the main tool at our disposal to attack the problem. In Sections V and VI we show that the maximum entropy principle indeed leads to a complete solution of the problem. Finally in Section VII we discuss the relation with the covariance selection results in Dempster’s paper [11].

Manuscript received November 03, 2009; revised April 22, 2010 and September 09, 2010; accepted November 18, 2010. Date of publication March 10, 2011; date of current version September 08, 2011. Recommended by Associate Editor E. Weyer.

F. P. Carli, A. Ferrante, and G. Picci are with the Department of Information Engineering (DEI), University of Padova, Padova 35131, Italy (e-mail: carlifra@dei.unipd.it; augusto@dei.unipd.it; picci@dei.unipd.it).

M. Pavon is with the Department of Pure and Applied Mathematics, University of Padova, Padova 35131, Italy (e-mail: pavon@math.unipd.it).

Digital Object Identifier 10.1109/TAC.2011.2125050

Band extension problems for block-circulant matrices of the type discussed in this paper occur in particular in applications to image modeling and simulation. For reasons of space, we do not provide details but rather refer the reader to the literature, see e.g. [6], [7] and [32].

II. STATIONARY PROCESSES ON A FINITE INTERVAL

In this paper, we work in the wide-sense setting of second-order, zero-mean random variables. For the benefit of the reader, we recall here that a second-order random vector (or more generally process) is just an equivalence class consisting of all zero-mean random vectors (or processes), each defined on some canonical probability space, say the space of their sample values, that have the same covariance matrix, see e.g. [29, Chap. X]. Hence, each second-order random vector contains in particular a Gaussian element which may be taken as the representative of the equivalence class, [12, p. 74]. All statements of this paper do, therefore, apply to the particular case of Gaussian distributions. In our setting, however, explicit assumptions of Gaussianness will not be needed. We also recall that there is a basic correspondence, established by Kolmogorov in the early 1940's, between probabilistic concepts depending only on second-order moments and geometric operations on certain subspaces of the Hilbert space of finite variance random variables, see e.g. [12, p. 636–637] for historical remarks on this. We assume henceforth that the reader is familiar with this correspondence.

Orthogonality of two random vectors will be understood as componentwise uncorrelation, i.e. $\mathbf{x} \perp \mathbf{y}$ means $\mathbb{E}\mathbf{x}\mathbf{y}^\top = 0$. The symbol $\hat{\mathbb{E}}[\cdot]$ denotes orthogonal projection (conditional expectation in the Gaussian case) onto the subspace spanned by a family of finite variance random variables listed in the second argument.

A m -dimensional stochastic process on a finite interval $[1, N]$, is just an ordered collection of (zero-mean) random m -vectors $\mathbf{y} := \{\mathbf{y}(k), k = 1, 2, \dots, N\}$ which will be written as a column vector with N , m -dimensional components. We say that \mathbf{y} is *stationary* if the covariances $\mathbb{E}\mathbf{y}(k)\mathbf{y}(j)^\top$ depend only on the difference of the arguments, namely

$$\mathbb{E}\mathbf{y}(k)\mathbf{y}(j)^\top = \Sigma_{k-j}, \quad k, j = 1, \dots, N$$

in which case the covariance matrix of \mathbf{y} has a symmetric block-Toeplitz structure; i.e.,¹

$$\Sigma_N := \mathbb{E}\mathbf{y}\mathbf{y}^\top = \begin{bmatrix} \Sigma_0 & \Sigma_1^\top & \dots & \Sigma_{N-1}^\top \\ \Sigma_1 & \Sigma_0 & \Sigma_1^\top & \dots \\ \vdots & \ddots & \ddots & \ddots \\ \Sigma_{N-1} & \dots & \Sigma_1 & \Sigma_0 \end{bmatrix}. \quad (1)$$

Processes \mathbf{y} which have a positive definite covariance Σ_N are called of *full rank* (or *minimal*). In this paper, we shall usually deal with full rank processes.

¹Boldface capitals, e.g. \mathbf{I}_N , Σ_N , etc. denote block matrices made of N blocks, each of dimension $m \times m$.

Definition 2.1: A *block-circulant* matrix with N blocks, is a finite block-Toeplitz matrix whose block-columns (or equivalently, block-rows) are shifted cyclically. It looks like

$$\mathbf{C}_N = \begin{bmatrix} C_0 & C_{N-1} & \dots & \dots & C_1 \\ C_1 & C_0 & C_{N-1} & \dots & \dots \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & C_{N-1} \\ C_{N-1} & C_{N-2} & \dots & C_1 & C_0 \end{bmatrix}.$$

where $C_k \in \mathbb{R}^{m \times m}$. A block-circulant matrix \mathbf{C}_N is fully specified by its first block-column (or row). It will be denoted by

$$\mathbf{C}_N = \text{Circ}\{C_0, C_1, \dots, C_{N-1}\}. \quad (2)$$

For an introduction to circulant matrices, we refer the reader to the monograph [8]. Block-circulant matrices of a fixed size form a real vector space which is actually an algebra with respect to the usual operations of sum and matrix multiplication. The invertible elements of this algebra form a group.

Consider now a stationary process $\tilde{\mathbf{y}}$ on the integer line \mathbb{Z} , which is periodic of period T , i.e. a process satisfying $\tilde{\mathbf{y}}(k + nT) := \tilde{\mathbf{y}}(k)$ (almost surely) for all $n \in \mathbb{Z}$. We can think of $\tilde{\mathbf{y}}$ as a process indexed on the *discrete circle* group, $\mathbb{Z}_T \equiv \{1, 2, \dots, T\}$ with arithmetics mod T .² Clearly, its covariance function $\tilde{\Sigma}$ must also be periodic of period T , namely, $\tilde{\Sigma}_{k+T} = \tilde{\Sigma}_k$ for all $k \in \mathbb{Z}$. Hence, we may also see the covariance sequence as a function on the isomorphic discrete group $\tilde{\mathbb{Z}}_T \equiv \{0, T-1\}$ with arithmetics mod T . But more must be true.

Proposition 2.1: A (second-order) stochastic process \mathbf{y} on $[1, T]$ is the restriction to the interval $[1, T]$ of a wide-sense stationary periodic process $\tilde{\mathbf{y}}$ of period T defined on \mathbb{Z} , if and only if its covariance matrix Σ_T is symmetric block-circulant.

Proof: (only if) Let $k \in [1, T]$. By assumption there is an m -dimensional stationary process $\tilde{\mathbf{y}}$ on the integer line \mathbb{Z} , which is periodic of period T , satisfying $\tilde{\mathbf{y}}(k + nT) := \mathbf{y}(k)$ (almost surely) for arbitrary $n \in \mathbb{Z}$. By wide-sense stationarity, the covariance function of $\tilde{\mathbf{y}}$ must depend only on the difference of the arguments, namely

$$\tilde{\Sigma}_{k,j} := \mathbb{E}\tilde{\mathbf{y}}(k)\tilde{\mathbf{y}}(j)^\top = \tilde{\Sigma}_{k-j}, \quad k, j = 1, \dots, T.$$

Moreover, it is a well-known fact that, for any wide-sense stationary process the following symmetry relation holds

$$\tilde{\Sigma}_{-\tau} = \tilde{\Sigma}_\tau^\top \quad \forall \tau \in \mathbb{Z} \quad (3)$$

that is the covariance matrix of $\tilde{\mathbf{y}}$ has a symmetric block-Toeplitz structure. Now since $\tilde{\mathbf{y}}$ is periodic of period T , its covariance function must also be periodic of period T ; i.e. $\tilde{\Sigma}_{k+nT} = \tilde{\Sigma}_k$ for arbitrary $k, n \in \mathbb{Z}$. Assume, just to fix the ideas, that T is an even number and consider the midpoint

²Whence $T + \tau = \tau$ so that T plays the role of the zero element.

$k = T/2$ of the interval $[1, T]$. The periodicity combined with the symmetry property (3) yields that

$$\tilde{\Sigma}_{\frac{T}{2}+\tau} = \tilde{\Sigma}_{\frac{T}{2}+\tau-T} = \tilde{\Sigma}_{\tau-\frac{T}{2}} = \tilde{\Sigma}_{\frac{T}{2}-\tau}^T \quad \forall \tau \in \mathbb{Z} \quad (4)$$

and since (4) holds for $\tau = 0, 1, \dots, (T/2) - 1$, we can say that the function $\tilde{\Sigma}$ must be *symmetric with respect to the midpoint $\tau = T/2$ of the interval*. Hence, we can conclude that the covariance matrix of the process $\tilde{\mathbf{y}}$ restricted to $[1, T]$; that is the covariance Σ_T of \mathbf{y} , is a symmetric block-circulant matrix, i.e. it must have the following structure:

$$\Sigma_T = \begin{bmatrix} \tilde{\Sigma}_0 & \tilde{\Sigma}_1^T & \dots & \tilde{\Sigma}_\tau^T & \dots & \tilde{\Sigma}_\tau & \dots & \tilde{\Sigma}_1 \\ \tilde{\Sigma}_1 & \tilde{\Sigma}_0 & \tilde{\Sigma}_1^T & \ddots & \tilde{\Sigma}_\tau^T & \dots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & & & & \tilde{\Sigma}_\tau \\ \tilde{\Sigma}_\tau & \dots & \tilde{\Sigma}_1 & \tilde{\Sigma}_0 & \tilde{\Sigma}_1^T & \dots & \ddots & \\ \vdots & \tilde{\Sigma}_\tau & \dots & & \tilde{\Sigma}_0 & \dots & & \tilde{\Sigma}_\tau^T \\ \tilde{\Sigma}_\tau^T & & \ddots & & & & & \vdots \\ \vdots & \ddots & & \ddots & & \ddots & \ddots & \tilde{\Sigma}_1^T \\ \tilde{\Sigma}_1^T & \dots & \tilde{\Sigma}_\tau^T & \dots & \tilde{\Sigma}_\tau & & \tilde{\Sigma}_1 & \tilde{\Sigma}_0 \end{bmatrix}$$

which we write

$$\Sigma_T = \text{Circ} \left\{ \tilde{\Sigma}_0, \tilde{\Sigma}_1^T, \dots, \tilde{\Sigma}_\tau^T, \dots, \tilde{\Sigma}_{\frac{T}{2}}, \dots, \tilde{\Sigma}_\tau, \dots, \tilde{\Sigma}_1 \right\}.$$

Similarly, if T is odd, it must hold that $\tilde{\Sigma}_{((T+1)/2)+\tau} = \tilde{\Sigma}_{((T-1)/2)-\tau}^T$, $\tau = 0, 1, \dots, ((T-1)/2) - 1$ and Σ_T can be written as

$$\Sigma_T = \text{Circ} \left\{ \tilde{\Sigma}_0, \tilde{\Sigma}_1^T, \dots, \tilde{\Sigma}_\tau^T, \dots, \tilde{\Sigma}_{\frac{T-1}{2}}, \dots, \tilde{\Sigma}_{\frac{T-1}{2}}, \dots, \tilde{\Sigma}_\tau, \dots, \tilde{\Sigma}_1 \right\}$$

which proves the first part of the statement.

(if) We want to prove that if \mathbf{y} is a process defined on a finite interval $[1, T]$ with a symmetric block-circulant covariance matrix Σ_T , then it admits a wide-sense stationary periodic extension, $\tilde{\mathbf{y}}$, defined on \mathbb{Z} of period T .

Let $\tilde{\mathbf{y}}$ be the process obtained by periodically extending the process \mathbf{y} to the whole integer line \mathbb{Z} by setting $\tilde{\mathbf{y}}(k + nT) := \mathbf{y}(k)$ for arbitrary $n \in \mathbb{Z}$ and let us denote by $\tilde{\Sigma}$ its (infinite) covariance matrix. Since $\tilde{\Sigma}$ is a covariance matrix, it must be positive semidefinite. What we need to show is that it is a symmetric block-Toeplitz matrix. By definition, $\tilde{\Sigma}$ is the covariance matrix of the infinite column vector formed by stacking $\tilde{\mathbf{y}}(0), \tilde{\mathbf{y}}(1), \dots, \tilde{\mathbf{y}}(T), \dots, \tilde{\mathbf{y}}(nT), \dots$ in that order, it is formed by subblocks which replicate Σ_T to produce a square matrix of infinite size. Since Σ_T is symmetric block-circulant, then $\tilde{\Sigma}$ is, in particular, symmetric block-Toeplitz, which implies that $\tilde{\mathbf{y}}$ is stationary. This concludes the proof. ■

Remark 2.1: The periodic extension to the whole line \mathbb{Z} of deterministic signals originally given on a finite interval $[1, T]$ is a common device in (deterministic) signal processing. This simple periodic extension does, however, not preserve the structure of a stationary random process since the covariance of a periodically extended process will not be stationary unless the

covariance function of the original process on $[1, T]$ was center-symmetric to start with. This counter-intuitive fact has to do with the quadratic dependence of the covariance of the process on its random variables.

Let for example \mathbf{y} be a scalar process on the finite interval $[1, 4]$; i.e. let $T = 4$ and $m = 1$. Suppose \mathbf{y} has covariance matrix $\Sigma_T = \text{Toepl}\{\sigma_0, \sigma_1, \sigma_2, \sigma_3\}$, the notation $\text{Toepl}\{a\}$ meaning that Σ_T is a symmetric Toeplitz matrix with first column given by the vector a . The upper-left $2T \times 2T$ corner of the covariance of the periodic extension of \mathbf{y} is

$$\begin{bmatrix} \sigma_0 & \sigma_1 & \sigma_2 & \sigma_3 & \sigma_0 & \sigma_1 & \sigma_2 & \sigma_3 \\ \sigma_1 & \sigma_0 & \sigma_1 & \sigma_2 & \sigma_1 & \sigma_0 & \sigma_1 & \sigma_2 \\ \sigma_2 & \sigma_1 & \sigma_0 & \sigma_1 & \sigma_2 & \sigma_1 & \sigma_0 & \sigma_1 \\ \sigma_3 & \sigma_2 & \sigma_1 & \sigma_0 & \sigma_3 & \sigma_2 & \sigma_1 & \sigma_0 \\ \sigma_0 & \sigma_1 & \sigma_2 & \sigma_3 & \sigma_0 & \sigma_1 & \sigma_2 & \sigma_3 \\ \sigma_1 & \sigma_0 & \sigma_1 & \sigma_2 & \sigma_1 & \sigma_0 & \sigma_1 & \sigma_2 \\ \sigma_2 & \sigma_1 & \sigma_0 & \sigma_1 & \sigma_2 & \sigma_1 & \sigma_0 & \sigma_1 \\ \sigma_3 & \sigma_2 & \sigma_1 & \sigma_0 & \sigma_3 & \sigma_2 & \sigma_1 & \sigma_0 \end{bmatrix}.$$

This matrix is clearly not Toeplitz unless $\sigma_3 = \sigma_1$, in which case Σ_T would be symmetric circulant. Hence, the extended process $\tilde{\mathbf{y}}$ is in general not stationary.

Remark 2.2: In many applications to signal and image processing, the signals under study naturally live on a finite interval of the time (or space) variable and modeling them as functions defined on the whole line appears just as an artifice introduced in order to use the standard tools of (causal) time-invariant systems and harmonic analysis on the line. It may indeed be more logical to describe these data as stationary processes \mathbf{y} defined on a finite interval $[1, T]$. The covariance function, say Σ_T , of such a process will be a symmetric positive definite block-Toeplitz matrix which has in general no block-circulant structure.

It is, however, always possible to extend the covariance function of \mathbf{y} to a larger interval so as to make it center-symmetric. This can be achieved by simply letting $\Sigma_{T+\tau} := \Sigma_{T-1-\tau}^T$ for $\tau = 0, 1, \dots, T-1$. In this way Σ_T is extended to a symmetric block-circulant matrix $\tilde{\Sigma}_T$ of dimension $(2T-1) \times (2T-1)$, but this operation does not necessarily preserve positivity. Positivity of a symmetric, block-circulant extension, however, can always be guaranteed provided the extension is done on a suitably large interval. The details on how to construct such an extension are postponed to Section V, see the proof of Theorem 5.1. The original process \mathbf{y} can then be seen as the restriction to the interval $[1, T]$ of an extended process, say $\tilde{\mathbf{y}}$, which lives on an interval $[1, N]$ of length $N \geq 2T-1$. Since the extended covariance is, in any case, completely determined by the entries of the original covariance matrix Σ_T , any statistical estimate thereof can be computed from the variables of the original process \mathbf{y} in the interval $[1, T]$ (or from their sample values). Hence, there is no need to know what the random vectors $\{\tilde{\mathbf{y}}(k); k = T+1, \dots, N\}$ look like. Indeed, as soon as we are given the covariance of the process \mathbf{y} defined on $[1, T]$, even if we may not ever see (sample values of) the “external” random vectors $\{\tilde{\mathbf{y}}(k); k = T+1, \dots, N\}$, we would in any case have a completely determined second-order description (covariance function) of $\tilde{\mathbf{y}}$.

In this sense, one can think of any stationary process \mathbf{y} given on a finite interval $[1, T]$ as the restriction to $[1, T]$ of a wide-

sense stationary *periodic* process, $\tilde{\mathbf{y}}$, of period $N \geq 2T - 1$, defined on the whole integer line \mathbb{Z} . This process naturally lives on the “discrete circle” \mathbb{Z}_N . Hence, dealing in our future study with the periodic extension $\tilde{\mathbf{y}}$, instead of the original process \mathbf{y} , will entail no loss of generality. \square

III. AR-TYPE RECIPROCAL PROCESSES

In this section, we describe a class of random processes on a finite interval which are a natural generalization of the reciprocal processes introduced in [27], discussed in [26] and, for the stationary case, especially in [33], [34], see also [15]. In a sense, they are an acausal “symmetric” generalization of auto-regressive (AR) processes on the integer line.

Let \mathbf{y} be a zero-mean m -dimensional stationary process on $[1, N]$ and let Σ_N denote its $mN \times mN$ covariance matrix. We assume that Σ_N is a symmetric block-circulant matrix, so that \mathbf{y} may be seen as a process on the discrete circle \mathbb{Z}_N . In line with what argued in Remark 2.2, we may, if we wish so, imagine that the matrix Σ_N was obtained by extending a positive block-Toeplitz matrix as (1) to make it symmetric block-circulant. Then $[1, N]$ will have to be identified with an enlarged interval on which \mathbf{y} is the periodic extension of some underlying stationary process.

Let n be a natural number such that $N > 2n$. This inequality will be assumed to hold throughout. We introduce the notation $\mathbf{y}_{[t-n, t]}$ for the nm -dimensional random vector obtained by stacking $\mathbf{y}(t-n), \dots, \mathbf{y}(t-1)$ in that order. Similarly, $\mathbf{y}_{(t, t+n]}$ is the vector obtained by stacking $\mathbf{y}(t+1), \dots, \mathbf{y}(t+n)$ in that order. Likewise, the vector $\mathbf{y}_{[t-n, t]}$ is obtained by appending $\mathbf{y}(t)$ as last block to $\mathbf{y}_{[t-n, t)}$, etc.. The sums $t-k$ and $t+k$ are to be understood modulo N . Consider a subinterval $(t_1, t_2) \subset [1, N]$ where $(t_1, t_2) := \{t | t_1 < t < t_2\}$ and $(t_1, t_2)^c$ denotes the complementary set in $[1, N]$.

Let $\mathcal{A}, \mathcal{B}, \mathcal{C}$ be subspaces of zero-mean second-order random variables in a certain common ambient Hilbert space. Recall that \mathcal{A} and \mathcal{B} are said to be *conditionally orthogonal*, given \mathcal{C} if

$$(\mathbf{a} - \hat{\mathbb{E}}[\mathbf{a}|\mathcal{C}]) \perp (\mathbf{b} - \hat{\mathbb{E}}[\mathbf{b}|\mathcal{C}]), \quad \forall \mathbf{a} \in \mathcal{A}, \forall \mathbf{b} \in \mathcal{B}. \quad (5)$$

Conditional orthogonality is the same as conditional uncorrelatedness (and, hence, conditional independence) in the Gaussian case. Various equivalent forms of this condition are discussed in [28]. When $\mathcal{A}, \mathcal{B}, \mathcal{C}$ are generated by finite-dimensional random vectors, condition (5) can equivalently be rewritten in terms of the generating vectors, which we shall normally do in the following. The following definition does not require stationarity.

Definition 3.1: A reciprocal process of order n on $[1, N]$ is characterized by the property that the random variables of the process in the interval (t_1, t_2) are conditionally orthogonal to the random variables in the exterior, $(t_1, t_2)^c$, given the $2n$ boundary values $\mathbf{y}_{(t_1-n, t_1]}$ and $\mathbf{y}_{[t_2, t_2+n)}$. Equivalently, it must hold that

$$\begin{aligned} \hat{\mathbb{E}}[\mathbf{y}_{(t_1, t_2)} | \mathbf{y}(s), s \in (t_1, t_2)^c] \\ = \hat{\mathbb{E}}[\mathbf{y}_{(t_1, t_2)} | \mathbf{y}_{(t_1-n, t_1]} \vee \mathbf{y}_{[t_2, t_2+n)}] \end{aligned} \quad (6)$$

for $t_1, t_2 \in [1, N]$.

In particular, we should have

$$\hat{\mathbb{E}}[\mathbf{y}(t) | \mathbf{y}(s), s \neq t] = \hat{\mathbb{E}}[\mathbf{y}(t) | \mathbf{y}_{[t-n, t)} \vee \mathbf{y}_{(t, t+n)}] \quad (7)$$

for $t \in [1, N]$, where the estimation error

$$\begin{aligned} \mathbf{d}(t) &:= \mathbf{y}(t) - \hat{\mathbb{E}}[\mathbf{y}(t) | \mathbf{y}(s), s \neq t] \\ &= \mathbf{y}(t) - \hat{\mathbb{E}}[\mathbf{y}(t) | \mathbf{y}_{[t-n, t)} \vee \mathbf{y}_{(t, t+n)}], \quad t \in [1, N] \end{aligned} \quad (8)$$

must clearly be orthogonal to all random variables $\{\mathbf{y}(s), s \neq t\}$; i.e.,

$$\mathbb{E} \mathbf{y}(t) \mathbf{d}(s)^\top = \Delta \delta_{ts}, \quad t, s \in [1, N] \quad (9)$$

where δ is the Kronecker function and Δ is a square matrix. The actual meaning of Δ will be clarified a few lines below. In the spirit of Masani's definition [31], \mathbf{d} is called the (unnormalized) *conjugate process*³ of \mathbf{y} . Since $\mathbf{d}(t+k)$ is a linear combination of the components of the random vector $\mathbf{y}_{[t+k-n, t+k+n]}$, it follows from (9) that both $\mathbf{d}(t+k)$ and $\mathbf{d}(t-k)$ are orthogonal to $\mathbf{d}(t)$ as soon as $k > n$. Hence, the process $\{\mathbf{d}(t)\}$ has correlation bandwidth n ; i.e.,

$$\mathbb{E} \mathbf{d}(t+k) \mathbf{d}(t)^\top = 0 \quad \text{for } n < |k| < N-n, k \in [0, N-1]. \quad (10)$$

It follows from (8) that a reciprocal process of order n on $[1, N]$, can always be described by a linear double-sided recursion of the form

$$\sum_{k=-n}^n F_k \mathbf{y}(t-k) = \mathbf{d}(t), \quad t \in [1, N] \quad (11)$$

where the F_k 's are $m \times m$ matrices, in general dependent on t , with $F_0 = I_m$ and \mathbf{d} a process of correlation bandwidth n , orthogonal to \mathbf{y} in the sense of (9). In fact, it follows from (9) that $\mathbb{E} \mathbf{d}(t) \mathbf{d}(t)^\top = \Delta$, and hence, Δ is the variance matrix of $\mathbf{d}(t)$, symmetric and positive semidefinite.

Equation (11) requires the specification of boundary values, which will be described in Theorem 3.1 below.

Lemma 3.1: If \mathbf{y} is stationary, the matrices $\{F_k\}$ in the representation (11) do not depend on t . If \mathbf{y} is full rank, they are uniquely determined by the covariance lags of the process up to order $2n$.

Proof: The $\{F_k(t)\}$'s are determined by the orthogonality condition $\mathbf{d}(t) \perp \mathbf{y}_{[t-n, t)} \vee \mathbf{y}_{(t, t+n]}$, which can be expressed as

$$\begin{aligned} [F_{-n}(t) \quad \dots \quad F_{-1}(t) \quad F_1(t) \quad \dots \quad F_n(t)] \begin{bmatrix} \Sigma_n & \mathbf{Q}_n \\ \mathbf{Q}_n^\top & \Sigma_n \end{bmatrix} \\ = -[\Sigma_n^\top \quad \dots \quad \Sigma_1^\top \quad \Sigma_1 \quad \dots \quad \Sigma_n] \end{aligned} \quad (12)$$

where

$$\Sigma_n := \begin{bmatrix} \Sigma_0 & \Sigma_1 & \dots & \Sigma_{n-1} \\ \Sigma_1^\top & \Sigma_0 & \dots & \vdots \\ \vdots & \vdots & \ddots & \Sigma_1 \\ \Sigma_{n-1}^\top & \dots & \Sigma_1^\top & \Sigma_0 \end{bmatrix} \quad (13)$$

³Also called *double-sided innovation*.

and

$$\mathbf{Q}_n := \begin{bmatrix} \Sigma_{n+1} & \Sigma_{n+2} & \cdots & \Sigma_{2n} \\ \Sigma_n & \Sigma_{n+1} & \cdots & \Sigma_{2n-1} \\ \vdots & & \ddots & \vdots \\ \Sigma_2 & \cdots & \Sigma_n & \Sigma_{n+1} \end{bmatrix}. \quad (14)$$

Note that, because of stationarity, none of the covariance matrices depends on t . The determinant of the large block-matrix in (12) is a principal minor of order $2n$ of Σ_N . If \mathbf{y} is full rank, it must be nonzero and the matrix must be invertible. Therefore, the matrices $\{F_k\}$ do not depend on t and are uniquely determined. ■

For stationary reciprocal processes on \mathbb{Z}_N , the boundary-values to be attached to the linear model (11) are a straightforward consequence of the fact that \mathbf{y} has a stationary periodic extension to the whole axis \mathbb{Z} .

Theorem 3.1: A stationary reciprocal process, \mathbf{y} , of order n on \mathbb{Z}_N satisfies a linear, constant-coefficients difference equation of the type (11), associated to the $2n$ cyclic boundary conditions

$$\mathbf{y}(k) = \mathbf{y}(N + k); \quad k = -n + 1, \dots, n. \quad (15)$$

The model can be rewritten in matrix form as

$$\mathbf{F}_N \mathbf{y} = \mathbf{d} \quad (16)$$

where \mathbf{F}_N is the N -block banded circulant matrix of bandwidth n

$$\mathbf{F}_N := \text{Circ}\{I, F_1, \dots, F_n, 0, \dots, 0, F_{-n}, \dots, F_{-1}\}. \quad (17)$$

If the process is full rank this description is unique.

Proof: By definition

$$\hat{\mathbb{E}}[\mathbf{y}(1)|\mathbf{y}(s), s \neq 1] = \hat{\mathbb{E}}[\mathbf{y}(1)|\mathbf{y}_{[1-n,1]} \vee \mathbf{y}_{(1,1+n)}]$$

which is a linear function of $\mathbf{y}_{[1-n,1]} \vee \mathbf{y}_{(1,1+n)}$, whereby we can express $\mathbf{y}(1)$ as

$$\mathbf{y}(1) = -\tilde{F}_- \mathbf{y}_{(1,1+n)} - \tilde{F}_+ \mathbf{y}_{[1-n,1]} + \mathbf{d}(1)$$

for some coefficient matrices \tilde{F}_-, \tilde{F}_+ . The process \mathbf{y} has a periodic extension of period N , and hence, the missing initial boundary vector $\mathbf{y}_{[-n+1,1]}$ is actually the same as $\mathbf{y}_{[N-n+1,N]}$, so that

$$\mathbf{y}(1) = -\tilde{F}_- \mathbf{y}_{(1,1+n)} - \tilde{F}_+ \mathbf{y}_{[N-n+1,N]} + \mathbf{d}(1).$$

By stationarity, the various $m \times m$ blocks in the matrices \tilde{F} must satisfy the same system of (12) which was derived by imposing the orthogonality condition $\mathbf{d}(t) \perp \mathbf{y}_{[t-n,t]} \vee \mathbf{y}_{(t,t+n]}$, for all times t . Since the solution is unique, it must hold that $\tilde{F}_k = F_k$, $k = \pm n, \dots, \pm 1$ where the F_k 's are the same block matrices introduced before for (18). Hence, we have

$$\mathbf{y}(1) = - \sum_{k=-n}^{-1} F_k \mathbf{y}(1-k) - \sum_{k=1}^n F_k \mathbf{y}(N-k+1) + \mathbf{d}(1)$$

which is the first ($t = 1$) block equation in (11) once the first set of boundary conditions in (15) is used to replace the missing random variables $\mathbf{y}_{[1-n,1]}$. Similar expressions can be derived for $\mathbf{y}(2), \dots, \mathbf{y}(n)$ and for $\mathbf{y}(N-n), \dots, \mathbf{y}(N)$. From this it readily follows that \mathbf{y} satisfies (16) where \mathbf{F}_N has the banded circulant structure (17). ■

Using the notations F_- and F_+ for $[F_{-n} \dots F_{-1}]$ and $[F_1 \dots F_n]$ respectively, the error covariance $\Delta = \text{Var}\{\mathbf{d}(t)\}$ can be expressed as

$$\Delta = \Sigma_0 - [F_- \quad F_+] \begin{bmatrix} \Sigma_n & \mathbf{Q}_n \\ \mathbf{Q}_n^\top & \Sigma_n \end{bmatrix}^{-1} [F_- \quad F_+]^\top. \quad (18)$$

The following proposition is a simple generalization of analogous statements in [27], [34] for $n = 1$.

Proposition 3.1: A stationary reciprocal process \mathbf{y} is full rank if and only if the variance matrix Δ of the conjugate process is positive definite.

Proof: (if) Suppose $\Delta > 0$. Multiplying both members of (16) from the right by \mathbf{y}^\top and taking expectations, in virtue of the orthogonality relation (9), we get

$$\mathbf{F}_N \Sigma_N = \mathbf{F}_N \mathbb{E} \mathbf{y} \mathbf{y}^\top = \mathbb{E} \mathbf{d} \mathbf{y}^\top = \text{diag}\{\Delta, \dots, \Delta\}. \quad (19)$$

Thus, $\Delta > 0$ implies that the square matrices \mathbf{F}_N and Σ_N are invertible which, combined with the positive semidefiniteness of Σ_N , implies $\Sigma_N > 0$.

(only if) Suppose now that Δ is only positive semidefinite. This implies that there exists $0 \neq a \in \mathbb{R}^m$ s.t. $\mathbb{E} a^\top \mathbf{d}(t) \mathbf{d}(t)^\top a = 0$, i.e. s.t. $a^\top \mathbf{d}(t) = 0$ a.s.. This means that the scalar components of $\mathbf{d}(t)$ are linearly dependent, which, by (11), implies that $\mathbf{y}(t-n), \dots, \mathbf{y}(t), \dots, \mathbf{y}(t+n)$ are linearly dependent. Thus, Σ_N must be singular, which contradicts the assumption $\Sigma_N > 0$. ■

Solving (19) we can express the inverse as

$$\mathbf{M}_N := \Sigma_N^{-1} = \text{diag}\{\Delta^{-1}, \dots, \Delta^{-1}\} \mathbf{F}_N \quad (20)$$

so that \mathbf{M}_N is symmetric block-circulant and positive definite, being the inverse of a matrix with the same properties. Furthermore, $M_k := \Delta^{-1} F_k$, $k = -n, \dots, n$ and $M_0 = \Delta^{-1}$, must form a *center-symmetric sequence* of bandwidth n ; i.e.,⁴

$$M_{-k} = M_k^\top, \quad k = 1, \dots, n. \quad (21)$$

If we normalize the conjugate process by setting

$$\mathbf{e}(t) := \Delta^{-1} \mathbf{d}(t) \quad (22)$$

so that $\text{Var}\{\mathbf{e}(t)\} = \Delta^{-1}$, the model (11) can be rewritten

$$\sum_{k=-n}^n M_k \mathbf{y}(t-k) = \mathbf{e}(t), \quad t \in \mathbb{Z}_N \quad (23)$$

for which the orthogonality relation (9) is replaced by

$$\mathbb{E} \mathbf{y} \mathbf{e}^\top = \mathbf{I}_N. \quad (24)$$

⁴That is to say that model (11) is *self-adjoint*.

We shall now show that \mathbf{M}_N is actually the covariance matrix of the normalized conjugate process \mathbf{e} . For, by the normalization (22), our reciprocal process \mathbf{y} satisfies the linear equation

$$\mathbf{M}_N \mathbf{y} = \mathbf{e} \quad (25)$$

which implicitly includes the cyclic boundary conditions (15). Multiplying this from the right by \mathbf{e}^\top and taking expectations, we get $\mathbf{M}_N \mathbb{E}\{\mathbf{y}\mathbf{e}^\top\} = \mathbb{E}\{\mathbf{e}\mathbf{e}^\top\}$ which, in force of (24), yields

$$\text{Var}\{\mathbf{e}\} = \mathbf{M}_N \quad (26)$$

as announced. We see that *the inverse of the covariance matrix of a full rank stationary reciprocal process of order n , must be a banded block-circulant matrix of bandwidth n .*

This is in fact a fundamental characterization of stationary reciprocal processes of order n . To prove it, we need to take up the (inverse) question of *well-posedness*, namely if an autoregressive model of the form (11) associated to the proper cyclic boundary conditions, determines uniquely a process \mathbf{y} which is stationary and reciprocal of order n .

To this end we may just as well examine the equivalent normalized model (25).

Theorem 3.2: Consider a linear model (25) where \mathbf{M}_N is a symmetric positive-definite banded block-circulant matrix of bandwidth n and the process $\{\mathbf{e}(t); t \in \mathbb{Z}_N\}$ is a stationary process on \mathbb{Z}_N with covariance matrix \mathbf{M}_N .

Then there is a unique full rank stationary reciprocal process \mathbf{y} of order n , solution of (25). This process satisfies the orthogonality condition (24) and \mathbf{e} is its normalized conjugate process.

Proof: Pick a finitely correlated process \mathbf{e} with covariance matrix \mathbf{M}_N (we can construct such a, say Gaussian, process on a suitable probability space) and let \mathbf{y} be a solution of the (23) with boundary conditions (15), equivalently a solution of (25). Then, since \mathbf{M}_N is invertible, the process \mathbf{y} is uniquely defined on the interval $[1, N]$, i.e. there is a unique random vector, \mathbf{y} , solution of (25). Let Σ_N be its covariance matrix. We have, $\Sigma_N := \mathbb{E}\{\mathbf{y}\mathbf{y}^\top\} = \mathbb{E}[\mathbf{M}_N^{-1}\mathbf{e}\mathbf{e}^\top\mathbf{M}_N^{-1}] = \mathbf{M}_N^{-1}$, so that Σ_N is a symmetric positive-definite block-circulant matrix and the process \mathbf{y} is stationary on \mathbb{Z}_N (Proposition 2.1).

By multiplying (25) by \mathbf{e}^\top and taking expectations, we find $\mathbf{M}_N \mathbb{E}\{\mathbf{y}\mathbf{e}^\top\} = \mathbf{M}_N$, so that $\mathbb{E}\{\mathbf{y}\mathbf{e}^\top\} = \mathbf{I}_N$, or equivalently $\mathbb{E}\{\mathbf{y}(t)\mathbf{e}(s)^\top\} = \mathbf{I}_m \delta_{ts}$. Therefore, the orthogonality (24) holds on \mathbb{Z}_N .

Next, we need to show that \mathbf{y} is reciprocal of order n . To this end we shall generalize an argument of [34]. Let $s < t$ be two points in $[1, N]$, which for the moment we choose such that $t - n > s + n$, which is always possible since by assumption $N > 2n$. Expanding (23) and rearranging terms, we get (27), shown at the bottom of the page, which can be compactly rewritten as

$$\tilde{\mathbf{M}}\mathbf{y}_{[t,s]} = \mathbf{e}_{[t,s]} - \begin{bmatrix} \mathbf{N} & 0 \\ 0 & 0 \\ 0 & \mathbf{N}^\top \end{bmatrix} \begin{bmatrix} \mathbf{y}_{[t-n,t]} \\ \mathbf{y}_{[s,s+n]} \end{bmatrix} \quad (28)$$

$$\begin{bmatrix} M_0 & M_1^\top & \dots & M_n^\top & 0 & \dots & 0 \\ M_1 & M_0 & M_1^\top & \ddots & M_n^\top & 0 & \\ \vdots & & \ddots & & \ddots & & \\ M_n & \dots & M_1 & M_0 & M_1^\top & \dots & M_n^\top & 0 \\ 0 & M_n & \dots & M_0 & \dots & & \ddots & \\ \vdots & & \dots & & \ddots & & & \\ 0 & & \dots & & \dots & & & M_n^\top \\ & & & & \vdots & & & \\ 0 & & \ddots & & \ddots & & M_1 & M_0 & M_1^\top \\ 0 & & 0 & \dots & 0 & M_n & \dots & M_1 & M_0 \end{bmatrix} \begin{bmatrix} \mathbf{y}(t) \\ \mathbf{y}(t+1) \\ \vdots \\ \mathbf{y}(t+n) \\ \vdots \\ \mathbf{y}(s-n) \\ \vdots \\ \mathbf{y}(s-1) \\ \mathbf{y}(s) \end{bmatrix} \\ = \begin{bmatrix} \mathbf{e}(s) \\ \mathbf{e}(s+1) \\ \vdots \\ \mathbf{e}(s+n) \\ \vdots \\ \mathbf{e}(t-n) \\ \vdots \\ \mathbf{e}(t-1) \\ \mathbf{e}(t) \end{bmatrix} - \begin{bmatrix} M_n & \dots & M_1 & 0 & \dots & 0 \\ 0 & M_n & \dots & M_2 & 0 & \dots & 0 \\ 0 & \vdots & \ddots & & 0 & \dots & 0 \\ 0 & & \ddots & M_n & 0 & \dots & 0 \\ 0 & \dots & 0 & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & \dots & 0 & \dots & 0 \\ 0 & \dots & 0 & & M_n^\top & 0 & \\ 0 & \dots & 0 & \dots & \vdots & & 0 \\ 0 & \dots & 0 & M_2^\top & \ddots & & 0 \\ 0 & \dots & 0 & M_1^\top & \dots & M_n^\top \end{bmatrix} \begin{bmatrix} \mathbf{y}(t-n) \\ \mathbf{y}(t-n+1) \\ \vdots \\ \mathbf{y}(t-1) \\ \mathbf{y}(s+1) \\ \vdots \\ \mathbf{y}(s+n-1) \\ \mathbf{y}(s+n) \end{bmatrix} \quad (27)$$

with an obvious meaning of the symbols. Note that $\tilde{\mathbf{M}}$ is nonsingular, its determinant being a principal minor of \mathbf{M}_N , and hence, nonzero; while the two random vectors on the right hand side are uncorrelated since all scalar components of $\mathbf{e}_{[t,s]}$ are orthogonal to the linear subspace spanned by (the scalar components of) $\{\mathbf{y}(\tau); \tau \in [t, s]^c\}$ and hence are in particular orthogonal to the boundary condition vectors $\mathbf{y}_{(s,s+n]}, \mathbf{y}_{[t-n,t)}$. Solving (28) we can express $\mathbf{y}_{[t,s]}$ as a sum of two linear functions of $\mathbf{e}_{[t,s]}$ and of $\mathbf{y}_{(s,s+n]} \vee \mathbf{y}_{[t-n,t)}$ so that the orthogonal projection onto the linear subspace spanned by (the scalar components of) $\{\mathbf{y}(\tau); \tau \in [t, s]^c\}$ results in a linear function of (the scalar components of) $\mathbf{y}_{[t-n,t)} \vee \mathbf{y}_{(s,s+n]}$ alone. This proves the conditional orthogonality of $\mathbf{y}_{[t,s]}$ to the other random variables of the process, given the boundary values $\mathbf{y}_{[t-n,t)}, \mathbf{y}_{(s,s+n]}$.

The argument remains valid also when the nonoverlapping condition $t-n > s+n$ does not hold; i.e. for an arbitrary interval $[t, s]$ of the discrete circle \mathbb{Z}_N . For, when $[t-n, t)$ and $(s, s+n]$ overlap clearly we have $[t, s]^c \subseteq [t-n, t) \cup (s, s+n]$, and hence, all random variables in the subspace spanned by $\{\mathbf{y}(\tau); \tau \in [t, s]^c\}$ are contained in the subspace spanned by the boundary conditions, say $\mathcal{C} := \{\mathbf{y}(\tau); \tau \in [t-n, t) \cup (s, s+n]\}$. This means that $\hat{\mathbb{E}}[\mathbf{y}(\tau)|\mathcal{C}] = \mathbf{y}(\tau)$, or equivalently that

$$\mathbf{y}(\tau) - \hat{\mathbb{E}}[\mathbf{y}(\tau)|\mathcal{C}] = 0, \quad \tau \in [t, s]^c$$

so that the second member in (5) is zero, and hence, the orthogonality condition trivially holds. ■

From this result, we obtain the following fundamental characterization of reciprocal processes on the discrete group \mathbb{Z}_N .

Theorem 3.3: A nonsingular $mN \times mN$ -dimensional matrix Σ_N is the covariance matrix of a reciprocal process of order n on the discrete group \mathbb{Z}_N if and only if its inverse is a positive-definite symmetric block-circulant matrix which is banded of bandwidth n .

Note that the second-order statistics of both \mathbf{y} and \mathbf{e} are encapsulated in the covariance matrix \mathbf{M}_N . In other words, the whole auto-regressive model of \mathbf{y} is defined in terms of the matrix \mathbf{M}_N . Note also that this result makes the stochastic realization problem for reciprocal processes of order n conceptually trivial. In fact, given the covariance matrix Σ_N (the external description of the process), assuming that it is in fact the covariance matrix of such a process, the model matrix \mathbf{M}_N can be computed by simply inverting Σ_N . This is the simplest answer one could hope for. The solution requires, however, a preliminary criterion to check whether a (full rank) symmetric block-circulant covariance matrix has a banded inverse. There seems to be no simple known answer to this question.

Finally, to make contact with the literature, we note that a full rank reciprocal process of order n can always be represented as a linear memoryless function of a reciprocal process of order 1. This reciprocal process, however, need not be of full rank. To see that this is the case, introduce the vectors

$$\mathbf{y}_t^+ := \begin{bmatrix} \mathbf{y}(t) \\ \vdots \\ \mathbf{y}(t+n-1) \end{bmatrix}, \quad \mathbf{y}_t^- := \begin{bmatrix} \mathbf{y}(t-n+1) \\ \vdots \\ \mathbf{y}(t) \end{bmatrix}. \quad (29)$$

Letting $\mathbf{x}(t)^\top := [(\mathbf{y}_t^-)^\top (\mathbf{y}_t^+)^\top]$, we find the representation

$$\mathbf{x}(t) = \begin{bmatrix} F_+ & 0 \\ 0 & 0 \end{bmatrix} \mathbf{x}(t-1) + \begin{bmatrix} 0 & 0 \\ 0 & F_- \end{bmatrix} \mathbf{x}(t+1) + \tilde{\mathbf{d}}(t) \quad (30)$$

$$\mathbf{y}(t) = [0 \dots 0 \ 1 \ 1 \ 0 \dots 0] \mathbf{x}(t) \quad (31)$$

where F_- and F_+ are the block-companion matrices

$$F_+ := \begin{bmatrix} 0 & I & 0 & \dots & 0 \\ 0 & 0 & I & \dots & 0 \\ & \dots & & & I \\ -F_n & \dots & & & -F_1 \end{bmatrix}$$

$$F_- := \begin{bmatrix} -F_{-1} & \dots & & & -F_{-n} \\ I & 0 & \dots & & 0 \\ 0 & I & 0 & \dots & 0 \\ & \dots & & I & 0 \end{bmatrix}$$

and $\tilde{\mathbf{d}}(t) = (1/2)[0 \dots 0 \mathbf{d}(t)^\top \mathbf{d}(t)^\top 0 \dots 0]^\top$ has a singular covariance matrix. This model is in general nonminimal [34].

IV. IDENTIFICATION

Assume that T independent realizations of one period of the process \mathbf{y} are available⁵ and let us denote the string of sample values by $\underline{y} := (y^{(1)}, \dots, y^{(T)})$. We want to solve the following

Problem 4.1: Given the observations \underline{y} of a reciprocal process \mathbf{y} of (known) order n , estimate the parameters $\{M_k\}$ of the underlying reciprocal model $\mathbf{M}_N \mathbf{y} = \mathbf{e}$.

Note first that if we are given $2n+1$ covariance data $\{\Sigma_k; k = 0, 1, \dots, 2n\}$, the identification of an order n reciprocal process can be carried out by a linear algorithm, namely by solving the Yule-Walker-type system of linear equations (12).

This procedure is, however, unsatisfactory since, due to the symmetry (21), there are actually only $n+1$ unknown M_k to be computed. Hence, one would expect only $n+1$ covariance lags to be needed, while the system (12) requires solving also for the negative order coefficients. Moreover, in practice, the Σ_k 's will have to be estimated from observed data and estimates of covariances with a large lag k will unavoidably be more uncertain and have a larger variance.

In an attempt to get asymptotically efficient estimates for the M_k 's, we consider maximum likelihood estimation. To this end, we set up a Gaussian likelihood function (which does not require to assume that \mathbf{y} has a Gaussian distribution, see [18, p. 112]), which uses the density function

$$p_{(M_0, \dots, M_n)}(\mathbf{y}) = \frac{1}{\sqrt{(2\pi)^{mN} \det(\mathbf{M}_N^{-1})}} \exp\left(-\frac{1}{2} \mathbf{y}^\top \mathbf{M}_N \mathbf{y}\right)$$

where $\mathbf{y} \in \mathbb{R}^{mN}$. Taking logarithms and neglecting terms which do not depend on the parameters, one can rewrite this expression as

$$\log p_{(M_0, \dots, M_n)}(\underline{y}) = -\frac{1}{2} \log \det(\mathbf{M}_N^{-1}) - \frac{1}{2} \text{tr}\{\mathbf{M}_N \underline{y} \underline{y}^\top\}. \quad (32)$$

⁵For example, a "movie" consisting of T successive images of the same texture.

Assuming that the T sample measurements are independent, the log-likelihood function, depending on the $n + 1$ matrix parameters $\{M_k; k = 0, 1, \dots, n\}$, can be written

$$L(M_0, \dots, M_n) = \log \det(\mathbf{M}_N) - \sum_{k=0}^n \text{tr} \{M_k T_k(\underline{y})\} \quad (33)$$

where each matrix-valued statistic $T_k(\underline{y})$ has the structure of a sample estimate of the lag k covariance of the process. For example, T_0 and T_1 are given by

$$\begin{aligned} T_0(\underline{y}) &= \frac{1}{T} \sum_{t=1}^T \left\{ \sum_{k=0}^{N-1} y^{(t)}(k) \left[y^{(t)}(k) \right]^\top \right\} \\ T_1(\underline{y}) &= \frac{2}{T} \sum_{t=1}^T \left\{ \sum_{k=1}^{N-1} y^{(t)}(k-1) \left[y^{(t)}(k) \right]^\top \right\} \\ &\quad + \frac{2}{T} \sum_{t=1}^T y^{(t)}(N-1) \left[y^{(t)}(0) \right]^\top. \end{aligned}$$

From exponential class theory [1], we see that the T_k 's are (matrix-valued) sufficient statistics. Indeed, we have the well-known characterization that the (suitably normalized) statistics T_0, T_1, \dots, T_n are maximum likelihood estimators of their expected values, namely

$$\begin{aligned} \hat{\Sigma}_0 &:= \frac{1}{N} T_0 = \text{M.L. Estimator of } \mathbb{E} \mathbf{y}(k) \mathbf{y}(k)^\top \\ &\vdots \\ \hat{\Sigma}_n &:= \frac{1}{N} T_n = \text{M.L. Estimator of } \mathbb{E} \mathbf{y}(k+n) \mathbf{y}(k)^\top. \end{aligned} \quad (34)$$

Let us now consider the following matrix completion problem, which, from now on, will be referred to as the *block-circulant band extension problem*.

Problem 4.2: (Block-Circulant Band Extension Problem): Given $n + 1$ initial data $m \times m$ matrices $\hat{\Sigma}_0, \dots, \hat{\Sigma}_n$, complete them with a sequence $\Sigma_{n+1}, \Sigma_{n+2}, \dots, \Sigma_{N-1}$, in such a way to form a positive definite symmetric block-circulant matrix Σ_N with a *block-circulant banded inverse of bandwidth n* .

Note that the model parameters (M_0, M_1, \dots, M_n) are the nonzero blocks of the (banded) inverse of the covariance matrix Σ_N of the process (Theorem 3.3). The invariance principle for maximum likelihood estimators [42] leads then to the following statement.

Theorem 4.1: The maximum likelihood estimates of (M_0, M_1, \dots, M_n) are the nonzero blocks of the banded inverse of the matrix $\hat{\Sigma}_N$ solving the block-circulant band extension problem with initial data the $n + 1$ covariance estimates (34).

Hence, solving the original identification problem 4.1 has been shown to lead to the solution of a block-circulant band extension problem. Note, however, that the extension problem 4.2 is nonlinear and it is hard to see what is going on by elementary means. Below we give a scalar example.

Example 4.1: Let $m = 1$, $N = 8$, $n = 2$ and assume we are given the covariance estimates $\hat{\sigma}_0, \hat{\sigma}_1, \hat{\sigma}_2$, forming a positive definite Toeplitz matrix. The three unknown coefficients in the reciprocal model (23) of order 2 are scalars, denoted

m_0, m_1, m_2 . Multiplying (25) from the right by \mathbf{y}^\top , we get $\mathbf{M}_N \Sigma_N = \mathbf{I}_N$, which leads to

$$\begin{bmatrix} m_0 & m_1 & m_2 & 0 & 0 & 0 & m_2 & m_1 \\ m_1 & m_0 & m_1 & m_2 & 0 & 0 & 0 & m_2 \\ m_2 & m_1 & m_0 & m_1 & m_2 & 0 & 0 & 0 \\ 0 & m_2 & m_1 & m_0 & m_1 & m_2 & 0 & 0 \\ 0 & 0 & m_2 & m_1 & m_0 & m_1 & m_2 & 0 \\ 0 & 0 & 0 & m_2 & m_1 & m_0 & m_1 & m_2 \\ m_2 & 0 & 0 & 0 & m_2 & m_1 & m_0 & m_1 \\ m_1 & m_2 & 0 & 0 & 0 & m_2 & m_1 & m_0 \end{bmatrix} \begin{bmatrix} \hat{\sigma}_0 \\ \hat{\sigma}_1 \\ \hat{\sigma}_2 \\ x_3 \\ x_4 \\ x_3 \\ \hat{\sigma}_2 \\ \hat{\sigma}_1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

where $x_3 := \sigma_3 = \sigma_5$ and $x_4 := \sigma_4$ are the unknown extended covariance lags. Rearranging and eliminating the last three redundant equations, one obtains

$$\begin{aligned} m_0 \hat{\sigma}_0 + 2m_1 \hat{\sigma}_1 + 2m_2 \hat{\sigma}_2 &= 1 \\ m_0 \hat{\sigma}_1 + m_1 (\hat{\sigma}_0 + \hat{\sigma}_2) + m_2 (\hat{\sigma}_1 + x_3) &= 0 \\ m_0 \hat{\sigma}_2 + m_1 (\hat{\sigma}_1 + x_3) + m_2 (\hat{\sigma}_0 + x_4) &= 0 \\ m_0 x_3 + m_1 (\hat{\sigma}_2 + x_4) + m_2 (\hat{\sigma}_1 + x_3) &= 0 \\ m_0 x_4 + 2m_1 x_3 + 2m_2 \hat{\sigma}_2 &= 0 \end{aligned}$$

which is a system of five quadratic equations in five unknowns whose solution already looks nontrivial. It may be checked that, under positivity of the matrix $\text{Toep}\{\hat{\sigma}_0, \hat{\sigma}_1, \hat{\sigma}_2\}$, it has a unique positive definite solution (i.e. making \mathbf{M}_N positive definite).

At first sight the circulant band extension problem of Theorem 4.2 recalls the classical band extension problems for Toeplitz matrices studied in [13], [17], which is solvable by factorization techniques. However, the banded algebra framework on which these papers rely does not apply here. The circulant band extension problem seems to be a new (and harder) extension problem.

General covariance extension problems are discussed in an illuminating paper by A. P. Dempster, [11]. Notice, however, that Dempster's procedures, having been conceived to solve a general covariance extension problem, do not exploit the circulant structure of the present setting and are computationally very intensive even for small scalar instances. A possible approximate approach to the circulant band extension problem was proposed in [6]. This approach, based on a result of B. Levy [25], exploits the fact that for $N \rightarrow \infty$ the problem becomes one of band extension for infinite positive definite symmetric block-Toeplitz matrices, for which satisfactory algorithms exist. For N finite, however, this approximation may in some cases turn out to be poor. In the next section, we propose a new approach to the circulant band extension problem.

V. MAXIMUM ENTROPY ON THE DISCRETE CIRCLE

Dempster's paper, which deals with general, unstructured covariance matrices, only considers Gaussian distributions. He solves the following extension problem: characterize, among all covariance matrices sharing a given set of entries, the one corresponding to the (zero-mean) maximum entropy Gaussian distribution. For our purposes, a key observation is Statement (b) in [11, p. 160]. In our setting, it reads as follows.

Proposition 5.1: Assume feasibility of the covariance extension problem. Among all covariance extensions of the data

$\hat{\Sigma}_0, \dots, \hat{\Sigma}_n$, there exists a unique such an extension whose inverse's entries are zero in all the positions complementary to those where the elements of the covariance are assigned. This extension corresponds to the Gaussian distribution with maximum entropy.

This principle of entropy maximization will lead us to a new convex optimization procedure for computing the band extension.

We hasten to remark that in this paper we are *not* restricting ourselves to the case of Gaussian distributions. We shall consider Σ_N to be the matrix variance of a Gaussian distribution only for the purpose of *interpreting* the following optimization problem in the light of Dempster's result. The far reaching implications of our maximum entropy principle for general probability distributions is provided in Theorem 7.2 below.

Notations: Let \mathbf{U}_N denote the block-circulant “shift” matrix with $N \times N$ blocks,

$$\mathbf{U}_N = \begin{bmatrix} 0 & I_m & 0 & \dots & 0 \\ 0 & 0 & I_m & \dots & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & 0 & \dots & I_m \\ I_m & 0 & 0 & \dots & 0 \end{bmatrix}$$

where I_m denotes the $m \times m$ identity matrix. Clearly, $\mathbf{U}_N^\top \mathbf{U}_N = \mathbf{U}_N \mathbf{U}_N^\top = I_{mN}$; i.e. \mathbf{U}_N is orthogonal. Note that a matrix C with $N \times N$ blocks is block-circulant if and only if it commutes with \mathbf{U}_N , namely if and only if it satisfies

$$\mathbf{U}_N^\top C \mathbf{U}_N = C. \quad (35)$$

Recall that the *differential entropy* $H(p)$ of a probability density function p on \mathbb{R}^n is defined by

$$H(p) = - \int_{\mathbb{R}^n} \log(p(x)) p(x) dx. \quad (36)$$

In case of a zero-mean Gaussian distribution p with covariance matrix Σ_N , we get

$$H(p) = \frac{1}{2} \log(\det \Sigma_N) + \frac{1}{2} n (1 + \log(2\pi)). \quad (37)$$

Let \mathfrak{S}_N denote the vector space of *symmetric* matrices with $N \times N$ square blocks of dimension $m \times m$. Let $\mathbf{T}_n \in \mathfrak{S}_{n+1}$ denote the Toeplitz matrix of *boundary data*

$$\mathbf{T}_n = \begin{bmatrix} \Sigma_0 & \Sigma_1^\top & \dots & \Sigma_n^\top \\ \Sigma_1 & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \Sigma_n & \dots & \dots & \Sigma_0 \end{bmatrix} \quad (38)$$

and let E_n denote the $N \times (n+1)$ block matrix

$$E_n = \begin{bmatrix} I_m & 0 & \dots & 0 \\ 0 & I_m & \dots & 0 \\ 0 & 0 & \ddots & \vdots \\ \vdots & & 0 & I_m \\ 0 & 0 & \dots & 0 \end{bmatrix}.$$

Maximum Entropy Problem on \mathbb{Z}_N : Consider the following Gaussian *maximum entropy* problem (MEP) on the discrete circle:

Problem 5.1:

$$\min \{-\text{tr} \log \Sigma_N | \Sigma_N \in \mathfrak{S}_N, \Sigma_N > 0\} \quad (39a)$$

subject to :

$$E_n^\top \Sigma_N E_n = \mathbf{T}_n, \quad (39b)$$

$$\mathbf{U}_N^\top \Sigma_N \mathbf{U}_N = \Sigma_N. \quad (39c)$$

Recalling that $\text{tr} \log \Sigma_N = \log \det \Sigma_N$ and (37), we see that the above problem indeed amounts to finding the maximum entropy Gaussian distribution with a block-circulant covariance, whose first $n+1$ blocks are precisely $\Sigma_0, \dots, \Sigma_n$. The circulant structure is equivalent to requiring this distribution to be stationary on the discrete circle \mathbb{Z}_N . We observe that in this problem we are minimizing a strictly convex function on the intersection of a convex cone (minus the zero matrix) with a linear manifold. Hence, we are dealing with a convex optimization problem.

Note that we are not imposing that the inverse of the solution Σ_N of Problem 5.1 should have a banded structure. We shall see that, whenever solutions exist, this property will be *automatically guaranteed*.

The first question to be addressed is feasibility of (MEP), namely the existence of a positive definite, symmetric matrix Σ_N satisfying (39b), (39c). Obviously, \mathbf{T}_n positive definite is a necessary condition for the existence of such a Σ_N . In general it turns out that, under such a necessary condition, feasibility holds for N large enough. The idea is that for $N \rightarrow \infty$, Toeplitz matrices can be approximated arbitrarily well by circulants [30], [39], and hence, existence of a positive block-circulant extension can be derived from the existence of positive extensions for Toeplitz matrices.

Theorem 5.1: Given the sequence $\Sigma_i \in \mathbb{R}^{m \times m}$, $i = 0, 1, \dots, n$, such that

$$\mathbf{T}_n = \mathbf{T}_n^\top > 0 \quad (40)$$

there exists \bar{N} such that for $N \geq \bar{N}$, the matrix \mathbf{T}_n can be extended to an $N \times N$ block-circulant, positive-definite symmetric matrix Σ_N .

Proof: A fundamental result in stochastic system theory is the so-called maximum entropy covariance extension. It states that, under condition (40), there exists a rational positive real function $\Phi_+(z) = (\Sigma_0/2) + C(zI - A)^{-1}B$ such that:

- 1) A has spectrum strictly inside the unit circle;
- 2) $\Sigma_i = CA^{i-1}B$, $i = 1, 2, \dots, n$;
- 3) the spectrum $\Phi(z) := \Phi_+(z) + \Phi_+^*(z)$ is coercive, i.e.,⁶

$$\exists \varepsilon > 0 \text{ such that } \Phi(e^{j\vartheta}) > \varepsilon I, \quad \forall \vartheta \in [0, 2\pi). \quad (41)$$

In fact, $\Phi(z)$ has no zeros on the unit circle since it can be expressed in the form $\Phi(z) = L_n(z^{-1})^{-1} \Lambda_n L_n(z)^{-\top}$ where $L_n(z^{-1})$ is the n th Levinson–Whittle matrix polynomial (also called n th matrix Szegő polynomial) of the block Toeplitz matrix \mathbf{T}_n , and $\Lambda_n = \Lambda_n^\top > 0$; see [40], [9] and [41].

⁶ Here, and in the following, j denotes the imaginary unit $\sqrt{-1}$.

Let $\Sigma_i := CA^{i-1}B$, $i = n+1, n+2, \dots$, so that $\Phi_+(z) = (\Sigma_0/2) + \sum_{i=1}^{infy} \Sigma_i z^{-i}$, and define

$$\Sigma_N := \text{Circ}\left(\Sigma_0, \Sigma_1^\top, \Sigma_2^\top, \dots, \Sigma_{\frac{N-1}{2}}^\top, \Sigma_{\frac{N-1}{2}}, \Sigma_{\frac{N-1}{2}-1}, \dots, \Sigma_1\right) \quad (42)$$

for N odd, and

$$\Sigma_N := \text{Circ}\left(\Sigma_0, \Sigma_1^\top, \Sigma_2^\top, \dots, \Sigma_{\frac{N-2}{2}}^\top, \Sigma_{\frac{N}{2}}^\top + \Sigma_{\frac{N}{2}}, \Sigma_{\frac{N-2}{2}}, \Sigma_{\frac{N-2}{2}-1}, \dots, \Sigma_1\right) \quad (43)$$

for N even. We need to show that there exists \tilde{N} such that $\Sigma_N > 0$ for $N \geq \tilde{N}$. To this aim, notice that Σ_N is, by definition, block-circulant so that, a similarity transformation induced by a unitary matrix \mathbf{V} reduces Σ_N to a block-diagonal matrix

$$\mathbf{V}^* \Sigma_N \mathbf{V} = \Psi_N := \text{diag}(\Psi_0, \Psi_1, \dots, \Psi_{N-1})$$

where \mathbf{V} is the Fourier block-matrix whose k, l th block is

$$V_{kl} = 1/\sqrt{N} \exp[-j2\pi(k-1)(l-1)/N] I_m$$

and Ψ_ℓ are the coefficients of the finite Fourier transform of the first block row of Σ_N

$$\Psi_\ell = \Sigma_0 + e^{j\vartheta_\ell} \Sigma_1^\top + (e^{j\vartheta_\ell})^2 \Sigma_2^\top + \dots \\ \dots + (e^{j\vartheta_\ell})^{N-2} \Sigma_2 + (e^{j\vartheta_\ell})^{N-1} \Sigma_1, \quad (44)$$

with $\vartheta_\ell := -2\pi\ell/N$, see e.g. [38, Sec. 3.4]. Clearly, $(e^{j\vartheta_\ell})^{N-i} = (e^{j\vartheta_\ell})^{-i}$ and hence

$$\Psi_\ell = \Phi(e^{j\vartheta_\ell}) - [\delta\Phi_N(e^{j\vartheta_\ell}) + \delta\Phi_N^*(e^{j\vartheta_\ell})] \quad (45)$$

where

$$\delta\Phi_N(z) := \sum_{i=h+1}^{\infty} \Sigma_i z^{-i} = \sum_{i=h+1}^{\infty} CA^{i-1}Bz^{-i} \\ = z^{-h}CA^h(zI - A)^{-1}B$$

with

$$h := \begin{cases} \frac{N-1}{2}, & N \text{ odd} \\ N/2, & N \text{ even.} \end{cases}$$

Since A is a stability matrix, if N , and hence, h , is large enough, $\delta\Phi_N(e^{j\vartheta_\ell}) + \delta\Phi_N^*(e^{j\vartheta_\ell})$ is dominated by εI , i.e. there exists \tilde{N} such that

$$\delta\Phi_N(e^{j\vartheta_\ell}) + \delta\Phi_N^*(e^{j\vartheta_\ell}) < \varepsilon I, \quad \forall \vartheta_\ell, \quad \forall N \geq \tilde{N} \quad (46)$$

so that it readily follows from (41) and (45) that if $N \geq \tilde{N}$, $\Psi_\ell > 0$ for all ℓ . ■

We observe that, given \mathbf{T}_n , the triple A, B, C can be explicitly computed so that we can compute ε and \tilde{N} for which (46) holds. In other words, Theorem 5.1 provides a sufficient condition that can be practically tested. Similar bounds, but valid only for the scalar case, were derived in [10].

VI. VARIATIONAL ANALYSIS

We shall introduce a suitable set of “Lagrange multipliers” for our constrained optimization problem. Consider the linear map $A : \mathfrak{S}_{n+1} \times \mathfrak{S}_N \rightarrow \mathfrak{S}_N$ defined by

$$A(\Lambda, \Theta) = E_n \Lambda E_n^\top + \mathbf{U}_N \Theta \mathbf{U}_N^\top - \Theta, \quad (\Lambda, \Theta) \in \mathfrak{S}_{n+1} \times \mathfrak{S}_N$$

and define the set

$$\mathcal{L}_+ := \left\{ (\Lambda, \Theta) \in (\mathfrak{S}_{n+1} \times \mathfrak{S}_N) \mid (\Lambda, \Theta) \in (\ker(A))^\perp \right. \\ \left. (E_n \Lambda E_n^\top + \mathbf{U}_N \Theta \mathbf{U}_N^\top - \Theta) > 0 \right\}.$$

Observe that \mathcal{L}_+ is an open, convex subset of $(\ker(A))^\perp$. For each $(\Lambda, \Theta) \in \mathcal{L}_+$, we consider the unconstrained minimization of the *Lagrangian function*

$$L(\Sigma_N, \Lambda, \Theta) := -\text{tr} \log \Sigma_N + \text{tr}(\Lambda(E_n^\top \Sigma_N E_n - \mathbf{T}_n)) \\ + \text{tr}(\Theta(\mathbf{U}_N^\top \Sigma_N \mathbf{U}_N - \Sigma_N)) \\ = -\text{tr} \log \Sigma_N + \text{tr}(E_n \Lambda E_n^\top \Sigma_N) - \text{tr}(\Lambda \mathbf{T}_n) \\ + \text{tr}(\mathbf{U}_N \Theta \mathbf{U}_N^\top \Sigma_N) - \text{tr}(\Theta \Sigma_N)$$

over $\mathfrak{S}_{N,+} := \{\Sigma_N \in \mathfrak{S}_N, \Sigma_N > 0\}$. For $\delta \Sigma_N \in \mathfrak{S}_N$, we get

$$\delta L(\Sigma_N, \Lambda, \Theta; \delta \Sigma_N) = -\text{tr}(\Sigma_N^{-1} \delta \Sigma_N) + \text{tr}(E_n \Lambda E_n^\top \delta \Sigma_N) \\ + \text{tr}((\mathbf{U}_N \Theta \mathbf{U}_N^\top - \Theta) \delta \Sigma_N).$$

We conclude that $\delta L(\Sigma_N, \Lambda, \Theta; \delta \Sigma_N) = 0$, $\forall \delta \Sigma_N \in \mathfrak{S}_N$ if and only if

$$\Sigma_N^{-1} = E_n \Lambda E_n^\top + \mathbf{U}_N \Theta \mathbf{U}_N^\top - \Theta.$$

Thus, for each fixed pair $(\Lambda, \Theta) \in \mathcal{L}_+$, the unique Σ_N^o minimizing the Lagrangian is given by

$$\Sigma_N^o = (E_n \Lambda E_n^\top + \mathbf{U}_N \Theta \mathbf{U}_N^\top - \Theta)^{-1}. \quad (47)$$

Consider next $L(\Sigma_N^o, \Lambda, \Theta)$. We get

$$L(\Sigma_N^o, \Lambda, \Theta) = -\text{tr} \log \left((E_n \Lambda E_n^\top + \mathbf{U}_N \Theta \mathbf{U}_N^\top - \Theta)^{-1} \right) \\ + \text{tr} \left[(E_n \Lambda E_n^\top + \mathbf{U}_N \Theta \mathbf{U}_N^\top - \Theta) \right. \\ \left. \times (E_n \Lambda E_n^\top + \mathbf{U}_N \Theta \mathbf{U}_N^\top - \Theta)^{-1} \right] \\ - \text{tr}(\Lambda \mathbf{T}_n) \\ = \text{tr} \log (E_n \Lambda E_n^\top + \mathbf{U}_N \Theta \mathbf{U}_N^\top - \Theta) \\ + \text{tr} I_{mN} - \text{tr}(\Lambda \mathbf{T}_n).$$

This is a strictly concave function on \mathcal{L}_+ whose maximization is the *dual problem* of (MEP). We can equivalently consider the convex problem

$$\min \{J(\Lambda, \Theta), (\Lambda, \Theta) \in \mathcal{L}_+\} \quad (48)$$

where J (henceforth called dual function) is given by

$$J(\Lambda, \Theta) = \text{tr}(\Lambda \mathbf{T}_n) - \text{tr} \log (E_n \Lambda E_n^\top + \mathbf{U}_N \Theta \mathbf{U}_N^\top - \Theta). \quad (49)$$

Existence for the Dual Problem: The minimization of the strictly convex function $J(\Lambda, \Theta)$ on the convex set \mathcal{L}_+ is a challenging problem as \mathcal{L}_+ is an *open* and *unbounded* subset of $(\ker(A))^\perp$. Nevertheless, the following existence result in the Byrnes-Lindquist spirit, [3], [14], [16] can be established.

Theorem 6.1: The function J admits a unique minimum point $(\bar{\Lambda}, \bar{\Theta})$ in \mathcal{L}_+ .

In order to prove this theorem, we need first to derive a number of auxiliary results. Let \mathfrak{C}_N denote the vector subspace of block-circulant matrices in \mathfrak{S}_N . We proceed to characterize the orthogonal complement of \mathfrak{C}_N in \mathfrak{S}_N .

Lemma 6.1: Let $M \in \mathfrak{S}_N$. Then $M \in (\mathfrak{C}_N)^\perp$ if and only if it can be expressed as

$$M = \mathbf{U}_N N \mathbf{U}_N^\top - N \quad (50)$$

for some $N \in \mathfrak{S}_N$.

Proof: By (35), \mathfrak{C}_N is the kernel of the linear map from \mathfrak{S}_N to \mathfrak{S}_N given by $M \mapsto \mathbf{U}_N^\top M \mathbf{U}_N - M$. Hence, its orthogonal complement is the range of the adjoint map. Since

$$\begin{aligned} \text{tr}((\mathbf{U}_N^\top M \mathbf{U}_N - M)N) &= \langle \mathbf{U}_N^\top M \mathbf{U}_N - M, N \rangle \\ &= \langle M, \mathbf{U}_N N \mathbf{U}_N^\top - N \rangle \end{aligned}$$

the conclusion follows. \blacksquare

Next we show that, as expected, feasibility of the primal problem (MEP) implies that the dual function J is bounded below.

Lemma 6.2: Assume that there exists $\bar{\Sigma}_N \in \mathfrak{S}_{N,+}$ satisfying (39b), (39c). Then, for any pair $(\Lambda, \Theta) \in \mathcal{L}_+$, we have

$$J(\Lambda, \Theta) \geq mN + \text{tr} \log \bar{\Sigma}_N. \quad (51)$$

Proof: By (39b), we have $\text{tr}(\Lambda \mathbf{T}_n) = \text{tr}(\Lambda E_n^\top \bar{\Sigma}_N E_n) = \text{tr}(E_n \Lambda E_n^\top \bar{\Sigma}_N)$. Using this fact and Lemma 6.1, we can now rewrite the dual function J , as follows:

$$\begin{aligned} J(\Lambda, \Theta) &= \text{tr}(\Lambda \mathbf{T}_n) - \text{tr} \log (E_n \Lambda E_n^\top + \mathbf{U}_N \Theta \mathbf{U}_N^\top - \Theta) \\ &= \text{tr} [(E_n \Lambda E_n^\top + \mathbf{U}_N \Theta \mathbf{U}_N^\top - \Theta) \bar{\Sigma}_N] \\ &\quad - \text{tr} \log (E_n \Lambda E_n^\top + \mathbf{U}_N \Theta \mathbf{U}_N^\top - \Theta). \end{aligned}$$

Define $M(\Lambda, \Theta) = (E_n \Lambda E_n^\top + \mathbf{U}_N \Theta \mathbf{U}_N^\top - \Theta)$ which is positive definite for (Λ, Θ) in \mathcal{L}_+ . Then

$$J(\Lambda, \Theta) = \text{tr} (M(\Lambda, \Theta) \bar{\Sigma}_N) - \text{tr} \log M(\Lambda, \Theta).$$

As a function of M , this is a strictly convex function on $\mathfrak{S}_{N,+}$, whose unique minimum occurs at $M = \bar{\Sigma}_N^{-1}$ where the minimum value is $\text{tr}(I_{mN}) + \text{tr} \log \bar{\Sigma}_N$. \blacksquare

Lemma 6.3: Let (Λ_k, Θ_k) , $n \geq 1$ be a sequence of pairs in \mathcal{L}_+ such that $\|(\Lambda_k, \Theta_k)\| \rightarrow \infty$. Then also $\|A(\Lambda_k, \Theta_k)\| \rightarrow \infty$. It then follows that $\|(\Lambda_k, \Theta_k)\| \rightarrow \infty$ implies that $J(\Lambda_k, \Theta_k) \rightarrow \infty$.

Proof: Notice that A is a linear operator between finite-dimensional linear spaces. Denote by σ_m the smallest singular

value of the restriction of A to $(\ker A)^\perp$ (the orthogonal complement of $\ker A$). Clearly, $\sigma_m > 0$, so that, since each element of the sequence (Λ_k, Θ_k) is in $(\ker A)^\perp$, $\|A(\Lambda_k, \Theta_k)\| \geq \sigma_m \|(\Lambda_k, \Theta_k)\| \rightarrow \infty$.

Assume now that

$$\|A(\Lambda_k, \Theta_k)\| = \|(E_n \Lambda_k E_n^\top + \mathbf{U}_N \Theta_k \mathbf{U}_N^\top - \Theta_k)\| \rightarrow \infty.$$

Since these are all positive definite matrices and all matrix norms are equivalent, it follows that

$$\text{tr} (E_n \Lambda E_n^\top + \mathbf{U}_N \Theta \mathbf{U}_N^\top - \Theta) \rightarrow \infty.$$

As a consequence, $\text{tr}((E_n \Lambda E_n^\top + \mathbf{U}_N \Theta \mathbf{U}_N^\top - \Theta) \bar{\Sigma}_N) \rightarrow \infty$ and, finally, $J(\Lambda_k, \Theta_k) \rightarrow \infty$. \blacksquare

We show next that the dual function tends to infinity also when approaching the boundary of \mathcal{L}_+ , namely

$$\begin{aligned} \partial \mathcal{L}_+ := \{ & (\Lambda, \Theta) \in (\mathfrak{S}_{n+1} \times \mathfrak{S}_N) | (\Lambda, \Theta) \in (\ker(A))^\perp \\ & (E_n \Lambda E_n^\top + \mathbf{U}_N \Theta \mathbf{U}_N^\top - \Theta) \geq 0 \\ & \det (E_n \Lambda E_n^\top + \mathbf{U}_N \Theta \mathbf{U}_N^\top - \Theta) = 0 \}. \end{aligned}$$

Lemma 6.4: Consider a sequence (Λ_k, Θ_k) , $k \geq 1$ in \mathcal{L}_+ such that the matrix $\lim_k (E_n \Lambda_k E_n^\top + \mathbf{U}_N \Theta_k \mathbf{U}_N^\top - \Theta_k)$ is singular. Assume also that the sequence (Λ_k, Θ_k) is bounded. Then, $J(\Lambda_k, \Theta_k) \rightarrow \infty$.

Proof: Simply write

$$J(\Lambda_k, \Theta_k) = -\log \det (E_n \Lambda_k E_n^\top + \mathbf{U}_N \Theta_k \mathbf{U}_N^\top - \Theta_k) + \text{tr}(\Lambda_k \mathbf{T}_k).$$

Since $\text{tr}(\Lambda_k \mathbf{T}_k)$ is bounded, the conclusion follows. \blacksquare

Proof of Theorem 6.1: Observe that the function J is a continuous, bounded below (Lemma 6.2) function that tends to infinity both when $\|(\Lambda, \Theta)\|$ tends to infinity (Lemma 6.3) and when it tends to the boundary $\partial \mathcal{L}_+$ with $\|(\Lambda, \Theta)\|$ remaining bounded (Lemma 6.4). It follows that J is *inf-compact* on \mathcal{L}_+ , namely it has compact sublevel sets. By Weierstrass' Theorem,⁷ it admits at least one minimum point. Since J is strictly convex, the minimum point is unique. \blacksquare

VII. RECONCILIATION WITH DEMPSTER'S COVARIANCE SELECTION

Let $(\bar{\Lambda}, \bar{\Theta})$ be the unique minimum point of J in \mathcal{L}_+ (Theorem 6.1). Then $\Sigma_N^o \in \mathfrak{S}_{N,+}$ given by

$$\Sigma_N^o = (E_n \bar{\Lambda} E_n^\top + \mathbf{U}_N \bar{\Theta} \mathbf{U}_N^\top - \bar{\Theta})^{-1} \quad (52)$$

satisfies (39b) and (39c). Hence, it is the unique solution of the primal problem (MEP). Since it satisfies (39c), Σ_N^o is in particular a block-circulant matrix and hence so is

$$(\Sigma_N^o)^{-1} = (E_n \bar{\Lambda} E_n^\top + \mathbf{U}_N \bar{\Theta} \mathbf{U}_N^\top - \bar{\Theta}).$$

⁷A continuous function on a compact set always achieves its maximum and minimum on that set.

Let $\pi_{\mathfrak{C}_N}$ denote the orthogonal projection onto the linear subspace of symmetric, block-circulant matrices \mathfrak{C}_N . It follows that, in force of Lemma 6.1

$$\begin{aligned} (\Sigma_N^o)^{-1} &= \pi_{\mathfrak{C}_N} \left((\Sigma_N^o)^{-1} \right) \\ &= \pi_{\mathfrak{C}_N} (E_n \bar{\Lambda} E_n^\top + \mathbf{U}_N \bar{\Theta} \mathbf{U}_N^\top - \bar{\Theta}) \\ &= \pi_{\mathfrak{C}_N} (E_n \bar{\Lambda} E_n^\top). \end{aligned} \quad (53)$$

Theorem 7.1: Let Σ_N^o be the maximum Gaussian entropy covariance given by (52). Then $(\Sigma_N^o)^{-1}$ is a symmetric block-circulant matrix which is banded of bandwidth n . Hence, the solution of (MEP) may be viewed as the covariance of a stationary reciprocal process of order n defined on \mathbb{Z}_N .

Proof: Let

$$\Pi_{\bar{\Lambda}} := \pi_{\mathfrak{C}_N}(E_n \bar{\Lambda} E_n^\top) = \begin{bmatrix} \Pi_0 & \Pi_1^\top & \Pi_2^\top & \dots & \Pi_1 \\ \Pi_1 & \Pi_0 & \Pi_1^\top & \dots & \Pi_2 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \Pi_2^\top & \dots & \Pi_1 & \Pi_0 & \Pi_1^\top \\ \Pi_1 & \Pi_2^\top & \dots & \Pi_1 & \Pi_0 \end{bmatrix}$$

be the orthogonal projection of $(E_n \bar{\Lambda} E_n^\top)$ onto \mathfrak{C}_N . Since $\Pi_{\bar{\Lambda}}$ is symmetric and block-circulant, it is characterized by the orthogonality condition

$$\text{tr}[(E_n \bar{\Lambda} E_n^\top - \Pi_{\bar{\Lambda}}) C] = \langle E_n \bar{\Lambda} E_n^\top - \Pi_{\bar{\Lambda}}, C \rangle = 0 \quad (54)$$

for all $C \in \mathfrak{C}_N$. Next observe that, if we write $C = \text{Circ}[C_0, C_1, C_2, \dots, C_2^\top, C_1^\top]$ and

$$\bar{\Lambda} = \begin{bmatrix} \bar{\Lambda}_{00} & \bar{\Lambda}_{01} & \dots & \bar{\Lambda}_{0n} \\ \bar{\Lambda}_{01}^\top & \bar{\Lambda}_{11} & \dots & \bar{\Lambda}_{1n} \\ \vdots & \ddots & \ddots & \vdots \\ \bar{\Lambda}_{0n}^\top & \bar{\Lambda}_{1n}^\top & \dots & \bar{\Lambda}_{nn} \end{bmatrix}$$

with $\bar{\Lambda}_{k,j} = \bar{\Lambda}_{j,k}^\top$, then

$$\begin{aligned} \text{tr}[E_n \bar{\Lambda} E_n^\top C] &= \text{tr}[\bar{\Lambda} E_n^\top C E_n] \\ &= \text{tr}[(\bar{\Lambda}_{00} + \bar{\Lambda}_{11} + \dots + \bar{\Lambda}_{nn}) C_0 \\ &\quad + (\bar{\Lambda}_{01} + \bar{\Lambda}_{12} + \dots + \bar{\Lambda}_{n-1,n}) C_1 + \dots \\ &\quad + \bar{\Lambda}_{0n} C_n \times (\bar{\Lambda}_{10} + \bar{\Lambda}_{21} + \dots, \bar{\Lambda}_{n,n-1}) \\ &\quad \times C_1^\top + \dots + \bar{\Lambda}_{n0} C_n^\top]. \end{aligned}$$

On the other hand, recalling that the product of two block-circulant matrices is block-circulant, we have that $\text{tr}[\Pi_{\bar{\Lambda}} C]$ is simply N times the trace of the first block row of $\Pi_{\bar{\Lambda}}$ times the first block column of C . We get

$$\begin{aligned} \text{tr}[\Pi_{\bar{\Lambda}} C] &= N \text{tr}[\Pi_0 C_0 + \Pi_1^\top C_1 + \Pi_2^\top C_2 + \dots \\ &\quad \dots + \Pi_2 C_2^\top + \Pi_1 C_1^\top]. \end{aligned}$$

Hence, the orthogonality condition (54), reads

$$\begin{aligned} &\text{tr}[(E_n \bar{\Lambda} E_n^\top - \Pi_{\bar{\Lambda}}) C] \\ &= \text{tr}[(\bar{\Lambda}_{00} + \bar{\Lambda}_{11} + \dots + \bar{\Lambda}_{nn}) - N \Pi_0] C_0 \\ &\quad + ((\bar{\Lambda}_{01} + \bar{\Lambda}_{12} + \dots + \bar{\Lambda}_{n-1,n}) - N \Pi_1^\top) C_1 \\ &\quad + ((\bar{\Lambda}_{10} + \bar{\Lambda}_{21} + \dots, \bar{\Lambda}_{n,n-1}) - N \Pi_1) C_1^\top \\ &\quad + \dots (\bar{\Lambda}_{0n} - N \Pi_1^\top) C_n + (\bar{\Lambda}_{n0} - N \Pi_1) C_n^\top \\ &\quad + N \Pi_{n+1}^\top C_{n+1} + N \Pi_{n+1} C_{n+1}^\top \\ &\quad + N \Pi_{n+2}^\top C_{n+2} + N \Pi_{n+2} C_{n+2}^\top + \dots \\ &= 0. \end{aligned}$$

Since this must hold true for all $C \in \mathfrak{C}_N$, we conclude that

$$\begin{aligned} \Pi_0 &= \frac{1}{N}(\bar{\Lambda}_{00} + \bar{\Lambda}_{11} + \dots + \bar{\Lambda}_{nn}) \\ \Pi_1 &= \frac{1}{N}(\bar{\Lambda}_{01} + \bar{\Lambda}_{12} + \dots + \bar{\Lambda}_{n-1,n})^\top \\ &\vdots \\ \Pi_n &= \frac{1}{N} \bar{\Lambda}_{0n}^\top \end{aligned}$$

while from the last equation we get $\Pi_i = 0$, for all i in the interval $n+1 \leq i \leq N-n-1$. From this it is clear that the inverse of the covariance matrix solving the primal problem (MEP), namely $\Pi_{\bar{\Lambda}} = (\Sigma_N^o)^{-1}$ has a circulant block-banded structure of bandwidth n . ■

Since the beginning of Section V, we have been dealing only with Gaussian distributions in order to facilitate the comparison with Dempster's classical results. It is now time to show that the Gaussian assumption can be dispensed with, and our solution is indeed optimal in the larger family of (zero-mean) second-order distributions.

Theorem 7.2: The Gaussian distribution with (zero-mean and) covariance Σ_N^o defined by (52) maximizes the entropy functional (36) over the set of all (zero-mean) probability densities whose covariance matrix satisfies the boundary conditions (39b), (39c).

Proof: Let $\mathfrak{C}_N(\mathbf{T}_n)$ be the set of (block-circulant) covariance matrices satisfying the boundary conditions (39b), (39c) and let p_{Σ} be a probability density with zero-mean and covariance Σ . In particular, we shall denote by g_{Σ} the Gaussian density with zero-mean and covariance Σ . Now, by a famous theorem of Shannon [37], the probability distribution having maximum entropy in the class of all distribution with a fixed mean vector (which we take equal to zero) and variance matrix Σ , is the Gaussian distribution g_{Σ} . Hence

$$\max_{\Sigma \in \mathfrak{C}_N(\mathbf{T}_n)} \left\{ \max_{p_{\Sigma}} [H(p_{\Sigma})] \right\} = \max_{\Sigma \in \mathfrak{C}_N(\mathbf{T}_n)} \{H(g_{\Sigma})\}$$

where the maximum in the right-hand side is attained by $g_{\Sigma_N^o}$. ■

The above can be interpreted as a particular covariance selection result in the vein of Dempster's paper; compare in particular [11, Proposition a]. In fact the results of this section substantiate also the maximum entropy principle of Dempster (Proposition 5.1). It is, however, important to note that none of our results follows as a particular case from Dempster's results, since [11]

deals with a very unstructured setting. In particular our main result (Theorem 7.1) that the solution, Σ_N^o , to our primal problem (MEP) has a block-circulant *banded* inverse, is completely original. Its proof uses in an essential way the characterization of the MEP solution provided by our variational analysis and cleverly exploits the block-circulant structure.

Actually, our results, *together* with Dempster's, may be used to show that the maximum entropy distribution, subject only to moment constraints (compatible with the circulant structure) on a block band and on the corners, is necessarily block-circulant, i.e. the underlying process is stationary.⁸

Because of the equivalence of reciprocal AR modeling and the underlying process covariance having an inverse with a banded structure, explained in Section III, we see that the maximum entropy principle leads in fact to (reciprocal) AR models. This makes contact with the ever-present problem in control an signal processing of (approximate) AR modeling from finite covariance data, whose solution dates back to the work of N. Levinson and P. Whittle. That AR modeling from finite covariance data is actually equivalent to a positive band extension problems for infinite Toeplitz matrices has been realized and studied in the past decades by Dym, Gohberg and co-workers, see e.g. [13], [17] as representative references of a very large literature. We should stress here that band extension problem for infinite Toeplitz matrices are invariably attacked and solved by factorization techniques, but circulant matrices do not fit in the "banded algebra" framework used in the literature. Also, one should note that the maximum entropy property is usually presented in the literature as a final embellishment of a solution which was already obtained by factorization techniques. Here, for the circulant band extension problem, factorization techniques do not work and the maximum entropy principle turns out to be the key to the solution of the problem.

This fact, together with Dempster's observation [11, Proposition b], may be taken as a proof (although referred to a very specific case) of a very much quoted general principle that maximum entropy distributions are distributions achieving maximum *simplicity of explanation* of the data.

Finally, we anticipate that the results of this section lead to an efficient iterative algorithm for the explicit solution of the MEP which is guaranteed to converge to a unique minimum. This solves the variational problem, and hence, the circulant band extension problem which subsumes maximum likelihood identification of reciprocal processes. This algorithm, which will not be described here for reasons of space limitations, compares very favorably with the best techniques available so far.

VIII. CONCLUSION

A new class of stationary reciprocal processes on a finite interval has been introduced which are the acausal analog of autoregressive (AR) processes on the integer line. Maximum likelihood identification of these AR-type reciprocal models is discussed. The computation of the estimates of the matrix parameters of the model turns out to be a particular instance of a *covariance selection problem* of the kind studied by the statistician

A.P. Dempster in the early seventies. In matrix terminology, the covariance selection for stationary reciprocal models is equivalent to a special *matrix band extension problem* for block-circulant matrices. We have shown that this band extension problem can be solved by maximizing an entropy functional.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers whose detailed comments and suggestions have helped to improve the manuscript.

REFERENCES

- [1] O. E. Barndorff-Nielsen, *Information and Exponential Families in Statistical Theory*. Hoboken, NJ: Wiley, 1978.
- [2] S. Bernstein, "Sur les liaisons entre le grandeurs aleatoires," in *Proc. Int. Congr. Math.*, Zürich, Switzerland, 1932, pp. 288–309.
- [3] C. Byrnes and A. Lindquist, "Interior point solutions of variational problems and global inverse function theorems," *Int. J. Robust Nonlin. Control*, vol. 17, pp. 463–481, 2007, Special Issue in Honor of V. A. Yakubovich on the Occasion of his 80th birthday.
- [4] F. Carli and T. T. Georgiou, "On the covariance completion problem under a circulant structure," *IEEE Trans. Autom. Control*, to be published.
- [5] F. Carli and G. Picci, "On the factorization approach to band extension of block-circulant matrices," in *Proc. 19th Int. Symp. Mathematical Theory of Networks and Systems (MTNS)*, Budapest, Hungary, Jul. 2010, pp. 907–914.
- [6] A. Chiuso, A. Ferrante, and G. Picci, "Reciprocal realization and modeling of textured images," presented at the 44th IEEE Conf. Decision and Control, Seville, Spain, Dec. 2005.
- [7] A. Chiuso and G. Picci, "Some identification techniques in computer vision (invited paper)," presented at the 47th IEEE Conf. Decision and Control, Cancun, Mexico, Dec. 2008.
- [8] P. Davis, *Circulant Matrices*. Hoboken, NJ: Wiley, 1979.
- [9] P. Delsarte, Y. V. Genin, and Y. G. Kamp, "Orthogonal polynomial matrices on the unit circle," *IEEE Trans. Circuits Syst.*, vol. CAS-25, no. 3, pp. 149–160, Mar. 1978.
- [10] A. Dembo, C. Mallows, and L. Shepp, "Embedding nonnegative definite Toeplitz matrices in nonnegative definite circulant matrices, with application to covariance estimation," *IEEE Trans. Inf. Theory*, vol. 35, no. 6, pp. 1206–1212, Nov. 1989.
- [11] A. P. Dempster, "Covariance selection," *Biometrics*, vol. 28, pp. 157–175, 1972.
- [12] J. L. Doob, *Stochastic Processes*. Hoboken, NJ: Wiley, 1990, Wiley Classics Library, Reprint of the 1953 original, A Wiley-Interscience Publication.
- [13] H. Dym and I. Gohberg, "Extension of band matrices with band inverses," *Lin. Alg. Appl.*, vol. 36, pp. 1–24, 1981.
- [14] A. Ferrante, M. Pavon, and F. Ramponi, "Further results on the Byrnes-Georgiou-Lindquist generalized moment problem," in *Modeling, Estimation and Control*, A. Ferrante, A. Chiuso, and S. Pinzoni, Eds. New York: Springer-Verlag, 2007, pp. 73–83, Festschrift in Honor of Giorgio Picci on the Occasion of His 65th Birthday.
- [15] R. Frezza, "Models of Higher-order and Mixed-Order Gaussian Reciprocal Processes With Application to the Smoothing Problem," Ph.D. dissertation, Applied Mathematics Program, Univ. California, Davis, 1990.
- [16] T. T. Georgiou and A. Lindquist, "Kullback-Leibler approximation of spectral density functions," *IEEE Trans. Inf. Theory*, vol. 49, no. 11, pp. 2910–2917, Nov. 2003.
- [17] I. Gohberg, Goldberg, and M. Kaashoek, *Classes of Linear Operators Vol II*. Boston, MA: Birkhauser, 1994.
- [18] E. J. Hannan and M. Deistler, *The Statistical Theory of Linear Systems*. Hoboken, NJ: Wiley, 1988.
- [19] B. Jamison, "Reciprocal processes: the stationary Gaussian case," *Ann. Math. Statist.*, vol. 41, pp. 1624–1630, 1970.
- [20] B. Jamison, "Reciprocal processes," *Zeitschrift. Wahrsch. Verw. Gebiete*, vol. 30, pp. 65–86, 1974.
- [21] B. Jamison, "The Markov process of Schrödinger," *Zeitschrift. Wahrsch. Verw. Gebiete*, vol. 32, pp. 323–331, 1975.
- [22] A. J. Krener, "Realization of reciprocal processes," presented at the IIASA Conf. Modeling and Adaptive Control, 1986, C. I. Byrnes, G. B. Di Masi, and A. Kurshanskij, Eds..

⁸An alternative proof of this fact can be constructed based on the invariance properties of the entropy functional and its strict concavity. This has recently been established (in a more general framework) in [4].

- [23] A. J. Krener, "Reciprocal processes and the stochastic realization problem for acausal systems," in *Modeling Identification and Robust Control*, C. I. Byrnes and A. Lindquist, Eds. Amsterdam, The Netherlands: North Holland, 1986, pp. 197–211.
- [24] N. Levinson, "The Wiener r.m.s. (root means square) error criterion in filter design and prediction," *J. Math. Phys.*, vol. 25, pp. 261–278, 1947.
- [25] B. C. Levy, "Regular and reciprocal multivariate stationary Gaussian reciprocal processes over Z are necessarily Markov," *J. Math. Syst. Estim. Control*, vol. 2, pp. 133–154, 1992.
- [26] B. C. Levy and A. Ferrante, "Characterization of stationary discrete-time Gaussian reciprocal processes over a finite interval," *SIAM J. Matrix Anal. Appl.*, vol. 24, pp. 334–355, 2002.
- [27] B. C. Levy, R. Frezza, and A. J. Krener, "Modeling and estimation of discrete-time Gaussian reciprocal processes," *IEEE Trans. Autom. Control*, vol. AC-35, no. 9, pp. 1013–1023, Sep. 1990.
- [28] A. Lindquist and G. Picci, "Realization theory for multivariate stationary Gaussian processes," *SIAM J. Control Optim.*, vol. 23, no. 6, pp. 809–857, 1985.
- [29] M. Loeve, *Probability Theory*. New York: Van Nostrand Reinhold, 1963, Reprinted in 1978 by Springer-Verlag, New York.
- [30] R. M. Gray, "Toeplitz and Circulant Matrices: A Review," Stanford Univ., Stanford, CA [Online]. Available: <http://ee.stanford.edu/gray/toeplitz.pdf>, 2002
- [31] P. Masani, "The prediction theory of multivariate stochastic processes, iii," *Acta Math.*, vol. 104, pp. 141–162, 1960.
- [32] G. Picci and F. Carli, "Modelling and simulation of images by reciprocal processes," in *Proc. 10th Int. Conf. Computer Modeling and Simulation UKSIM*, 2008, pp. 513–518.
- [33] J. A. Sand, "Four Papers in Stochastic Realization Theory," Ph.D. dissertation, Dept. Math., Royal Inst. Technol. (KTH), Stockholm, Sweden, 1994.
- [34] J. A. Sand, "Reciprocal realizations on the circle," *SIAM J. Control Optim.*, vol. 34, pp. 507–520, 1996.
- [35] E. Schrödinger, "Über die umkehrung der naturgesetze," *Sitzungsber. Preuss. Akad. Wiss Berlin Phys. Math. Kl.* 8, vol. 9, pp. 144–153, 1931.
- [36] E. Schrödinger, "Sur la theorie relativiste de l'electron et l'interpretation de la mecanique quantique," *Ann. Inst. H. Poincare*, vol. 2, pp. 269–310, 1932.
- [37] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 1948, 623–656, Reprinted in: C. E. Shannon, W. Weaver, *The Mathematical Theory of Communication* Univ. of Illinois, Press 1949.
- [38] G. J. Tee, "Eigenvectors of block-circulant and alternating circulant matrices," *Re. Lett. Inf. Math. Sci.*, vol. 8, pp. 123–142, 2005.
- [39] E. E. Tyrtyshnikov, "A unifying approach to some old and new theorems on distribution and clustering," *Lin. Alg. Appl.*, vol. 232, pp. 1–43, 1968.
- [40] P. Whittle, "On the fitting of multivariate autoregressions and the approximate spectral factorization of a spectral density matrix," *Biometrika*, vol. 50, pp. 129–134, 1963.
- [41] D. C. Youla and N. N. Kazanjian, "Bauer-type factorization of positive matrices and the theory of matrix polynomials orthogonal on the unit circle," *IEEE Trans. Circuits Syst.*, vol. CAS-25, pp. 57–65, 1978.
- [42] P. W. Zehna, "Invariance of maximum likelihood estimators," *Ann. Math. Statist.*, vol. 37, p. 744, 1966.



Francesca P. Carli received the M.Sc. engineering degree (*cum laude*) from the University of Padova, Padova, Italy, in 2007. She is currently pursuing the Ph.D. degree in information engineering at the Department of Information Engineering, University of Padova.

From September 2009 to January 2010, she was a visiting scholar with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN. Her current research interests are in the areas of identification and estimation theory

and statistical signal processing.



optimal control and optimal filtering, quantum control, and stochastic realization.

Augusto Ferrante was born in Piove di Sacco, Italy, on August 5, 1967. He received the "Laurea" degree (*cum laude*) in electrical engineering in 1991 and the Ph.D. degree in control systems engineering in 1995, both from the University of Padova, Padova, Italy.

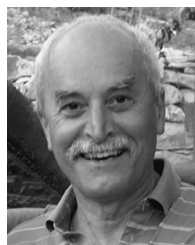
He has been on the faculty of the Colleges of Engineering of the University of Udine, Italy, and of the "Politecnico di Milano", Italy. He is presently a Professor in the "Department of Information Engineering", University of Padova. His research interests are in the areas of linear systems, spectral estimation,



Michele Pavon was born in Venice, Italy, on October 12, 1950. He received the Laurea degree from the University of Padova, Padova, Italy, in 1974, and the Ph.D. degree from the University of Kentucky, Lexington, in 1979, both in mathematics.

After service in the Italian Army, he was on the research staff of LADSEB-CNR, Padova, for six years. Since July 1986, he has been a Professor at the College of Engineering, University of Padova. He has visited several institutions in Europe and Northern America. His present research interests include max-

imum entropy problems and quantum control.



Giorgio Picci (S'67–M'70–SM'91–F'94–LF'08) graduated (*cum laude*) from the University of Padova, Padova, Italy, in 1967.

He has been a full Professor with the Department of Information Engineering, University of Padova, since 1980. He has held several long-term visiting appointments with various American, Japanese and European universities, including Brown University, Providence, RI; Massachusetts Institute of Technology, Cambridge; University of Kentucky; Arizona State University; Center for Mathematics

and Computer Sciences (C.W.I.), Amsterdam, The Netherlands; Royal Institute of Technology, Stockholm, Sweden; Kyoto University, Japan; and Washington University, St. Louis, Mo. He has been contributing to systems and control theory mostly in the areas of modeling, estimation, and identification of stochastic systems and has published over 100 papers and edited three books in these areas. He has been involved in various joint research projects with industry and state agencies.

Dr. Picci has been chairman of the IFAC Technical Committee on Stochastic Systems, past member of the EUCA council, project manager of the Italian team for the Commission of the European Communities Network of Excellence System Identification (ERNSI), and general coordinator of the Commission of European Communities IST project RECSYS, in the fifth Framework Program. He is an IFAC Fellow and a foreign member of the Swedish Royal Academy of Engineering Sciences.