A Maximum Entropy approach to the Covariance Extension Problem for Reciprocal Processes

Francesca Carli, Augusto Ferrante, Michele Pavon, and Giorgio Picci

***Abstract*— This paper addresses the problem of completing a partially specified symmetric matrix, where the specified entries lie on a single band centered on the main diagonal, in such a way that the completed matrix is positive definite, block-circulant and with a banded inverse. This particular completion has the meaning of the covariance matrix of a reciprocal process stationary on the discrete circle $\mathbb{Z}_N$. This problem, called the *block-circulant band extension problem*, arises in the context of maximum likelihood identification for such processes. This paper shows that the block-circulant band extension problem can in fact be solved as a maximum entropy problem. Indeed, the constraint that the inverse be banded can be removed with a considerable theoretical and computational simplification, as the maximum entropy block-circulant extension can be shown to always enjoy this property. Conditions for the feasibility of the problem are also provided.**

I. Introduction

This paper is a shortened version without detailed proofs of [1]. It deals with the problem of completing a partially specified block-circulant symmetric matrix, where the specified elements lie on a single band centered on the main diagonal, in such a way that the completion is a positive definite block-circulant matrix with a banded inverse. From now on this matrix completion problem will be referred to as the *block-circulant band extension problem*. This particular completion has the meaning of a covariance matrix of a vector valued reciprocal process stationary on the discrete circle $\mathbb{Z}_N$. The completion problem arises in the context of estimation and identification for such processes.

First of all, we show that the block-circulant band extension problem can be solved as a maximum entropy problem where the constraint that enforces the inverse of the completion to be banded can be removed. Moreover comparing this problem with its well-known unconstrained version [5], [8], i.e. the version where no additional constraints are placed on the overall structure of the matrix to be completed, it can be shown that, provided the specified (block-)entries are consistent with the circulant structure, then the maximum entropy completion is necessarily circulant, see [2].

No feasibility conditions are known for the general maximum entropy completion problem, except for the case where the graph underlying the specified pattern is chordal [8]. In this paper we show that, under the obvious necessary condition that the principal submatrix composed by the specified entries is positive definite, the problem is feasible provided that the dimension of the extension is sufficiently large.

Work partially supported by the Italian Ministry for Education and Resarch (MIUR) under PRIN grant "Identification and Adaptive Control of Industrial Systems".

F. Carli, A. Ferrante and G. Picci are with the Department of Information Engineering (DEI), University of Padova, via Gradenigo 6/B, 35131 Padova, Italy. 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, pavon@math.unipd.it

II. Maximum entropy on the discrete circle

Let $\mathcal{I}$ and $\mathcal{I}_b$ denote the sets of pairs of block indices consistent, respectively, with a generic symmetric structure, i.e.

$$(i,j) \in \mathcal{I} \Rightarrow (j,i) \in \mathcal{I}$$

and with a banded-symmetric circulant structure of bandwidth n, i.e.

$$\begin{aligned} &|i-j| \leq n \Rightarrow (i,j) \in \mathcal{I}_b \\ &(i,j) \in \mathcal{I}_b \Rightarrow (j,i) \in \mathcal{I}_b \\ &(i,j) \in \mathcal{I}_b \Rightarrow \Big((j+1)_{\text{mod N}}, (i+1)_{\text{mod N}}\Big) \in \mathcal{I}_b. \end{aligned}$$

Let $\mathcal{R}$ and $\mathcal{R}_b$ be the corresponding sets of values, also assumed to be consistent with the generic symmetric structure and with the banded-symmetric circulant structure, respectively, i.e.

$$\begin{aligned} &\mathcal{R} := \left\{r_{ij} \mid (i,j) \in \mathcal{I},\, r_{ij} = r_{ji}^*\right\} \\ &\mathcal{R}_b := \left\{r_{ij} \mid (i,j) \in \mathcal{I}_b, r_{ij} = r_{ji}^*, r_{ij} = r^*_{(j+1)_{\text{mod N}}(i+1)_{\text{mod N}}}\right\} \end{aligned}$$

where r_{ij} are $m \times m$ matrices and r^* denotes the complex-conjugate-transpose. Recall that 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 the C_k's are $m \times m$ matrices. It is apparent that a block-circulant matrix $\mathbf{C}_N$ is fully specified by its first block-column (or row). Thus, for ease of notation, we will write

$$\mathbf{C}_N = \text{Circ}\{C_0, C_1, \dots, C_{N-1}\}. \tag{1}$$

The set of $N \times N$ m-block-circulant matrices will be denoted as $\mathcal{C}_{mN}$. Finally, let e_k be the $m \times mN$ block-matrix with the identity in the k-th block position,

$$e_k = \begin{bmatrix} 0 & \dots & 0 & I_m & 0 & \dots & 0 \end{bmatrix}$$

ISBN 978-963-311-370-7

and $\mathfrak{S}_N$ the vector space of *symmetric* matrices with $N \times N$ square blocks of dimension $m \times m$. We are interested in is the following problem.

Problem 2.1:

$$\text{Find } \mathbf{\Sigma}_N \in \mathfrak{S}_N \text{ such that} \tag{2a}$$

$$e_i^\top \mathbf{\Sigma}_N e_j = r_{ij}, \text{ for } (i,j) \in \mathcal{I}_b \text{ and } r_{ij} \in \mathcal{R}_b, \tag{2b}$$

$$\mathbf{\Sigma}_N > 0, \tag{2c}$$

$$e_i^\top \mathbf{\Sigma}_N^{-1} e_j = 0, \text{ for } (i,j) \notin \mathcal{I}_b, \tag{2d}$$

$$\mathbf{\Sigma}_N \in \mathcal{C}_{mN}. \tag{2e}$$

A special case of the problem addressed in Dempster's paper reads as follows.

Problem 2.2:

$$\text{Find } \mathbf{\Sigma}_N \in \mathfrak{S}_N \text{ such that} \tag{3a}$$

$$e_i^\top \mathbf{\Sigma}_N e_j = r_{ij}, \text{ for } (i,j) \in \mathcal{I} \text{ and } r_{ij} \in \mathcal{R}, \tag{3b}$$

$$\mathbf{\Sigma}_N > 0, \tag{3c}$$

$$e_i^\top \mathbf{\Sigma}_N^{-1} e_j = 0, \text{ for } (i,j) \notin \mathcal{I}. \tag{3d}$$

For our purposes, a key observation is Statement (b) in [5, p. 160]. In our setting, it reads as follows.

Proposition 2.1: Assume that Problem 2.2 is feasibile. Among all the positive definite extensions, there exists a unique one 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.

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

$$H(p) = -\int_{\mathbb{R}^n} \log(p(x))p(x)dx. \tag{4}$$

In the case of a zero-mean Gaussian distribution p with covariance matrix $\mathbf{\Sigma}_N$, we get

$$H(p) = \frac{1}{2}\log(\det \mathbf{\Sigma}_N) + \frac{1}{2}n\left(1 + \log(2\pi)\right). \tag{5}$$

Thus, exploiting this maximum entropy principle, Problem 2.2 can be restated as follows.

Problem 2.3:

$$\max\left\{\log\det \mathbf{\Sigma}_N \mid \mathbf{\Sigma}_N \in \mathfrak{S}_N,\ \mathbf{\Sigma}_N > 0\right\} \tag{6a}$$

$$\text{subject to :}$$

$$e_i^\top \mathbf{\Sigma}_N e_j = r_{ij}, \text{ for } (i,j) \in \mathcal{I} \text{ and } r_{ij} \in \mathcal{R}. \tag{6b}$$

At first sight, our problem seems to be a particular instance of Dempster's where the assigned elements are those on the main diagonals and in the NE and SW corners. Notice, however, that the linear constraint that forces the completed matrix to be circulant is not present in the Dempster's setting. Neverthless, inspired by this maximum entropy principle, we can switch to consider the following problem.

Problem 2.4:

$$\max\left\{\log\det \mathbf{\Sigma}_N \mid \mathbf{\Sigma}_N \in (\mathfrak{S}_N \cap \mathcal{C}_{mN}),\ \mathbf{\Sigma}_N > 0\right\} \tag{7a}$$

$$\text{subject to :}$$

$$e_i^\top \mathbf{\Sigma}_N e_j = r_{ij}, \text{ for } (i,j) \in \mathcal{I}_b \text{ and } r_{ij} \in \mathcal{R}_b. \tag{7b}$$

In the sequel, we solve Problem 2.4 and show that it is equivalent to our original Problem 2.1. To this aim, let us introduce the following notation. Let $\mathbf{U}_N$ denote the block-circulant "shift" matrix with $N \times N$ blocks,

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

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 $\mathbf{C}_N$ 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

$$\mathbf{U}_N^\top \mathbf{C}_N \mathbf{U}_N = \mathbf{C}_N. \tag{8}$$

Finally, let $\mathbf{T}_n \in \mathfrak{S}_{n+1}$ denote the Toeplitz matrix of *boundary data*:

$$\mathbf{T}_n = \begin{bmatrix} \Sigma_0 & \Sigma_1^\top & \dots & \Sigma_n^\top \\ \Sigma_1 & \dots & & \dots \\ \dots & \dots & & \dots \\ \Sigma_n & \dots & & \Sigma_0 \end{bmatrix} \tag{9}$$

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

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

In this notation, Problem 2.4 can be restated as:

Problem 2.5:

$$\max\left\{\log\det \mathbf{\Sigma}_N \mid \mathbf{\Sigma}_N \in \mathfrak{S}_N,\ \mathbf{\Sigma}_N > 0\right\} \tag{10a}$$

$$\text{subject to :}$$

$$E_n^\top \mathbf{\Sigma}_N E_n = \mathbf{T}_n, \tag{10b}$$

$$\mathbf{U}_N^\top \mathbf{\Sigma}_N \mathbf{U}_N = \mathbf{\Sigma}_N. \tag{10c}$$

The above problem indeed amounts to finding the maximum entropy Gaussian distribution with a block-circulant covariance, whose first $n+1$ blocks are precisely $\Sigma_0, \dots, \Sigma_n$. 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 $\mathbf{\Sigma}_N$ of Problem 2.5 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 Problem 2.5, namely the existence of a positive definite, symmetric matrix $\mathbf{\Sigma}_N$ satisfying (10b)-(10c). Obviously, $\mathbf{T}_n$ positive definite is a necessary condition for the existence of such a $\mathbf{\Sigma}_N$. In general it turns out that, under such a necessary condition, feasibility holds for N large enough. The idea is that for $N \to \infty$, Toeplitz matrices can be approximated arbitrarily well by circulants ([9], [10]) and hence existence of a positive block-circulant extension can be derived from the existence of positive extensions for Toeplitz matrices.

Theorem 2.1: *Given the sequence $\Sigma_i \in \mathbb{R}^{m\times m}$, $i = 0, 1, \dots, n$, such that*

$$\mathbf{T}_n = \mathbf{T}_n^\top > 0, \tag{11}$$

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 $\mathbf{\Sigma}_N$.

Proof: A fundamental result in stochastic system theory is the so-called maximum entropy covariance extension. It states (see [11], [3] and [12]) that, under condition (11), there exists a rational positive real function

$$\Phi_+(z) = \frac{\Sigma_0}{2} + C(zI - A)^{-1}B$$

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. the spectral density $\Phi(z)$ is bounded away from zero on the unit circle:

$$\exists \varepsilon > 0 \text{ such that } \Phi(\mathrm{e}^{\mathrm{j}\vartheta}) > \varepsilon I, \ \forall \vartheta \in [0, 2\pi). \tag{12}$$

Define $\Sigma_i := CA^{i-1}B$, $i = n+1, n+2, \dots$, so that

$$\Phi_+(z) = \frac{\Sigma_0}{2} + \sum_{i=1}^{\infty} \Sigma_i z^{-i},$$

and let

$$\mathbf{\Sigma}_N := \mathrm{Circ}\left(\Sigma_0, \Sigma_1^\top, \dots, \Sigma_{\frac{N-1}{2}}^\top, \Sigma_{\frac{N-1}{2}}, \Sigma_{\frac{N-1}{2}-1}, \dots \Sigma_1\right)$$

for N odd, and

$$\mathbf{\Sigma}_N := \mathrm{Circ}\left(\Sigma_0, \Sigma_1^\top, \dots, \Sigma_{\frac{N-2}{2}}^\top, \Sigma_{\frac{N}{2}}^\top + \Sigma_{\frac{N}{2}}, \Sigma_{\frac{N-2}{2}}, \dots \Sigma_1\right)$$

for N even.

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

$$\mathbf{V}^*\mathbf{\Sigma}_N\mathbf{V} = \mathbf{\Psi}_N := \mathrm{diag}\left(\Psi_0, \Psi_1, \dots, \Psi_{N-1}\right),$$

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

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

and Ψ_ℓ are the coefficients of the finite Fourier transform of the first block row of $\mathbf{\Sigma}_N$:

$$\Psi_\ell = \Sigma_0 + \mathrm{e}^{\mathrm{j}\vartheta_\ell}\Sigma_1^\top + \left(\mathrm{e}^{\mathrm{j}\vartheta_\ell}\right)^2\Sigma_2^\top + \dots + \left(\mathrm{e}^{\mathrm{j}\vartheta_\ell}\right)^{N-1}\Sigma_1, \tag{13}$$

with $\vartheta_\ell := -2\pi\ell/N$. Clearly, $\left(\mathrm{e}^{\mathrm{j}\vartheta_\ell}\right)^{N-i} = \left(\mathrm{e}^{\mathrm{j}\vartheta_\ell}\right)^{-i}$ and hence

$$\Psi_\ell = \Phi\left(\mathrm{e}^{\mathrm{j}\vartheta_\ell}\right) - \left[\delta\Phi_N\left(\mathrm{e}^{\mathrm{j}\vartheta_\ell}\right) + \delta\Phi_N^*\left(\mathrm{e}^{\mathrm{j}\vartheta_\ell}\right)\right] \tag{14}$$

where,

$$\delta\Phi_N(z) := \sum_{i=h+1}^{\infty} \Sigma_i z^{-i} \tag{15}$$

with

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

The quantity $\delta\Phi_N(z)$ may be rewritten in the form

$$\delta\Phi_N(z) = \sum_{i=h+1}^{\infty} CA^{i-1}Bz^{-i} = z^{-h}CA^h(zI - A)^{-1}B. \tag{16}$$

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

$$\delta\Phi_N\left(\mathrm{e}^{\mathrm{j}\vartheta_\ell}\right) + \delta\Phi_N^*\left(\mathrm{e}^{\mathrm{j}\vartheta_\ell}\right) < \varepsilon I, \quad \forall \vartheta_\ell, \quad \forall N \geq \bar{N} \tag{17}$$

so that it readily follows from (12) and (14) that if $N \geq \bar{N}$, $\Psi_\ell > 0$ 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 $\bar{N}$ for which (17) holds. In other words, Theorem 2.1 provides a sufficient condition that can be practically tested. Similar bounds, but valid only for the scalar case, were derived in [4].

III. Variational analysis

We shall introduce a suitable set of "Lagrange multipliers" for our constrained optimization Problem 2.5. Consider the linear map $A : \mathfrak{S}_{n+1} \times \mathfrak{S}_N \to \mathfrak{S}_N$ 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

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

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*

$$\begin{aligned} L(\mathbf{\Sigma}_N, \Lambda, \Theta) \ &:= \ -\mathrm{tr}\log\mathbf{\Sigma}_N + \mathrm{tr}\left(\Lambda\left(E_n^\top\mathbf{\Sigma}_N E_n - \mathbf{T}_n\right)\right) \\ &\quad + \mathrm{tr}\left(\Theta\left(\mathbf{U}_N^\top\mathbf{\Sigma}_N\mathbf{U}_N - \mathbf{\Sigma}_N\right)\right) \\ &= \ -\mathrm{tr}\log\mathbf{\Sigma}_N + \mathrm{tr}\left(E_n\Lambda E_n^\top\mathbf{\Sigma}_N\right) \\ &\quad - \mathrm{tr}\left(\Lambda\mathbf{T}_n\right) + \mathrm{tr}\left(\mathbf{U}_N\Theta\mathbf{U}_N^\top\mathbf{\Sigma}_N\right) \\ &\quad - \mathrm{tr}\left(\Theta\mathbf{\Sigma}_N\right) \end{aligned}$$

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

$$\begin{aligned} \delta L(\mathbf{\Sigma}_N, \Lambda, \Theta; \delta\mathbf{\Sigma}_N) = &-\mathrm{tr}\left(\mathbf{\Sigma}_N^{-1}\delta\mathbf{\Sigma}_N\right) + \mathrm{tr}\left(E_n\Lambda E_n^\top\delta\mathbf{\Sigma}_N\right) \\ &+ \mathrm{tr}\left(\left(\mathbf{U}_N\Theta\mathbf{U}_N^\top - \Theta\right)\delta\mathbf{\Sigma}_N\right). \end{aligned}$$

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

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

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

$$\mathbf{\Sigma}_N^o = \left(E_n\Lambda E_n^\top + \mathbf{U}_N\Theta\mathbf{U}_N^\top - \Theta\right)^{-1}. \tag{18}$$

Consider next $L(\mathbf{\Sigma}_N^o, \Lambda, \Theta)$. We get

$$L(\mathbf{\Sigma}_N^o, \Lambda, \Theta) = \operatorname{tr}\log\left(E_n \Lambda E_n^\top + \mathbf{U}_N \Theta \mathbf{U}_N^\top - \Theta\right) + \operatorname{tr} I_{mN} - \operatorname{tr}\left(\Lambda \mathbf{T}_n\right).$$

This is a strictly concave function on $\mathcal{L}_+$ whose maximization is the *dual problem* of Problem 2.5. We can equivalently consider the convex problem

$$\min\left\{J(\Lambda,\Theta), (\Lambda,\Theta) \in \mathcal{L}_+\right\}, \tag{19}$$

where J (henceforth called dual function) is given by

$$J(\Lambda,\Theta) = \operatorname{tr}\left(\Lambda \mathbf{T}_n\right) - \operatorname{tr}\log\left(E_n \Lambda E_n^\top + \mathbf{U}_N \Theta \mathbf{U}_N^\top - \Theta\right). \tag{20}$$

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, whose complete proof is provided in [1], can be established.

Theorem 3.1: *The function J admits a unique minimum point $(\bar\Lambda, \bar\Theta)$ in $\mathcal{L}_+$.*

Sketch of the proof: First, we need to characterize the orthogonal complement of the vector subspace of block-circulant matrices $\mathfrak{C}_N$ in $\mathfrak{S}_N$. More precisely we need to show that $M \in (\mathfrak{C}_N)^\perp$ if and only if it can be expressed as

$$M = \mathbf{U}_N N \mathbf{U}_N^\top - N \tag{21}$$

for some $N \in \mathfrak{S}_N$. This will allow us to show that the function J (which is continuous) is bounded below. The third step is to show that J tends to infinity both when $\|(\Lambda,\Theta)\|$ tends to infinity and when it tends to the boundary $\partial\mathcal{L}_+$ with $\|(\Lambda,\Theta)\|$ remaining bounded. It follows that J is *inf-compact* on $\mathcal{L}_+$, namely it has compact sublevel sets. By Weierstrass' Theorem, it admits at least one minimum point. Since J is strictly convex, the minimum point is unique. □

IV. Reconciliation with Dempster's Covariance Selection

In this section, we first show that the solution of Problem 2.5 has inverse that features a banded structure. This leads us to discuss the connections of our results with that in [5].

Let $(\bar\Lambda, \bar\Theta)$ be the unique minimum point of J in $\mathcal{L}_+$ (Theorem 3.1). Then $\mathbf{\Sigma}_N^o \in \mathfrak{S}_{N,+}$ given by

$$\mathbf{\Sigma}_N^o = \left(E_n \bar\Lambda E_n^\top + \mathbf{U}_N \bar\Theta \mathbf{U}_N^\top - \bar\Theta\right)^{-1} \tag{22}$$

satisfies (10b) and (10c). Hence, it is the unique solution of the primal Problem 2.5. Since it satisfies (10c), $\mathbf{\Sigma}_N^o$ is in particular a block-circulant matrix and hence so is

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

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 (21),

$$\begin{aligned}(\mathbf{\Sigma}_N^o)^{-1} &= \pi_{\mathfrak{C}_N}((\mathbf{\Sigma}_N^o)^{-1}) \\ &= \pi_{\mathfrak{C}_N}\left(E_n \bar\Lambda E_n^\top + \mathbf{U}_N \bar\Theta \mathbf{U}_N^\top - \bar\Theta\right) \\ &= \pi_{\mathfrak{C}_N}\left(E_n \bar\Lambda E_n^\top\right).\end{aligned} \tag{23}$$

We are now ready to state and prove the main result of this section.

Theorem 4.1: *Let $\mathbf{\Sigma}_N^o$ be the maximum Gaussian entropy covariance given by (22). Then $(\mathbf{\Sigma}_N^o)^{-1}$ is a symmetric block-circulant matrix which is banded of bandwidth n. Hence the solution of Problem 2.5 may be viewed as the covariance of a stationary reciprocal process of order n defined on $\mathbb{Z}_N$.*

Proof: Let

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

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

$$\operatorname{tr}\left[\left(E_n \bar\Lambda E_n^\top - \Pi_{\bar\Lambda}\right) C\right] = \langle E_n \bar\Lambda E_n^\top - \Pi_{\bar\Lambda}, C\rangle = 0 \quad \forall C \in \mathfrak{C}_N. \tag{24}$$

Next observe that, if we write

$$C = \operatorname{Circ}\left[C_0, C_1, C_2, \dots, C_2^\top, C_1^\top\right]$$

and

$$\bar\Lambda = \begin{bmatrix} \bar\Lambda_{00} & \bar\Lambda_{01} & \dots & \dots & \bar\Lambda_{0n} \\ \bar\Lambda_{10}^\top & \bar\Lambda_{11} & \dots & & \bar\Lambda_{1n} \\ \dots & & \dots & & \dots \\ \bar\Lambda_{n0}^\top & \bar\Lambda_{n1}^\top & \dots & & \bar\Lambda_{nn} \end{bmatrix}, \qquad \bar\Lambda_{k,j} = \bar\Lambda_{j,k}^\top$$

then

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

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

$$\operatorname{tr}[\Pi_{\bar\Lambda} C] = N \operatorname{tr}[\,\Pi_0 C_0 + \Pi_1^\top C_1 + \Pi_2^\top C_2 + \dots + \Pi_2 C_2^\top + \Pi_1 C_1^\top\,].$$

Hence, the orthogonality condition (24), reads

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

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

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

while from the last equation we get $\Pi_i = 0$, forall 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 2.5, namely $\Pi_{\bar{\Lambda}} = (\mathbf{\Sigma}_N^o)^{-1}$ has a circulant block-banded structure of bandwidth n. ■

The above results can be interpreted as a particular covariance selection result in the vein of Dempster's paper; compare in particular [5, Proposition a]. In fact the results of this section substantiate also the maximum entropy principle of Dempster (Proposition 2.1). It is however important to note that none of our results follows as a particular case from Dempster's results, since [5] deals with a very unstructured setting. In particular, our result (Theorem 4.1) that the solution, $\mathbf{\Sigma}_N^o$, to our primal Problem 2.5 has a block-circulant *banded* inverse, is completely original. Its proof uses in an essential way the characterization of the solution of Problem 2.5 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, cf. [2] (an alternative proof of this property could also be constructed based on the invariance properties of the entropy functional and its strict concavity). This fact, together with Theorem 4.1 (stating that the solution, $\mathbf{\Sigma}_N^o$, to our primal Problem 2.5 has *banded* inverse) implies that Problem 2.1 is indeed equivalent to Problem 2.5.

Because of the equivalence of reciprocal AR modeling and the underlying process covariance having an inverse with a banded structure, 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. [6], [7] as representative references of a very large literature. We should stress here that band extension problems 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 [5, 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.

V. Conclusions

In the present paper, maximum likelihood identification of AR-type reciprocal models is discussed. The computation of the estimates of the matrix parameters of the model, corresponding to the nonzero blocks of the inverse of the covariance matrix, turns out to be a particular instance of a *Covariance selection problem* of the kind studied by 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.

References

[1] F. Carli, A. Ferrante, M. Pavon, and G. Picci. A maximum entropy solution of the covariance extension problem for reciprocal processes. *preprint*, submitted November 2009.

[2] F. Carli and T. T. Georgiou. On the covariance completion problem under a circulant structure. *IEEE Transactions on Automatic Control* (to appear), submitted December 2009.

[3] P. Delsarte, Y. V. Genin, and Y. G. Kamp. Orthogonal polynomial matrices on the unit circle. *IEEE Trans. Circuits Systems*, CAS 25:149–160, 1978.

[4] A. Dembo, C. Mallows, and L. Shepp. Embedding nonnegative definite Toeplitz matrices in nonnegative definite circulant matrices, with application to covariance estimation. *IEEE Trans. Information Theory*, IT-35:1206–1212, 1989.

[5] A.P. Dempster. Covariance selection. *Biometrics*, 28:157–175, 1972.

[6] H. Dym and I. Gohberg. Extension of band matrices with band inverses. *Linear Algebra and Applications*, 36:1–24, 1981.

[7] I. Gohberg, Goldberg, and M. Kaashoek. *Classes of Linear Operators vol II*. Birkhauser, Boston, 1994.

[8] Robert Grone, Charles R. Johnson, Eduardo M. Sa, and Henry Wolkowicz. Positive definite completions of partial hermitian matrices. *Linear Algebra and its Applications*, 58:109 – 124, 1984.

[9] Gray R. M. *Toeplitz and Circulant Matrices: A Review*. Stanford University, http://ee.stanford.edu/ gray/toeplitz.pdf, 2002.

[10] E. E. Tyrthyshnikov. A unifying approach to some old and new theorems on distribution and clustering. *Linear Algebra and its Applpications*, 232:1–43, 1968.

[11] P. Whittle. On the fitting of multivariate autoregressions and the approximate spectral factorization of a spectral density matrix. *Biometrica*, 50:129–134, 1963.

[12] 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 and Systems*, CAS-25:57–65, 1978.