Non-Euclidean Geometrical Aspects of the Schur and Levinson–Szegö Algorithms

François Desbouvries, *Member, IEEE*

***Abstract*—In this paper, we address non-Euclidean geometrical aspects of the Schur and Levinson–Szegö algorithms. We first show that the Lobachevski geometry is, by construction, one natural geometrical environment of these algorithms, since they necessarily make use of automorphisms of the unit disk. We next consider the algorithms in the particular context of their application to linear prediction. Then the Schur (resp., Levinson–Szegö) algorithm receives a direct (resp., polar) spherical trigonometry (ST) interpretation, which is a new feature of the classical duality of both algorithms.**

***Index Terms*—Interpolation theory, linear regression, Lobachevski geometry, partial correlation coefficients, Schur and Levinson–Szegö algorithms, spherical trigonometry (ST).**

I. Introduction

Linear prediction and interpolation is a major tool in time series analysis and in signal processing. In this context, the Schur and Levinson–Szegö algorithms compute the partial autocorrelation function of a wide-sense stationary process. As such, they have become very popular and are now described in standard signal processing textbooks (see, e.g., [35], [48]). They have found a large variety of electrical engineering applications [5], [30], [39], [40], including spectral estimation [18], circuit and network synthesis [24], geophysics [6], and speech modeling and coding (the Schur algorithm is used in the GSM European mobile telephone system [53]).

Although these algorithms are mainly known in the signal processing community as linear regression algorithms, they originally stem from different mathematical disciplines, as we now briefly recall. At the beginning of the century, Schur, Carathéodory, and Toeplitz were active in such fields as analytic function theory, Toeplitz forms, and moment problems. In 1917, Schur developed a recursive algorithm for checking whether a given function

$$s(z) = \sum_{k=0}^{\infty} s_k z^k$$

is analytic and bounded by one in the unit disk [52]. Such functions are characterized by a sequence of parameters of modulus less than one (the Schur parameters) which are computed recursively from the power series coefficients s_k by an elegant algorithm. On the other hand, Carathéodory and Toeplitz showed that

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

is analytic and has positive real part for $|z| < 1$ if and only if the Toeplitz forms $\sum_{i,\,j=0}^{n} a_i b_j^* c_{j-i}$, with $c_{-n} = c_n^*$, are positive for all n.[1] Let

$$s(z) = \frac{c(z) - c_0}{c(z) + c_0} \iff c(z) = c_0 \frac{1 + s(z)}{1 - s(z)} \qquad \text{(I.1)}$$

since $|s(z)| \leq 1$ if and only if $c(z)$ has a positive real part, the Schur algorithm implicitly enables a test of whether a Toeplitz form is positive.

On the other hand, Toeplitz forms were studied independently by Szegö, who introduced a set of orthogonal polynomials with respect to an (absolutely continuous) positive measure on the unit circle. These polynomials obey a two-terms recursion [54], [55] involving a set of parameters of modulus bounded by one, which later on were recognized to be equal to the Schur parameters [31], [32]. In the 1940s, Toeplitz forms received a revived interest in view of their natural occurrence in the Kolmogorov–Wiener prediction and interpolation theory of stationary processes (see, e.g., [34, Ch. 10], as well as the survey paper [38] and the references therein). Working on Wiener's solution of the continuous-time prediction problem, Levinson [45] proposed a fast algorithm for solving Toeplitz systems; later on, the Levinson recursions were recognized as being the recurrence relations of Szegö.

Finally, there was an intense activity in these fields beginning in the late 1970s, mainly toward the development of fast algorithms for numerical linear algebra, on the one hand, and in the domain of interpolation theory, on the other. Through these new developments and extensions, new connections with other mathematical topics and disciplines were developed, including, among others, displacement rank theory, J-lossless transfer functions, reproducing kernel Hilbert spaces, the commutant lifting problem, modern analytic function theory, and operator theory. The literature on these connections and extensions is vast; the reader may refer for instance to the papers [17], [18], [21]–[23], [33], [42] and books [3], [5], [13], [25], [27], [28] (this list is not at all exhaustive).

Manuscript received April 12, 2000; revised April 3, 2003. The material in this paper was presented in part at the International Symposium on the Mathematical Theory of Networks and Systems (MTNS), Perpignan, France, June 2000.

The author is with GET, Institut National des Télécommunications, dpt. CITI, and with CNRS UMR 5157, 91011 Evry, France (e-mail: Francois.Desbouvries@int-evry.fr).

Communicated by J. A. O'Sullivan, Associate Editor for Detection and Estimation.

Digital Object Identifier 10.1109/TIT.2003.814478

[1]Throughout this paper, $(\ )^*$ denotes complex conjugation and $(\ )^H$ Hermitian transposition.

The mathematical environment of these algorithms is thus very rich, and these various interactions have already been thoroughly investigated in many outstanding contributions. In this wealthy context, our contribution in this paper consists in exhibiting new unnoticed connections with spherical trigonometry (ST).

As far as geometry is concerned, the Lobachevski geometry was already known to be the natural geometrical environment of the Schur and Levinson–Szegö algorithms, since the core of these algorithms mainly consists in a linear fractional transformation (LFT) leaving the unit circle invariant. However, a new point of view is obtained when considering the algorithms (via positive-definite Toeplitz forms) in the particular context of their application to linear prediction. Then, up to an appropriate normalization, the Schur and Levinson–Szegö algorithms become trigonometric identities in a spherical triangle. Since the real projective 2-space $\mathbb{P}^2$ is the quotient space obtained from the sphere by identifying antipodal points, we see that the alternate non-Euclidean geometry with constant curvature (i.e., the elliptic one) is indeed another natural geometrical environment of the Schur and Levinson–Szegö algorithms as well.

Let us briefly outline the underlying mechanisms leading to this new interpretation. Let $\{X_i\}$ be real, zero-mean, square-integrable random variables, $\hat{X}_j^{1:n}$ the best linear mean-square estimate of X_j in terms of $\{X_i\}_{i=1}^n$, and $\tilde{X}_j^{1:n} = X_j - \hat{X}_j^{1:n}$ the corresponding estimation error. The partial correlation coefficient (or parcor) of X_0 and X_{n+1}, given $\{X_i\}_{i=1}^n$, is defined as

$$\rho_{0,n+1}^{1:n} = \frac{E\left(\tilde{X}_0^{1:n}\tilde{X}_{n+1}^{1:n}\right)}{\sqrt{E\left(\tilde{X}_0^{1:n}\right)^2}\sqrt{E\left(\tilde{X}_{n+1}^{1:n}\right)^2}}.$$

It is bounded by 1 in magnitude and is classically interpreted as the correlation coefficient of X_0 and X_{n+1}, once the influence of $\{X_i\}_{i=1}^n$ has been removed. In 1907, G. U. Yule [58] showed that the parcors could be computed recursively

$$\rho_{0,\,n+1}^{1:n} = \frac{\rho_{0,n+1}^{1:n-1} - \rho_{0,n}^{1:n-1}\,\rho_{n,n+1}^{1:n-1}}{\sqrt{1-(\rho_{0,n}^{1:n-1})^2}\,\sqrt{1-(\rho_{n,n+1}^{1:n-1})^2}}. \tag{I.2}$$

It happens that this well-known formula is formally equal to the fundamental ST cosine law

$$\cos A = \frac{\cos a - \cos b \cos c}{\sin b\, \sin c} \tag{I.3}$$

which gives an angle of a spherical triangle in terms of its three sides (see Fig. 1). This observation establishes a link between statistics and time-series analysis, on the one hand, and ST, on the other.

In earlier papers [19], [20], ST was shown also to admit a close connection with the topic of recursive least-squares adaptive filtering. Now, the Schur and Levinson–Szegö algorithms can be written as algebraic recursions within a covariance matrix or its inverse; due to the identification (I.2) = (I.3), they admit a connection with ST as well.

The existence of such an algebraic link with ST is not totally unexpected; in fact, trigonometric relations naturally appear when one deals with the structure of a positive-definite matrix, as we briefly recall from the following two simple examples.

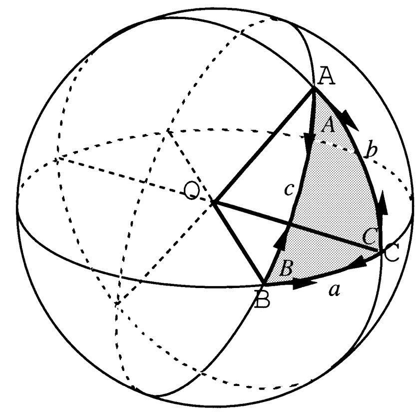

Fig. 1. The spherical triangle ABC.

Let us first consider a (Schur complement) recursive procedure for testing the positivity of a covariance matrix. Let

$$R_{0:2} = \begin{bmatrix} 1 & r_{0,1} & r_{0,2} \\ r_{0,1} & 1 & r_{1,2} \\ r_{0,2} & r_{1,2} & 1 \end{bmatrix}$$

be the covariance matrix of $\{X_i\}_{i=0}^2$. $R_{0:2}$ is positive definite if and only if the Schur complement

$$R_{0:2}^0 \stackrel{\text{def}}{=} \begin{bmatrix} 1 & r_{1,\,2} \\ r_{1,\,2} & 1 \end{bmatrix} - \begin{bmatrix} r_{0,\,1} \\ r_{0,\,2} \end{bmatrix} \begin{bmatrix} r_{0,\,1} & r_{0,\,2} \end{bmatrix}$$
$$= \begin{bmatrix} 1 - r_{0,\,1}^2 & r_{1,\,2} - r_{0,\,1}r_{0,\,2} \\ r_{1,\,2} - r_{0,\,1}r_{0,\,2} & 1 - r_{0,\,2}^2 \end{bmatrix}$$

is positive definite. Due to the normalization $r_{i,\,i} = 1$, $r_{1,\,2}$, $r_{0,\,1}$, and $r_{0,\,2}$ are lower than 1 in absolute value, and can thus be considered as the cosine of some angles, say a, b, and c. $R_{0,\,2}^0$, in turn, is positive definite if and only if $1 - r_{0,\,1}^2 > 0$, $1 - r_{0,\,2}^2 > 0$, and

$$1 - \left[\frac{r_{1,\,2} - r_{0,\,1}r_{0,\,2}}{\sqrt{1 - r_{0,\,1}^2}\,\sqrt{1 - r_{0,\,2}^2}}\right]^2 > 0$$

and this last constraint means that there exists some angle A, such that

$$\cos(A) = \frac{r_{1,\,2} - r_{0,\,1}r_{0,\,2}}{\sqrt{1 - r_{0,\,1}^2}\,\sqrt{1 - r_{0,\,2}^2}} = \frac{\cos a - \cos b\, \cos c}{\sin b\, \sin c}.$$

On the other hand, trigonometry also stems from dilation theoretic results, and more precisely from the multiplicative structure of the Kolmogorov decomposition of a positive-definite kernel [11], [12], [41], [13, Ch. 1]. For illustrative purposes let us consider the following example. Let $e_1 = [1\,0\,0]^T$, and for $|\rho| < 1$, $|\rho'| < 1$, let $\mathcal{H}(\rho,\, \rho')$ be the orthogonal lower Hessenberg matrix

$$\mathcal{H}(\rho, \rho') = \begin{bmatrix} \rho & \sqrt{1-\rho^2} & 0 \\ \sqrt{1-\rho^2} & -\rho & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \rho' & \sqrt{1-\rho'^2} \\ 0 & \sqrt{1-\rho'^2} & -\rho' \end{bmatrix}.$$

Then the identity

$$r_{0,2} = e_1^T \mathcal{H}\left(\rho_{0,1}, \rho_{0,2}^1\right) \mathcal{H}\left(\rho_{1,2}, \rho_{1,3}^2\right) e_1,$$

say, reads

$$r_{0,2} = \rho_{0,1}\rho_{1,2} + \sqrt{1-(\rho_{0,1})^2}\rho_{0,2}^1\sqrt{1-(\rho_{1,2})^2}$$

which is a particular case of (I.2) = (I.3).

In the two previous connections, positivity plays a major role. Now, the spherical nature of these trigonometric relations finds its source in recursive projection identities. In linear regression, one recognizes that the mean-square error to be minimized is a distance, so the projection theorem can be applied in the Hilbert space generated by the random variables. Introducing a new variable in the regression problem amounts to updating a projection operator, and the problem can indeed be described in terms of projections in a space generated by three vectors. But three unit-length vectors form a tetrahedron in three–dimensiona (3-D) space, and deriving projective identities in a normalized tetrahedron results in deriving trigonometric relations in the spherical triangle determined by this tetrahedron (see Fig. 1, and [20] for details).

This paper is organized as follows. Non-Euclidean hyperbolic aspects of the Schur algorithm are implicit in [10] but do not seem otherwise to be well known. Yet the Lobachevski geometry is, by construction, an essential feature of the algorithm, which deserves to be better appreciated. So we feel useful, in the context of the present paper, to begin with a brief section on this topic (which is partially of tutorial nature). More precisely, we show in Section II that Schur's layer-peeling type solution to the Carathéodory problem necessarily makes use of automorphisms of the unit disk which, on the other hand, happen to be the direct isometries of the Lobachevski plane.

The next three sections are devoted to the new geometrical interpretations in terms of ST. So in Section III, we begin with briefly recalling the general projection identities, as well as their ST counterparts, which will be used in the rest of this paper. Next, in Section IV, we relate recursive regressions within a set of $(n+1)$ random variables, algebraic manipulations in a covariance matrix or in its inverse, and ST. We show that adding (resp., removing) a new variable in the regression problem which, in terms of Schur complements on the covariance matrix (resp., on its inverse), amounts to using the quotient property [46, p. 279], corresponds in terms of ST to applying the law of cosines (resp., the polar law of cosines).

Finally, in Section V, we further assume that the random variables are taken out of a discrete time, wide-sense stationary time series, and we use the results of Sections III and IV to interpret in parallel the Schur and Levinson–Szegö algorithms in terms of ST. The Schur (resp., Levinson–Szegö) relations consist in two Schur complement recursions (in the forward and backward sense) on the original covariance matrix (resp., on its inverse), and can indeed be interpreted in dual spherical triangles. Up to an appropriate normalization, the Schur (resp., order decreasing Levinson–Szegö) recursions coincide with two coupled occurrences of the law of cosines (resp., of the polar law of cosines), and the Levinson–Szegö recursions with two coupled occurrences of the polar five-elements formula.

II. Non-Euclidean *(Hyperbolic)* Geometrical Aspects of the Schur and Levinson–Szegö Algorithms

In this section, we first briefly recall the mechanisms underlying the Schur algorithm. We next show that the choice by Schur of a recursive solution to the Carathéodory problem naturally sets the algorithm in a non-Euclidean hyperbolic environment.

A. The Schur Algorithm

The Carathéodory analytic interpolation problem consists in finding all functions s such that 1) $s(z) = \sum_{k=0}^n a_k z^k + \mathcal{O}(z^{n+1})$, and 2) $s \in \mathcal{S} \stackrel{\text{def}}{=} \{f(z)$ analytic in $|z| < 1$, and $|f(z)| \leq 1$ for $|z| < 1\}$. In 1917, Schur proposed a "layer-peeling" type algorithm [52] (i.e., in which the interpolation data are processed recursively) which we briefly recall.

Let us first consider the case where there is only one interpolation point a_0. Due to the maximum principle, the problem has no solution if $|a_0| > 1$, and admits the unique solution $s(z) = a_0$ if $|a_0| = 1$. If $|a_0| < 1$, let

$$s_1(z) = \frac{1}{z}\,\frac{s(z)-a_0}{1-a_0^* s(z)} \Longleftrightarrow s(z) = \frac{zs_1(z)+a_0}{1+a_0^* z s_1(z)}. \quad \text{(II.1)}$$

Then the key property of this transformation is that $s \in \mathcal{S} \Longleftrightarrow s_1 \in \mathcal{S}$, so that

$$\begin{cases} s \in \mathcal{S} \\ s(0) = a_0 \end{cases} \Longleftrightarrow \begin{cases} s(z) = \frac{zs_1(z)+a_0}{1+a_0^* z s_1(z)} \\ s_1 \in \mathcal{S}. \end{cases} \quad \text{(II.2)}$$

In the case of a single interpolation point a_0, (II.2) provides a parametrization of all solutions to the Carathéodory problem in terms of an arbitrary Schur function s_1. A new interpolation point a_1 can be taken into account by further restricting this set of possible functions s_1. From (II.2), we see that [$s \in \mathcal{S}$ and has interpolation constraints $(a_0 \cdots a_n)$] if and only if [$s_1 \in \mathcal{S}$ and has interpolation constraints $(a_0^1 \cdots a_{n-1}^1)$], in which the a_i^1 depend on the data a_i. In particular, $(s'(z)/1!)|_{z=0} = a_1$ if and only if $s_1(0) = a_0^1$: we are thus led back to the same metric-constrained interpolation problem, but now of order $n-1$. These considerations lead by induction to the Schur algorithm[2]

$$s_0(z) = s(z);\ \text{if}\ |s_p(0)| < 1,\ s_{p+1}(z) = \frac{1}{z}\,\frac{s_p(z)-s_p(0)}{1-s_p^*(0)s_p(z)}. \quad \text{(II.3)}$$

B. The Schur Mechanism and Automorphisms of the Unit Circle

In this section, we analyze the design of the Schur algorithm in terms of automorphisms of the unit disk. Automorphisms of a domain G are bi-holomorphic mappings of G

$$\text{Aut}(G) \stackrel{\text{def}}{=} \{f \in \mathcal{H}(G,G) \text{ s.t. } f^{-1} \text{ exists, and } f^{-1} \in \mathcal{H}(G,G)\}$$

where $\mathcal{H}(G_1, G_2)$ denotes the set of holomorphic functions of G_1 onto G_2. The Schur class $\mathcal{S}$ coincides with $\mathcal{H}(\mathbb{D}, \overline{\mathbb{D}})$, where $\mathbb{D}$ is the open unit disk and $\overline{\mathbb{D}}$ its closure.

[2] In Section II-C, we will deal with geometrical aspects of recursion (II.3). This is why the regular case only is considered in this brief review; all details of the general case can be found in [52].

The mapping (II.1) can be decomposed into two steps

$$\tilde{s}_1(z)=\frac{s(z)-a_0}{1-a_0^* s(z)} \Longleftrightarrow s(z)=\frac{\tilde{s}_1(z)+a_0}{1+a_0^*\tilde{s}_1(z)} \qquad \text{(II.4)}$$

and $s_1(z)=\frac{1}{z}\,\tilde{s}_1(z)$. Since the transform $s\mapsto\tilde{s}_1$ in (II.4) is an LFT, we begin with recalling elementary (Euclidean) properties of these mappings [26], [8], [1]. Let $\mathrm{Möb}$ denote the Möbius group $\mathrm{Möb}\stackrel{\mathrm{def}}{=}\{z\mapsto(az+b)/(cz+d)$, with $ad-bc\neq 0\}$. Then the mapping

$$\phi: GL(2,\mathbb{C})\rightarrow \mathrm{Möb}$$
$$M=\begin{bmatrix} a & b\\ c & d\end{bmatrix}\mapsto\phi_M, \qquad \text{with } \phi_M(z)=\frac{az+b}{cz+d}$$

is a group homomorphism. Since $\phi_M=\phi_{\lambda M}$ for all $\lambda\in\mathbb{C}\backslash\{0\}$, there is no loss of generality in supposing that $\det(M)=1$. From now on we shall thus restrict to $SL(2,\,\mathbb{C})$. Then the kernel of ϕ reduces to I and $-I$, and $\mathrm{Möb}$ is isomorphic to the associated quotient group

$$\mathrm{Möb}\cong(SL(2,\mathbb{C})/\pm I).$$

On the other hand,

$$(az+b)/(cz+d)=a/c+[(bc-ad)/c^2]/(z+d/c)$$

(resp., $=az/d+b/d$) if $c\neq 0$ (resp., $c=0$), so any LFT is a succession of translations, inversions, rotations, and/or homotheties. Since all these geometrical transformations preserve circles, an LFT maps any circle in the complex plane (possibly of infinite radius) into another circle (possibly of infinite radius). In particular, ϕ_M maps the unit circle $\mathbb{T}$ onto itself if and only if M belongs to the subgroup $SU(1,\,1)$ of $SL(2,\,\mathbb{C})$ consisting of Σ-unitary matrices, with $\Sigma=diag(+1,\,-1)$

$$SU(1,\,1)=\left\{M\in SL(2,\,\mathbb{C}),\ \text{s.t. } M\begin{bmatrix}1&0\\0&-1\end{bmatrix}M^H=\begin{bmatrix}1&0\\0&-1\end{bmatrix}\right\}$$
$$=\left\{\begin{bmatrix}a&b\\b^*&a^*\end{bmatrix}\ \text{with } |a|^2-|b|^2=1\right\}.$$

Let $\mathcal{N}$ be the set of such Möbius transforms; $\mathcal{N}$ is a subgroup of $\mathrm{Möb}$, and

$$\mathcal{N}\stackrel{\mathrm{def}}{=}\left\{z\mapsto\frac{az+b}{b^*z+a^*},\ \text{with } |a|^2-|b|^2=1\right\}$$
$$\cong(SU(1,1)/\pm I).$$

Finally, observe that any $\phi_M\in\mathcal{N}$ also maps the interior $\mathbb{D}$ of the unit circle onto itself (and similarly for its exterior).

Let us turn back to the Schur algorithm. Since $|a_0|<1$, we have

$$\left(s(z)\in\mathcal{H}(\mathbb{D},\overline{\mathbb{D}})\text{ and } s(0)=a_0\right)\Leftrightarrow \qquad \text{(II.5)}$$
$$\left(s(z)\in\mathcal{H}(\mathbb{D},\mathbb{D})\text{ and } s(0)=a_0\right)\Leftrightarrow \qquad \text{(II.6)}$$
$$\left(\tilde{s}_1(z)\in\mathcal{H}(\mathbb{D},\mathbb{D})\text{ and } \tilde{s}_1(0)=0\right)\Leftrightarrow \qquad \text{(II.7)}$$
$$s_1(z)\in\mathcal{H}\left(\mathbb{D},\overline{\mathbb{D}}\right) \qquad \text{(II.8)}$$

whence (II.2). Equivalence (II.5) $\Leftrightarrow$ (II.6) is due to the maximum principle and (II.7) $\Leftrightarrow$ (II.8) to the Schwarz lemma. As for (II.6) $\Leftrightarrow$ (II.7), it holds because $\tilde{s}_1=\phi_M\circ s$, where the mapping

$$\phi_M: z\mapsto\phi_M(z)=(z-a_0)/(1-a_0^*z)$$
$$=\frac{(z-a_0)/\sqrt{1-|a_0|^2}}{(1-a_0^*z)/\sqrt{1-|a_0|^2}}$$

belongs to $\mathcal{N}\subset\mathrm{Aut}\,(\mathbb{D})$. In fact, it is interesting to notice that (II.4) was the only possible choice, because, as is well known [51], [2], the automorphisms of $\mathbb{D}$ indeed coincide with the group of LFTs which leave the unit-disk invariant

$$\mathcal{N}=\mathrm{Aut}\,(\mathbb{D}).$$

C. Hyperbolic Geometry of the Schur and Levinson–Szegö Algorithms

We now turn from analytical to geometrical considerations. These are obtained naturally in the framework of the theory developed in [37], which aims at describing the holomorphic structure of a domain G of $\mathbb{C}^n$ in terms of geometric properties of the space $(G,\,d_G)$. If the distance d_G is chosen such that any holomorphic mapping of G onto itself is a contraction, then automorphisms of G are isometries with respect to d_G and can thus be interpreted (in the spirit of F. Klein) as rigid motions with respect to the geometry specified by d_G.

The Schwarz–Pick lemma [37], [2], [8], [7] provides a nice illustration of this general methodology to the present situation; as expected, we shall meet non-Euclidean hyperbolic geometry, since G reduces to the unit disk $\mathbb{D}$, which is a Euclidean model of the Lobachevski plane (see, e.g., [49]). Let $d_M(.,\,.)$ denote the Möbius distance in $\mathbb{C}$: for all $z,\,z'\in\mathbb{C}$, $d_M(z,\,z')=|(z-z')/(1-z'^*z)|$. The Schwarz–Pick lemma states that any function f belonging to $\mathcal{H}(\mathbb{D},\,\mathbb{D})$ is a contraction with respect to the Möbius distance

$$f\in\mathcal{H}(\mathbb{D},\,\mathbb{D})\Rightarrow\left|\frac{f(z)-f(z')}{1-(f(z'))^*f(z)}\right|\leq\left|\frac{z-z'}{1-z'^*z}\right|, \quad \text{for all } z,\,z'\in\mathbb{D};$$

equality holds if and only if f belongs to $\mathrm{Aut}\,(\mathbb{D})$.

Let us now turn back to the discussion at the end of Section II-B. Schur had to chose the functional $s\mapsto\tilde{s}_1$ (or, in general, $s_p\mapsto zs_{p+1}$) within $\mathrm{Aut}\,(\mathbb{D})$, and automorphisms of $\mathbb{D}$ preserve the Möbius distance d_M, and thus the Poincaré distance $d_P(.,\,.)$, with

$$d_P(z,\,z')=\log([1+d_M(z,\,z')]/[1-d_M(z,\,z')]).$$

More precisely, they are well known to coincide with the direct isometries of the Lobachevski plane $(\mathcal{H}^2,\,d_P)$ [8], [26][3]

$$\mathcal{N}=\mathrm{Aut}\,(\mathbb{D})\cong\{\text{direct isometries of } (\mathcal{H}^2,d_P)\}.$$

This geometry is thus, by construction, the natural geometrical environment of the Schur algorithm.

[3]The full isometry group of $(\mathcal{H}^2,\,d_P)$ is obtained by including the map $z\mapsto z^*$ as a generator.

Finally, let us briefly consider the Levinson–Szegö algorithm. It is not a solution to an analytic interpolation problem, but can nevertheless be rephrased (via the Schur–Cohn stability test) in the framework of Section II-A [30], and thus shares the same geometrical environment. For let $a(z) = \sum_{i=0}^{n} a_i^n z^i$, $b(z) = z^n(a(1/z^*))^*$, and $f(z) = b(z)/a(z)$. Then $f(z)$ is rational and has modulus 1 on $\mathbb{T}$. So, by the maximum modulus theorem, [$f(z)$ is analytic in $\mathbb{D}$ and $|f(z)| = 1$ on $\mathbb{T}$] if and only if [$|f(z)| \le 1$ in $\mathbb{D}$ and $|f(z)| = 1$ on $\mathbb{T}$], i.e., if and only if f is a rational bounded function of the lossless type (a Blashke product). But this can be checked via the Schur algorithm, because $f = f_0$ is a Blashke product of order n if and only if $|f_p(0)| < 1$ for $0 \le p \le n-1$ and $|f_n(0)| = 1$; the recursions coincide with the order-decreasing Levinson–Szegö algorithm.

III. Recursive Projections in Hilbert Spaces and ST

From now on, we shall deal with ST aspects of the Schur and Levinson–Szegö algorithms. In this intermediate section, we first recall some elementary projective identities. We then give an ST interpretation to Yule's parcor identity (an elementary linear regression recursion), and bring back from ST some relations among parcors which we will refer to in Sections IV and V.

A. Partial Correlation Coefficients, Recursive Projections, and Yule's Parcor Identity

Our geometrical results are based on the properties of orthogonal projectors and can thus be formalized in any Hilbert space $\mathcal{H}$. However, the natural framework in this paper is the space $\mathcal{H} = \mathbb{L}^2(\Omega, \mathcal{A}, P)$ of complex, zero-mean, square-integrable random variables defined on $(\Omega, \mathcal{A}, P)$, endowed with the scalar product $(X, Y) = E(XY^*)$.

Let $P_{\mathcal{M}}$ denote the orthogonal projector on the Hilbert space $\mathcal{H}(\mathcal{M})$ generated by $\mathcal{M}$, $P_{\mathcal{M}}^{\perp} = I - P_{\mathcal{M}}$, $\hat{A}^{\mathcal{M}}$ the projection of A onto $\mathcal{H}(\mathcal{M})$, and $\tilde{A}^{\mathcal{M}} = A - \hat{A}^{\mathcal{M}}$. For any $X \in \mathcal{H} \setminus \{0\}$, $\overline{X}$ denotes normalization to unit norm: $\overline{X} = X(X, X)^{-1/2}$. Let A and B belong to $\mathcal{H} \setminus \{0\}$ (resp., to $\mathcal{H} \setminus \{\mathcal{M}\}$). The (sometimes called total) correlation coefficient $\rho_{A,B}$ (resp., partial correlation coefficient $\rho_{A,B}^{\mathcal{M}}$) of A and B (resp., of A and B, with respect to a common subspace $\mathcal{M}$) is defined as $\rho_{A,B} = (\overline{A}, \overline{B}) = (\rho_{B,A})^*$ (resp., $\rho_{A,B}^{\mathcal{M}} = (\overline{\tilde{A}^{\mathcal{M}}}, \overline{\tilde{B}^{\mathcal{M}}}) = (\rho_{B,A}^{\mathcal{M}})^*$).

Let us now consider recursive projections. It is well known that

$$P_{\mathcal{M},A} = P_{\mathcal{M}} + P_{\tilde{A}^{\mathcal{M}}},\ P_{\mathcal{M},A}^{\perp} = P_{\mathcal{M}}^{\perp} - P_{\tilde{A}^{\mathcal{M}}} \tag{III.1}$$

where $P_{\mathcal{M},A}$, say, is the orthogonal projector onto the closed subspace generated by $\mathcal{M}$ and A. These identities are of utmost importance in recursive least squares (RLS) adaptive filtering as well as in Kalman filtering. From (III.1), we get

$$\tilde{B}^{\mathcal{M},A} = \tilde{B}^{\mathcal{M}} - \left(\tilde{B}^{\mathcal{M}}, \tilde{A}^{\mathcal{M}}\right)\left(\tilde{A}^{\mathcal{M}}, \tilde{A}^{\mathcal{M}}\right)^{-1}\tilde{A}^{\mathcal{M}} \tag{III.2}$$

which gives the useful relation

$$\frac{\left(\tilde{B}^{\mathcal{M},A}, \tilde{B}^{\mathcal{M},A}\right)}{\left(\tilde{B}^{\mathcal{M}}, \tilde{B}^{\mathcal{M}}\right)} = 1 - \left|\rho_{A,B}^{\mathcal{M}}\right|^2. \tag{III.3}$$

Using (III.3), it is easy to see that (III.2) leads to the two following relations among (unit-length) normalized projection residuals:

$$\overline{\tilde{B}^{\mathcal{M},A}} = \left(\overline{\tilde{B}^{\mathcal{M}}} - \overline{\tilde{A}^{\mathcal{M}}}\,\rho_{B,A}^{\mathcal{M}}\right)\left(1 - |\rho_{B,A}^{\mathcal{M}}|^2\right)^{-1/2} \tag{III.4}$$

$$= \overline{\tilde{A}^{\mathcal{M},B}}\left(-\rho_{B,A}^{\mathcal{M}}\right) + \overline{\tilde{B}^{\mathcal{M}}}\left(1 - |\rho_{B,A}^{\mathcal{M}}|^2\right)^{1/2}. \tag{III.5}$$

Next, from (III.1) and (III.3), we get

$$\left(\overline{\tilde{A}^{\mathcal{M},B}}, \overline{\tilde{B}^{\mathcal{M},A}}\right) = -\rho_{A,B}^{\mathcal{M}} \tag{III.6}$$

which, in the space spanned by $\overline{\tilde{A}^{\mathcal{M}}}$ and $\overline{\tilde{B}^{\mathcal{M}}}$, can be interpreted as $\cos(\pi - \theta) = -\cos(\theta)$ [20, p. 306].

In order to get ST relations, we need to consider three projection residuals. From (III.2) we have

$$\left(\tilde{C}^{\mathcal{M},A}, \tilde{B}^{\mathcal{M},A}\right) = \left(\tilde{C}^{\mathcal{M}}, \tilde{B}^{\mathcal{M}}\right) - \left(\tilde{C}^{\mathcal{M}}, \tilde{A}^{\mathcal{M}}\right)\left(\tilde{A}^{\mathcal{M}}, \tilde{A}^{\mathcal{M}}\right)^{-1}\left(\tilde{A}^{\mathcal{M}}, \tilde{B}^{\mathcal{M}}\right). \tag{III.7}$$

Dividing by $(\tilde{C}^{\mathcal{M}}, \tilde{C}^{\mathcal{M}})^{1/2}(\tilde{B}^{\mathcal{M}}, \tilde{B}^{\mathcal{M}})^{1/2}$, and using (III.3), we get

$$\rho_{C,B}^{\mathcal{M},A} = \frac{\rho_{C,B}^{\mathcal{M}} - \rho_{C,A}^{\mathcal{M}}\rho_{A,B}^{\mathcal{M}}}{\sqrt{1 - |\rho_{C,A}^{\mathcal{M}}|^2}\sqrt{1 - |\rho_{A,B}^{\mathcal{M}}|^2}} \tag{III.8}$$

which is formally equal to (I.3), up to a straightforward identification of variables.[4]

B. New Relations Among Parcors Induced by ST

We now briefly recall some ST principles (and, in particular, the duality principle), and derive some ST-like relations among parcors, which will be useful in Sections IV and V.

Three points A, B, and C on the sphere $(0, 1)$ determine the spherical triangle ABC, which by definition consists of the three arcs of great circles AB, AC, and BC obtained by intersecting the sphere and the planes OAB, OAC, and OBC (see Fig. 1). A spherical triangle has six elements: the three sides a, b, and c, and the three angles A, B, and C. The side a, say, is defined as the angle $\widehat{\mathrm{BOC}}$ and is equal to the length of the arc BC. The angle A, say, is defined as the dihedral angle between the planes OAB and OAC, and is also equal to the angle made by the tangents to the spherical triangle ABC at point A.

There are three degrees of freedom in a spherical triangle, so there cannot be more than three distinct relations among the six elements. All the ST relations can thus be derived from the three cosine laws obtained by permuting variables into (I.3), and, similarly, the relations among parcors below can all be derived from (III.8). Some of them (such as the cosine law in the

[4]This is a slight abuse of language: since sines and cosines are real valued, identification indeed holds only in the case of scalar and real parcors. However, as we will see in the next section, this trigonometric interpretation enables to hint at the existence of "ST-like" relations among parcors which, though purely algebraic, are not necessarily intuitive. These relations naturally hold in the real as well as in the complex case, and can also be extended from the scalar to the matrix case. We chose in this paper to stick to the scalar case; but since, on the other hand, the Schur and Levinson–Szegö algorithms are naturally complex data algorithms, we had to deal with complex-valued parcors and algorithms, whence this slight discrepancy (we should say that (III.8) is the complex-valued extension of a real-valued formula which is formally equal to (I.3)).

polar triangle) are already known, and in this case we only bring a geometrical interpretation, but some others (such as the sine law) seem to be new.

We first need to briefly evoke the duality principle of ST. Let $\mathtt{A}'$ be the pole (with respect to the equator passing through $\mathtt{B}$ and $\mathtt{C}$) which is in the same hemisphere as $\mathtt{A}$; $\mathtt{B}'$ and $\mathtt{C}'$ are defined similarly. The spherical triangle $\mathtt{A}'\mathtt{B}'\mathtt{C}'$ is the polar triangle of $\mathtt{ABC}$. In $\mathtt{A}'\mathtt{B}'\mathtt{C}'$, the elements a' and A', say, are equal, respectively, to $\pi - A$ and $\pi - a$ (see, e.g., [43], [44], [47]). So, for any ST formula there exists a dual relation, obtained by replacing (a, b, c, A, B, C) by $(\pi - A, \pi - B, \pi - C, \pi - a, \pi - b, \pi - c)$, respectively.

Among any four elements there exists one and only one relation. These 15 relations are the three cosine laws, the three cosine laws in the polar triangle, the three self-dual sine formulas, and the six self-dual cotangent formulas. They all have a parcor equivalent. However, there are many different relations among any five elements (or between the six), and it always seems possible to find new ones. Thus, we shall give only one of them, the five-elements formula.

1) The Cosine Law in the Polar Triangle: In the polar triangle, the cosine law reads

$$\cos a = \frac{\cos A + \cos B \cos C}{\sin B \, \sin C}. \tag{III.9}$$

Similarly, (III.8) admits the polar version

$$\rho^{\mathcal{M}}_{C,\,B} = \frac{\rho^{\mathcal{M},A}_{C,B} + \rho^{\mathcal{M},B}_{C,A}\rho^{\mathcal{M},C}_{A,B}}{\sqrt{1 - \left|\rho^{\mathcal{M},B}_{C,A}\right|^2}\sqrt{1 - \left|\rho^{\mathcal{M},C}_{A,B}\right|^2}} \tag{III.10}$$

which was already known to Yule [58, eq. (19), p. 93].

Proof: Using (III.5), we have

$$\overline{\tilde{C}^{\mathcal{M},\,A,\,B}} = \overline{\tilde{A}^{\mathcal{M},\,B,\,C}}\left(-\rho^{\mathcal{M},\,B}_{C,\,A}\right) + \overline{\tilde{C}^{\mathcal{M},\,B}}\left(1 - \left|\rho^{\mathcal{M},\,B}_{C,\,A}\right|^2\right)^{1/2}$$

$$\overline{\tilde{B}^{\mathcal{M},\,C,\,A}} = \overline{\tilde{A}^{\mathcal{M},\,B,\,C}}\left(-\rho^{\mathcal{M},\,C}_{B,\,A}\right) + \overline{\tilde{B}^{\mathcal{M},\,C}}\left(1 - \left|\rho^{\mathcal{M},\,C}_{B,\,A}\right|^2\right)^{1/2}.$$

Taking the inner product and using (III.6), we get (III.10), because from (III.4) we have

$$(\overline{\tilde{A}^{\mathcal{M},\,B,\,C}}, \overline{\tilde{C}^{\mathcal{M},\,B}}) = (\overline{\tilde{A}^{\mathcal{M},\,B,\,C}}, \overline{\tilde{B}^{\mathcal{M},\,C}}) = 0. \qquad \square$$

2) The Sine Law: The spherical triangle self-dual sine law is the following formula:

$$\frac{\sin A}{\sin a} = \frac{\sin B}{\sin b} = \frac{\sin C}{\sin c}. \tag{III.11}$$

Similarly, the following relation holds among parcors:

$$\sqrt{\frac{1 - |\rho^{\mathcal{M},A}_{B,C}|^2}{1 - |\rho^{\mathcal{M}}_{B,C}|^2}} = \sqrt{\frac{1 - |\rho^{\mathcal{M},B}_{C,A}|^2}{1 - |\rho^{\mathcal{M}}_{C,A}|^2}} = \sqrt{\frac{1 - |\rho^{\mathcal{M},C}_{A,B}|^2}{1 - |\rho^{\mathcal{M}}_{A,B}|^2}}. \tag{III.12}$$

Proof: Using (III.3) recursively, we have

$$\begin{aligned}
&\left(\tilde{C}^{\mathcal{M},\,A,\,B}, \tilde{C}^{\mathcal{M},\,A,\,B}\right) \\
&= \left(1 - \left|\rho^{\mathcal{M},\,A}_{B,\,C}\right|^2\right)\left(\tilde{C}^{\mathcal{M},\,A}, \tilde{C}^{\mathcal{M},\,A}\right) \\
&= \left(1 - \left|\rho^{\mathcal{M},\,A}_{B,\,C}\right|^2\right)\left(1 - |\rho^{\mathcal{M}}_{C,\,A}|^2\right)\left(\tilde{C}^{\mathcal{M}}, \tilde{C}^{\mathcal{M}}\right) \\
&\left(\tilde{C}^{\mathcal{M},\,A,\,B}, \tilde{C}^{\mathcal{M},\,A,\,B}\right) \\
&= \left(1 - |\rho^{\mathcal{M},\,B}_{C,\,A}|^2\right)\left(\tilde{C}^{\mathcal{M},\,B}, \tilde{C}^{\mathcal{M},\,B}\right) \\
&= \left(1 - \left|\rho^{\mathcal{M},\,B}_{C,\,A}\right|^2\right)\left(1 - |\rho^{\mathcal{M}}_{B,\,C}|^2\right)\left(\tilde{C}^{\mathcal{M}}, \tilde{C}^{\mathcal{M}}\right).
\end{aligned}$$

This leads the first equation of (III.12), in which it remains to permute variables. $\square$

3) The Cotangent Formulas: These are the six self-dual formulas obtained by permuting variables into the equation

$$\cot b \sin a = \cos C \cos a + \sin C \cot B. \tag{III.13}$$

Similarly, the following relation among parcors holds:

$$\begin{aligned}
\frac{\rho^{\mathcal{M}}_{A,C}}{\sqrt{1 - |\rho^{\mathcal{M}}_{A,C}|^2}}\sqrt{1 - |\rho^{\mathcal{M}}_{B,\,C}|^2} &= \rho^{\mathcal{M},\,C}_{A,\,B}\rho^{\mathcal{M}}_{B,\,C} \\
&+ \sqrt{1 - |\rho^{\mathcal{M},\,C}_{A,\,B}|^2}\left[\rho^{\mathcal{M},\,B}_{A,\,C} \Big/ \sqrt{1 - |\rho^{\mathcal{M},\,B}_{A,\,C}|^2}\right].
\end{aligned} \tag{III.14}$$

Proof: Using (III.8) twice, we get

$$\begin{aligned}
\rho^{\mathcal{M}}_{A,\,C} = &\left[\rho^{\mathcal{M}}_{A,\,C}\rho^{\mathcal{M}}_{C,\,B} + \sqrt{1 - |\rho^{\mathcal{M}}_{A,\,C}|^2}\rho^{\mathcal{M},\,C}_{A,\,B}\sqrt{1 - |\rho^{\mathcal{M}}_{C,\,B}|^2}\right] \\
&\cdot\rho^{\mathcal{M}}_{B,\,C} + \sqrt{1 - |\rho^{\mathcal{M}}_{A,\,B}|^2}\rho^{\mathcal{M},\,B}_{A,\,C}\sqrt{1 - |\rho^{\mathcal{M}}_{B,\,C}|^2}.
\end{aligned}$$

It remains to move the first term of the right-hand side to the left-hand side, to divide by

$$\sqrt{1 - |\rho^{\mathcal{M}}_{A,\,C}|^2}\sqrt{1 - |\rho^{\mathcal{M}}_{C,\,B}|^2}$$

and to use the sine law (III.12). $\square$

4) The Five-Elements Formula: These are the six formulas obtained by permuting variables into

$$\cos b \sin c = \sin b \cos A \cos c + \sin a \cos B. \tag{III.15}$$

The dual equations are

$$\cos B \sin C = -\sin B \cos a \cos C + \sin A \cos b. \tag{III.16}$$

Similarly, the following relations among parcors hold:

$$\begin{aligned}
\rho^{\mathcal{M}}_{C,\,A}\sqrt{1 - \left|\rho^{\mathcal{M}}_{B,\,A}\right|^2} &= \sqrt{1 - \left|\rho^{\mathcal{M}}_{C,\,A}\right|^2}\rho^{\mathcal{M},\,A}_{C,\,B}\rho^{\mathcal{M}}_{B,\,A} \\
&+ \sqrt{1 - \left|\rho^{\mathcal{M}}_{C,\,B}\right|^2}\rho^{\mathcal{M},\,B}_{C,\,A}
\end{aligned} \tag{III.17}$$

$$\rho_{C,A}^{\mathcal{M},B}\sqrt{1-\left|\rho_{B,A}^{\mathcal{M},C}\right|^2} = -\sqrt{1-\left|\rho_{C,A}^{\mathcal{M},B}\right|^2}\rho_{C,B}^{\mathcal{M}}\rho_{B,A}^{\mathcal{M},C} + \sqrt{1-|\rho_{C,B}^{\mathcal{M},A}|^2}\rho_{C,A}^{\mathcal{M}}. \quad \text{(III.18)}$$

Proof: Using (III.8) twice, we get

$$\begin{aligned}\rho_{C,B}^{\mathcal{M}}\rho_{B,A}^{\mathcal{M}}+\rho_{C,A}^{\mathcal{M}} &= \left[\rho_{C,A}^{\mathcal{M}}\rho_{A,B}^{\mathcal{M}}+\sqrt{1-\left|\rho_{C,A}^{\mathcal{M}}\right|^2}\rho_{C,B}^{\mathcal{M},A}\sqrt{1-\left|\rho_{A,B}^{\mathcal{M}}\right|^2}\right]\rho_{B,A}^{\mathcal{M}} \\ &\quad + \left[\rho_{C,B}^{\mathcal{M}}\rho_{B,A}^{\mathcal{M}}+\sqrt{1-\left|\rho_{C,B}^{\mathcal{M}}\right|^2}\rho_{C,A}^{\mathcal{M},B}\sqrt{1-\left|\rho_{B,A}^{\mathcal{M}}\right|^2}\right].\end{aligned}$$

It remains to divide by $\sqrt{1-|\rho_{B,A}^{\mathcal{M}}|^2}$. Equation (III.18) is obtained similarly from (III.10). □

IV. Schur Complements in $R_{0:N}/R_{0:n}^{-1}$ and ST

In view of Section V, we now consider Schur complementation in a covariance matrix or in its inverse, because Schur complements provide the connection between the Schur and Levinson–Szegö algorithms and ST. The reason why is that algebraically the elementary recursion with pivot $t_{k,k}$:

$$t_{i,j} \to t'_{i,j} = t_{i,j} - t_{i,k}t_{k,k}^{-1}t_{k,j}$$

reduces to a cosine law when normalized by $\sqrt{t_{i,i}}\sqrt{t_{j,j}}$.

Let $\{X_i\}_{i=0}^n$ be scalar random variables. For $p \le q$, let $X_{p:q} = [X_p \cdots X_q]^T$. In all of this section, we will assume that $\{X_i\}_{i=0}^n$ belong to $\mathbb{L}^2(\Omega, \mathcal{A}, P)$, and that the covariance matrix $R_{0:n} = E(X_{0:n}X_{0:n}^H)$ of $X_{0:n}$ is invertible. We shall focus on the Hilbert space $\mathcal{H}(X_0, \ldots, X_n)$ generated by $\{X_i\}_{i=0}^n$, which is a subspace of $\mathbb{L}^2(\Omega, \mathcal{A}, P)$.

In the sequel, the general notation $\widetilde{X_i}^{\mathcal{H}(X_{l_1},\ldots,X_{l_k})}$ of Section III-A is simplified to $\tilde{X}_i^{l_1,\ldots,l_k}$. Let $p \le r,\ s \le q$. Since we will essentially use contiguous sets of indexes (without loss of generality), we also replace

$$\tilde{X}_i^{\mathcal{H}(\{X_m\}_{p\le m\le q})},\ \tilde{X}_i^{\mathcal{H}(\{X_m\}_{p\le m\le q}^{m\ne r})}.\ \text{and}\ \tilde{X}_i^{\mathcal{H}(\{X_m\}_{p\le m\le q}^{m\ne r,\, m\ne s})}$$

respectively, by $\tilde{X}_i^{p:q}$, $\tilde{X}_i^{[p:q]\backslash r}$, and $\tilde{X}_i^{[p:q]\backslash r,s}$. Similar notations are adopted for the correlation coefficients, so that the parcor (of order $q-p$) $\rho_{X_i,X_j}^{\mathcal{H}(\{X_m\}_{p\le m\le q}^{m\ne r})}$, say, is denoted simply by $\rho_{i,j}^{[p:q]\backslash r}$. In our conventions, the order of the secondary (upper) indexes p and q is meaningful: $\tilde{X}_i^{p:q}$, $\rho_{i,j}^{p:q}$ (and later on $R_{i,j}^{p:q}$, $r_{i,j}^{p:q}$, $P_{i,j}^{p:q}$, and $p_{i,j}^{p:q}$) reduce, respectively, to X_i, $\rho_{i,j}$, $R_{i,j}$, $r_{i,j}$, $P_{i,j}$, and $p_{i,j}$ if $p > q$. In this way, the notation changes continuously from the total to the partial situation. For instance, there is no conceptual need to distinguish between total and partial correlation coefficients since a total correlation coefficient is simply a partial correlation coefficient of order 0.

In this section, we shall first recall (and slightly extend) some results giving the covariance matrix of $X_{0:n}$ (resp., its inverse) in terms of covariances of the random variables $\{X_i\}_{i=0}^n$ (resp., of the random variables $\{\tilde{X}_i^{[0:n]\backslash i}\}_{i=0}^n$). We thus get Lemmas 4.1 and 4.2, which are generalized to Theorem 4.1 by considering Schur complements in the covariance matrix and in its inverse.[5] Finally, these recursions receive an ST interpretation. We begin with the following elementary results.

Lemma 4.1: Let

$$R_{0:n} = (r_{i,j})_{i,j=0}^n$$

and

$$P_{0:n} = R_{0:n}^{-1} = (p_{i,j})_{i,j=0}^n.$$

Then, for all $i, j \in [0 \cdots n]$

$$r_{i,j} = (X_i, X_j) \quad \text{(IV.1)}$$

$$p_{i,j} = \left(\frac{\tilde{X}_i^{[0:n]\backslash i}}{\left(\tilde{X}_i^{[0:n]\backslash i}, \tilde{X}_i^{[0:n]\backslash i}\right)}, \frac{\tilde{X}_j^{[0:n]\backslash j}}{\left(\tilde{X}_j^{[0:n]\backslash j}, \tilde{X}_j^{[0:n]\backslash j}\right)}\right). \quad \text{(IV.2)}$$

Proof: Equation (IV.1) holds by definition. As for (IV.2), Whittaker noticed that $p_{i,i} = (\tilde{X}_i^{[0:n]\backslash i}, \tilde{X}_i^{[0:n]\backslash i})^{-1}$ [56, p. 143], but indeed (IV.2) holds for all i, j, as we now see. Let $\{-b_{i,j}\}_{j\ne i}$ denote the coefficients of the optimal (in the mean-square sense) linear interpolator of X_i in terms of $\{X_j\}_{j\ne i}$, and let $\tilde{X}_i^{[0:n]\backslash i}$ denote the associated estimation error

$$\tilde{X}_i^{[0:n]\backslash i} = X_i + \sum_{\substack{j=0 \\ j\ne i}}^n b_{i,j}X_j = [b_{i,0} \cdots b_{i,n}]X_{0:n}$$

where we set $b_{i,i} = 1$. The coefficients $b_{i,j}$ minimize the mean-square error $(\tilde{X}_i^{[0:n]\backslash i}, \tilde{X}_i^{[0:n]\backslash i})$ and can thus be found from the orthogonality principle: $(\tilde{X}_i^{[0:n]\backslash i}, X_j) = 0$ for all $j \in [0, \ldots, n]$, $j \ne i$. Taking successively the inner product of $\tilde{X}_i^{[0:n]\backslash i}$ with $\tilde{X}_j^{[0:n]\backslash j}$ and with $X_{0:n}^H$, we see that

$$(\tilde{X}_i^{[0:n]\backslash i}, \tilde{X}_j^{[0:n]\backslash j}) = b_{i,j}(\tilde{X}_j^{[0:n]\backslash j}, \tilde{X}_j^{[0:n]\backslash j})$$

and (as is well known) that the ith row of $R_{0:n}^{-1}$ contains, up to normalization, the terms

$$\{b_{i,j}\}_{j=0}^n: b_{i,j} = (\tilde{X}_i^{[0:n]\backslash i}, \tilde{X}_i^{[0:n]\backslash i})p_{i,j}.$$

Equation (IV.2) follows immediately. □

Lemma 4.2: Let

$$R_{0:n} = (r_{i,j})_{i,j=0}^n$$

and

$$P_{0:n} = R_{0:n}^{-1} = (p_{i,j})_{i,j=0}^n.$$

Then, for all $i, j \in [0 \cdots n]$

$$\frac{r_{i,j}}{\sqrt{r_{i,i}r_{j,j}}} = (\overline{X_i}, \overline{X_j}) = \rho_{i,j} \quad \text{(IV.3)}$$

$$\frac{p_{i,j}}{\sqrt{p_{i,i}p_{j,j}}} = \left(\overline{\tilde{X}_i^{[0:n]\backslash i}}, \overline{\tilde{X}_j^{[0:n]\backslash j}}\right) = -\rho_{i,j}^{[0:n]\backslash i,j}. \quad \text{(IV.4)}$$

Proof: Equation (IV.3) holds by definition. Equation (IV.4) is a well-known determinantal formula [57, p. 94],

[5] To the best of our knowledge, out of the six formulas (IV.1) to (IV.6), only (IV.2) and (IV.6) are original. However, (IV.4) and (IV.5) need to be recalled since they will be used in Section V.

[16, p. 306], [56, p. 143]. In the context of this paper, it is also a direct consequence of (IV.2) and of (III.6). □

We are now ready to extend Lemma 4.2. We refer to the annex for the definition and elementary properties of Schur complements. Let $R_{0:n}^{0:p}$ (resp., $P_{0:n}^{0:p}$) be the ($(p+1)$th-order) Schur complement of $R_{0:p}$ in $R_{0:n}$ (of the $(p+1)\times(p+1)$ top left corner $[I_{p+1}0]P_{0:n}[I_{p+1}0]^T$ of $P_{0:n}$ in $P_{0:n}$). The reader should notice that $P_{0:n}^{0:p}\neq(R_{0:n}^{0:p})^{-1}$. In the same way that a total correlation coefficient is a partial correlation coefficient of order zero, the original covariance matrix and its inverse are Schur complements of order zero. From this point of view, the following theorem encompasses and generalizes Lemma 4.2 (which corresponds to the particular case $p=-1$).

Theorem 4.1: Let

$$R_{0:n}=(r_{i,j})_{i,j=0}^{n}$$
$$P_{0:n}=R_{0:n}^{-1}=(p_{i,j})_{i,j=0}^{n}.$$

Let moreover

$$R_{0:n}^{0:p}=(r_{i,j}^{0:p})_{i,j=p+1}^{n}$$

(resp., $P_{0:n}^{0:p}=(p_{i,j}^{0:p})_{i,j=p+1}^{n}$) be the Schur complement of $R_{0:p}$ in $R_{0:n}$ (of the $(p+1)\times(p+1)$ top left corner of $P_{0:n}$ in $P_{0:n}$). For $p=-1$, we set $R_{0:n}^{0:p}=R_{0:n}$, $P_{0:n}^{0:p}=P_{0:n}$, $r_{i,j}^{0:p}=r_{i,j}$, $p_{i,j}^{0:p}=p_{i,j}$, and $\rho_{i,j}^{0:p}=\rho_{i,j}$. Then, for all $p\in[-1,0,\ldots,n-1]$, and for all $i,j\in[p+1\cdots n]$

$$\frac{r_{i,j}^{0:p}}{\sqrt{r_{i,i}^{0:p}r_{j,j}^{0:p}}}=\rho_{i,j}^{0:p} \tag{IV.5}$$

$$\frac{p_{i,j}^{0:p}}{\sqrt{p_{i,i}^{0:p}p_{j,j}^{0:p}}}=-\rho_{i,j}^{[p+1:n]\backslash i,j}. \tag{IV.6}$$

Proof: A proof of (IV.5) is given a few lines below. Note that (IV.5) is actually given in [4, p. 37] as the definition of the parcor $\rho_{i,j}^{0:p}$, under the assumption that the probability law of $X_{0:n}$ is Gaussian. Of course, it is a standard result that the theory of conditioning in the Gaussian case algebraically leads to the same results as the theory of linear regression in $\mathbb{L}^2(\Omega,\mathcal{A},P)$; this is because in the Gaussian case, the conditional law of $X_{p+1:n}$ given $X_{0:p}$ is Gaussian with covariance matrix $(R_{0:n}/R_{0:p})$.

On the other hand, (IV.6) is a direct application of (IV.4), after it has been observed, with the help of (A2), that

$$P_{0:n}^{0:p}=P_{p+1:n}. \tag{IV.7}$$

□

We now turn to the connection with the ST cosine laws.

Corollary 4.1: Up to normalization, an elementary (i.e., rank 1) Schur complement step on $R_{0:n}^{0:p}$ (resp., on $P_{0:n}^{0:p}$) performs the law of cosines (I.3) (resp., the polar law of cosines (III.9)): For all $p\in[-1,0,\ldots,n-2]$ and for all $i,j\in[p+2\cdots n]$

$$\rho_{i,j}^{0:p+1}=\frac{\rho_{i,j}^{0:p}-\rho_{i,p+1}^{0:p}\rho_{p+1,j}^{0:p}}{\sqrt{1-\left|\rho_{i,p+1}^{0:p}\right|^2}\sqrt{1-\left|\rho_{p+1,j}^{0:p}\right|^2}} \tag{IV.8}$$

$$\rho_{i,j}^{[p+2:n]\backslash i,j}=\frac{\rho_{i,j}^{[p+1:n]\backslash i,j}+\rho_{i,p+1}^{[p+1:n]\backslash i,p+1}\rho_{p+1,j}^{[p+1:n]\backslash p+1,j}}{\sqrt{1-\left|\rho_{i,p+1}^{[p+1:n]\backslash i,p+1}\right|^2}\sqrt{1-\left|\rho_{p+1,j}^{[p+1:n]\backslash p+1,j}\right|^2}}. \tag{IV.9}$$

Proof: We begin with (IV.8). As we now see, it happens that the Schur complementation step $R_{0:n}^{0:p}\rightarrow(R_{0:n}^{0:p}/r_{p+1,p+1}^{0:p})$ provides the loop of a mathematical induction proof of (IV.5), which we get as a by-product; since, on the other hand, this loop is indeed one of the recursive projective identities of Section III-A, the link with ST is immediate. So, let us assume that $r_{i,j}^{0:p}=(\tilde{X}_i^{0:p},\tilde{X}_j^{0:p})$ (for $p=-1$, $r_{i,j}=(X_i,X_j)$ holds by definition). Due to the quotient property (A.3)

$$(R_{0:n}^{0:p}/r_{p+1,p+1}^{0:p})=R_{0:n}^{0:p+1}.$$

This equality reads componentwise

$$r_{i,j}^{0:p+1}=\left(\tilde{X}_i^{0:p},\tilde{X}_j^{0:p}\right)-\left(\tilde{X}_i^{0:p},\tilde{X}_{p+1}^{0:p}\right)\left(\tilde{X}_{p+1}^{0:p},\tilde{X}_{p+1}^{0:p}\right)^{-1}\left(\tilde{X}_{p+1}^{0:p},\tilde{X}_j^{0:p}\right)$$

and thus

$$r_{i,j}^{0:p+1}=(\tilde{X}_i^{0:p+1},\tilde{X}_j^{0:p+1})$$

due to (III.7). Normalizing as in Section III-A we get both (IV.5) and (IV.8).

We next consider (IV.9). Similarly, due to the quotient property (A.3),

$$(P_{0:n}^{0:p}/p_{p+1,p+1}^{0:p})=P_{0:n}^{0:p+1}.$$

But this equality reads componentwise

$$p_{i,j}^{0:p}-p_{i,p+1}^{0:p}(p_{p+1,p+1}^{0:p})^{-1}p_{p+1,j}^{0:p}=p_{i,j}^{0:p+1}.$$

Dividing this equation by $\sqrt{p_{i,i}^{0:p}}\sqrt{p_{j,j}^{0:p}}$ and using (IV.6), we get

$$\rho_{i,j}^{[p+1:n]\backslash i,j}+\rho_{i,p+1}^{[p+1:n]\backslash i,p+1}\rho_{p+1,j}^{[p+1:n]\backslash p+1,j}=\sqrt{p_{i,i}^{0:p+1}/p_{i,i}^{0:p}}\,\rho_{i,j}^{[p+2:n]\backslash i,j}\sqrt{p_{j,j}^{0:p+1}/p_{j,j}^{0:p}}.$$

Remarking from (IV.2) and (IV.7) that

$$p_{i,i}^{0:p}=(\tilde{X}_i^{[p+1:n]\backslash i},\tilde{X}_i^{[p+1:n]\backslash i})^{-1}$$

and using (III.3), we get

$$\sqrt{p_{i,i}^{0:p+1}/p_{i,i}^{0:p}}=\sqrt{1-|\rho_{i,p+1}^{[p+1:n]\backslash i,p+1}|^2}$$

and thus (IV.9), which indeed is the polar cosine law (III.10) = (III.9). □

V. Non-Euclidean *(Spherical)* Geometrical Aspects of the Schur and Levinson–Szegö Algorithms

Notations are as in Section IV. From now on, we shall further assume that $[X_0X_1\cdots X_n]=[X_tX_{t-1}\cdots X_{t-n}]$, where $\{X_t,\,t\in\mathbb{Z}\}$ is a zero-mean, discrete-time, wide-sense stationary time series. As a consequence, $R_{0:n}$ is a Toeplitz matrix. For simplicity, let us denote $R_{0:n}$ by R_n and $r_{i,j}$ by r_{j-i}.

The parcors satisfy a shift-invariance property: for all $i,\ j,\ k,\ p,\ q \in \mathbb{Z},\ p \le q,\ i,\ j \notin \{p,\ldots,q\},\ \rho_{i,j}^{p:q} = \rho_{i+k,\,j+k}^{p+k:q+k}$. Among all correlation coefficients (total or partial), the function $\{\rho(p) = \rho_{0,\,p}^{1:p-1}\}_{p\in\mathbb{N}\setminus\{0\}}$ (with $\rho(1) = \rho_{0,\,1}$, as in Theorem 4.1) is the partial autocorrelation function of the process. It is well known to be in one-to-one relation with the autocorrelation function [50] and is of particular interest in signal processing.

Let us turn back to the Schur and Levinson–Szegö algorithms. In this final section, we shall use the results of Sections III and IV to propose a new interpretation of the algorithm in terms of ST.

More precisely, we shall write the common (lattice) recursions of both algorithms as two Schur complement recursions (in the forward and backward directions), but acting on the covariance matrix (in the Schur case) or on its inverse, i.e., on the covariance matrix of the normalized interpolation process (in the Levinson–Szegö case). From Section IV, the link with ST will follow immediately: up to normalization, the Schur (resp., inverse Levinson–Szegö) algorithm performs the law of cosines (resp., the polar law of cosines). This is a new feature of the classical duality of the Schur and Levinson–Szegö algorithms. As for the Levinson–Szegö algorithm, it is an implementation of the polar five-elements formula.

A. Spherical Geometry of the Schur Algorithm

The new (spherical) geometrical interpretation of the algorithm stems from the connection between the Schur algorithm and linear regression (recall from Section I that the algorithm can be used to check whether a given sequence $\{r_k\}$ is the covariance function of a wide-sense stationary process). Let us thus initialize (II.3), via (I.1), with

$$s_0(z) = (r_1 z + r_2 z^2 + \cdots)/(r_0 + r_1 z + r_2 z^2 + \cdots).$$

In this case, for all $p \ge 1$, the Schur parameter $s_p(0)$ is equal to the (partial) correlation coefficient $\rho_{0,\,p}^{1:p-1}$. It is convenient [39], [40] to write the algorithm in vector form: for $p \ge 1$,

$$\begin{bmatrix} u_{p-1}(0) & v_{p-1}(1) \\ u_{p-1}(1) & v_{p-1}(2) \\ \vdots & \vdots \\ u_{p-1}(k) & v_{p-1}(k+1) \\ \vdots & \vdots \end{bmatrix} \begin{bmatrix} 1 & -s_p(0) \\ -s_p^*(0) & 1 \end{bmatrix} = \begin{bmatrix} u_p(0) & 0 \\ u_p(1) & v_p(1) \\ \vdots & \vdots \\ u_p(k) & v_p(k) \\ \vdots & \vdots \end{bmatrix} \qquad \text{(V.1)}$$

with initialization $u_0(0) = r_0$, and $u_0(i) = v_0(i) = r_i$ for $i > 0$.

From the point of view of analytic interpolation theory (which was that of Section II-A), this pth step of the algorithm incorporates the new data r_p in the covariance extension problem. This problem is recursive and "hierarchical" by nature: given $(r_0,\ldots,r_{p-1})$ such that R_{p-1} is positive definite $(>)0$, $R_p > 0$ if and only if r_p belongs to a disk (of decreasing radius $r_0 \prod_{i=1}^{p-1}(1 - |\rho_{0,\,i}^{1:i-1}|^2)$), the center of which depends on $(r_0,\ldots,r_{p-1})$. So for all $k \ge 0$, the row number k of (V.1) integrates the contribution of the correlation lag r_p in the subsequent (possible) compatibility of r_{p+k} with $(r_0,\ldots,r_{p+k-1})$. In particular, the row number zero tells whether r_p is compatible with the data $(r_0,\ldots,r_{p-1})$ via the following test: assuming that $R_{p-1} > 0$, $R_p > 0$ if and only if

$$|s_p(0) = v_{p-1}(1)/u_{p-1}(0)| < 1.$$

This progressive incorporation of the constraints $r_2,\ldots,r_p,\ldots$ in the analytic interpolation problem corresponds to the progressive incorporation of the random variables $X_{t-1},\ldots,X_{t-p+1},\ldots$ in the linear prediction problem, and thus to the progressive updating of the associated projection operator (this, of course, is nothing but the classical lattice or Gram–Schmidt interpretation of the Schur algorithm [29]). To see this, let us rewrite the Schur algorithm in terms of projective identities. It is easily seen (by induction) that for $k \ge 0$, the two recursions of the row number k of (V.1) are two coupled occurrences of the same identity (III.7): see (V.2) at the bottom of the page. Since all these quantities are covariances of estimation errors, they reduce to parcors when appropriately normalized; so a connection of the recursive equations (V.2) with ST is expected.

In fact, both equations are easily seen to be Schur complement recursions in

$$R_{0:n}^{1:p-1} \stackrel{\text{def}}{=} ((\tilde{X}_{t-i}^{t-p+1:t-1},\ \tilde{X}_{t-j}^{t-p+1:t-1}))_{i,\,j=0}^{n}.$$

These two Schur complementation steps correspond to augmenting the set of variables $\{X_{t-i}\}_{i=1}^{p-1}$ in the projective space in its two (contiguous) opposite directions: the forward X_t

$$\left[\left(\tilde{X}_{t-p}^{t-p+1:t-1}, \tilde{X}_{t-p-k}^{t-p+1:t-1}\right)\left(\tilde{X}_{t}^{t-p+1:t-1}, \tilde{X}_{t-p-k}^{t-p+1:t-1}\right)\right] \times \begin{bmatrix} 1 & -\frac{\left(\tilde{X}_{t}^{t-p+1:t-1}, \tilde{X}_{t-p}^{t-p+1:t-1}\right)}{\left(\tilde{X}_{t-p}^{t-p+1:t-1}, \tilde{X}_{t-p}^{t-p+1:t-1}\right)} \\ -\frac{\left(\tilde{X}_{t-p}^{t-p+1:t-1}, \tilde{X}_{t}^{t-p+1:t-1}\right)}{\left(\tilde{X}_{t}^{t-p+1:t-1}, \tilde{X}_{t}^{t-p+1:t-1}\right)} & 1 \end{bmatrix}$$
$$= \left[\left(\tilde{X}_{t-p}^{t-p+1:t},\ \tilde{X}_{t-p-k}^{t-p+1:t}\right)\left(\tilde{X}_{t}^{t-p:t-1},\ \tilde{X}_{t-p-k}^{t-p:t-1}\right)\right]. \qquad \text{(V.2)}$$

and the backward X_{t-p}. Because of stationarity, the resulting quantities still are covariances of estimation residuals with respect to the same subspace, because the right-hand side of (V.2) also reads

$$[(\tilde{X}_{t-p-1}^{t-p:t-1}, \tilde{X}_{t-p-k-1}^{t-p:t-1})(\tilde{X}_{t}^{t-p:t-1}, \tilde{X}_{t-p-k}^{t-p:t-1})]$$

and the two coefficients in the transformation matrix reduce to $-\rho_{0,p}^{1:p-1}$ and $-\left(\rho_{0,p}^{1:p-1}\right)^*$. From the discussion in section Section III, the link with ST is immediate

Propostion 5.1: Up to normalization, an elementary step of the Schur algorithm performs two coupled occurrences of the law of cosines: for all $p \geq 1$, and for all $k \geq 0$,

$$[\rho_{p,p+k}^{1:p-1}, \rho_{0,p+k}^{1:p-1}] \begin{bmatrix} \frac{1}{\sqrt{1-|\rho_{p,0}^{1:p-1}|^2}} & \frac{-\rho_{0,p}^{1:p-1}}{\sqrt{1-|\rho_{0,p}^{1:p-1}|^2}} \\ \frac{-\rho_{p,0}^{1:p-1}}{\sqrt{1-|\rho_{p,0}^{1:p-1}|^2}} & \frac{1}{\sqrt{1-|\rho_{0,p}^{1:p-1}|^2}} \end{bmatrix} = \left[\rho_{p,p+k}^{0:p-1}\sqrt{1-\left|\rho_{0,p+k}^{1:p-1}\right|^2}, \rho_{0,p+k}^{1:p}\sqrt{1-\left|\rho_{p,p+k}^{1:p-1}\right|^2}\right]. \tag{V.3}$$

Proof: Divide (V.2) by

$$(\tilde{X}_{t}^{t-p+1:t-1}, \tilde{X}_{t}^{t-p+1:t-1})^{1/2}(\tilde{X}_{t-p-k}^{t-p+1:t-1}, \tilde{X}_{t-p-k}^{t-p+1:t-1})^{1/2}$$

which is equal to

$$(\tilde{X}_{t-p}^{t-p+1:t-1}, \tilde{X}_{t-p}^{t-p+1:t-1})^{1/2}(\tilde{X}_{t-p-k}^{t-p+1:t-1}, \tilde{X}_{t-p-k}^{t-p+1:t-1})^{1/2}$$

and use (III.3). □

B. Spherical Geometry of the Levinson–Szegö Algorithm

We now turn to the spherical geometry of the Levinson–Szegö algorithm. Recall from Theorem 4.1 that successive Schur complements in $R_{0:n}$ (resp., in $R_{0:n}^{-1}$) correspond to an increase (resp., a reduction) in the number of variables in the regression problem. So, as was already the case at the end of Section II, the comparison with the Schur algorithm indeed proves easier when dealing with order-decreasing recursions.

Let us introduce the forward jth-order linear prediction coefficients $\{-a_i^j\}_{i=1}^j$ by

$$\tilde{X}_t^{t-j:t-1} = \sum_{i=0}^{j} a_i^j X_{t-i}$$

with $a_0^j = 1$. From the Wiener–Hopf equations and Theorem 4.1, we get

$$a_i^j = (\tilde{X}_{t-i}^{[t-j:t]\setminus t-i}, \tilde{X}_t^{[t-j:t]\setminus t})/(\tilde{X}_{t-i}^{[t-j:t]\setminus t-i}, \tilde{X}_{t-i}^{[t-j:t]\setminus t-i}).$$

So the order-decreasing Levinson–Szegö recursions read as shown in (V.4) at the bottom of the page [6] (the first equation is valid for $1 \leq k \leq p$, with $p \geq 2$, and the second for $0 \leq k \leq p-1$ with $p \geq 2$).

These equations are Schur complement recursions in $R_{0:p}^{-1}$; they correspond to reducing the set of variables $\{X_{t-i}\}_{i=0}^p$ in the projective space in its two (extremum) opposite directions: the forward one X_t and the backward one X_{t-p}. From the discussion in Section IV, we thus expect that appropriate normalization of the covariances of the estimation errors will reduce (V.4) to some ST polar law.

This hint is enforced when looking at the random variables in the left-hand side of (V.2) and (V.4). Let

$$\mathcal{Z} = (\{X_{t-i}\}_{0 \leq i \leq p}, X_{t-m})$$
$$(A, B, C) = (X_t, X_{t-p}, X_{t-m})$$

and

$$\mathcal{M} = \mathcal{H}(\mathcal{Z} \setminus \{X_t, X_{t-p}, X_{t-m}\}).$$

So

$$\mathcal{M} = \mathcal{H}(\{X_{t-i}\}_{1 \leq i \leq p-1}), \qquad \text{if } m \geq p$$

and

$$\mathcal{M} = \mathcal{H}(\{X_{t-i}\}_{1 \leq i \leq p-1}^{i \neq m}), \qquad \text{if } 1 \leq m \leq p-1.$$

Then (V.2) can be visualized as projective identities within the tetrahedron $(\tilde{A}^{\mathcal{M}}, \tilde{B}^{\mathcal{M}}, \tilde{C}^{\mathcal{M}})$, and (V.4) as projective identities within the tetrahedron $(\tilde{A}^{\mathcal{M}, B, C}, \tilde{B}^{\mathcal{M}, C, A}, \tilde{C}^{\mathcal{M}, A, B})$,

[6] In this subsection (as well as in Section V-A), we are only interested in the real core of the algorithm, i.e., in (V.4). This is the reason why we do not talk of the way $\rho_{0:p}^{1:p-1}$ is computed.

$$\Bigg[\underbrace{\frac{\left(\tilde{X}_{t-k}^{[t-p:t]\setminus t-k}, \tilde{X}_t^{t-p:t-1}\right)}{\left(\tilde{X}_{t-k}^{[t-p:t]\setminus t-k}, \tilde{X}_{t-k}^{[t-p:t]\setminus t-k}\right)}}_{a_k^p}, \underbrace{\frac{\left(\tilde{X}_{t-k}^{[t-p:t]\setminus t-k}, \tilde{X}_{t-p}^{t-p+1:t}\right)}{\left(\tilde{X}_{t-k}^{[t-p:t]\setminus t-k}, \tilde{X}_{t-k}^{[t-p:t]\setminus t-k}\right)}}_{(a_{p-k}^p)^*}\Bigg] \times \begin{bmatrix} \frac{1}{1-|\rho_{p,0}^{1:p-1}|^2} & \frac{\rho_{0,p}^{1:p-1}}{1-|\rho_{0,p}^{1:p-1}|^2} \\ \frac{\rho_{p,0}^{1:p-1}}{1-|\rho_{p,0}^{1:p-1}|^2} & \frac{1}{1-|\rho_{0,p}^{1:p-1}|^2} \end{bmatrix}$$
$$= \Bigg[\underbrace{\frac{\left(\tilde{X}_{t-k}^{[t-p+1:t]\setminus t-k}, \tilde{X}_t^{t-p+1:t-1}\right)}{\left(\tilde{X}_{t-k}^{[t-p+1:t]\setminus t-k}, \tilde{X}_{t-k}^{[t-p+1:t]\setminus t-k}\right)}}_{a_k^{p-1}}, \underbrace{\frac{\left(\tilde{X}_{t-k}^{[t-p:t-1]\setminus t-k}, \tilde{X}_{t-p}^{t-p+1:t-1}\right)}{\left(\tilde{X}_{t-k}^{[t-p:t-1]\setminus t-k}, \tilde{X}_{t-k}^{[t-p:t-1]\setminus t-k}\right)}}_{(a_{p-k}^{p-1})^*}\Bigg]. \tag{V.4}$$

which can be shown [20] to be the polar tetrahedron of $(\overline{\tilde{A}^{\mathcal{M}}}, \overline{\tilde{B}^{\mathcal{M}}}, \overline{\tilde{C}^{\mathcal{M}}})$.

Proposition 5.2: Up to normalization, an elementary step of the order-decreasing Levinson–Szegö algorithm (resp., of the Levinson–Szegö algorithm) performs two coupled occurrences of the polar law of cosines (resp., of the polar five-elements formula): for all $p \geq 2$ and for all $1 \leq k \leq p-1$

$$\left[\rho_{k,0}^{[0:p]\backslash k,0}, \rho_{k,p}^{[0:p]\backslash k,p}\right]\begin{bmatrix}\frac{1}{\sqrt{1-\left|\rho_{p,0}^{[0:p]\backslash p,0}\right|^2}} & \frac{\rho_{0,p}^{[0:p]\backslash 0,p}}{\sqrt{1-\left|\rho_{0,p}^{[0:p]\backslash 0,p}\right|^2}} \\ \frac{\rho_{p,0}^{[0:p]\backslash p,0}}{\sqrt{1-\left|\rho_{p,0}^{[0:p]\backslash p,0}\right|^2}} & \frac{1}{\sqrt{1-\left|\rho_{0,p}^{[0:p]\backslash 0,p}\right|^2}}\end{bmatrix}$$
$$= \left[\rho_{k,0}^{[0:p]\backslash k,0,p}\sqrt{1-\left|\rho_{k,p}^{[0:p]\backslash k,p}\right|^2}, \rho_{k,p}^{[0:p]\backslash k,p,0}\sqrt{1-|\rho_{k,0}^{[0:p]\backslash k,0}|^2}\right] \tag{V.5}$$

and

$$\left[\rho_{k,0}^{[0:p]\backslash k,0,p}\sqrt{1-\left|\rho_{k,p}^{[0:p]\backslash k,p}\right|^2}, \rho_{k,p}^{[0:p]\backslash k,p,0}\sqrt{1-\left|\rho_{k,0}^{[0:p]\backslash k,0}\right|^2}\right]$$
$$\times\begin{bmatrix}\frac{1}{\sqrt{1-\left|\rho_{p,0}^{[0:p]\backslash p,0}\right|^2}} & \frac{-\rho_{0,p}^{[0:p]\backslash 0,p}}{\sqrt{1-\left|\rho_{0,p}^{[0:p]\backslash 0,p}\right|^2}} \\ \frac{-\rho_{p,0}^{[0:p]\backslash p,0}}{\sqrt{1-\left|\rho_{p,0}^{[0:p]\backslash p,0}\right|^2}} & \frac{1}{\sqrt{1-\left|\rho_{0,p}^{[0:p]\backslash 0,p}\right|^2}}\end{bmatrix}$$
$$= \left[\rho_{k,0}^{[0:p]\backslash k,0}, \rho_{k,p}^{[0:p]\backslash k,p}\right]. \tag{V.6}$$

Proof: Using (III.6), (V.4) can be rewritten as

$$\left[\frac{\left(\tilde{X}_t^{[t-p:t]\backslash t}, \tilde{X}_t^{[t-p:t]\backslash t}\right)^{1/2}}{\left(\tilde{X}_{t-k}^{[t-p:t]\backslash t-k}, \tilde{X}_{t-k}^{[t-p:t]\backslash t-k}\right)^{1/2}}\left(-\rho_{k,0}^{[0:p]\backslash k,0}\right), \frac{\left(\tilde{X}_{t-p}^{[t-p:t]\backslash t-p}, \tilde{X}_{t-p}^{[t-p:t]\backslash t-p}\right)^{1/2}}{\left(\tilde{X}_{t-k}^{[t-p:t]\backslash t-k}, \tilde{X}_{t-k}^{[t-p:t]\backslash t-k}\right)^{1/2}}\left(-\rho_{k,p}^{[0:p]\backslash k,p}\right)\right]$$
$$\times\begin{bmatrix}\frac{1}{1-\left|\rho_{p,0}^{[0:p]\backslash p,0}\right|^2} & \frac{\rho_{0,p}^{[0:p]\backslash 0,p}}{1-\left|\rho_{0,p}^{[0:p]\backslash 0,p}\right|^2} \\ \frac{\rho_{p,0}^{[0:p]\backslash p,0}}{1-\left|\rho_{p,0}^{[0:p]\backslash p,0}\right|^2} & \frac{1}{1-\left|\rho_{0,p}^{[0:p]\backslash 0,p}\right|^2}\end{bmatrix}$$
$$= \left[\frac{\left(\tilde{X}_t^{[t-p+1:t]\backslash t}, \tilde{X}_t^{[t-p+1:t]\backslash t}\right)^{1/2}}{\left(\tilde{X}_{t-k}^{[t-p+1:t]\backslash t-k}, \tilde{X}_{t-k}^{[t-p+1:t]\backslash t-k}\right)^{1/2}}\left(-\rho_{k,0}^{[0:p]\backslash k,0,p}\right), \frac{\left(\tilde{X}_{t-p}^{[t-p:t-1]\backslash t-p}, \tilde{X}_{t-p}^{[t-p:t-1]\backslash t-p}\right)^{1/2}}{\left(\tilde{X}_{t-k}^{[t-p:t-1]\backslash t-k}, \tilde{X}_{t-k}^{[t-p:t-1]\backslash t-k}\right)^{1/2}}\left(-\rho_{k,p}^{[0:p]\backslash k,p,0}\right)\right].$$

Next divide by

$(\tilde{X}_t^{t-p+1:t-1}, \tilde{X}_t^{t-p+1:t-1})^{1/2} \big/ (\tilde{X}_{t-k}^{[t-p:t]\backslash t-k}, \tilde{X}_{t-k}^{[t-p:t]\backslash t-k})^{1/2}$

which is equal to

$(\tilde{X}_{t-p}^{t-p+1:t-1}, \tilde{X}_{t-p}^{t-p+1:t-1})^{1/2} \big/ (\tilde{X}_{t-k}^{[t-p:t]\backslash t-k}, \tilde{X}_{t-k}^{[t-p:t]\backslash t-k})^{1/2}$.

Using (III.3), we get (V.5) = (V.6). □

VI. Conclusion

In this paper, we addressed non-Euclidean geometrical aspects of the Schur and Levinson–Szegö algorithms. We showed that the Lobachevski geometry is, by construction, the natural geometrical environment of these algorithms, since they call for automorphisms of the unit disk. By considering the algorithms in the particular context of their application to linear prediction, we next gave them a new interpretation in terms of ST. The role of Schur complementation in linear regression analysis was emphasized, because of the natural link between this basic algebraic mechanism and the ST cosine laws. Finally, the Schur (resp., Levinson–Szegö) algorithm received a direct (resp., polar) ST interpretation, which is a new feature of the classical duality of both algorithms.

Finally, let us briefly mention that these interpretations provide the algorithms with structural constraints of a geometrical nature. The Lobachevski invariants are the Poincaré distance and the cross ratio (because of the use of LFT), and those of ST are expressed by the relations among parcors which were derived in Section III-B. These constraints could prove useful in the design of practical algorithms.

Appendix
Some Results on Schur Complements

In this appendix, we briefly recall some well-known results [9], [14], [46] on Schur complements. Let the matrix M be partitioned as

$$M_{(n+1)\times(n+1)} = \begin{bmatrix} A_{(p+1)\times(p+1)} & B \\ C & D\end{bmatrix}. \tag{A1}$$

Then the Schur complements (M/A) of A in M and (M/D) of D in M, if they exist, are defined, respectively, as $(M/A) = D - CA^{-1}B$ and $(M/D) = A - BD^{-1}C$ (these definitions can obviously be generalized to any pivot or block pivot). Schur complements appear in particular when computing the inverse of a partitioned matrix

$$M^{-1} = \begin{bmatrix} (M/D)^{-1} & -(M/D)^{-1}BD^{-1} \\ -D^{-1}C(M/D)^{-1} & D^{-1} + D^{-1}C(M/D)^{-1}BD^{-1}\end{bmatrix}. \tag{A2}$$

It is well known [15] that Schur complements can be obtained recursively: if A in (A1) is itself partitioned as

$$A = \begin{bmatrix} E & F \\ G & H\end{bmatrix}$$

then the "quotient formula" holds

$$(M/A) = ((M/E)/(A/E)). \tag{A3}$$

References

[1] L. V. Ahlfors, *Complex Analysis*, 3rd ed. New York: McGraw-Hill, 1979.

[2] ——, *Conformal Invariants—Topics in Geometric Function Theory*. New York: McGraw, 1973.

[3] D. Alpay, "The Schur algorithm, reproducing kernel Hilbert spaces, and system theory," in *SMF/AMS Texts and Monographs*. Providence, RI: Amer. Math. Soc., 2001, vol. 5.

[4] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, 2nd ed. New York: Wiley, 1984.

[5] M. Bakonyi and T. Constantinescu, *Schur's Algorithm and Several Applications*, ser. Pitman Research Notes in Mathematics no. 261. White Plains, NY: Longman, 1992.

[6] A. Bruckstein and T. Kailath, "An inverse scattering framework for several problems in signal processing," *IEEE ASSP Mag.*, vol. 4, pp. 6–20, Jan. 1987.

[7] R. B. Burckel, *An Introduction to Classical Complex Analysis*. Basel, Switzerland: Birkäuser-Verlag, 1979.

[8] C. Carathéodory, *Theory of Functions of a Complex Variable*. New York: Chelsea, 1954.

[9] D. Carlson, "What are Schur complements, anyway?," *Linear Alg. Its Applic.*, vol. 74, pp. 257–275, 1986.

[10] C. Chang and T. T. Georgiou, "Geometric aspects of the Carathéodory extension problem," *Linear Alg. Its Applic.*, vol. 203-04, pp. 209–251, 1994.

[11] T. Constantinescu, "On the structure of the Naimark dilation," *J. Operator Theory*, vol. 12, pp. 159–175, 1984.

[12] ——, "Schur analysis of positive block-matrices," in *I. Schur Methods in Operator Theory and Signal Processing*. ser. Operator Theory: Advances and Applications, I. Gohberg, Ed. Basel, Switzerland: Birkäuser-Verlag, 1986, vol. 18, pp. 191–206.

[13] ——, "Schur parameters, factorization and dilation problems," in *Operator Theory: Advances and Applications*. Basel, Switzerland: Birkäuser-Verlag, 1991, vol. 82.

[14] R. W. Cottle, "Manifestations of the Schur complement," *Linear Alg. Its Applic.*, vol. 8, pp. 189–211, 1974.

[15] D. E. Crabtree and E. V. Haynsworth, "An identity for the Schur complement of a matrix," in *Proc. Amer. Math. Soc.*, vol. 22, 1969, pp. 364–366.

[16] H. Cramer, *Mathematical Methods of Statistics*, 13th ed. Princeton, NJ: Princeton Univ. Press, 1974.

[17] P. Delsarte, Y. Genin, and Y. Kamp, "Orthogonal polynomial matrices on the unit circle," *IEEE Trans. Circuits Syst.*, vol. CAS-25, pp. 149–160, 1978.

[18] P. Delsarte, Y. Genin, Y. Kamp, and P. Van Dooren, "Speech modeling and the trigonometric moment problem," *Philips J. Res.*, vol. 37, no. 5/6, pp. 277–292, 1982.

[19] F. Desbouvries, "Spherical trigonometry, Yule's PARCOR identity and FRLS fully-normalized lattice," *IEEE Trans. Signal Processing*, vol. 44, pp. 427–430, Feb. 1996.

[20] ——, "Trigonométrie sphérique, identité de Yule entre PARCOR's et algorithmes MCRR," *Traitement du Signal*, vol. 13, no. 4, pp. 303–318, Dec. 1996.

[21] P. Dewilde and H. Dym, "Lossless chain scattering matrices and optimum linear prediction: The vector case," *Int. J. Circuit Theory Appl.*, vol. 9, pp. 135–175, 1981.

[22] ——, "Schur recursions, error formulas, and convergence of rational estimators for stationary stochastic sequences," *IEEE Trans. Inform. Theory*, vol. IT-27, pp. 446–461, July 1981.

[23] ——, "Lossless inverse scattering, digital filters, and estimation theory," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 644–661, July 1984.

[24] P. Dewilde, A. C. Vieira, and T. Kailath, "On a generalized Szegö-Levinson realization algorithm for optimal linear predictors on a network synthesis approach," *IEEE Trans. Circuits Syst*,, vol. CAS-25, pp. 663–675, Sept. 1978.

[25] V. K. Dubovoj, B. Fritzsche, and B. Kirstein, *Matricial Version of the Classical Schur Problem*, ser. Texte zur Mathematik. Leipzig, Germany: Teubner, 1992, vol. 129.

[26] B. A. Dubrovin, A. T. Fomenko, and S. P. Novikov, *Modern Geometry—Methods and Applications. Part I*. New York: Springer-Verlag, 1993.

[27] H. Dym, "J contractive matrix functions, reproducing kernel Hilbert spaces and interpolation," in *Coll. CBMS Regional Conference Series in Mathematics*. Providence, RI: AMS, 1989, vol. 71.

[28] C. Foiaş and A. E. Frazho, "The Commutant lifting approach to interpolation problems," in *Operator Theory: Advances and Applications*. Basel, Switzerland: Birkäuser-Verlag, 1990, vol. 44.

[29] B. Friedlander, "Lattice filters for adaptive processing," *Proc. IEEE*, vol. 70, pp. 829–868, Aug. 1982.

[30] Y. Genin, "An introduction to the modern theory of positive functions and some of its today applications to signal processing, circuits and systems problems," in *Proc. ECCTD Conf.*, Paris, France, 1987, pp. 195–233.

[31] J. Geronimus, "On polynomials orthogonal on the circle, on trigonometric moment-problem and on allied Carathéodory and Schur functions," *C.R. (Dokl.) Acad. Sci. U.S.S.R. (N.S.)*, vol. 39, pp. 291–295, 1943.

[32] ——, "On polynomials orthogonal on the circle, on trigonometric moment-problem and on allied Carathéodory and Schur functions," *Mat. Sbornik*, vol. 57, pp. 99–130, 1944.

[33] I. Gohberg, Ed., "I. Schur methods in operator theory and signal processing," in *Operator Theory: Advances and Applications*. Basel, Switzerland: Birkhäuser-Verlag, 1986, vol. 18.

[34] U. Grenander and G. Szegö, *Toeplitz Forms and Their Applications*. Berkeley, CA: Univ. California Press, 1958.

[35] S. Haykin, *Modern Filters*. New York: Macmillan , 1990.

[36] J. W. Helton, "Non-Euclidean functional analysis and electronics," *Bull. (New Series) Amer. Math. Soc.*, vol. 7, no. 1, pp. 1–64, July 1982.

[37] M. Jarnicki and P. Pflug, *Invariant Distances and Metrics in Complex Analysis*. Berlin, Germany: Walter de Gruyter, 1993.

[38] T. Kailath, "A view of three decades of linear filtering theory," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 146–181, Mar. 1974.

[39] ——, "A theorem of I. Schur and its impact on modern signal processing," in *Operator Theory: Advances and Applications*, I. Gohberg, Ed. Basel, Switzerland: Birkäuser-Verlag, 1986, vol. 18, pp. 9–30.

[40] ——, "Signal processing applications of some moment problems," *Proc. Symp. Appl. Math.*, vol. 37, pp. 71–109, 1987.

[41] T. Kailath and A. M. Bruckstein, "Naimark dilations, state-space generators and transmission lines," in *Operator Theory: Advances and Applications*. Basel, Switzerland: Birkäuser-Verlag, 1986, vol. 17, pp. 173–186.

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

[43] L. M. Kells, W. F. Kerns, and J. R. Bland, *Plane and Spherical Trigonometry*. New York: McGraw-Hill, 1951.

[44] E. G. Kogbeliantz, *Fundamentals of Mathematics From an Advanced Viewpoint*. New York: Gordon and Breach, 1969, vol. 4, Solid Geometry and Spherical Trigonometry.

[45] N. Levinson, "The Wiener r. m. s. (root mean square) error criterion in filter design and prediction," *J. Math. and Phys.*, vol. 25.4, pp. 261–278, 1947.

[46] D. V. Ouelette, "Schur complements and statistics," *Linear Alg. Its Applic.*, vol. 36, pp. 187–295, 1981.

[47] G. Papelier, *Éléments de Trigonométrie Sphérique*, 3rd ed. Paris, France: Vuibert, 1956.

[48] B. Porat, *Digital Processing of Random Signals*. Englewood Cliffs, NJ: Prentice-Hall, 1994.

[49] J. G. Ratcliffe, *Foundations of Hyperbolic Manifolds*. New York: Springer-Verlag, 1994.

[50] F. L. Ramsey, "Characterization of the partial autocorrelation function," *Ann. Statist.*, vol. 2, no. 6, pp. 1296–1301, 1974.

[51] R. Remmert, *Theory of Complex Functions*. New York: Springer-Verlag, 1991.

[52] I. Schur, "Über Potenzreihen, die im Innern des Einheitskreises beschränkt sind," *J. Reine und Angewandte Math.*, vol. 147, pp. 205–232, 1917; and vol. 148, pp. 122–145, 1918. English translation: "On power series which are bounded in the interior of the unit circle," in *Operator Theory, Advances and Applications*, vol. 18, I. Gohberg, Ed. Basel, Switzerland: Birkhäuser-Verlag, 1986, pt. I, pp. 31–59; and pt. II, pp. 61–88..

[53] R. Steele, *Mobile Radio Communications*. London, U.K.: Pentech, 1992.

[54] G. Szegö, "Beiträge zur Theorie der Toeplitzschen Formen, I, II," *Math. Z*, vol. 6, pp. 167–202, 1920. and vol. 9, pp. 167–190, 1921; also in *Collected Papers*, vol. 1, pp. 237–272 and 279–302. Boston, MA: Birkhäuser, 1982.

[55] ——, *Orthogonal Polynomials*. Providence, RI: Amer. Math. Soc., 1939. 4th edition with revisions: Providence, RI: Amer. Math. Soc., 1975.

[56] J. Whittaker, *Graphical Models in Applied Multivariate Statistics*. Chichester, U.K.: Wiley, 1990.

[57] S. S. Wilks, *Mathematical Statistics*. New York: Wiley, 1962.

[58] G. U. Yule, "On the theory of correlation for any number of variables, treated by a new system of notations," in *Proc. Roy. Soc.*, vol. 79A, 1907, pp. 182–193. also in *Statistical Papers of George Udny Yule*, London, U.K.: Griffin, pp. 85–96, 1978.