

On signal and extraneous roots in Singular Spectrum Analysis

Konstantin Usevich
 Department of Mathematics and Mechanics,
 St.Petersburg State University, Russia
 E-mail: konstantin.usevich@statmod.ru

June 18, 2010

Abstract

In the present paper we study properties of roots of characteristic polynomials for the linear recurrent formulae (LRF) that govern time series. We also investigate how the values of these roots affect Singular Spectrum Analysis implications, in what concerns separation of components, SSA forecasting and related signal parameter estimation methods. The roots of the characteristic polynomial for an LRF comprise the signal roots, which determine the structure of the time series, and extraneous roots. We show how the separability of two time series can be characterized in terms of their signal roots. All possible cases of exact separability are enumerated. We also examine properties of extraneous roots of the LRF used in SSA forecasting algorithms, which is equivalent to the Min-Norm vector in subspace-based estimation methods. We apply recent theoretical results for orthogonal polynomials on the unit circle, which enable us to precisely describe the asymptotic distribution of extraneous roots relative to the position of the signal roots.

Keywords: Singular Spectrum Analysis; SSA; separability; linear recurrent formula; LRF; continuation; extraneous roots; min-norm; subspace methods; orthogonal polynomials on the unit circle;

1 Introduction

In the theory of Singular Spectrum Analysis [11, 17] the time series that are governed by a *linear recurrent formula* (LRF) are of great importance. For these time series $F_N = (f_0, \dots, f_{N-1})$ there exist coefficients a_k such that the relation

$$f_{\rho+n} = \sum_{k=0}^{\rho-1} a_k f_{k+n}, \quad (1)$$

$a_0 \neq 0$, holds for all appropriate n ; these time series are called the *time series of finite difference dimension*. In the present paper we consider complex-valued time series and LRFs with complex coefficients.

The roots of the *characteristic polynomial* $A(z) \stackrel{\text{def}}{=} z^\rho - a_{\rho-1}z^{\rho-1} + \dots + a_1z + a_0$ of the LRF (1) can be divided into two groups: the *signal roots* which determine the representation of the time series as a sum of polynomially modulated exponential signals

$$f_n = \sum_{k=1}^m P_k(n) \lambda_k^n, \quad (2)$$

and other roots, called *extraneous roots*. In this paper we show how the signal roots of the LRFs determine several properties of time series, in what concerns the Singular Spectrum Analysis. Specifically, we consider separability and the behaviour of extraneous roots for the LRF used in SSA forecasting.

The latter LRF is of particular interest since the extraneous roots affect the results of SSA continuation. The obtained results also concern *subspace-based* methods of signal processing [4, 9, 12, 19] (in particular, the *Min-Norm* method), which share their structure with SSA.

The first part of this paper is devoted to a novel outlook at separability [17, Ch. 6]. A new criterion of exact weak separability for the time series of finite difference dimension is introduced. This criterion enables us to present all the standard examples in a unified manner, elucidating the meaning of the conditions of separability in these examples. Moreover, it enables us to enumerate all possible cases of exact weak separability for arbitrary time series; this is also a new result. In fact, we develop a criterion for the one-sided separability, which has been recently shown to be important for the SSA theory [23]. The criterion is simple, illustrative and based solely on the signal roots of the time series. Specific features of the case where the time series is real are clarified as well.

The second part of the paper deals with the LRF used in SSA forecasting (shortly, the SSA LRF), see [17, Ch. 5]. This LRF coincides with the *Min-Norm prediction vector* [4, 9, 12]. Properties of the extraneous roots of the SSA LRF are examined in the noise-free case. Recently, the basic properties were proved by various authors. In particular, it was proved that all extraneous roots lie inside the unit circle and that the extraneous roots of the forward and backward SSA LRF coincide, see [5, 15]. The asymptotic distribution of the extraneous roots was independently addressed in [7, 16], but only some particular cases were considered. In the present paper we show the correspondence between the extraneous roots of the SSA LRF and *orthogonal polynomials on the unit circle* [18]. This correspondence enables us to provide elegant and short proofs for the basic properties of the extraneous roots and to describe the asymptotic distribution of the roots. By the way, this approach was used in [7], but in the present paper we try to provide a more accurate, comprehensive and integral exposition. We describe the asymptotic distribution for the general case of time series of finite difference dimension with the help of the most recent advances in the theory of orthogonal polynomials [20, 21].

Both parts of the present paper are based on the same ground of a revised theory of time series of finite difference dimension. In fact, this revision can be useful in the SSA theory on its own. Surprisingly, the two parts of the present paper are linked from another side: the approximate separability of certain time series can be somehow described with the help of the distribution of extraneous roots of a specific SSA LRF.

This paper is organized as follows. In Section 2 we provide all necessary background on the linear recurrent formulae and continuation, the time series of finite difference dimension and their trajectory spaces. Section 3 is devoted to exact weak separability. In Section 4 we investigate the SSA forecasting LRF and the properties of its extraneous roots. At the end of Section 4 we discuss the connection between the asymptotic distribution of extraneous roots and separability; we also make practical conclusions and examine the behaviour of roots in the presence of noise as well.

2 Basic facts and notation

2.1 Time series of finite difference dimension and signal roots

Let \mathbb{N}_0 denote the set of all nonnegative integers. An *infinite time series*

$$F_\infty = (f_0, f_1, \dots), \quad f_n \in \mathbb{C}, \quad (3)$$

is said to *satisfy a linear recurrent formula (LRF)* of order ρ if there exist coefficients $a_0, \dots, a_{\rho-1} \in \mathbb{C}$ such that the relation

$$f_{\rho+n} = \sum_{k=0}^{\rho-1} a_k f_{k+n} \quad (4)$$

holds for all $n \in \mathbb{N}_0$. Note that in the case $\rho = 0$ we have $f_n = 0$ for all $n \in \mathbb{N}_0$. Once a time series satisfies an LRF (4), its form can be described by the roots of the *characteristic polynomial* of the

LRF

$$A(z) = z^\rho - a_{\rho-1}z^{\rho-1} - \dots - a_1z - a_0. \quad (5)$$

Theorem 2.1 ([14, Th. 3.1.1]). *Assume that an infinite time series F_∞ satisfies an LRF (4) with $a_0 \neq 0$. Then it can be represented as*

$$f_n = \sum_{k=1}^m P_k(n)\lambda_k^n, \quad (6)$$

where $\lambda_k \in \mathbb{C} \setminus \{0\}$ are distinct numbers, and P_k are non-zero polynomials. All λ_k in the representation (6) are roots of the characteristic polynomial $A(z)$, with multiplicity not less than $\nu_k \stackrel{\text{def}}{=} \deg P_k + 1$, where $\deg \cdot$ is the degree of a polynomial.

The coefficients of P_k are determined by the first d values of the time series, where d is defined as

$$d = \nu_1 + \dots + \nu_m \leq \rho. \quad (7)$$

Remark 2.1. If a time series admits a representation of type (6), then this representation is unique. This follows from the linear independence of the time series of type $g_n = n^k \lambda^n$ for different $\lambda \in \mathbb{C} \setminus \{0\}$ and $k \in \mathbb{N}_0$.

For a time series of type (6), by Remark 2.1, one can unambiguously define the polynomial

$$\begin{aligned} P(z) &\stackrel{\text{def}}{=} (z - \lambda_1)^{\nu_1} \cdot \dots \cdot (z - \lambda_m)^{\nu_m} = \\ &= p_d z^d + \dots + p_1 z + p_0, \end{aligned} \quad (8)$$

where $p_d = 1$. This polynomial is called the *characteristic polynomial of the time series*. The characteristic polynomial determines the set of all LRF that are satisfied by the time series.

Theorem 2.2. *Let F_∞ be a time series of the form (6). Then any polynomial*

$$B(z) = b_r z^r + \dots + b_1 z + b_0 \quad (9)$$

of degree r (i.e. $b_r \neq 0$) is a multiple of the characteristic polynomial (8), i.e. $B(z) = P(z)Q(z)$, if and only if the time series satisfies the LRF

$$f_{n+r} = \sum_{k=0}^{r-1} -\frac{b_k}{b_r} f_{n+k}, \quad n \in \mathbb{N}_0. \quad (10)$$

Proof. \Rightarrow If F satisfies an LRF $f_{r+n} = \sum_{k=0}^{r-1} b_k f_{k+n}$, which has the characteristic polynomial $B(z) =$

$z^r - b_{r-1}z^{r-1} - \dots - b_0$, then it satisfies the LRF $f_{(r+l)+n} = \sum_{k=l}^{(r+l)-1} b_k f_{k+n} = 0$, which corresponds to

$B(z)z^l$. Therefore, for any $A(z) = B(z)Q(z)$, F satisfies the corresponding LRF.

The proof that a time series of the form (6) satisfies the LRF with the characteristic polynomial $P(z)$ can be found in the proof of [14, Th. 3.1.1].

\Leftarrow Note that if F_∞ satisfies the LRF with the characteristic polynomial $Q(z) = S(z)z^m$, $S(0) \neq 0$, then the time series $G_\infty = (f_m, f_{m+1}, \dots)$ satisfies the LRF corresponding to $S(z)$. Let $R(z) = \mathbf{GCD}(P(z), S(z))$ be the greatest common divisor of $P(z)$ and $S(z)$. Then $R(z)$ can be represented in the form $R(z) = P(z)c(z) + S(z)d(z)$ (see [10, Ch.III, §17]), and by \Rightarrow part, G_∞ satisfies the LRF with characteristic polynomial $R(z)$. If $R(z)$ is not $P(z)$ then $R(z) = (z - \lambda_1)^{d_1} \cdot \dots \cdot (z - \lambda_m)^{d_m}$ with $d_k \leq \nu_k$ for all k and at least one $d_l < \nu_l$. Then by Theorem 2.1 $g_n = \sum_{k=1}^m Q_k(n)\lambda_k^n$, where $Q_l(n)$ has the degree less than $d_k - 1$. By the linear independence of time series $n^k \lambda^n$ we obtain the contradiction with the representation (6). \square

Remark 2.2. Theorems 2.1 and 2.2 establish the one-to-one correspondence between the time series of type (6) and the time series satisfying at least one LRF (4) with non-zero last coefficient ($a_0 \neq 0$).

Now assume that a time series F_∞ satisfies an LRF (4) with $a_0 \neq 0$. By Theorem 2.1 it has the representation (6) and the characteristic polynomial (8) is uniquely determined. By Theorem 2.2, the relation

$$A(z) = P(z)V(z) \quad (11)$$

holds. Moreover, the time series satisfies all LRFs with characteristic polynomials of form $B(z) = P(z)Q(z)$, and hence the polynomial $V(z)$ (11) (and its roots) has no effect on the form of the time series. Thus, the ρ roots of the characteristic polynomial $A(z)$ can be divided into two groups:

1. the d *signal roots* (i.e. the roots of $P(z)$), which determine the structure of the time series,
2. the $\rho - d$ *extraneous roots*,

where d is defined in (7). We also say that the signal roots λ_k of $A(z)$ are the *signal roots of the time series* and ν_k are their *multiplicities*.

The number d of signal roots has an important interpretation in terms of LRFs; this interpretation follows from Theorem 2.2.

Corollary 2.1. *If a time series F_∞ satisfies an LRF (4) with $a_0 \neq 0$, then the LRF corresponding to the characteristic polynomial (8) of F_∞*

$$f_{d+n} = - \sum_{k=0}^{d-1} p_k f_{k+n} \quad (12)$$

has the minimal order d among all LRFs satisfied by F_∞ .

Note that Corollary 2.1 is a characterization of $P(z)$, and can be taken as an alternative definition of the characteristic polynomial $P(z)$. It also validates the following notation.

Definition 2.1. We say that F_∞ is a time series of finite difference dimension (an *f.d.d. time series*) if it satisfies at least one LRF (4) with $a_0 \neq 0$. The degree d of the characteristic polynomial, defined in (7), is called the *difference dimension* of F_∞ .

Remark 2.3. The case of LRFs with $a_0 = 0$ can be considered within the same framework along with generalizations of Theorems 2.1 and 2.2, but we omit this consideration.

For clarity, let us consider a real-valued time series F_∞ of finite difference dimension. By Theorem 2.1, it has the form

$$f_n = \sum_{l=1}^{s'} P_l(n) \rho_l^n + \sum_{l=s'+1}^s P_l(n) \rho_l^n \cos(2\pi\omega_l n + \varphi_l), \quad (13)$$

where ω_l, ρ_l are distinct, $|\omega_l| < 0.5$ and P_l are real polynomials of degree $n_l - 1$. If we denote $\lambda_l = \rho_l, \nu_l = r_l$ for $l \leq s'$ and $\lambda_{2l-s'-1} = \rho_l e^{2\pi i \omega_l}, \lambda_{2l-s} = \rho_l e^{-2\pi i \omega_l}, \nu_{2l-s'-1} = \nu_{2l-s} = n_l$ for $s' < l \leq s$, then we obtain the representation (6) with $m = 2s - s'$ modulated exponents.

2.2 Hankel matrices and trajectory spaces

Let

$$F = F_N = (f_0, \dots, f_{N-1})^T \in \mathbb{C}^N \quad (14)$$

be a (finite) *time series*. The *Hankel matrix* generated by the time series (or the *trajectory matrix*) is the matrix

$$\mathbf{X}^{(L)}(F_N) \stackrel{\text{def}}{=} \begin{pmatrix} f_0 & f_1 & f_2 & \dots & f_{K-1} \\ f_1 & f_2 & f_3 & \dots & f_K \\ f_2 & f_3 & f_4 & \dots & f_{K+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ f_{L-1} & f_L & f_{L+1} & \dots & f_{N-1} \end{pmatrix}, \quad (15)$$

where the parameter L is called the *window length*, $1 < L < N$ and $K = N - L + 1$.

Definition 2.2. If F_N is a subseries of an infinite time series (3) of difference dimension $d \leq N/2$, then F_N is called a time series of (finite) *difference dimension* d (with characteristic polynomial $P(z)$).

This definition agrees with the definition given in [17, Ch. 2], see remarks at the beginning of Section 2.3. In particular, the following theorem states that the time series of finite difference dimension are *time series of finite rank*, see also [17, Ch. 5, Prop. 5.4].

Proposition 2.1. *Let F_N be of difference dimension d .*

1. *For the window length L such that $d \leq L \leq N - d + 1$ the trajectory matrix (14) is of rank d .*
2. *If $L < d$ or $L > N - d + 1$ then $\mathbf{X}^{(L)}$ has maximal possible rank (L or $N - L + 1$, respectively).*

The proposition immediately follows from [13, Ch. XVI, §10, Th. 7] and its corollary.

Let us show how the structure of a time series is connected to LRFs which are satisfied by the time series. The structure of a time series in SSA is described by its *trajectory space*

$$\mathfrak{L}^{(L)} = \mathfrak{L}^{(L)}(F_N) \stackrel{\text{def}}{=} \text{span}(X_1^{(L)}, \dots, X_K^{(L)}) \subseteq \mathbb{C}^L,$$

where

$$X_i^{(L)} = (f_{i-1}, \dots, f_{i+L-2})^T, \quad 1 \leq i \leq K, \quad (16)$$

are the columns of the matrix $\mathbf{X}^{(L)} = \mathbf{X}^{(L)}(F_N)$, see [17, Ch. 1] for a detailed discussion. In what follows, the following subspace of \mathbb{C}^L is very useful.

Definition 2.3. The *relations space* is defined as

$$\mathfrak{R}^{(L)} = \mathfrak{R}^{(L)}(F_N) \stackrel{\text{def}}{=} \overline{\mathfrak{L}_\perp^{(L)}},$$

where $\overline{\mathfrak{L}}$ denotes the complex conjugation of \mathfrak{L} and $\mathfrak{L}_\perp^{(L)}$ is the orthogonal complement to the trajectory space.

The relations space consists of all linear relations on rows of $\mathbf{X}^{(L)}$. Indeed, the vector

$$(a_0, \dots, a_r, 0, \dots, 0)^T, \quad a_r \neq 0,$$

belongs to $\mathfrak{R}^{(L)}$ if and only if

$$f_{n+r} = - \sum_{k=0}^{r-1} \frac{a_k}{a_r} f_{n+k}, \quad 0 \leq n \leq N - L + 1. \quad (17)$$

Remark 2.4. In complex vector spaces the inner product involves complex conjugation. This explains the presence of the conjugation in the definition of the relations space. If F_N is real-valued, then $\mathfrak{R}^{(L)} = \mathfrak{L}_\perp^{(L)}$.

The following proposition shows that the relations space of a time series of finite difference dimension is generated by all LRF of order less than L , satisfied by its infinite time series (c.f. Definition 2.2).

Proposition 2.2. *Let F_N be a time series of difference dimension d with characteristic polynomial $P(z)$ (8). For the window length L , $d < L \leq N - d + 1$, we have the following.*

1. The columns of the $L \times (L - d)$ matrix

$$\mathbf{P} = \mathbf{P}^{(L)} = \begin{pmatrix} p_0 & 0 & \dots & 0 \\ \vdots & p_0 & \ddots & \vdots \\ p_d & \vdots & \ddots & 0 \\ 0 & p_d & \ddots & p_0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & p_d \end{pmatrix} \quad (18)$$

form a basis of the space $\mathfrak{R}^{(L)}$.

2. A vector $B = (b_0, \dots, b_{L-1})^\top$ belongs to $\mathfrak{R}^{(L)}$ if and only if $B(z) = b_{L-1}z^{L-1} + \dots + b_1z + b_0$ is a multiple of $P(z)$.

Proof. Since $p_d \neq 0$, the columns of the matrix \mathbf{P} (18) are linearly independent. By Theorem 2.2, the time series F_N satisfies the LRF (12) and the $L - d$ columns of \mathbf{P} belong to $\mathfrak{R}^{(L)}$. By Proposition 2.1, the dimension of $\mathfrak{R}^{(L)}$ is $L - d$, and hence the columns of \mathbf{P} form a basis of this space. \square

Remark 2.5. By Remark 2.4, for a real time series the basis given in Proposition 2.2 is a basis of $\mathfrak{L}_\perp^{(L)}$ as well.

We also note that a similar basis can be introduced for arbitrary (not only f.d.d.) time series (see [6]) and can be very useful in the SSA theory. In the present paper we demonstrate several applications of this basis in the case of time series of finite difference dimension. We will also need a (“Vandermonde”-like) basis of the trajectory space.

Proposition 2.3. *Let F_N be an f.d.d. time series with the characteristic polynomial defined in (8). Let also $d < L \leq N - d + 1$. Then a basis of $\mathfrak{L}^{(L)}$ is given by the vectors*

$$\ell_L^0(\lambda_1), \dots, \ell_L^{\nu_1-1}(\lambda_1), \dots, \ell_L^0(\lambda_m), \dots, \ell_L^{\nu_m-1}(\lambda_m),$$

where

$$\begin{aligned} \ell_L^k(\lambda) &= \frac{\partial^k}{\partial \lambda^k} (1, \lambda, \lambda^2, \dots, \lambda^{L-1})^\top = \\ &= \underbrace{(0, \dots, 0)}_k, k!, \dots, \frac{(k+j)!}{j!} \lambda^j, \dots, \frac{(L-1)!}{(L-k-1)!} \lambda^{L-k-1} \end{aligned}$$

Proof. Let $(G_\infty)^{\lambda, k}$ be the formal continuation (as an infinite time series) of $\ell_L^k(\lambda)$, e.g.

$$(G_\infty)^{\lambda, 0} = (1, \lambda, \lambda^2, \dots).$$

Then by Theorem 2.2 any time series $(G_\infty)^{\lambda_j, k}$, $0 \leq k < \nu_j$, satisfies the LRF (12) and all LRFs with characteristic polynomials of type $P(z)Q(z)$, where $Q(z) \not\equiv 0$. Therefore, by the second assertion of Proposition 2.2, each vector $\ell_L^k(\lambda_j)$ is orthogonal to $\mathfrak{L}_\perp^{(L)}$. Since these d vectors are linearly independent, the assertion is proved. \square

2.3 LRFs and continuation

In this section we discuss the time series that can be continued within the SSA framework. A time series F_N defined in (14) admits the (forward) L -continuation (is L -continuable) if there exists unique $\alpha \in \mathbb{C}$ such that $\mathfrak{L}^{(L)}(F_N) = \mathfrak{L}^{(L)}(f_0, \dots, f_{N-1}, \alpha)$, see [17, Ch. 5] for details on continuation. First, we show a connection between these time series and time series of finite difference dimension.

Proposition 2.4 ([17, Th. 5.4]). *If a time series F_N satisfies some LRF*

$$f_{n+d_0} = \sum_{k=0}^{d_0-1} a_k f_{n+k}, \quad 0 \leq n \leq N - d_0 - 1, \quad (19)$$

with $d_0 \leq \min(L - 1, K)$, then it is L -continuable and the continuation is achieved by the same LRF (i.e. (19) holds for $n = N - d_0$ if we set $f_N = \alpha$).

Remark 2.6. By Proposition 2.4 one can continue F_N to an infinite time series $F_\infty = (f_0, f_1, \dots)$, which satisfies (19) for all n (i.e. F_∞ is of finite difference dimension). This fact is the base of SSA forecasting, see [17, Ch. 2].

The condition (19) means that the time series F_N is f.d.d. in the sense of [17, Ch. 2], see also a remark after Definition 2.2. However, Remark 2.6 together with the following assertion shows that (19) is equivalent to the definition adopted in the present paper.

Corollary 2.2. *Any time series F_N of difference dimension d is L -continuable if $d < L \leq N - d + 1$.*

Evidently, the infinite continuation (see Remark 2.6) of a finite subseries F_N of an f.d.d. time series F_∞ (with $d \leq N/2$) coincides with the original time series F_∞ . This observation removes the ambiguity from Definition 2.2: a finite time series F_N of finite difference dimension cannot be a subseries of more than one infinite f.d.d. time series due to the uniqueness of continuation.

The following result, which is the converse to Proposition 2.4, can be found in [6, Ch. 5].

Proposition 2.5. *If a time series F_N is L -continuable, then there exists $d_0 \leq \min(L - 1, K)$ such that F_N satisfies an LRF (19).*

Proof. The proposition is a direct consequence of [6, Th. 5.6] and [6, Prop. 5.8]. \square

For convenience, we recall the well-known necessary and sufficient conditions for F_N to be L -continuable.

Proposition 2.6 ([17, §5.3]). *If F_N is L -continuable, then $e_L \notin \mathfrak{L}^{(L)}(F_N)$, where $e_L \stackrel{\text{def}}{=} (0, \dots, 1)^\top \in \mathbb{C}^L$.*

Proposition 2.7 ([17, Th. 5.4]). *Let $L \leq N/2$. If $e_L \notin \mathfrak{L}^{(L)}(F_N)$, then F_N is L -continuable.*

We also need the notion of the backward L -continuation.

Definition 2.4. A time series F_N admits the *backward L -continuation* (it is *backward L -continuable*) if there exists unique $\alpha \in \mathbb{C}$ such that $\mathfrak{L}^{(L)}(F_N) = \mathfrak{L}^{(L)}(\alpha, f_0, \dots, f_{N-1})$.

It is clear that the backward L -continuation is equivalent to the forward L -continuation of the reversed time series $\widehat{F_N} \stackrel{\text{def}}{=} (f_{N-1}, \dots, f_0)$, and the following assertion holds.

Corollary 2.3. 1. *If F_N is backward L -continuable, then $e_1 \notin \mathfrak{L}^{(L)}(F_N)$, where $e_1 \stackrel{\text{def}}{=} (1, \dots, 0)^\top \in \mathbb{C}^L$.*

2. *Let $L \leq N/2$. If $e_1 \notin \mathfrak{L}^{(L)}(F_N)$ then F_N is backward L -continuable.*

We conclude this subsection with a result which is complementary to Corollary 2.2 and will be very useful in the next section.

Proposition 2.8. *If F_N is both forward and backward L -continuable, then it is a time series of finite difference dimension $d \leq \min(L - 1, K)$.*

Proof. If F_N is forward L -continuable then by Proposition 2.4 there exist coefficients $a_0, \dots, a_{d_1-1} \in \mathbb{C}$ such that F_N satisfies

$$f_{n+d_1} = \sum_{k=0}^{d_1-1} a_k f_{n+k}, \quad 0 \leq n \leq N - d_1 - 1,$$

where $d_1 \leq \min(L - 1, K) \leq N/2$. If $a_0 \neq 0$ then F_N is an f.d.d. time series by the definition. If $a_0 = 0$ then we use the fact that F_N is backward L -continuable and there exist coefficients $b_1, \dots, b_{d_1} \in \mathbb{C}$ such that

$$f_n = \sum_{k=1}^{d_2} b_k f_{n+k}, \quad 0 \leq n \leq N - d_2 - 1.$$

Without loss of generality, assume that $d_2 \leq d_1$. Then there exists a constant $\alpha \neq 0$ such that the last component c_{d_1} of

$$\begin{aligned} (c_0, \dots, c_{d_1})^T &\stackrel{\text{def}}{=} \alpha(1, -b_1, \dots, -b_{d_2}, \overbrace{0, \dots, 0}^{d_1-d_2})^T + \\ &+ (0, -a_1, \dots, -a_{d_1-1}, 1)^T \end{aligned}$$

is non-zero. Then F_N satisfies

$$f_{n+d_1} = - \sum_{k=0}^{d_1-1} \frac{c_k}{c_{d_1}} f_{n+k}, \quad 0 \leq n \leq N - d_1 - 1.$$

Since $c_0 = \alpha \neq 0$, the result follows from Theorem 2.1. □

3 Separability

The main result of this section is a necessary and sufficient condition of separability in terms of characteristic polynomials and signal roots of f.d.d. time series. In particular, it enables us to enumerate all possible cases of weak separability and present standard examples (see [17, §6.1]) in a unified manner. Moreover, we develop the new theory for the *one-sided separability*, the results on the conventional separability are its consequence.

3.1 One-sided separability criterion

Definition 3.1. Time series $F_N^{(1)}, F_N^{(2)}$ are called *weakly left-separable (right-separable)* if $\mathfrak{L}^{(L)}(F_N^{(1)}) \perp \mathfrak{L}^{(L)}(F_N^{(2)})$ ($\mathfrak{L}^{(K)}(F_N^{(1)}) \perp \mathfrak{L}^{(K)}(F_N^{(2)})$), respectively).

For any vector $B = (b_0, \dots, b_{L-1})^T \in \mathbb{C}^L$ we denote by $B(z) = b_{L-1}z^{L-1} + \dots + b_1z + b_0$ its *generating polynomial*. We shall also need a notation for the polynomial with conjugate coefficients, $\overline{B}(z) = \overline{b_{L-1}}z^{L-1} + \dots + \overline{b_1}z + \overline{b_0}$.

Proposition 3.1. Assume that $F_N^{(1)}$ is a non-zero time series and $F_N^{(2)}$ is a time series of difference dimension d with characteristic polynomial $P^{(2)}(z)$. Let L satisfy $d < L \leq N - d + 1$ and $\{U_1, \dots, U_r\}$ be an arbitrary basis of $\mathfrak{L}^{(L)}(F_N^{(1)})$. Then $F_N^{(1)}$ and $F_N^{(2)}$ are separable if and only if $P^{(2)}(z)$ is a common divisor of $\overline{U}_i(z)$, $1 \leq i \leq d$ (in other words, all the signal roots of $F_N^{(2)}$ are common roots of all $\overline{U}_i(z)$, at least with respective multiplicities).

Proof. Note that $\mathfrak{L}^{(L)}(F_N^{(1)}) \perp \mathfrak{L}^{(L)}(F_N^{(2)})$ if and only if $\mathfrak{L}^{(L)}(F_N^{(1)}) \subseteq \mathfrak{L}_\perp^{(L)}(F_N^{(2)})$. The subspace inclusion can be reformulated as $\overline{U}_i \in \mathfrak{R}^{(L)}(F_N^{(2)})$ for all $1 \leq i \leq r$ (c.f. Definition 2.3). By Proposition 2.2, $\overline{U}_i \in \mathfrak{R}^{(L)}(F_N^{(2)})$ if and only if $\overline{U}_i(z) = P^{(2)}(z)Q_i(z)$, where $P^{(2)}(z)$ is the characteristic polynomial of $F_N^{(2)}$ and $Q_i(z)$ is a nonzero polynomial. This proves the proposition. \square

Remark 3.1. Proposition 3.1 is valid even for $1 < L \leq N - d + 1$, and thus it is valid for the essential case $L \leq N/2$. Indeed, if $L \leq d$, then by Proposition 2.1 $\dim(\mathfrak{L}_\perp^{(L)}(F_N^{(2)})) = L - L = 0$ and only zero time series can be left separable from $F_N^{(2)}$. At the same time, any $\overline{U}_i(z)$ is of degree less than d and hence it cannot be a multiple of $P^{(2)}(z)$.

In what follows we assume that $L \leq N/2$, unless indicated. Proposition 3.1 together with Remark 3.1 points out a method of finding all f.d.d. time series $F_N^{(1)}$ that are left-separable from given $F_N^{(2)}$.

Proposition 3.2. Assume that $F_N^{(1)}$ is a non-zero time series and $\{U_1, \dots, U_r\}$ is a basis of its trajectory space $\mathfrak{L}^{(L)}(F_N^{(1)})$, $L \leq N/2$. Let μ_1, \dots, μ_l be all distinct nonzero common roots of $\overline{U}_1(z), \dots, \overline{U}_r(z)$ and d_1, \dots, d_l be their common multiplicities, i.e. d_k is the minimal multiplicity of μ_k in polynomials $\overline{U}_i(z)$. Then all f.d.d. time series that are left-separable from $F_N^{(1)}$ are given by

$$f_n^{(2)} = \sum_{k=1}^l Q_k(n) \mu_k^n, \quad (20)$$

where $Q_k(n)$ are (possibly zero) polynomials of degree less than d_k and not all Q_k are zero.

Proof. Polynomials $\overline{U}_i(z)$ may have an additional zero root $\mu_0 = 0$ of multiplicity $d_0 \in \mathbb{N}_0$. Then the polynomial

$$R(z) = (z - \mu_1)^{d_1-1} \dots (z - \mu_l)^{d_l-1} z^{d_0}$$

is the greatest common divisor of $\overline{U}_i(z)$. Therefore, by Remark 3.1, $F_N^{(1)}$ and $F_N^{(2)}$ are left-separable if and only if the characteristic polynomial $P^{(2)}(z)$ is a divisor of $R(z)$. By Theorem 2.1, the time series having this property are precisely the time series of form (20). \square

Remark 3.2. The statement of Proposition 3.2 does not depend on the choice of the basis. Indeed, if μ is a common root for the set of polynomials $\{\overline{U}_1(z), \dots, \overline{U}_r(z)\}$, then it is a common root of all linear combinations of $\overline{U}_i(z)$.

Let $F_N^{(1)}$ be an f.d.d. time series of the form (6). Then a basis of $\mathfrak{L}^{(L)}(F_N^{(1)})$ is given by the vectors $\ell_L^i(\lambda_k)$ from Proposition 2.3, $1 \leq k \leq m$, $0 \leq i < \nu_k$, and we can use Proposition 3.2 to determine all time series that are left-separable from $F_N^{(1)}$. Moreover, we can check the left separability condition separately for each summand in the representation (6).

Remark 3.3. Let $F_N^{(1)}$ be a time series of type (6). Then an f.d.d. time series $F_N^{(2)}$ is left-separable from $F_N^{(1)}$ if and only if it is left-separable from each $P_k(n)\lambda_k^n$ given in (6).

Indeed, each summand $P_k(n)\lambda_k^n$ has the vectors $\ell_L^i(\lambda_k)$, $0 \leq i < \nu_k$, as a basis of its trajectory space, and the conditions for the signal roots of $F_N^{(2)}$ to be separable from $F_N^{(1)}$ coincide with the set of conditions for the separability from each summand of $F_N^{(1)}$.

3.2 Separability of f.d.d. time series

By Remark 3.3, to study the separability of arbitrary f.d.d. time series it is sufficient to consider the separability from a time series $F_1(n)\lambda^n$, i.e. the time series with only one (possibly multiple) signal root. We are going to examine this case gradually in separate examples. In the present subsection we again assume $L \leq N/2$.

Example 3.1. [Separability from a constant] Let $F_N^{(1)} \equiv c \neq 0$. Then the space $\mathfrak{L}^{(L)}(F_N^{(1)})$ is spanned by the single vector $(1, \dots, 1)^T \in \mathbb{C}^L$. Consider its generating polynomial

$$W_1(z) = z^{L-1} + \dots + 1 = \frac{z^L - 1}{z - 1}. \quad (21)$$

The roots μ_k of $W_1(z)$ are simple; they are the L th roots of unity, excluding 1:

$$\mu_k = \exp\left(\frac{2\pi ik}{L}\right), \quad 0 < k < L.$$

In Fig. 1 the roots μ_k are represented by dots.

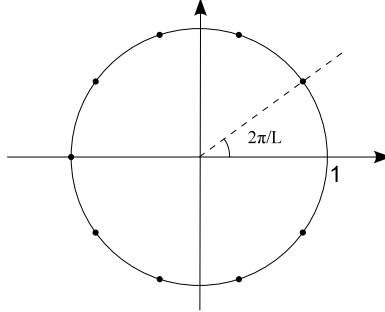


Figure 1: Roots of the series left-separable from a constant.

By Proposition 3.2, all left-separable from $F_N^{(1)}$ time series have the form

$$f_n^{(2)} = \sum_{k=1}^{L-1} c_k \left(\exp\left(\frac{2\pi ik}{L}\right) \right)^n.$$

Note that $F_N^{(2)}$ is an L -periodic series with

$$\sum_{n=0}^{L-1} f_n^{(2)} = 0 \quad (22)$$

if and only if it has the representation as above, see [17, §6.1]. In the case of real-valued time series, $F_N^{(2)}$ is a sum of harmonics with period L .

Example 3.2. [Separability from a complex exponent] Let $f_n^{(1)} = \lambda^n$. Then $\mathfrak{L}^{(L)}(F_N^{(1)})$ is spanned by the single vector $(1, \lambda, \dots, \lambda^{L-1})^T$, and all signal roots of $F_N^{(2)}$ have to be roots of the polynomial

$$W_\lambda(z) = \bar{\lambda}^{L-1} z^{L-1} + \dots + \bar{\lambda} z + 1.$$

Making the change of variables $z = x/\bar{\lambda}$, we have to solve $W_1(x) = 0$, where $W_1(x)$ is defined in (21). Therefore, the roots of $W_\lambda(z)$ are simple and have the form

$$\mu_k = \frac{\exp(2\pi ik/L)}{\bar{\lambda}} = \frac{\lambda}{|\lambda|^2} \exp\left(\frac{2\pi ik}{L}\right),$$

where $0 < k < L$. Hence, the time series is of form

$$f_n^{(2)} = \sum_{k=1}^{L-1} c_k \mu_k^n = \sum_{k=1}^{L-1} c_k \left(\frac{\exp(2\pi ik/L)}{\bar{\lambda}}\right)^n,$$

where $c_k \in \mathbb{C}$ are not simultaneously zero. This representation is more illustrative if we represent the roots μ_k in the polar form. If $\lambda = \rho \exp(2\pi i\omega)$, $\rho > 0$, $\omega \in [0; 1)$, then

$$\mu_k = \rho^{-1} \exp\left(2\pi i \left(\frac{k}{L} + \omega\right)\right).$$

The roots μ_k are shown in Fig. 2

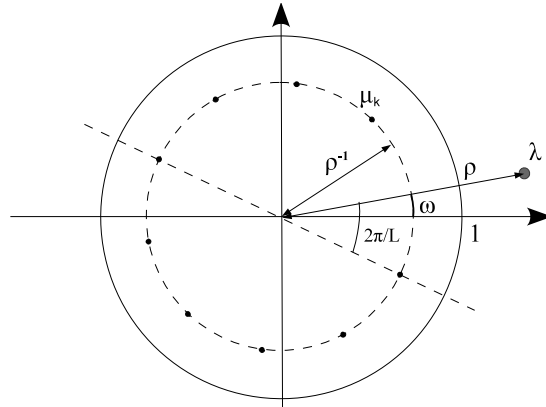


Figure 2: Roots of the series left-separable from a complex exponent.

Let us write out the results for the case $\text{Im } \lambda = 0$, when $F_N^{(1)}$ is a real exponential time series. If $\lambda = \text{Re } \lambda > 0$ (i.e. $\omega = 0$), then $f_n^{(1)} = \rho^n$ is left-separable from all time series of type $f_n^{(2)} \stackrel{\text{def}}{=} \rho^{-n} f_n^{(3)}$, where $f_n^{(3)}$ has period L and zero sum (in the sense of (22)). If $\lambda < 0$ ($\omega = 0.5$) then $f_n^{(1)} = (-\rho)^n$ is also an exponentially modulated cosine with period 2 (an exponentially modulated *saw-tooth* series), and the set of all left-separable series can be described in a similar way.

Let us consider the important case of a sum of conjugate exponents, which corresponds to real cosine time series with periods other than 2, multiplied by a complex constant. This case has some peculiarities.

Example 3.3. [Separability from a sum of conjugate exponents] Let $f_n^{(1)} = c\lambda^n + d\bar{\lambda}^n$, $\text{Im } \lambda \neq 0$. We choose $\text{Im } \lambda > 0$, without loss of generality. Then, by Example 3.2 and Remark 3.3, any signal root μ of $F_N^{(2)}$ must satisfy

$$\mu = \exp\left(\frac{2\pi ik}{L}\right) / \bar{\lambda} = \exp\left(\frac{2\pi il}{L}\right) / \lambda,$$

where $0 < k, l < L$. Hence,

$$\frac{\lambda}{\bar{\lambda}} = \exp\left(\frac{2\pi im}{L}\right),$$

where $m \in \mathbb{Z}$. Therefore, λ has to be of form

$$\lambda = \rho \exp\left(\frac{2\pi im}{2L}\right), \quad \rho > 0, \quad (23)$$

for $F_N^{(1)}$ to be separable from $F_N^{(2)}$. Since $\text{Im } \lambda > 0$ we can choose m such that $0 < m < L$. Then the signal roots of $F_N^{(2)}$ are of the form

$$\mu_k = \rho^{-1} \exp\left(\frac{2\pi i}{2L}(2k - m)\right),$$

where $0 < k < L$, $k \neq m$, and the time series has the representation

$$f_n^{(2)} = \sum_{k \in \{1, \dots, L-1\} \setminus \{m\}} c_k \rho^{-n} \exp\left(\frac{2\pi i}{2L}(2k - m)n\right).$$

Note that, in contrast to Example 3.2, the constraint (23) on the signal roots of $F_N^{(1)}$ naturally arises. In the real-valued case, if two exponentially modulated cosine series are left-separable, then these cosines should have integer periods related to L . Indeed, if $f_n^{(1)} = \rho^n \cos(\varphi + 2\pi\omega n) = c\lambda^n + d\bar{\lambda}^n$ where $\varphi \in \mathbb{R}$, $\rho > 0$ and $\omega \in (0, 0.5)$, then ω must satisfy $\omega = m/2L$, $0 < m < L$, for $F_N^{(1)}$ to be separable from another time series. Thus all real-valued time series $F_N^{(2)}$ that are left-separable from $F_N^{(1)}$ have the following representation

$$f_n^{(2)} = \sum_{k \in \{1, \dots, L-\delta\} \setminus \{k_0\}} c_k \rho^{-n} \cos\left(\varphi_k + 2\pi \frac{(2k + \delta)}{2L} n\right),$$

where $k_0 = \lfloor \frac{m}{2} \rfloor$,

$$\delta = \begin{cases} 0, & m \text{ is even,} \\ 1, & m \text{ is odd,} \end{cases}$$

c_k , φ_k are arbitrary reals and not all c_k are zero. This is in agreement with considerations in [17, §6.1], but the present exposition is more illustrative.

Note that we have two slightly different situations for even and odd m . In essence, $\rho^n f_n^{(2)}$ has to be $2L$ periodic; however, for even m it necessarily has the period L . This effect has been already observed (for the cosine time series) in [17, §6.1], but in the present paper, again, the nature of this effect is more evident. Examples of locations of signal roots of $F_N^{(2)}$ for these two situations and $\rho = |\lambda| = 1$ are depicted in Fig. 3.

Finally, let us show that in the case of multiple roots no time series is separable from a given time series.

Example 3.4. [Separability from an exponentially modulated polynomial] Let $f_n^{(1)} = P_1(n)\lambda^n$, $\lambda \neq 0$, and $P_1(n)$ be a polynomial of nonzero degree. Then, by Proposition 2.3, the basis of $F_N^{(1)}$ includes the vectors

$$\begin{aligned} W_\lambda &= \ell_L^0(\lambda) = (1, \lambda, \dots, \lambda^{L-1})^\top, \\ \ell_L^1(\lambda) &= (0, 1, 2\lambda, \dots, (L-1)\lambda^{L-2})^\top. \end{aligned}$$

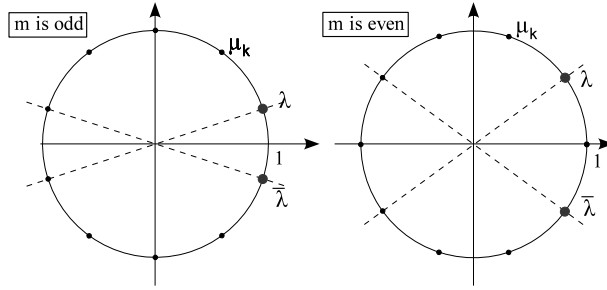


Figure 3: Roots of the series (small dots) separable from a sum of conjugate exponents (large dots), $\rho = 1$.

Note that in this basis we can replace $\ell_L^1(\lambda)$ by

$$W_\lambda^{(1)} = \ell_L^0(\lambda) + \lambda \ell_L^1(\lambda) = (1, 2\lambda, 3\lambda^2, \dots, L\lambda^{L-1})^T.$$

By Proposition 3.2, all signal roots of $F_N^{(2)}$ should be at least common roots of the generating polynomials $W_\lambda(z)$ and $W_{\bar{\lambda}}^{(1)}(z)$. Making the change of variables $z = x/\bar{\lambda}$, as in Example 3.2, we have to solve

$$W_1(x) = W_1^{(1)}(x) = 0,$$

where $W_1(x)$ is defined in (21) and $W_1^{(1)}(x) = Lx^{L-1} + \dots + 3x^2 + 2x + 1$. $W_1(x)$ has the roots on the unit circle, whereas $W_1^{(1)}(x)$ has decreasing positive coefficients, and therefore, by the Eneström-Kakeya theorem [1], its roots lie strictly inside the unit disc, and hence $W_1(x)$ and $W_1^{(1)}(x)$ have no common roots. Thus, no f.d.d. time series is left-separable from $F_N^{(1)}$.

3.3 Enumeration of separability cases

First, let us explicitly enumerate all cases of the separability between nonzero f.d.d. time series. By Example 3.4, only time series with the simple signal roots can be left-separable. Consider two time series

$$\begin{aligned} f_n^{(1)} &= \sum_{k=1}^r c_k \lambda_k^n, \\ f_n^{(2)} &= \sum_{j=1}^l d_j \mu_j^n, \end{aligned} \tag{24}$$

such that $c_k, d_j \neq 0$ and $\lambda_k \neq \lambda_i, \mu_j \neq \mu_m$ for $k \neq i, j \neq m$. Then, by Remark 3.3, these time series are left-separable if and only if each summand $c_k \lambda_k^n$ is separable from each $d_j \mu_j^n$. Summarizing all the above and using Example 3.2, we obtain the following proposition.

Proposition 3.3. *Two f.d.d. time series $F_N^{(1)}$ and $F_N^{(2)}$ are left-separable if and only if they have the form (24), and there exist $\rho > 0$ and $\omega \in [0; 1/L)$ such that*

$$\begin{aligned} \lambda_k &= \rho \exp\left(2\pi i \left(\frac{m_k}{L} + \omega\right)\right), \\ \mu_j &= \rho^{-1} \exp\left(2\pi i \left(\frac{n_j}{L} + \omega\right)\right), \end{aligned}$$

where $0 \leq m_k, n_k < L$ are distinct numbers.

As in Example 3.3, Proposition 3.3 can be specialized to real-valued time series. In this case a constraint on L appears again (either $\omega = 0$ or $\omega = 1/2L$).

Let us examine the separability of backward and forward non-continuable time series. Note that hereafter we do not assume $L \leq N/2$.

Proposition 3.4. *Let $F_N^{(1)}$ and $F_N^{(2)}$ be two nonzero time series, $1 < L < N$, and $e_L \in \mathfrak{L}^{(L)}(F_N^{(1)})$. These time series are left-separable if and only if both are “border” time series, specifically,*

$$\begin{aligned} F_N^{(1)} &= (0, \dots, 0, f_{N-d}^{(1)}, \dots, f_{N-1}^{(1)}), \\ F_N^{(2)} &= (f_0^{(2)}, \dots, f_{L-d-1}^{(2)}, 0, \dots, 0), \end{aligned} \quad (25)$$

where $1 < d < L$.

Proof. Let $e_L \in \mathfrak{L}^{(L)}(F_N^{(1)})$. Then the columns $X_i^{(L,2)}$ of the trajectory matrix $\mathbf{X}^{(L)}(F_N^{(2)})$ satisfy $X_i^{(L,2)} \perp e_L$, $1 \leq i \leq K$. Hence the last component of each vector $X_i^{(L,2)}$ is zero and finally $F_N^{(2)} = (f_0^{(2)}, \dots, f_{L-2}^{(2)}, \underbrace{0, \dots, 0}_K)$.

Let us consider the reversed time series $\widehat{F_N^{(1)}}$ and $\widehat{F_N^{(2)}}$. Since $F_N^{(2)}$ is nonzero, we have $e_L \in \mathfrak{L}^{(L)}(\widehat{F_N^{(2)}})$ and, applying the first part of the proof, we obtain $\widehat{F_N^{(1)}} = (f_{N-1}^{(1)}, \dots, f_K^{(1)}, \underbrace{0, \dots, 0}_K)$. The

proof is completed by noting that the sum of the lengths of the non-zero parts at the beginning of $F_N^{(2)}$ and at the end of $F_N^{(1)}$ cannot exceed L ; the details are left to the reader. \square

From Propositions 2.7 and 3.4 we immediately obtain the following result.

Corollary 3.1. *Let $F_N^{(1)}$ and $F_N^{(2)}$ be two nonzero time series, $L \leq N/2$, and either $F_N^{(1)}$ is forward non-continuable or $F_N^{(2)}$ is backward non-continuable. Then the time series are left-separable if and only if they have the form (25).*

Remark 3.4. By Proposition 2.8, the complete classification of all the cases of left separability is given by Proposition 3.3 and Corollary 3.1.

Let us finish this section with the consideration of the conventional, two-sided, separability. Two time series are called *weakly (two-sided) separable* if they are both left- and right-separable [17, Ch. 6].

Proposition 3.5. *Corollary 3.1 remains valid if we replace the left separability with the two-sided separability, and L with $\min(L, K)$.*

Proof. Without loss of generality we can choose $L \leq N/2$. Corollary 3.1 implies that the $\boxed{\Rightarrow}$ part takes place. The $\boxed{\Leftarrow}$ also takes place since all time series of form (25) with $L = L_0$ are of the same form for all $L \geq L_0$. \square

Therefore, we should again consider only the cases of separability of f.d.d. time series. Let $F_N^{(1)}$ and $F_N^{(2)}$ be time series of difference dimension $d^{(1)}$ and $d^{(2)}$, respectively. Then they are separable only if their trajectory matrices do not have full column or row rank, and hence, by Proposition 2.1 the window length L satisfies the inequality $\max(d^{(1)}, d^{(2)}) < L < N - \max(d^{(1)}, d^{(2)}) + 1$. One can see that Proposition 3.2 and Remark 3.3 can be extended to handle these window lengths and hence can be applied for both L and K .

Proposition 3.6. *Let L^* denote the greatest common divisor of L and K . All examples from Section 3.2 and Proposition 3.3 are valid if we replace the left separability with the (two-sided) weak separability and L with L^* .*

Proof. It is sufficient to prove the proposition only for Example 3.1, since the other assertions are based on it. In this case the roots μ_k from Example 3.1 have to be common roots of $W_1(z)$ defined in (21) and the polynomial $z^{K-1} + \dots + 1 = (z^K - 1)/(z - 1)$. One can immediately observe that μ_k are of form

$$\mu_k = \exp\left(\frac{2\pi ik}{L^*}\right), \quad 0 < k < L^*,$$

which completes the proof. \square

4 Extraneous roots of SSA continuation LRF

In the present section we study the behaviour of the extraneous roots for the specific LRF which is used in SSA forecasting, see [17, §5.2]. For an f.d.d. time series we express this LRF through the characteristic polynomial of the time series. Then we show the correspondence between the extraneous roots and a special system of orthogonal polynomials. Using this correspondence, first, we demonstrate that several main properties of the extraneous roots are easily proved and, second, we derive the asymptotic behaviour of the extraneous roots for the f.d.d. time series in the noise-free case.

4.1 SSA LRF and its basic properties

Let Λ be a subspace of \mathbb{C}^L such that $e_L \notin \Lambda$. Let $\{U_1, \dots, U_d\} \subseteq \mathbb{C}^L$ be the orthonormal basis of Λ and $U_k = \begin{pmatrix} U_k^\nabla \\ \pi_k \end{pmatrix}$ where $U_k^\nabla \in \mathbb{C}^{L-1}$ and $\pi_k \in \mathbb{C}$. Define $\mathcal{R} = (a_0, \dots, a_{L-1})^\top \in \mathbb{C}^{L-1}$ as

$$\mathcal{R} = \frac{1}{1 - \nu^2} \sum_{k=1}^d \pi_k \overline{U_k^\nabla}, \quad (26)$$

where $\nu^2 = |\pi_1|^2 + \dots + |\pi_d|^2 = \sum_{i=1}^d |\langle U_i, e_L \rangle|^2 < 1$. The last inequality holds since $e_L \notin \Lambda$. The following proposition is a version of [17, Ch. 5, Th. 5.2] for complex-valued time series.

Proposition 4.1. *Let F_N be a time series of difference dimension d , $1 \leq d < L$. Let $\Lambda = \mathfrak{L}^{(L)}(F_N)$. Then F_N satisfies an LRF*

$$f_{n+L} = \sum_{k=0}^{L-1} a_k f_{n+k}, \quad 0 \leq n \leq N - L - 1, \quad (27)$$

with the coefficients a_k given by (26).

The LRF (27) is abbreviated as the *SSA LRF*. Proposition 4.1 provides the base of the SSA continuation algorithms, see [17, Ch. 2]. In these algorithms one finds Λ that approximates $\mathfrak{L}^{(L)}(F_N)$ and employs the SSA LRF obtained from Λ , which approximates the SSA LRF obtained from $\mathfrak{L}^{(L)}(F_N)$. A detailed discussion of this approximation can be found in [22, 23].

Let \mathbb{R} be the conjugate to the orthogonal complement of Λ (c.f. Definition 2.3). The next proposition is a version of [17, Ch. 5, Prop. 5.5] for the complex-valued case.

Proposition 4.2. *The vector*

$$A = (-\mathcal{R}^\top, 1)^\top = (-a_0, \dots, -a_{L-1}, 1)^\top, \quad (28)$$

with \mathcal{R} given by (26), can be expressed as

$$A = c \Pi_{\mathbb{R}} e_L, \quad (29)$$

where $\Pi_{\mathbb{R}}$ is the orthogonal projector on the space \mathbb{R} and $c = (1 - \nu^2)^{-1} = \langle \Pi_{\mathbb{R}} e_L, e_L \rangle^{-1}$.

By Proposition 4.2, the SSA continuation vector (28) is equivalent to the Min-Norm prediction vector [5, 9]. One of its well known properties, providing the name *Min-Norm*, is the following; the proof can be found, for example, in [4].

Proposition 4.3. *The vector (28) yields the minimum of*

$$|a_0|^2 + |a_1|^2 + \dots + |a_{L-1}|^2 \quad (30)$$

among the vectors from \mathbb{R} .

4.2 SSA LRF and the characteristic polynomial

Let us fix a time series F_N of difference dimension d with characteristic polynomial $P(z)$. Let L be such that $d < L \leq N - d + 1$ and $\Lambda = \mathfrak{L}^{(L)}(F_N)$, i.e. the noise-free case is treated. Then $\Pi_{\mathfrak{R}} = \Pi_{\mathbb{R}}$ is the projector on the relations space \mathfrak{R} , see Definition 2.3. Consider the vector $B = \Pi_{\mathfrak{R}} e_L$. By Proposition 4.2, $A = cB$. Let $A(z)$, $B(z)$ denote the generating polynomials of A , B , see the notation in Section 3.1. By Proposition 2.2 we obtain

$$A(z)/c = B(z) = P(z)H_n(z), \quad (31)$$

where $H_n(z) = h_n^{(n)} z^n + \dots + h_1^{(n)} z + h_0^{(n)}$ is a polynomial of degree $n+1$, $n \stackrel{\text{def}}{=} L - d - 1$. The n extraneous roots of $A(z)$ are exactly the roots of the polynomial $H_n(z)$. Below we study the properties of these polynomials.

Proposition 4.4. *The vector*

$$H_n = (h_0^{(n)}, \dots, h_n^{(n)})^T \quad (32)$$

is given by

$$H_n = (\mathbf{P}^* \mathbf{P})^{-1} e_{n+1}$$

where $\mathbf{P} = \mathbf{P}^{(L)}$ is defined in (18) and \mathbf{P}^* denotes the Hermitian conjugate of \mathbf{P} . In other words, H_n is the unique solution of

$$\mathbf{T}_n H_n = e_{n+1}, \quad (33)$$

where

$$\mathbf{T}_n \stackrel{\text{def}}{=} \mathbf{P}^* \mathbf{P} \in \mathbb{C}^{(n+1) \times (n+1)}. \quad (34)$$

Proof. By Propositions 4.2 and 2.2, $B = \mathbf{P}^{(L)} H_n$ and H_n minimizes the Euclidean norm $\|\mathbf{P}^{(L)} V - e_L\|_2$ among all vectors $V \in \mathbb{C}^{L-d}$. Therefore, H_n can be expressed as the least squares solution of $\mathbf{P}^{(L)} V \approx e_L$:

$$H_n = (\mathbf{P}^* \mathbf{P})^{-1} \mathbf{P}^* e_L = (\mathbf{P}^* \mathbf{P})^{-1} e_{n+1}, \quad (35)$$

where the last equality holds since $p_d = 1$. □

Remark 4.1. In fact, the projector $\Pi_{\mathfrak{R}}$ itself can be explicitly expressed through the characteristic polynomial as $\Pi_{\mathfrak{R}} = \mathbf{P}(\mathbf{P}^* \mathbf{P})^{-1} \mathbf{P}^*$.

By construction, the matrix \mathbf{T}_n defined in (34) is a Toeplitz Hermitian matrix

$$\mathbf{T}_n = (t_{i-j})_{i,j=0}^{n,n}$$

with

$$t_k = \begin{cases} \sum_{j=0}^{d-k} \overline{p_j} p_{k+j}, & 0 \leq k \leq d, \\ 0, & k > d, \\ \overline{t_{-k}}, & k < 0. \end{cases} \quad (36)$$

Note that the coefficients t_k do not depend on n . Also, any matrix \mathbf{T}_n has no more than $2d + 1$ non-zero diagonals

$$\mathbf{T}_n = \begin{pmatrix} t_0 & \dots & t_{-d} & & & \\ \vdots & \ddots & \ddots & \ddots & & \\ t_d & \ddots & t_0 & \ddots & t_{-d} & \\ & \ddots & \ddots & \ddots & \ddots & \\ & & t_d & \ddots & t_0 & \end{pmatrix}.$$

By the way, t_k coincide with the values of the covariance function for the moving average process with coefficients p_k , see [2, Ch. 5]. If we rewrite the equation (33) as

$$\mathbf{T}G_n = e_1, \quad (37)$$

where $G = (\overline{h_n^{(n)}}, \dots, \overline{h_0^{(n)}})^T$, then we obtain the Yule-Walker equation for this process.

4.3 Orthogonal polynomials. Basic properties

In this subsection we apply the powerful theory of orthogonal polynomials on the unit circle and obtain short proofs for the basic properties of extraneous roots. Let us first rewrite the equation (33)

in terms of orthogonal polynomials. Let $t(z) = \sum_{k=-d}^d t_k z^k$, where t_k are given by (36). It is easy to see that $t(z) = P(z)\overline{P}(1/z)$ and the relation

$$t(z) = P(z)\overline{P}(\overline{z}) = |P(z)|^2 \geq 0 \quad (38)$$

holds for all $z \in \mathbb{T}_1$, where $\mathbb{T}_r = \{z \in \mathbb{C} : |z| = r\}$ denotes the circle of radius r .

For a non-negative function $w(z) \in \mathbf{L}_1(\mathbb{T}_1)$ (i.e. a Lebesgue integrable over the contour \mathbb{T}_1 function) one can define the inner product in the space of complex polynomials

$$\langle p(z), q(z) \rangle_w \stackrel{\text{def}}{=} \frac{1}{2\pi} \int_{-\pi}^{\pi} p(z)\overline{q(z)}w(z)d\theta, \quad (39)$$

where $z = e^{i\theta}$. The function $w(z)$ is called the *weight*. In particular, one can define the inner product $\langle \cdot, \cdot \rangle_t$ for the weight $t(z)$.

Proposition 4.5. *Polynomials $H_n(z)$, $n \geq 0$, defined by (31) form an orthogonal system with respect to $\langle \cdot, \cdot \rangle_t$.*

Proof. It is evident that $t_{k-l} = \langle 1, z^{k-l} \rangle_t = \langle z^l, z^k \rangle_t$ for all $k, l \in \mathbb{Z}$. Therefore, we can rewrite the equation (33) as

$$\langle H_n(z), z^k \rangle_t = \sum_{l=0}^n h_l^{(n)} t_{k-l} = \begin{cases} 0, & 0 \leq k < n, \\ 1, & k = n. \end{cases}$$

Hence $\langle H_n(z), H_m(z) \rangle_t = 0$ for $m \neq n$ and

$$\|H_n(z)\|_t^2 \stackrel{\text{def}}{=} \langle H_n(z), H_n(z) \rangle_t = h_n^{(n)} \neq 0$$

for all $n \geq 0$, where $h_n^{(n)}$ is given in (32), and the assertion is proved. \square

Remark 4.2. Let $\Phi_n(z)$, $n \geq 0$, be a system of polynomials of degree n , which are orthogonal with respect to some weight $w(z)$ (i.e. $\langle \Phi_n(z), \Phi_m(z) \rangle_w = 0$ for $m \neq n$, $\|\Phi_n(z)\|_w^2 \neq 0$ for all n). Then Φ_n are defined uniquely up to constant factors. This follows from the properties of the standard orthogonalization process applied to the sequence $\{1, x, x^2, \dots\}$, see [13, Ch. IX, §6].

Thus, Proposition 4.5 is a characterization of $H_n(z)$.

Remark 4.3. The system Φ_n from Remark 4.2 is also orthogonal with respect to the weight $\alpha w(z)$ for all $\alpha > 0$.

Now let us recall the well-known property of orthogonal polynomials, which has a consequence for the extraneous roots. We provide the proof just to show its simplicity, it also can be found in [18, Ch. 1].

Theorem 4.1. *Let $\Phi_n(z)$ be an orthogonal polynomial of degree n with respect to $\langle \cdot, \cdot \rangle_w$. Then $|z_0| < 1$ if $\Phi_n(z_0) = 0$, i.e. the roots of the polynomial Φ_n are located inside the unit circle.*

Proof. If Φ_n has a root z_0 then $\Phi_n(z) = Q(z)(z - z_0)$, where $\deg Q = n - 1$, and hence $\langle \Phi_n, Q \rangle_w = 0$. Then

$$\begin{aligned} \|Q(z)\|_w^2 &= \langle zQ(z), zQ(z) \rangle_w = \|zQ(z)\|_w^2 = \\ &= \|Q(z)z_0 + \Phi_n(z)\|_w^2 = |z_0|^2 \|Q(z)\|_w^2 + \|\Phi_n(z)\|_w^2, \end{aligned}$$

where the last equality follows from the orthogonality of polynomials. Rewriting we have $(1 - |z_0|^2) \|Q(z)\|_w^2 = \|\Phi_n(z)\|_w^2 > 0$, which completes the proof. \square

In particular, we obtain the following result for the SSA LRF (Min-Norm prediction), which is well known [4, 5, 15], but the proof is often too complicated.

Corollary 4.1. *All extraneous roots of SSA LRF lie inside the unit circle.*

Next, we provide another clear proof for the well-known result [15] on the correspondence between the extraneous roots for the forward and backward SSA LRFs. By the backward SSA LRF we mean the SSA LRF for the reversed time series \widehat{F}_N ; the forward LRF means the standard SSA LRF. It is easy to observe (for example, using the basis in Proposition 2.2) that \widehat{F}_N is an f.d.d. time series with the characteristic polynomial $p_0^{-1} \widehat{P}(z)$, where $\widehat{P}(z)$ denotes the reversed polynomial $p_d + p_{d-1}z + \dots + p_0z^d$. This polynomial has roots λ_k^{-1} with multiplicities ν_k (c.f. (8)). For convenience, for any polynomial $B(z) = b_r z^r + \dots + b_1 z + b_0$, $b_r \neq 0$, we introduce the notation

$$B^*(z) \stackrel{\text{def}}{=} \widehat{B}(z) = \overline{b_0} z^r + \dots + \overline{b_{r-1}} z + \overline{b_r}. \quad (40)$$

Then

$$|P(z)|^2 = |P^*(z)|^2 = |\widehat{P}(\bar{z})|^2, \quad |z| = 1. \quad (41)$$

This fact enables us to easily show the following.

Proposition 4.6. *The extraneous roots of the backward and forward SSA LRFs are conjugate.*

Proof. Let $u(z) = |p_0|^{-2}|\widehat{P}(z)|^2$ denote the weight for the reversed time series. Then for any $Q(z)$ and $S(z)$ we obtain

$$\langle Q(z), S(z) \rangle_u = |p_0|^{-2} \langle \overline{P}(z), \overline{Q}(z) \rangle_t,$$

which follows from (39). Therefore, each polynomial from a set of polynomials orthogonal with respect to $u(z)$ has the roots which are conjugate to the roots of the corresponding $H_n(z)$, and the assertion is proved. \square

Note that it is common to define the backward SSA LRF with conjugation of the coefficients [5]. In this case, the assertion of Proposition 4.6 changes; specifically, the extraneous roots of the backward and forward SSA LRFs coincide (as stated in Introduction).

4.4 Asymptotic properties

In this section we present a review of recent results on the asymptotic distribution of the roots of orthogonal polynomials which can be used to study the extraneous roots for the SSA LRF. We mainly follow [20] and [21] and present all the facts in a unified and simplified manner, specializing the results for the weight $t(z)$ defined in (38).

For convenience, in addition to \mathbb{T}_r we define

$$\begin{aligned} \mathbb{D}_r(a) &\stackrel{\text{def}}{=} \{z \in \mathbb{C} : |z - a| < r\}, \\ \mathbb{D}_r &\stackrel{\text{def}}{=} \mathbb{D}_r(0), \\ \mathbb{D}_r^c &\stackrel{\text{def}}{=} \{z \in \mathbb{C} : |z| \leq r\}, \end{aligned}$$

for $r > 0$ and $a \in \mathbb{C}$.

Let us consider the characteristic polynomial $P(z)$ defined in (8) and the orthogonal polynomials $H_n(z)$ defined in (31). Using a transformation similar to (41) we can “transfer” all the roots inside the closed unit disk \mathbb{D}_1^c . Define

$$C(z) = \prod_{k:|\lambda_k|\leq 1} (z - \lambda_k)^{\nu_k} \prod_{l:|\lambda_l|>1} (z - \overline{\lambda_l^{-1}})^{\nu_l}. \quad (42)$$

Then

$$|P(z)|^2 = c|C(z)|^2, \quad (43)$$

where c is some positive constant. Hence, by Remark 4.3, the weight $t(z) = |P(z)|^2$ and the weight $u(z) = |C(z)|^2$ generate the same system of orthogonal polynomials H_n , and the roots of $C(z)$ are inside the closed unit disk.

Note that in the representation (42) the pairs of roots of $P(z)$ which are related by $\lambda_k = \overline{\lambda_l^{-1}}$ are glued together. The *normalized representation* of $C(z)$, defined in (44), takes this into account. We set

$$C(z) = (z - a_1)^{m_1} \dots (z - a_s)^{m_s} \quad (44)$$

where all a_k , $1 \leq k \leq s$, are distinct.

Let us also introduce several definitions. The radius $\rho = \max_{1 \leq k \leq s} |a_k| \leq 1$ is called the *critical radius* for the polynomial $P(z)$. The circle \mathbb{T}_ρ is called the *critical circle*. Moreover, let the roots of $C(z)$ be ordered such that the first u roots are on the critical circle and the other roots are inside it, i.e. $|a_1| = \dots = |a_u| = \rho$ and $|a_k| < \rho$ for $k > u$. We will call a_1, \dots, a_u the *leading roots* of $C(z)$. Let also the first ℓ roots a_1, \dots, a_ℓ be of greatest multiplicity M among a_1, \dots, a_u , i.e. $m_s = M$ for all $1 \leq s \leq \ell$ and $m_j < M$ for $\ell < j \leq u$.

Proposition 4.7 ([20, Prop. 1, Th. 3]). *Let $\rho < 1$.*

1. *For any $\varepsilon > 0$ there exists $N_1(\varepsilon)$ such that all zeros of $H_n(z)$ are inside $\mathbb{D}_{\rho+\varepsilon}$ for all $n \geq N_1$.*
2. *For any $\varepsilon > 0$ there exists $N_2(\varepsilon)$ such that the closed disk $\mathbb{D}_{\rho-\varepsilon}^c$ contains at most $\ell - 1$ roots of $H_n(z)$ for all $n \geq N_2$.*

Proposition 4.8 ([21, Cor. 1]). *If $\rho = 1$ then for any $\varepsilon > 0$ there exists $N_2(\varepsilon)$ such that for all $n \geq N_2$ the closed disk $\mathbb{D}_{1-\varepsilon}^c$ contains at most $u - 1$ roots of $H_n(z)$.*

These two propositions imply that the majority of roots tend uniformly to the critical circle \mathbb{T}_ρ (they are called *general roots*), and a bounded number of roots stay strictly inside \mathbb{D}_ρ . (called *spurious roots*). The spurious roots chaotically float inside \mathbb{D}_ρ (when changing n), but, roughly speaking, they are asymptotically close to zeros of the functions G_n defined by

$$G_n(z) = \begin{cases} \sum_{k=1}^{\ell} \frac{a_k^{n-M+d+1} C^*(a_k)}{(z - a_k) C^{(M)}(a_k)}, & \rho < 1, \\ \sum_{k=1}^u \frac{a_k^{n+d+1} (-1)^{\nu_k} \nu_k (C^*)^{(\nu_k)}(a_k)}{(z - a_k) C^{(\nu_k)}(a_k)}, & \rho = 1, \end{cases}$$

where $f^{(m)}$ stands for the m -th derivative of a function f and $C^*(z)$ is defined as in (40). See [3, 20, 21] for exact formulations; see also an example with spurious roots at the end of this section.

Let us look at the behaviour of general extraneous roots. We present only an informal summary in order not to overload the exposition with technicalities. Precise and mathematically strict results can be found in [20, Th. 4] and [21, Th. 5] (for $\rho < 1$ and $\rho = 1$, respectively).

- The asymptotics for the absolute values of the general roots are

$$|z_i^{(n)}| = \begin{cases} \rho \left(1 + M \frac{\log(n)}{n} + O\left(\frac{1}{n}\right) \right), & \rho < 1, \\ 1 + \frac{\log(n)}{n} + O\left(\frac{1}{n}\right), & \rho = 1. \end{cases}$$

- For a function f , define by $Z_\varepsilon(f) = \bigcup_{a: f(a)=0} \mathbb{D}_\varepsilon(a)$ the ε -vicinity of its zero set. Let us denote

$$\mathcal{B}_{C,\varepsilon} = \begin{cases} \bigcup_{k=1}^{\ell} \mathbb{D}_\varepsilon(a_k), & \rho < 1, \\ \bigcup_{k=1}^u \mathbb{D}_\varepsilon(a_k), & \rho = 1. \end{cases}$$

Then for large n and small δ there are no spurious roots in the area

$$\mathcal{A}_{\varepsilon,\delta,n} = \{z : ||z| - \rho| < \delta\} \setminus (\mathcal{B}_{C,\varepsilon} \cup Z_\varepsilon(G_n)).$$

- Let \mathcal{D} be a connected sub-area of $\mathcal{A}_{\varepsilon,\delta,n}$ and $z_1^{(n)}, \dots, z_r^{(n)}$ denote the roots of $H^n(z)$ in \mathcal{D} , ordered by the magnitude of their arguments. Then for i, j such that $1 \leq i, (i + j) \leq r$ we have

$$\arg(z_{i+j}^{(n)}) - \arg(z_i^{(n)}) = \frac{2\pi j}{n} + O\left(\frac{1}{n^2}\right).$$

These results can also be found in [3] in a weaker form. Let us make some comments.

- The general roots show an asymptotic angular equidistribution. The order of their convergence to the critical circle differs for $\rho = 1$ and $\rho < 1$. In the latter case the convergence rate depends on the maximum multiplicity M of the leading roots of $C(z)$.

- For $\rho < 1$ only the roots a_1, \dots, a_ℓ (i.e. the leading roots with maximum multiplicity M) significantly affect the asymptotic behaviour of extraneous roots: they determine the number of spurious roots, and their vicinities are excluded from $\mathcal{A}_{\varepsilon, \delta, n}$; however for $\rho = 1$ all leading roots of $C(z)$ affect the asymptotic behaviour.
- In the real-valued case each signal root is usually accompanied with its conjugate (when the signal is a sum of modulated cosines). Therefore, there is likely to be more than one leading root of $C(z)$, and the spurious roots are likely to appear.

Example 4.1. [General and spurious roots] Consider a sum of two exponentially modulated (with the same exponent) cosine time series. Since each cosine corresponds to two conjugate roots, we have $\ell = u = 4$ and $M = 1$. The roots of the SSA LRF are plotted in Fig. 4 (recall that $L = n + d + 1$).

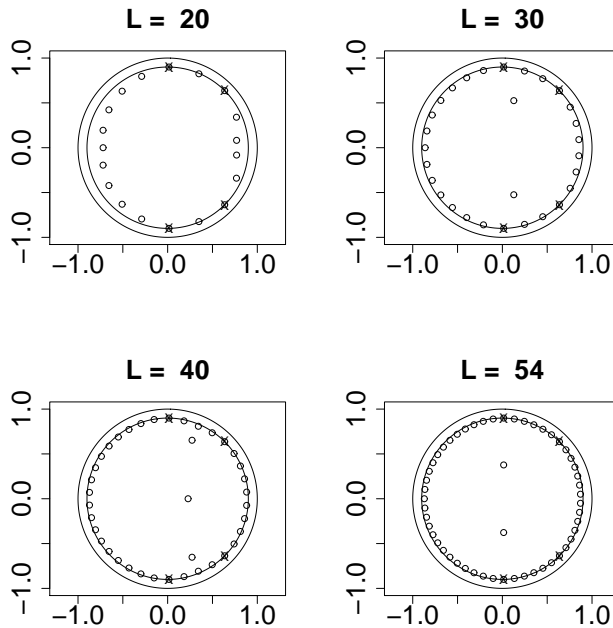


Figure 4: Signal (x) and extraneous (o) roots, $f_n = 0.9^n (\cos(\frac{2\pi}{8}n) + \cos(2\pi \sin(0.25)n))$

As one can see in Fig. 4, the spurious roots float inside the critical circle; for some L the spurious roots disappear. The coefficient $\sin(0.25)$ was chosen to ensure the rational independence of the leading roots of $C(z)$. If the majority of leading roots is rationally dependent (i.e. some roots are expressed through linear combinations of others with rational coefficients), then the spurious roots float more regularly, see [3, 20, 21] for details.

4.5 Several applications and remarks

First, let us show a connection between separability and the behaviour of extraneous roots. We demonstrate that the approximate [17, Ch. 6] (left) separability of a linear time series from a periodic time series can be informally justified by the condition of separability developed in Section 3. Let $W^{(L,0)}(x) \stackrel{\text{def}}{=} W_1(x)$ and $W^{(L,1)}(z) \stackrel{\text{def}}{=} W_1^{(1)}(z)$ be the polynomials from Example 3.4. The roots of $W^{(L,0)}$ are always on the unit circle and uniformly distributed with equal angles between adjacent roots on it (without number 1). One can show that $W^{(L,1)}(x)$ for $L = 1, 2, \dots$ form an orthogonal system with respect to the weight $P(z) = |z - 1|^2$, and by results in Section 4.4 the roots of $W^{(L,1)}(x)$ are

distributed asymptotically in the same way as the roots of $W^{(L,0)}(x)$. In Fig. 5 one can see the roots of $W^{(L,0)}(x)$ and $W^{(L,1)}(x)$ for large L .

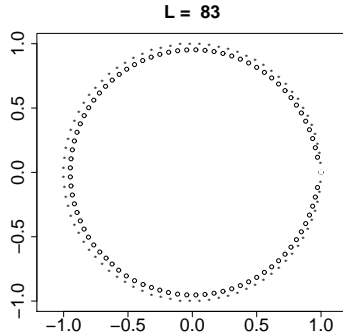


Figure 5: The roots of $W^{(L,0)}$ and $W^{(L,1)}$.

We might say that this validates the asymptotic separability of an arbitrary periodic series from a linear function. The same can be shown for any polynomial time series. Note that this observation does not pretend to be a rigorous assertion. A recent theoretical study of asymptotic and approximate separability can be found in [23].

Let us now consider the problem of determining the signal roots from the approximate SSA LRF (the Min-Norm estimation of the main roots). Theorem 4.1 is the base of the *root-Min-Norm* [4, 9, 12] method for estimation of exponents. Under the assumption that all signal roots have modulus not less than 1, one selects the greatest by absolute value d roots from the estimated LRF to be the signal roots.

If all signal roots have the absolute values greater than 1, we conclude from the asymptotic distribution of the roots that the root-Min-Norm approach is applicable. In the presence of moderate noise, the main and extraneous roots of the estimated LRF are close to those of the noiseless LRF, see, for example, [4]. In Fig. 6 a noisy time series with its SSA LRF roots are depicted. If some signal

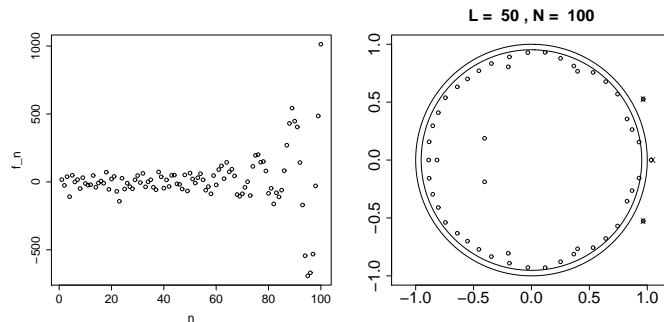


Figure 6: Left: the time series, Right: the signal roots (x) and the extraneous roots of the SSA LRF (o), $f_n = 1.05^n + \frac{1}{10}1.1^n \cos(0.5n) + \varepsilon_n$, ε_n are i.i.d. $\mathcal{N}(0, 50^2)$,

roots are on the unit circle, then this approach behaves worse.

If all signal roots have modulus less than 1, then the naive root-Min-Norm approach is not applicable. Even if the signal roots are only on the critical circle, the extraneous roots may have larger

modulus, as shown in Fig. 7. In this case one can use the backward SSA LRF to get the main and extraneous roots separated, see [4].

Assume now that within signal roots there are roots both outside and inside the unit circle. In this case one can consider the backward and forward SSA LRF simultaneously and use the relations between the extraneous and signal roots of both LRFs.

Finally, we discuss the notion of multiplicity. Consider the case of a single real exponent time series, modulated by a quadratic polynomial. There is one signal root of multiplicity 3. In Fig. 7 the roots of SSA LRF for the noise-free case and the noise case are depicted. As one can see in

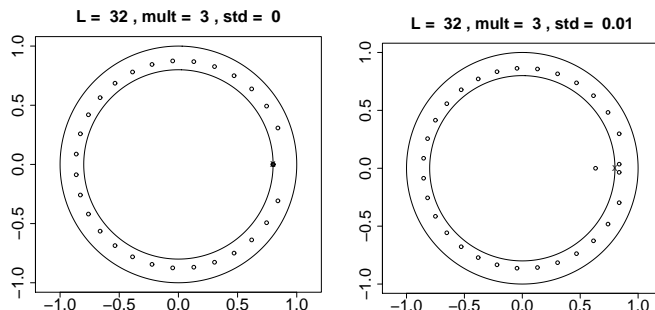


Figure 7: The signal roots (x) and the extraneous roots of the SSA LRF (o), $f_n = n^2 0.8^n + \text{std} \cdot \varepsilon_n$, ε_n are i.i.d. $\mathcal{N}(0, 1)$, $N = 150$.

Fig. 7, in the presence of noise a multiple signal root splits into three separate roots. Nevertheless, the extraneous roots behave as if there was a multiple signal root. Indeed, the extraneous roots are close to those in the noise-free case. In addition, no “spurious” roots appear inside the critical circle.

References

- [1] GOVIL, N.K. and RAHMAN, Q.I., (1968). On the Eneström-Keakeya theorem. *Tohoku Mathematical Journal*, **20:2**, 126–136.
- [2] ANDERSON, T.W., (1971). *The statistical analysis of time series*, John Wiley & Sons, New York.
- [3] SZABADOS, J., (1979). On some problems connected with orthogonal polynomials. *Acta Mathematica Academiae Scientiarum Hungaricae*, **33:1–2**, 197–210.
- [4] KUMARESAN, R. and TUFTS, D.W., (1982). Estimating the Parameters of Exponentially Damped Sinusoids and Pole-Zero Modelling in Noise. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **30:6**, 833–840.
- [5] KUMARESAN, R., (1983). On the Zeros of the Linear Prediction-Error Filter for Deterministic Signals. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **31:1**, 217–220.
- [6] HEINIG, G. and ROST, K., (1984). *Algebraic methods for Toeplitz-like matrices and operators*, Akademie Verlag, Berlin.
- [7] PAKULA, L., (1987). Asymptotic Zero Distribution of Orthogonal Polynomials in Sinusoidal Frequency Estimation. *IEEE Transactions on Information Theory*, **33:4**, 569–576
- [8] MHASKAR, H.N. and SAFF, E.B., (1990). On the distribution of zeros of polynomials orthogonal on the unit circle. *Journal of Approximation Theory*, **63**, 30–38.

- [9] KRIM, H., FORSTER, P. and PROAKIS, J.G., (1992). Operator approach to performance analysis of root-MUSIC and root-min-norm. *IEEE Transactions on Signal Processing*, **40:7**, 1687–1696.
- [10] VAN DER WAERDEN, B.L., (1994). *Algebra I*, 9th ed., Springer Verlag.
- [11] ELSNER, J. and TSONIS, A., (1996). *Singular Spectrum Analysis. A New Tool in Time Series Analysis*, Plenum Press, New York.
- [12] STOICA, P. and MOSES, R.L., (1997). *Introduction to spectral analysis*, Prentice Hall.
- [13] GANTMACHER, F.R., (1998). *The Theory of Matrices*, Chelsea, Providence, RI.
- [14] HALL, M. JR., (1998). *Combinatorial theory*, Wiley, New York.
- [15] BEZERRA, L.H. and BAZAN, F.S.V., (1998). Eigenvalue Locations of Generalized Companion Predictor Matrices. *SIAM Journal on Matrix Analysis and Applications*, **19:4**, 886–897.
- [16] NEKRUTKIN, V.V., (1999). Approximation spaces and continuation of time series, In S. M. Ermakov and Yu. N. Kashtanov (Eds.), *Statistical Models with Applications in Econometrics and Neighbouring Fields*, University of St.Petersburg, St.Petersburg, 3–32 (in russian).
- [17] GOLYANDINA, N., NEKRUTKIN, V. and ZHIGLJAVSKY, A., (2001). *Analysis of Time Series Structure: SSA and Related Techniques*, Chapman & Hall/CRC, Boca Raton.
- [18] SIMON, B., (2005). *Orthogonal Polynomials on the Unit Circle, Part 1: Classical Theory*, AMS Colloquium Series, American Mathematical Society, Providence, RI.
- [19] BADEAU, R., DAVID, B. and RICHARD, G., (2006). High-Resolution Spectral Analysis of Mixtures of Complex Exponentials Modulated by Polynomials. *IEEE Transactions on Signal Processing*, **54:4**, 1341–1350.
- [20] MARTÍNEZ-FINKELSHEIN, A., MCCLAUGHLIN, K.T.-R. and Saff, E.B., (2006). Szegő Orthogonal Polynomials with Respect to an Analytic Weight: Canonical Representation and Strong Asymptotics. *Constructive Approximation*, **24:3**, 319–363.
- [21] MARTÍNEZ-FINKELSHEIN, A., MCCLAUGHLIN, K.T.-R. and Saff, E.B., (2006). Asymptotics of orthogonal polynomials with respect to an analytic weight with algebraic singularities on the circle. *International Mathematics Research Notices*, **2006:91426**, 1–47.
- [22] GOLYANDINA, N., (2010). On the choice of parameters in Singular Spectrum Analysis and related subspace-based methods. <http://arxiv.org/abs/1005.4374>.
- [23] NEKRUTKIN, V., (2010). Perturbation expansions of signal subspaces for long signals. <http://arxiv.org/abs/1001.1051>.