Chilean Journal of Statistics
Vol. 3, No. 1, April 2012, 43–56

Time Series
Research Paper

On the singular values of the Hankel matrix with application in singular spectrum analysis

Rahim Mahmoudvand[1,*] and Mohammad Zokaei[1]
[1]Department of Statistics, Shahid Beheshti University, Tehran, Iran

(Received: 21 June 2011 · Accepted in final form: 24 August 2011)

Abstract

Hankel matrices are an important family of matrices that play a fundamental role in diverse fields of study, such as computer science, engineering, mathematics and statistics. In this paper, we study the behavior of the singular values of the Hankel matrix by changing its dimension. In addition, as an application, we use the obtained results for choosing the optimal values of the parameters of singular spectrum analysis, which is a powerful technique in time series based on the Hankel matrix.

Keywords: Eigenvalues · Singular spectrum analysis.

Mathematics Subject Classification: Primary 15A18 · Secondary 37M10.

1. Introduction

A Hankel matrix can be finite or infinite and its (i, j) entry is a function of $i+j$; see Widom (1966). In other words, a matrix whose entries are the same along the anti-diagonals is called the Hankel matrix. Specifically, an $L \times K$ Hankel matrix $\mathbf{H}$ is a rectangular matrix of the form

$$\mathbf{H} = \begin{pmatrix} h_1 & h_2 & \dots & h_K \\ h_2 & h_3 & \dots & h_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ h_L & h_{L+1} & \dots & h_N \end{pmatrix}, \tag{1}$$

where $K = N - L + 1$.

Hankel matrices play many roles in diverse areas of mathematics, such as approximation and interpolation theory, stability theory, system theory, theory of moments and theory

*Corresponding author: Rahim Mahmoudvand. Department of Statistics, Shahid Beheshti University, PO. Box 1983963113, Evin, Tehran, Iran. Email: r.mahmodvand@gmail.com

ISSN: 0718-7912 (print)/ISSN: 0718-7920 (online)

http://www.soche.cl/chjs

of orthogonal polynomials, as well as in communication and control engineering, including filter design, identification, model reduction and broadband matching; for more details, see Peller (2003). Thus, this type of matrices has been subjected to intensive study with respect to its spectrum (collection of eigenvalues) and many interesting results were derived. However, closed form computation of eigenvalues is not known and, consequently, the effect of changing the dimension of the matrix on its eigenvalues have not been investigated in detail.

In recent years, singular spectrum analysis (SSA), a relatively novel, but powerful technique in time series analysis, has been developed and applied to many practical problems; see, e.g., Golyandina et al. (2001), Hassani et al. (2009), Hassani and Thomakos (2010) and references therein. The SSA decomposes the original time series into a sum of small numbers of interpretable components, such as slowly varying trend, oscillatory component and noise. The basic SSA method consists of two complementary stages: decomposition and reconstruction; each stage includes two separate steps. At the first stage, we decompose the series and, at the second stage, we reconstruct the noise free series by using the reconstructed series for forecasting new data points.

A short description of the SSA technique is given in the next section. For more explanations and comparison with other time series analysis techniques, refer to Hassani (2007).

The whole procedure of the SSA technique depends upon two parameters:

(i) The window length, which is usually denoted by L.
(ii) The number of needed singular values, denoted by r, for reconstruction.

Improper choice values of parameters L or r may yield incomplete reconstruction and misleading results in forecasting.

Considering a series of length N, Elsner and Tsonis (1996) provided some discussion and remarked that choosing $L = N/4$ is a common practice. Golyandina et al. (2001) recommended that L should be large enough, but not larger than $N/2$. Large values of L allow longer period oscillations to be resolved, but choosing L too large leaves too few observations from which to estimate the covariance matrix of the L variables. It should be noted that variations in L may influence separability feature of the SSA technique: the orthogonality and closeness of the singular values. There are some methods for selecting L. For example, the weighted correlation between the signal and noise component has been proposed in Golyandina et al. (2001) to determine the suitable value of L in terms of separability.

Although considerable attempt and various techniques have been taken into account for selecting the proper value of L, but there is not enough algebraic and theoretical materials for choosing L and r. The aim of his paper is to obtain some theoretical properties of the singular values of the Hankel matrix that can be used directly for choosing proper values of the two parameters of the SSA.

The outline of this paper is as follows. Section 2 describes the SSA technique and also shows the importance of a Hankel matrix for this technique. Section 3 provides the main results of the paper. Section 4 discusses some examples and an application of the obtained results. Section 5 sketches some conclusion of this work.

2. Singular Spectrum Analysis

In this section, we briefly introduce stages of the SSA method and discuss the importance of using a Hankel matrix in the development of this technique.

2.1 Stage I: decomposition

1st step: embedding. Embedding is as a mapping that transfers a one-dimensional time series $Y_N = (y_1, \ldots, y_N)$ into the multi-dimensional series $X_1, \ldots, X_K$ with vectors $X_i = (y_i, \ldots, y_{i+L-1})^\top \in \mathbf{R}^L$, where L $(2 \leq L \leq N-1)$ is the window length and $K = N-L+1$. The result of this step is the trajectory matrix

$$\mathbf{X} = (X_1, \ldots, X_K) = (x_{ij})_{i,j=1}^{L,K}. \tag{2}$$

Note that the matrix given in Equation (2) is a Hankel matrix as defined in Equation (1).

2nd step: singular value decomposition (SVD). In this step, we perform the SVD of $\mathbf{X}$. Denote by $\lambda_1, \ldots, \lambda_L$ the eigenvalues of $\mathbf{X}\mathbf{X}^\top$ arranged in the decreasing order $(\lambda_1 \geq \cdots \geq \lambda_L \geq 0)$ and by $U_1, \ldots, U_L$ the corresponding eigenvectors. The SVD of $\mathbf{X}$ can be written as $\mathbf{X} = \mathbf{X}_1 + \cdots + \mathbf{X}_L$, where $\mathbf{X}_i = \sqrt{\lambda_i} U_i {V_i}^\top$ and $V_i = \mathbf{X}^\top U_i / \sqrt{\lambda_i}$ (if $\lambda_i = 0$ we set $\mathbf{X}_i = 0$).

2.2 Stage II: reconstruction

1st step: grouping. The grouping step corresponds to splitting the elementary matrices into several groups and summing the matrices within each group. Let $I = \{i_1, \ldots, i_p\}$, for $p < L$, be a group of indices $i_1, \ldots, i_p$. Then, the matrix $\mathbf{X}_I$ corresponding to the group I is defined as $\mathbf{X}_I = \mathbf{X}_{i_1} + \cdots + \mathbf{X}_{i_p}$. The split of the set of indices $\{1, \ldots, L\}$ into disjoint subsets $I_1, \ldots, I_m$ corresponds to the representation $\mathbf{X} = \mathbf{X}_{I_1} + \cdots + \mathbf{X}_{I_m}$. The procedure of choosing the sets $I_1, \ldots, I_m$ is called the grouping. For a given group I, the contribution of the component $\mathbf{X}_I$ is measured by the share of the corresponding eigenvalues $\sum_{i \in I} \lambda_i / \sum_{i=1}^{d} \lambda_i$, where d is the rank of $\mathbf{X}$.

2nd step: diagonal averaging. The purpose of diagonal averaging is to transform a matrix $\mathbf{Z}$ to the form of a Hankel matrix $\mathbf{HZ}$, which can be subsequently converted to a time series. If z_{ij} stands for an element of a matrix $\mathbf{Z}$, then the kth term of the resulting series is obtained by averaging z_{ij} for all i, j such that $i + j = k + 1$. Hankelization $\mathbf{HZ}$ is an optimal procedure, which is nearest to $\mathbf{Z}$ with respect to the matrix norm.

3. Theoretical Results

Along the paper, the matrices to be considered are over the field of the real numbers. In addition, we consider different values of L, whereas N is supposed to be fixed. Recall that, for any operator $\mathbf{A}$, the operator $\mathbf{A}\mathbf{A}^\top$ is always positive, and its unique positive square root is denoted by $|\mathbf{A}|$. The eigenvalues of $|\mathbf{A}|$ counted with multiplicities are called the singular values of $\mathbf{A}$. In this section, we provide the main results of the paper.

3.1 On sum of square of the singular values of a Hankel matrix

Let $\lambda_j^{L,N}$ denote the jth ordered eigenvalue of $\mathbf{H}\mathbf{H}^\top$. Then, for a fixed value of L, the trace of $\mathbf{H}\mathbf{H}^\top$ is given by

$$T_{\mathbf{H}}^{L,N} = \mathrm{tr}(\mathbf{H}\mathbf{H}^\top) = \sum_{j=1}^{L} \lambda_j^{L,N}. \tag{3}$$

The behavior of $T_{\mathbf{H}}^{L,N}$ given in Equation (3), with respect to different values of L, is considered in the following theorem.

THEOREM 3.1 Consider the Hankel matrix $\mathbf{H}$ as defined in Equation (1). Then,

$$T_{\mathbf{H}}^{L,N} = \sum_{j=1}^{N} w_j^{L,N} h_j^2,$$

where $w_j^{L,N} = \min\{\min\{L, K\}, j, L + K - j\} = w_j^{K,N}$.

PROOF Applying definition of $\mathbf{H}$ as given in Equation (1), we have

$$T_{\mathbf{H}}^{L,N} = \sum_{i=1}^{L} \sum_{j=i}^{N-L+i} h_j^2. \tag{4}$$

Changing the order of the summations in Equation (4), we get

$$T_{\mathbf{H}}^{L,N} = \sum_{j=1}^{N} C_{j,L,N} h_j^2,$$

where $C_{j,L,N} = \min\{j, L\} - \max\{1, j - N + L\} + 1$. Therefore, we only need to show that $C_{j,L,N} = w_j^{L,N}$, for all j and L. We consider two cases: $L \leq K$ and $L > K$. For the first case, we have

$$C_{j,L,N} = \begin{cases} j, & 1 \leq j \leq L; \\ L, & L + 1 \leq j \leq K; \\ N - j + 1, & K + 1 \leq j \leq N, \end{cases}$$

which is exactly equals to $w_j^{L,N}$. Similarly for the second case, we get

$$C_{j,L,N} = \begin{cases} j, & 1 \leq j \leq K; \\ K, & K + 1 \leq j \leq L; \\ N - j + 1, & L + 1 \leq j \leq N; \end{cases}$$

and again is equal to $w_j^{L,N}$, for $L > K$. ■

The weight $w_j^{L,N}$ defined in Theorem 3.1 can be written in the functional form

$$w_j^{L,N} = \frac{N+1}{2} - \frac{\left|\frac{N+1}{2} - L\right|}{2} - \frac{\left|\frac{N+1}{2} - j\right|}{2} - \frac{\left|\left|\frac{N+1}{2} - L\right| - \left|\frac{N+1}{2} - j\right|\right|}{2}. \tag{5}$$

Equation (5) shows that

- $w_j^{L,N}$ is a concave function of L for all j, where $j \in \{1, \ldots, N\}$;
- $w_j^{L,N}$ is a concave function of j for all L, where $L \in \{2, \ldots, N - 1\}$;
- $w_j^{L,N}$ is a symmetric function around line $(N + 1)/2$ with respect to j and L.

The above mentioned results imply that the behavior of the quantity $T_{\mathbf{H}}^{L,N}$ is similar on two intervals $2 \leq L \leq [(N + 1)/2]$ and $[(N + 1)/2] + 1 \leq L \leq N - 1$, where, as usual, $[x]$ denotes the integer part of the number x. Therefore, we only need to consider one of these intervals.

Theorem 3.2 Let $T_{\mathbf{H}}^{L,N}$ be defined as in Equation (3). Then, $T_{\mathbf{H}}^{L,N}$ is an increasing function of L on $\{2, \ldots, [(N+1)/2]\}$, a decreasing function on $\{[(N+1)/2]+1, \ldots, N-1\}$, and

$$\max T_{\mathbf{H}}^{L,N} = T_{\mathbf{H}}^{\left[\frac{N+1}{2}\right],N}.$$

Proof First, we show that $w_j^{L,N}$ is an increasing function of L on $\{2, \ldots, [(N+1)/2]\}$. Let L_1 and L_2 be two arbitrary values, where $L_1 < L_2 \leq [(N+1)/2]$. From the definition of $w_j^{L,N}$, we have

$$w_j^{L_2,N} - w_j^{L_1,N} = \begin{cases} 0, & 1 \leq j \leq L_1; \\ j - L_1, & L_1 + 1 \leq j \leq L_2; \\ L_2 - L_1, & L_2 + 1 \leq j \leq N - L_2 + 1; \\ N - j + 1 - L_1, & N - L_2 + 2 \leq j \leq N - L_1 + 1; \\ 0, & N - L_1 + 2 \leq j \leq N. \end{cases}$$

Therefore, $w_j^{L_2,N} - w_j^{L_1,N} \geq 0$, for all j, and inequality is strict for some j. Thus,

$$T_{\mathbf{H}}^{L_2,N} - T_{\mathbf{H}}^{L_1,N} = \sum_{j=1}^{N} \left(w_j^{L_2,N} - w_j^{L_1,N} \right) h_j^2 > 0. \quad (6)$$

This confirms that $T_{\mathbf{H}}^{L,N}$ is an increasing function of L on $\{2, \ldots, [(N+1)/2]\}$. Similar approach for the set $\{[(N+1)/2]+1, \ldots, N-1\}$ implies that $T_{\mathbf{H}}^{L,N}$ is a decreasing function of L on this interval. Note also that $T_{\mathbf{H}}^{L_2,N} - T_{\mathbf{H}}^{L_1,N}$ in Equation (6) increases as the value of L_2 increases too proving that $T_{\mathbf{H}}^{L,N}$ is an increasing function on $\{2, \ldots, [(N+1)/2]\}$. Therefore, the maximum value of $T_{\mathbf{H}}^{L_2,N}$ is attained at the maximum value of L, which is $[(N+1)/2]$. ■

Corollary 3.3 Let $L_{\max}$ denote the value of L such that $T_{\mathbf{H}}^{L,N} \leq T_{\mathbf{H}}^{L_{\max},N}$, for all L, and the inequality to be strict for some values of L. Then,

$$L_{\max} = \begin{cases} \dfrac{N+1}{2}, & \text{if } N \text{ is odd;} \\ \dfrac{N}{2} \quad \text{and} \quad \dfrac{N}{2}+1, & \text{if } N \text{ is even.} \end{cases}$$

Corollary 3.3 shows that $L = \text{median}\{1, \ldots, N\}$ maximizes the sum of squares of the Hankel matrix singular values with fixed values of N. Applying Corollary 3.3 and Equation (5), we can show that

$$w_j^{L_{\max},N} = \frac{N+1}{2} - \left| \frac{N+1}{2} - j \right|. \quad (7)$$

Equation (7) shows that $h_{[(N+1)/2]}$ has maximum weight at $T_{\mathbf{H}}^{L,N}$.

3.2 Eigenvalues of $\mathbf{HH}^\top$ and rank of $\mathbf{H}$

Here, some inequalities between the ordered eigenvalues for different values of L are derived. According to Cauchy's interlacing theorem, it can be given the following theorem; see Bhatia (1997).

Theorem 3.4 Let $\mathbf{H}$ be an $L \times K$ Hankel matrix as defined in Equation (1). Then, we have

$$\lambda_j^{L,N} \geq \lambda_j^{L-m,N-m} \geq \lambda_{j+m}^{L,N}, \quad j = 1, \ldots, L-m,$$

where m is a number belonging to the set $\{1, \ldots, L-1\}$.

Proof Consider the partition

$$\mathbf{HH}^\top = \begin{pmatrix} \mathbf{HH}^\top_{(1)} & \mathbf{HH}^\top_{(3)} \\ \mathbf{HH}^\top_{(2)} & \mathbf{HH}^\top_{(4)} \end{pmatrix},$$

where

$$\mathbf{HH}^\top_{(1)} = \begin{pmatrix} \sum_{j=1}^K h_j^2 & \sum_{j=1}^K h_j h_{j+1} & \cdots & \sum_{j=1}^K h_j h_{j+L-m-1} \\ \sum_{j=1}^K h_{j+1} h_j & \sum_{j=1}^K h_{j+1}^2 & \cdots & \sum_{j=1}^K h_{j+1} h_{j+L-m-1} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{j=1}^K h_{j+L-m-1} h_j & \sum_{j=1}^K h_{j+L-m-1} h_{j+1} & \cdots & \sum_{j=1}^K h_{j+L-m-1}^2 \end{pmatrix}.$$

Using this partitioning form, we can say that the sub-matrix $\mathbf{HH}^\top_{(1)}$ is obtained from a Hankel matrix corresponding to the sub-series $H_{N-m} = (h_1, \ldots, h_{N-m})$ and its eigenvalues are $\lambda_1^{L-m,N-m} \geq \lambda_2^{L-m,N-m} \geq \cdots \geq \lambda_{L-m}^{L-m,N-m} \geq 0$. Therefore, the proof is completed using Cauchy's interlacing theorem. ■

Now, we would like to find a relationship between $\lambda_j^{L-m,N}$ and $\lambda_j^{L,N}$. Therefore, Theorem 3.4 should not use directly. Next, we consider four cases and show that we can find general relationships for some classes of the Hankel matrix.

3.2.1 Case 1: $L \geq 1$, rank of $\mathbf{H} = 1$

In this case, it is obvious that we have one positive eigenvalue. Therefore, we can write

$$\lambda_1^{L,N} = \sum_{j=1}^{L} \lambda_j^{L,N} = \mathrm{tr}(\mathbf{HH}^\top) = \sum_{l=1}^{L} \sum_{j=l}^{K+l-1} h_j^2.$$

According to Theorem 3.2, eigenvalue $\lambda_1^{L,N}$ increases with L till $[(N+1)/2]$ and then decreases for $L \geq [(N+1)/2]+1$. Therefore, we have $\lambda_1^{L-m,N} \leq \lambda_1^{L,N}$ and $L \leq [(N+1)/2]$ providing that the conditions of Case 1 are satisfied.

3.2.2 Case 2: $L = 2$, rank of $\mathbf{H} = 2$

In this case, $\mathbf{HH}^\top$ has at most two eigenvalues which are the solution of the quadratic equation

$$\lambda^2 - \left(\sum_{j=1}^{N-1} h_j^2 + \sum_{j=2}^{N} h_j^2\right)\lambda + \sum_{j=1}^{N-1} h_j^2 \sum_{j=2}^{N} h_j^2 - \left(\sum_{j=1}^{N-1} h_j h_{j+1}\right)^2 = 0. \tag{8}$$

Equation (8) has two real solutions so that we have two real eigenvalues. The first eigenvalue (larger one) is given by

$$\lambda_1^{2,N} = \frac{\sum\limits_{j=1}^{N-1} h_j^2 + \sum\limits_{j=2}^{N} h_j^2 + \sqrt{\left(h_1^2 - h_N^2\right)^2 + 4\left(\sum\limits_{j=1}^{N-1} h_j h_{j+1}\right)^2}}{2}. \tag{9}$$

Equation (9) shows that

$$\lambda_1^{2,N} = \begin{cases} \geq \lambda_1^{1,N}, \left(\sum_{j=1}^{N-1} h_j h_{j+1}\right)^2 \geq h_1^2 h_N^2; \\ \leq \lambda_1^{1,N}, \left(\sum_{j=1}^{N-1} h_j h_{j+1}\right)^2 \leq h_1^2 h_N^2; \end{cases} \tag{10}$$

where $\lambda_1^{1,N} = \sum_{j=1}^{N} h_j^2$, when $L = 1$. Practically, it seems that the first condition of Equation (10) is usually satisfied for a wide classes of models. For example, it can be seen that the condition is equivalent to monotonicity of the sequence $\{h_j, j = 1, \ldots, N\}$. For a non-negative (or non-positive) monotone sequence, we have $\sum_{j=1}^{N-1} h_j h_{j+1} \geq h_1 h_N$. Applying Equation (9), it follows $\lambda_1^{2,N} \geq \sum_{j=1}^{N-1} h_j^2 = \lambda_1^{1,N-1}$. A greater class is obtained if we consider positive data, where all observations are bigger that the first one and $h_1 \geq h_N/(N-1)$. Under this condition, it is easy to show that $\sum_{j=1}^{N-1} h_j h_{j+1} \geq h_1 h_N$ and therefore $\lambda_1^{2,N} \geq \lambda_1^{1,N}$. In the next section, we see some examples of models that have not these conditions but $\lambda_1^{2,N} \geq \lambda_1^{1,N}$.

It is worth mention that we can state a geometrical display of Equation (8) as

$$\lambda^2 - \left(||h_{1:N-1}||^2 + ||h_{2:N}||^2\right)\lambda + ||h_{1:N-1}||^2 ||h_{2:N}||^2 (\sin(\theta_{1,2}))^2 = 0, \tag{11}$$

where $h_{1:N-1}$ and $h_{2:N}$ denote the first and second rows of $\mathbf{H}$, $||.||$ the Euclidean norm and $\theta_{1,2}$ the angle between two rows of $\mathbf{H}$. Notice that last expression in Equation (11) is the magnitude of the cross product between two first rows of $\mathbf{H}$. Since $(\sin(\theta_{1,2}))^2 \leq 1$, it is easy to obtain the inequality $\lambda_1^{2,N} \geq \lambda_1^{1,N-1}$, which is a direct result of Theorem 3.4 from characteristics given in Equation (11).

3.2.3 Case 3: $L > 2$, rank of $\mathbf{H} = 2$

In this case, $\mathbf{HH}^\top$ has two positive eigenvalues. To obtain the eigenvalues, first of all note that

$$\det\left(\lambda \mathbf{I} - \mathbf{HH}^\top\right) = \lambda^L + c_1 \lambda^{L-1} + \cdots + c_{L-1}\lambda + c_L, \tag{12}$$

where the coefficients of c_j can be obtained from following lemma.

Lemma 3.5 (Horn and Johnson, 1985, Theorem 1.2.12) Let $\mathbf{A}$ be an $n \times n$ real or complex matrix with eigenvalues $\lambda_1, \ldots, \lambda_n$. Then, for $1 \le k \le n$,

(i) $s_k(\lambda) = (-1)^k c_k$, and
(ii) $s_k(\lambda)$ is the sum of all the $k \times k$ principal minors of $\mathbf{A}$.

Equation (12) shows that the eigenvalues of $\mathbf{HH}^\top$ in this case are the solution of

$$\lambda^2 - \lambda \sum_{l=1}^{L} \sum_{j=l}^{K+l-1} h_j^2 + \sum_{l=1}^{L-1} \sum_{i=1}^{L-l} \left\{ \left(\sum_{j=l}^{K+l-1} h_j^2 \right) \left(\sum_{j=l}^{K+l-1} h_{j+i}^2 \right) - \left(\sum_{j=l}^{k+l-1} h_j h_{j+i} \right)^2 \right\} = 0. \tag{13}$$

The first eigenvalue (larger one) is given by

$$\lambda_1^{L,N} = \frac{1}{2} \left\{ \sum_{l=1}^{L} \sum_{j=l}^{K+l-1} h_j^2 + \sqrt{\Delta_L} \right\}, \tag{14}$$

where Δ_L is the discriminant of the quadratic expression given in Equation (13). According to Equation (14), it is easy to see that, for $L \le [(N+1)/2]$,

$$\lambda_j^{L,N} \ge \lambda_j^{L-1,N} \iff \left| \sqrt{\Delta_{L-1}} - \sqrt{\Delta_L} \right| \le \sum_{l=L}^{N-L+1} h_l^2, \quad j = 1, 2.$$

Similar to the previous case, Equation (13) may be reformulated in the language of multivariate geometry for the L-lagged vectors given by

$$\lambda^2 - \sum_{j=1}^{L} ||h_{j:K+j-1}||^2 \lambda + \sum_{i=1}^{L-1} \sum_{j=i+1}^{L} ||h_{i:K+i-1}||^2 ||h_{j:K+j-1}||^2 (\sin(\theta_{i,j}))^2 = 0,$$

where notations are defined similarly as mentioned in Case 3.

3.2.4 Case 4: $L > 2$, rank of $\mathbf{H} > 2$

Applying Equation (12), it can be obtained the characteristic equation whose solution gives the eigenvalues of $\mathbf{HH}^\top$. However, their functional forms are very sophisticated in this case and, therefore, we consider several series to check the interesting relationship between eigenvalues.

3.3 Contribution of eigenvalues in $T_{\mathbf{H}}^{L,N}$

The ratio $\sum_{j=1}^{r} \lambda_j^{L,N} / \sum_{j=1}^{L} \lambda_j^{L,N}$ is the characteristic of the best r-dimensional approximation of the lagged vectors in the SSA technique. Furthermore, this ratio is an obvious criterion for choosing the proper values of the parameters r and L in the SSA. Therefore, the study for changing this ratio with respect to L and r is important for the SSA technique. First of all, note that, if let $C_j^{L,N} = \lambda_j^{L,N} / \sum_{j=1}^{L} \lambda_j^{L,N}$, then it is easy to see that, for some $j \in \{1, \ldots, L-m\}$ and all values of m belonging to $\{1, \ldots, L-1\}$, we have

$$C_j^{L,N} \le C_j^{L-m,N}. \tag{15}$$

Since inequality given in Equation (15) is satisfied for all values of m belonging to $\{1, \ldots, L-1\}$, it appears that $C_1^{L,N}$ is decreasing on $L \in \{2, \ldots, [(N+1)/2]\}$. In the next section, we see examples that show whether such a behavior is true for polynomial models or not.

4. Examples and Application

In this section, we discuss some examples related to the theoretical results obtained in Section 3. Also, we provide an application of these results.

4.1 Examples

Example 4.1 Let $h_t = \exp(\alpha_0 + \alpha_1 t)$, for $t = 1 \ldots, N$. It is easy to see that the corresponding Hankel matrix $\mathbf{H}$ has rank one. Figure 1 shows first singular value of $\mathbf{H}$ for this model with $\alpha_0 = 0.1, \alpha_1 = 0.2$ and $N = 20$, which is convex with respect to L and attains maximum value at $L = 10, 11$, i.e., the median of $\{1, \ldots, 20\}$.

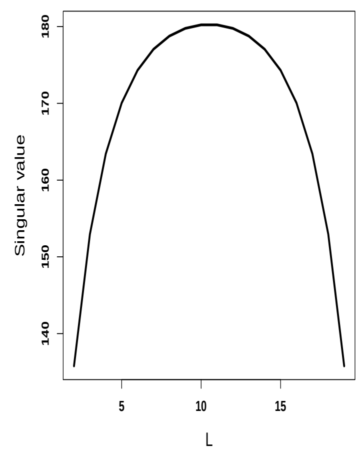

Figure 1. Plot of the first singular value of $\mathbf{H}$ for different values of L: example 4.1.

Now, we consider two different examples where their corresponding Hankel matrices have rank two. The first one is a simple linear model and the second is a cosine model. As we see for both models, roughly speaking, we can say that the results are somewhat similar to Example 4.1.

Example 4.2 Let $h_t = \alpha_0 + \alpha_1 t$, for $t = 1, \ldots, N$. It is easy to show that rank of the corresponding Hankel matrix $\mathbf{H}$ is two. Figure 2 shows the first and second singular values of $\mathbf{H}$ for $\alpha_0 = 1, \alpha_1 = 2$ and $N = 20$. From this figure, we can say that both first and second singular values of $\mathbf{H}$ increase for $L \leq [(N + 1)/2]$ and then decrease.

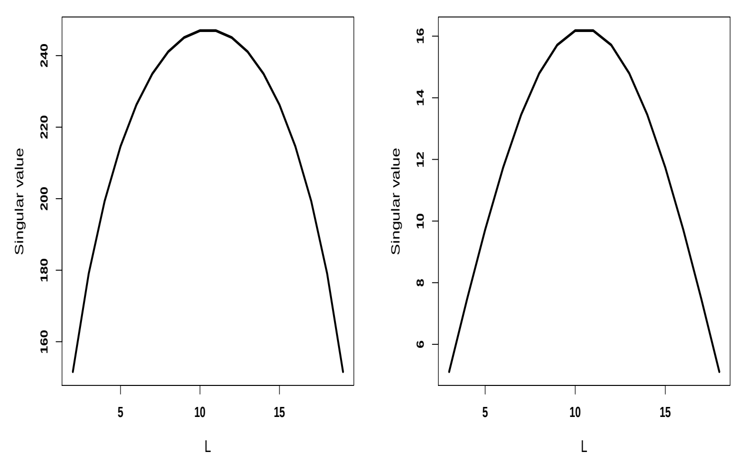

Figure 2. Plots of the first (left) and second (right) singular values of $\mathbf{H}$ for different values of L: example 4.2.

Example 4.3 Let $h_t = \cos(\pi t/12)$, for $t = 1, \ldots, N$. First and second singular values of $\mathbf{H}$ are depicted in Figure 3 for series length 100. If we connive some small fluctuations in the plots, we can say that behavior of singular values of $\mathbf{H}$ is similar to Example 4.2.

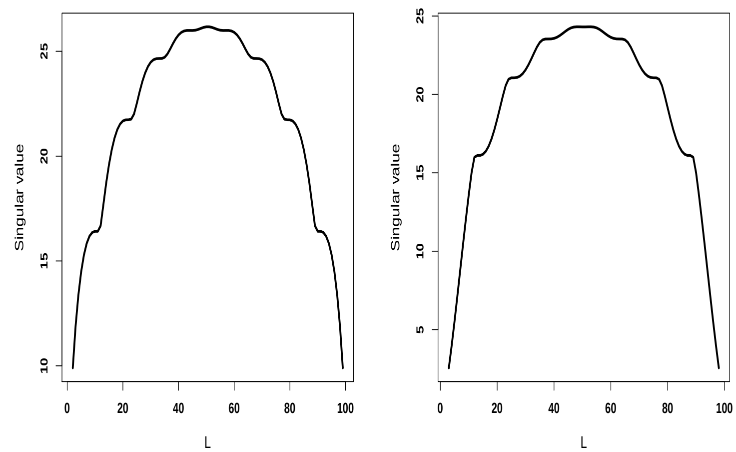

Figure 3. Plots of the first (left) and second (right) singular values of $\mathbf{H}$ for different values of L: example 4.3.

Example 4.4 Let $h_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2$, for $t = 1, \ldots, N$. It is easy to show that rank of the corresponding Hankel matrix $\mathbf{H}$ is 3. Figure 4 shows the singular values of $\mathbf{H}$ for $\alpha_0 = 1, \alpha_1 = 2, \alpha_2 = 3$ and $N = 20$. From this figure, we note that all the singular values of $\mathbf{H}$ increase for $L \leq [(N+1)/2]$ and then decrease, which coincides with Theorem 3.2.

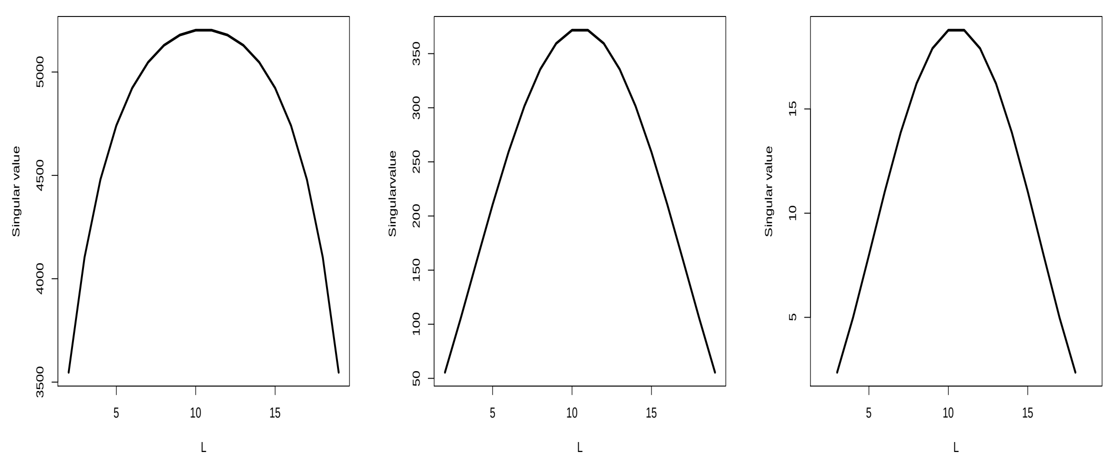

Figure 4. Plots of the three largest singular values of $\mathbf{H}$ for different values of L: example 4.4.

Example 4.5 Let $h_t = \log(t)$, for $t = 1, \ldots, N$. Then, it can be seen that rank of the corresponding Hankel matrix $\mathbf{H}$ is four. Singular values of $\mathbf{H}$ are shown in Figure 5 for $N = 20$. The results of this example are in concordance with Example 4.4.

Figure 6 shows two singular values for models $h_t = \cos(\pi t/12)$ (left) and $h_t = \log(t)$ (right), for $N = 5, \ldots, 100$. Solid and dashed lines in Figure 6 denote the singular values for $L = 2$ and $L = 1$, respectively. Both of these values confirm our expectation for discrepancy between two singular values. Notice that the cosine model is not monotone, but $\lambda_1^{2,N} \geq \lambda_1^{1,N}$.

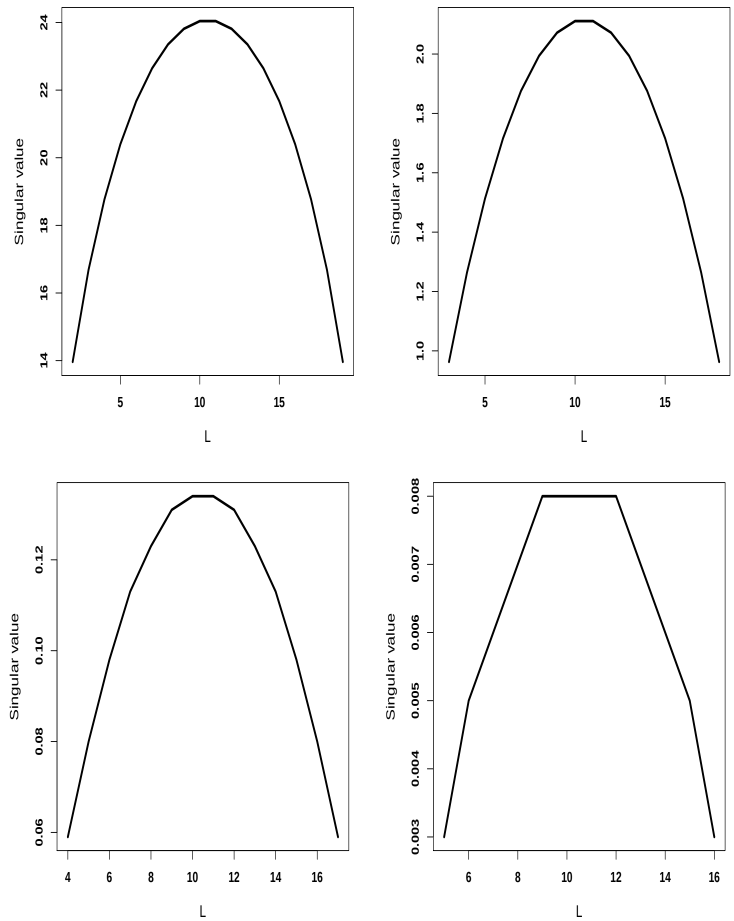

Figure 5. Plots of the four largest singular values of $\mathbf{H}$ with respect to different values of L: example 4.5.

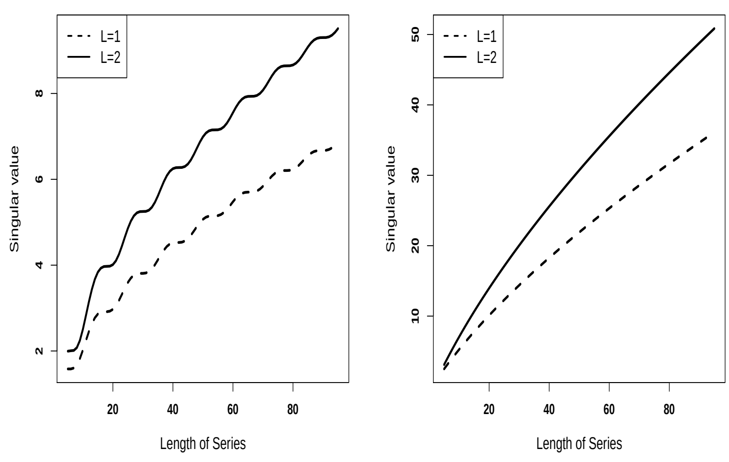

Figure 6. Plots of the first singular value for values of L and N in cosine (left) and logarithm (right) models.

Example 4.6 Let $h_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2$, for $t = 1, \ldots, N$ (a polynomial model). Figure 7 shows the ratio $C_j^{L,N}$ for $\alpha_0 = 1, \alpha_1 = 2, \alpha_2 = 3, N = 20$ and $j = 1, 2, 3$. From this figure, we note that $C_1^{L,N}$ decreases for the values of L less than $[(N+1)/2]$ and then increases on the set $L \in \{[(N+1)/2] + 1, \ldots, N-1\}$. Whereas $C_2^{L,N}$ and $C_3^{L,N}$ increase on the set $\{1, \ldots, [(N+1)/2]\}$ and decrease on $\{[(N+1)/2] + 1, \ldots, N\}$.

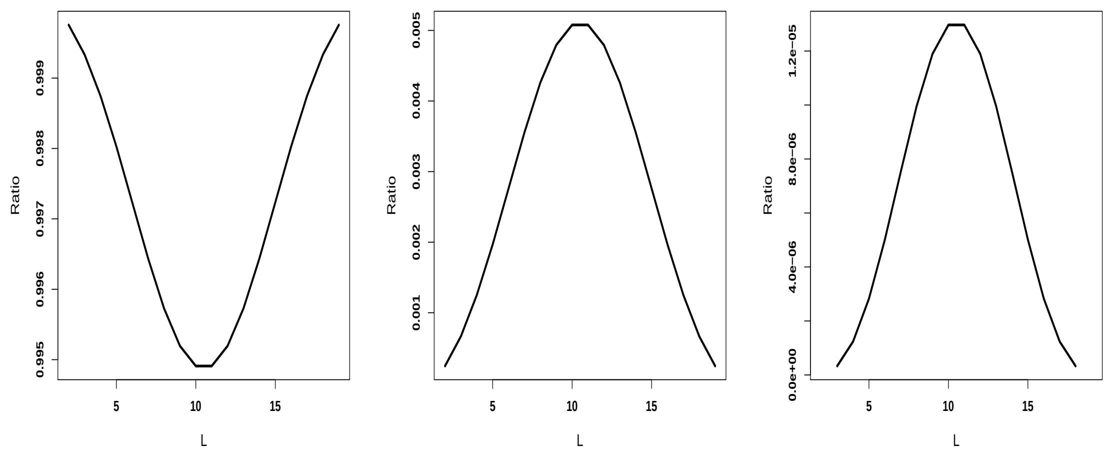

Figure 7. Plots of $C_j^{L,N}$ with respect to L for $N = 20$ and $j = 1$ (left), $j = 2$ (center), $j = 3$ (right): example 4.6.

Next, we examine the cases where the degree of the polynomial is greater than two. Furthermore, different coefficients are considered. The results are similar to Example 4.6 and thus we do not report them here. As a general result, we can say that inequality given in Equation (15) is satisfied for $j = 1$ in the polynomial models. Now, we consider the ratio $C_{1:r}^{L,N} = \sum_{j=1}^{r} C_j^{L,N}$. Since $C_1^{L,N}$ is bigger than $C_j^{L,N}$, for $j > 1$, and the discrepancy between them usually is so much (see the polynomial model of Example 4.6), we expect that the ratio $C_{1:r}^{L,N}$ has a behavior such as $C_1^{L,N}$. In the following example, the behavior of this ratio is depicted for the polynomial model with degree four.

Example 4.7 Let $h_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \alpha_3 t^3 + \alpha_4 t^4$, for $t = 1, \ldots, N$. Figure 8 shows the ratio $C_{1:r}^{L,N}$ for $\alpha_0 = 1, \alpha_1 = 2, \alpha_2 = 3, \alpha_3 = 4, \alpha_4 = 5$ and $N = 20$. From this figure, we note that $C_{1:r}^{L,N}$ decreases on $L \in \{2, \ldots, [(N+1)/2]\}$, for $r \geq 1$, and then increases on $L > [(N+1)/2]$, as expected.

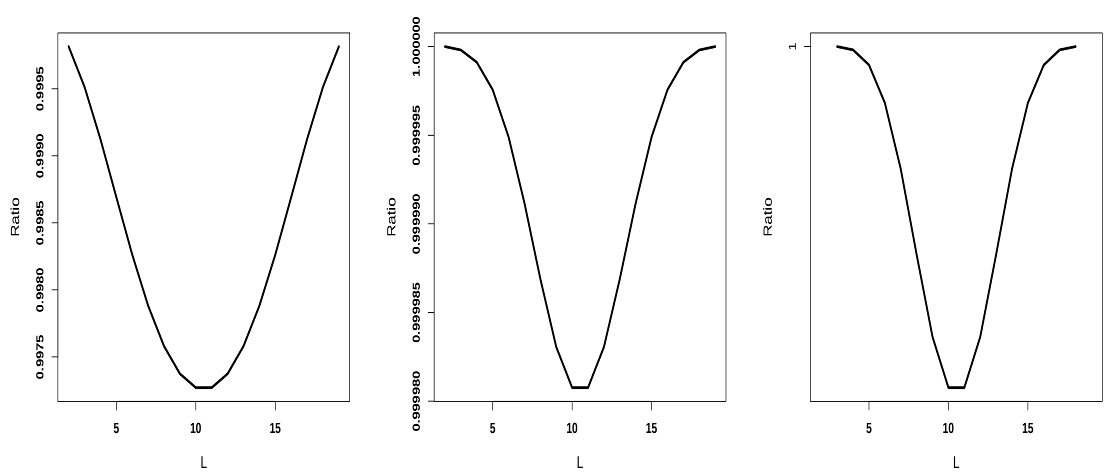

Figure 8. Plot of $C_{1:r}^{L,N}$ with respect to L for $N = 20$ and $r = 1$ (left), $r = 2$ (center), $r = 3$ (right): example 4.7.

4.2 Choosing the SSA parameters

Several rules have been proposed in the literature for choosing the SSA parameters; see, e.g., Golyandina et al. (2001) and Hassani et al. (2011). However, the list is by no means exhaustive. Certainly, the choice of parameters depends on the data collected and on the analysis we have performed. Anyway one important note is that singular values give most effective information for choosing parameters in the SSA. In previous subsections, several criteria and theorems were considered to investigate the behavior of singular values of the Hankel matrix. Considering theoretical results about the structure of the Hankel matrix, trajectory matrix and relationship with their dimensions, enable us to state that the choice of L close to one-half of the time series length is a suitable choice for decomposition stage in most cases. The previous empirical and theoretical results also confirm the results obtained by us here. However, by using definition of the criteria $T_{\mathbf{H}}^{L,N}$, it can be seen that

$$T_{\mathbf{H}}^{L,N} - T_{\mathbf{H}}^{L-1,N} = \sum_{j=L}^{K} h_j^2. \tag{16}$$

Equation (16) is the rate of change in $\mathrm{tr}(\mathbf{H}\mathbf{H}^\top)$ for each unit change in the window length. This rate is large for small values of the window length and decreases to attain minimum value at $L = K$, where it is equivalent to $L = L_{\max} = \text{median}\{1, \ldots, N\}$. This motivate us to choose smaller values than $L_{\max}$ when the rate given in Equation (16) is small. To support this motivation, Golyandina et al. (2001) said that series with a complex structure and too large window length L can produce an undesirable decomposition of the series components of interest, which may lead, in particular, to their mixing with other series components. Sometimes, in these circumstances, even a small variation in the value of L can reduce mixing and lead to a better separation of the components, i.e., it provides a transition from weak to strong separability.

Another important parameter to be chosen is the number of needed singular values r for grouping in the reconstruction step. Election of this parameter is similar to the procedure for obtaining the cutoff value in principal component analysis. It is known that there is not a general way to choose an optimal value of the cutoff number and it depends on the data; for a complete description and review of this topic, see Jollife (2002). Perhaps the most obvious criterion for choosing the cutoff value is to select a (cumulative) percentage of the total variation, which one desires that the selected singular values contribute, say a 80% or 90%. The required number of singular values is then the smallest value of r for which this chosen percentage is exceeded. This criterion is equivalent to the ratio $C_{1:r}^{L,N}$ previously defined.

5. Conclusions

We have considered one of the main and most important issues in the singular spectrum analysis, that is, the selection of parameters. As stated, singular values of the trajectory matrix in the singular spectrum analysis play an important role. Specifically, election of the parameters values of the window length (L) and the number of needed singular values for reconstruction of series (r) depend on the behavior of the singular values of the trajectory matrix. In this paper, we have studied the behavior of the singular values of a Hankel matrix ($\mathbf{H}$) with respect to its dimension. We have shown that, for a wide classes of time series, the singular value of $\mathbf{H}\mathbf{H}^\top$ ($\lambda_j^{L,N}$) increases with L in $L \in \{1, \ldots, [(N+1)/2]\}$ and decreases in $L \in \{[(N+1)/2]+1, \ldots, N\}$. In addition, we have investigated the behavior of the sum of square and the contribution of each singular value. The results based on these criteria have shown that the choice of L close to one-half of the time series length is a suitable choice for decomposition stage in most cases for the singular spectrum analysis.

Acknowledgements

Authors would like to thank anonymous referees for their valuable comments that improved the exposition of the paper.

References

Bhatia, R., 1997. Matrix Analysis. Springer, New York.

Elsner, J.B., Tsonis, A.A., 1996. Singular Spectrum Analysis: A New Tool in Time Series Analysis. Plenum Press, New York.

Golyandina, N., Nekrutkin, V., Zhigljavsky, A., 2001. Analysis of Time Series Structure: SSA and Related Techniques. Chapman & Hall/CRC, New York.

Hassani, H., 2007. Singular spectrum analysis: methodology and comparision. Journal of Data Science, 5, 239–257.

Hassani, H., Heravi, S., Zhigljavsky, A., 2009. Forecasting European industrial production with singular spectrum analysis. International Journal of Forecasting, 25, 103–118.

Hassani, H., Mahmoudvand, R., Zokaei, M., 2011. Separability and window length in singular spectrum analysis. Comptes Rendus Mathematique (in press).

Hassani, H., Thomakos, D., 2010. A Review on singular spectrum analysis for economic and financial time series. Statistics and its Interface, 3, 377–397.

Horn, R.A., Johnson, C.R., 1985. Matrix Analysis. Cambridge University Press, Cambridge.

Jolliffe, I.T., 2002. Principal Component Analysis. Springer, New York.

Peller, V., 2003. Hankel Operators and Their Applications. Springer, New York.

Widom, H., 1966. Hankel matrix. Transactions of the American Mathematical Society, 121, 1–35.