

## APPROXIMATE PROJECTORS IN SINGULAR SPECTRUM ANALYSIS\*

V. MOSKVINA<sup>†</sup> AND K. M. SCHMIDT<sup>†</sup>

**Abstract.** Singular spectrum analysis (SSA) is a method of time-series analysis based on the singular value decomposition of an associated Hankel matrix. We present an approach to SSA using an effective and numerically stable high-degree polynomial approximation of a spectral projector, which also provides a means of time-series forecasting. Several numerical examples illustrating the algorithm are given.

**Key words.** Hankel matrix, singular value decomposition, spectral projector, polynomial approximation

**AMS subject classifications.** 15A60, 41A10, 65F15, 62-07

**PII.** S0895479801398967

**1. Introduction.** Singular spectrum analysis (SSA) is a well-established method of time-series analysis (see [2], [3], [15], [16], and the recent monographs [4] and [5]). The main idea of SSA is to select a number of significant principal components from the singular value decomposition (SVD) of the so-called trajectory matrix of a given time series, and hence to reconstruct a time series showing characteristic traits, e.g., the trend, periodicities, or signal (as opposed to random noise) of the original series.

SVD is a fundamental and very well studied process of numerical linear algebra with a long history (cf. [12], [1]). However, it is computationally expensive and thus problematic in real-time signal processing; therefore truncated forms of the SVD which only provide partial information (see [14], [17]) and alternative methods (see [7], [11], [13], and, specifically for Hankel-type matrices, [9]) have been proposed. Recently, combinations of SSA/SVD with a wavelet transform have attracted some attention [6], [18], [19].

The procedure of classical SSA is as follows. Let  $x_1, x_2, \dots, x_N \in \mathbb{R}$ ,  $N \in \mathbb{N}$ , be (part of) a time series, let  $M \leq N/2$  be a positive integer, and let  $K = N - M + 1$ . Set

$$(1.1) \quad \mathbf{X} = (x_{ij})_{i,j=1}^{M,K} = \begin{pmatrix} x_1 & x_2 & x_3 & \dots & x_K \\ x_2 & x_3 & x_4 & \dots & x_{K+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_M & x_{M+1} & x_{M+2} & \dots & x_N \end{pmatrix}.$$

$\mathbf{X}$  is called the *trajectory matrix*. Obviously  $x_{ij} = x_{i+j-1}$ , so that the matrix  $\mathbf{X}$  has identical entries on the diagonals  $i + j = \text{const}$ , i.e., it is a Hankel matrix. One can consider  $\mathbf{X}$  as multivariate data with  $M$  characteristics and  $K = N - M + 1$  observations  $X_1, X_2, \dots, X_K$ , where

$$X_j = \begin{pmatrix} x_j \\ \vdots \\ x_{j+M-1} \end{pmatrix} \in \mathbb{R}^M \quad (j \in \{1, \dots, K\}).$$

---

\*Received by the editors November 29, 2001; accepted for publication (in revised form) by A. H. Sayed August 23, 2002; published electronically February 12, 2003.

<http://www.siam.org/journals/simax/24-4/39896.html>

<sup>†</sup>School of Mathematics, Cardiff University, Senghennydd Rd., Cardiff, UK, CF24 4YH (MoskvinaV1@cardiff.ac.uk, SchmidtKM@cardiff.ac.uk).

The SVD of  $\mathbf{X}$  is based on the spectral decomposition of the *lag-covariance matrix*  $\mathbf{R} = \mathbf{X}\mathbf{X}^T \in \mathbb{R}^{M \times M}$ . Note that  $\mathbf{R}$  is symmetric and positive semidefinite. Therefore, it has a complete set of eigenvectors and can be diagonalized in the form

$$(1.2) \quad \mathbf{R} = U\Lambda U^T,$$

where  $\Lambda$  is the diagonal  $M \times M$  matrix of eigenvalues  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_M \geq 0$ , and

$$U = (U_1, U_2, \dots, U_M) = \begin{pmatrix} u_{11} & u_{21} & \dots & u_{M1} \\ u_{12} & u_{22} & \dots & u_{M2} \\ \vdots & \vdots & \ddots & \vdots \\ u_{1M} & u_{2M} & \dots & u_{MM} \end{pmatrix}$$

is an orthogonal matrix of eigenvectors of the matrix  $\mathbf{R}$ . Denoting  $d = \max\{i \in \{1, \dots, M\} \mid \lambda_i > 0\}$  and  $V_i = X^T U_i / \sqrt{\lambda_i}$  ( $i = 1, \dots, d$ ), we can write the SVD of the trajectory matrix  $\mathbf{X}$ ,

$$(1.3) \quad \mathbf{X} = \sum_{i=1}^d \sqrt{\lambda_i} U_i V_i^T = \mathbf{X}_1 + \dots + \mathbf{X}_d,$$

where  $\mathbf{X}_i = \sqrt{\lambda_i} U_i V_i^T$  are rank-one biorthogonal matrices; we have  $\text{rank } \mathbf{X} = d$ .

Now a subset of the *SVD components*  $\mathbf{X}_1, \dots, \mathbf{X}_d$  is selected by choosing a set of indices  $I \subset \{1, \dots, d\}$ , resulting in the decomposition

$$\mathbf{X} = \mathbf{X}_I + \mathbf{X}_{\bar{I}}, \quad \text{where } \mathbf{X}_I = \sum_{i \in I} \mathbf{X}_i \quad \text{and} \quad \mathbf{X}_{\bar{I}} = \sum_{i \notin I} \mathbf{X}_i.$$

If  $I$  is suitably chosen,  $\mathbf{X}_I$  will represent some characteristic feature of the original time series which can be exhibited by removing  $\mathbf{X}_{\bar{I}}$ . Unfortunately, however,  $\mathbf{X}_I$  itself is not in general the trajectory matrix of some time series, as it does not necessarily have Hankel structure. This obstacle is overcome by *diagonal averaging* over the diagonals  $i + j = \text{const}$ , which allows us to extract a time series  $\tilde{x}_k$  ( $k \in \{1, \dots, N\}$ ) from any  $M \times K$  matrix  $Y$  by the formula

$$(1.4) \quad \tilde{x}_k = \begin{cases} \frac{1}{k} \sum_{i=1}^k y_{i, k-i+1} & \text{for } 1 \leq k \leq M-1, \\ \frac{1}{M} \sum_{i=1}^M y_{i, k-i+1} & \text{for } M \leq k \leq K, \\ \frac{1}{N-k+1} \sum_{i=k-K+1}^M y_{i, k-i+1} & \text{for } K+1 \leq k \leq N. \end{cases}$$

Applying this to  $\mathbf{X}_I$  to construct a time series  $(z_t)$ , we obtain the SSA decomposition of the original series

$$(1.5) \quad x_t = z_t + \varepsilon_t, \quad t \in \{1, \dots, N\}.$$

(It is not difficult to verify that the residual series  $(\varepsilon_t)$  results from diagonal averaging of  $\mathbf{X}_{\bar{I}}$ .)

An interesting practical application of SSA is the extraction of a signal from a time series perturbed by noise. Since one expects, in light of such asymptotic results as Corollary 6.1 in [5], that the signal will correspond to larger eigenvalues of the lag-covariance matrix, while eigenvalues associated with noise components should be small, this means that the first SVD components will be selected, cutting off at a certain eigenvalue size  $\lambda_{\text{cut}} > 0$ . Thus the index set will have the structure  $I = \{1, \dots, l\}$  with some  $l \in \{1, \dots, d\}$  such that  $\lambda_l \geq \lambda_{\text{cut}} > \lambda_{l+1}$ . Ideally, the series  $(z_t)$  in (1.5) can then be associated with signal and the residual series  $(\varepsilon_t)$  with noise. For an extensive discussion of the problem of choosing the values for the two SSA parameters, viz. the lag  $M$  and the number  $l$  of SVD components included in the reconstruction, see [5]. A common choice for  $M$  is the maximal value  $M = \lfloor N/2 \rfloor$ . The value of  $l$  (or, equivalently, of the cut-off point  $\lambda_{\text{cut}}$ ) must depend on the properties of the given time series. If  $l$  is too small (underfitting), then we miss part of the signal; alternatively, if  $l$  is too large (overfitting), then we approximate a part of noise together with the signal.

In the present paper we develop a method of computing  $\mathbf{X}_I$  (and hence the reconstructed series  $(x_t)$ ) in this situation without actually performing the spectral decomposition of the lag-covariance matrix, i.e., without calculating its eigenvalues and eigenvectors. For large time series and correspondingly large matrices, this will offer a faster alternative and open the way for noise-reduction applications of the SSA method.

This paper is organized as follows. We first observe that the selection of the part  $\mathbf{X}_I$  from the SVD of the trajectory matrix can be replaced by applying a spectral projector of the lag-covariance matrix; in the case at hand, this will be the orthogonal projector onto the eigenspace for eigenvalues in the interval  $[\lambda_{\text{cut}}, \infty)$ . In section 3, we then proceed to find a polynomial approximation of the characteristic function of this interval, which permits a direct approximate calculation of the spectral projector. We use an iterative method which avoids the problems inherent in a naive evaluation of the approximating polynomial, which would be inefficient and highly unstable. Section 4 presents a geometric forecasting algorithm based on the approximate spectral projector. The examples studied in section 5 demonstrate the workings and the practical applicability of our method. It turns out that even a relatively rough and inexpensive approximation, corresponding to an SSA with a “fuzzy cut-off,” can yield a very high degree of noise suppression and an excellent reconstruction of the signal.

**2. Polynomial approximation of the spectral projector.** For the construction of the matrix  $\mathbf{X}_I$  after choosing the index set  $I$ , it is sufficient to find the orthogonal projector  $P$  onto the subspace of  $\mathbb{R}^M$  spanned by the eigenvectors  $U_j$  with  $j \in I$ ,

$$(2.1) \quad P = \sum_{j \in I} U_j U_j^T.$$

Indeed, one then has, using (1.3) and the orthonormality of the eigenvectors,

$$P\mathbf{X} = \sum_{j \in I} \sum_{i=1}^d \sqrt{\lambda_i} U_j U_j^T U_i V_i^T = \sum_{j \in I} \sqrt{\lambda_j} U_j V_j^T = \mathbf{X}_I.$$

Thus,  $\mathbf{X}_I$  is obtained by a simple matrix multiplication once  $P$  is known. The matrix  $P$ , on the other hand, can be represented as a function of the matrix  $\mathbf{R}$ . Generally,

given a function  $f : \mathbb{R} \rightarrow \mathbb{R}$ , one can define the matrix

$$(2.2) \quad f(\mathbf{R}) = \sum_{j=1}^M f(\lambda_j) U_j U_j^T.$$

(This holds for general symmetric matrices and, in analogous form, for self-adjoint operators in Hilbert space, provided  $f$  is measurable with respect to the spectral measure; see [10, Theorem VII.2].)

If  $S \subset \mathbb{R}$  contains the eigenvalues with indices in  $I$  but no other, then clearly  $P = \chi_S(\mathbf{R})$ , where

$$\chi_S(\lambda) = \begin{cases} 1, & \lambda \in S, \\ 0, & \lambda \notin S \end{cases}$$

is the characteristic function of the set  $S$ .

Of course, rewriting the definition of  $P$  in this way seems of little benefit. Note, however, that for a polynomial function  $p$ ,  $p(\mathbf{R})$  can be evaluated directly, interpreting the powers  $\mathbf{R}^n$  in the sense of matrix multiplication. Using the spectral representation

$$\mathbf{R} = \sum_{j=1}^M \lambda_j U_j U_j^T,$$

it is not hard to see that this calculation, for which no knowledge of the eigenvalues and eigenvectors of  $\mathbf{R}$  is required, gives the same result as formula (2.2).

Even if  $f$  is not a polynomial, we can use this to find approximations for  $f(\mathbf{R})$  based on the following observation. Let  $(p_n)_{n \in \mathbb{N}}$  be a sequence of polynomials such that

$$\lim_{n \rightarrow \infty} \sup_{j \in \{1, \dots, M\}} |f(\lambda_j) - p_n(\lambda_j)| = 0;$$

then  $\lim_{n \rightarrow \infty} p_n(\mathbf{R}) = f(\mathbf{R})$  in the Euclidean operator norm. Indeed, denoting by  $\|\cdot\|$  the Euclidean norm on  $\mathbb{R}^M$ , we have for all  $v \in \mathbb{R}^M$

$$\begin{aligned} \|(p_n(\mathbf{R}) - f(\mathbf{R}))v\|^2 &= \left\| \sum_{j=1}^M (p_n(\lambda_j) - f(\lambda_j)) U_j U_j^T v \right\|^2 \\ &= \sum_{j=1}^M |p_n(\lambda_j) - f(\lambda_j)|^2 |U_j^T v|^2 \\ &\leq \sup_{j \in \{1, \dots, M\}} |p_n(\lambda_j) - f(\lambda_j)|^2 \|v\|^2 \rightarrow 0 \quad (n \rightarrow \infty). \end{aligned}$$

In the situation at hand, we wish to omit eigenvalues below  $\lambda_{\text{cut}}$  and include all others, so here  $f = \chi_{[\lambda_{\text{cut}}, \infty)}$ . In order to obtain an approximation of the corresponding spectral projector  $P$ , we replace  $f$  by an approximating polynomial  $p$ . Of course, we can expect to find only a good polynomial approximation on a compact interval, as the polynomial will grow rapidly near  $\pm\infty$ . However, we need only to approximate  $f$  at the eigenvalues of  $\mathbf{R}$  or, as these are unknown, on an interval which contains them. We already know that the eigenvalues are nonnegative; furthermore, if  $\|\cdot\|$  denotes any matrix norm satisfying

$$\|\mathbf{R}v\| \leq \|\mathbf{R}\| \|v\| \quad (v \in \mathbb{R}^M)$$

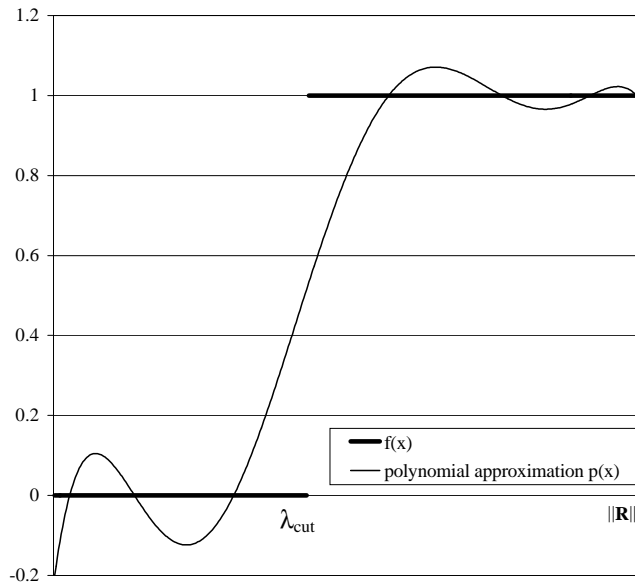


FIG. 2.1. Function  $f(x)$  and a polynomial approximation  $p(x)$  of degree 6.

for some norm  $\|\cdot\|$  on  $\mathbb{R}^M$ , it follows that  $\lambda_1 \leq \|\mathbf{R}\|$ . Hence it is sufficient to approximate  $f$  on the interval  $[0, \|\mathbf{R}\|]$  (see Figure 2.1).

Then we substitute the matrix  $\tilde{P} = p(\mathbf{R})$  for the actual projector  $P$  to construct  $\tilde{\mathbf{X}}_I = \tilde{P}X$ , and hence reconstruct the time series without noise. In general,  $\tilde{P}$  will not be a projector, but if  $p$  is a good approximation to  $f$  on  $[0, \|\mathbf{R}\|]$ , it will be close to the projector  $P$ . Moreover,  $\tilde{P}$  can always be interpreted as a weighted sum of the spectral projectors  $U_j U_j^T$  onto the eigenspaces for the individual  $\lambda_j$  by means of (2.2),

$$\tilde{P} = \sum_{j=1}^M p(\lambda_j) U_j U_j^T.$$

(In our approximation, we shall actually have weights  $p(\lambda_j) \in [0, 1]$ —see below.) Thus, even if  $p$  is only a coarse approximation of  $f$ , the resulting reconstructed time series is meaningful as a mixture of SVD components with a “fuzzy cut-off,” which may even have advantages over the usual SSA reconstruction that uses a sharp selection of SVD components.

The number of components included in the projector  $P$  is  $l = \text{tr} P$ ; similarly we take  $\text{tr} \tilde{P}$  as an indicator of how much of the original series is included in the reconstruction. The size of  $\text{tr} \tilde{P}$  and its closeness to an integer can be used to assess whether the cut-off point  $\lambda_{\text{cut}}$  was suitably chosen. As a further guide to the choice of the cut-off point, one can obtain a rough estimate of the spectrum of  $\mathbf{R}$  by performing either standard SSA or the approximate procedure outlined in the present paper with a small lag  $M$ ; if this smaller window covers a stable structure (periodics) in the original time series, the corresponding eigenvalues will not change very much with increasing  $M$  and can thus give an idea of the spectral structure of  $\mathbf{R}$ . However, this preliminary step will miss out on large-scale structures, which will become apparent only when the approximate SSA with the large lag  $M$  is calculated.

It is common to express the eigenvalues of the lag-covariance matrix as a frac-

tion of their sum, called *eigenvalue share* (see [5] for details). Fortunately, the sum of the eigenvalues of  $P$  is known without calculating them, since by virtue of the diagonalization (1.2) we have

$$\sum_{j=1}^M \lambda_j = \text{tr } \mathbf{R},$$

and the trace can easily be obtained as the sum of diagonal elements of  $\mathbf{R}$ .

**3. The iterative approximation procedure.** The approximation of functions of a real variable by polynomials is notoriously problematic. An accurate approximation requires polynomials of high degree, which in turn leads to strong oscillations and quick growth outside the interval of approximation, and to costly and numerically unstable computations. Therefore, one often prefers the use of localized substitutes for polynomial approximation, e.g., splines. In some situations, however, such as the present case, polynomials are the only type of function whose values can be calculated.

The following observation provides a numerically stable iterative method of calculating highly accurate polynomial approximations of a characteristic function (see Figure 3.1).

**PROPOSITION 3.1.** *Let  $p_1(x) = 3x^2 - 2x^3$  and let  $p_n$  be the  $n$ th iteration of  $p_1$ , i.e.,  $p_n(x) = p_1(p_1(\dots p_1(x)))$  ( $p_1$  applied  $n$  times). Then for  $n \in \mathbb{N}$ ,  $p_n$  is strictly increasing on  $[0, 1]$  and fixes the points  $0, \frac{1}{2}$ , and  $1$ . Furthermore,*

$$\lim_{n \rightarrow \infty} p_n(x) = \begin{cases} 0, & x \in \left(-\frac{\sqrt{3}-1}{2}, \frac{1}{2}\right), \\ \frac{1}{2}, & x = \frac{1}{2}, \\ 1, & x \in \left(\frac{1}{2}, \frac{1+\sqrt{3}}{2}\right); \end{cases}$$

*the convergence is uniform on  $[-\frac{\sqrt{3}-1}{2} + \varepsilon, \frac{1}{2} - \varepsilon] \cup [\frac{1}{2} + \varepsilon, \frac{1+\sqrt{3}}{2} - \varepsilon]$  for any  $\varepsilon > 0$ .*

The base polynomial  $p_1$  is characterized as the lowest-degree polynomial fixing the points  $0, \frac{1}{2}$ , and  $1$ , and with zero derivative at  $0$  and  $1$ . The steepness of the flank for the  $n$ th iteration is

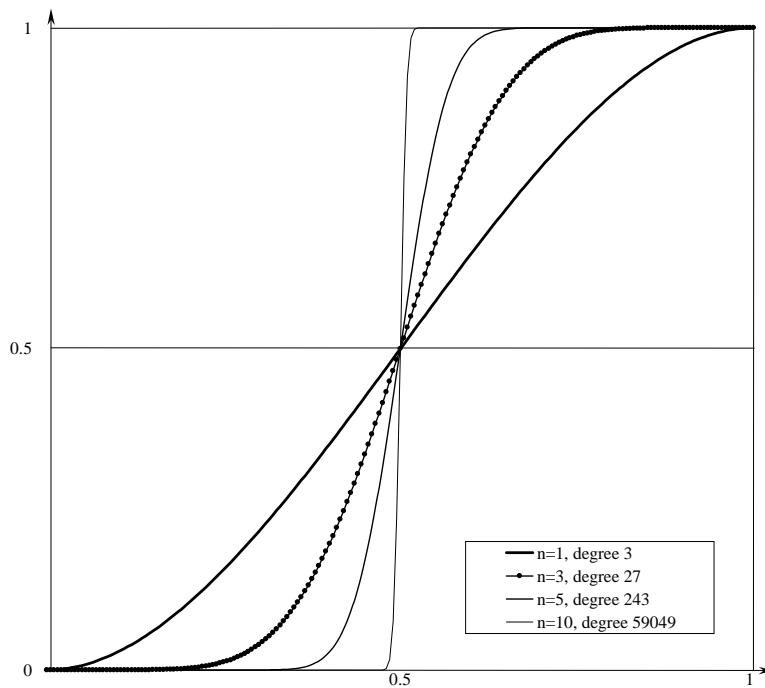
$$p'_n\left(\frac{1}{2}\right) = \left(\frac{3}{2}\right)^n.$$

The above proposition gives an approximation of the characteristic function  $\chi_{[1/2, \infty)}$  on the interval  $[0, 1]$ . By a suitable scaling transformation, we can always assume that we are dealing with a matrix with eigenvalues in  $[0, 1]$  and a cut-off point at  $\frac{1}{2}$ . Indeed, we can replace the matrix  $\mathbf{R}$  by the matrix

$$B = \begin{cases} \frac{1}{2\lambda_{\text{cut}}}\mathbf{R} & \text{if } \lambda_{\text{cut}} \geq \|\mathbf{R}\|/2, \\ \frac{1}{2(\|\mathbf{R}\| - \lambda_{\text{cut}})}(\mathbf{R} + (\|\mathbf{R}\| - 2\lambda_{\text{cut}})\mathbf{I}) & \text{otherwise;} \end{cases}$$

then  $\chi_{[\lambda_{\text{cut}}, \infty)}(\mathbf{R}) = \chi_{[1/2, \infty)}(B)$ . (Here  $\mathbf{I}$  is the identity  $M \times M$  matrix.)

*Remark.* One may be tempted to consider calculating and storing the coefficients of the polynomial  $p_n$  instead of applying the above iterative procedure to the individual matrix  $B$  (or number  $x$  when evaluating  $p_n(x)$ ). However, this alternative approach has severe disadvantages which make it wholly impractical. Indeed,  $p_n$  is

FIG. 3.1. The approximating polynomials  $p_n$ .

a polynomial of degree  $3^n$ , and hence the calculation of  $p_n(B)$  by the Horner scheme requires  $3^n - 1$  matrix multiplications, as compared to the mere  $2n$  matrix multiplications in the iterative method. Furthermore, the evaluation of the polynomial is highly unstable in view of its very large coefficients; thus even for numbers  $x \in [0, 1]$ ,  $p_4(x)$  cannot be correctly calculated in double precision (in single precision, the problem already appears in  $p_3(x)$ ).

In contrast, the iterative method, in addition to its numerical stability, provides the possibility of monitoring the progress of the approximation, stopping when the desired accuracy is reached, rather than fixing the degree of the approximating polynomial in advance.

**4. Geometric forecasting.** The (approximate) projector  $P$  can also be used to forecast the given time series.

A simple geometric method of forecasting a time series  $x_1, x_2, \dots, x_N$  using components  $\mathbf{X}_i$ ,  $i \in I$ , of the SVD of its trajectory matrix is based on the principle of choosing the next term  $x_{N+1}$  in the series in such a way that the vector  $\mathbf{x} = (x_{N+2-M}, \dots, x_N, x_{N+1})^T$  is closest to the subspace of  $\mathbb{R}^M$  spanned by the eigenvectors  $U_j$ ,  $j \in I$ . Taking the Euclidean norm in  $\mathbb{R}^M$  as a measure for the closeness, we can express this in terms of the orthogonal projector  $P$  (see (2.1)) as the problem of minimizing the norm of the difference vector between  $\mathbf{x}$  and its orthogonal projection  $P\mathbf{x}$ ,

$$\|(\mathbf{I} - P)\mathbf{x}\|^2 \rightarrow \min,$$

varying  $x_{N+1}$ .

The minimum satisfies

$$0 = \frac{d}{dx_{N+1}} \|(\mathbf{I} - P)\mathbf{x}\|^2 = 2\mathbf{x}^T(\mathbf{I} - P)(\mathbf{I} - P)_M,$$

where  $(\mathbf{I} - P)_M$  denotes the last column of the matrix  $(\mathbf{I} - P)$ . In other words, the minimizing vector  $\mathbf{x}$  is orthogonal to the *forecast vector*  $f = (\mathbf{I} - P)(\mathbf{I} - P)_M$ . If  $f_M \neq 0$ , we thus find the recurrence for  $x_{N+1}$ ,

$$x_{N+1} = -\frac{1}{f_M} \sum_{j=1}^{M-1} f_j x_{N+1-M+j}.$$

This recurrence formula is meaningful and can be used to forecast the time series even when the orthogonal projector  $P$  is replaced by the approximate projector  $\tilde{P}$  calculated according to the method outlined in section 3. Note that the resulting approximate forecast vector converges to the exact one as  $\tilde{P}$  approaches  $P$  in the Euclidean operator norm.

**5. Numerical examples.** Let us study two examples to illustrate the approximate SSA algorithm and geometric forecasting described above.

First consider the simulated time series

$$z_t = \sin(0.1t) + e_t,$$

where the  $e_t$  are independent identically distributed random variables  $e_t \sim N(0, 16)$  for  $t = 1, \dots, 2000$  (white noise).

Figure 5.1 shows the result of the approximate SSA with maximal lag  $M = 1000$ , cut-off point  $\lambda_{\text{cut}} = 1\%$ , and an approximating polynomial of degree 243 (i.e., 5 iterations). The trace of the approximate projector  $\tilde{P}$  is 21.56, indicating that SSA components with a total weight corresponding to about 22 eigenvectors are included in the reconstruction.

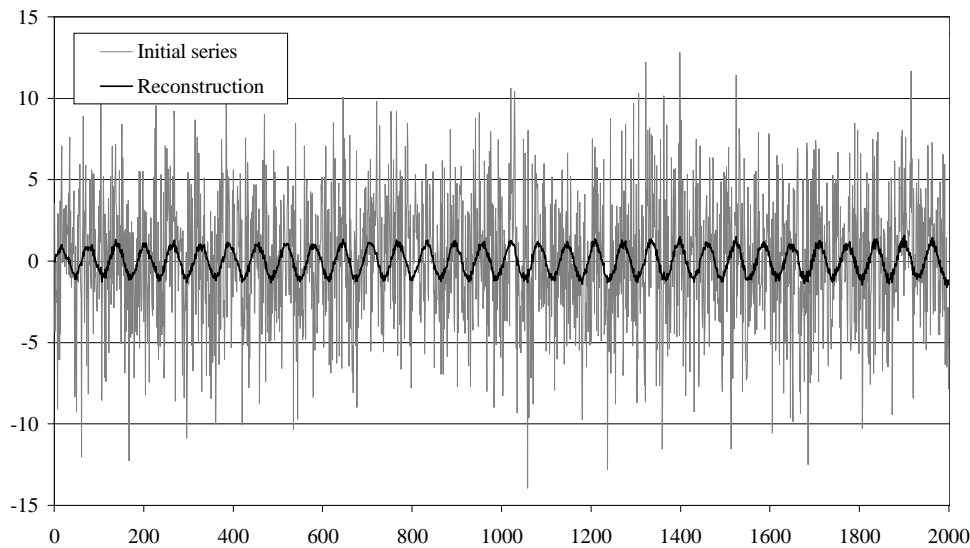


FIG. 5.1.  $\sin(0.1t) + e_t$ ,  $e_t \sim N(0, 16)$  and its reconstruction.



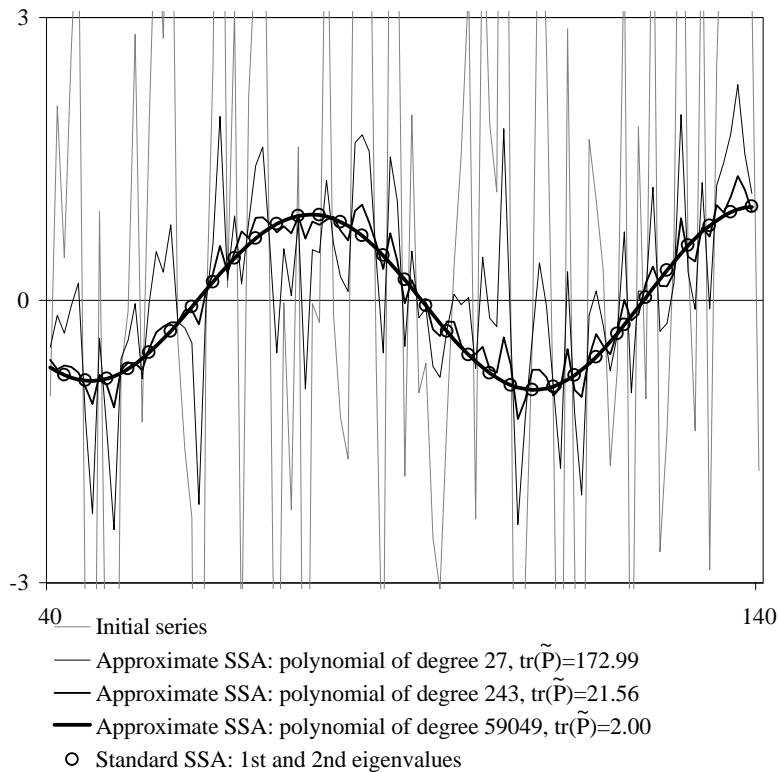


FIG. 5.2. Comparison of reconstructions using different polynomials.

In Figure 5.2 a zoomed part of the time series is presented. One can clearly observe how the degree of the approximating polynomial, and hence the sharpness of the cut-off in the calculation of the approximate projector, influences the quality of the reconstruction. In the case of the polynomial of degree 59049 (i.e., 10 iterations) the reconstruction by our approximate method reproduces the result of standard SSA; then the number of principal components included in the latter coincides (within some tolerance) with  $\text{tr}(\tilde{P})$ .

Note, however, that already the reconstruction based on the coarser approximation of Figure 5.1, with  $\tilde{P}$  still far from the actual projector, clearly picks out the signal from the very noisy time series.

Our second example is based on the well-known real-life data (see [8]) of monthly averages of hotel rooms occupied from 1963–1976; this example was studied in detail in [5]. We demonstrate by this example how different choices of the cut-off point  $\lambda_{\text{cut}}$  exhibit various features of the time series in the resulting reconstructions (see Figure 5.3).

To pick up the first eigenvalue (corresponding to the trend) we have chosen lag  $M = 84$  and  $\lambda_{\text{cut}} = 2\%$ , and 15 iterations were enough to separate it from the rest of the spectrum ( $\text{tr} \tilde{P} = 1.0000$ ). With  $\lambda_{\text{cut}} = 0.51\%$  and 19 iterations, we also include the second and third eigenvalues ( $\text{tr} \tilde{P} = 3.0000$ ), corresponding to the one year cycle. For a more precise reconstruction of the initial data we have taken the cut-off point  $\lambda_{\text{cut}} = 0.11\%$  and 23 iterations ( $\text{tr} \tilde{P} = 5.0000$ ). By this choice we include exactly the first five components, which describe the main structure of the series: linear

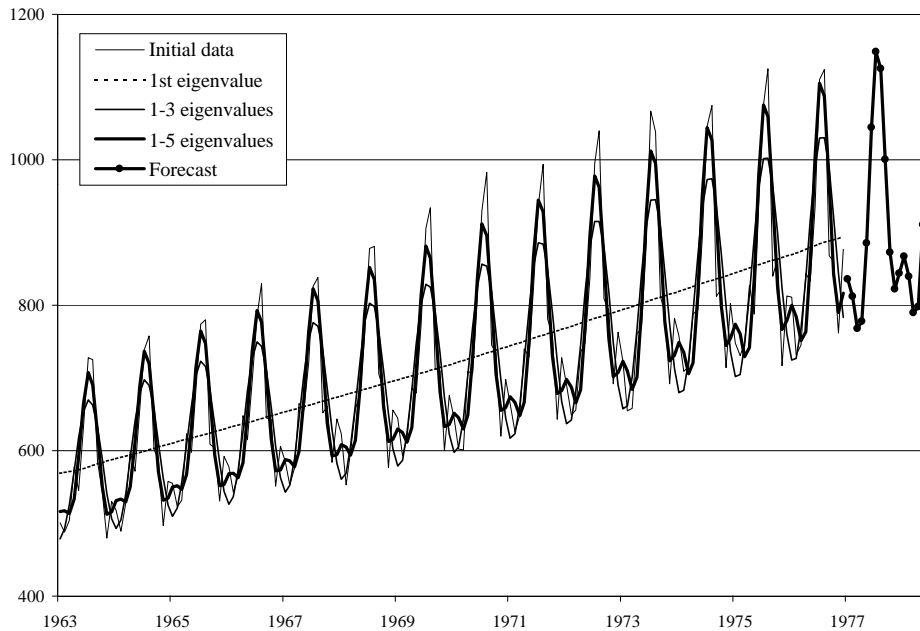


FIG. 5.3. Occupied hotel rooms (average per month), 1963–1976.

trend, one-year and half-year periodicities. The geometric forecast based on these components does not contradict the structure of the series.

**Acknowledgments.** The authors thank V. Nekrutkin and A. Zhigljavsky for their interest and advice.

#### REFERENCES

- [1] E. BIGLIERI AND K. YAO, *Some properties of singular value decomposition and their applications to digital signal processing*, Signal Process., 18 (1989), pp. 277–289.
- [2] D.S. BROOMHEAD, R. JONES, AND G.P. KING, *Topological dimension and local coordinates from time series data*, J. Phys. A, 20 (1987), pp. L563–L569.
- [3] D.S. BROOMHEAD AND G.P. KING, *Extracting qualitative dynamics from experimental data*, Phys. D, 20 (1986), pp. 217–236.
- [4] J. ELSNER AND A. TSONIS, *Singular Spectrum Analysis: A New Tool in Time Series Analysis*, Plenum Press, New York, 1996.
- [5] N. GOLJANDINA, V. NEKRUTKIN, AND A. ZHIGLJAVSKY, *Analysis of Time Series Structure: SSA and Related Techniques*, Chapman and Hall, London, 2001.
- [6] R. KAKARALA AND P.O. OGUNBONA, *Signal analysis using a multiresolution form of the singular value decomposition*, IEEE Trans. Image Process., 10 (2001), pp. 723–735.
- [7] T. MORITA AND T. KANADE, *A sequential factorization method for recovering shape and motion from image streams*, IEEE Trans. Pattern Anal. and Machine Intell., 19 (1997), pp. 858–867.
- [8] T.M. O'DONOVAN, *Short Term Forecasting: An Introduction to the Box-Jenkins Approach*, Wiley, New York, 1983.
- [9] H. PARK, L. ZHANG, AND J.B. ROSEN, *Low rank approximation of a Hankel matrix by structured total least norm*, BIT, 39 (1999), pp. 757–779.
- [10] M. REED AND B. SIMON, *Methods of Modern Mathematical Physics I. Functional Analysis*, Academic Press, New York, 1980.
- [11] G.X. RITTER AND P. SUSSNER, *The minimax eigenvalue transform*, in Image Algebra and Morphological Image Processing, III (San Diego, CA, 1992), Proc. SPIE 1769, SPIE, Bellingham, WA, 1992, pp. 276–282.

- [12] G.W. STEWART, *On the early history of the singular value decomposition*, SIAM Rev., 35 (1993), pp. 551–566.
- [13] A.-J. VAN DER VEEN, *A Schur method for low-rank approximation*, SIAM J. Matrix Anal. Appl., 17 (1996), pp. 139–160.
- [14] S. VAN HUFFEL, *Partial singular value decomposition algorithm*, J. Comput. Appl. Math., 33 (1990), pp. 105–112.
- [15] R. VAUTARD AND M. GHIL, *Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series*, Phys. D, 35 (1989), pp. 395–424.
- [16] R. VAUTARD, P. YIOU, AND M. GHIL, *Singular-spectrum analysis: A toolkit for short, noisy chaotic signals*, Phys. D, 58 (1992), pp. 95–126.
- [17] C.R. VOGEL AND J.G. WADE, *Iterative SVD-based methods for ill-posed problems*, SIAM J. Sci. Comput., 15 (1994), pp. 736–754.
- [18] F. VOGT AND M. TACKE, *Fast principal component analysis of large data sets*, Chemometrics and Intell. Lab. Systems, 59 (2001), pp. 1–18.
- [19] P. YIOU, D. SORNETTE, AND M. GHIL, *Data-adaptive wavelets and multi-scale singular-spectrum analysis*, Phys. D, 142 (2000), pp. 254–290.