Application of the singular spectrum analysis for change-point detection in time series

Anatoly Zhigljavsky

School of Mathematics, Cardiff University
CF24 4AG, Cardiff, UK
ZhigljavskyAA@cardiff.ac.uk

Abstract. Singular-spectrum analysis (SSA) is a powerful technique of time series analysis. SSA is based on a singular value decomposition of a 'trajectory matrix' obtained from the original time series with subsequent reconstruction of the series. A methodology of change-point detection in time series based on sequential application of the singular-spectrum analysis is proposed and studied. The underlying idea is that if at a certain time τ the mechanism generating the time series x_t has changed, then an increase in the distance between the l-dimensional hyperplane spanned by the eigenvectors of the so-called lag-covariance matrix, and the M-lagged vectors $(x_{\tau+1}, \ldots, x_{\tau+M})$ is to be expected. Under certain conditions, the proposed algorithm can be considered as a proper statistical procedure with the moving sum of weighted squares of random variables being the detection statistic. The correlation structure of the moving sums is studied. Several asymptotic expressions for the significance level of the algorithm are compared.

Keywords: boundary crossing probability, change-point detection, moving sum, singular-spectrum analysis, singular value decomposition.

1 Introduction

Let us briefly describe the main idea of the method. Let $x_1, x_2, \ldots$ be a time series, M and N be two integers ($M \le N/2$), and set $K = N - M + 1$. Define the vectors $X_j = (x_j, \ldots, x_{j+M-1})^T$ ($j = 1, 2, \ldots$) and the matrix

$$\mathbf{X} = (x_{i+j-1})_{i,j=1}^{M,K} = (X_1, \ldots, X_K),$$

which is called the trajectory matrix.

We consider $\mathbf{X}$ as multivariate data with M characteristics and K observations. The columns X_j of $\mathbf{X}$, considered as vectors, lie in the M-dimensional space $\mathbb{R}^M$. The singular value decomposition (SVD) of the so-called lag-covariance matrix $\mathbf{R} = \mathbf{X}\mathbf{X}^T$ (and of the trajectory matrix $\mathbf{X}$ itself) provides us with a collection of M eigenvalues and eigenvectors. A particular combination of a certain number $l < M$ of these eigenvectors determines an l-dimensional hyperplane in $\mathbb{R}^M$. According to the SSA algorithm, the M-dimensional data is projected onto this l-dimensional subspace and the subsequent averaging over the diagonals gives us an approximation to the original series.

One of the features of the SSA algorithm is that the distance between the vectors X_j $(j = 1, \ldots, K)$ and the l-dimensional hyperplane is controlled by the choice of l and can be reduced to a rather small value. If the time series $\{x_t\}_{t=1}^N$ is continued for $t > N$ and there is no change in the mechanism which generates the values x_t, then this distance should stay reasonably small for X_j, $j \ge K$ (for testing, we take Q such vectors). However, if at a certain time $N + \tau$ the mechanism generating x_t $(t \ge N + \tau)$ has changed, then an increase in the distance between the l-dimensional hyperplane and the vectors X_j for $j \ge K + \tau$ is to be expected.

SSA expansion tends to pick up the main structure of the time series, if there is one. (This happens when the l-dimensional subspace approximates well the M-dimensional vectors $X_1, \ldots, X_K$.) If this structure is being found and there are no structural changes, then the SSA continuation of the time series should agree with the continued series. (That is, the Q vectors X_j for $j \ge K$ should stay close to the l-dimensional subspace.) A change in structure of the time series should force the corresponding vectors X_j out of the subspace. This is the central idea of the method we propose.

SSA performs the analysis of the time series structure in a nonsequential (off-line) manner. However, change-point detection is typically a sequential (on-line) problem, and we aim to develop an algorithm

that can be used in the on-line regime. This can be achieved by sequentially applying the SVD to the lag-covariance matrices computed in a sequence of time intervals, either $[n+1, n+N]$ or $[1, n+N]$. Here $n = 0, 1, \ldots$ is the iteration number and N is the length of the time interval where the trajectory matrix is computed. The latter version produces a CUSUM-type algorithm. We, however, prefer the former version, with the sequence of time intervals $[n+1, n+N]$: this version is better accommodated to the presence of slow changes in the time series structure, to outliers and to the case of multiple changes. (The price for that is a smaller size of the sample used to construct the trajectory matrices, and therefore some loss in efficiency in the ideal situation.)

SSA and the proposed change-point detection algorithm are model–free tools and generally are not intended for precise statistical inferences; they are essentially model-building procedures. However, under certain conditions, the proposed algorithm can be considered as a proper statistical procedure. Studying properties of this procedure is the main purpose of the talk, see also Chapter 3 in Golyandina et al (2001), Moskvina and Zhigljavsky (2003).

2 Algorithm

Let $x_1, x_2, \ldots, x_{\mathbb{T}}$ be a time series with $\mathbb{T} \le \infty$. Let us choose two integers: the window width N ($N \le \mathbb{T}$), and the lag parameter M ($M \le N/2$). Also, set $K = N - M + 1$.

For each suitable $n \ge 0$ we consider the time interval $[n+1, n+N]$ and construct the trajectory matrix (which will be called *base matrix*)

$$\mathbf{X}^{(n)} = (x_{n+i+j-1})_{i,j=1}^{M,K} = \begin{pmatrix} x_{n+1} & x_{n+2} & \ldots & x_{n+K} \\ x_{n+2} & x_{n+3} & \ldots & x_{n+K+1} \\ \vdots & \vdots & \vdots & \ddots \\ x_{n+M} & x_{n+M+1} & \ldots & x_{n+N} \end{pmatrix} . \tag{1}$$

The columns of $\mathbf{X}^{(n)}$ are the vectors $X_j^{(n)}$ ($j = 1, \ldots, K$), where

$$X_j^{(n)} = (x_{n+j}, \ldots, x_{n+M+j-1})^T , \quad j \ge -n+1 .$$

For each $n = 0, 1, \ldots$ we define the lag-covariance matrix $\mathbf{R}_n = \mathbf{X}^{(n)}\big(\mathbf{X}^{(n)}\big)^T$. The SVD of $\mathbf{R}_n$ gives us a collection of M eigenvectors, and a particular group I of $l < M$ of them determines an l-dimensional subspace $\mathcal{L}_{n,I}$ of the M-dimensional space $\mathbb{R}^M$ of vectors $X_j^{(n)}$.

We denote the l eigenvectors that form the basis of the subspace $\mathcal{L}_{n,I}$ by U_{i_1}, ...,U_{i_l} and the sum of squares of the (Euclidean) distances between the vectors $X_j^{(n)} (j = p+1, \ldots, q)$ and this l-dimensional subspace by $\mathcal{D}_{n,I,p,q}$ (the choice of p and $q = p + Q$ is discussed in the talk). The matrix with columns $X_j^{(n)}$ ($j = p+1, \ldots, q$) is called *test matrix*; the location of the base and test matrices is depicted in Figure 1.

Since the eigenvectors of $\mathbf{R}_n$ are orthonormal, the squared Euclidean distance between any vector $Z \in \mathbb{R}^M$ and the subspace $\mathcal{L}_{n,I}$ spanned by the l eigenvectors $U_{i_1}, \ldots, U_{i_l}$, is just

$$||Z||^2 - ||U^T Z||^2 = Z^T Z - Z^T U U^T Z ,$$

where $||\cdot||$ is the usual Euclidean norm and U is the $(M \times l)$-matrix with columns $U_{i_1}, \ldots, U_{i_l}$. It is also the difference between the squared norms of the vector Z and the projection of Z to the space $\mathcal{L}_{n,I}$. The squared distance $\mathcal{D}_{n,I,p,q}$ is the sum of these differences for the vectors $X_j^{(n)}$ constituting the test matrix. That is,

$$\mathcal{D}_{n,I,p,q} = \sum_{j=p+1}^{q} \left((X_j^{(n)})^T X_j^{(n)} - (X_j^{(n)})^T U U^T X_j^{(n)} \right) . \tag{2}$$

If a change in the mechanism generating x_t occurs at a certain point τ, then we expect that the vectors $X_j = X_{j-n}^{(n)}$ with $j > \tau$ lie further away from the l-dimensional subspace $\mathcal{L}_{n,I}$ than the vectors X_j with

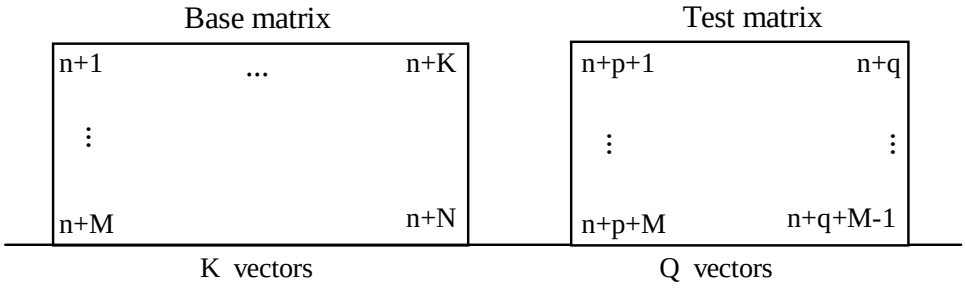

Fig. 1. Construction of the base and test matrices.

$j \le \tau$. This means that we expect that as n changes, the sequence $\mathcal{D}_{n,I,p,q}$ starts growing somewhere around $\hat{n}$ such that $\hat{n}+q+M-1=\tau$. (This value $\hat{n} = \tau-q-M+1$ is the first value of n such that the test sample $x_{n+p+1}, \dots, x_{n+q+M-1}$ contains a point with a change.) This growth continues for some time; the expected time of the growth depends on the duration of change and the relations between p, q and N. In a particular case when $p = N$ and $Q = q - p \le M$ and for an abrupt single change, the sequence $\mathcal{D}_{n,I,p,q}$ stops growing after Q iterations, around the point $n = \tau - p - M$. Then during the following $M - Q$ iterations one would expect reasonably high values of this sequence, which must be followed by its decrease to, perhaps, a new level. (This relates to the fact that the SSA decomposition should incorporate the new signal at the intervals $[n + 1, n + N]$ with $n \ge \tau - M$.)

The detection statistics are:

- $\mathcal{D}_{n,I,p,q}$, the sum of squared Euclidean distances between the vectors $X_j^{(n)}$ $(j=p+1, \dots, q)$ and the $l-$dimensional subspace $\mathcal{L}_{n,I}$ of $\mathbb{R}^M$;
- the normalized sum of squared distances (the normalization is made with respect to the number of elements in the test matrix);
$$\tilde{\mathcal{D}}_{n,I,p,q} = \frac{1}{M(q-p)} \mathcal{D}_{n,I,p,q}\,;$$
- $S_n = \tilde{\mathcal{D}}_{n,I,p,q}/\upsilon_n$.

Here υ_j is an estimate of the normalized sum of squared distances $\tilde{\mathcal{D}}_{j,I,p,q}$ at the time intervals $[j+1, j+N]$ where the hypothesis of no change can be accepted. We suggest to use $\upsilon_n = \tilde{\mathcal{D}}_{\bar{n},I,0,K}$, where $\bar{n}$ is the largest value of $j < n$ such that the null hypothesis of no change in the interval $[j + 1, j + N]$ has been accepted. S_n is the squared distance normalized to the number of elements in the test and base matrices and to the variance of the residuals (which are associated with noise); this statistic is shown in graphs. A natural decision rule in the algorithm is to announce a change if for some n we have $S_n \ge H$, where H is a fixed threshold.

3 The detection statistic as a moving quadratic form

Under the null hypothesis that there is no change in the signal and the signal is fully recovered by the SSA, in the change-point detection algorithm we have at iteration n

$$\mathcal{D}_{n,I,p,q} = \sum_t w_{M,n+p,n+q}(t) e_t^2\,, \tag{3}$$

where

$$w_{M,p,q}(t) = \begin{cases} t-p & \text{for } p < t \le p+Q, \\ Q & \text{for } p+Q < t \le p+M, \\ p+M+Q-t & \text{for } p+M < t < p+M+Q, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

The form of the weight function $w_{M,p,q}(t)$ is related to the structure of the trajectory matrix (1), where x_{n+1} appears once, x_{n+2} – twice, and so on.

Obviously, (3) is a quadratic form $e^T Be$, where $e = (e_1, e_2, \ldots, e_N)^T$ and B is a diagonal matrix with diagonal elements $B_{tt} = w_{M,n+p,n+q}(t)$. The first two moments of this quadratic form can easily be calculated:

$$E\mathcal{D}_{n,I,p,q} = \sigma^2 MQ, \;\; \mathrm{var}(\mathcal{D}_{n,I,p,q}) = \frac{1}{3}Q(\mu_4 - \sigma^4)(3MQ - Q^2 + 1)\,, \tag{5}$$

where $\sigma^2 = Ee_i^2$ and $\mu_4 = Ee_i^4$, the second and the forth moments of the error distribution. In the case when the errors e_i are normal $N(0, \sigma^2)$ we have $\mu_4 = 3\sigma^4$. In this case the distribution of the quadratic form $\mathcal{D}_{n,I,p,q} = e^T Be$ can be thought of as a modification of the χ^2-distribution for the weight function (4); this distribution is studied in Moskvina (2000); it can also be considered as a particular case of the distribution (3.3.1.3) in Richter (1992).

Using the Central Limit Theorem we obtain asymptotically, as $M \to \infty$,

$$\xi_n = \frac{\mathcal{D}_{n,I,p,q} - E\mathcal{D}_{n,I,p,q}}{\sqrt{\mathrm{var}(\mathcal{D}_{n,I,p,q})}} \sim N(0,1)\,. \tag{6}$$

We could have ignored the dependence structure of the sequence of squared distances $\mathcal{D}_{n,I,p,q}$ and use either the asymptotic normality (6) alone or the limiting extreme value distribution to choose the threshold H. We adopt another approach which is based on approximating the sequence $\mathcal{D}_{n,I,p,q}$ by a continuous time random process. Numerical results (demonstrated in the talk) show that this approach leads to much more precise approximations for the boundary crossing probabilities than the Durbin's tangent approximation, Cramer and Leadbetter approximations, see Durbin (1985), Cramer (1965) and Theorem 8.2.7 in Leadbetter, Lindgren and Rootzen (1983).

References

Cramer H., (1965) A Limit Theorem for fhe Maximum Values of Certain Stochastic Processes. Theory of Probability and Applications **10**, 137-139.

Durbin, J. (1985) The First–Passage Density of a Continuous Gaussian Process to a General Boundary. *Applied Probability* **22**, 99-122.

Golyandina, N., Nekrutkin, V. and Zhigljavsky, A. (2001) *Analysis of Time Series Structure: SSA and Related Techniques*, London: Chapman And Hall.

Leadbetter, M.R., Lindgren, G., And Rootzen, H. (1983) *Extremes and Related Properties of Random Sequences and Processes*, Springer Series in Statistics, Springer-Verlag.

Moskvina, V. (2000) Distribution of Random Quadratic Forms Arising in Singular-Spectrum Analysis, *Mathematical Communications*, **5**, 161-171.

Moskvina, V. and Zhigljavsky A.A. (2003) An Algorithm Based on Singular-Spectrum Analysis for Change-Point Detection, *Communication in Statistics. Statistics and Simulations*, **32**, 319-352.

Richter, M. (1992), *Approximation of Gaussian Random Elements and Statistics*, B.G. Teubner Verlagsgesellscaft, Leipzig.