A Novel and Effective Approach to Shape Analysis: Nonparametric Representation, De-noising and Change-Point Detection, Based on Singular-Spectrum Analysis

Vasile Georgescu

Department of Mathematical Economics, University of Craiova, A.I.Cuza str. 13, 200585 Craiova, Romania
vasile.georgescu@feaa.ucv.ro

Abstract. This paper proposes new very effective methods for building nonparametric, multi-resolution models of 2D closed contours, based on Singular Spectrum Analysis (SSA). Representation, de-noising and change-point detection to automate the landmark selection are simultaneously addressed in three different settings. The basic one is to apply SSA to a shape signature encoded by sampling a real-valued time series from a radius-vector contour function. However, this is only suited for star-shaped contours. A second setting is to generalize SSA so as to apply to a complex-valued trajectory matrix in order to directly represent the contour as a time series path in the complex plan, along with detecting change-points in a complex-valued time series. A third setting is to consider the pairs (x, y) of coordinates as a co-movement of two real-valued time series and to apply SSA to a trajectory matrix defined in such a way to span both of them.

Keywords: Statistical shape analysis, Transforming planar closed contours into time series, Singular-spectrum analysis, Real- and complex-valued trajectory matrices, SSA-based change-point detection.

1 The Classical Computational Geometry Approach to Sampling Time Series from Planar Closed Contours

Statistical Shape Analysis involves methods for the geometrical study of random objects where location, rotation and scale information can be removed. By contrast, time series analysis is a widely spread technique that takes into consideration the temporal nature of data. However, despite their differences in nature, statistical shape analysis may benefit from methods commonly used in time series analysis. Indeed, there are certain ways of transforming a closed planar contour into a shape signature, represented by a contour function, and subsequently it may be possible to sample a time series from the contour function. Such functions are defined with respect to either simple or complex geometrical considerations: from metrics induced by symmetry relationships or periodicity, to the formal study of shapes based on computational differential geometry, where the quantification of differences between shapes can be achieved via a Riemannian metric on a shape manifold (namely, a

V. Torra et al. (Eds.): MDAI 2011, LNAI 6820, pp. 162–173, 2011.

finite-dimensional Riemannian manifold), and the interest naturally focuses on computing geodesic distances and geodesic paths between shapes.

Assuming that the contour has some desirable properties such as convexity or star-shapedness (i.e., given a figure A, for each point $(x, y) \in A$, the line segment connecting (x, y) with the centroid is contained in A), relatively simple contour functions, such as the radius-vector or support functions, can be introduced. Otherwise more complex contour functions should be considered, by representing the curves in a parameterized form. Their geometric properties and various quantities associated with them, such as the arc-length and the curvature, can then be expressed via derivatives and integrals using vector calculus.

The radius-vector function $r(\theta)$ is the distance from the reference point O (usually the center of gravity) to the contour in the direction of the θ-ray where $0 \le \theta \le 2\pi$. An example of a star-shaped figure and its radius-vector function is given in Figure 1. If the shape is inferred from noisy data, as it is the case with the figure below, the availability of a de-noising method becomes important. Furthermore, a change-point detection algorithm to automate the selection of salient landmarks may be of great interest.

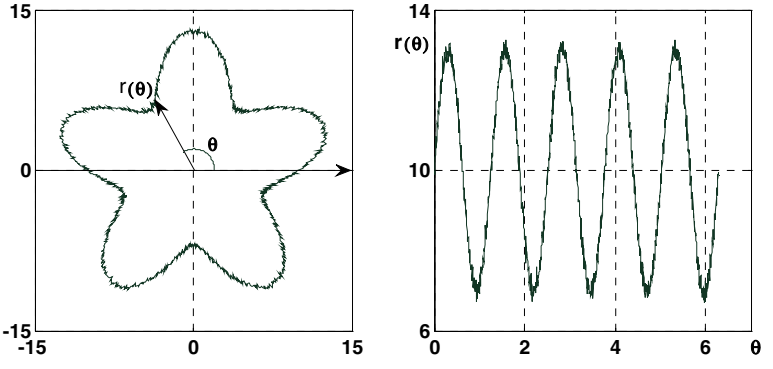

Fig. 1. A noisy star-shaped closed contour (left) and the radius-vector function $r(\theta)$ (right). The centroid has been used as the origin to generate the radius-vector function.

One can now choose a discrete sequence of (equally-spaced) values in $[0, 2\pi]$, i.e., $0 = \theta_1 < \theta_2 < \ldots < \theta_N = 2\pi$. The ordered sequence of radius-vector function values $\{r_t\}_{t=1,\ldots,N}$, with $r_t = r(\theta_t)$, can be regarded as a "time series" sampled from the contour function. The radius-vector function $r(\theta)$ is called a continuous shape signature, whereas $\{r_t\}_{t=1,\ldots,N}$ is called a discrete shape signature.

In the general case, however, description of a shape signature by the radius-vector function is not suitable for non-star-shaped contours (Figure 2).

Alternatively, there are at least two ways of representing planar curves in a parameterized form: one is using the angle (direction) function and another is using

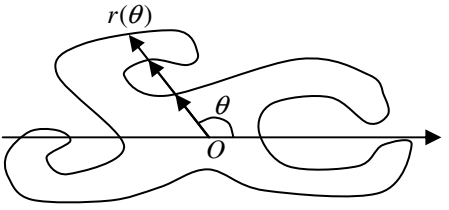

Fig. 2. Problems with the radius-vector function occur if the contour is not star-shaped

the curvature function. These involve using differential geometry, which provides a set of powerful tools for shape analysis. However, from a practical viewpoint, it is difficult to infer planar contour representations in a parameterized form.

The next section introduces a new approach to representation and de-noising of closed planar contours, along with a change-point detection algorithm to automate the selection of salient landmarks. It consists of a nonparametric, multi-resolution method, based on Singular-Spectrum Analysis.

2 A Novel and Effective Approach to Shape Analysis: Nonparametric Representation, De-noising and Change-Point Detection, Based on Singular-Spectrum Analysis

2.1 An overview of Singular-Spectrum Analysis

Singular-Spectrum Analysis (SSA) is a nonparametric method for time series structure recognition and identification. It tries to overcome the problems of finite sample length and noisiness of sampled time series not by fitting an assumed model to the available series, but by using a data-adaptive basis set.

The SSA algorithm has two basic stages: decomposition and reconstruction.

The decomposition stage is carried out in two steps:

(D1) The *Embedding* step maps the original one-dimensional time series $\{x_1, x_2, \ldots, x_N\}$ to a sequence of $K = N - M + 1$ lagged vectors of dimension M (where M is called *window length*):

$$X_i = (x_i, \ldots, x_{i+M-1})', \quad i = 1, \ldots, K, \qquad 1 < M < N. \tag{1}$$

This lagged vectors form the columns of the *trajectory matrix* X, which is actually a Hankel matrix (i.e., it has equal elements on the diagonals $i + j - 1$ =const.): $X = [X_1 : X_2 : \ldots : X_K]$.

(D2) The SVD step is the *singular value decomposition* of the trajectory matrix. Let $R = \frac{1}{N} X \cdot X'$ be an $M \times M$ matrix, called *lag-covariance matrix*. Denote by $\lambda_1, \ldots, \lambda_M$ the eigenvalues of R taken in the decreasing order of magnitude

$(\lambda_1 \geq \ldots \geq \lambda_M \geq 0)$ and by $U_1, \ldots, U_M$ the orthonormal system of the eigenvectors of the matrix R corresponding to these eigenvalues. Let $d = \max\{ i$, such that $\lambda_i > 0 \}$. If we denote $V_i = X'U_i / \sqrt{\lambda_i}$ ($i = 1, \ldots, d$), then the SVD of the trajectory matrix X can be written as $X = X_1 + \ldots + X_d$, where $X_i = \sqrt{\lambda_i} U_i \otimes V_i'$ and $\otimes$ is the outer product. The matrices X_i are elementary matrices (have rank one).

The reconstruction stage is also carried out in two steps:

(R1) The *grouping* step consists of partitioning the set of indices $\{1, \ldots, d\}$ into m disjoint subsets $I_1, \ldots, I_m$. The case of practical interest for our application is that of a dichotomic partitioning: split the set of indices into two groups, $\{1, \ldots, d\} = I + \bar{I}$, where $I = \{i_1, \ldots, i_\ell\}$ and $\bar{I} = \{1, \ldots, d\} \setminus I$, and sum the matrices X_i within each group:

$$X = X_I + X_{\bar{I}} \tag{2}$$

where $X_I = \sum_{i \in I} X_i$ and $X_{\bar{I}} = \sum_{i \in \bar{I}} X_i$.

The choice of the ℓ most contributing eigenvalues λ_i, $i \in I$, and thus of the corresponding ℓ eigenvectors is an appropriate way to control and reduce the distance between the *M*-dimensional vectors that form the columns of trajectory matrix and the ℓ-dimensional hyperplane determined by the ℓ eigenvectors. For example, we can choose the index set I such that $\sum_{j=1}^{\ell} \lambda_{i_j} \Big/ \sum_{j=1}^{d} \lambda_j > 0.95$ corresponding to the set of eigenvalues whose cumulated contribution exceeds 95%.

(R2) The last step transforms each matrix of the grouped decomposition (2) into a new series of length N, by *diagonal averaging*. It consists of averaging over the diagonals $i + j - 1 =$const. ($i = 1, \ldots, M$, $j = 1, \ldots, K$) of the matrices X_I and $X_{\bar{I}}$. Applying then twice the one-to-one correspondence between the series of length N and the Henkel matrices of size $M \times K$ (with $K = N - M + 1$), we obtain the SSA decomposition of the original series $\{x_t\}$ into a sum of two series: $x_t = z_t + \varepsilon_t$, $t = 1, \ldots, N$. In this context, the series z_t (obtained from the diagonal averaging of X_I) can often be associated with signal and the residual series ε_t with noise.

2.2 The First Setting: Applying SSA to a Shape Signature Encoded by Sampling a Real-Valued Time Series from a Radius-Vector Contour Function

This setting is well suited for star-shaped planar closed contours and starts with sampling a "time series" $\{r_t\}_{t=1, \ldots, N}$ from the radius-vector contour function, i.e., $r_t = r(\theta_t)$, $0 = \theta_1 < \theta_2 < \ldots < \theta_N = 2\pi$. The trajectory matrix X is then constructed

by mapping the time series $\{r_t\}_{t=1,\ldots,N}$ to a sequence of $K = N - M + 1$ lagged vectors of dimension M :

$$X = \begin{pmatrix} r_1 & r_2 & \cdots & r_K \\ r_2 & r_3 & \cdots & r_{K+1} \\ \vdots & \vdots & \vdots & \vdots \\ r_M & r_{M+1} & \cdots & r_N \end{pmatrix} \tag{3}$$

Noise reduction is attained by reducing the rank of the trajectory matrix. In the absence of noise one should be able to recover the data trajectory matrix with the first L singular vectors. Thus, the SVD reconstruction of the trajectory matrix X can be truncated to obtain an estimate of the noise-reduced trajectory matrix, i.e., a reduced-rank form:

$$\hat{X} = \sum_{i=1}^{L} X_i, \qquad X_i = \sqrt{\lambda_i}\, U_i \otimes V_i', \qquad L < M \tag{4}$$

where $\otimes$ is the outer product and the matrices X_i are rank one matrices.

It is important to stress that in the absence of noise, $\hat{X} = X$. Thus, in the presence of noise the L strongest singular values and their associated singular vectors span the noise-free signal. It is clear that we are not interested in recovering the trajectory matrix but the signal itself. For this purpose we average the elements of the filtered trajectory matrix along the anti-diagonals of $\hat{X}$ to obtain an estimate of the enhanced signal, denoted by $\{\hat{r}_t\}_{t=1,\ldots,N}$.

Let us consider the time series $\{r_t\}_{t=1,\ldots,N}$ corresponding to the noisy star-shaped closed contour depicted in Fig.1, where $N = 1441$, $M = 54$, $K = N - M + 1 = 1388$. Figure 3 shows the contribution of each of the 54 singular values.

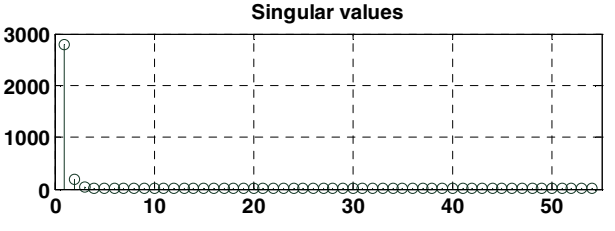

Fig. 3. The contribution of each of the 54 singular values

One can easily see that only the first two singular values have a significant contribution in recovering the smooth part of the signal (noise reduction). However, Figures 4 and 5 show that the signal can not be consistently recovered using only the first largest singular value. The second largest singular value is also needed.

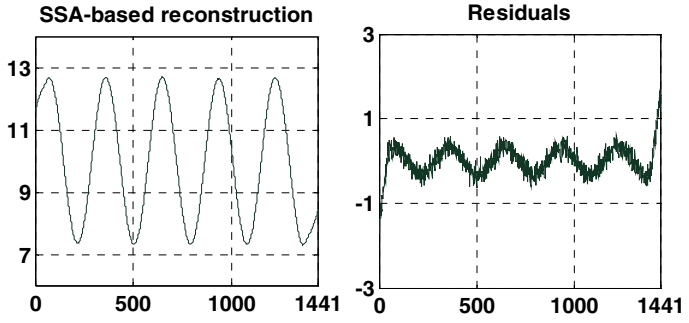

Fig. 4. SSA-based reconstruction of time series $\{\hat{r}_t\}_{t=1,\ldots,N}$ using only the first largest singular value (left) and the corresponding residuals (right)

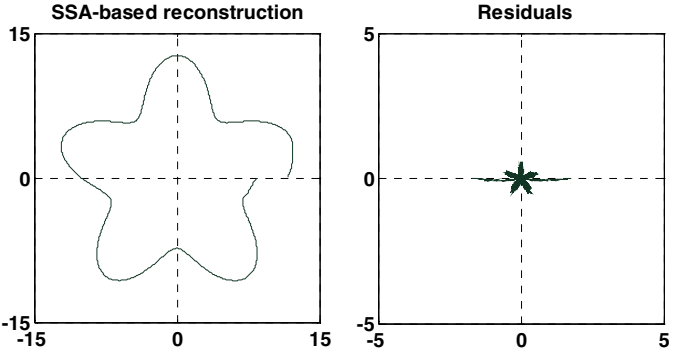

Fig. 5. SSA-based reconstruction of the contour $(\hat{x}_t, \hat{y}_t)_{t=1,\ldots,N}$ using only the first largest singular value (left) and the corresponding residuals, at a scale magnified 3 times (right)

Figures 6 and 7 show that the best choice for de-noising both the time series $\{\hat{r}_t\}_{t=1,\ldots,N}$ and the planar contour $(\hat{x}_t, \hat{y}_t)_{t=1,\ldots,N}$ is to recover the data trajectory matrix with the first 2 singular vectors. This results in a smooth reconstruction.

The exact reconstruction of the initial noisy closed contour can be also performed if all the singular vectors corresponding to non-zero singular values are used when recovering the trajectory matrix.

In the final part of this section, a SSA-based change-point detection algorithm is presented, with the aim of automating the landmark selection.

SSA-Based Change-Point Detection. A frequentist, non-parametric algorithm for multiple change-point detection in time series based on sequential application of the Singular Spectrum Analysis was developed in [5]. The idea behind the algorithm is to apply SSA to a windowed portion of the signal in order to pick up its structure through an ℓ -dimensional subspace spanned by the eigenvectors of the lag-covariance matrix,

computed in a sequence of moving time intervals [n+1; n+m] of a given length m, where n = 0, 1, ... is the iteration number. If at a certain time moment τ the mechanism generating the time series x_t has changed then an increase in the distance between the ℓ-dimensional hyperplane and the M-lagged vectors ($x_{\tau+1}, \ldots, x_{\tau+M}$) of trajectory matrix is to be expected. This increase will indicate the change. However, if the generating mechanism does not change further along the signal, then the corresponding lagged vectors will stay close to this hyperplane.

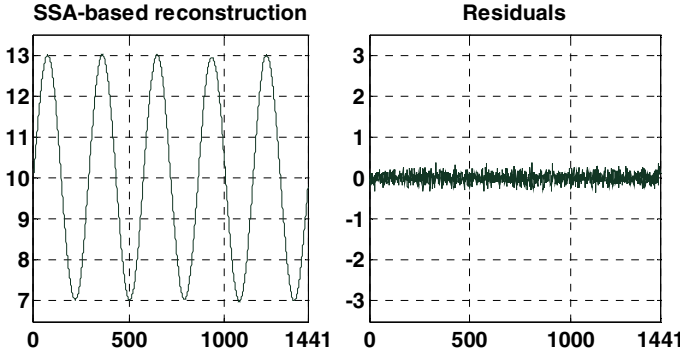

Fig. 6. SSA-based reconstruction of time series $\{\hat{r}_t\}_{t=1,\ldots,N}$ using the first two largest singular values (left) and the corresponding residuals (right)

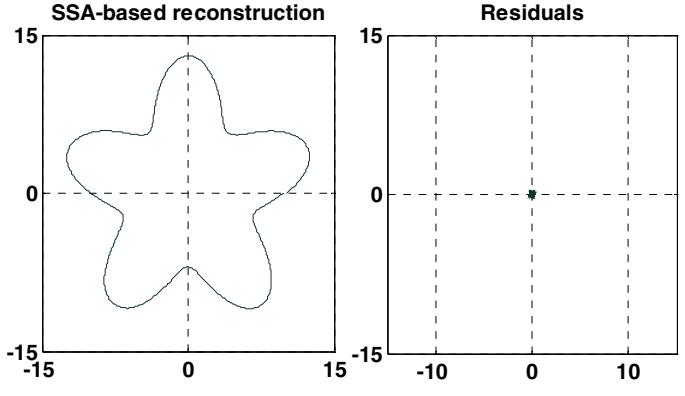

Fig. 7. SSA-based reconstruction of the contour $(\hat{x}_t, \hat{y}_t)_{t=1,\ldots,N}$ using the first two largest singular values (left) and the corresponding residuals (right)

Let $\{x_1, x_2, \ldots, x_N\}$ be a time series, where N is large enough. Two parameters have to be chosen: the window width m ($m < N$), and the lag parameter M ($M \le m/2$). Define also $K = m - M + 1$.

For each n = 0, 1, ..., $N - m$, a three-stage procedure is executed:

Stage 1. Perform the SSA algorithm for the time interval $[n + 1, n + m]$.

1. Construct the trajectory matrix $X^{(n)}$ (here called *base matrix*), whose columns are the vectors $X_j^{(n)}$:

$$X^{(n)} = \left(x_{n+i+j-1}\right)_{i=1:M;\ j=1:K} = \begin{pmatrix} x_{n+1} & x_{n+2} & \cdots & x_{n+K} \\ x_{n+2} & x_{n+3} & \cdots & x_{n+K+1} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n+M} & x_{n+M+1} & \cdots & x_{n+m} \end{pmatrix} \tag{5}$$

$$X_j^{(n)} = \left(x_{n+j}, \ldots, x_{n+M+j-1}\right)', \quad j = -n+1, -n+2, \ldots, N-n-M+1$$

2. Perform the SDV of the lag-covariance matrix $R_n = 1/K \cdot X^{(n)}\left(X^{(n)}\right)'$. This gives us a collection of M eigenvectors.

3. Select a particular group I of $\ell < M$ of these eigenvectors; this determines an ℓ-dimensional subspace $S_{n,\ell}$ in the M-dimensional space of vectors $X_j^{(n)}$. Denote the ℓ eigenvectors that determine the subspace $S_{n,\ell}$ by $U_{i_1}, \ldots, U_{i_\ell}$.

Stage 2. Construction of the test matrix.

Denote $Q = q - p$ (thus $q = p + Q$) and construct the following $M \times Q$ trajectory matrix (called *test matrix*):

$$X_{test}^{(n)} = \left(x_{n+p+i+j-1}\right)_{i=1:M;\ j=1:Q} = \begin{pmatrix} x_{n+p+1} & x_{n+p+2} & \cdots & x_{n+q} \\ x_{n+p+2} & x_{n+p+3} & \cdots & x_{n+q+1} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n+p+M} & x_{n+p+M+1} & \cdots & x_{n+q+M-1} \end{pmatrix} \tag{6}$$

The part of sample $x_{n+1}, \ldots, x_{n+m}$ that is used to construct the (base) trajectory matrix $X^{(n)}$ will be called 'base sample', and another part, $x_{n+p+1}, \ldots, x_{n+q+M-1}$, which is used to construct the vectors $X_j^{(n)}$ $(j = p+1, \ldots, q)$ and thus to compute the sum of squared distances $\mathcal{D}_{n,I,p,q}$ will be called "test sample".

Stage 3. Computation of the detection statistics

The detection statistics are:

$\mathcal{D}_{n,I,p,q}$, the sum of squares of the (Euclidean) distances between the vectors $X_j^{(n)}$ $(j = p+1, \ldots, q)$ and the ℓ-dimensional subspace $S_{n,\ell}$. Since the eigenvectors are orthogonal, the square of the Euclidean distance between an M–vector $ZY = X_j^{(n)}$ and the subspace $S_{n,\ell}$ spanned by the ℓ eigenvectors $P_1, \ldots, P_\ell$, is just $\|Z\|^2 - \|P'Z\|^2 = Z'Z - Z'PP'Z$, where $\|\cdot\|$ is the Euclidean norm and P is the $M \times \ell$-matrix with columns $P_1, \ldots, P_\ell$. Therefore

$$\mathcal{D}_{n,I,p,q} = \sum_{j=p+1}^{q} \left(X_j^{(n)}\right)' X_j^{(n)} - \left(X_j^{(n)}\right)' P P' X_j^{(n)} \tag{7}$$

The normalized sum of squared distances

$$\mathcal{D}_{n,\ell,p,q} / \mu_{n,\ell,p,q} \geq h \tag{8}$$

$S_n = \mathcal{D}_{n,I,p,q} / v_n$. Here v_j is an estimate of the sum of squared distances $\mathcal{D}_{n,I,p,q}$ at the time intervals $[\, j+1,\ j+r\,]$ where the hypothesis of no change can be accepted. Actually, $v_n = \dfrac{1}{n-m/2} \sum_{i=0}^{n-m/2-1} \mathcal{D}_{n,I,p,q}$ or $v_n = \mathcal{D}_{r,I,0,K}$ can be two alternative choices for v_n, where r is the largest value of $r \leq n$ so that the hypothesis of no change is accepted.

The decision rule in the algorithm, denoted by $A(M,m,\ell,p,q,h)$, is to announce that a change in the mechanism generating x_t occurs at a certain point τ, if for a certain n

$$\mathcal{D}_{n,I,p,q} / v_n \geq h \tag{9}$$

where h is a fixed threshold. Then we would expect than the vectors $X_j = X_{j-n}^{(n)}$ with $j > \tau$ lie further away from the ℓ-dimensional subspace $S_{n,l}$ than the vectors X_j with $j \leq \tau$. This means that the sequence $D(n) = \mathcal{D}_{n,I,p,q}$, considered as a function of n, is expected to start growing somewhere around $\hat{n}$, such that $\hat{n} + q + M - 1 = \tau$. The value $\hat{n} = \tau - q - M + 1$ is the first value of n such that the test sample $x_{n+p+1}, \ldots, x_{n+q+M-1}$ contain the change point.

In other words, $q + M - 1$ should be interpreted as a latency of test statistic in detecting change-points. Therefore, a corresponding backshift of the starting point on the contour with respect to the first position should be considered. Since closed contours are periodic in nature, such a task is easy to be done. For the time series $\{\hat{r}_t\}_{t=1,\ldots,N}$ encoding the shape signature of our star-shaped contour, a backshift of $q + M - 1 = 161$ positions is required. The test statistic is depicted in Figure 8. Actually, the location of change-points is in the local minima of test statistic function.

Figure 9 shows the landmark positions, automatically selected through change-point detection. Here, the detection statistics have been computed as normalized sum of squares of the distances between the vectors $X_j^{(n)}$ and the ℓ-dimensional subspace $S_{n,\ell}$, assuming $\ell = 1$. Increasing ℓ results in an increasing number of change-points.

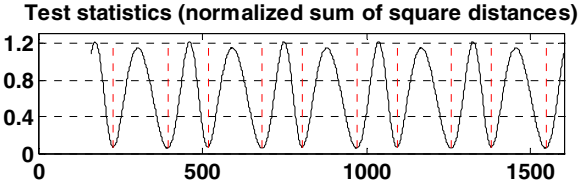

Fig. 8. Detecting change points from Distance detection statistic. The location of change points is in the local minima of test statistic function.

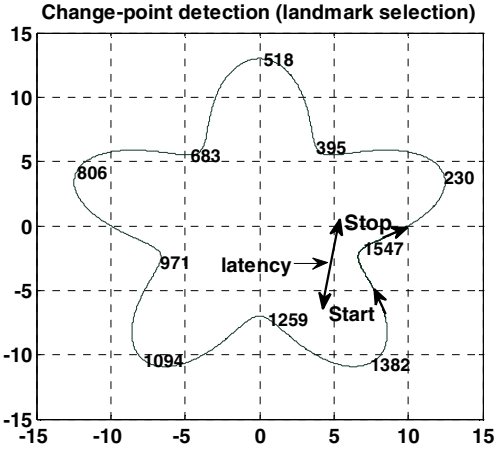

Fig. 9. Automatic selection of landmark positions through change-point detection

2.3 The Second Setting: Applying SSA to a Shape Signature Encoded by Sampling a Complex-Valued Time Series from the Contour Itself, Represented in the Complex Plane

For closed contours that are not star-shaped, a shape signature encoded by means of radius-vector function is inappropriate. In the general case, a complex-valued time series can be sampled from the contour itself, whose trace can be encoded as a sequence of complex numbers: $\{z_t\}_{t=1,\ldots,N}$, where $z_t = x_t + i\, y_t \in C$. Now, the trajectory matrix has complex elements, too:

$$X = \begin{pmatrix} z_1 & z_2 & \cdots & z_K \\ z_2 & z_3 & \cdots & z_{K+1} \\ \vdots & \vdots & \vdots & \vdots \\ z_M & z_{M+1} & \cdots & z_N \end{pmatrix}, \quad z_t \in C \tag{10}$$

The generalization of SSA for this setting is founded on the ACM Algorithm 358 for Singular Value Decomposition of a complex matrix. The decomposition theorem (Businger and Golub, [1]) can be stated as follows: each and every $M \times K$ complex-valued matrix X can be reduced to diagonal form by unitary transformations U and V, $X = U\, diag[\sigma_1, \ldots, \sigma_K]V^H$, where $\sigma_1 \geq \ldots \geq \sigma_K \geq 0$ are real-valued scalars,

called the singular values of X. Here U is an $M \times K$ column orthogonal matrix, V an $K \times K$ unitary matrix and V^H is a Hermitian transpose of V. The columns of U and V are called the left and right singular vectors of X, respectively.

As concerning the SSA-based change point detection algorithm, the sum of squared distances in equation (7) changes accordingly:

$$\mathcal{D}_{n,I,p,q} = \sum_{j=p+1}^{q} abs\left[\left(X_j^{(n)}\right)' X_j^{(n)} - \left(X_j^{(n)}\right)' P\, P'\, X_j^{(n)}\right] \tag{11}$$

Figures 10 and 11 show that the SSA-based algorithm I proposed for complex-valued trajectory matrices can be successfully applied to non star-shaped contours.

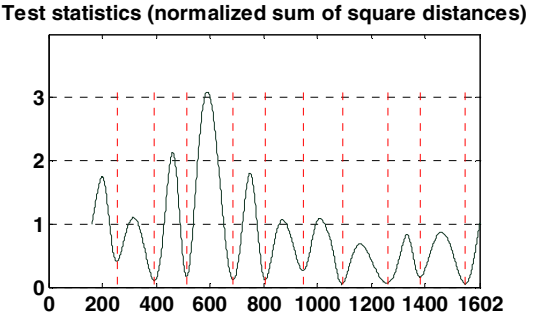

Fig. 10. Detecting change points from Distance detection statistic. The location of change points is in the local minima of test statistic function.

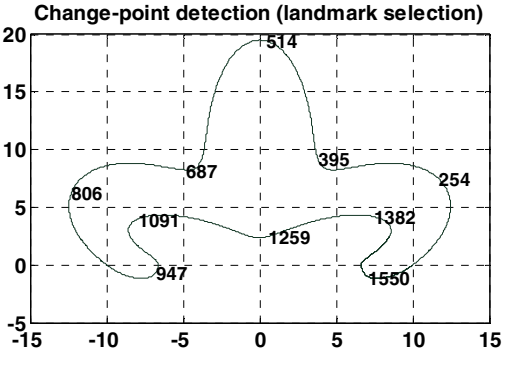

Fig. 11. Automatic selection of landmark positions through change-point detection

2.4 The Third Setting: Applying SSA to a Shape Signature Encoded by Sampling Two Real-Valued Time Series from the *x* and *y* Coordinates

Given the coordinate pairs $(x_t, y_t)_{t=1,\ldots,N}$, the trajectory matrix can now be written as follows:

$$X = \begin{pmatrix} x_1 & x_2 & \cdots & x_K \\ y_1 & y_2 & \cdots & y_K \\ x_2 & x_3 & \cdots & x_{K+1} \\ y_2 & y_3 & \cdots & y_{K+1} \\ \cdots & \cdots & \cdots & \cdots \\ x_M & x_{M+1} & \cdots & x_N \\ y_M & y_{M+1} & \cdots & y_N \end{pmatrix}, \quad x_t, y_t \in \Re \tag{12}$$

The generalization is straightforward; however, in my experiments, this version of the SSA-based change-point detection algorithm underperforms when comparing with that presented in section 2.3. A possible explanation is provided next.

3 Conclusion

For star-shaped contours the first and the second settings to my approach give identical and very accurate results, whereas the third setting produces comparable results when using for de-noising, but less accurate results when using for change-point detection.

Both the second and third settings are suitable for general-purpose applications. However, the second setting proved to be more reliable in practical cases, presumably because the number of rows in the trajectory matrix is twice as less as in the third case, when its reconstruction have to collect contributions from higher dimensional subspaces.

References

1. Businger, P.A., Golub, G.H.: Algorithm 358: Singular value decomposition of a complex matrix. Comm. ACM 12, 564–565 (1969)
2. Dryden, I.L., Mardia, K.V.: Statistical shape analysis. John Wiley and Sons, Chichester (1998)
3. Georgescu, V.: Clustering of Fuzzy Shapes by Integrating Procrustean Metrics and Full Mean Shape Estimation into K-Means Algorithm. In: 13th IFSA World Congress and 6th Conference of EUSFLAT, Lisbon, Portugal, pp. 1791–1796 (2009)
4. Goljadina, N., Nekrutkin, V., Zhigljavky, A.: Analysis of Time Series Structure: SSA and related techniques. Chapman and Holl, London (2001)
5. Moskvina, V.: Applications of the singular-spectrum analysis for change point detection in time series. Ph.D. thesis, School of Mathematics, Cardiff University, Cardiff (2001)