embedded space, projective filtering,
nonlinear noise reduction,
PCA, L1*-norm PCA*

Tomasz PRZYBYŁA[1], Janusz WRÓBEL[2], Tomasz PANDER[1],
Robert CZABAŃSKI[1], Janusz JEŻEWSKI[2], Adam MATONIA[2]

PROJECTIVE FILTERING BASED ON *L1*-NORM PCA

The paper presents a modification of nonlinear state-space projections (NSSP) method. The proposed approach deals with the sub-space estimation problem. In the original NSSP method, the principal component analysis (PCA) is used for the subspace determination. The classical PCA uses *L2*-norm. It is well known that the *L2*-norm is sensitive to outliers. Thus, in this paper the *L1*-norm PCA is proposed a subspace determination. In numerical experiments an analytic signal and real ECG signals are processed with the proposed method. The signals are contaminated with Gaussian distributed noise with different signal to noise ratio (SNR). Obtained results confirm the usefulness of the proposed modification.

1. INTRODUCTION

In many cases recorded biomedical signals contain unwanted component i.e. noise. A filtering process is applied for the noise suppression. Linear filters are successfully applied when the a noise frequency band does not overlap a frequency band of the processed signal. Significant progress in biomedical signal processing was achieved by applications of digital filters [11]. Unfortunately, the wide frequency band of the noise that overlaps the frequency band of the processed signal makes the filters practically useless. The synchronized averaging technique was introduced to cope with overlapping frequency bands [4]. Due to increasing computational power of computers, methods from the field of nonlinear dynamics are taken into account. One of these methods is the nonlinear state-space projection method (NSSP). The nonlinear state-space projection filtering was successfully applied to the ECG noise reduction [14], [8], the fetal ECG extraction [12] or the noise reduction in hydrologic time series [1]. For the projection subspace estimation, the NSSP methods involve principal component analysis (PCA). The classical PCA is based on the *L2*-norm, which is sensitive to outliers. Thus, the purpose of this work is to apply the *L1*-norm principal component analysis. In such a way the determination of the projective subspace becomes more robust.

This paper is organized as follows. Section 2 introduces the rules of the phase-space reconstruction as well as the principal component analysis based on either the *L2*-norm or the *L1*-norm. In Section 3 the performance of proposed approach is investigated. Section 4 contains a summary.

2. METHODS

The following steps are carried out to perform nonlinear projective filtering [3], [6]. First, the phase-space is reconstructed by applying the Takens theory. After the phase-space reconstruction, each point

[1]Silesian University of Technology, Institute of Electronics, Akademicka St. 16, 44-100 Gliwice, Poland.
[2]Institute of Medical Technology and Equipment, Biomedical Signal Processing Department, Roosevelta St. 118, 41-800 Zabrze, Poland.

of the trajectory is corrected. The correction stage consists of the following steps. In the first step, a vicinity is determined for each point of the trajectory. Next, based on the vicinity, a new subspace is determined and the corrected point is projected on it. The dimension of the sub-space is lower than the dimension of the phase-space. In the successive step, a reconstruction in the phase-space is performed on modified point. Finally, the time domain representation of the processed signal is computed. Below, the mentioned stages of the projective filtering are described in more details.

2.1. STATE-SPACE RECONSTRUCTION METHOD

The applied technique is an outcome of the theory of nonlinear dynamical systems. In deterministic dynamical systems, the post-transient trajectory of the system is frequently confined to a set of points in the state-space, called an attractor [6]. The state-space can be reconstructed by the Takens embedding operation [13]. For a given signal x_N, the point in the reconstructed state-space is given by

$$\mathbf{x}_n = \left[x(n), x(n+\tau), \ldots, x(n+(m-1)\tau)\right], \tag{1}$$

where $x(n)$ is the processed signal, N is the length of the signal, τ is the time lag and m is the embedding dimension. The product $(m-1)\tau$ is the embedding window. In many applications, the time lag $\tau = 1$ is advantageous [14], [8]. So, henceforth this time lag value will be used in this study. The whole processed signal x_N which is represented by a vector in the time domain, in the phase-space is represented by the matrix $\mathbf{X}$ given by

$$\mathbf{X} = \begin{bmatrix} \vdots & & & & \vdots \\ x_{n-m+1} & x_{n-m+2} & \cdots & x_{n-1} & x_n \\ \vdots & & & & \vdots \\ x_{n-1} & x_n & \cdots & x_{n+m-3} & x_{n+m-2} \\ x_n & x_{n+1} & \cdots & x_{n+m-2} & x_{n+m-1} \\ \vdots & & & & \vdots \end{bmatrix} \tag{2}$$

where: x_n denotes the amplitude of a processed signal at the time stamp n. It is essential for the time-domain reconstruction, that the element x_n occurs m times in the phase-space representation.

2.2. VICINITY DETERMINATION AND PCA

Let $\mathbf{x}^{(n)}$ be a point in the reconstructed phase-space. The vicinity $\mathcal{X}^{(n)}$ of $\mathbf{x}^{(n)}$ contains points of the trajectory which are close to $\mathbf{x}^{(n)}$, i.e.

$$\mathcal{X}^{(n)} = \left\{\mathbf{x}_k \middle| \left\|\mathbf{x}_k - \mathbf{x}^{(n)}\right\| < \varepsilon\right\}, \tag{3}$$

where ε is a radius of a hypersphere which has the center at $\mathbf{x}^{(n)}$ and $\|\cdot\|$ denotes the Euclidean distance. The radius value ε should be selected in such way, that the cardinal number of $\mathcal{X}^{(n)}$ should not be lower than an assumed value N_{min}.

Principal component analysis (PCA) is a technique for extracting a structure from high-dimensional data sets [5]. PCA is an orthogonal transformation of the coordinate systems in which the data are described. The new coordinate system (known as the principal coordinates) is obtained by the projection onto the so-called principal axes of the data. Let $\mathcal{X}^{(n)} = \left\{\mathbf{x}_1^{(n)}, \mathbf{x}_2^{(n)}, \ldots, \mathbf{x}_N^{(n)},\right\}$ be a vicinity of $\mathbf{x}^{(n)}$, where $\mathbf{x}_i^{(n)} \in \mathbb{R}^m$. Each element from the dataset is described by m features associated with the time stamp (i). Determination of the principal axes begins with centering the data samples and then computing the sample covariance matrix, i.e.

$$\mathbf{C}_X = \frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{x}_i^{(n)} - \bar{\mathbf{x}}^{(n)}\right)\left(\mathbf{x}_i^{(n)} - \bar{\mathbf{x}}^{(n)}\right)^T, \tag{4}$$

where $\bar{\mathbf{x}}^{(n)}$ is the sample mean,

$$\bar{\mathbf{x}}^{(n)} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i^{(n)},$$

and $N = |\mathcal{X}^{(n)}|$ is the cardinal number of the dataset.

The principal axes $\mathbf{w}_i$ ($1 \leq i \leq l$) are equal to the eigenvectors that correspond to the largest eigenvalues of the covariance matrix $\mathbf{C}_X$. The projection onto l-dimensional principal space is a linear transformation of $\mathbf{x}^{(n)}$ which is performed according to the following equation

$$\mathbf{y}^{(n)} = \mathbf{W}^T\left(\mathbf{x}^{(n)} - \bar{\mathbf{x}}^{(n)}\right), \tag{5}$$

where $\mathbf{y}^{(n)} \in \mathbb{R}^l$ is the l-dimensional representation of $\mathbf{x}^{(n)}$ and $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_l]$ is the projection matrix.

A reconstruction of the corrected point $\mathbf{y}^{(n)}$ in the m-dimensional space is given by

$$\hat{\mathbf{x}}^{(n)} = \mathbf{W}\mathbf{y}^{(n)} = \mathbf{W}\mathbf{W}^T\left(\mathbf{x}^{(n)} - \bar{\mathbf{x}}^{(n)}\right) + \bar{\mathbf{x}}^{(n)}. \tag{6}$$

The signal sample x_n occurs m times in the phase-space representation. Hence, the value of the filtered signal at the time stamp n is determined as a sample mean from m subsequent vector components in the phase-space, i.e.

$$\hat{x}(n) = \frac{1}{m}\sum_{i=1}^{m} x_i^{n-i+1}, \tag{7}$$

where $x_i^{(n-i+1)}$ denotes the i-th component of the vector $\hat{\mathbf{x}}^{(n-i+1)}$ in the phase space.

2.3. *L1*-NORM PCA

The PCA can be viewed in several different, but equivalent ways: as the decomposition of a covariance matrix based on its eigenvectors and eigenvalues, as a method for finding successive directions of maximum variation in data, and as a method for linear subspace estimation [5]. *L2*-norm PCA tries to find an l dimensional subspace ($l < m$). It is done by solving the following optimization problem [9], [10]

$$\begin{aligned} \mathbf{W}^* = \arg\max_{\mathbf{W}} \|\mathbf{W}^T\mathbf{S}_X\mathbf{W}\|_2 &= \arg\max_{\mathbf{W}} \|\mathbf{W}^T\mathbf{X}\|_2 \\ \text{subject to} \quad \mathbf{W}^T\mathbf{W} = \mathbf{1}_m & \end{aligned} \tag{8}$$

where $\mathbf{S}_X = \mathbf{X}^T\mathbf{X}$ is the covariance matrix and $\mathbf{1}_m$ is the $m \times m$ identity matrix. The solution of (8) is provided by the singular value decomposition (SVD) [2]. The norm $L2$ is sensitive to outliers [10], [7] while the $L1$-norm PCA resolves this problem. Similarly, applying the *L1*-norm to the optimization task (8), the $L1$ PCA can be regarded as the following optimization problem [9]

$$\begin{aligned} \mathbf{W}^* = \arg\max_{\mathbf{W}} \|\mathbf{W}^T\mathbf{S}_X\mathbf{W}\|_1 &= \arg\max_{\mathbf{W}} \|\mathbf{W}^T\mathbf{X}\|_1 \\ \text{subject to} \quad \mathbf{W}^T\mathbf{W} = \mathbf{1}_m & \end{aligned} \tag{9}$$

The first principal $\mathbf{w}_1$ can be computed by applying the following algorithm [10]:

1. Initialize $\mathbf{w}_0 \in \mathbb{R}^m$, such that $\mathbf{w}_0^T\mathbf{w}_0 = 1$, and fix $p = 1$,
2. $\mathbf{w}_p = \sum_{i=1}^{N} \operatorname{sgn}\left(\mathbf{w}_{p-1}^T\mathbf{x}_i\right)\mathbf{x}_i$,
   $\mathbf{w}_p = \frac{\mathbf{w}_p}{\|\mathbf{w}_p\|}$,
   $p = p + 1$
3. if $\mathbf{w}_p \neq \mathbf{w}_{p-1}$ then go to step 2, otherwise $\mathbf{w}_1 = \mathbf{w}_p$

The above algorithm allows to determine the first principal component. However, this method can also be used for a computation of an arbitrary number of principal components. The arbitrary number of principal components l can be computed using the algorithm presented below [10]:

1. for a given dataset $\mathbf{X}$, set $\mathbf{w}_0 = \mathbf{0}$,
2. for j=1 to l,
3. modify the dataset $\mathbf{X}^{(j)}$ such as $\mathbf{x}_i^{(j)} = \mathbf{x}_i^{(j-1)} - \mathbf{w}_{j-1}\left(\mathbf{w}_{j-1}^T \mathbf{x}_i^{(j-1)}\right)$, $(1 \leq i \leq N)$,
4. Apply the $L1$-norm PCA procedure to $\mathbf{X}^{(j)}$.

Summarizing, the following step are carried out in the proposed nonlinear projective filtering method.

First, the phase-space representation is determined (1).

Next, a vicinity is determined for each point of the trajectory (3).

Based on the *L1*-norm PCA a new subspace with lower dimensionality is computed and the corrected point is projected onto determined subspace (5).

Next, a modified point in the phase-space is computed according to (6).

Finally, the time domain representation is determined by virtue of (7).

3. EXPERIMENTS

In our numerical experiments, an additive noise model is used, i.e.

$$x(n) = s(n) + av(n),$$

where: $x(n)$ in the contaminated signal, $s(n)$ is the original (noise-less, noise-free) signal, a is an amplitude of the noise component and $v(n)$ is the noise. For the assumend model, the signal to noise ratio (SNR) is defined as follows

$$SNR = 10\ \log \frac{\sigma_s^2}{a^2 \sigma_v^2},$$

where: σ_s^2 is the variance of the original signal and σ_v^2 is the noise variance. In the conducted experiments, the noise variance is set up to $\sigma_v^2 = 4$.

In the numerical experiment two kinds of the signal are selected. Both signals are contaminated with the Gaussian distributed noise. The proposed method is tested for different levels of signal to noise ratio (SNR) and for different dimensions of the subspace. Noise reduction factor (NRF) is used to quantify the filtering efficacy. The NRF parameter is defined as follows [14]:

$$NRF = \sqrt{\frac{\|y_N - x_N\|^2}{\|c_N - x_N\|^2}},$$

where: x_N is the original signal, y_N is the artificially contaminated signal, and c_N is the result of filtering y_N. The nonlinear projective filtering method with the classical PCA is used as the reference method.

In the first experiment a fully deterministic signal which is the sinusoidal signal with the amplitude $A = 1$ and the frequency $f = 5Hz$ is used. This signal is sampled with the sampling frequency

$fs = 100Hz$. Fig. 1 depicts an example of the contaminated signal as well as the obtained signals from the proposed method and the reference one. For the both methods, the dimension of subspace is $l = 1$. The embedding dimension is selected in such a way that the vector contains at least one period of the processed signal. For the filtering process, the embedding dimension is $m = 30$. For the vicinity determination, the radius is chosen in such a way that the cardinal number of the vicinity is greater than $N_{min} = 50$. On the top of Fig. 1 the contaminated sinusoidal signal with Gaussian noise for $SNR = 0dB$. Middle of Fig. 1 depicts the obtained signal from the proposed method, and on the bottom the obtained signal from the reference method is presented. Tab. 1 contains values of NRF parameter for different values of SNR and different subspace dimensions.

Table 1. Noise reduction factor obtained for the proposed method and the reference one applied to the artificial signal. The embedding dimension is $m = 30$.

SNR [dB]	NRF			
	NSSP *L1*		NSSP *L2*	
	$l = 2$	$l = 1$	$l = 2$	$l = 1$
10	3.5675	4.0708	3.5604	4.0529
0	3.1135	3.3663	2.2350	2.7072

In the second experiment, the real ECG signal is used. This signal is a record *100.dat* from the MITBIH database. This signal is stored with the sampling frequency $fs = 360Hz$. Similarly as in the previous experiment, the computations are conducted for two different SNRs and two different subspace dimension. The embedding dimension is chosen in such way that the vector in the phasespace contains QRS complex. During the filtering, the embedding dimension is $m = 110$, which corresponds to $t = 300ms$. As in the previous experiment, the vicinity radius is selected such that the vicinity includes at least $N_{min} = 50$ neighbours. Fig. 2 presents the ECG signal. The original signal is presented on the top. Below, the corrupted signal is presented, where signal to noise ratio is $SNR = 0dB$. Fig. 2C: the obtained signal from the proposed method is presented. The signal obtained from reference method is presented on the bottom. Tab. 2 shows obtained values of the *NRF* parameter for different values of *SNR* and different dimensions of the subspace.

Table 2. Noise reduction factor obtained for the proposed method and the reference one applied to the ECG signal. The embedding dimension is $m = 110$.

SNR [dB]	NRF			
	NSSP *L1*		NSSP *L2*	
	$l = 2$	$l = 1$	$l = 2$	$l = 1$
10	2.4056	2.2918	1.8358	1.8362
0	4.3088	4.3054	2.3071	2.4699

4. CONCLUSIONS

The nonlinear phasespace projective filtering method is successfully applied in many fields of applications such as: de-noising of biomedical signals or the noise reduction in hydrologic time series. The idea of projective filtering rely on processing of the time series in the phasespace. Next, a subspace is determined and then each point of the trajectory is projected onto this subspace. The dimension of the subspace is much lower than the dimension of the phasespace. In the successive step of filtering, based on the subspace the phasespace is reconstructed. Finally, the representation of the processed signal is determined in time domain. The projecting subspace is determined by the principal component analysis (PCA) method. The traditional PCA uses the *L2*-norm. It is well known that the *L2*-norm is sensitive

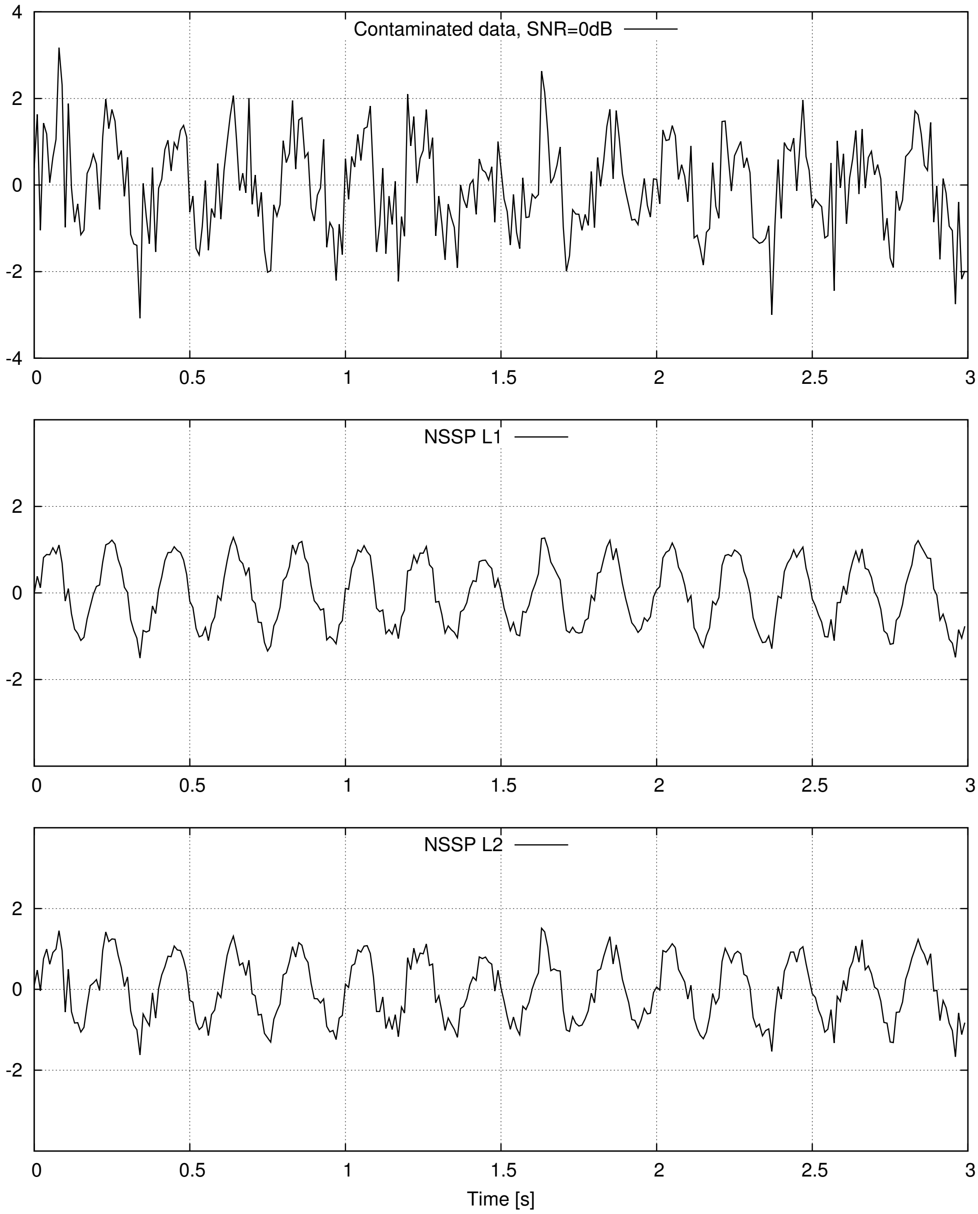

Fig. 1. Noise reduction of an analytical signal. Upper: the sinusoidal signal with the amplitude $1V$ and the frequency $f = 5Hz$. Middle: the test signal after nonlinear filtering with the proposed method. Lower: the test signal after filtering with the reference method. The embedding dimension is $m = 30$.

to the outliers in the input data. Thus, in this paper we propose the *L1*-norm PCA as the method for subspace computation. The proposed method is applied to two kinds of signals: a pure analytic signal and the real ECG signal. The noise reduction factor (NRF) is used as the quantity of filtering efficacy. The proposed method, with the *L1*-norm PCA, gives higher values of NRF parameter than the reference method that uses the *L2*-norm PCA. This means that the noise suppression efficacy is better for the proposed method.

ACKNOWLEDGEMENT

Calculations were carried out using the computer cluster Ziemowit (http://www.ziemowit.hpc.polsl.pl) funded by the Silesian BIO-FARMA project No. POIG.02.01.00-00-166/08 in the Computational Biology and Bioinformatics Laboratory of the Biotechnology Centre in the Silesian University of Technology.

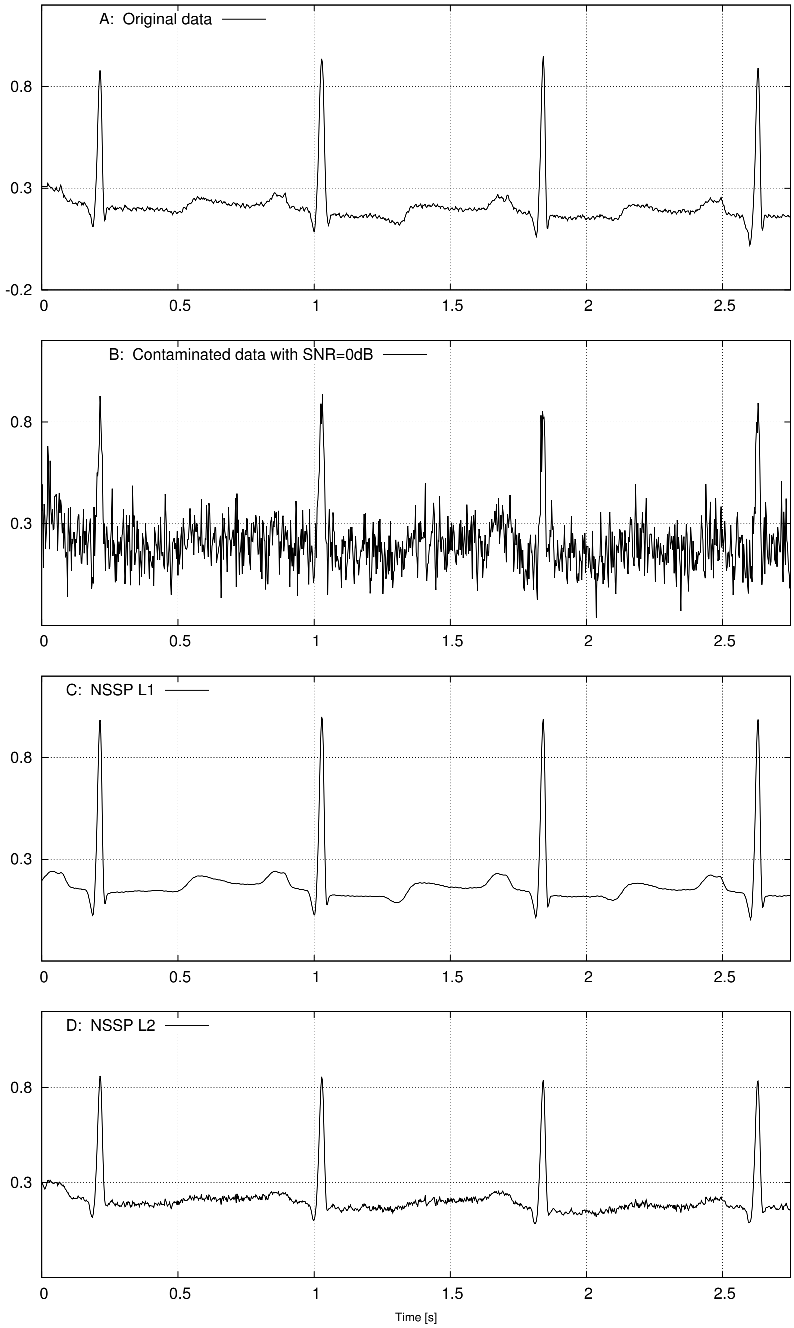

Fig. 2. Noise reduction on an electrocardiogram. A: the original signal, B: the contaminated signal with SNR=0dB. C: The filtered signal obtained with the proposed method. D: The filtered signal obtained with the reference method. The embedding dimension is $m = 110$.

BIBLIOGRAPHY

[1] ELSHORBAGY A., SIMONOVIC S. P., PANU U. S., Noise reduction in chaotic hydrologic time series: facts and doubts, J. Hydrology, 2002, No. 256, pp. 147-165.

[2] GOLUB G., Van LOAN Ch., Matrix computation, The Johns Hopkins Univ. Press, 1996.

[3] GRASSBERGER P., HEGGER R. et. al., On noise reduction methods for chaotic data, Chaos 3, 1992, pp. 127-141.

[4] JANE R., RIX H. at.al. Aligment methods for averaging of high resolution cardiac signals: a comparative study of performance, IEEE Trans. Biomed. Eng., 1991, Vol. 38, pp. 571-579.

[5] JOLLIFE I. T., Principal component analysis, Springer, New York, 2002.

[6] KANTZ H., SCHREIBER T., Nonlinear time series analysis, Cambridge Univ. Press, 2004.

[7] KOTAS M., Robust projective filtering of time-warped ECG beats, Comp. Methods and Programs in Biomedicne, 2008, No. 92, pp. 161-172.

[8] KOTAS M., Projective filtering of time warped ECG beats, Comp. in Biology and Medicine, 2008, No. 38, pp. 127-137.

[9] NIE F., HUANG H. et al, Robust principal component analysis with non-greedy l_1-norm maximization, Proc. 22nd Int'l Conf.

Artificial Intelligence, 2011.
[10] KWAK N., Principal component analysis based on $L1$-norm maximization, IEEE Trans. Pattern Analysis and Machine Learning, 2008, Vol. 30, No. 9, pp. 1672-1680.
[11] PAHLM O., SORNMO L., Data processing of exercisse ECG's, IEEE Trans. Biomed. Eng., 1987, Vol. BME-34, pp. 158-165.
[12] RICHTER M., SCHREIBER T., Fetal ECG extraction with nonlinear state space projections, IEEE Trans. Biomed. Eng., 1998, Vol. 45, pp. 133-137.
[13] TAKENS E., Detecting strange attractors in turbulence, Lecture Notes in Math, 1981, Vol. 898, pp. 366-381.
[14] SCHREIBER T., KAPLAN D., Nonlinear noise reduction for electrocardiograms, Chaos, 1996, Vol. 6, pp. 87-92.