Approximation of quasiperiodic signal phase trajectory using directional regression

Abstract: This paper solves the phase trajectory approximation problem. Quasiperiodic time series form its trajectory in high dimensional space. The trajectory is represented in the spherical coordinate system. To approximate the trajectory the authors use a directional regression technique. It finds space of minimal dimension with the phase trajectory has no self-intersections. Its self-intersections defined within the standard deviation of the reconstructed trajectory. The experiment was conducted on two data sets: data of electricity consumption during the year and sensor data of the accelerometer while walking and running.

Key words: time series, phase trajectory, directional regression, dimention of eigenspace, trajectory matrix

1 Introduction

This paper studies the quasiperiodic time series approximation problem. Examples of such series are sensor data of the accelerometer while walking, the average temperature per day during the year, data of electricity consumption [1, 2].

The phase trajectory of time series is described by its trajectory matrix or Hankel matrix. This matrix is used to analyze and predict time series. For example, in the singular spectrum analysis method (SSA), the forecast of time series is based on the singular value decomposition of the trajectory matrix [3–5]. The rows of the trajectory matrix form trajectory space of the time series. Therefore this matrix is used to study trajectory space properties. The Convergent Cross Mapping method (CCM) [6, 7] uses the trajectory matrices of two time series to check if there is a Lipschitz mapping between their phase trajectories [8].

The dimensionality of trajectory space may be excessive. This leads to instability of the forecasting model and complicates a description of the time series. In this case, it is necessary to to reduce the dimensionality or select features [9, 10]. Dimension reduction of the trajectory space is equivalent to constructing the projection of the phase trajectory into some its subspace. In this paper, we use directional regression to reduce dimensionality [11]. For a sample set $\{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)\}$ this method extracts all the required information about desired subspace from the set of empirical directions

$$\{\mathbf{x}_i - \mathbf{x}_j \mid i < j\} \tag{1}$$

and their first two conditional moments. The elements of the sample set are points in spherical coordinates. The contour regression method uses a similar approach [12]. The paper [13] describes a generalized method of directional regression, which uses generalized empirical directions

$$\{\mathbf{x}_i - c\mathbf{x}_j \mid i < j, c > 0\}$$

instead of (1). The advantages of the directional regression are high accuracy and computational efficiency.

Let the time series $\mathbf{s}$ is given. Build a projection of its phase trajectory into some trajectory subspace. We call the eigensubspace the trajectory subspace of minimal dimensionality for which some approximating model constructs an adequate approximation of the series $\mathbf{s}$.

The phase trajectory of a periodic time series is a bunch of lines. Each line describes a series for one period. It is proposed to describe the trajectory with the diameter of this bunch and also with a line approximating its expected value. We find the expected value of the bunch by averaging in-phase segements of the trajectory. The bunch diameter is defined as its dispersion around the average value. Thus, if the phase trajectory has no self intersections, then its expected value and diameter define the torus in the trajectory space. The fig. 1 shows the time series, its phase trajectory, and its expected value, highlighted by a dotted line.

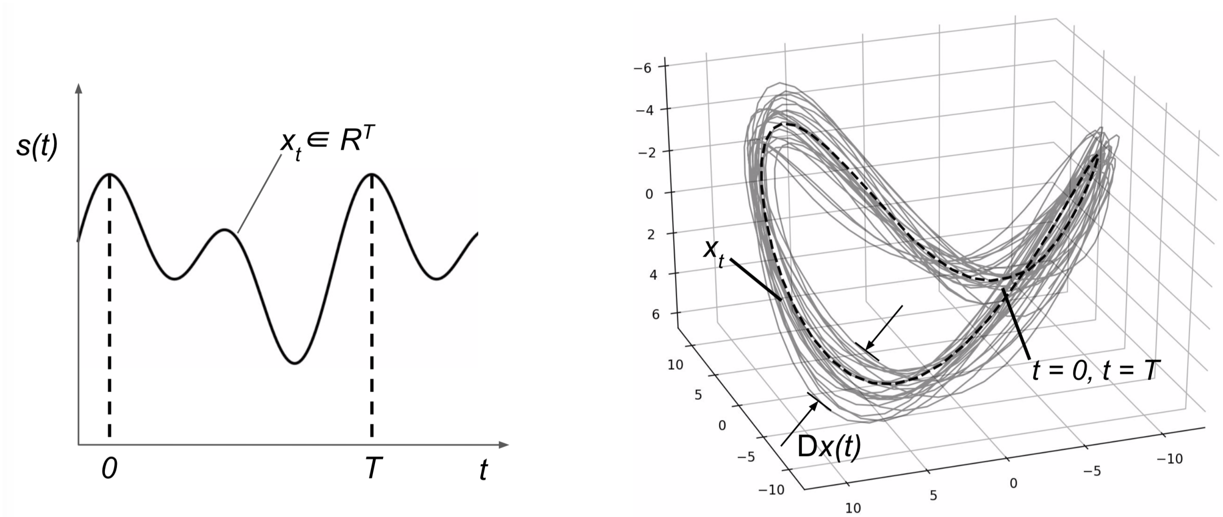


Figure 1: Time series and phase trajectory

For time series with a time-variable structure the problem of time series segementation arises [14, 15]. We can use a torus representation of the phase trajectory to solve this problem. Let the expected value of one phase trajectory statistically significantly extends beyond the torus of another phase trajectory over the entire period. Then we conclude that these trajectories describe time series of different structure or that eigensubspaces of two trajectories are significantly different.

In this paper, we solve the problem of finding the minimum necessary dimension of the phase trajectory space to construct an adequate approximation of a series. The authors also investigate self-intersections of the phase trajectory. The presence of self-intersections means that the series has more than one fundamental period. This complicates the analysis of the series and the study of the relationship between the series. If in some subspace there are no self-intersections, then the series, on the contrary, has a simple structure.

2 Representation of quasiperiodic signal in spherical coordinates

Let $\mathbf{s} = [s_1, \dots, s_N]^\mathsf{T}$ be given time series. Define its trajectory matrix as

$$\mathbf{H_s} = \begin{bmatrix} s_1 & s_2 & \dots & s_{n-1} & s_n \\ s_2 & s_3 & \dots & s_n & s_{n+1} \\ \vdots & \vdots & & \vdots & \vdots \\ s_{N-n+1} & s_{N-n+2} & \dots & s_{N-1} & s_N \end{bmatrix}, \tag{2}$$

where n is its window length. The length is not less than some supposed period. Let $\mathbf{s}_t$ be a t-th row of $\mathbf{H_s}$. Write the matrix $\mathbf{H_s}$ as

$$\mathbf{H_s} = \begin{bmatrix} \mathbf{s}_1 \\ \vdots \\ \mathbf{s}_m \end{bmatrix}, \quad \mathbf{s}_t = [s_t, s_{t+1}, \dots, s_{t+n-1}], \quad m = N - n + 1. \tag{3}$$

Note that all vectors $\mathbf{s}_t$ belong to n-dimensional trajectory space $\mathbb{H}_\mathbf{s} \subseteq \mathbb{R}^n$ of the series $\mathbf{s}$. Form its phase trajectory $\mathbf{x}(t) \in \mathbb{R}^n$. Suppose that the dimensionality of trajectory space is excessive. This leads to instability of forecasting models for the series $\mathbf{s}$. Project the trajectory $\mathbf{x}(t)$

$$\mathbf{x}_p(t) \in \mathbb{R}^p, p \leq n,$$

to some trajectory subspace. Analysis of trajectory subspaces simplifies the structure of the phase trajectory.

Approximate trajectory $\mathbf{x}(t)$ by its diameter and a scalar time function $\mathbf{x}^*(t)$ to simplify the description of the series $\mathbf{s}$ (fig. 1). Average values of $\mathbf{x}(t)$ on time segments $[1, T]$, $[T+1, 2T], \dots, [(K-1)T+1$ to find the mean trajectory value:

$$\mathbf{x}^*(t) = \frac{1}{K}\sum_{k=1}^{K} \mathbf{x}(t + kT), \tag{4}$$

where T is a period, K is a number of whole periods in $[1, N]$. Thus $\mathbf{x}^*(t)$ is an expected value of the phase trajectory:

$$\mathbf{x}^*(t) = \mathsf{E}\mathbf{x}(t). \tag{5}$$

Diameter of the trajectory is defined as a standard deviation within a bunch of traces in the trajectory:

$$\mathsf{D}\mathbf{x}(t) = \frac{2}{T}\sum_{t=1}^{T}\sqrt{\frac{1}{K}\sum_{k=1}^{K}\Big(\mathbf{x}(t+kT) - \mathbf{x}^*(t)\Big)^2}.$$

As a result, the trajectory $\mathbf{x}(t)$ is described by its expected value $\mathsf{E}\mathbf{x}(t)$ and its diameter $\mathsf{D}\mathbf{x}(t)$.

Suppose, that dimensionality of trajectory $\mathbf{x}(t)$ is lower and structure of its forecasting model is simpler in spherical coordinates. Construct a mapping of $\mathbf{x}_p(t)$ from cartesian coordinates to spherical:

$$\varphi\ :\mathbf{x}_p(t)\ \to\ \mathbf{z}_p(t) = [\alpha_1(t), \alpha_2(t), \dots\ \alpha_{p-1}(t), r(t)]. \tag{6}$$

Construct regression model $f(\cdot)$ to reconstruct the variable r by variables $\boldsymbol{\alpha} = [\alpha_1, \dots\ \alpha_{p-1}]$. Directional regression reduces dimensionality of the space of the variables $\boldsymbol{\alpha}$ by constructing a map

$$g\ :\ \underbrace{[\alpha_1, \alpha_2, \dots\ \alpha_{p-1}]}_{\boldsymbol{\alpha}}\ \to\ \underbrace{[\beta_1, \beta_2, \dots\ \beta_{q-1}]}_{\boldsymbol{\beta}}, \quad q \leq p, \tag{7}$$

such, that the variable r can be constructed using $\boldsymbol{\beta}$:

$$f \; : \; \boldsymbol{\beta} \mapsto \hat{r} \tag{8}$$

$$\hat{r} = f(\hat{\mathbf{w}}, \beta_1, \beta_2, \dots \beta_{q-1})$$
$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}} (r - \mathbf{w}^\mathsf{T}\boldsymbol{\beta})^\mathbf{2}$$

We study the dependence of the model complexity q (10) on the trajectory $\mathbf{x}(t)$. The model complexity q is defined as a minimal complexity at which the model constructs an adequate approximation

$$\hat{\mathbf{x}}_p(t) = \varphi^{-1}(\hat{\mathbf{z}}_p(t)) = \varphi^{-1}\left([\boldsymbol{\alpha}, \hat{r}(t)]\right).$$

of the phase trajectory $\mathbf{x}_p(t)$ according to the mean squared error

$$\mathrm{MSE}(\mathbf{x}_p, \hat{\mathbf{x}}_p) = \frac{1}{N}\sum_{t=1}^{N} ||\mathbf{x}_p(t) - \hat{\mathbf{x}}_p(t)||^2, \tag{9}$$

$$\min_{q\in[1,\dots,p]} : \mathrm{MSE}(\mathbf{x}_p, \hat{\mathbf{x}}_p) < \theta. \tag{10}$$

The model complexity q defines dimensionality of eigensubspace of series $\mathbf{s}$:

Definition 2.1. *Consider all subspaces of trajectory space $\mathbb{H}_\mathbf{s}$ of series $\mathbf{s}$ in which model* (8) *constructs an adequate approximation $\hat{\mathbf{x}}$ according to* (10)*. Subspace of minimal dimensionality is called eigensubspace of the series $\mathbf{s}$.*

3 Directional regression

In this section we describe the directional regression technique used in mapping (7). Let $\{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)\}$ be a random independent sample set from the set $(\mathbf{X}, \mathbf{y})$, where $\mathbf{X} \in \mathbb{R}^{(p-1)\times n}$, $\mathbf{y} \in \mathbb{R}^n$. Here a pair $(\mathbf{x}, y)$ denotes not Cartesian, but spherical coordinates (6). In the other words, a pair $(\mathbf{x}, y)$ corresponds to a vector

$$\mathbf{z}_p(t) = [\alpha_1(t), \alpha_2(t), \dots \alpha_{p-1}(t), r(t)]$$

from (6). For any given dimensionality $q \le p$ directional regression constructs projection of $\mathbf{X}$ to some space $\mathbb{S}^q_\mathbf{X} \subseteq \mathbb{R}^{(q-1)\times n}$. This space is called central subspace.

Directional regression constructs central subspace using a set of empirical directions

$$\{\mathbf{x}_i - \mathbf{x}_j \,|\, i < j\}.$$

Standardize $\mathbf{X}$:

$$\mathbf{Z} = \boldsymbol{\Sigma}^{-1/2}(\mathbf{X} - \boldsymbol{\mu}), \quad \boldsymbol{\mu} = \mathsf{E}(\mathbf{X}),\; \boldsymbol{\Sigma} = \mathbf{cov}(\mathbf{X}).$$

Let $(\tilde{\mathbf{Z}}, \tilde{\mathbf{y}})$ be a sample set from $(\mathbf{Z}, \mathbf{y})$, independent from $\{(\mathbf{z}_1, y_1), \dots, (\mathbf{z}_1, y_n)\}$. Denote

$$\mathbf{A}(\mathbf{y}, \tilde{\mathbf{y}}) = \mathsf{E}\left[(\mathbf{Z} - \tilde{\mathbf{Z}})(\mathbf{Z} - \tilde{\mathbf{Z}})^\mathsf{T} \,|\, \mathbf{y}, \tilde{\mathbf{y}}\right] \in \mathbb{R}^{n\times n}$$

In this equation each empirical direction $(\mathbf{x}_i-\mathbf{x}_j)$ is represented by $(\mathbf{z}_i-\tilde{\mathbf{z}}_j)(\mathbf{z}_i-\tilde{\mathbf{z}}_j)^{\mathsf{T}}$. Intuitively the directions from $\mathbf{Z}-\tilde{\mathbf{Z}}$, which belong to $\mathbb{S}^q_{\mathbf{Z}}$ are more dependent on $\mathbf{y}$ than the directions, which belong to orthogonal complement of $\mathbb{S}^q_{\mathbf{Z}}$. Paper [11] shows, that eigenspace of the matrix

$$\mathbf{G}=\mathsf{E}\left[2\mathbf{I}_p-\mathbf{A}(\mathbf{y},\tilde{\mathbf{y}})\right]^2$$

is an estimator of central subspace $\mathbb{S}^q_{\mathbf{X}}$. Rewrite the matrix $\mathbf{G}$ as

$$\mathbf{G}=2\mathsf{E}\Big[\mathsf{E}(\mathbf{Z}\mathbf{Z}^{\mathsf{T}}|\mathbf{y})\Big]^2+2\mathsf{E}^2\Big[\mathsf{E}(\mathbf{Z}|\mathbf{y})\mathsf{E}(\mathbf{Z}^{\mathsf{T}}|\mathbf{y})\Big]+$$

$$+2\mathsf{E}\Big[\mathsf{E}(\mathbf{Z}^{\mathsf{T}}|\mathbf{y})\mathsf{E}(\mathbf{Z}|\mathbf{y})\Big]\cdot\mathsf{E}\Big[\mathsf{E}(\mathbf{Z}|\mathbf{y})\mathsf{E}(\mathbf{Z}^{\mathsf{T}}|\mathbf{y})\Big]-2\mathbf{I}_p.$$

Let $\Omega_{\mathbf{y}}$ be a set of values of $\mathbf{y}$ in sample $\{(\mathbf{z}_1,y_1),\ldots,(\mathbf{z}_1,y_n)\}$ and let $\{J_1,\ldots,J_l\}$ be a partition of $\Omega_{\mathbf{y}}$. The matrixes $\mathbf{G}$ and $\mathbf{A}$ are estimated as follows:

$$\hat{\mathbf{G}}=\binom{l}{2}^{-1}\sum_{k<l}\Big[2\mathbf{I}_p-\hat{\mathbf{A}}(J_k,J_{k'})\Big]^2,\quad\text{where}$$

$$\hat{\mathbf{A}}(J_k,J_{k'})=\frac{\sum_{i<j}\left(\mathbf{Z}_i-\mathbf{Z}_j\right)\left(\mathbf{Z}_i-\mathbf{Z}_j\right)^{\mathsf{T}}\mathbf{I}\left(y_i\in J_k,y_j\in J_{k'}\right)}{\sum_{i<j}\mathbf{I}\left(y_i\in J_k,y_j\in J_{k'}\right)}.$$

Let $\lambda_1\geq\ldots\geq\lambda_p$ be the eigenvalues of $\hat{\mathbf{G}}$, and let $\mathbf{v}_1,\ldots,\mathbf{v}_p$ be the eigenvectors of $\hat{\mathbf{G}}$. We estimate space $\mathbb{S}^q_{\mathbf{Z}}$ as a linear span of the first q eigenvectors of $\hat{\mathbf{G}}$. We then estimate space $\mathbb{S}^q_{\mathbf{X}}$ as a linear span of the vectors $\mathbf{\Sigma}^{-1/2}\mathbf{v}_1,\ldots,\mathbf{\Sigma}^{-1/2}\mathbf{v}_q$.

4 Phase trajectory self-intersection analysis

Construct eigensubspace $\mathbb{S}$ for the series $\mathbf{s}$ using directional regression. Check whether the projection of the phase trajectory into this subspace has self-intersections. For this, we consider the expected value $\mathsf{E}\mathbf{x}(t)$ of the trajectory $\mathbf{x}^*(t)$ and check whether it has self-intersections within the diameter $\mathsf{D}\mathbf{x}(t)$. Thus, checking the self-intersections of the projection of the phase trajectory delivers us the second, empirical definition of eigensubspace.

Definition 4.1. *Let $\mathbf{x}_p$ be a projection of the phase trajectory $\mathbf{x}$ into some trajectory subspace $\mathbb{S}\subset\mathbb{R}^p$. Call subspace $\mathbb{S}$ an eigensubspace if the expected value $\mathsf{E}\mathbf{x}_p$ of the trajectory does not have any self-intersections within the trajectory diameter $\mathsf{D}\mathbf{x}_p$.*

If the above definition is satisfied, but the found subspace $\mathbb{S}_1$ is not the eigenspace $\mathbb{S}_2$ according to the definition (10)

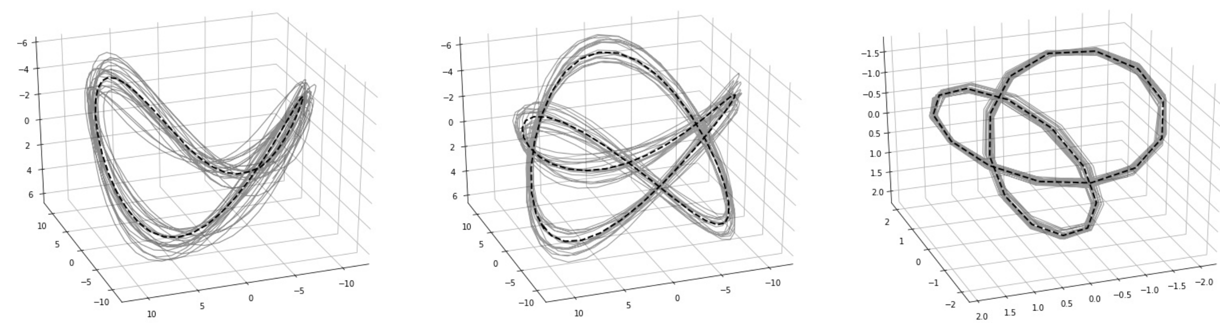
$$\mathbb{S}_1\neq\mathbb{S}_2,$$

then we increase the dimension of the found subspace $\mathbb{S}_1$. And vice versa, if some subspace satisfying the definition (10) is found, but the phase trajectory has self-intersections in this subspace, then we increase the dimensionality of the found subspace.

5 Experiment

In the experiment we study dimensionality of eigensubspace and check if phase trajectories have self-intersections. In the other words, the experiment compares definitions 2.1 and 4.1 of the eigensubspace from sections 3 and 4 correspondingly. The experiment was carried out on two data sets: sensor data of the accelerometer while walking $\mathbf{s}_1(t)$ and running $\mathbf{s}_2(t)$ [1] and the data of electricity consumption during the year $\mathbf{s}_3(t)$ [2].

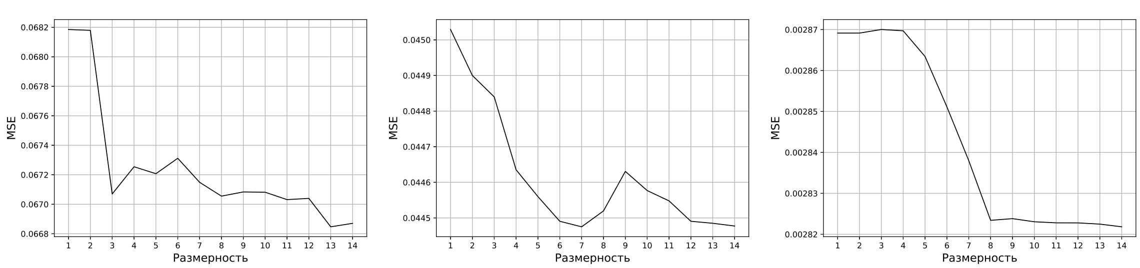
For each time series $\mathbf{s}_i, i \in [1, 3]$ build projection of its phase trajectory to the 3d space and denote projection as $\mathbf{x}_i$. Find the expected value $\mathsf{E}\mathbf{x}_i$ for each trajectory $\mathbf{x}_i$. Built trajectories and their expected values Fig. 2 shows found trajectories with the gray lines and their expected values with the black dotted line.



(a) (b) (c)

Figure 2: Phase trajectory projections and its expected values: (a) accelerometer, walking, (b) accelerometer, running, (c) energy consumption

Construct eigensubspace for each series according to (10). Go over its dimensionality $q \in [1, 15]$. For each dimensionality q, apply directional regression and find the approximation error (9). The obtained error values are shown in the fig. 3 for all three series.



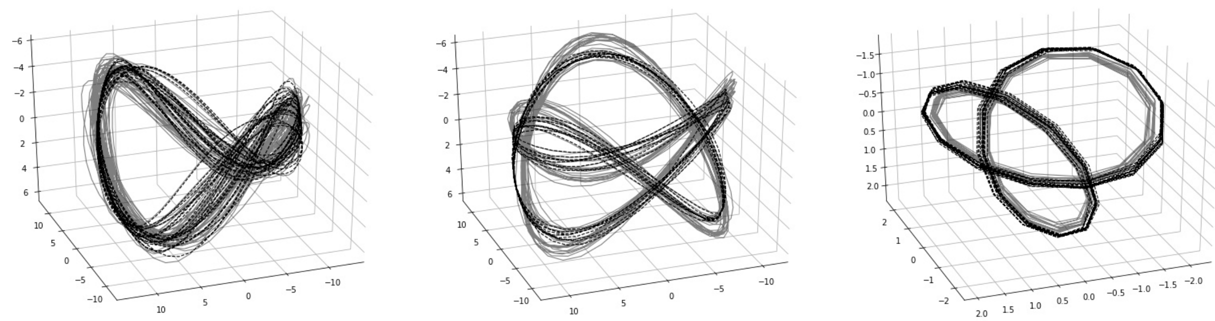
(a) (b) (c)

Figure 3: Error value depending on the dimensionality of the trajectory space: (a) accelerometer, walking, (b) accelerometer, running, (c) energy consumption

According to the definitions 2.1, for the series $\mathbf{s}_1$ dimensionality of the eigenspace equals 3, for the series $\mathbf{s}_2$ — 6 and for the series $\mathbf{s}_3$ — 8. Phase trajectories of the $\mathbf{s}_1$ and $\mathbf{s}_3$ series have no self-intersections in three-dimensional space. Therefore, according to the both definitions 2.1 and 4.1, dimensionality of the eigenspaces of the $\mathbf{s}_1$ and $\mathbf{s}_3$ series equals $q_1 = 3$ and $q_3 = 8$

respectively. Phase trajectory of the $\mathbf{s}_2$ series stops having self-intersections in four-dimensional space. Therefore, $q_2 = 6$.

Approximate phase trajectory for each time series using the model (8) found in the previous paragraph. The initial and approximated phase trajectories are shown in fig. 4 with the gray and the black dotted lines respectively.



(a) (b) (c)

Figure 4: Initial and approximated phase trajectories: (a) accelerometer, walking, (b) accelerometer, running, (c) energy consumption

6 Conclusion

This paper solved phase trajectory approximation problem for quasiperiodic time series. For the studied time series their eigensubspaces were found. Directional regression was used to reduce dimensionality of trajectory space. For this, firstly, a phase trajectory was represented in spherical coordinates and then directional regression was applied. In the found subspace, the phase trajectory was additionally checked for the presence of self-intersections.

The experiment was conducted for three time series: accelerometer sensor data during walking and running, electricity consumption during the year. For each time series, its eigensubspace was found and an approximation model of the corresponding dimensionality was constructed. The dimensionality of the phase trajectory of walking is three and for running is six, which corresponds to the expectations of experts.

The code for constructing the expected value of trajectories can be found at the link [16]. For each time series used in the experiment its trajectory was projected into three-dimensional space. And then the expected value of the trajectory was constructed. Directional regression method definition is presented in [17]. It was used for determination of the eigensubspace dimensionality. The phase trajectories were approximated according to the found dimensionality.

References

[1] M. Zhang and A. A. Sawchuk, “Usc-had: A daily activity dataset for ubiquitous activity recognition using wearable sensors,” in ACM International Conference on Ubiquitous Computing (Ubicomp) Workshop on Situation, Activity and Goal Awareness (SAGAware), (Pittsburgh, Pennsylvania, USA), September 2012.

[2] Energy consumption data https://sourceforge.net/p/mvr/code/HEAD/tree/data/EnergyConsumption.xls

[3] Golyandina N., Stepanov D. SSA-based approaches to analysis and forecast of multidimensional time series // 2005. Proceedings of the 5th St. Petersburg workshop on simulation. P. 293–298.

[4] Golyandina N., Nekrutkin V., Zhigljavsky A. A. Analysis of time series structure: SSA and related techniques. – Chapman and Hall, 2002. 320 p.

[5] Golyandina N., Zhigljavsky A. Singular Spectrum Analysis for time series. - Springer Science & Business Media, 2013. 120 p.

[6] Sugihara G., May R., Hao Ye, Chih-hao H., Deyle E., Fogarty M., Munch S. Detecting causality in complex ecosystems // Science, 2012. Vol. 338. P. 1227079

[7] Sugihara G., May R. Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series // Nature, 1990. Vol. 344. No. 6268. P. 734–741.

[8] Усманова, Карина Равилевна, Вадим Викторович Стрижов. "Модели обнаружения зависимостей во временных рядах в задачах построения прогностических моделей." Системы и средства информатики 29.2 (2019): 12-30.

[9] A. Katrutsa and V. Strijov, "Stress test procedure for feature selection algorithms," Chemometrics and Intelligent Laboratory Systems, vol. 142, pp. 172–183, 2015.

[10] J. Li, K. Cheng, S. Wang, F. Morstatter, R. P. Trevino, J. Tang, and H. Liu, "Feature selection: A data perspective," ACM Computing Surveys (CSUR), vol. 50, no. 6, p. 94, 2018.

[11] Li, Bing and Wang, Shaoli // On directional regression for dimension reduction. // Taylor & Francis, 2007

[12] Li, Bing, Hongyuan Zha, and Francesca Chiaromonte. "Contour regression: a general approach to dimension reduction." The Annals of Statistics 33.4 (2005): 1580-1616.

[13] Yu, Z., Dong, Y. and Huang, M., 2014. General directional regression. Journal of Multivariate Analysis, 124, pp.94-104

[14] Hušková, Marie, Zuzana Prášková, and Josef Steinebach. "On the detection of changes in autoregressive time series I. Asymptotics." Journal of Statistical Planning and Inference 137.4 (2007): 1243-1259.

[15] Chen, Bin, and Yongmiao Hong. "Testing for smooth structural changes in time series models via nonparametric regression." Econometrica 80.3 (2012): 1157-1183.

[16] Code for constructing trajectory expected value: https://github.com/Intelligent-Systems-Phystech/SphericalRegression/tree/master/code/approx_trajectory_with_mean.ipynb

[17] Code for applying directional regression: https://github.com/Intelligent-Systems-Phystech/SphericalRegression/tree/master/code/apply_dr.ipynb