

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
МОСКОВСКИЙ ФИЗИКО-ТЕХНИЧЕСКИЙ ИНСТИТУТ (ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ)
ФАКУЛЬТЕТ УПРАВЛЕНИЯ И ПРИКЛАДНОЙ МАТЕМАТИКИ
КАФЕДРА «ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ»

Усманова Карина Равиловна

Аппроксимация фазовой траектории квазипериодических сигналов методом сферической регрессии

03.04.01 Прикладные математика и физика (магистратура)

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА МАГИСТРА

Научный руководитель:

д.ф.-м.н. Стрижов Вадим Викторович

Москва

2020

Содержание

1	Введение	4
2	Аппроксимация фазовой траектории методом сферической регрессии	6
2.1	Траекторное пространство ряда	6
2.2	Ожидаемое значение и диаметр траектории	6
2.3	Представление квазипериодического сигнала в сферических координатах	7
3	Сферическая регрессия	8
4	Анализ самопересечений фазовой траектории	9
5	Обнаружение связи между временными рядами	9
5.1	Метод сходящегося перекрестного отображения	9
5.2	Анализ связи рядов на траекторных подпространствах	10
6	Экспериментальный анализ собственного подпространства	11
7	Заключение	14

Аннотация

Решается задача аппроксимации фазовой траектории построенной по квазипериодическому временному ряду. Фазовая траектория представлена в сферической системе координат. Для ее аппроксимации используется метод сферической регрессии. Восстанавливается регрессия координат фазовой траектории на расстояние до центра координат. Учитывается зависимость от фазы квазипериодического сигнала. Находится пространство минимальной размерности, в котором фазовая траектория не имеет самопересечений с точностью до стандартного отклонения восстановленной траектории. Эксперимент проведен на двух наборах данных: показатели потребления электроэнергии в течение года и показатели акселерометра во время ходьбы.

Ключевые слова: *временной ряд, фазовая траектория, сферическая регрессия, собственная размерность пространства, траекторная матрица*

1 Введение

Работа посвящена аппроксимации квазипериодических временных рядов. Примерами таких рядов являются показатели акселерометра во время ходьбы, средняя температура за день в течение года, показатели потребления электроэнергии [1, 2].

Фазовая траектория временного ряда описывается его траекторной матрицей или матрицей Ганкеля. Эта матрица часто используется для анализа и прогноза временных рядов. Например, в методе анализа спектральных компонент (singular spectrum analysis, SSA) прогноз временного ряда основан на сингулярном разложении траекторной матрицы [3–5]. Так как на строки траекторной матрицы натянута траекторное пространство временного ряда, то она может использоваться для исследования его траекторного пространства. В методе сходящегося перекрестного отображения (convergent cross mapping, CCM) [16, 17] траекторные матрицы временных рядов используются для проверки наличия Липшицева отображения между фазовыми траекториями [8].

Размерность траекторного пространства может оказаться избыточна. Это приводит к неустойчивости прогностической модели и усложнению описания ряда. В этом случае необходимо производить отбор признаков [9, 10]. Снижение размерности траекторного пространства эквивалентно построению проекции фазовой траектории в некоторое его подпространство. В данной работе предлагается снизить размерность траекторного пространства с помощью сферической регрессии [11]. Согласно этому методу, необходимая информация об искомом подпространстве для выборки $\{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)\}$ извлекается из множества эмпирических направлений

$$\{\mathbf{x}_i - \mathbf{x}_j \mid i < j\} \quad (1)$$

и их квадратичных моментов. Элементами выборки являются точки в сферических координатах. Аналогичный подход используется в методе контурной регрессии [12]. В [13] описан обобщенный метод сферической регрессии, где вместо множества (1) используются обобщенные эмпирические направления $\{\mathbf{x}_i - c\mathbf{x}_j \mid i < j, c > 0\}$. Преимуществами метода сферической регрессии является высокая точность и вычислительная эффективность.

Пусть для временного ряда \mathbf{s} построена проекция фазовой траектории в некоторое траекторное подпространство и задана аппроксимирующая модель. Назовем собственным подпространством траекторное подпространство минимальной размерности, при которой аппроксимирующая модель строит адекватное приближение ряда \mathbf{s} .

Фазовая траектория периодического временного ряда представляет собой пучок линий, где каждая линия описывает ряд в течение одного периода. Предлагается описывать траекторию диаметром этого пучка, а также линией, аппроксимирующей его ожидаемое значение. Ожидаемое значение пучка находится усреднением синфазных участков траектории. Диаметр пучка задается как его дисперсия вокруг среднего значения. Таким образом, если фазовая траектория не имеет самопересечений, то ее ожидаемое значение и диаметр задают тор в траекторном пространстве ряда. На рис. 1 представлен временной ряд, его фазовая траектория и ее ожидаемое значение, выделенное красным.

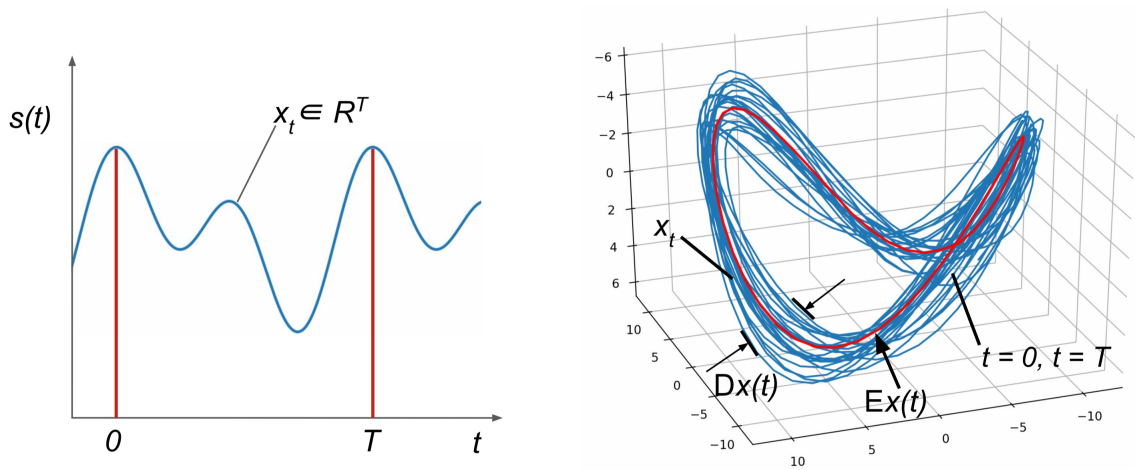


Рис. 1: Временной ряд и его фазовая траектория

Для временных рядов с переменной во времени структурой возникает задача разбиения их на сегменты, на которых их свойства не меняются [14, 15]. Для решения этой задачи используется описание фазовой траектории в виде тора. Если центр одной фазовой траектории статистически значимо выходит за пределы тора другой фазовой траектории на всем периоде, то делаем вывод, что эти траектории описывают различные по структуре ряды или что собственные пространства двух траекторий значимо различны.

В работе решается задача нахождения минимально необходимой размерности пространства фазовой траектории для построения адекватной аппроксимации ряда. Также исследуется наличие самопересечений фазовой траектории ряда. Наличие самопересечений означает, что ряд имеет более одного фундаментального периода. Это осложняет анализ ряда и изучение связи между рядами. Если в некотором подпространстве отсутствуют самопересечения фазовой траектории, то ряд, напротив, имеет простую структуру.

Поиск собственного подпространства применяется в задаче установления связи между временными рядами методом сходящегося перекрестного отображения (convergent cross mapping CCM) [16, 17]. Этот метод позволяет исследовать наличие связи не только на траекторных пространствах рядов, но и на некоторых их подпространствах.

Метод CCM основан на оценке того, насколько хорошо один ряд может быть восстановлен с использованием второго. Согласно методу сходящегося перекрестного отображения, ряд s_1 может быть восстановлен по ряду s_2 , только если ряд s_2 связан с рядом s_1 . Ряды s_1 и s_2 считаются связанными, если окрестность фазовой траектории x^1 ряда s_1 отображается в окрестность фазовой x^2 траектории ряда s_2 . Проверка наличия такого отображения осуществляется с помощью рассмотрения ближайших соседей вектора $x^1(t)$ на фазовой траектории ряда s_1 и соответствующих моменту времени t ближайших соседей на фазовой траектории ряда s_2 . Данный подход можно применить не только к исходным траекторным пространствам рядов, но и их подпространствам. В работе [8] описан метод обнаружения связи между рядами на их траекторных подпространствах. В данной работе предлагается проверить наличие связи временных рядов на их собственных подпространствах.

2 Аппроксимация фазовой траектории методом сферической регрессии

В данном разделе приводится определение и алгоритм построения собственного подпространства временного ряда.

2.1 Траекторное пространство ряда

Для заданного временного ряда $\mathbf{s} = [s_1, \dots, s_N]^T$ построим траекторную матрицу

$$\mathbf{H}_s = \begin{bmatrix} s_1 & s_2 & \dots & s_{n-1} & s_n \\ s_2 & s_3 & \dots & s_n & s_{n+1} \\ \vdots & \vdots & & \vdots & \vdots \\ s_{N-n+1} & s_{N-n+2} & \dots & s_{N-1} & s_N \end{bmatrix}, \quad (2)$$

где n – ширина окна, не меньшая, чем предполагаемый период. Обозначим t -ю строку матрицы \mathbf{H}_s как \mathbf{x}_t . Тогда матрица \mathbf{H}_s принимает вид

$$\mathbf{H}_s = \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_m \end{bmatrix}, \quad \mathbf{x}_t = [s_t, s_{t+1}, \dots, s_{t+n-1}], \quad m = N - n + 1. \quad (3)$$

Заметим, что все векторы \mathbf{x}_t принадлежат n -мерному траекторному пространству $\mathbb{H}_s \subseteq \mathbb{R}^n$ ряда \mathbf{s} , а последовательность $\{\mathbf{x}_1, \dots, \mathbf{x}_m\}$ образует фазовую траекторию $\mathbf{x}(t) \in \mathbb{R}^n$ ряда \mathbf{s} . Предполагается, что размерность траекторного пространства избыточна. Это приводит к неустойчивости прогностических моделей ряда \mathbf{s} . Поэтому предлагается рассматривать не саму траекторию $\mathbf{x}(t)$, а ее проекцию

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

в траекторное подпространство. Переход к траекторным подпространствам позволяет упростить структуру фазовой траектории и исследование ее свойств.

2.2 Ожидаемое значение и диаметр траектории

Для более компактного описания ряда \mathbf{x} предлагается аппроксимировать траекторию $\mathbf{x}(t)$ одной скалярной функцией от времени $\mathbf{x}^*(t)$ (см. рис. 1) и диаметром траектории. Для нахождения ожидаемого значения траектории усредним значения $\mathbf{x}(t)$ на отрезках времени $[1, T], [T + 1, 2T], \dots, [(K - 1)T + 1, KT]$:

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

где T – период ряда \mathbf{s} , K – число целых периодов, укладывающихся в $[1, N]$. Таким образом $\mathbf{x}^*(t)$ – это матожидание фазовой траектории

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

Диаметр траектории определим как дисперсию внутри пучка линий в траектории:

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

в результате траектория $\mathbf{x}(t)$ описывается с помощью матожидания $E\mathbf{x}(t)$ и диаметра $D\mathbf{x}(t)$.

2.3 Представление квазипериодического сигнала в сферических координатах

Предполагается, что собственная размерность фазовой траектории $\mathbf{x}(t)$ ниже, а структура ее модели проще при ее представлении в сферических координатах. Построим отображение $\mathbf{x}_p(t)$ из декартовых координат в сферические:

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

Необходимо построить регрессионную модель $f(\cdot)$, восстанавливающую переменную r по переменным $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_{p-1}]$. Метод сферической регрессии позволяет понизить размерность задачи, строя отображение

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

такое, что переменная r может быть восстановлена по $\boldsymbol{\beta}$:

$$f : \boldsymbol{\beta} \mapsto \hat{r}, \quad (8)$$

$$\hat{r} = f(\hat{\mathbf{w}}, \beta_1, \beta_2, \dots, \beta_{q-1}),$$

$$\hat{\mathbf{w}} = \arg \min_{\mathbf{w}} (r - \mathbf{w}^\top \boldsymbol{\beta})^2$$

Предлагается исследовать зависимость сложности модели q (10) от фазовой траектории $\mathbf{x}(t)$. Сложность модели q определяется как минимальная сложность, при которой модель строит адекватную аппроксимацию

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

фазовой траектории $\mathbf{x}_p(t)$ в смысле квадратичной ошибки

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

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

Сложность модели q определяет размерность собственного подпространства ряда $\mathbf{s}(t)$:

Определение 2.1. *Собственное подпространство ряда \mathbf{s} определяется как подпространство траекторного пространства $\mathbb{H}_{\mathbf{s}}$, имеющее минимальную размерность среди тех, в которых модель (8) строит адекватную аппроксимацию $\hat{\mathbf{x}}$ согласно (10).*

3 Сферическая регрессия

В данном разделе представлено описание сферической регрессии, использующейся в отображении (7).

Пусть задана случайная независимая выборка $\{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)\}$ из распределения (\mathbf{X}, \mathbf{y}) , где $\mathbf{X} \in \mathbb{R}^{(p-1) \times n}$, $\mathbf{y} \in \mathbb{R}^n$. Здесь под (\mathbf{x}, y) понимаются не декартовы координаты, а сферические (6). Иными словами пара (\mathbf{x}, y) соответствует вектору

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

из (6). Для заданной размерности $q \leq p$ метод сферической регрессии строит проекцию \mathbf{X} в некоторое пространство $\mathbb{S}_{\mathbf{X}}^q$ размерности q . Обозначим его $\mathbb{S}_{\mathbf{X}}^q \subset \mathbb{R}^{q \times n}$.

Метод сферической регрессии находит центральное подпространство с помощью рассмотрения множества эмпирических направлений:

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

Стандартизируем \mathbf{X} :

$$\mathbf{Z} = \Sigma^{-1/2}(\mathbf{X} - \boldsymbol{\mu}), \quad \boldsymbol{\mu} = \mathbb{E}[\mathbf{X}], \quad \Sigma = \text{cov}(\mathbf{X}).$$

Обозначим $(\tilde{\mathbf{Z}}, \tilde{\mathbf{y}})$ выборку из распределения (\mathbf{Z}, \mathbf{y}) , семплированную равномерно, равномошно и независимо от $\{(\mathbf{z}_1, y_1), \dots, (\mathbf{z}_n, y_n)\}$. Введем $\mathbf{A}(\mathbf{y}, \tilde{\mathbf{y}}) \in \mathbb{R}^{n \times n}$:

$$\mathbf{A}(\mathbf{y}, \tilde{\mathbf{y}}) = \mathbb{E} \left[(\mathbf{Z} - \tilde{\mathbf{Z}})(\mathbf{Z} - \tilde{\mathbf{Z}})^\top \mid \mathbf{y}, \tilde{\mathbf{y}} \right].$$

Здесь каждое эмпирическое направление $(\mathbf{x}_i - \mathbf{x}_j)$ представлено с помощью $(\mathbf{z}_i - \tilde{\mathbf{z}}_j)(\mathbf{z}_i - \tilde{\mathbf{z}}_j)^\top$. Интуитивно направления из $\mathbf{Z} - \tilde{\mathbf{Z}}$, принадлежащие $\mathbb{S}_{\mathbf{Z}}^q$, будут сильнее зависеть от \mathbf{y} , чем направления, принадлежащие ортогональному дополнению $\mathbb{S}_{\mathbf{Z}}^q$. В [11] показано, что собственное пространство матрицы

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

является оценкой центрального подпространства $\mathbb{S}_{\mathbf{X}}^q$. Матрица \mathbf{G} представима в следующем виде:

$$\begin{aligned} \mathbf{G} &= 2\mathbb{E} \left[\mathbb{E}(\mathbf{Z}\mathbf{Z}^\top \mid \mathbf{y}) \right]^2 + 2\mathbb{E}^2 \left[\mathbb{E}(\mathbf{Z} \mid \mathbf{y}) \mathbb{E}(\mathbf{Z}^\top \mid \mathbf{y}) \right] + \\ &+ 2\mathbb{E} \left[\mathbb{E}(\mathbf{Z}^\top \mid \mathbf{y}) \mathbb{E}(\mathbf{Z} \mid \mathbf{y}) \right] \cdot \mathbb{E} \left[\mathbb{E}(\mathbf{Z} \mid \mathbf{y}) \mathbb{E}(\mathbf{Z}^\top \mid \mathbf{y}) \right] - 2\mathbf{I}_p. \end{aligned}$$

Пусть $\Omega_{\mathbf{y}}$ — множество значений \mathbf{y} в выборке $\{(\mathbf{z}_1, y_1), \dots, (\mathbf{z}_n, y_n)\}$. Обозначим $\{J_1, \dots, J_l\}$ — разбиение $\Omega_{\mathbf{y}}$ на подмножества. Оценка матриц \mathbf{G} и \mathbf{A} по выборке $\{(\mathbf{z}_1, y_1), \dots, (\mathbf{z}_n, y_n)\}$ делается следующим образом:

$$\begin{aligned} \hat{\mathbf{G}} &= \binom{l}{2}^{-1} \sum_{k < l} \left[2\mathbf{I}_p - \hat{\mathbf{A}}(J_k, J_{k'}) \right]^2, \quad \text{где} \\ \hat{\mathbf{A}}(J_k, J_{k'}) &= \frac{\sum_{i < j} (\mathbf{z}_i - \mathbf{z}_j)(\mathbf{z}_i - \mathbf{z}_j)^\top \mathbf{I}(y_i \in J_k, y_j \in J_{k'})}{\sum_{i < j} \mathbf{I}(y_i \in J_k, y_j \in J_{k'})}. \end{aligned}$$

Пусть $\lambda_1 \geq \dots \geq \lambda_p$ — собственные числа, v_1, \dots, v_p — собственные векторы матрицы \mathbf{G} . Тогда пространство $\mathbb{S}_{\mathbf{Z}}^q$ можно найти как пространство, натянутое на первые q собственных векторов матрицы \mathbf{G} , а пространство $\mathbb{S}_{\mathbf{X}}^q$ — как пространство натянутое на векторы $\Sigma^{-1/2}v_1, \dots, \Sigma^{-1/2}v_q$.

4 Анализ самопересечений фазовой траектории

Пусть для ряда $\mathbf{s}(t)$ найдено собственное подпространство \mathbb{S} с помощью сферической регрессии. Предлагается проверить, является ли проекция фазовой траектории \mathbf{x} в это подпространство самопересекающейся. Для этого необходимо рассмотреть матожидание траектории $\mathbf{x}^*(t)$ и проверить, имеет ли траектория самопересечения в пределах диаметра. Таким образом, проверка самопересечений проекции фазовой траектории дает нам второе, эмпирическое определение собственного подпространства:

Определение 4.1. Пусть задана проекция \mathbf{x}_p фазовой траектории \mathbf{x} в некоторое траекторное подпространство $\mathbb{S} \subset \mathbb{R}^p$. Будем называть \mathbb{S} собственным, если ожидаемое значение траектории $\mathbf{E}\mathbf{x}_p$ не имеет пересечений в пределах диаметра траектории $D\mathbf{x}_p$.

Если приведенное определение выполняется, но при этом по (10) подпространство не является собственным, то будем увеличивать размерность подпространства. И наоборот, если найдено подпространство, удовлетворяющее (10), но при этом в нем фазовая траектория имеет самопересечения, то также необходимо увеличить его размерность.

5 Обнаружение связи между временными рядами

В данном разделе описывается метод сходящегося перекрестного отображения, а также его применение на траекторных подпространствах.

5.1 Метод сходящегося перекрестного отображения

Рассмотрим алгоритм обнаружения зависимости между рядами \mathbf{s}_1 и \mathbf{s}_2 . Обозначим фазовые траектории рядов $\mathbf{x}^1(t) = \{\mathbf{x}_1^1, \dots, \mathbf{x}_m^1\}$ и $\mathbf{x}^2(t) = \{\mathbf{x}_1^2, \dots, \mathbf{x}_m^2\}$ соответственно. Выберем момент времени $t^* \in [1, T]$ и найдем k ближайших соседей вектора $\mathbf{x}_{t^*}^1$ в траекторном пространстве $\mathbb{H}_{\mathbf{s}_1}$ ряда \mathbf{s}_1 . Обозначим их множество как $U_k(\mathbf{x}_{t^*}^1) = \{\mathbf{x}_{t_1}^1, \dots, \mathbf{x}_{t_k}^1\}$, где

$$\mathbf{x}_{t_i}^1 = [s_{t_i}^1, s_{t_i+1}^1, \dots, s_{t_i+n-1}^1], \quad i = 1, \dots, k. \quad (11)$$

Так как оба ряда \mathbf{s}_1 и \mathbf{s}_2 определены на единой временной оси, отыщем $U_k(\mathbf{x}_{t^*}^2)$, см. рис. 2, в пространстве $\mathbb{H}_{\mathbf{s}_2}$ с целью анализа связей между этими двумя рядами. Для этого повторим в пространстве $\mathbb{H}_{\mathbf{s}_2}$ построения (2), (3), (11). Аналогично строим матрицу $\mathbf{H}_{\mathbf{s}_2}$. Поставим в соответствие каждому вектору $\mathbf{x}_{t_i}^1 \in U_k(\mathbf{x}_{t^*}^1)$ вектор $\mathbf{x}_{t_i}^2$:

$$\varphi : \mathbf{x}_{t_i}^1 \rightarrow \mathbf{x}_{t_i}^2, \quad i = 1, \dots, k.$$

Найденные векторы $\mathbf{x}_{t_i}^2$ образуют множество $U_k(\mathbf{x}_{t^*}^2)$. Утверждается, что ряды \mathbf{s}_1 и \mathbf{s}_2 связаны, если отображение φ из пространства $\mathbb{H}_{\mathbf{s}_1}$ в пространство $\mathbb{H}_{\mathbf{s}_2}$ липшицево:

$$\rho_{\mathbb{H}_{\mathbf{s}_2}}(\varphi(\mathbf{x}_i^1), \varphi(\mathbf{x}_j^1)) \leq n \cdot \rho_{\mathbb{H}_{\mathbf{s}_1}}(\mathbf{x}_i^1, \mathbf{x}_j^1), \quad \forall \mathbf{x}_i^1, \mathbf{x}_j^1 \in \mathbb{H}_{\mathbf{s}_1},$$

где $\rho_{\mathbb{H}_{\mathbf{s}_1}}(\cdot, \cdot), \rho_{\mathbb{H}_{\mathbf{s}_2}}(\cdot, \cdot)$ – метрики в пространствах $\mathbb{H}_{\mathbf{s}_1}$ и $\mathbb{H}_{\mathbf{s}_2}$ соответственно. Проверим наличие такого отображения следующим образом. Введем меру близости векторов в окрестностях $U_k(\mathbf{x}_{t^*}^1)$

и $U_k(\mathbf{x}_{t^*}^2)$:

$$L(\mathbf{s}_1, \mathbf{s}_2) = \frac{R(U_k(\mathbf{x}_{t^*}^1))}{R(U_k(\mathbf{x}_{t^*}^2))}, \quad R(U_k(\mathbf{x}_{t^*}^1)) = \frac{1}{k} \sum_{i=1}^k \rho_{\mathbb{H}_x}(\mathbf{x}_{t^*}^1, \mathbf{x}_{t_i}^1). \quad (12)$$

Если $L(\mathbf{s}_1, \mathbf{s}_2)$ больше некоторого порога $L(n)$, то ряд \mathbf{s}_2 зависит от ряда \mathbf{s}_1 .

На рис. 2 показан описанный способ обнаружения связи между рядами. На верхнем рисунке ряд \mathbf{s}_2 зависит от ряда \mathbf{s}_1 , на нижнем – не зависит.

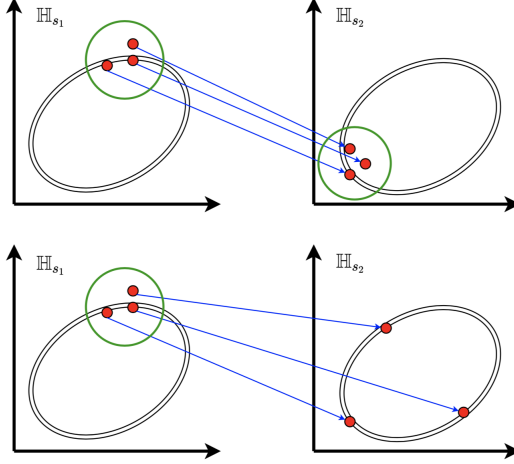


Рис. 2: Отображение из траекторного пространства ряда \mathbf{s}_1 в траекторное пространство ряда \mathbf{s}_2

5.2 Анализ связи рядов на траекторных подпространствах

Предлагается обнаружить связь между рядами \mathbf{s}_1 и \mathbf{s}_2 не в траекторных пространствах $\mathbb{H}_{\mathbf{s}_1}$ и $\mathbb{H}_{\mathbf{s}_2}$, а только в некоторых их подпространствах, натянутых на главные компоненты траекторных матриц $\mathbf{H}_{\mathbf{s}_1}$ и $\mathbf{H}_{\mathbf{s}_2}$, не обязательно первые. Сингулярное разложение матрицы $\mathbf{H}_{\mathbf{s}_1}$:

$$\mathbf{H}_{\mathbf{s}_1} = \mathbf{U}_{\mathbf{s}_1} \mathbf{\Lambda}_{\mathbf{s}_1} \mathbf{V}_{\mathbf{s}_1}.$$

Обозначим $\mathcal{J}_{\mathbf{s}_1} = \{1, \dots, n\}$ – индексы компонент ряда \mathbf{s}_1 . Подпространство траекторного пространства зададим с помощью набора индексов компонент $\mathcal{A}_{\mathbf{s}_1} \subseteq \mathcal{J}_{\mathbf{s}_1}$, на которые оно будет натянуто. Построим проекцию ряда \mathbf{s}_1 на подпространство, натянутое на компоненты с номерами из $\mathcal{A}_{\mathbf{s}_1}$. Обозначим это подпространство $\mathbb{H}_{\mathcal{A}_{\mathbf{s}_1}} \subseteq \mathbb{H}_{\mathbf{s}_1}$. Заменяем в матрице $\mathbf{\Lambda}_{\mathbf{s}_1}$ элементы, находящиеся в строках с индексами не из $\mathcal{A}_{\mathbf{s}_1}$, нулями. Обозначим полученную матрицу $\tilde{\mathbf{\Lambda}}_{\mathbf{s}_1}$. Тогда проекция ряда \mathbf{s}_1 на подпространство, натянутое на компоненты с индексами из $\mathcal{A}_{\mathbf{s}_1}$, задается траекторной матрицей

$$\mathbf{P}_{\mathcal{A}_{\mathbf{s}_1}} = \mathbf{U}_{\mathbf{s}_1} \tilde{\mathbf{\Lambda}}_{\mathbf{s}_1} \mathbf{V}_{\mathbf{s}_1}.$$

Аналогично по некоторому набору $\mathcal{A}_{\mathbf{s}_2}$ строится подпространство $\mathbb{H}_{\mathcal{A}_{\mathbf{s}_2}} \subseteq \mathbb{H}_{\mathbf{s}_2}$ и траекторная матрица $\mathbf{P}_{\mathcal{A}_{\mathbf{s}_2}}$. Предлагается искать ближайших соседей не в полных траекторных пространствах $\mathbb{H}_{\mathbf{s}_1}$ и $\mathbb{H}_{\mathbf{s}_2}$, задающихся траекторными матрицами $\mathbf{H}_{\mathbf{s}_1}$ и $\mathbf{H}_{\mathbf{s}_2}$ соответственно, а в подпространствах $\mathbb{H}_{\mathcal{A}_{\mathbf{s}_1}}$ и $\mathbb{H}_{\mathcal{A}_{\mathbf{s}_2}}$, задающихся матрицами $\mathbf{P}_{\mathcal{A}_{\mathbf{s}_1}}$ и $\mathbf{P}_{\mathcal{A}_{\mathbf{s}_2}}$. Предполагается, что переход к траекторным подпространствам меньшей размерности повышает устойчивость прогностической модели и позволяет более подробно изучить связь между рядами.

Рассмотрев различные подпространства, выбираем то, которое будет наилучшим образом описывать исследуемый временной ряд и иметь минимальную размерность. Перебор различных подпространств позволяет установить, между какими именно компонентами рядов \mathbf{s}_1 и \mathbf{s}_2 существует зависимость.

Будем перебирать различные комбинации индексов главных компонент и соответствующие им подпространства $\mathbb{H}_{\mathcal{A}_{\mathbf{s}_1}}$ и $\mathbb{H}_{\mathcal{A}_{\mathbf{s}_2}}$. Для каждой пары $(\mathcal{A}_{\mathbf{s}_1}, \mathcal{A}_{\mathbf{s}_2})$ индексов главных компонент рядов \mathbf{s}_1 и \mathbf{s}_2 соответственно будем находить среднее расстояние между k ближайшими соседями для ряда \mathbf{s}_1 и между ближайшими соседями для ряда \mathbf{s}_2 . Введем меру близости векторов, аналогичную (12) и нормированную на размерность траекторных подпространств:

$$L(\mathbf{s}_1, \mathbf{s}_2 | \mathcal{A}_{\mathbf{s}_1}, \mathcal{A}_{\mathbf{s}_2}) = \frac{R(U_k(\mathbf{x}_{t^*}^1) | \mathcal{A}_{\mathbf{s}_1})}{R(U_k(\mathbf{x}_{t^*}^2) | \mathcal{A}_{\mathbf{s}_2})} \cdot \frac{|\mathcal{A}_{\mathbf{s}_2}|}{|\mathcal{A}_{\mathbf{s}_1}|}, \quad R(U_k(\mathbf{x}_{t^*}) | \mathcal{A}_{\mathbf{x}}) = \frac{1}{k} \sum_{i=1}^k \rho_{\mathbb{H}_{\mathbf{x}}}(\mathbf{x}_{t^*}, \mathbf{x}_{t_i}). \quad (13)$$

Тогда задача поиска подпространств $\mathbb{H}_{\mathcal{A}_{\mathbf{s}_1}}$ и $\mathbb{H}_{\mathcal{A}_{\mathbf{s}_2}}$ эквивалентна поиску номеров главных компонент $(\mathcal{A}_{\mathbf{s}_1}, \mathcal{A}_{\mathbf{s}_2})$ и имеет вид

$$(\mathcal{A}_{\mathbf{s}_1}, \mathcal{A}_{\mathbf{s}_2}) = \arg \max_{\mathcal{A}_{\mathbf{s}_1}, \mathcal{A}_{\mathbf{s}_2}} L(\mathbf{s}_1, \mathbf{s}_2 | \mathcal{A}_{\mathbf{s}_1}, \mathcal{A}_{\mathbf{s}_2}), \quad \mathcal{A}_{\mathbf{s}_1}, \mathcal{A}_{\mathbf{s}_2} \subseteq \{1, \dots, n\}. \quad (14)$$

6 Экспериментальный анализ собственного подпространства

Целью эксперимента является анализ размерности собственного подпространства и проверка наличия самопересечений траектории в этом пространстве. Другими словами, в эксперименте проведено сравнение определений 2.1 и 4.1 собственного подпространства, приведенных в разделах 3 и 4. Эксперимент проведен на трех временных рядах: показатели акселерометра во время ходьбы $\mathbf{s}_1(t)$ и во время бега $\mathbf{s}_2(t)$ [1], почасовые данные потребления электроэнергии в течение года $\mathbf{s}_3(t)$ [2]. Для ряда \mathbf{s}_1 показателей акселерометра во время ходьбы исследована его связь с рядом показателей гироскопа \mathbf{s}_4 во время ходьбы на собственных подпространствах этих рядов.

Построим проекции фазовых траекторий рядов $\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3$ в трехмерное пространство, обозначим их $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ соответственно. Для каждой фазовой траектории определим ее матожидание. На рис. 3 траектории и их ожидаемые значения изображены синим и красным цветом соответственно.

Для каждого ряда найдем собственное пространство удовлетворяющее (10). Для этого переберем его размерность $q \in [1, 15]$. Для каждого значения q применим сферическую регрессию и найдем ошибку (9). Полученные значения ошибки для всех трех рядов представлены на рис. 4.

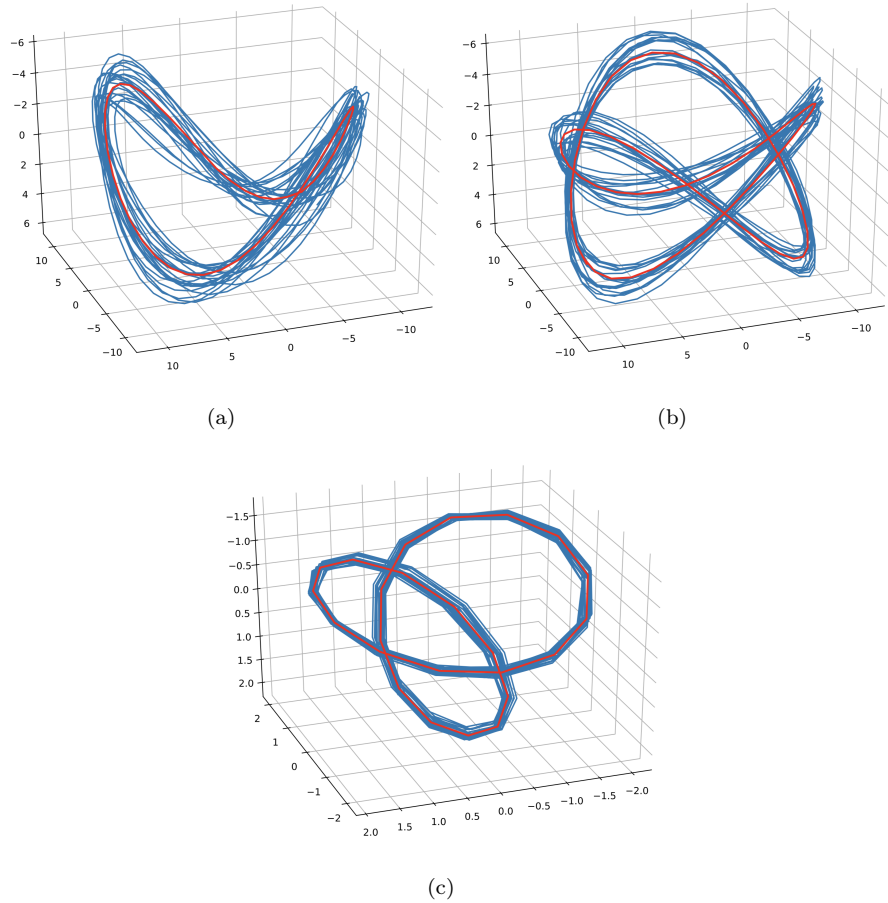


Рис. 3: Проекция фазовых траекторий и их матожидания: (а) акселерометр, ходьба, (б): акселерометр, бег, (с): потребление электроэнергии

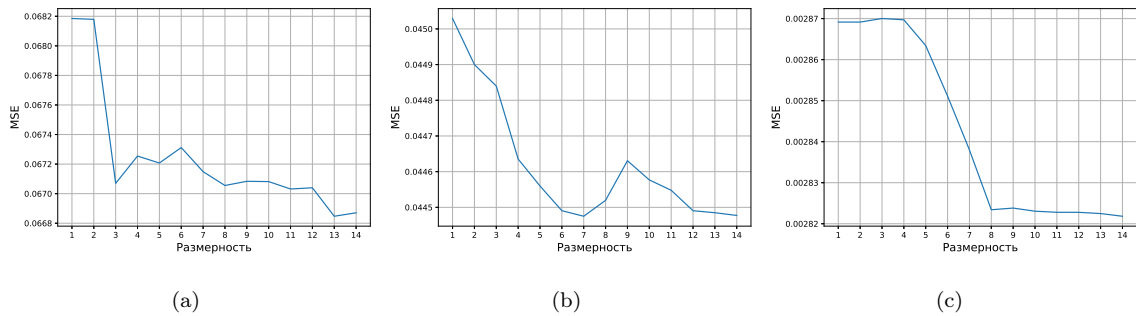


Рис. 4: Значение ошибки в зависимости от размерности пространства траектории: (а) акселерометр, ходьба, (б): акселерометр, бег, (с): потребление электроэнергии

Таким образом для ряда s_1 собственная размерность равна 3, для ряда s_2 — 6, для ряда s_3 — 8. При этом фазовые траектории рядов s_1 , s_3 не имеют самопересечений в трехмерном пространстве. Значит, согласно определениям из 3 и 4 собственная размерность s_1 равна $q_1 = 3$, а собственная размерность s_3 — $q_3 = 8$. Фазовая траектория ряда s_2 перестает самопересекаться в четырехмерном пространстве, значит $q_2 = 6$.

Для каждого из рядов построим аппроксимацию фазовой траектории моделью с найденной в предыдущем пункте размерностью. Исходные и восстановленные траектории изображены на рис. 5

синим и красным цветом соответственно.

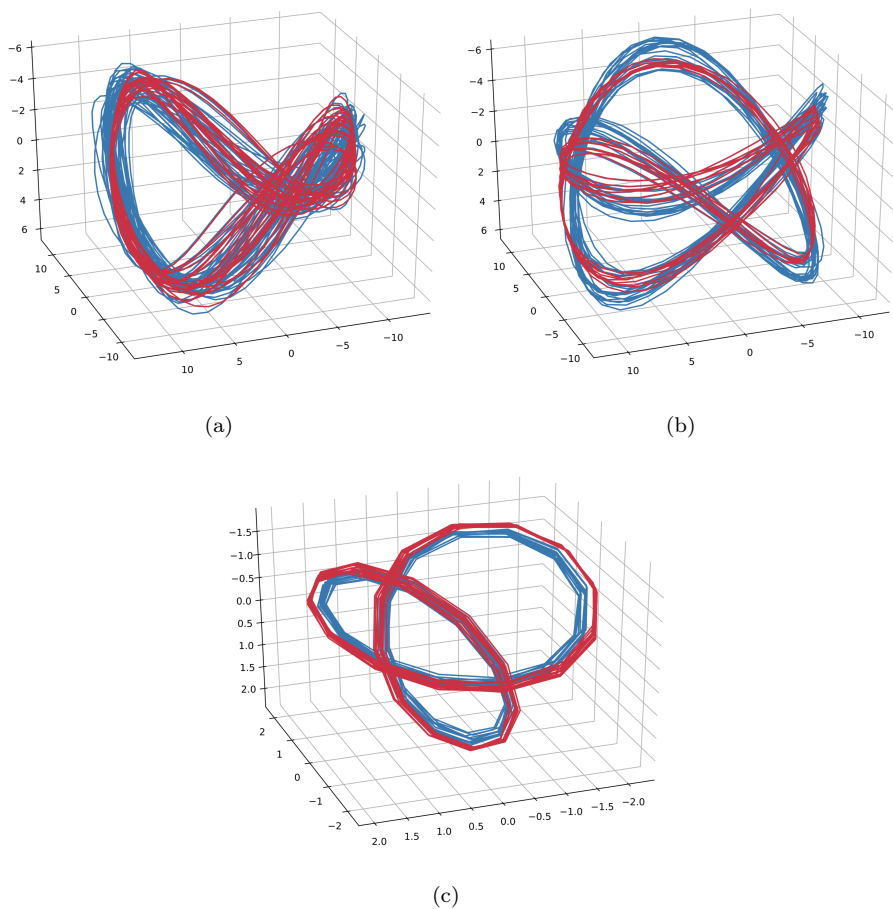


Рис. 5: Исходные и восстановленные фазовые траектории: (а) акселерометр, ходьба, (b): акселерометр, бег, (с): потребление электроэнергии

Найдем собственное подпространство для ряда s_4 показателей гироскопа во время ходьбы. Исследуем связь между рядами s_3 и s_4 на их собственных подпространствах с методом ССМ согласно описанию в разделе 5. На рис. 6 представлена проекция траектории ряда s_4 в трехмерное пространство.

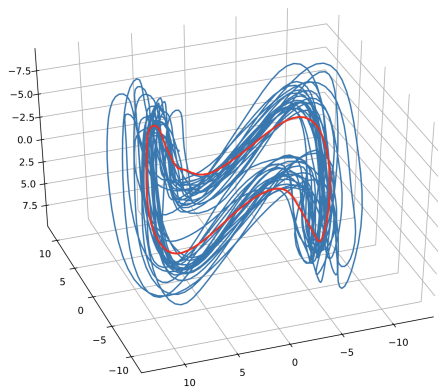


Рис. 6: Фазовая траектория и ее ожидаемое значения ряда s_4

Фиксируем ширину окна n и некоторый момент времени $t^* \in [1, N]$. Найдем k ближайших соседей векторов $\mathbf{x}_{t^*}^3$ и $\mathbf{x}_{t^*}^4$ и их расположение в траекторных пространствах рядов.

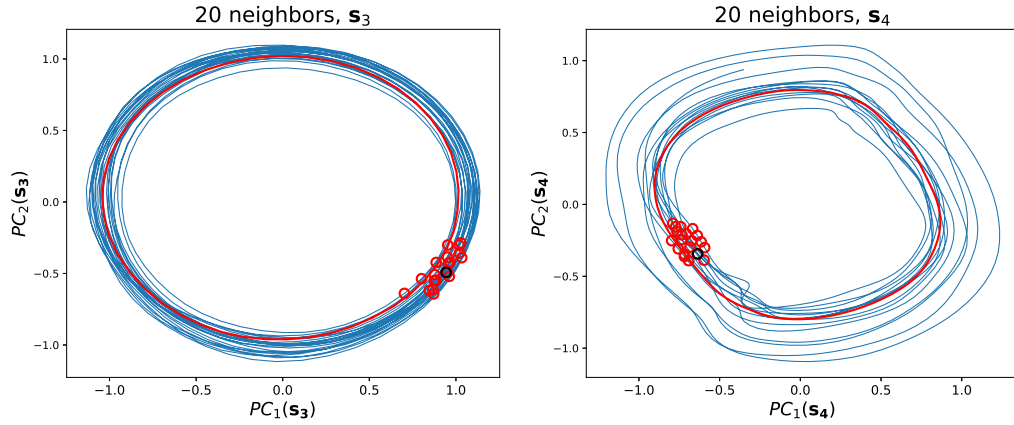


Рис. 7: Вектора $\mathbf{x}_{t^*}^3, \mathbf{x}_{t_1}^3, \dots, \mathbf{x}_{t_k}^3$ (ближайшие соседи вектора $\mathbf{x}_{t^*}^3$) и соответствующие вектора $\mathbf{x}_{t^*}^4, \mathbf{x}_{t_1}^4, \dots, \mathbf{x}_{t_k}^4$ на фазовых траекториях

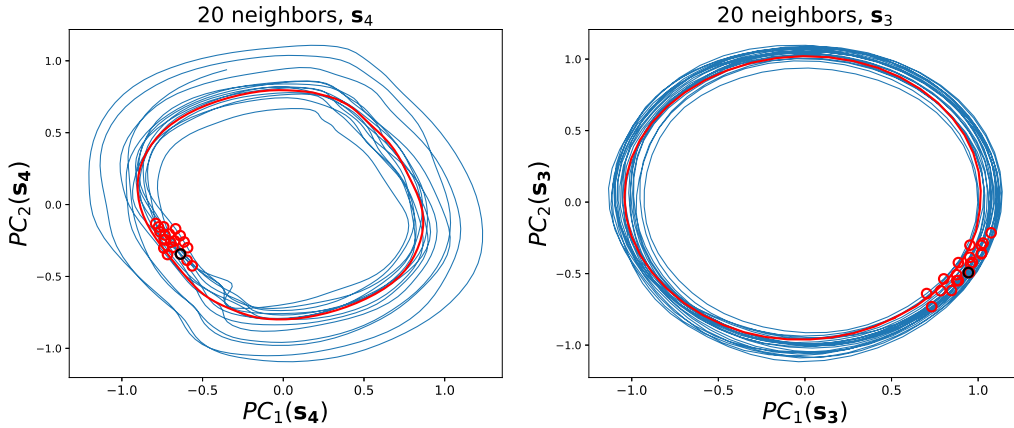


Рис. 8: Вектора $\mathbf{x}_{t^*}^4, \mathbf{x}_{t_1}^4, \dots, \mathbf{x}_{t_k}^4$ (ближайшие соседи вектора $\mathbf{x}_{t^*}^4$) и соответствующие вектора $\mathbf{x}_{t^*}^3, \mathbf{x}_{t_1}^3, \dots, \mathbf{x}_{t_k}^3$ на фазовых траекториях

Видно, что при обоих отображениях точки, близкие на одной фазовой траектории, переходят в точки, близкие на второй фазовой траектории. Значит, ряды \mathbf{s}_3 и \mathbf{s}_4 связаны.

7 Заключение

В работе решена задача аппроксимации квазипериодических временных рядов, а также задача поиска собственного подпространства временного ряда. Для снижения размерности траекторного подпространства к фазовой траектории, предварительно представленной в сферических координатах, применялся метод сферической регрессии. В найденном подпространстве фазовая траектория дополнительно исследовалась на наличие самопересечений.

Эксперимент проводился на трех временных рядах: показатели акселерометра во время ходьбы и бега, потребление электроэнергии в течение года. Для каждого временного ряда найдено собственное подпространство и построена аппроксимирующая модель соответствующей размерности.

Показано, что размерность фазовой траектории ходьбы равна трем, для бега шести, что соответствует ожиданиям экспертов. Для ряда акселерометра во время ходьбы показана его связь с рядом показателей гироскопа во время ходьбы.

Код построения ожидаемого значения траектории находится по ссылке [18]. Здесь для каждого временного ряда, использовавшегося в эксперименте, строится проекция фазовой траектории в трехмерное пространство, а также определяется ее ожидаемое значение. В коде [19] представлена реализация метода сферической регрессии. С его помощью для каждого временного ряда определяется размерность собственного подпространства. Для найденной размерности строится аппроксимация фазовой траектории.

Список литературы

- [1] Zhang M., Sawchuk A. A., 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. 2012.
- [2] Данные потребления электроэнергии: <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 // Proceedings of the 5th St. Petersburg workshop on simulation. 2005. P. 293–298.
- [4] Golyandina N., Nekrutkin V., Zhigljavsky A. A. Analysis of time series structure: SSA and related techniques // Chapman and Hall. 2002. P. 320
- [5] Golyandina N., Zhigljavsky A. Singular Spectrum Analysis for time series // Springer Science & Business Media. 2013. P. 120
- [6] Sugihara G., May R., Hao Ye, Chih-hao H., Deyle E., Fogarty M., Munch S. Detecting causality in complex ecosystems // Science. 2012. **338**. P. 1227079
- [7] Sugihara G., May R. Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series // Nature. 1990. **344**. N 6268. P. 734–741.
- [8] Усманова К. Р., Стрижов В. В. Модели обнаружения зависимостей во временных рядах в задачах построения прогностических моделей // Системы и средства информатики. 2019. **29**. N 2. P. 12-30.
- [9] Katrutsa A. and Strijov V. Stress test procedure for feature selection algorithms // Chemometrics and Intelligent Laboratory Systems. 2015. **142**, P. 172–183.
- [10] Li J., Cheng K., Wang S., Morstatter F., Trevino R. P., Tang J., Liu H. Feature selection: A data perspective. ACM Computing Surveys (CSUR). 2018. **50**. N 6. P. 94.
- [11] Li B., Wang S. On directional regression for dimension reduction. // Taylor & Francis. 2007.

- [12] Li B., Zha H., Chiaromonte F. Contour regression: a general approach to dimension reduction // The Annals of Statistics. 2005. **33**. N 4. P. 1580-1616.
- [13] Yu Z., Dong Y., Huang M. General directional regression // Journal of Multivariate Analysis. 2014. **124**. P. 94-104
- [14] Hušková M., Prášková Z., Steinebach J. On the detection of changes in autoregressive time series I. Asymptotics. // Journal of Statistical Planning and Inference. 2007. **137**. N 4. P. 1243-1259.
- [15] Chen B., Hong Y. Testing for smooth structural changes in time series models via nonparametric regression // Econometrica. 2012. **80**. N 3. P. 1157-1183.
- [16] Sugihara G., May R., Hao Ye, Chih-hao H., Deyle E., Fogarty M., Munch S. Detecting causality in complex ecosystems // Science, 2012. **338**. P. 1227079.
- [17] 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.
- [18] Ссылка на код построения ожидаемого значения траектории: https://github.com/Intelligent-Systems-Phystech/SphericalRegression/tree/master/code/approx_trajectory_with_mean.ipynb
- [19] Ссылка на код применение сферической регрессии: https://github.com/Intelligent-Systems-Phystech/SphericalRegression/tree/master/code/apply_dr.ipynb