

# Модели согласования скрытого пространства в задаче корреляционного анализа

Роберт Сафиуллин

Научный руководитель:  
д.ф.-м.н. В. В. Стрижов

Московский физико-технический институт  
Факультет управления и прикладной математики  
Кафедра Интеллектуальных систем

Москва, 2023

# Задача декодирования временного ряда

## Задача

Дана пара синхронизированных рядов из двух фазовых пространств одной системы. Требуется построить устойчивую предсказательную модель для восстановления одного ряда по известным значениям второго

## Проблема

Сложная структура рядов - наличие нелинейных зависимостей и высокая размерность данных.

## Предлагается

Построить модели для получения низкоразмерной аппроксимации измерений состояния системы в исходном и целевом пространствах. Между полученными скрытыми представлениями построить согласующую модель для отображения одного пространства на другое.

- Hadwiger, M., Theußl, T., et al. (2023). **“Riemannian Geometry for Scientific Visualization.”** In ACM SIGGRAPH Asia 2022 Courses (SIGGRAPH Asia '22). Association for Computing Machinery, New York, NY, USA, Article 5, 1–82.
- Barachant, A., Bonnet, S., et al. (2012). **“Multiclass Brain–Computer Interface Classification by Riemannian Geometry”**. In IEEE Transactions on Biomedical Engineering, vol. 59, no. 4, pp. 920–928.
- Isachenko, R., Vladimirova, M., et al. (2018). **“Dimensionality Reduction for Time Series Decoding and Forecasting Problems.”** DEStech Transactions on Computer Science and Engineering, 2018. No. optim: 286–296.

# Постановка задачи

Дано

$\{\mathbf{X}, \mathbf{Y}\} = \{x_i, y_i\}_{i=1}^N$  - выборка из сигналов.

$x_i, y_i \in \mathbb{R}^{a \times T}$

где  $a$  - число осей датчика,  $T$  - число временных отсчетов.

Требуется построить функции аппроксимации:

$$\phi : \mathbb{R}^{a \times T} \rightarrow \mathbb{R}^h$$

$$\psi : \mathbb{R}^{a \times T} \rightarrow \mathbb{R}^p$$

для получения признаков описаний исходных рядов и модель согласования этих представлений:

$$\mathbf{f} : \mathbb{R}^h \rightarrow \mathbb{R}^p$$

Затем, по полученному представлению  $\mathbf{f}(\phi(\mathbf{X}))$  строится прогноз  $\hat{\mathbf{Y}}$  ряда  $\mathbf{Y}$ :

$$\hat{\mathbf{Y}} = \mathbf{g}(\mathbf{f}(\phi(\mathbf{X})))$$

$$\mathbf{g} : \mathbb{R}^p \rightarrow \mathbb{R}^{a \times T}$$

## Определение

Симметрично положительно определенные (СПО) матрицы  $\{\mathbf{C}_j\}_{j=1}^S \in \mathbb{R}^{a \times a}$  формируют риманово многообразие  $\mathbf{M}$ , расстояние между двумя точками  $\mathbf{C}_1, \mathbf{C}_2$  которого определяется как:

$$\delta_R(\mathbf{C}_1, \mathbf{C}_2) = \|\log(\mathbf{C}_1^{-1}\mathbf{C}_2)\|_F = (\sum_{i=1}^n \log^2 \lambda_i)^{1/2} \quad (1)$$

где  $\lambda_1, \dots, \lambda_n$  - положительные собственные значения матрицы  $\mathbf{C}_1^{-1}\mathbf{C}_2$ .

Для каждой ковариационной матрицы  $\mathbf{C}_i$  существует проекция  $\mathbf{S}_i$  в касательную плоскость в точке среднего геометрического  $\mathbf{C}$ :

$$\begin{aligned} \mathbf{S}_i &= \mathbf{C}^{1/2} \log(\mathbf{C}^{-1/2} \mathbf{C}_i \mathbf{C}^{-1/2}) \mathbf{C}^{1/2} \\ \mathbf{C} &= \arg \min_{\mathbf{C}} \sum_i \delta_R(\mathbf{C}, \mathbf{C}_i)^2 \end{aligned}$$

## Гипотеза

С помощью признакового описания:

$$\text{vec}(\mathbf{S}) = [\mathbf{S}_{1,1}; \sqrt{2}\mathbf{S}_{1,2}; \sqrt{2}\mathbf{S}_{1,3}; \mathbf{S}_{2,2}; \sqrt{2}\mathbf{S}_{2,3}; \mathbf{S}_{3,3}]$$

составленного из элементов проекций ковариационных матриц на касательное к их средней точке пространство, можно построить устойчивую модель прогнозирования.

Пусть  $\mathbf{X} = \{x_i\}_{i=1}^N$  и  $\mathbf{Y} = \{y_i\}_{i=1}^M$  обозначают временные ряды, полученные от двух различных устройств, согласованных таким образом, что они представляют собой отображения одной и той же системы. Пусть представления этих временных рядов в виде ковариационных матриц обозначены как  $\{\mathbf{C}_i^{\mathbf{X}}\}_{i=1}^N$  и  $\{\mathbf{C}_i^{\mathbf{Y}}\}_{i=1}^M$  соответственно.

## Теорема 1. (Сафиуллин, 2023)

Эти представления лежат на одном и том же Римановом многообразии, но в разных его точках, определяющихся средними ковариационных матриц данных измерений.

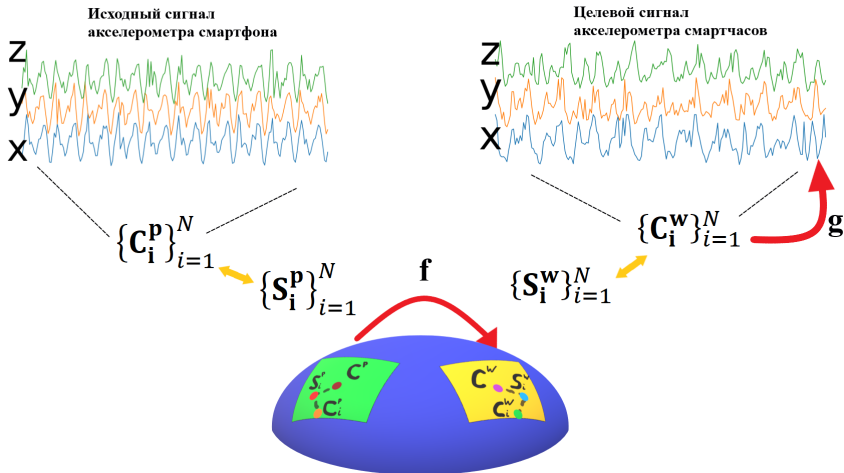
В соответствии с *Теоремой 1*,  $\phi(\mathbf{X})$  и  $\psi(\mathbf{Y})$  являются касательными пространствами к поверхности Римана, а значит между ними существует диффеоморфизм  $\mathbf{f}$ . Для моделирования  $\mathbf{f}$ ,  $\mathbf{g}$  использовалась полносвязная нейронная сеть

Прогноз  $\hat{\mathbf{Y}}$  ряда  $\mathbf{Y}$  строится как:

$$\hat{\mathbf{Y}} = \mathbf{g}(\mathbf{f}(\phi(\mathbf{X})))$$
$$\mathbf{g} : \mathbb{R}^p \rightarrow \mathbb{R}^{3 \times 200}$$



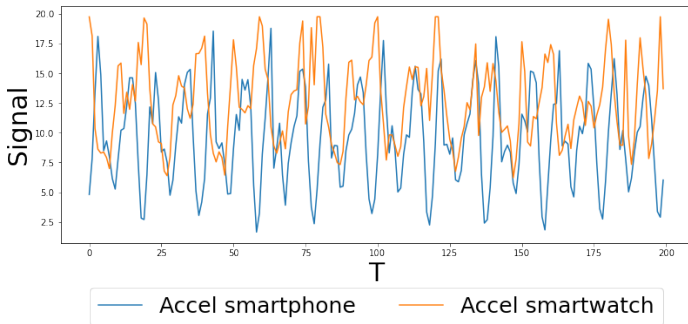
# Схема метода



## Датасет: Smartphone and Smartwatch Activity and Biometrics Dataset

$\{\mathbf{X}, \mathbf{Y}\} = \{x_i, y_i\}_{i=1}^N$  - выборка из сигналов с акселерометра смартфона и смартчасов соответственно.

$x_i, y_i \in \mathbb{R}^{3 \times 200}$



1

<sup>1</sup>Gary M. Weiss, Kenichi Yoneda, and Thaier Hayajneh. Smartphone and Smartwatch-Based Biometrics Using Activities of Daily Living. IEEE Access, 7:133190-133202, Sept. 2019.

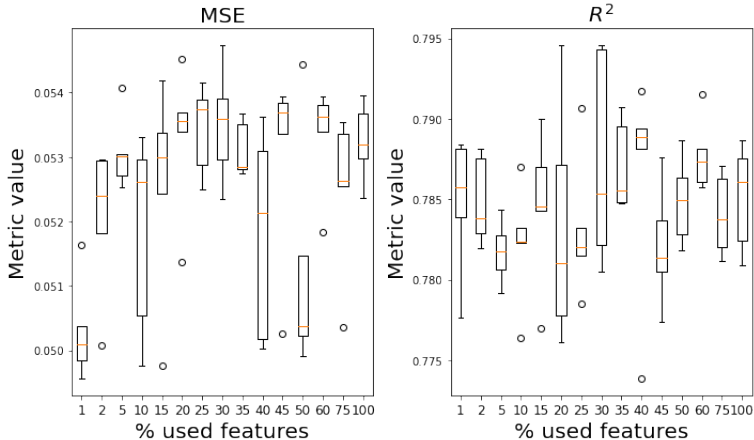
## Цели эксперимента

- 1 Проверить качество предсказания с помощью предложенного метода по метрикам:  $R^2$ ,  $MSE$
- 2 Изучить зависимость качества от количества выбранных признаков в сравнении с другим способом аппроксимации измерений.

Для проверки устойчивости метода использован отбор признаков на основе **QPFS**

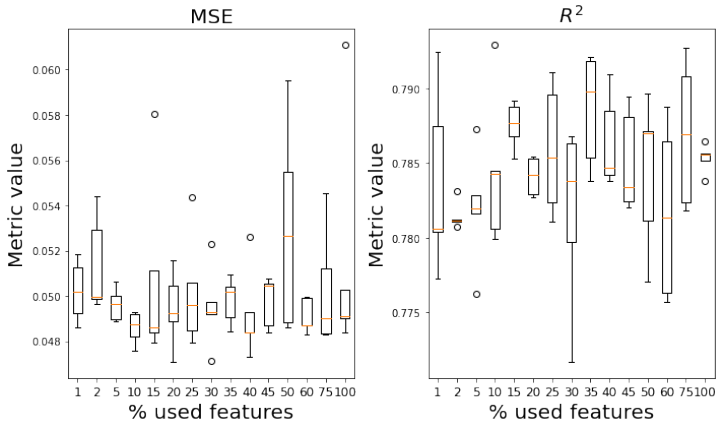
Сравнение метода производится с аппроксимациями сигналов, полученными с помощью автокодировщика

## Скрытое представление автокодировщика



Высокие показатели коэффициента корреляции. Большой разброс ошибки

## Проецирование на многообразии



Использование проекций ковариационных матриц в качестве описания систем позволило достичь более низких показателей ошибки.

- 1 Предложен и теоретически обоснован метод восстановления временных рядов с помощью мультимodelей
- 2 Выполнен численный эксперимент для проверки качества метода
- 3 Показано, что предложенный метод успешно решает задачу корреляционного анализа и показывает высокую устойчивость

## Публикации по специальности

- Boginskaya, I.; Safiullin, R.; et al.  
*Human Angiotensin I-Converting Enzyme Produced by Different Cells: Classification of the SERS Spectra with Linear Discriminant Analysis*. Biomedicines 2022, 10, 1389.
- Slipchenko, E.A.; Boginskaya, I.A.; Safiullin, R.R.; et.al.  
*SERS Sensor for Human Glycated Albumin Direct Assay Based on Machine Learning Methods*. Chemosensors 2022, 10, 520.