

# Выбор тензорных представлений для прогнозирования по мультимодальным измерениям.

Надежда Алсаханова

Научный руководитель:

д.ф.-м.н. В. В. Стрижов

Московский физико-технический институт  
Физтех-школа прикладной математики и информатики  
Кафедра «Интеллектуальные системы»

21 июня 2023 г.

# Задача декодирования временного ряда

## Цель

Исследовать зависимости в пространствах объектов и ответов и построить устойчивую модель декодирования временных рядов в случае коррелированного описания данных с учетом внутренних корреляций.

## Проблема

Данные высоко коррелированы, поэтому для устойчивости моделей применяют снижение размерности. Применение снижения размерности ведёт к потери информации ввиду не учёта скрытых корреляций по времени и пространству.

## Решение

Предлагается применять ганкелизацию перед снижением размерности по временной и пространственной координатам для учета внутренних корреляций.

## Литература

- Lu, H., Plataniotis, K. N., Venetsanopoulos, A. N. (2006, August). Multilinear principal component analysis of tensor objects for recognition. // *In 18th International Conference on Pattern Recognition (ICPR'06)* (Vol. 2, pp. 776-779). IEEE.
- Zhao, Q., Caiafa, C. F., Mandic, D. P., Chao, Z. C., Nagasaka, Y., Fujii, N., ... Cichocki, A. (2012). Higher order partial least squares (HOPLS): A generalized multilinear regression method. // *IEEE transactions on pattern analysis and machine intelligence*, 35(7), 1660-1673.
- Ran, X., Chen, W., Yvert, B., Zhang, S. (2022). A hybrid autoencoder framework of dimensionality reduction for brain-computer interface decoding. // *Computers in Biology and Medicine*, 148, 105871.
- Shi, Q., Yin, J., Cai, J., Cichocki, A., Yokota, T., Chen, L., ... Zeng, J. (2020, April). Block Hankel tensor ARIMA for multiple short time series forecasting. // *In Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 34, No. 04, pp. 5758-5766).
- Shen, Z., Liu, B., Zhou, Q., Liu, Z., Xia, B., Li, Y. (2023). Cost-sensitive Tensor-based Dual-stage Attention LSTM with Feature Selection for Data Center Server Power Forecasting. // *ACM Transactions on Intelligent Systems and Technology*, 14(2), 1-20.

## Постановка задачи

Дано

Пусть дана выборка  $\{\underline{\mathbf{X}}_m, \underline{\mathbf{Y}}_m\}_{m=1}^M$ , где  $\underline{\mathbf{X}}_m \in \mathbb{R}^{I_1 \times \dots \times I_N}$  – тензор независимых переменных,  $\underline{\mathbf{Y}}_m \in \mathbb{R}^{J_1 \times \dots \times J_N}$  – тензор целевых переменных.

Модель

$$\underline{\mathbf{Y}}_m = \underline{\mathbf{X}}_m \times_1 \mathbf{W}^{(1)} \dots \times_N \mathbf{W}^{(N)} + \underline{\mathbf{E}}.$$

Функция потерь

$$\sum_{m=1}^M \|\underline{\mathbf{Y}}_m - \underline{\mathbf{X}}_m \times_1 \mathbf{W}^{(1)} \dots \times_N \mathbf{W}^{(N)}\|_F^2 \rightarrow \min_{\mathbf{W}^{(1)}, \dots, \mathbf{W}^{(N)}},$$

$$\mathbf{W}_m^{(n)} = \left( \underline{\mathbf{Y}}_m^{(n)} \underline{\mathbf{X}}_{m(n)}^{(n)\top} \right) \left( \underline{\mathbf{X}}_{m(n)}^{(n)} \underline{\mathbf{X}}_{m(n)}^{(n)\top} \right)^{-1}$$

Проблема

Линейная зависимость столбцов матрицы  $\underline{\mathbf{X}}_{m(n)}^{(n)}$  приводит к неустойчивому решению  $\mathbf{W}_m^{(n)}$ ,  $1 \leq m \leq M$ .

## Multilinear Principal Component Analysis (MPCA)

Основная цель MPCA состоит в том, чтобы определить многолинейный преобразования  $\{\tilde{\mathbf{U}}^{(d)} \in \mathbb{R}^{I_d \times L_d}, d = 1, \dots, D\}$ , отображающие

$$\mathbb{R}^{M \times I_1 \times I_2 \times \dots \times I_D} \rightarrow \mathbb{R}^{M \times L_1 \times L_2 \times \dots \times L_D}, L_d < I_d, d = 1, \dots, D, m = 1, \dots, M$$

следующим образом:

$$\hat{\mathbf{X}}_m \approx \tilde{\mathbf{X}}_m \times_1 \tilde{\mathbf{U}}^{(1)\top} \times_2 \tilde{\mathbf{U}}^{(2)\top} \dots \times_D \tilde{\mathbf{U}}^{(D)\top}, m = 1, \dots, M,$$

где  $\tilde{\mathbf{X}}_m \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_D}$  – центрированный исходный тензор,  
 $\hat{\mathbf{X}}_m \in \mathbb{R}^{L_1 \times L_2 \times \dots \times L_D}$  – малоразмерное представление  $\tilde{\mathbf{X}}_m$ .

Проекционные матрицы  $\{\tilde{\mathbf{U}}^{(d)} \in \mathbb{R}^{I_d \times L_d}, d = 1, \dots, D\}$  находятся следующим образом:

$$\sum_{m=1}^M \|\hat{\mathbf{X}}_m\|_F^2 \rightarrow \max_{\tilde{\mathbf{U}}^{(1)}, \dots, \tilde{\mathbf{U}}^{(D)}}.$$

Здесь размерность  $L_d$  для каждой моды предполагается известной или заданной.

## Higher-Order Partial Least Squares (HOPLS)

$$\underline{\mathbf{X}} = \sum_{1 \leq r \leq R} \underline{\mathbf{G}}_{xr} \times_1 \mathbf{t}_r \times_2 \mathbf{P}_r^{(1)} \dots \times_{D+1} \mathbf{P}_r^{(D)} + \underline{\mathbf{E}}_R,$$

$$\underline{\mathbf{Y}} = \sum_{1 \leq r \leq R} \underline{\mathbf{G}}_{yr} \times_1 \mathbf{t}_r \times_2 \mathbf{Q}_r^{(1)} \dots \times_{D+1} \mathbf{Q}_r^{(D)} + \underline{\mathbf{F}}_R,$$

где

- $R$  – количество скрытых векторов,  $\mathbf{t}_r \in \mathbb{R}^M$  –  $r$ -ый скрытый вектор;
- $\{\mathbf{P}_r^{(d)}\}_{d=1}^D \in \mathbb{R}^{I_d \times L_d}$  и  $\{\mathbf{Q}_r^{(d)}\}_{d=1}^D \in \mathbb{R}^{J_d \times K_d}$  – матрицы перехода из латентного пространства в исходное;
- $\underline{\mathbf{G}}_{xr} \in \mathbb{R}^{1 \times L_1 \times \dots \times L_D}$  и  $\underline{\mathbf{G}}_{yr} \in \mathbb{R}^{1 \times K_1 \times \dots \times K_D}$  – ядра разложения Таксера.

Определив  $\mathbf{T} = [\mathbf{t}_1, \dots, \mathbf{t}_R]$ ,  $\bar{\mathbf{P}}^{(d)} = [\mathbf{P}_1^{(d)}, \dots, \mathbf{P}_R^{(d)}]$ ,  $\bar{\mathbf{Q}}^{(d)} = [\mathbf{Q}_1^{(d)}, \dots, \mathbf{Q}_R^{(d)}]$  и

$$\underline{\mathbf{G}}_x = \text{blockdiag}(\underline{\mathbf{G}}_{x1}, \dots, \underline{\mathbf{G}}_{xR}) \in \mathbb{R}^{R \times RL_1 \times \dots \times RL_D},$$

$$\underline{\mathbf{G}}_y = \text{blockdiag}(\underline{\mathbf{G}}_{y1}, \dots, \underline{\mathbf{G}}_{yR}) \in \mathbb{R}^{R \times RK_1 \times \dots \times RK_D},$$

HOPLS можно переписать как

$$\underline{\mathbf{X}} = \underline{\mathbf{G}}_x \times_1 \mathbf{T} \times_2 \bar{\mathbf{P}}^{(1)} \dots \times_{D+1} \bar{\mathbf{P}}^{(D)} + \underline{\mathbf{E}}_R,$$

$$\underline{\mathbf{Y}} = \underline{\mathbf{G}}_y \times_1 \mathbf{T} \times_2 \bar{\mathbf{Q}}^{(1)} \dots \times_{D+1} \bar{\mathbf{Q}}^{(D)} + \underline{\mathbf{F}}_R.$$

# Autoencoder ReducedNet

Состоит из энкодера ( $h$ ), декодера ( $\psi$ ) и модуля декодирования ( $\omega$ ).

$$\underline{\mathbf{X}} \in \mathbb{R}^{M \times I_1 \times \dots \times I_G} \xrightarrow{\text{conv1D blocks}} \tilde{\underline{\mathbf{X}}} \in \mathbb{R}^{M \times I_1 \times L_2 \times \dots \times L_G} \xrightarrow{\text{LSTM block}} \hat{\underline{\mathbf{X}}} \in \mathbb{R}^{M \times L_1 \times L_2 \times \dots \times L_G}.$$

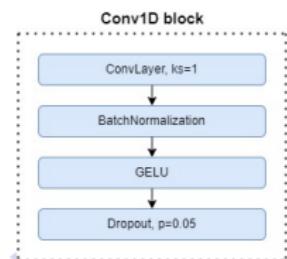
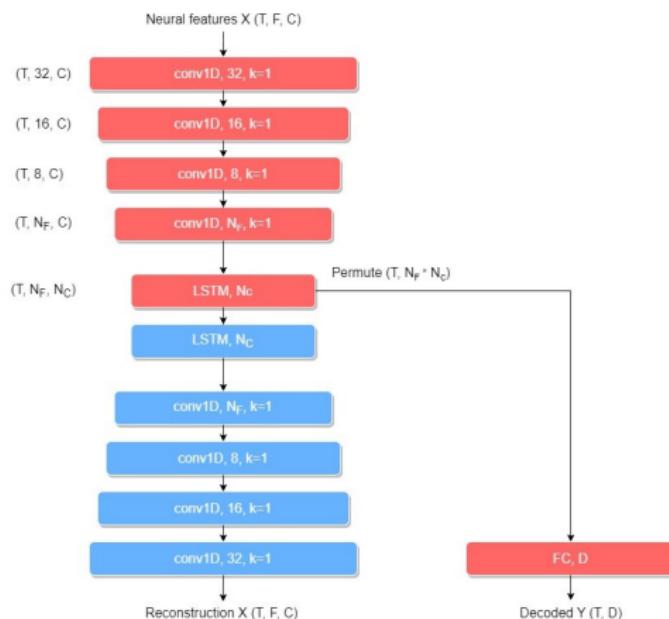
Функция потерь:

$$\mathcal{L}_1 = \mathcal{L}_{rec} + \alpha \cdot \mathcal{L}_{dec},$$

где  $\mathcal{L}_{rec}$  и  $\mathcal{L}_{dec}$ :

$$\mathcal{L}_{rec} = \frac{1}{M} \sum_{m=1}^M \|\underline{\mathbf{X}}_m - h \circ \psi(\underline{\mathbf{X}}_m)\|^2,$$

$$\mathcal{L}_{dec} = \frac{1}{M} \sum_{m=1}^M \|\mathbf{Y}_m - \omega \circ h(\underline{\mathbf{X}}_m)\|^2.$$



## Скрытые корреляции в BCI данных

Данные BCI имеют скрытые корреляции по временной и пространственной координате:

- **По времени:** измерения зависят от прошлых измерений как в сигналах с мозга, так и целевых временных рядах.
- **По пространству:** имеется зависимость между временными рядами, полученными с соседних электродов.

Выявить скрытые корреляции в BCI данных может помочь ганкелизации из-за связи ее со свертками.

## Ганкелизация

Дано

Пусть дан временной ряд  $\mathbf{x} = (x_1, \dots, x_I)^\top \in \mathbb{R}^I$ . Ганкелизация  $\mathbf{x}$  с параметром  $\tau$  задается

$$\mathcal{H}_\tau(\mathbf{x}) = \begin{pmatrix} x_1 & x_2 & \dots & x_{I-\tau+1} \\ x_2 & x_3 & \dots & x_{I-\tau+2} \\ \vdots & \vdots & \ddots & \vdots \\ x_\tau & x_{\tau+1} & \dots & x_I \end{pmatrix} \in \mathbb{R}^{\tau \times (I-\tau+1)},$$

где  $\mathcal{H}_\tau(\mathbf{x})$  является Ганкелевой матрицей вектора  $\mathbf{x}$ .

1D convolution

Рассмотрим одномерную свертку для вектора  $\mathbf{x} \in \mathbb{R}^I$  и ядра  $h \in \mathbb{R}^k$ , тогда 1D свертка может быть представлена в виде

$$\mathbf{x} * h = \begin{pmatrix} x_1 & x_2 & \dots & x_k \\ x_2 & x_3 & \dots & x_{k+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{I-k+1} & x_{I-k+2} & \dots & x_I \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \\ \vdots \\ h_k \end{pmatrix} = \mathcal{H}_{I-k+1}(\mathbf{x})h.$$

## Ганкелизация

### Полезные свойства Ганкелизации

- Связь ганкелизации со сверткой делает ганкелизацию полезный инструмент для выявления скрытых корреляций в ВСІ данных.
- Ганкелизация – это эффективный способ преобразования данных из более низкого порядка в более высокий порядок.

### Ганкелизация для тензоров (MDT)

Прямой MDT для  $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times \dots \times I_D}$  определяется следующим образом:

$$\hat{\underline{\mathbf{X}}} = \mathcal{H}_\tau(\underline{\mathbf{X}}) = \text{Fold}_{(I, \tau)}(\underline{\mathbf{X}} \times_1 \mathbf{S}_1 \dots \times_D \mathbf{S}_D) \in \mathbb{R}^{\tilde{I}_1 \times \dots \times \tilde{I}_G} \quad D < G,$$

т.е. строит блочный Ганкелевый тензор более высокого порядка  $\hat{\underline{\mathbf{X}}}$  из входного  $\underline{\mathbf{X}}$ , где

- $\mathbf{S}_d \in \mathbb{R}^{\tau_d \cdot (I_d - \tau_d + 1) \times I_d}, \quad d = 1, \dots, D$
- $\text{Fold}_{(I, \tau)} : \mathbb{R}^{\tau_1 \cdot (I_1 - \tau_1 + 1) \times \dots \times \tau_D \cdot (I_D - \tau_D + 1)} \rightarrow \mathbb{R}^{\tau_1 \times (I_1 - \tau_1 + 1) \times \dots \times \tau_D \times (I_D - \tau_D + 1)}$

Обратный MDT для  $\hat{\underline{\mathbf{X}}} \in \mathbb{R}^{\tilde{I}_1 \times \dots \times \tilde{I}_G}$  определяется следующим образом:

$$\underline{\mathbf{X}} = \mathcal{H}_\tau^{-1}(\hat{\underline{\mathbf{X}}}) = \text{Unfold}_{(I, \tau)}(\hat{\underline{\mathbf{X}}}) \times_1 \mathbf{S}_1^\dagger \dots \times_N \mathbf{S}_N^\dagger.$$

## Основной алгоритм

Дано

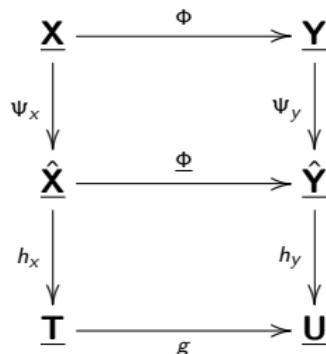
Пусть дана выборка  $(\underline{\mathbf{X}}, \underline{\mathbf{Y}})$ , где  $\underline{\mathbf{X}} \in \mathbb{R}^{M \times I_1 \times \dots \times I_{D_x}}$  – тензор независимых переменных,  $\underline{\mathbf{Y}} \in \mathbb{R}^{M \times J_1 \times \dots \times J_{D_y}}$  – тензор целевых переменных.

Цель

Основная цель работы состоит в том, чтобы найти набор моделей, таких как:

$$\Phi = \Psi_x \circ h_x \circ g \circ h_y^{-1} \circ \Psi_y^{-1} : \mathbb{R}^{I_1 \times \dots \times I_{D_x}} \rightarrow \mathbb{R}^{J_1 \times \dots \times J_{D_y}},$$

где  $\Psi_x, h_x, g, h_y$  и  $\Psi_x$ :



# Основной алгоритм

## Модели основного алгоритма

- Ганкелизация  $\Psi : \mathbb{R}^{I_1 \times \dots \times I_D} \rightarrow \mathbb{R}^{\tilde{I}_1 \times \dots \times \tilde{I}_G}$ , где  $D < G$ ;
- $h : \mathbb{R}^{\tilde{I}_1 \times \dots \times \tilde{I}_G} \rightarrow \mathbb{R}^{L_1 \times \dots \times L_G}$ , где  $L_i < I_i$  называется моделью снижения размерности, которая уменьшает размерность с  $\mathbb{R}^{\tilde{I}_1 \times \dots \times \tilde{I}_G}$  до  $\mathbb{R}^{L_1 \times \dots \times L_G}$ , а именно  $\underline{T}_m = h_x(\underline{X}_m, \theta_y)$ , где  $\theta$  – параметры модели;
- $g : \mathbb{R}^{L_1 \times \dots \times L_{G_x}} \rightarrow \mathbb{R}^{K_1 \times \dots \times K_{G_y}}$  называется моделью прогнозирования в скрытом латентом пространстве  $\underline{T}_m \in \mathbb{R}^{L_1 \times \dots \times L_{G_x}}$  до  $\underline{U}_m \in \mathbb{R}^{K_1 \times \dots \times K_{G_y}}$ , а именно  $\underline{U}_m = g(\underline{T}_m, \eta)$ , где  $\eta$  – параметры модели;
- В данной работе используются следующие модели снижения размерности пространства: HOPLS, MPCA, Autoencoder ReducedNet.

## Модель

Модель  $\Phi$  называется оптимальной, если минимизируется функционал ошибки  $\mathcal{L}$ :

$$\Phi^* = \arg \min_{\{\theta, \eta, \zeta\}} \mathcal{L}(\Phi(\underline{X}, \theta, \eta, \zeta), \underline{Y}),$$

где  $\theta$  – параметры модели снижения размерности,  $\eta$  – это параметры модели прогнозирования и  $\zeta$  – тип ганкелизации.

## Критерии качества

### Нормированное RMSE

Качество предсказания:

$$sRMSE(\underline{\mathbf{Y}}, \hat{\underline{\mathbf{Y}}}_p) = \sqrt{\frac{MSE(\underline{\mathbf{Y}}, \hat{\underline{\mathbf{Y}}}_p)}{MSE(\underline{\mathbf{Y}}, \bar{\underline{\mathbf{Y}}})}} = \frac{\|\underline{\mathbf{Y}} - \hat{\underline{\mathbf{Y}}}_p\|_2}{\|\underline{\mathbf{Y}} - \bar{\underline{\mathbf{Y}}}\|_2},$$

где  $\hat{\underline{\mathbf{Y}}}_p = \underline{\mathbf{X}}_p \underline{\Theta}_p^T$  – это предсказание модели, а  $\bar{\underline{\mathbf{Y}}}$  является константным прогнозом.

### Количество элементов тензора в латентном пространстве

Для тензора в латентном пространстве  $\hat{\underline{\mathbf{X}}} \in \mathbb{R}^{L_1 \times \dots \times L_G}$  количество его элементов,  $N_{values}$ , оценивается как:

$$N_{values} = L_1 \cdot \dots \cdot L_G.$$

## Данные для вычислительного эксперимента

### Данные Neurotycho

Данные были сняты с мозга обезьян с помощью 32 и 64 канальных ЭКоГ систем. Целевая переменная – траектория движения руки.

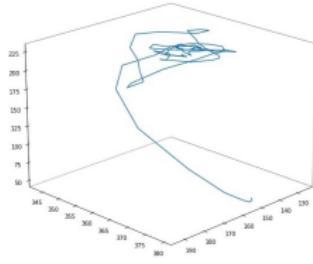
**Без ганкелизации:**  $\underline{\mathbf{X}} \in \mathbb{R}^{T \times 32 \times 27}$ ,  $\underline{\mathbf{Y}} \in \mathbb{R}^{T \times 3}$

**С ганкелизацией по времени:**  $\underline{\mathbf{X}} \in \mathbb{R}^{T' \times 10 \times 32 \times 27}$ ,  $\underline{\mathbf{Y}} \in \mathbb{R}^{T' \times 10 \times 3}$

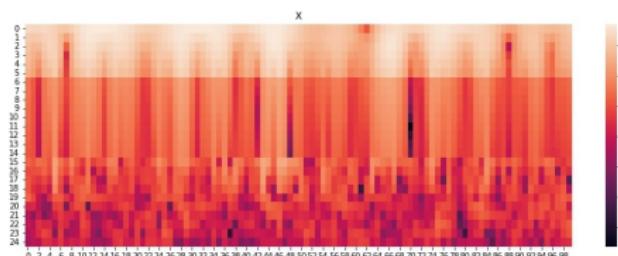
**С ганкелизацией по пространству:**  $\underline{\mathbf{X}} \in \mathbb{R}^{T \times 31 \times 2 \times 27}$ ,  $\underline{\mathbf{Y}} \in \mathbb{R}^{T \times 3}$

**С ганкелизацией по обеим координатам:**

$\underline{\mathbf{X}} \in \mathbb{R}^{T' \times 10 \times 31 \times 2 \times 27}$ ,  $\underline{\mathbf{Y}} \in \mathbb{R}^{T' \times 10 \times 3}$



Траектория  
движения руки



Данные одного канала в  
частотно-временных координатах

## Данные для вычислительного эксперимента

### Данные EEG

Данные были сняты с мозга человека с помощью 60 канальной ЭЭГ системы. Целевая переменная – 6-канальные ЭМГ данные с руки.

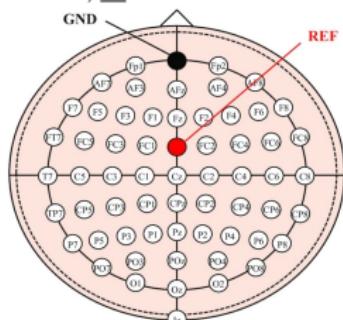
**Без ганкелизации:**  $\underline{X} \in \mathbb{R}^{T \times 60 \times 24}$ ,  $\underline{Y} \in \mathbb{R}^{T \times 6}$

**С ганкелизацией по времени:**  $\underline{X} \in \mathbb{R}^{T' \times 10 \times 60 \times 24}$ ,  $\underline{Y} \in \mathbb{R}^{T' \times 10 \times 6}$

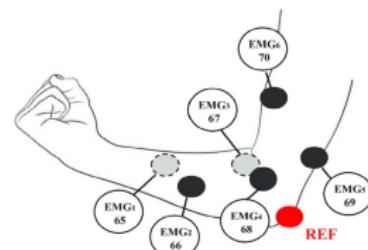
**С ганкелизацией по пространству:**  $\underline{X} \in \mathbb{R}^{T \times 59 \times 2 \times 24}$ ,  $\underline{Y} \in \mathbb{R}^{T \times 6}$

**С ганкелизацией по обеим координатам:**

$\underline{X} \in \mathbb{R}^{T' \times 10 \times 59 \times 2 \times 24}$ ,  $\underline{Y} \in \mathbb{R}^{T' \times 10 \times 6}$

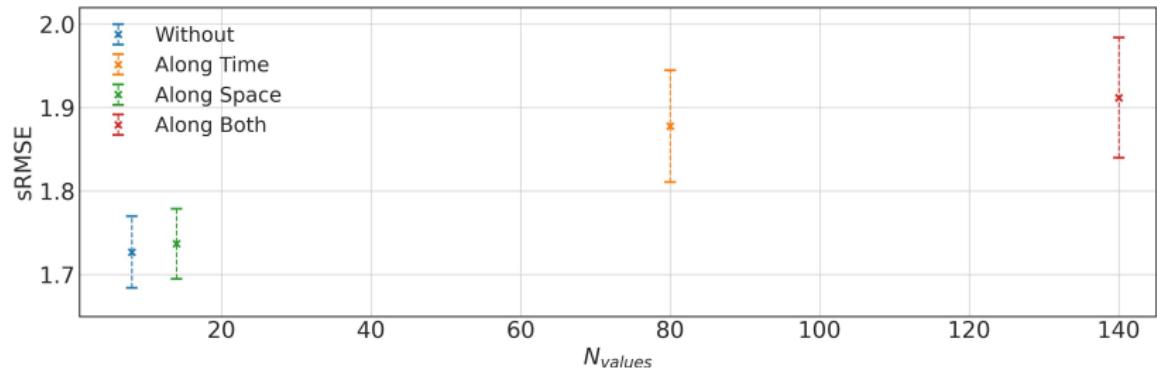


EEG электроды

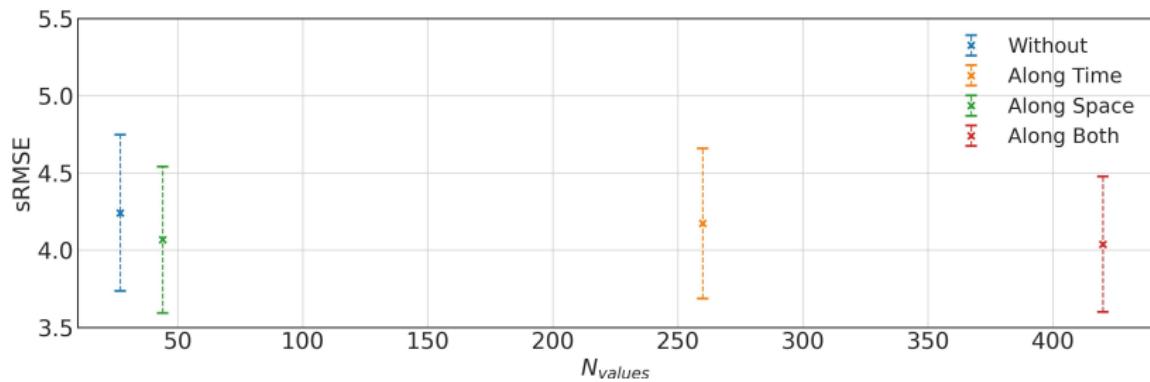


EMG electrodes positions

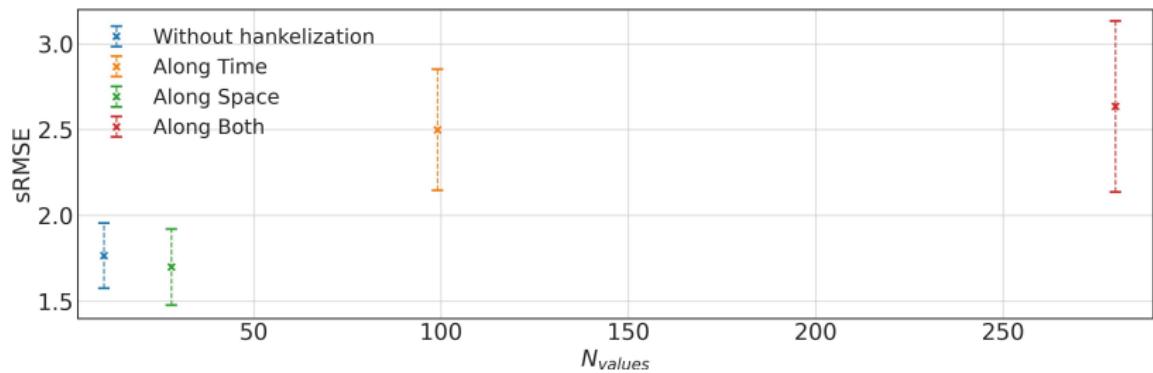
## Результаты для NeuroTycho (ch=32) с MPCA



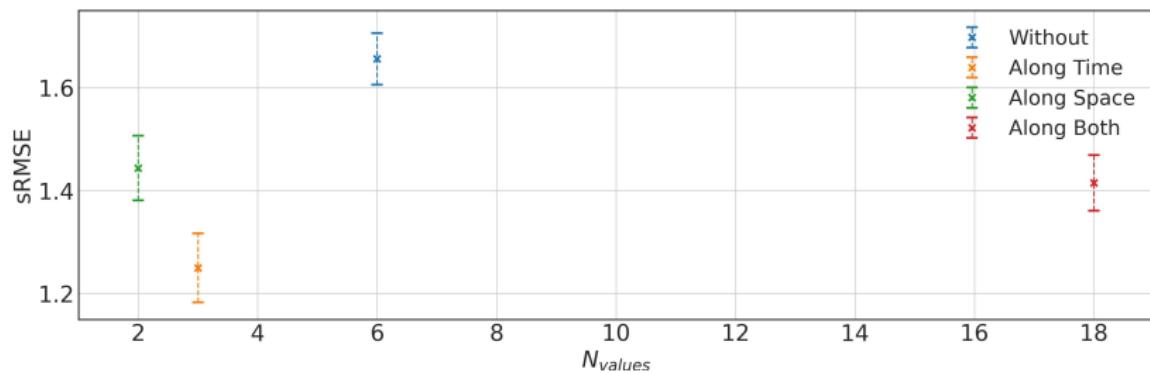
## Результаты для NeuroTycho (ch=64) с MPCA



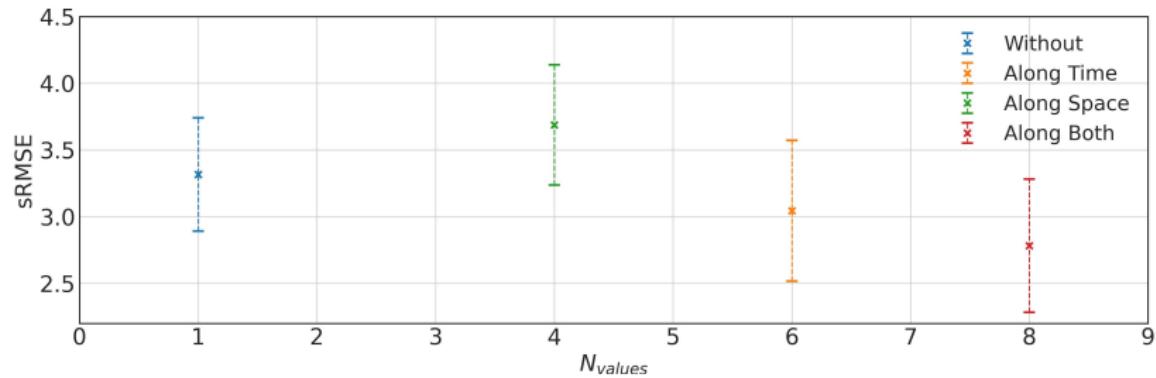
## Результаты для EEG (ch=60) с MPCA



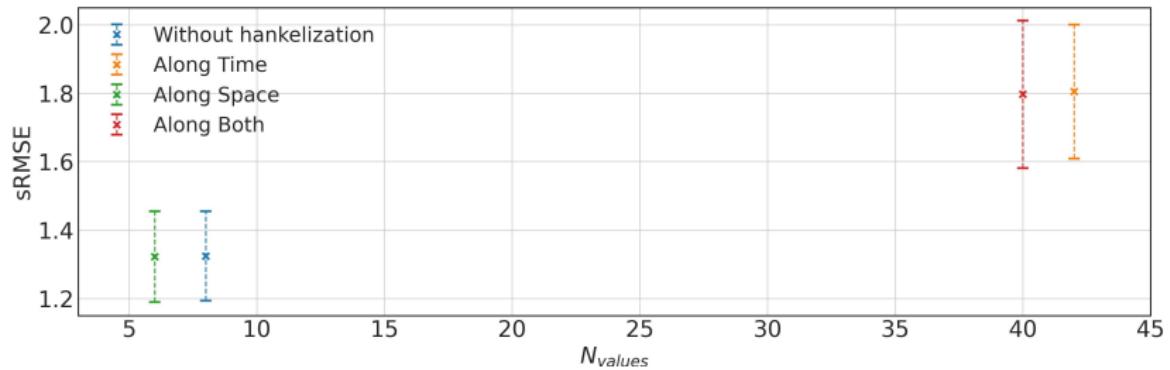
## Результаты для NeuroTycho (ch=32) с HOPLS



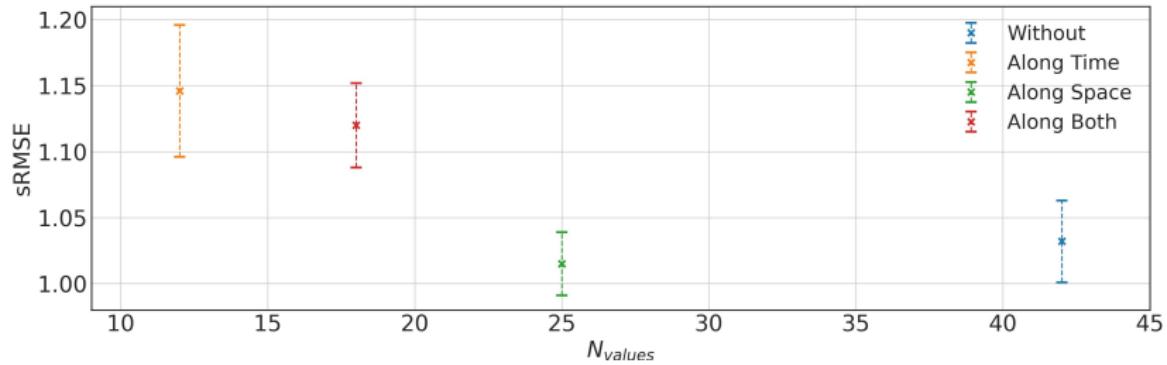
## Результаты для NeuroTycho (ch=64) с HOPLS



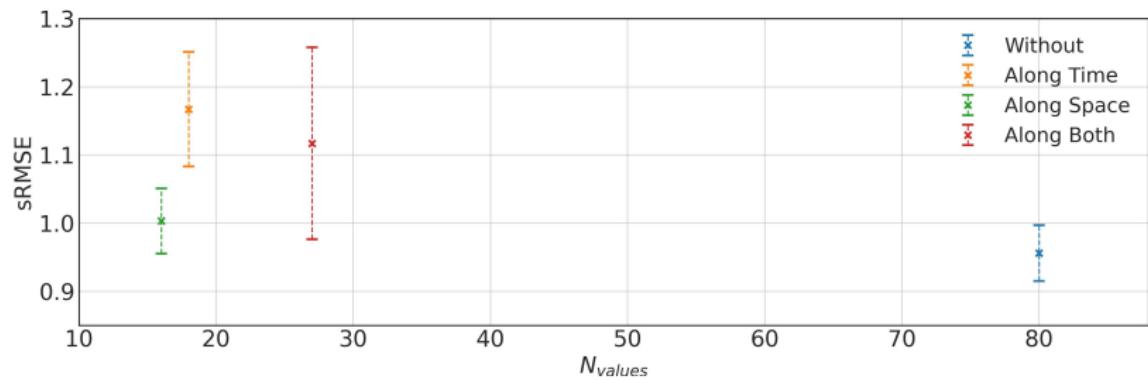
## Результаты для EEG (ch=60) с HOPLS



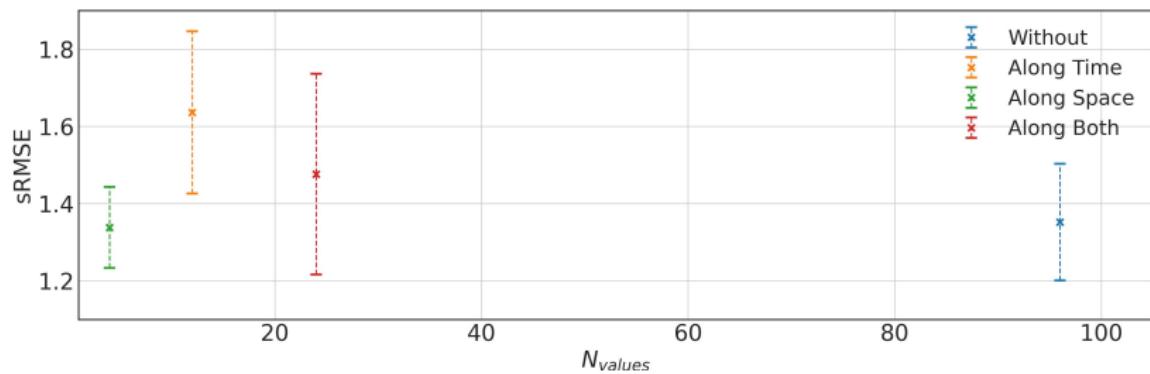
## Результаты для NeuroTycho (ch=32) с AutoEncoder



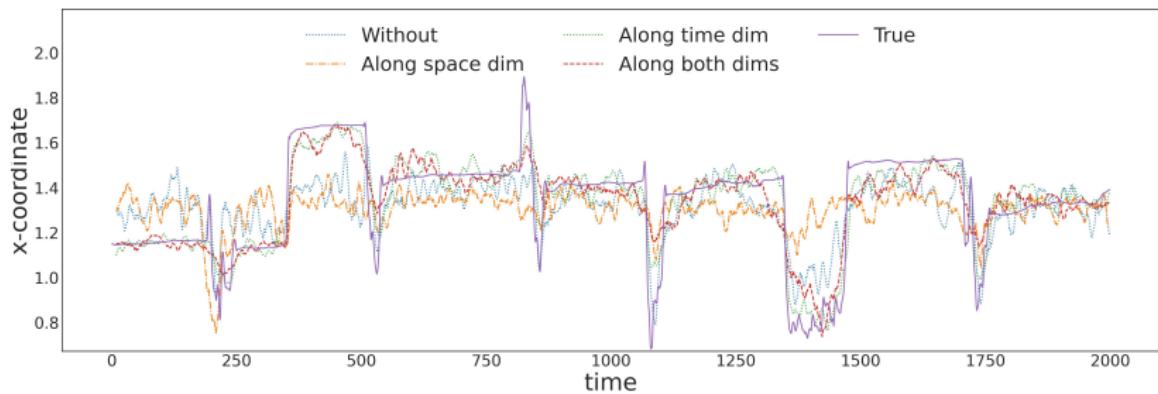
## Результаты для NeuroTycho (ch=64) с AutoEncoder



## Результаты для EEG (ch=60) с AutoEncoder



## Пример прогнозов для AutoEncoder



## Результаты, выносимые на защиту

- Исследована задача декодирования сигналов в пространствах высокой размерности.
- Исследованы различные типы ганкелизации для учета скрытых корреляций для задачи нейрокомпьютерных интерфейсов.
- Предложенная комбинация ганкелизации с различными методами снижения размерности показала преимущество использования ганкелизации для задачи нейрокомпьютерных интерфейсов.
- Показана важность ганкелизации вдоль пространственной координаты, так как она дает улучшение в точности предсказаний.

## Notation

### Тензор

Под тензором мы понимаем многомерный массив чисел. Тензор будем обозначать так:

$$\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N},$$

где  $N$  называется размерностью тензора.

Элементы тензора  $\underline{\mathbf{X}}$  в позиции  $(i_1, \dots, i_d)$  будем обозначать так:

$$x_{i_1, \dots, i_d} = \underline{\mathbf{X}}_{i_1, \dots, i_d}.$$

- индекс  $i_k$  соответствует  $k$ -му направлению или  $k$ -ой моде;
- индекс  $i_k$  принимает значение от  $1 \leq i_k \leq I_k$ , где  $I_k$  называется размером  $k$ -ой моды.

### Скалярное произведение

Определим скалярное произведение двух тензоров  $\underline{\mathbf{X}}$  и  $\underline{\mathbf{Y}}$ :

$$\langle \underline{\mathbf{X}}, \underline{\mathbf{Y}} \rangle = \sum_{i_1, \dots, i_N} \underline{\mathbf{X}}(i_1, i_2, \dots, i_N) \underline{\mathbf{Y}}(i_1, i_2, \dots, i_N) = \sum_{i_1, i_2, \dots, i_N} x_{i_1, i_2, \dots, i_N} y_{i_1, i_2, \dots, i_N}.$$



## Notation

### Матрицизация

Определим  $k$ -развертку тензора  $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times \dots \times I_N}$  следующим образом:

$$\mathbf{X}_{(k)} \in \mathbb{R}^{I_k \times I_1 \cdot \dots \cdot I_{k-1} \cdot I_{k+1} \cdot \dots \cdot I_N},$$

со  $I_k$  строками и  $I_1 \cdot I_2 \cdot \dots \cdot I_{k-1} \cdot I_{k+1} \cdot \dots \cdot I_N$  столбцами.

### Произведение тензора на матрицу по $k$ -й моде

Определим произведение тензора  $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times \dots \times I_N}$  и матрицы  $\mathbf{U} \in \mathbb{R}^{J \times I_k}$  вдоль направления  $k$ :

$$\underline{\mathbf{Y}} = \underline{\mathbf{X}} \times_k \mathbf{U} \in \mathbb{R}^{I_1 \times \dots \times I_{k-1} \times J \times I_{k+1} \times \dots \times I_N}.$$

Можно записать поэлементно следующим образом:

$$c_{i_1 \dots i_{k-1} j i_{k+1} \dots i_N} = \sum_{i_n} x_{i_1 i_2, \dots, i_N} u_{j i_k}.$$

Эквивалентная матричная форма:

$$\mathbf{C}_{(k)} = \mathbf{U} \mathbf{X}_{(k)} \in \mathbb{R}^{I_k \times I_1 \cdot \dots \cdot I_{k-1} \cdot I_{k+1} \cdot \dots \cdot I_N}.$$

## Notation

### Таккеровское разложение

Тензор  $\underline{\mathbf{X}}$  представляется в формате Таккера, если он может быть записан в следующем виде:

$$\underline{\mathbf{X}}(i_1, \dots, i_N) \approx \sum_{\alpha_1, \dots, \alpha_N} \underline{\mathbf{G}}(\alpha_1, \dots, \alpha_N) \mathbf{U}^{(1)}(i_1, \alpha_1) \dots \mathbf{U}^{(N)}(i_N, \alpha_N),$$

где  $1 \leq \alpha_k \leq R_k$  или в другой форме:

$$\underline{\mathbf{X}} \approx \underline{\mathbf{G}} \times_1 \mathbf{U}^{(1)} \dots \times_N \mathbf{U}^{(N)},$$

где  $\times_k$  означает произведение тензора на матрицу по  $k$ -й моде:

$$(\underline{\mathbf{G}} \times_k \mathbf{U}^{(k)})(\alpha_1, \dots, i_k, \dots, \alpha_N) = \sum_{i_k=1}^{R_k} \underline{\mathbf{G}}(\alpha_1, \dots, \alpha_k, \dots, \alpha_N) \mathbf{U}^{(k)}(i_k, \alpha_k).$$

- Числа  $R_1, R_2, \dots, R_N$  называются Таккеровскими рангами тензора;
- $\underline{\mathbf{G}} \in \mathbb{R}^{R_1 \times R_2 \times \dots \times R_N}$  – ядро разложения Таккера;
- Матрицы  $\mathbf{U}^{(k)} \in \mathbb{R}^{I_k \times R_k}$  называются Таккеровскими факторами.

# Principal component analysis (PCA)

Дано

Пусть дана выборка  $(\mathbf{X}, \mathbf{Y})$ , где  $\mathbf{X} \in \mathbb{R}^{m \times n}$  – матрица независимых переменных,  $\mathbf{Y} \in \mathbb{R}^{m \times r}$  – матрица целевых переменных.

В основе метода РСА лежат следующие принципы

- снижение размерности данных, сохраняющее максимальную дисперсию.
- ортогональное линейное преобразование исходного признакового пространства  $\mathbb{R}^{m \times n}$  в новое пространство меньшей размерности  $\mathbb{R}^{m \times p}$ , где  $p < n$ .

$$\tilde{\mathbf{X}} = \mathbf{X}\mathbf{Q}^\top \in \mathbb{R}^{m \times p},$$

где  $\mathbf{Q} = [\mathbf{q}_1, \dots, \mathbf{q}_p]$  – базисные векторы вдоль которых выборочная дисперсия исходных данных максимальна.

## Partial Least Squares (PLS)

Дано

Пусть дана выборка  $(\mathbf{X}, \mathbf{Y})$ , где  $\mathbf{X} \in \mathbb{R}^{m \times n}$  – матрица независимых переменных,  $\mathbf{Y} \in \mathbb{R}^{m \times r}$  – матрица целевых переменных.

В основе метода PLS лежат следующие принципы:

- спроектировать исходные матрицы  $\mathbf{X}$  и  $\mathbf{Y}$  в общее латентное пространство;
- максимизировать ковариацию между образами с целью сохранения исходной информации об матрицах  $\mathbf{X}$  и  $\mathbf{Y}$ .

$$\mathbf{X}_{m \times n} = \mathbf{T}_{m \times \ell} \cdot \mathbf{P}_{\ell \times n}^T + \mathbf{F}_{m \times n} = \sum_{k=1}^{\ell} \mathbf{t}_k \mathbf{p}_k^T + \mathbf{F},$$

$$\mathbf{Y}_{m \times r} = \mathbf{U}_{m \times \ell} \cdot \mathbf{Q}_{\ell \times r}^T + \mathbf{E}_{m \times r} = \sum_{k=1}^{\ell} \mathbf{u}_k \mathbf{q}_k^T + \mathbf{E},$$

$$\mathbf{U} = \mathbf{T}\mathbf{D}, \quad \mathbf{D} = \text{diag}(d_k), \quad d_k = \mathbf{u}_k^T \mathbf{t}_k / \|\mathbf{t}_k\|_2^2.$$