

# Dimensionality reduction for time series decoding and forecasting problems\*

R. V. Isachenko<sup>1</sup>, M. R. Vladimirova<sup>2</sup>, V. V. Strijov<sup>3</sup>

**Abstract:** (Решается задача обнаружения зависимостей в прогнозируемой переменной.) The paper is devoted to the problem of detecting the relation between independent and target variables. The problem is to detect the relation between independent and target variables. (Вместо построения прогноза одного момента времени предлагается прогнозировать многомерный вектор прогноза.) We propose to predict a multidimensional target vector instead of predicting one timestamp point. (Рассматривается линейная модель метода частных наименьших квадратов.) We consider the linear model of partial least squares (PLS). (Описан метод нахождения матрицы совместного описания для исходных матриц объектов и ответов. Strange sentence:) The method finds the matrix of a joint description for the design matrix and the outcome matrix. (Найденное описание является низкоразмерным, что позволяет построить устойчивую, простую модель.) The description is low-dimensional and allows to build a simple, stable model. (Проводится вычислительный эксперимент на реальных данных объемов потребления электроэнергии и данных сигналов кортикограмм.) We conducted computational experiments on the real data of energy consumption and electrocorticograms signals (ECoG).

**Keywords:** time series decoding, forecast, partial least squares, dimensionality reduction

## 1 Introduction

(В работе рассматривается задача восстановления зависимости между независимой и прогнозируемой переменными.) Paper investigates the problem of dependence recovering between an input data and a model outcome. (Предлагаемая модель позволяет восстанавливать зависимость в случае многомерной прогнозируемой переменной.) The

---

\*The work was financially supported by the Russian Foundation for Basic Research (project 16-07-01155).

<sup>1</sup>Moscow Institute of Physics and Technology, isa-ro@yandex.ru

<sup>2</sup>Moscow Institute of Physics and Technology, vladimirova.maria@phystech.edu

<sup>3</sup>A. A. Dorodnicyn Computing Centre, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, strijov@ccas.ru

6 proposed model is suitable for predicting a multidimensional target variable. (В случае за-  
7 дачи прогнозирования пространства объектов и ответов имеют одну и ту же природу.)  
8 In the case of the forecasting problem the object and target spaces have the same nature.  
9 (Для построения модели по исходному временному ряду строятся авторегрессионные  
10 матрицы объектов и ответов.) To build the model we need to construct autoregressive  
11 matrices for input objects and target variables. (Объектом является локальная история  
12 сигнала, ответом — значение сигнала в следующие моменты времени.) The object is a  
13 local signal history, the outcome is signal values in the next timestamps. (Авторегресси-  
14 онная модель предполагает, что значение сигнала в данный момент времени линейно  
15 зависит от предыдущих значений этого же сигнала. ) An autoregressive model makes a  
16 consumption that the current signal values depend linearly on the previous signal values.

17 (В задаче декодирования временных рядов пространство объектов имеет значи-  
18 мо большую размерность.) ( В этом случае пространства объектов и ответов имеют  
19 разную природу, ответом является отклик системы на входной сигнал.) In the case  
20 of time series decoding problem the object and target spaces are different in nature, the  
21 outcome is a system response to the input signal. (Авторегрессионная матрица объектов  
22 содержит локальную историю входного сигнала, а авторегрессионная матрица отве-  
23 тов содержит локальную историю отклика.) The autoregressive design matrix contains  
24 local history of the input signal. The autoregressive target matrix contains the local history  
25 of the response.

26 (Исходное признаковое пространство является многомерным.) The object space in  
27 time series decoding problems is high dimensional. (Избыточная размерность признако-  
28 вого пространства приводит к неустойчивости модели.) Excessive dimensionality of the  
29 feature description leads to instability of the model. (Для решения этой проблемы ис-  
30 пользуются методы отбора признаков [1, 2].) To solve this problem the feature selection  
31 procedures are used [1, 2].

32 (В работе исследуется метод частных наименьших квадратов (PLS) [3–5].) The  
33 paper considers the partial least squares regression (PLS) model [3–5]. (Метод частных  
34 наименьших квадратов снижает размерности матрицы признаков и выделяет линей-  
35 ные комбинации признаков, которые оказывают наибольшее влияние на вектор отве-  
36 тов. ) The PLS model reduces the dimension of the input data and extracts the linear  
37 combination of features which have the greatest impact on the response vector. (Выделе-  
38 ние признаков происходит итеративно, в порядке уменьшения их влияния на вектор  
39 ответов. ) Feature extraction is an iterative process in order of decreasing the influence on  
40 the response vector. (Рассматриваются только значимые комбинации признаков, незна-  
41 чительно потеряв в точности прогноза. Методы PLS регрессии подробно описаны в  
42 работах [6–8]. ) PLS regression methods are described in detail in [6–8]. (Разницу меж-  
43 ду методом PLS и связанными с ним подходами, различные разновидности регрессии  
44 PLS можно найти в [9].) The difference between various PLS approaches, different kinds  
45 of the PLS regression could be found in [9].

46 (Современное состояние области и обзор нелинейных модификаций метода PLS  
47 описаны [10].) The current state of the field and the overview of nonlinear PLS method  
48 modifications are described in [10]. Нелинейное расширение метода PLS регрессии впер-

49 вые введено в [11]. A nonlinear PLS method extension was introduced in [11]. (В литера-  
50 туре были разработаны различные модификации PLS. ) There has been developed the  
51 variety of PLS modifications. (Предложены нелинейные методы PLS, основанные на  
52 различных моделях: сглаживающих сплайнах [12], нейронных сетях [13], базисных  
53 радиальных функциях [14], генетическом алгоритме [15]. ) The proposed nonlinear  
54 PLS methods are based on smoothing splines [12], neural networks [13], radial basis  
55 functions [14], genetic algorithms [15].

56 (Результатом отбора признаков является снижение размерности задачи и повы-  
57 шение устойчивости моделей без существенной потери точности прогноза.) The result  
58 of the feature selection is the dimensionality reduction and the increasing model stability  
59 without significant loss of the prediction quality. (Метод решения задачи применяется к  
60 двум наборам данных, имеющих избыточную информацию. ) The proposed method is  
61 used on two datasets with the redundant input and target spaces. (Первый набор дан-  
62 ных — почасовые временные ряды объёмов потребления электроэнергии. Временные  
63 ряды соответствуют периоду 1999-2004 годов. ) The first dataset consists of hourly time  
64 series of energy consumption. Time series were collected in Poland from 1999 to 2004.

65 (Второй набор данных взят из проекта NeuroTycho [16], в котором проектиру-  
66 ется нейрокомпьютерный интерфейс (BCI) [17, 18] для обмена информацией между  
67 мозгом и электронным устройством. ) The second dataset comes from the NeuroTycho  
68 project [16] that designs Brain-Computer Interface (BCI) [17, 18] for information transmitting  
69 between brains and electronic devices. (Система BCI повышает умственные и физиче-  
70 ские способности пользователя, обеспечивая прямую связь между мозгом и компьюте-  
71 ром [19]. ) Brain-Computer Interface (BCI) system enhances its user's mental and physical  
72 abilities, providing a direct communication mean between the brain and a computer [19].  
73 (BCI восстанавливает поврежденную функциональность пациентов с нарушениями  
74 двигательного или когнитивного развития. ) BCIs aim at restoring damaged functionality  
75 of motorically or cognitively impaired patients. (Цель анализа моторных изображений  
76 заключается в распознавании предполагаемых движений по записанной активности  
77 мозга. ) The goal of motor imagery analysis is to recognize intended movements from  
78 the recorded brain activity. (Существуют различные методы измерения кортикальных  
79 данных для BCI описанные в [20, 21]. В работе рассматриваются сигналы электро-  
80 кортикографии (ECoG) [22]. ) While there are various techniques for measuring cortical  
81 data for BCI [20, 21], we concentrate on the ElectroCorticoGraphic (ECoG) signals [22].  
82 (ECoG, как и другие инвазивные методы, обеспечивает более стабильную запись и  
83 лучшее разрешение в временных и пространственных областях, чем ее неинвазивные  
84 аналоги.) ECoG, as well as other invasive techniques, provides more stable recordings and  
85 better resolution in temporal and spatial domains than its non-invasive counterparts. (Дан-  
86 ные ECoG являются многомерными и измерения коррелируют как во временной, так  
87 и в пространственной областях. Задача состоит в предсказании траектории движения  
88 руки по временным рядам напряжения кортикальной активности. Сигналы ECoG из-  
89 меряются через 32 канала, в то время как субъект перемещает руку. ) We address  
90 the problem of continuous hand trajectory reconstruction. The subdural ECoG signals are  
91 measured across 32 channels as the subject is moving its hand. Once the ECoG signals

are transformed into informative features, the problem of trajectory reconstruction is an autoregression problem. Feature extraction involves application of some spectro-temporal transform to the ECoG signals from each channel [23].

(В работах по прогнозированию сложных пространственных временных рядов прогноз выполняется поточно.) In papers which are devoted to forecasting of complex spatial time series the forecast is built pointwise.[[сюда напрашиваются ссылки, вставить 2-3 штуки](#)] (При необходимости прогнозировать несколько точек одновременно задачу предлагается решать последовательно вычисляя прогноз по точкам. ) If one need to predict multiple points simultaneously, it is proposed to compute forecasted points sequentially . (При этом используются предыдущие спрогнозированные значения для получения последующих.) During this process the previous predicted values are used to obtain a subsequent ones. (Предлагаемый метод позволяет получать временной ряд прогнозируемых значений одновременно с учётом скрытых зависимостей не только в пространстве объектов, но и в пространстве ответов с согласованием этих двух зависимостей.) The proposed method allows to obtain multiple predicted time series values at the same time taking into account hidden dependencies not only in the object space, but also in the target space. (Предлагаемый метод позволил существенно снизить размерность признакового пространства. ) The proposed method significantly reduces the dimensionality of the feature space.

## 2 Problem statement

Задана выборка  $\mathcal{D} = (\mathbf{X}, \mathbf{Y})$ , где  $\mathbf{X} \in \mathbb{R}^{m \times n}$  — матрица объектов,  $\mathbf{Y} \in \mathbb{R}^{m \times r}$  — матрица ответов. Способ построения выборки под определенную прикладную задачу описан в разделе *Вычислительный эксперимент*.

Предполагается, что между объектами  $\mathbf{x} \in \mathbb{R}^n$  и ответами  $\mathbf{y} \in \mathbb{R}^r$  существует линейная зависимость

$$\underset{1 \times r}{\mathbf{y}} = \underset{1 \times n}{\mathbf{x}} \cdot \underset{n \times r}{\boldsymbol{\Theta}} + \underset{1 \times r}{\boldsymbol{\varepsilon}}, \quad (1)$$

где  $\boldsymbol{\Theta}$  — матрица параметров модели, а  $\boldsymbol{\varepsilon}$  — вектор регрессионных остатков.

Необходимо по известной выборке  $\mathcal{D}$  восстановить матрицу параметров модели (1). Оптимальные параметры находятся минимизацией функции ошибки. Введем квадратичную функцию ошибки  $S$  на выборке  $\mathcal{D}$ :

$$S(\boldsymbol{\Theta}|\mathcal{D}) = \left\| \underset{m \times n}{\mathbf{X}} \cdot \underset{n \times r}{\boldsymbol{\Theta}} - \underset{m \times r}{\mathbf{Y}} \right\|_2^2 = \sum_{i=1}^m \left\| \underset{1 \times n}{\mathbf{x}_i} \cdot \underset{n \times r}{\boldsymbol{\Theta}} - \underset{1 \times r}{\mathbf{y}_i} \right\|_2^2 \rightarrow \min_{\boldsymbol{\Theta}}. \quad (2)$$

Линейная зависимость столбцов матрицы  $\mathbf{X}$  приводит к неустойчивому решению задачи оптимизации (2). Для устранения линейной зависимости применяются методы отбора признаков.

### 3 Partial Least Squares method

Для устранения линейной зависимости и снижения размерности пространства применяется метод главных компонент PCA. Основным недостатком данного метода является то, что он не учитывает взаимосвязь между объектами и ответами. Метод частных наименьших квадратов PLS проецирует матрицу объектов  $\mathbf{X}$  и матрицу ответов  $\mathbf{Y}$  в латентное пространство  $\mathbb{R}^l$  меньшей размерности ( $l < r < n$ ). Алгоритм PLS находит в латентном пространстве матрицу  $\mathbf{T} \in \mathbb{R}^{m \times l}$ , наилучшим образом описывающую исходные матрицы  $\mathbf{X}$  и  $\mathbf{Y}$ .

Матрица объектов  $\mathbf{X}$  и матрица ответов  $\mathbf{Y}$  проецируются в латентное пространство следующим образом:

$$\mathbf{X}_{m \times n} = \mathbf{T}_{m \times l} \cdot \mathbf{P}_{l \times n}^T + \mathbf{F}_{m \times n} = \sum_{k=1}^l \mathbf{t}_k \cdot \mathbf{p}_k^T + \mathbf{F}_{m \times n}, \quad (3)$$

$$\mathbf{Y}_{m \times r} = \mathbf{T}_{m \times l} \cdot \mathbf{Q}_{l \times r}^T + \mathbf{E}_{m \times r} = \sum_{k=1}^l \mathbf{t}_k \cdot \mathbf{q}_k^T + \mathbf{E}_{m \times r}, \quad (4)$$

где  $\mathbf{T}$  — матрица совместного описания объектов и ответов в латентном пространстве, причём столбцы матрицы  $\mathbf{T}$  ортогональны;  $\mathbf{P}$ ,  $\mathbf{Q}$  — матрицы перехода из латентного пространства в исходные пространства;  $\mathbf{E}$ ,  $\mathbf{F}$  — матрицы невязок.

Псевдокод метода регрессии PLS приведен в алгоритме 1. Алгоритм итеративно на каждом из  $l$  шагов вычисляет по одному столбцу  $\mathbf{t}_k$ ,  $\mathbf{p}_k$ ,  $\mathbf{q}_k$  матриц  $\mathbf{T}$ ,  $\mathbf{P}$ ,  $\mathbf{Q}$  соответственно. После вычисления следующего набора векторов из матриц  $\mathbf{X}$ ,  $\mathbf{Y}$  вычитаются очередные одноранговые аппроксимации. Первым шагом необходимо произвести нормировку столбцов исходных матриц (вычесть среднее и разделить на стандартное отклонение). На этапе тестирования необходимо провести нормировку тестовых данных, вычислить предсказание модели 1, а затем провести обратную нормировку.

Вектора  $\mathbf{t}_k$  и  $\mathbf{u}_k$  из внутреннего цикла алгоритма 1 содержат информацию о матрице объектов  $\mathbf{X}$  и матрице ответов  $\mathbf{Y}$  соответственно. Блоки из шагов (6)-(7) и шагов (8)-(9) — аналоги алгоритма PCA для матриц  $\mathbf{X}$  и  $\mathbf{Y}$  [5]. Последовательное выполнение блоков позволяет учесть взаимную связь между матрицами  $\mathbf{X}$  и  $\mathbf{Y}$ .

Теоретическое обоснование алгоритма PLS следует из следующих утверждений.

**Утверждение 1.** *Наилучшее описание матриц  $\mathbf{X}$  и  $\mathbf{Y}$  с учётом их взаимосвязи достигается при максимизации ковариации между векторами  $\mathbf{t}_k$  и  $\mathbf{u}_k$ .*

Утверждение следует из равенства

$$\text{cov}(\mathbf{t}_k, \mathbf{u}_k) = \text{corr}(\mathbf{t}_k, \mathbf{u}_k) \cdot \sqrt{\text{var}(\mathbf{t}_k)} \cdot \sqrt{\text{var}(\mathbf{u}_k)}.$$

Максимизация дисперсий векторов  $\mathbf{t}_k$  и  $\mathbf{u}_k$  отвечает за сохранение информации об исходных матрицах, корреляция между векторами отвечает взаимосвязи между  $\mathbf{X}$  и  $\mathbf{Y}$ . ■

Во внутреннем цикле алгоритма вычисляются нормированные вектора весов  $\mathbf{w}_k$  и  $\mathbf{c}_k$ . Из данных векторов строятся матрицы весов  $\mathbf{W}$  и  $\mathbf{C}$  соответственно.

---

**Algorithm 1** PLSR algorithm
 

---

**Require:**  $\mathbf{X}, \mathbf{Y}, l$ ;

**Ensure:**  $\mathbf{T}, \mathbf{P}, \mathbf{Q}$ ;

```

1: normalize matrices  $\mathbf{X}$  и  $\mathbf{Y}$  by columns
2: initialize  $\mathbf{u}_0$  (the first column of  $\mathbf{Y}$ )
3:  $\mathbf{X}_1 = \mathbf{X}; \mathbf{Y}_1 = \mathbf{Y}$ 
4: for  $k = 1, \dots, l$  do
5:   repeat
6:      $\mathbf{w}_k := \mathbf{X}_k^\top \mathbf{u}_{k-1} / (\mathbf{u}_{k-1}^\top \mathbf{u}_{k-1}); \quad \mathbf{w}_k := \frac{\mathbf{w}_k}{\|\mathbf{w}_k\|}$ 
7:      $\mathbf{t}_k := \mathbf{X}_k \mathbf{w}_k$ 
8:      $\mathbf{c}_k := \mathbf{Y}_k^\top \mathbf{t}_k / (\mathbf{t}_k^\top \mathbf{t}_k); \quad \mathbf{c}_k := \frac{\mathbf{c}_k}{\|\mathbf{c}_k\|}$ 
9:      $\mathbf{u}_k := \mathbf{Y}_k \mathbf{c}_k$ 
10:   until  $\mathbf{t}_k$  stabilizes
11:    $\mathbf{p}_k := \mathbf{X}_k^\top \mathbf{t}_k / (\mathbf{t}_k^\top \mathbf{t}_k), \quad \mathbf{q}_k := \mathbf{Y}_k^\top \mathbf{t}_k / (\mathbf{t}_k^\top \mathbf{t}_k)$ 
12:    $\mathbf{X}_{k+1} := \mathbf{X}_k - \mathbf{t}_k \mathbf{p}_k^\top$ 
13:    $\mathbf{Y}_{k+1} := \mathbf{Y}_k - \mathbf{t}_k \mathbf{q}_k^\top$ 

```

---

155 **Утверждение 2.** В результате выполнения внутреннего цикла вектора  $\mathbf{w}_k$  и  $\mathbf{c}_k$   
 156 будут собственными векторами матриц  $\mathbf{X}_k^\top \mathbf{Y}_k \mathbf{Y}_k^\top \mathbf{X}_k$  и  $\mathbf{Y}_k^\top \mathbf{X}_k \mathbf{X}_k^\top \mathbf{Y}_k$ , соответству-  
 157 ющими максимальным собственным значениям.

$$\begin{aligned} \mathbf{w}_k &\propto \mathbf{X}_k^\top \mathbf{u}_{k-1} \propto \mathbf{X}_k^\top \mathbf{Y}_k \mathbf{c}_{k-1} \propto \mathbf{X}_k^\top \mathbf{Y}_k \mathbf{Y}_k^\top \mathbf{t}_{k-1} \propto \mathbf{X}_k^\top \mathbf{Y}_k \mathbf{Y}_k^\top \mathbf{X}_k \mathbf{w}_{k-1}, \\ \mathbf{c}_k &\propto \mathbf{Y}_k^\top \mathbf{t}_k \propto \mathbf{Y}_k^\top \mathbf{X}_k \mathbf{w}_k \propto \mathbf{Y}_k^\top \mathbf{X}_k \mathbf{X}_k^\top \mathbf{u}_{k-1} \propto \mathbf{Y}_k^\top \mathbf{X}_k \mathbf{X}_k^\top \mathbf{Y}_k \mathbf{c}_{k-1}, \end{aligned}$$

158 где символ  $\propto$  означает равенство с точностью до мультипликативной константы.

159

160 Утверждение следует из того факта, что правила обновления векторов  $\mathbf{w}_k, \mathbf{c}_k$  сов-  
 161 падают с итерацией алгоритма поиска максимального собственного значения. Данный  
 162 алгоритм основан на следующем факте.

163 Если матрица  $\mathbf{A}$  диагонализуема,  $\mathbf{x}$  — некоторый вектор, то

$$\lim_{k \rightarrow \infty} \mathbf{A}^k \mathbf{x} = \lambda_{\max}(\mathbf{A}) \cdot \mathbf{v}_{\max},$$

164 где  $\lambda_{\max}(\mathbf{A})$  — максимальное собственное значение матрицы  $\mathbf{A}$ ,  $\mathbf{v}_{\max}$  — собственный  
 165 вектор матрицы  $\mathbf{A}$ , соответствующий  $\lambda_{\max}(\mathbf{A})$ . ■

166 **Утверждение 3.** Обновление векторов по шагам (6)–(9) алгоритма 1 соответству-  
 167 ет максимизации ковариации между векторами  $\mathbf{t}_k$  и  $\mathbf{u}_k$ .

Максимальная ковариация между векторами  $\mathbf{t}_k$  и  $\mathbf{u}_k$  равна максимальному собственному значению матрицы  $\mathbf{X}_k^\top \mathbf{Y}_k \mathbf{Y}_k^\top \mathbf{X}_k$ : The maximum covariance between the vectors  $\mathbf{t}_k$  and  $\mathbf{u}_k$  is equal to the maximum eigenvalue of the matrix  $\mathbf{X}_k^\top \mathbf{Y}_k \mathbf{Y}_k^\top \mathbf{X}_k$ :

$$\begin{aligned} \max_{\mathbf{t}_k, \mathbf{u}_k} \text{cov}(\mathbf{t}_k, \mathbf{u}_k)^2 &= \max_{\substack{\|\mathbf{w}_k\|=1 \\ \|\mathbf{c}_k\|=1}} \text{cov}(\mathbf{X}_k \mathbf{w}_k, \mathbf{Y}_k \mathbf{c}_k)^2 = \max_{\substack{\|\mathbf{w}_k\|=1 \\ \|\mathbf{c}_k\|=1}} \text{cov}\left(\mathbf{c}_k^\top \mathbf{Y}_k^\top \mathbf{X}_k \mathbf{w}_k\right)^2 = \\ &= \max_{\|\mathbf{w}_k\|=1} \text{cov}\left\|\mathbf{Y}_k^\top \mathbf{X}_k \mathbf{w}_k\right\|^2 = \max_{\|\mathbf{w}_k\|=1} \mathbf{w}_k^\top \mathbf{X}_k^\top \mathbf{Y}_k \mathbf{Y}_k^\top \mathbf{X}_k \mathbf{w}_k = \\ &= \lambda_{\max}\left(\mathbf{X}_k^\top \mathbf{Y}_k \mathbf{Y}_k^\top \mathbf{X}_k\right), \end{aligned}$$

где  $\lambda_{\max}(\mathbf{A})$  — максимальное собственное значение матрицы  $\mathbf{A}$ . Применяя утверждение 2, получаем требуемое. where  $\lambda_{\max}(\mathbf{A})$  is the maximum eigenvalue of the matrix  $\mathbf{A}$ . Applying the statement 2, we obtain the required result. ■

После завершения внутреннего цикла на шаге (11) вычисляются вектора  $\mathbf{p}_k$ ,  $\mathbf{q}_k$  проецированием столбцов матриц  $\mathbf{X}_k$  и  $\mathbf{Y}_k$  на вектор  $\mathbf{t}_k$ . Для перехода на следующий шаг необходимо вычесть из матриц  $\mathbf{X}_k$  и  $\mathbf{Y}_k$  одноранговые аппроксимации  $\mathbf{t}_k \mathbf{p}_k^\top$  и  $\mathbf{t}_k \mathbf{q}_k^\top$

$$\begin{aligned} \mathbf{X}_{k+1} &= \mathbf{X}_k - \mathbf{t}_k \mathbf{p}_k^\top = \mathbf{X} - \sum_k \mathbf{t}_k \mathbf{p}_k^\top, \\ \mathbf{Y}_{k+1} &= \mathbf{Y}_k - \mathbf{t}_k \mathbf{q}_k^\top = \mathbf{Y} - \sum_k \mathbf{t}_k \mathbf{q}_k^\top. \end{aligned}$$

Тогда каждый следующий вектор  $\mathbf{t}_k$  оказывается ортогонален всем векторам  $\mathbf{t}_i$ ,  $i = 1, \dots, k$ .

На Рис. 1 продемонстрирован результат работы алгоритма PLS для случая, когда размерности пространств объектов, ответов и латентного пространства равны 2 ( $n = r = l = 2$ ). Синими и зелёными точками изображены строки матриц  $\mathbf{X}$  и  $\mathbf{Y}$ . Точки были сгенерированы из нормального распределения с нулевым матожиданием. Красным контуром показаны линии уровня матриц ковариаций распределений. Черным изображены единичные окружности. Красные стрелки соответствуют главным компонентам. Черные стрелки соответствуют векторам матриц  $\mathbf{W}$  и  $\mathbf{C}$  алгоритма PLS. Вектора  $\mathbf{t}_k$  и  $\mathbf{u}_k$  равны проекциям матриц  $\mathbf{X}_k$  и  $\mathbf{Y}_k$  на вектора  $\mathbf{w}_k$  и  $\mathbf{c}_k$  соответственно и изображены черными плюсами. Учёт взаимной связи между матрицами  $\mathbf{X}$  и  $\mathbf{Y}$  отклоняет вектора  $\mathbf{w}_k$  и  $\mathbf{c}_k$  от направления главных компонент. Вектора  $\mathbf{w}_k$  отклоняются незначительно. На первой итерации  $\mathbf{c}_1$  близок к  $pc_1$ , но вектора  $\mathbf{c}_k$ , найденные на следующих итерациях могут оказаться сильно коррелированными. Это происходит в следствие того, что из матрицы  $\mathbf{Y}$  на каждом шаге вычитается одноранговая аппроксимация, найденная в пространстве матрицы  $\mathbf{X}_k$ .

Fig. 1 shows the result of the PLS algorithm for the case when the space dimensions of objects, responses and latent variables are equal to 2 ( $n = r = l = 2$ ). Blue and green dots show the lines of matrices  $\mathbf{X}$  and  $\mathbf{Y}$ , respectively. The dots were generated from a normal distribution with zero expectation. A red outline shows the covariances matrix level lines

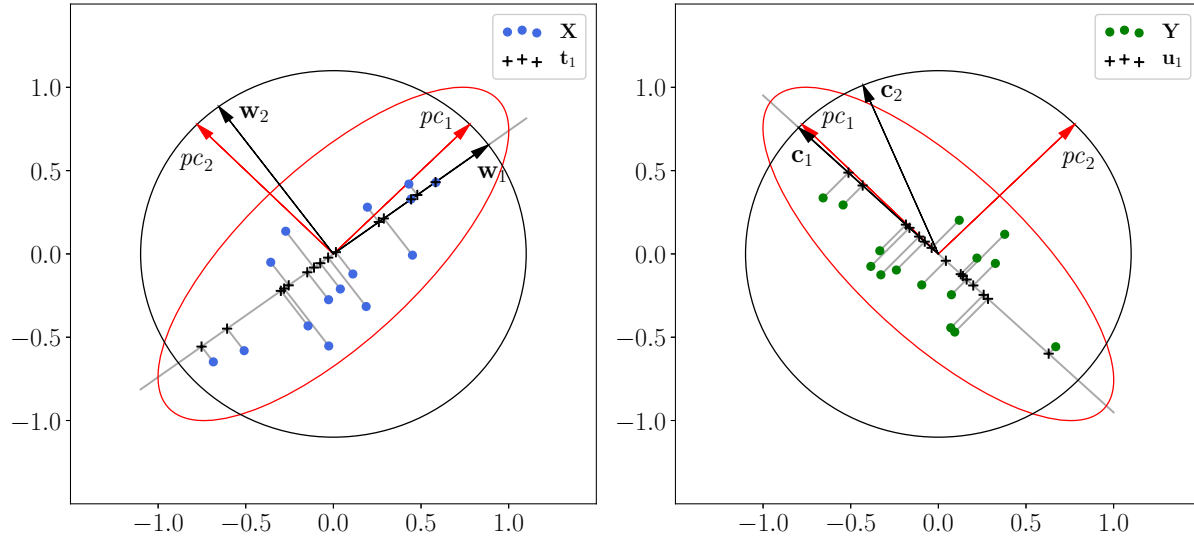


Рис. 1: Иллюстрация алгоритма PLS

of the distributions. A black outline is a unit circle. Red arrows correspond to principal components. Black arrows correspond to the vectors of matrices  $\mathbf{W}$  and  $\mathbf{C}$  from the PLS algorithm. Vectors  $\mathbf{t}_k$  and  $\mathbf{u}_k$  are equal to the projection matrices  $\mathbf{X}_k$  and  $\mathbf{Y}_k$  on the vectors  $\mathbf{w}_k$  and  $\mathbf{c}_k$ , respectively, and are shown by black pluses. Account the mutual coupling between the matrices  $\mathbf{X}$  and  $\mathbf{Y}$  rejects vectors  $\mathbf{w}_k$  and  $\mathbf{c}_k$  from the directions of the main components. Vectors  $\mathbf{w}_k$  deviate slightly. In the first iteration,  $\mathbf{c}_1$  is close to the principal component  $pc_1$ , but vectors  $\mathbf{c}_k$  found in the next iterations can be strongly correlated. This is due to the fact that from the matrix  $\mathbf{Y}$  at each step the peer approximation found in space of the matrix  $\mathbf{X}_k$  is subtracted.

Для получения прогнозов модели и нахождения параметров модели домножим справа формулу (3) на матрицу  $\mathbf{W}$ . Строки матрицы невязок  $\mathbf{E}$  ортогональны столбцам матрицы  $\mathbf{W}$ , поэтому

$$\mathbf{XW} = \mathbf{TP}^T \mathbf{W}.$$

Линейное преобразование между объектами в исходном и латентном пространстве имеет вид

$$\mathbf{T} = \mathbf{XW}^*, \quad (5)$$

где  $\mathbf{W}^* = \mathbf{W}(\mathbf{P}^T \mathbf{W})^{-1}$ .

To obtain the model predictions and find the model parameters, we multiply the formula (3) from the right by the matrix  $\mathbf{W}$ . Since the rows of the residual matrix  $\mathbf{E}$  are orthogonal to the columns of the matrix  $\mathbf{W}$ , we have

$$\mathbf{XW} = \mathbf{TP}^T \mathbf{W}.$$

The linear transformation between objects in the source and latent spaces has the form

$$\mathbf{T} = \mathbf{XW}^*, \quad (6)$$



210 where  $\mathbf{W}^* = \mathbf{W}(\mathbf{P}^\top \mathbf{W})^{-1}$ .

Матрица параметров модели 1 находится из уравнений (4), (6)

$$\mathbf{Y} = \mathbf{T}\mathbf{Q}^\top + \mathbf{E} = \mathbf{X}\mathbf{W}^*\mathbf{Q}^\top + \mathbf{E} = \mathbf{X}\mathbf{\Theta} + \mathbf{E}.$$

211 Таким образом, параметры модели (1) равны

$$\mathbf{\Theta} = \mathbf{W}(\mathbf{P}^\top \mathbf{W})^{-1}\mathbf{Q}^\top. \quad (7)$$

The parameter matrix of the model 1 is found from equations (4), (6)

$$\mathbf{Y} = \mathbf{T}\mathbf{Q}^\top + \mathbf{E} = \mathbf{X}\mathbf{W}^*\mathbf{Q}^\top + \mathbf{E} = \mathbf{X}\mathbf{\Theta} + \mathbf{E}.$$

212 Thus, the parameters of the model (1) are equal to

$$\mathbf{\Theta} = \mathbf{W}(\mathbf{P}^\top \mathbf{W})^{-1}\mathbf{Q}^\top. \quad (8)$$

213 После нахождения параметров модели на этапе тестирования для нахождения пред-  
214 сказания модели необходимо

- 215 • нормировать тестовые данные;
- 216 • вычислить предсказание модели с помощью линейного преобразования с матри-  
217 цей  $\mathbf{\Theta}$  из (8);
- 218 • провести обратную нормировку.

219 To find the model predictions after finding the model parameters during the testing phase  
220 we have to

- 221 • normalize the test data;
- 222 • compute the prediction of the model using the linear transformation with the matrix  
223  $\mathbf{\Theta}$  from (8);
- 224 • perform the inverse normalization.

## 225 4 Computational experiment

226 Временные ряды энергии состоят из почасовых записей (всего 52512 наблюдений),  
227 а погодные измерения проводились раз в день и содержат 2188 наблюдений. Строка  
228 матрицы  $\mathbf{X}$  — локальная история сигнала за одну неделю  $n = 24 \times 7$ . Строка матрицы  
229  $\mathbf{Y}$  — локальный прогноз потребления электроэнергии в следующие 24 часа  $r = 24$ . В  
230 этом случае матрицы  $\mathbf{X}$  и  $\mathbf{Y}$  являются авторегрессионными матрицами(можно ссыл-  
231 ку).

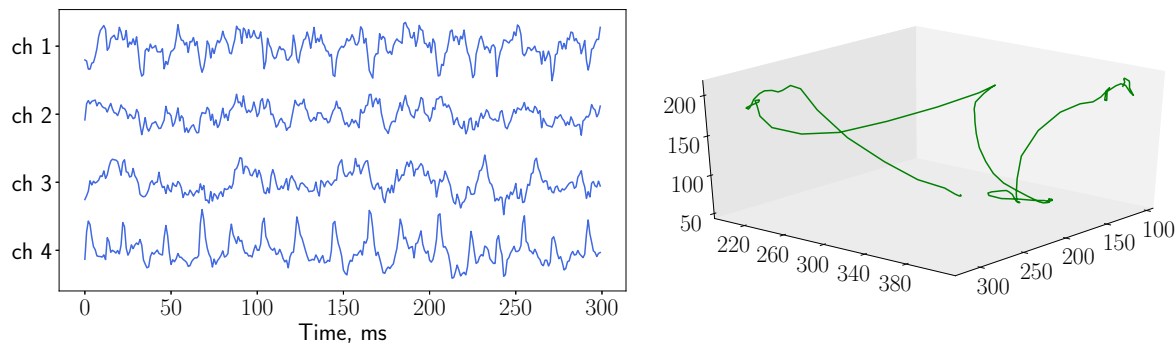


Рис. 2: Пример данных ECoG. Слева изображены данные напряжения, снятые с нескольких каналов, справа — координаты руки по трём осям. The ECoG data example. On the left there is voltage data taken from multiple channels, on the right there are coordinates of the hand in three axes. [как-то неочень](#)

Energy time series contain hourly records (total of 52512 observations), while weather time series were measured daily and contain 2188 observations. A row of the matrix  $\mathbf{X}$  is the local history of the signal for one week  $n = 24 \times 7$ . A row of the matrix  $\mathbf{Y}$  is the local forecast of electricity consumption for the next 24 hours  $r = 24$ . In this case, the matrices  $\mathbf{X}$  and  $\mathbf{Y}$  are autoregressive matrices ([можно ссылку](#)).

В случае данных ECoG матрица  $\mathbf{X}$  состоит из пространственно-временного спектрального представления временных рядов напряжения, а матрица  $\mathbf{Y}$  содержит информацию о положении руки. Процесс генерации матрицы  $\mathbf{X}$  из значений напряжения описан в [23]. Признаковое описание в каждый момент времени имеет размерность 864, положение руки описывается координатами по трём осям. Пример данных напряжения с разных каналов и соответствующих пространственных координат руки представлен на Рис. 2. Для прогнозирования движения руки используется авторегрессионный подход. Один объект состоит из признакового описания в несколько отсчётов времени. Ответом является положение руки в несколько следующих моментов времени. Требуется предсказать положение руки в несколько следующих моментов времени.

In the case of the ECoG data, the matrix  $\mathbf{X}$  consists of the space-time spectral representation of voltage time series, and the matrix  $\mathbf{Y}$  contains information about the position of the hand. The generation process of the matrix  $\mathbf{X}$  from the voltage values described in [23]. Feature descriptions in each moment of time has dimension equal to 864, the hand position is described by the coordinates in three axes. An example of voltage data samples with the different channels and corresponding spatial coordinates of the hand are shown in Fig. 2. To predict the movements of the hands we used an autoregressive approach. One object consists of a feature description in a few time. The answer is the hand position in the next few moments of time. It is required to predict the hand position in the next few moments of time.

Введём среднеквадратичную ошибку для некоторых матриц  $\mathbf{A} = [a_{ij}]$  и  $\mathbf{B} = [b_{ij}]$

$$\text{MSE}(\mathbf{A}, \mathbf{B}) = \sum_{i,j} (a_{ij} - b_{ij})^2.$$

Для оценивания качества аппроксимации вычисляется значение нормированной среднеквадратичной ошибки

$$\text{NMSE}(\mathbf{Y}, \hat{\mathbf{Y}}) = \frac{\text{MSE}(\mathbf{Y}, \hat{\mathbf{Y}})}{\text{MSE}(\mathbf{Y}, \bar{\mathbf{Y}})}, \quad (9)$$

где  $\hat{\mathbf{Y}}$  — прогноз модели,  $\bar{\mathbf{Y}}$  — константный прогноз средним значением по столбцам матрицы.

We introduce the mean-squared error for certain matrices  $\mathbf{A} = [a_{ij}]$  and  $\mathbf{B} = [b_{ij}]$

$$\text{MSE}(\mathbf{A}, \mathbf{B}) = \sum_{i,j} (a_{ij} - b_{ij})^2.$$

To estimate the approximation quality, we compute the value of the normalized MSE

$$\text{NMSE}(\mathbf{Y}, \hat{\mathbf{Y}}) = \frac{\text{MSE}(\mathbf{Y}, \hat{\mathbf{Y}})}{\text{MSE}(\mathbf{Y}, \bar{\mathbf{Y}})}, \quad (10)$$

where  $\hat{\mathbf{Y}}$  is the model forecast,  $\bar{\mathbf{Y}}$  is the constant forecast by the average value over the columns of the matrix.

## 4.1 Energy consumption dataset

Для нахождения оптимальной размерности  $l$  латентного пространства все данные потребления электроэнергии были разбиты на обучающую и валидационную части. Обучающая выборка состоит из 700 объектов, валидационная из 370. Зависимость нормированной квадратичной ошибки (10) от размерности  $l$  латентного пространства представлена на Рис. 3. Сначала ошибка резко падает при увеличении размерности скрытого пространства, а затем меняется незначительно.

To find the optimal dimensionality  $l$  of the latent space, the electricity consumption dataset was divided into training and validation parts. The training sample consists of 700 objects, the validation part is of 370 ones. The dependence of the normalized mean-squared error (10) on the latent space with dimensionality  $l$  is shown in Fig. 3. First, the error drops sharply with increasing the latent space dimensionality and then changes slightly.

Минимальная ошибка наблюдается при  $l = 14$ . Построим прогноз потребления электроэнергии при данном  $l$ . Результат аппроксимации изображен на Рис. 4. Алгоритм PLS восстановил авторегрессионную зависимость и обнаружил дневную сезонность.

The minimum error occurs when  $l = 14$ . Let us build a forecast of electricity consumption for a given  $l$ . The approximation result is shown in Fig. 4. The PLS algorithm restored the autoregressive dependence and found the daily seasonality.

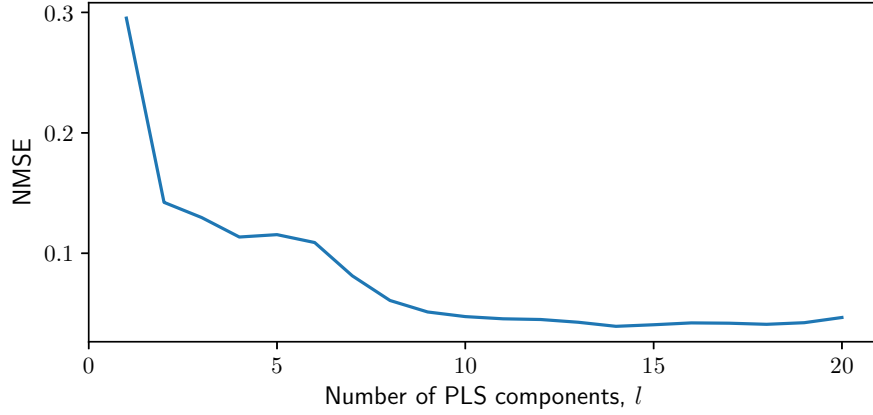


Рис. 3: Прогноз потребления электроэнергии алгоритмом PLS при размерности латентного пространства  $l=14$ . The forecast of electricity consumption by the PLS algorithm when the latent space dimensionality is equal to  $l=14$ .

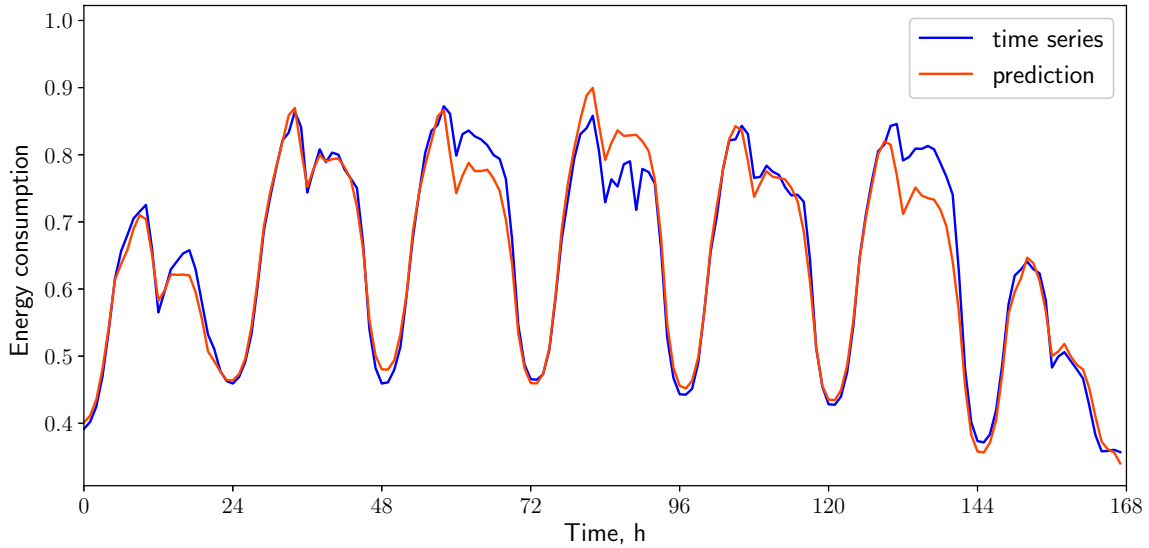


Рис. 4: Зависимость ошибки от размерности латентного пространства для данных потребления электроэнергии. The dependence of the error on the latent space dimensionality for the energy consumption dataset.

## 4.2 ECoG dataset

На Рис. 5 представлена зависимость нормированной квадратичной ошибки (10) от размерности латентного пространства. Ошибка аппроксимации меняется незначительно при  $l > 5$ . Таким образом совместное описание пространственно-временного спектрального представления объектов и пространственного положения руки может быть

представлено вектором размерности  $l \ll n$ . Зафиксируем  $l = 5$ . Пример аппроксимации положения руки изображен на Рис. 6. Сплошными линиями изображены истинные координаты руки по всем осям, пунктирными линиями показана аппроксимация методом PLS.

Fig. 5 shows the dependence of the normalized quadratic error (10) on the latent space dimensionality. The approximation error changes slightly when  $l > 5$ . Thus, the joint spatial-temporal spectrum representation of objects and the spatial position of the hand can be represented by a vector of dimensionality  $l \ll n$ . Let us fix  $l = 5$ . An example of the approximation of the hand position is shown in Fig. 6. Solid lines represent the true coordinates of the hand on all axes, the dotted lines show the approximation by the PLS method.

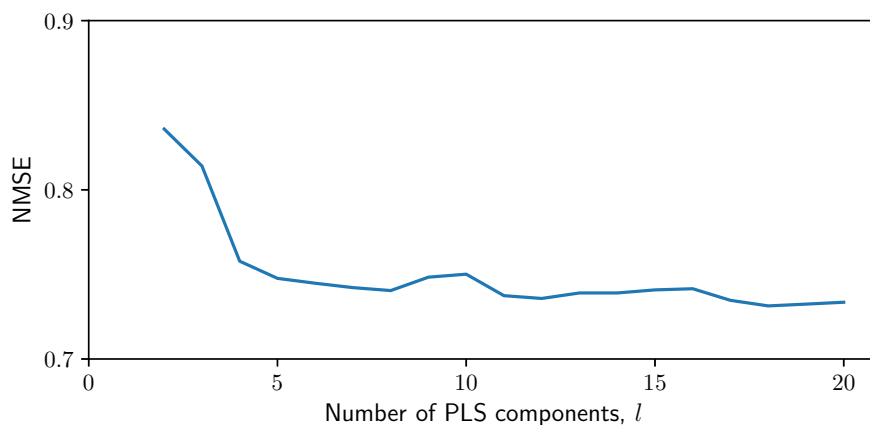


Рис. 5: Зависимость ошибки от размерности латентного пространства для данных ECoG. The dependence of the error on the latent space dimensionality for the ECoG dataset.

## 5 Conclusion

В данной работе предложен подход к решению задачи декодирования временных рядов и прогнозирования. Алгоритм частичных наименьших квадратов позволяет построить векторный прогноз целевой переменной. Латентное пространство содержит информацию об объектах и ответах и снижает размерности исходных матриц на порядки. Вычислительный эксперимент продемонстрировал применимость предложенного метода к задачам прогнозирования временных рядов объёмов потребления электроэнергии и проектирования нейрокомпьютерного интерфейса.

In the paper we proposed the approach for solving the problem of decoding and forecasting time-series. The algorithm of partial least squares allows to build a target variable vector prediction. The latent space contains information about the objects and the answers and reduces the dimensions of the source (? initial) matrices by orders of magnitude. The

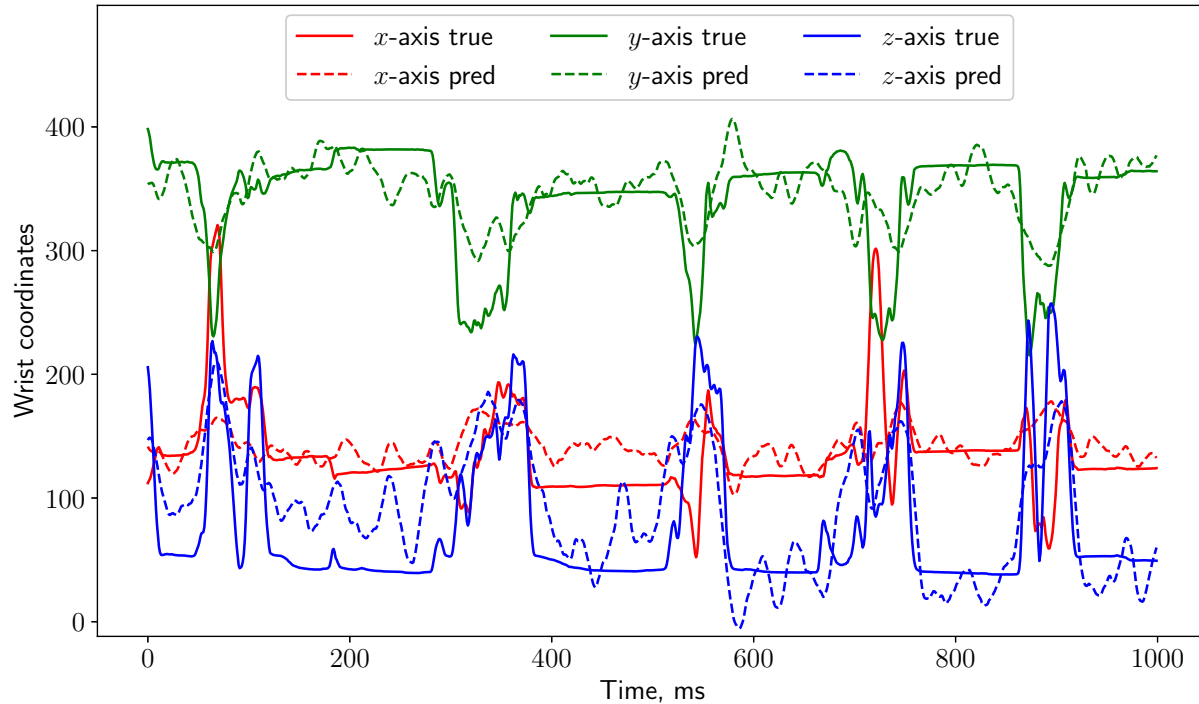


Рис. 6: Прогноз движения руки данных ECoG алгоритмом PLS при размерности латентного пространства  $l = 5$ . The forecast of the hand movement by the PLS algorithm when the latent space dimensionality is equal to  $l=5$ .

computational experiment demonstrated the applicability of the proposed method to the tasks of electricity consumption forecasting and brain-computer interface designing.

## Список литературы

- [1] AM Katrutsa and VV Strijov. Stress test procedure for feature selection algorithms. *Chemometrics and Intelligent Laboratory Systems*, 142:172–183, 2015.
- [2] Jundong Li, Kewei Cheng, Suhang Wang, Fred Morstatter, Robert P Trevino, Jiliang Tang, and Huan Liu. Feature selection: A data perspective. *arXiv preprint arXiv:1601.07996*, 2016.
- [3] Jacob A Wegelin et al. A survey of partial least squares (pls) methods, with emphasis on the two-block case. *University of Washington, Department of Statistics, Tech. Rep*, 2000.
- [4] Hervé Abdi. Partial Least Squares (PLS) Regression. *Encyclopedia for research methods for the social sciences*, pages 792–795, 2003.

- [5] Paul Geladi and Bruce R Kowalski. Partial least-squares regression: a tutorial. *Analytica chimica acta*, 185:1–17, 1986.
- [6] Paul Geladi. Notes on the history and nature of partial least squares (PLS) modelling. *Journal of Chemometrics*, 2(January):231–246, 1988.
- [7] Agnar Höskuldsson. PLS regression. *Journal of Chemometrics*, 2(August 1987):581–591, 1988.
- [8] Sijmen De Jong. Simpls: an alternative approach to partial least squares regression. *Chemometrics and intelligent laboratory systems*, 18(3):251–263, 1993.
- [9] Roman Rosipal and Nicole Kramer. Overview and Recent Advances in Partial Least Squares. *C. Saunders et al. (Eds.): SLSFS 2005, LNCS 3940*, pages 34–51, 2006.
- [10] Roman Rosipal. Nonlinear partial least squares: An overview. *Chemoinformatics and Advanced Machine Learning Perspectives: Complex Computational Methods and Collaborative Techniques*, pages 169–189, 2011.
- [11] Svante Wold, Nouna Kettaneh-Wold, and Bert Skagerberg. Nonlinear pls modeling. *Chemometrics and Intelligent Laboratory Systems*, 7(1-2):53–65, 1989.
- [12] Ildiko E. Frank. A nonlinear PLS model. *Chemometrics and Intelligent Laboratory Systems*, 8(2):109–119, 1990.
- [13] S Joe Qin and Thomas J McAvoy. Nonlinear pls modeling using neural networks. *Computers & Chemical Engineering*, 16(4):379–391, 1992.
- [14] Xuefeng F. Yan, Dezhaoh Z. Chen, and Shangxu X. Hu. Chaos-genetic algorithms for optimizing the operating conditions based on RBF-PLS model. *Computers and Chemical Engineering*, 27(10):1393–1404, 2003.
- [15] Hugo Hiden, Ben McKay, Mark Willis, and Gary Montague. Non-linear partial least squares using genetic. In *Genetic Programming 1998: Proceedings of the Third*, pages 128–133. Morgan Kaufmann, 1998.
- [16] Project tycho <http://neurotycho.org/food-tracking-task>.
- [17] José del R Millán, Rüdiger Rupp, Gernot Mueller-Putz, Roderick Murray-Smith, Claudio Giugliemma, Michael Tangermann, Carmen Vidaurre, Febo Cincotti, Andrea Kubler, Robert Leeb, et al. Combining brain–computer interfaces and assistive technologies: State-of-the-art and challenges. *Frontiers in Neuroscience*, 4:161, 2010.
- [18] SG Mason, A Bashashati, M Fatourehchi, KF Navarro, and GE Birch. A comprehensive survey of brain interface technology designs. *Annals of biomedical engineering*, 35(2):137–169, 2007.

- 360 [19] José del R Millán, Frédéric Renkens, Josep Mouriño, and Wulfram Gerstner. Brain-  
361 actuated interaction. *Artificial Intelligence*, 159(1-2):241–259, 2004.
- 362 [20] Luis Fernando Nicolas-Alonso and Jaime Gomez-Gil. Brain computer interfaces, a  
363 review. *Sensors*, 12(2):1211–1279, 2012.
- 364 [21] Setare Amiri, Reza Fazel-Rezai, and Vahid Asadpour. A review of hybrid brain-  
365 computer interface systems. *Advances in Human-Computer Interaction*, 2013:1, 2013.
- 366 [22] Andrey Eliseyev and Tetiana Aksenova. Penalized multi-way partial least squares for  
367 smooth trajectory decoding from electrocorticographic (ecog) recording. *PloS one*,  
368 11(5):e0154878, 2016.
- 369 [23] Motrenko A. Gasanov I. Creation of approximating scalogram description in a problem  
370 of movement prediction. *Journal of Machine Learning and Data Analysis*, 3(2):160–  
371 169, 2017.