

# Нелинейное снижение размерности в задачах декодирования временных рядов и прогнозирования\*

Мария Владимировна, Роман Исаченко

vladimirova.maria@phystech.edu, isa-ro@yandex.ru

Московский физико-технический институт

В работе решается задача обнаружения зависимостей в прогнозируемой переменной. Используются набор гомогенных моделей, восстанавливающих прогноз по общему для всех переменных описанию объектов. Рассматривается линейная модель метода частных наименьших квадратов и ее предложенная нелинейная модификация. Находятся оптимальные параметрические преобразования исходных пространств объектов и ответов. Проводится вычислительный эксперимент на реальных данных объемов потребления электроэнергии и данных сигналов кортикограмм.

**Ключевые слова:** прогнозирование временных рядов; мультиколлинеарность; метод частных наименьших квадратов; PLS; нелинейный PLS

## 1. Введение

В работе рассматривается задача прогнозирования временных рядов в случае наличия мультиколлинеарности в данных. Методы решения данной задачи сравниваются на двух наборах данных, имеющих избыточную информацию.

Первый набор данных представляет собой временные ряды объема потребления электроэнергии в Варшаве. Электрическая энергия является важной движущей силой экономического развития, а точность прогнозов спроса является важным фактором, который ведет к успешному эффективному планированию. По этой причине энергетическим анализам необходимо руководство для лучшего выбора наиболее подходящих методов прогнозирования, чтобы обеспечить точные прогнозы тенденций потребления электроэнергии. Предполагается, что значение сигнала в данный момент времени линейно зависит от предыдущих значений этого же сигнала, поэтому данные являются мультиколлинеарными.

Второй набор данных взят из проекта Project Tycho, в котором изучалась проблема проектирования нейро-компьютерного интерфейса (BCI) для обмена информацией между мозгом и электронным устройством. Решается задача выбора функций в моделях регрессии в приложении к декодированию движения на основе электрокардиограмм (ECoG). Проблема состоит в том, чтобы предсказать траектории руки из временных рядов напряжения кортикальной активности. Описание функции каждой точки находится в пространственно-временной частотной области включает в себя сами временные ряды напряжения и их спектральные характеристики. Выбор функции имеет решающее значение для адекватного решения проблемы регрессии, поскольку электрокортикальные данные являются высокоразмерными и измерения коррелируют как во временной, так и в пространственной областях.

Система BCI улучшает умственные и физические возможности пользователя, обеспечивая прямую связь между мозгом и компьютером. BCI направлены на восстановление поврежденных функциональных возможностей пациентов с механическими или когнитивными нарушениями. В данной статье предлагается новый метод выбора признаков

29 в прогнозировании движения и его реконструкции. Первый шаг к прогнозированию  
30 предполагаемых движений — научиться реконструировать фактические перемещения из  
31 кортикальной активности. Рассматривается проблема непрерывной реконструкции траек-  
32 тории. Субдуральные сигналы ECoG измеряются через 32 или 64 канала, когда субъект  
33 перемещает руку. Когда сигналы ECoG трансформируются в информационные функции,  
34 проблема восстановления траектории является проблемой регрессии. Извлечение функ-  
35 ции включает в себя применение некоторого спектрально-временного преобразования к  
36 сигналам ECoG с каждого канала. Так как результирующее пространственно-временное  
37 спектральное представление сильно избыточно, используются различные методы выбора  
38 объектов и уменьшения размерности, чтобы извлечь только наиболее важные функции.

39 Для решения задачи прогнозирования используется авторегрессионная модель. Авто-  
40 регрессионная модель является неустойчивой в случае наличия мультиколлинеарности в  
41 исторических данных. Для решения этой проблемы необходимо используются методы от-  
42 бора признаков [21], в результате чего повышается устойчивость модели без существенного  
43 снижения качества прогноза.

44 В работе исследуются методы отбора признаков: метод частных наименьших квад-  
45 ратов (PLS) [11] и предложенная его нелинейная модификация (cnlPLS). Метод частных  
46 наименьших квадратов основан на снижении размерности матрицы признаков и выделя-  
47 ет линейные комбинации признаков, которые оказывают наибольшее влияние на вектор  
48 ответов. Выделение признаков происходит итеративно, в порядке уменьшения их влия-  
49 ния на вектор ответов [11]. Рассматриваются только значимые комбинации признаков,  
50 незначительно потеряв в точности прогноза.

51 Методы PLS регрессии подробно описаны в работах [8,9]. Разницу между методом PLS  
52 и связанными с ним подходами, различные разновидности регрессии PLS можно найти  
53 в [17].

54 Нелинейное расширение метода PLS регрессии впервые введено в [6]. В литературе  
55 были разработаны различные модификации PLS. Предложены нелинейные методы PLS,  
56 основанные на различных моделях: искусственных нейронных сетей [4], функции актива-  
57 ции радиальных оснований [5], логистическая функция активации и методы оптимизации  
58 роевых частиц [7], используют прямые нейронные сети [3], искусственную нейронную сеть  
59 Эльмана [10].

60 Предлагается провести модификацию алгоритма PLS: совершить криволинейное и  
61 нелинейное преобразования пространства целевой переменной для учета зависимостей  
62 между сигналами в разные моменты времени.

63 В работе проведено сравнение двух методов отбора признаков в задаче авторегрессион-  
64 ного прогнозирования сигналов (PLSR и cnlPLSR). Цель регрессии PLS [18] — предсказать  
65  $\mathbf{Y}$  по  $\mathbf{X}$  и описать их общую структуру. Когда  $\mathbf{Y}$  — вектор, а  $\mathbf{X}$  — матрица полного ран-  
66 га, эта цель может быть выполнена с использованием обычной линейной регрессии. Если  
67 число предикторов велико по сравнению с числом наблюдений, то  $\mathbf{X}$  будет сингулярной и  
68 регрессионный подход в этом случае невозможен из-за наличия мультиколлинеарности.

69 В качестве практической проверки данных методов в ходе вычислительного экспери-  
70 мента решается задача прогнозирования на реальных данных. Результатом применения  
71 отбора признаков является снижение размерности задачи и повышение устойчивости мо-  
72 делей без существенной потери точности прогноза.

## Постановка задачи

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

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

$$\mathbf{y} = \mathbf{x}\Theta + \boldsymbol{\varepsilon}, \quad (1)$$

где  $\Theta \in \mathbb{R}^{n \times r}$  — матрица параметров модели, а  $\boldsymbol{\varepsilon} \in \mathbb{R}^r$  — 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  $\mathcal{D}$ :

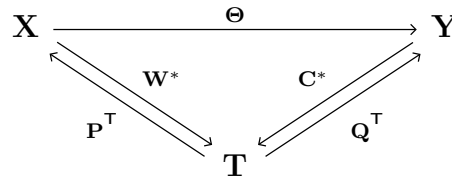
$$S(\Theta|\mathcal{D}) = \|\mathbf{X}\Theta - \mathbf{Y}\|_2^2 = \sum_{i=1}^m \|\mathbf{x}_i\Theta - \mathbf{y}_i\|_2^2 \rightarrow \min_{\Theta}. \quad (2)$$

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

## Метод частных наименьших квадратов (PLS)

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 ( $l < n$ ) с сохранением взаимосвязи между объектами и ответами.

Схема алгоритма PLS изображена на следующей коммутативной диаграмме



Каждая стрелка соответствует линейному отображению с матрицей параметров, указанной на диаграмме. Необходимо найти линейное отображение из пространства объектов в пространство ответов. Данное отображение соответствует модели (1). 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 на латентное пространство следующим образом:

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

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

101 где  $\mathbf{T}$  — матрица совместного описания объектов и ответов в латентном пространстве,  
 102 причём  $\mathbf{T}^\top \mathbf{T} = \mathbf{I}_l$ ;  $\mathbf{P}$ ,  $\mathbf{Q}$  — матрицы перехода из латентного пространства в исходные  
 103 пространства;  $\mathbf{E}$ ,  $\mathbf{F}$  — матрицы невязок.

104 Псевдокод метода регрессии PLS приведен в алгоритме 1. Алгоритм итеративно на  
 105 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.  
 106 После вычисления следующего набора векторов из матриц  $\mathbf{X}$ ,  $\mathbf{Y}$  вычитаются очередные  
 107 одноранговые аппроксимации.

---

#### Алгоритм 1 Алгоритм PLSR

---

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

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

```

1: нормировать матрицы  $\mathbf{X}$  и  $\mathbf{Y}$ 
2: инициализировать  $\mathbf{u}_0$  (первый столбец матрицы  $\mathbf{Y}$ )
3:  $\mathbf{X}_1 = \mathbf{X}$ ;  $\mathbf{Y}_1 = \mathbf{Y}$ 
4: для  $k = 1, \dots, l$ 
5:   повторять
6:      $\mathbf{w}_k := \mathbf{X}_k^\top \mathbf{u}_{k-1} / (\mathbf{u}_{k-1}^\top \mathbf{u}_{k-1})$ ;  $\mathbf{w}_k := \frac{\mathbf{w}_k}{\|\mathbf{w}_k\|}$ 
7:      $\mathbf{t}_k := \mathbf{X}_k \mathbf{w}_k$ 
8:      $\mathbf{c}_k := \mathbf{Y}_k^\top \mathbf{t}_k / (\mathbf{t}_k^\top \mathbf{t}_k)$ ;  $\mathbf{c}_k := \frac{\mathbf{c}_k}{\|\mathbf{c}_k\|}$ 
9:      $\mathbf{u}_k := \mathbf{Y}_k \mathbf{c}_k$ 
10:   пока  $\mathbf{t}_k$  не стабилизируется
11:    $\mathbf{p}_k := \mathbf{X}_k^\top \mathbf{t}_k / (\mathbf{t}_k^\top \mathbf{t}_k)$ ,  $\mathbf{q}_k := \mathbf{Y}_k^\top \mathbf{t}_k / (\mathbf{t}_k^\top \mathbf{t}_k)$ 
12:    $\mathbf{X}_{k+1} := \mathbf{X}_k - \mathbf{t}_k \mathbf{p}_k^\top$ 
13:    $\mathbf{Y}_{k+1} := \mathbf{Y}_k - \mathbf{t}_k \mathbf{q}_k^\top$ 

```

---

108 Вектора  $\mathbf{t}_k$  и  $\mathbf{u}_k$  из внутреннего цикла алгоритма 1 содержат информацию о матрице  
 109 объектов  $\mathbf{X}$  и матрице ответов  $\mathbf{Y}$  соответственно. *Использование вектора  $\mathbf{t}_k$  при вычисле-*  
 110 *нии вектора  $\mathbf{u}_k$  и наоборот позволяет извлечь взаимосвязь.*

111 Теоретическое обоснование алгоритма PLS следует из следующих утверждений.

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

114 Утверждение следует из следующего равенства

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

115 Максимизация дисперсий векторов  $\mathbf{t}_k$  и  $\mathbf{u}_k$  сохраняет информацию об исходных матрицах,  
 116 корреляция отвечает взаимосвязи между  $\mathbf{X}$  и  $\mathbf{Y}$ .

117 Во внутреннем цикле алгоритма вычисляются вектора  $\mathbf{w}_k$  и  $\mathbf{c}_k$ . Из данных векторов стро-  
 118 ятся матрицы  $\mathbf{W}$  и  $\mathbf{C}$  соответственно.

**Утверждение 2.** Обновление векторов по шагам (6)–(9) алгоритма 1 соответствует максимизации ковариации между векторами  $\mathbf{t}_k$  и  $\mathbf{u}_k$

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

119 где  $\lambda_{\max}(\cdot)$  — максимальное собственное значение матрицы.

120 **Утверждение 3.** В результате выполнения внутреннего цикла вектора  $\mathbf{w}_k$  и  $\mathbf{c}_k$  будут  
121 являться собственными векторами матриц  $\mathbf{X}_k^\top \mathbf{Y}_k \mathbf{Y}_k^\top \mathbf{X}_k$  и  $\mathbf{Y}_k^\top \mathbf{X}_k \mathbf{X}_k^\top \mathbf{Y}_k$ , соответствующими  
122 максимальным собственным значениям.

123 Вектора  $\mathbf{w}_k$ ,  $\mathbf{c}_k$  являются собственными векторами матриц  $\mathbf{X}_k^\top \mathbf{Y}_k \mathbf{Y}_k^\top \mathbf{X}_k$  и  $\mathbf{Y}_k^\top \mathbf{X}_k \mathbf{X}_k^\top \mathbf{Y}_k$   
124 соответственно

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

125 где символ  $\propto$  означает равенство с точностью до мультипликативной константы. Утвер-  
126 ждение следует из того факта, что правила обновления векторов  $\mathbf{w}_k$ ,  $\mathbf{c}_k$  совпадают с ите-  
127 рацией алгоритма поиска максимального собственного значения [20].

После завершения внутреннего цикла 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  $\mathbf{t}_k \mathbf{q}_k^\top$

$$\mathbf{X}_{k+1} = \mathbf{X}_{k-1} - \mathbf{t}_k \mathbf{p}_k^\top = \mathbf{X} - \sum_k \mathbf{t}_k \mathbf{p}_k^\top,$$

$$\mathbf{Y}_{k+1} = \mathbf{Y}_{k-1} - \mathbf{t}_k \mathbf{q}_k^\top = \mathbf{Y} - \sum_k \mathbf{t}_k \mathbf{q}_k^\top,$$

128 Этим обеспечивает факт того, что каждый следующий вектор  $\mathbf{t}_k$  оказывается ортогонален  
129 всем векторам  $\mathbf{t}_i$ ,  $i = 1, \dots, k$ .

130 На Рис. 1 продемонстрирован результат работы алгоритма PLS для случая, когда раз-  
131 мерности пространств объектов, ответов и латентного пространства равны 2 ( $n = r = l =$   
132  $= 2$ ). Синими и зелёными точками изображены матрицы  $\mathbf{X}$  и  $\mathbf{Y}$ . Точки были сгенери-  
133 рованы из нормального распределения с нулевым мат. ожиданием. Красным контуром  
134 показаны линии уровня матриц ковариаций распределений. Черным проведены единич-  
135 ные окружности. Красные стрелки соответствуют главным компонентам. Черные стрелки  
136 соответствуют векторам матриц  $\mathbf{W}$  и  $\mathbf{C}$  алгоритма PLS. Вектора  $\mathbf{t}_k$  и  $\mathbf{u}_k$  равны проекциям  
137 матриц  $\mathbf{X}_k$  и  $\mathbf{Y}_k$  на вектора  $\mathbf{w}_k$  и  $\mathbf{c}_k$  соответственно.

138 Домножим справа формулу (6) на матрицу  $\mathbf{W}$ . Строки матрицы невязок  $\mathbf{E}$  ортого-  
139 нальны столбцам матрицы  $\mathbf{W}$ , поэтому

$$\mathbf{XW} = \mathbf{TW}.$$

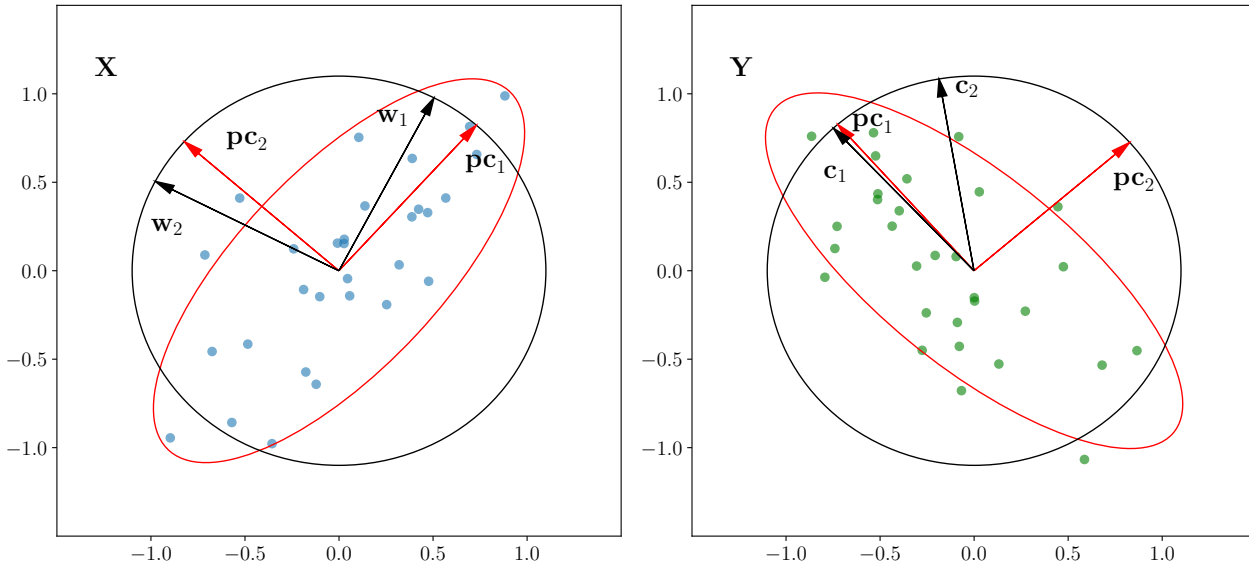


Рис. 1. Иллюстрация алгоритма PLS

140 Линейное преобразование между объектами в исходном и латентном пространстве имеет  
141 вид

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

142 где  $\mathbf{W}^* = \mathbf{W}(\mathbf{P}^T \mathbf{W})^{-1}$ . Аналогичным образом может быть получена матрица перехода из  
143 пространства ответов в латентное пространство  $\mathbf{C}^* = \mathbf{C}(\mathbf{Q}^T \mathbf{C})^{-1}$ .

Матрица параметров модели 1 находится из уравнений (7), (5)

$$\mathbf{Y} = \mathbf{TQ}^T + \mathbf{F} = \mathbf{XW}^* \mathbf{Q}^T + \mathbf{F} = \mathbf{X}\boldsymbol{\Theta} + \mathbf{F}.$$

Таким образом,

$$\boldsymbol{\Theta} = \mathbf{W}(\mathbf{P}^T \mathbf{W})^{-1} \mathbf{Q}^T.$$

## 144 Модификация метода частных наименьших квадратов (cnlPLS)

145 Предлагается провести модификацию алгоритма PLS: применить нелинейные пара-  
146 метрические преобразования пространства объектов и ответов для выявления сложных  
147 зависимостей.

$$\tilde{\mathbf{X}} = F_x(\mathbf{X}, \mathbf{v}_x) = \mathbf{T} \cdot \mathbf{P}^T + \mathbf{E} \quad (6)$$

$$\tilde{\mathbf{Y}} = F_y(\mathbf{Y}, \mathbf{v}_y) = \mathbf{T} \cdot \mathbf{Q}^T + \mathbf{F}, \quad (7)$$

## 148 Нелинейные преобразования

149 Рассматриваются нелинейное параметрическое преобразование пространства зависи-  
150 мой переменной  $\mathbf{Y}$  и независимой переменной  $\mathbf{X}$  (примеры преобразований представлены  
151 в табл. 1). Преобразование и вектор параметров, относящиеся к зависимой переменной и  
152 независимой переменной, обозначим соответственно  $F_y(\mathbf{Y}, \mathbf{v}_y)$  и  $F_x(\mathbf{X}, \mathbf{v}_x)$  и введем пере-  
153 менные для преобразованных пространств

$$\tilde{\mathbf{Y}} = F_y(\mathbf{Y}, \mathbf{v}_y), \quad \tilde{\mathbf{X}} = F_x(\mathbf{X}, \mathbf{v}_x). \quad (8)$$

154 Функции для криволинейных преобразований удовлетворяют следующим условиям:

- 155 —  $F : \mathbb{R} \rightarrow \mathbb{R}$ ,
- 156 —  $F(0) = 0$ ,
- 157 —  $F$  дифференцируется по параметрам  $\mathbf{v}_y$ ,
- 158 — существует  $F^{-1}$ .

№	Функция	Параметры
1	$F(x) = \text{sign}(x) \exp(a)(\exp(b x ) - 1)$	$a, b > 0$
2	$F(x) = \text{sign}(x) \exp(a)(\exp(b \ln(1 +  x ) - 1)$	$a, b > 0$
3	$F(x) = \text{sign}(x) \exp(a)(\exp(b x ^{1/2}) - 1)$	$a, b > 0$
4	$F(x) = \text{sign}(x) \exp(a)(\exp(b x ^{1/3}) - 1)$	$a, b > 0$
5	$F(x) = \text{sign}(x) \exp(a)(\exp(b x ^{1/4}) - 1)$	$a, b > 0$
6	$F(x) = \text{sign}(x) \exp(a)(\exp(b x ^2) - 1)$	$a, b > 0$

Таблица 1. Нелинейные преобразования

Для обучения параметров  $\mathbf{v}_y$  используется градиентный метод. Предлагается подход для обновления весов  $\mathbf{v}_y$ , основанный на линеаризации функции преобразования. Разложим (8) в ряд Тейлора до второго порядка:

$$\mathbf{u} \approx \mathbf{u}_0 + \frac{\partial \mathbf{u}}{\partial \mathbf{v}_y} \Delta \mathbf{v}_y.$$

Для вычисления  $\Delta \mathbf{v}_y$  предложены следующие шаги. Рассматривается разница  $\mathbf{u} - \mathbf{u}_0 = \frac{\partial \mathbf{u}}{\partial \mathbf{v}_y} \Delta \mathbf{v}_y$ . Определется рассогласование

$$\mathbf{u} - \mathbf{u}_0 \approx \frac{\partial \mathbf{u}}{\partial \mathbf{v}_y} \Delta \mathbf{v}_y = \mathbf{J}_u \Delta \mathbf{v}_y,$$

159 где матрица  $\mathbf{J}_u$  состоит из частных производных  $\left\{ \frac{\partial \mathbf{u}}{\partial \mathbf{v}_y} \right\}$ , вычисленных при известном зна-  
160 чении переменной  $\mathbf{u}$ :

$$\mathbf{J}_u = \frac{\partial \mathbf{u}}{\partial \mathbf{v}_y} = \frac{\partial}{\partial \mathbf{v}_y} (\tilde{\mathbf{Y}} \mathbf{c}) = \frac{1}{(\mathbf{t}^\top \mathbf{t})} \frac{\partial}{\partial \mathbf{v}_y} (\tilde{\mathbf{Y}} \tilde{\mathbf{Y}}^\top \mathbf{t}) = \frac{1}{(\mathbf{t}^\top \mathbf{t})} \left( \frac{\partial \tilde{\mathbf{Y}}}{\partial \mathbf{v}_y} \cdot \tilde{\mathbf{Y}}^\top + \tilde{\mathbf{Y}} \cdot \frac{\partial \tilde{\mathbf{Y}}^\top}{\partial \mathbf{v}_y} \right) \mathbf{t}.$$

161 Правило обновления для вектора  $\Delta \mathbf{v}$  является решением задачи регрессии рассогла-  
162 сования

$$\Delta \mathbf{v}_y = (\mathbf{J}_u^\top \mathbf{J}_u)^{-1} \mathbf{J}_u^\top (\mathbf{u} - \mathbf{u}_0). \quad (9)$$

163 Аналогично преобразованию зависимой переменной сводим задачу обновления вектора  
164 параметров  $\mathbf{v}_x$  к задаче линейной регрессии:

$$\begin{aligned} \mathbf{t} - \mathbf{t}_0 &\approx \frac{\partial \mathbf{t}}{\partial \mathbf{v}_x} \Delta \mathbf{v}_x = \mathbf{J}_t \Delta \mathbf{v}_x \\ \Delta \mathbf{v}_x &= (\mathbf{J}_t^\top \mathbf{J}_t)^{-1} \mathbf{J}_t^\top (\mathbf{t} - \mathbf{t}_0). \end{aligned}$$

## Алгоритм cnlPLSR

165 В данном разделе представлен модифицированный метод PLSR, содержащий шаги  
166 преобразования целевой переменной. Аналогично методу PLSR (алгоритм ??), алгоритм  
167 ?? начинается с инициализации вектора  $\mathbf{u}$ , а обновления весов преобразования считается  
168 с помощью рассогласования  $\mathbf{e}$  для вектора  $\mathbf{u}$ , вычисленного в цикле и на предыдущей  
169 итерации.  
170

---

**Алгоритм 2** Алгоритм snlPLSR с преобразованием пространства объектов 2
 

---

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

**Выход:**  $\mathbf{T}, \mathbf{P}, \mathbf{Q}, \mathbf{v}_x, \mathbf{v}_y$ ;

```

1: инициализировать  $\mathbf{v}_x$  и  $\mathbf{v}_y$ 
2: нормировать матрицы  $\mathbf{X}, \mathbf{Y}$ 
3:  $\mathbf{X}_1 = \mathbf{X}; \mathbf{Y}_1 = \mathbf{Y}$ 
4: для  $k = 1, \dots, l$ 
5:   инициализировать  $\mathbf{t}_0$  (первый столбец матрицы  $\mathbf{X}$ )
6:   инициализировать  $\mathbf{u}_0$  (первый столбец матрицы  $\mathbf{Y}$ )
7:   повторять
8:      $\tilde{\mathbf{X}}_k := F_x(\mathbf{X}_k, \mathbf{v}_x); \quad \tilde{\mathbf{Y}}_k = F_y(\mathbf{Y}_k, \mathbf{v}_y)$ 
9:      $\mathbf{w}_k := \tilde{\mathbf{X}}_k^\top \mathbf{u}_{k-1} / (\mathbf{u}_{k-1}^\top \mathbf{u}_{k-1}); \quad \mathbf{w}_k := \frac{\mathbf{w}_k}{\|\mathbf{w}_k\|}$ 
10:     $\mathbf{t}_k := \tilde{\mathbf{X}}_k \mathbf{w}_k$ 
11:     $\Delta \mathbf{v}_x = (\mathbf{J}_t^\top \mathbf{J}_t)^{-1} \mathbf{J}_t^\top (\mathbf{t}_k - \mathbf{t}_{k-1})$ , где  $\mathbf{J}_t := \frac{\partial \mathbf{t}_k}{\partial \mathbf{v}_x}$ 
12:     $\mathbf{v}_x := \mathbf{v}_x + \Delta \mathbf{v}_x$ 
13:     $\mathbf{c}_k := \tilde{\mathbf{Y}}_k^\top \mathbf{t}_k / (\mathbf{t}_k^\top \mathbf{t}_k); \quad \mathbf{c}_k := \frac{\mathbf{c}_k}{\|\mathbf{c}_k\|}$ 
14:     $\mathbf{u}_k := \tilde{\mathbf{Y}}_k \mathbf{c}_k$ 
15:     $\Delta \mathbf{v}_y = (\mathbf{J}_u^\top \mathbf{J}_u)^{-1} \mathbf{J}_u^\top (\mathbf{u}_k - \mathbf{u}_{k-1})$ , где  $\mathbf{J}_u := \frac{\partial \mathbf{u}_k}{\partial \mathbf{v}_y}$ 
16:     $\mathbf{v}_y := \mathbf{v}_y + \Delta \mathbf{v}_y$ 
17:   пока  $\mathbf{t}_k$  не стабилизируется
18:    $\mathbf{p}_k := \tilde{\mathbf{X}}_k^\top \mathbf{t}_k / (\mathbf{t}_k^\top \mathbf{t}_k); \quad \mathbf{q}_k := \tilde{\mathbf{Y}}_k^\top \mathbf{t}_k / (\mathbf{t}_k^\top \mathbf{t}_k)$ 
19:    $\tilde{\mathbf{X}}_k := \tilde{\mathbf{X}}_k - \mathbf{t}_k \mathbf{p}_k^\top$ 
20:    $\tilde{\mathbf{Y}}_k := \tilde{\mathbf{Y}}_k - \mathbf{t}_k \mathbf{q}_k^\top$ 
21:    $\mathbf{X}_k = F_x^{-1}(\tilde{\mathbf{X}}_k, \mathbf{v}_x); \mathbf{Y}_k = F_y^{-1}(\tilde{\mathbf{Y}}_k, \mathbf{v}_y)$ 

```

---

## Вычислительный эксперимент

В случае данных электроэнергии  $i$ -ая строка матрицы  $\mathbf{X}$  — локальная история сигнала ( $n$  значений сигнала, начиная с момента  $i$ ), а  $i$ -ая строка матрицы  $\mathbf{Y}$  — локальный прогноз, то есть  $r$  значений сигнала, начиная с момента  $n + 1$ . 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 (ссылка на Мотренко).

В рамках вычислительного эксперимента строится прогноз временных рядов. В ходе эксперимента сравниваются методы PLSR, нелинейных автоэнкодеров и snlPLS. Сравнение проводится на реальных данных объемов потребления электроэнергии в Польше.

Вычислительный эксперимент, продемонстрированный в этом разделе, основан на данных электроэнергии. Данные состоят из временного ряда польских электрических нагрузок и временных рядов погоды в Варшаве (долгота: 21,25, широта: 52,30, высота над уровнем моря: 94). Временные ряды энергии состоят из почасовых записей (всего 52512 наблюдений), а погодные измерения проводились раз в день и содержат 2188 наблюдений. Многомасштабные временные ряды соответствуют периоду 1999-2004 годов. Результаты, полученные с этим набором данных, являются иллюстрацией предлагаемых методов, по-



Алгоритм	N=3	N=5	N=10	N=20
PLS	0,00404	0,00337	<b>0,00151</b>	0,00135
cnlPLS $g(x) = \text{sign}(x) \exp(a)(\exp(b x ) - 1)$	0.00529	0.00514	0.00536	0.00506
cnlPLS $g(x) = \text{sign}(x) \exp(a)(\exp(b \ln(1 +  x ) - 1)$	0.00362	0.00386	0.00326	0.00317
cnlPLS $g(x) = \text{sign}(x) \exp(a)(\exp(b x ^{1/2}) - 1)$	0.00272	0.00236	0.00287	<b>0.00128</b>
cnlPLS $g(x) = \text{sign}(x) \exp(a)(\exp(b x ^{1/3}) - 1)$	<b>0.00241</b>	<b>0.00233</b>	0.00221	0.00173
cnlPLS $g(x) = \text{sign}(x) \exp(a)(\exp(b x ^{1/4}) - 1)$	0.00796	0.00768	0.00737	0.00803
cnlPLS $g(x) = \text{sign}(x) \exp(a)(\exp(b x ^2) - 1)$	0.00816	0.00798	0.00796	0.00775

**Таблица 2.** Значения ошибки MSE для разных чисел компонент и разных функций

скольку данные содержат многомасштабные временные ряды, имеющие различный характер.

Примеры работы алгоритма приведены на рис. ???. Метод успешно делает краткосрочный прогноз (до 10 дней). С увеличением горизонта прогнозирования предсказание смещается.

Результаты вычислительного эксперимента для предложенного модифицированного алгоритма cnlPLS представлены на рис. ??. На графиках изображены сглаженные зависимости ошибки MSE от числа компонент в алгоритме для разных функций. Из графиков видно, что для функций (a) – (e) ошибка при увеличении числа компонент падает, затем колеблется, слабо меняясь. Ошибка алгоритма с функцией (f) увеличивается при увеличении числа компонент. Это означает, что преобразование, выполненное в пространстве целевой переменной с помощью функции (f), плохо описывает зависимость. Меньшую ошибку имеют функции, растущие медленнее, а именно (d) и (e).

В табл. 2 продемонстрировано увеличение точности прогнозирования при использовании криволинейного преобразования в пространстве зависимой переменной, но увеличение точности в пределах погрешности алгоритма (0.0005-0.0010). Функции с быстрым ростом не позволяют описать зависимость.

## Заключение

В данной работе предложен новый подход к обнаружению зависимостей в пространстве зависимой переменной задачи прогнозирования временных рядов. Сравнивались результаты прогнозирования временных рядов, полученных с помощью метода частных наименьших квадратов и предложенной модификации. Проведен вычислительный эксперимент на реальных данных потребления электроэнергии в Варшаве. Построенная прогностическая модель показала высокое качество предсказания электрической нагрузки.

## Литература

- [1] *Thrun, Sebastian and Pratt, Lorien* Learning to learn // Springer Science & Business Media, 2012.
- [2] *Chong, Il Gyo and Jun, Chi Hyuck* Performance of some variable selection methods when multicollinearity is present // *Chemometrics and Intelligent Laboratory Systems*, 2005. Vol. 78. No. 1. P. 103–112.
- [3] *Xuefeng, Yan* Hybrid artificial neural network based on BP-PLSR and its application in development of soft sensors // *Chemometrics and Intelligent Laboratory Systems*, 2010. Vol. 103. No. 2. P. 152–159.
- [4] *Mcavovt, J. and Process, Chemical Title* // *Journal name*, 2005. Vol. 16. No. 4. P. 379–391.
- [5] *Yan, Xuefeng F. and Chen, Dezhao Z. and Hu, Shangru X.* Chaos-genetic algorithms for optimizing the operating conditions based on RBF-PLS model // *Computers and Chemical Engineering*, 2003. Vol. 27. No. 10. P. 1393–1404.
- [6] *Frank, Ildiko E.* A nonlinear PLS model // *Chemometrics and Intelligent Laboratory Systems*, 1990. Vol. 8. No. 2. P. 109–119.
- [7] *Zhou, Yan Ping and Jiang, Jian Hui and Lin, Wei Qi and Xu, Lu and Wu, Hai Long and Shen, Guo Li and Yu, Ru Qin* Artificial neural network-based transformation for nonlinear partial least-square regression with application to QSAR studies // *Talanta*, 2007. Vol. 71. No. 2. P. 848–853.
- [8] *Chong, Il Gyo and Jun, Chi Hyuck* Notes on the history and nature of partial least squares (PLS) modelling // *Journal of Chemometrics*, 1988. Vol. 2. No. January. P. 231–246.
- [9] *Höskuldsson, Agnar* PLS regression // *Chemometrics and Intelligent Laboratory Systems*, 1987. Vol. 2. No. August. P. 581–591.
- [10] *Bulut, Elif and Egrioglu, Erol* A New Partial Least Square Method Based on Elman Neural Network // *Chemometrics and Intelligent Laboratory Systems*, 2005. Vol. 4. No. 4. P. 154–158.
- [11] *Ng, Kee Siong* A Simple Explanation of Partial Least Squares // *Journal title*, 2013. Vol. volume. No. number. P. 1–10.
- [12] *Rosipal, Roman* Nonlinear partial least squares: An overview // *Chemoinformatics and Advanced Machine Learning Perspectives: Complex Computational Methods and Collaborative Techniques*, 2011. Vol. number. No. number. P. 169–189.
- [13] *Wold, Svante and Kettaneh-Wold, Nouna and Skagerberg, Bert* Nonlinear PLS modeling // *Chemometrics and Intelligent Laboratory Systems*, 1989. Vol. 7. No. 1-2. P. 53–65.
- [14] *Rosipal, Roman and Kramer, Nicole* Overview and Recent Advances in Partial Least Squares // *????? C. Saunders et al. (Eds.): SLSFS 2005, LNCS 3940*, 2006. Vol. ?. No. ?. P. 34–51.
- [15] *Lu, Wen-Cong and Chen, Nian-Yi and Li, Guo-Zheng and Yang, Jie* Multitask Learning Using Partial Least Squares Method // *Proceedings of the Seventh International Conference on Information Fusion; International Society of Information Fusion*, 2004. Vol. 1. P. 79–84.
- [16] *Varnek, Alexandre and Baskin, Igor* Machine learning methods for property prediction in chemoinformatics: Quo Vadis? // *Journal of Chemical Information and Modeling*, 2012. Vol. 52. No. 6. P. 1413–1437.
- [17] *Lehky, Sidney R. and Kiani, Roozbeh and Esteky, Hossein and Tanaka, Keiji* Dimensionality of object representations in monkey inferotemporal cortex // *Neural computation*, 2014. Vol. 1872. No. 10. P. 1840–1872.
- [18] *Abdi, Hervé* Partial Least Squares (PLS) Regression // *Encyclopedia for research methods for the social sciences*, 2003. P. 792–795.
- [19] *Caruana, Rich and de Sa, Virginia R.* Benefitting from the Variables that Variable Selection Discards // *Journal of Machine Learning Research*, 2003. Vol. 3. No. 7-8. P. 1245–1264.

- 257 [20] *Mises R. V., Pollaczek-Geiringer H.* Praktische Verfahren der Gleichungsauflösung // ZAMM-  
258 Journal of Applied Mathematics and Mechanics/Zeitschrift für Angewandte Mathematik und  
259 Mechanik, 1929. Vol. 9. No. 1. P. 58–77.
- 260 [21] *Li J. et al.* Feature selection: A data perspective // arXiv preprint arXiv:1601.07996, 2016.