

На правах рукописи



Беляев Михаил Геннадьевич

**Построение аппроксимации многомерных
зависимостей по выборкам с факторным
планом**

05.13.17 — Теоретические основы информатики

АВТОРЕФЕРАТ
диссертации на соискание учёной степени
кандидата физико-математических наук

Москва — 2013

Работа выполнена в Федеральном государственном бюджетном учреждении науки Институте проблем передачи информации им. А.А. Харкевича Российской академии наук (ИППИ РАН)

Научный руководитель:

доктор технических наук, академик РАН

Кулешов Александр Петрович

Официальные оппоненты:

доктор физико-математических наук, профессор

Магарил-Ильяев Георгий Георгиевич,

профессор кафедры Общих проблем управления механико-математического факультета Московского государственного университета имени М.В. Ломоносова

кандидат физико-математических наук, доцент

Гасников Александр Владимирович,

доцент кафедры Математических основ управления

Московского физико-технического института

(государственного университета)

Ведущая организация:

Федеральное государственное бюджетное учреждение науки Институт прикладной математики им. М.В. Келдыша Российской академии наук

Защита состоится 16 декабря 2013 г. в 16.30 на заседании диссертационного совета Д 212.156.04 при Московском физико-техническом институте (ГУ) по адресу: 141700, г. Долгопрудный, Московская обл., Институтский пер., д. 9, ауд. 204 Нового корпуса.

С диссертацией можно ознакомиться в библиотеке Московского физико-технического института (государственного университета).

Автореферат разослан 14 ноября 2013 года.

Ученый секретарь

диссертационного совета Д 212.156.04

к.ф.-м.н.



Стрыгин Л. В.

Общая характеристика работы

Актуальность темы. В задачах инженерного проектирования часто необходимо работать со сложными физическими явлениями. Эти явления, как правило, моделируются с помощью систем дифференциальных уравнений, которые в большинстве случаев не решаются аналитически. Для нахождения решения применяются численные методы, использование которых существенно повышает время моделирования. С целью уменьшения временных затрат на этапе предварительного проектирования используют математические модели, построенные с помощью методов анализа данных по заданной обучающей выборке, которая состоит из пар “точка” (описание объекта и условий функционирования, входной вектор) – “значение функции в точке” (характеристика объекта, выходной вектор). Это позволяет получить лишь приближенное описание физического явления, но существенно понижает время моделирования.

В таких задачах все переменные (компоненты входного вектора) можно разбить на несколько групп с разным физическим смыслом, причем при проведении экспериментов бывает целесообразно изменять только переменные из одной группы при фиксированных значениях переменных из второй группы. Например, при исследовании аэродинамики самолета часто используется следующая постановка эксперимента: для заданной геометрии крыла (первая группа переменных) проводится ряд экспериментов при разных условий полета (вторая группа переменных), а затем для каждой следующей геометрии используется тот же самый набор условий полета. Это означает, что план эксперимента будет полным факторным (множество точек обучающей выборки будет содержать все возможные попарные комбинации значений переменных из первой и второй групп). Если же часть точек полного плана пропущена (эксперимент не завершен, генератор данных не может выдать значения в некоторых точках и т.д.), то план эксперимента будет неполным факторным.

При выборе плана эксперимента, как правило, принимаются во внимание знания из предметной области, которые говорят о неравнозначной зависимости функции от различных групп переменных. Для тех групп переменных, которые вносят существенный вклад в изменчивость функции, используются большие мощности факторов, чем для переменных, от которых функция зависит достаточно

гладко. Построенная по данным модель должна соответствовать этим представлениям, то есть необходим механизм контроля гладкости модели относительно различных групп переменных.

Задача построения аппроксимации по выборкам с полным факторным планом изучается достаточно давно, в том числе большое количество исследований посвящено вопросу создания вычислительно эффективных алгоритмов аппроксимации. С. de Boor¹ рассмотрел задачу интерполяции по полному факторному плану с одномерными факторами и предложил эффективный итеративный алгоритм нахождения решения.

P. Dierckx² поставил задачу построения бикубической аппроксимации с явным контролем гладкости. Задача свелась к решению системы линейных уравнений вида $(M \otimes N + A \otimes I + I \otimes B)x = y$, которая эффективно не решалась (символ \otimes обозначает произведение Кронекера, т.е. тензорное произведение матриц). Была предложена модифицированная штрафная функция, которая теряла содержательный смысл, но позволяла строить модель эффективно. Использование такого штрафа приводило к системе вида $(M+A) \otimes (N+B)x = y$, которая легко решалась с помощью использования базовых свойств произведения Кронекера.

B. Marx и P. Eilers³ рассмотрели важный случай штрафа в двумерной задаче аппроксимации, а именно ввели двумерный параметр регуляризации, который определяет гладкость модели вдоль разных направлений.

Наконец, L. Xiao с соавторами⁴ предложили подход, позволяющий строить аппроксимацию для многомерной задачи с одномерными факторами и явно контролировать изменчивость модели с помощью вектора параметров регуляризации. Из недостатков следует

¹De Boor C. Efficient computer manipulation of tensor products // ACM Transactions on Mathematical Software. 1979. V. 5, no. 2. P. 173–182.

²Dierckx P. A fast algorithm for smoothing data on a rectangular grid while using spline functions // SIAM Journal on Numerical Analysis. 1982. V. 19, no. 6. P. 1286–1304.

³Marx B. D., Eilers P. H. C. Multidimensional penalized signal regression // Technometrics. 2005. V. 47, no. 1. P. 13–22.

⁴Xiao L., Li Y., Ruppert D. Fast bivariate P-splines: the sandwich smoother // Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2013. V. 75, no. 3. P. 577–599.

отметить способ выбора штрафной функции: он в первую очередь осуществляется таким образом, чтобы получить низкую вычислительную сложность, а не исходя из содержательных соображений (аналогичный прием использовался в работе R. Dierckx, указанной выше).

Из методов, лежащих несколько в стороне от тематики описанных подходов, следует отметить работу R. Paulo⁵ по регрессии на гауссовских процессах, в которой предполагается, что данные имеют полный факторный план (который, в том числе, может содержать многомерные факторы). Работа посвящена разработке ковариационной функции специального вида, учитывающей этот вид данных. Однако, такой подход не позволяет контролировать гладкость модели.

Результатов по неполному факторному плану существенно меньше. В частности, G. Pisinger и A. Zimmermann⁶ рассмотрели задачу построения бикубической интерполяции по неполной двумерной решетке (без какого-либо штрафа на изменчивость модели) и предложили решение, основанное на использовании техники псевдообращения. Метод легко обобщается на многомерный случай, однако помимо отсутствия штрафа у него есть еще один существенный недостаток: необходимо решать систему из M уравнений (M – число пропущенных точек). Эта система не имеет какой-то специальной структуры, что не позволяет использовать этот метод при большом числе пропущенных точек.

Кроме того, существует ряд вычислительно-эффективных методов аппроксимации, предназначенных для выборок произвольной структуры. Среди них следует отметить метод MARS⁷, основанный на разложении по словарю, сформированному с помощью тензорного произведения линейных сплайнов. Словарь строится жадным образом с помощью специального алгоритма включения-исключения. В этом методе, как и ряде других, для уменьшения вычислительной сложности заметно упрощается модель. В случае достаточно

⁵R. Paulo *Default Priors for Gaussian Processes* // The Annals of Statistics. 2005. V. 33, no. 2. P. 556–582.

⁶G. Pisinger, A. Zimmermann. *Linear least squares problems with data over incomplete grid* // BIT Numerical Mathematics. 2007. V. 47. P. 809–824.

⁷Friedman J. H. Multivariate adaptive regression splines // The Annals of statistics. 1991. V.19, no. 1. P. 1-67.

сложных зависимостей (и соответствующих больших объемов выборки) такое упрощение может существенно повлиять на качество аппроксимации.

Таким образом, востребовано решение задачи аппроксимации для выборок с факторным планом эксперимента, в том числе неполным. Разработанные на данные момент методы для полных факторных планов могут решать такую задачу, однако требуется обобщение класса штрафных функций. В случае неполного факторного плана задача построения аппроксимации с контролем гладкости не решена. Соответственно, **актуальна** задача создания метода построения аппроксимации многомерной зависимости, учитывающего структуру плана эксперимента. **Целью** данной работы является разработка вычислительно-эффективного метода аппроксимации многомерных зависимостей для выборок с факторным планом, который позволял бы контролировать изменчивость модели относительно различных групп переменных. Для достижения поставленной цели необходимо было решить следующие **задачи**:

1. Формализовать задачу построения аппроксимации, предложив класс функций и штраф на изменчивость, учитывающие особенности факторного плана эксперимента.
2. Разработать вычислительно-эффективный метод нахождения оптимального решения задачи аппроксимации для выборок с полным факторным планом.
3. Обобщить метод нахождения оптимального решения задачи аппроксимации на случай выборок с неполным факторным планом.
4. Предложить методы выбора вектора параметров регуляризации для разработанного способа решения задачи аппроксимации.

Общая методика исследования. Для решения поставленных задач в работе используются методы матричной и тензорной алгебры, теории оптимизации, статистического анализа данных и вычислительной математики.

Научная новизна работы состоит в том, что в ней впервые

- разработан новый вычислительно-эффективный метод решения задачи аппроксимации, учитывающий неоднородную структуру факторного плана;
- предложен новый способ ограничения изменчивости моделей, являющихся линейным разложением по тензорному произведению словарей функций меньшей размерности;
- получено вычислительно-эффективное оптимальное решение задачи аппроксимации для выборок с неполным факторным планом.

Практическая значимость диссертационной работы определяется широким применением разработанных алгоритмов, которые были реализованы в программном продукте Macros компании Datadvice, для решения ряда задач в аэрокосмической отрасли.

Основные положения, выносимые на защиту:

1. Разработан новый метод решения задачи аппроксимации для случая факторных планов.
2. Предложен класс штрафных функций, учитывающий структурные особенности факторного плана эксперимента.
3. Для нахождения оптимального решения задачи аппроксимации в случае полного факторного плана эксперимента разработана явная вычислительно-эффективная формула.
4. Для нахождения оптимального решения задачи аппроксимации в случае неполного факторного плана эксперимента предложен вычислительно-эффективный итеративный метод.

Апробация работы. Результаты работы докладывались и обсуждались на следующих российских и международных конференциях:

- International workshop on advances in predictive modeling and optimization (2013, Берлин, Германия);
- 9-я Международная конференция “Интеллектуализация обработки информации” (2012, Будва, Черногория);

- 15-я Всероссийская конференция “Математические методы распознавания образов” (2011, Петрозаводск, Россия);
- Конференции молодых ученых “Информационные Технологии и Системы” (2011, Геленджик, Россия; 2012, Петрозаводск, Россия; 2013, Калининград, Россия);
- Third International Workshop on Surrogate Modelling and Space Mapping For Engineering Optimization (2012, Рейкьявик, Исландия);
- 53-я, 54-я и 55-я научные конференции Московского физико-технического института (2010-2012, Долгопрудный, Россия).

Кроме того, полученные в работе результаты обсуждались на научных семинарах лаборатории структурных методов анализа данных в предсказательном моделировании МФТИ (2012, 2013).

Публикации. Основные результаты по теме диссертации изложены в 6 печатных изданиях, 3 из которых изданы в журналах, рекомендованных ВАК ([1]-[3]), 3 — в тезисах докладов. Основные результаты изложены в работах [1], [2], написанных без соавторов, и совместной работе [3], в которой вклад диссертанта был определяющим.

Объем и структура работы. Диссертация состоит из введения, трех глав, заключения и приложения. Полный объем диссертации **98 страниц**, включая **7 рисунков**. Список литературы содержит **89 наименований**.

Диссертационная работа была выполнена при поддержке лаборатории структурных методов анализа данных в предсказательном моделировании МФТИ, грант правительства РФ дог. 11.G34.31.0073.

Содержание работы

Во **введении** описывается концепция использования моделей, построенных по данным, и формулируются характерные особенности плана эксперимента в задачах аппроксимации, возникающих в инженерном проектировании. В соответствии с описанными особенностями ставятся цели и формулируются задачи исследования.

Первая глава посвящена постановке задачи аппроксимации. В **первом** разделе формулируется задача аппроксимации по конеч-

ному набору данных в заданном классе функций. Назовем обучающей выборкой S совокупность конечного множества точек $\Sigma = \{x_i \in X, X \subset R^d\}_{i=1}^N$ и значений некоторой неизвестной функции $g : X \rightarrow R^1$ в точках из этого множества:

$$S = \{x_i \in X, y_i = g(x_i)\}_{i=1}^N = \{\Sigma, g(\Sigma)\}.$$

Задача (Аппроксимации). Пусть заданы обучающая выборка S , некоторый класс функций F и штрафная функция $P_\lambda : F \rightarrow R_+^1$. Необходимо найти $f^* \in F$, минимизирующую сумму выборочной функции ошибки и штрафа:

$$f^* = \operatorname{argmin}_{f \in F} \sum_{x \in \Sigma} (g(x) - f(x))^2 + P_\lambda(f).$$

Для конкретной постановки задачи аппроксимации необходимо задать класс функций F и штрафную функцию $P_\lambda(f)$. В следующих разделах мы выберем их, принимая во внимание особенности плана эксперимента Σ .

Во **втором** разделе вводится определение факторного плана эксперимента и детально описываются особенности плана эксперимента в задачах инженерного проектирования. Назовем факторами следующие множества

$$\sigma_k = \{x_{i_k}^k \in X_k\}_{i_k=1}^{n_k}, \quad X_k \subset R^{d_k}, \quad k = 1, \dots, K.$$

Под полным факторным планом будем понимать множество точек обучающей выборки вида $\Sigma = \sigma_1 \times \dots \times \sigma_K \stackrel{\text{def}}{=} \Sigma_{full}$ ($d = \sum d_k$). В случае неполного факторного плана множество точек обучающей выборки Σ не равно множеству Σ_{full} , а составляет его достаточно большое подмножество.

В **третьем** разделе рассматривается вопрос выбора класса функций F . Построим словарь функций Δ_{tensor} с помощью тензорного произведения заданных словарей $\Delta_k = \{\psi_{j_k}^k\}_{j_k=1}^{p_k}$ функций меньшей размерности:

$$\Delta_{tensor} = \{\psi_{j_1}^1 \otimes \psi_{j_2}^2 \otimes \dots \otimes \psi_{j_K}^K, \{j_k = 1, \dots, p_k\}_{k=1}^K\}.$$

Будем искать решение задачи в классе функций F_{tensor} , который состоит из всех возможных линейных комбинаций функций из Δ_{tensor} : $f(x) = \sum_{j_1, j_2, \dots, j_K} a_{j_1, j_2, \dots, j_K} \psi_{j_1}^1(x^1) \psi_{j_2}^2(x^2) \dots \psi_{j_K}^K(x^K)$. Перешипим это выражение с помощью тензорных операций:

$$f(x) = \mathcal{A} \otimes_1 \psi^1(x^1) \otimes_2 \psi^2(x^2) \dots \otimes_K \psi^K(x^K).$$

где $x = (x^1, x^2, \dots, x^K)$, $\vec{\psi}^k(x^k)$ — это вектор-столбец значений функций из словаря Δ_k в точке x^k , а \mathcal{A} — тензор коэффициентов разложения по словарю Δ_{tensor} размера $n_1 \times \dots \times n_K$. Символ \otimes_1 используется для обозначения умножения тензора на матрицу вдоль направления⁸. Для двумерных тензоров (т.е. матриц) эта операция сводится к умножению слева или справа: $\mathcal{A} \otimes \mathcal{B} = \mathcal{B}^T \mathcal{A}$, $\mathcal{A} \otimes \mathcal{B} = \mathcal{A} \mathcal{B}$.

Четвертый раздел посвящен выбору штрафной функции $P_\lambda(f)$, которая позволила бы контролировать гладкость модели относительно различных групп переменных. Рассмотрим сначала способ ограничения изменчивости модели вдоль срезов.

Определение. Будем называть срезами модели (функции) $f(x)$ вдоль переменных x^1 следующее множество функций

$$\{f_{i_2, \dots, i_K}^1(x^1) = f(x^1, x_{i_2}^2, \dots, x_{i_K}^K); i_2 = \overline{1, n_2}, \dots, i_K = \overline{1, n_K}\}.$$

где $f(x^1, x_{i_2}^2, \dots, x_{i_K}^K)$ функция от x^1 при зафиксированных переменных $x^2 = x_{i_2}^2, \dots, x^K = x_{i_K}^K$.

Определение естественным образом обобщается на срезы вдоль переменных из нескольких факторов.

Мы будем использовать штраф, основанный на ограничении изменчивости модели вдоль срезов. Рассмотрим в качестве примера штраф для срезов вдоль x^1 . Под изменчивостью будем понимать интегральную норму второй производной, что является весьма распространенным подходом и используется, например, в thin-plate сплайнах⁹. Введем обозначения для переменных, составляющих этот фактор: $x^1 = (x^{1,1}, \dots, x^{1,d_1})$. Тогда определим штраф

⁸ см., например, Kolda T. G., Bader B. W. Tensor decompositions and applications // SIAM Review. 2009. V. 51. №. 3. P. 455-500.

⁹ Wood S. N. Thin plate regression splines // Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2003. V. 65, no. 1. P. 95–114.

$p_{e_1}(f)$ как сумму по всем срезам и всем возможным частным производным порядка 2 (индекс e_1 описывается ниже):

$$\begin{aligned} p_{e_1}(f) &= \sum_{i_2, \dots, i_K} \sum_{q=1}^{d_1} \sum_{r=1}^{d_1} \left\| \frac{\partial^2 f_{i_2, \dots, i_K}^1(x^1)}{\partial x^{1,q} \partial x^{1,r}} \right\|^2 = \\ &= \sum_{i_2, \dots, i_K} \sum_{q=1}^{d_1} \sum_{r=1}^{d_1} \int_{X_1} \left(\frac{\partial^2 f_{i_2, \dots, i_K}^1(x^1)}{\partial x^{1,q} \partial x^{1,r}} \right)^2 dx^{1,1} \dots dx^{1,d_1}; \end{aligned}$$

где $e_1 = (1, 0, \dots, 0) \in \Theta$ является бинарным вектором длины K , $\Theta \stackrel{\text{def}}{=} \{0, 1\}^K$, а $p_{e_1}(f)$ — скаляр, поскольку срезы не зависят от x^2, \dots, x^K , а по x^1 проводится интегрирование.

Штраф можно обобщить на случай среза вдоль переменных из нескольких факторов. Введем бинарный вектор $\theta \in \Theta \setminus 0$, ненулевые компоненты которого будут определять штрафуемые факторы. Возьмем в качестве меры изменчивости все возможные суммы смешанных частных производных порядка 2 $\sum \theta_k$ (по 2 переменные из каждого штрафуемого фактора и $\sum \theta_k$ таких факторов). В примере выше штрафуется изменчивость только по переменным из x^1 , поэтому $\theta = e_1 = (1, 0, \dots, 0)$. Можно показать, что для произвольного θ штраф $p_\theta(f)$ является квадратичным и представим в виде

$$\begin{aligned} p_\theta(f) &= \langle \mathcal{A} \otimes_1 (\theta_1 \Omega_1 + (1 - \theta_1) \Psi_1^\top \Psi_1) \dots \otimes_K (\theta_K \Omega_K + (1 - \theta_K) \times \\ &\quad \times \Psi_K^\top \Psi_K, \mathcal{A}) \rangle = \text{vec}(\mathcal{A})^\top \bigotimes_{k=1}^K (\theta_k \Omega_k + (1 - \theta_k) \Psi_k^\top \Psi_k) \text{vec}(\mathcal{A}), \end{aligned} \quad (1)$$

где

- каждая матрица Ω_k (размера $p_k \times p_k$) — это сумма интегральных норм произведения вторых частных производных функций словаря Δ_k ,
- $\langle \mathcal{A}, \mathcal{B} \rangle = \sum_{i_1, \dots, i_K} a_{i_1, \dots, i_K} b_{i_1, \dots, i_K}$ — скалярное произведение тензоров,
- $\bigotimes_{k=1}^K$ используется для сокращенной записи произведения Кронекера K матриц,

- $\text{vec}(\mathcal{A})$ означает переупорядочивание элементов тензора в вектор-столбец.

Рассмотрим класс штрафных функций, который состоит из взвешенных сумм $p_\theta(f)$ для всех возможным $\theta \in \Theta \setminus 0$, причем веса $\gamma_\theta(\lambda)$ слагаемых этой суммы параметризованы вектором параметров регуляризации $\lambda \geq 0$ длины K :

$$\mathbb{P}_\lambda = \left\{ P_\lambda(f) : P_\lambda(f) = \sum_{\theta \in (\Theta \setminus 0)} \gamma_\theta(\lambda) p_\theta(f), \gamma_\theta(\lambda) \geq 0 \right\}. \quad (2)$$

В алгоритме мы воспользуемся штрафной функцией $P_\lambda^1(f)$, в которой $\gamma_\theta(\lambda) = \lambda_k$ для случая $\theta_k = 1$, $\sum_{l \neq k} \theta_l = 1$ и $\gamma_\theta(\lambda) = 0$ для $\sum_{l \neq k} \theta_l \neq 1$:

$$P_\lambda^1(f) = \sum_{k=1}^K \lambda_k \sum_{i_l, l \neq k} \sum_{q=1}^{d_1} \sum_{r=1}^{d_1} \left\| \frac{\partial^2 f_{\{i_l, l \neq k\}}^k}{\partial x^{k,q} \partial x^{k,r}} \right\|^2.$$

Такой выбор $\gamma_\theta(\lambda)$ дает возможность с помощью вектора регуляризации λ явно определять вклад изменчивости по переменным из каждого фактора в общую функцию ошибки $R(f, \Sigma, g(\Sigma))$.

Вторая глава посвящена задаче формирования словарей Δ_k . **Первый** раздел содержит алгоритм построения аппроксимации для выборок с факторным планом. Рассмотрим этап построения словарей $\Delta_k, k = 1, \dots, K$ по заданной обучающей выборке S . Зададим словари $\Delta_k, k \neq l$ и рассмотрим задачу выбора словаря Δ_l . Выборочная функция ошибки в этом случае может быть переписана следующим образом:

$$Q(f, \Sigma, g(\Sigma)) = \text{trace} \left((\mathcal{Y}^{(l)} - \Psi_l \mathbf{B}_l)^T (\mathcal{Y}^{(l)} - \Psi_l \mathbf{B}_l) \right),$$

где $\mathcal{Y}^{(l)}$ обозначает развертку тензора вдоль направления l (переупорядочивание элементов в матрицу размера $n_l \times \prod_{k \neq l} n_k$). Задача минимизации Q по функциям $\{\psi_{j_l}^l\}_{j_l=1}^{p_l} = \Delta_l$ (точнее, по их параметрам) может быть заменена построением аппроксимации по выборке $S_l = \{\sigma_l, \mathcal{Y}^{(l)}\}$. Таким образом, словари $\Delta_k, k = 1, \dots, K$ предлагаются строить, решая задачу аппроксимации отдельно в каждом из

факторов. Следует отметить характерную особенность этой задачи: размерность выходных данных (равная $\prod_{k \neq l} n_k$) может существенно превышать не только размерность входных данных d_l , но и объем обучающей выборки n_l .

Во *втором* разделе рассматривается случай одномерных факторов, в которых в качестве словаря Δ_k мы будем использовать базис В-сплайнов. Раздел содержит краткое описание этого типа сплайнов.

Во *третьем* разделе исследуется случай многомерных факторов. Основное отличие от одномерных факторов заключается в том, что в этом случае нет таких естественных словарей функций как сплайны.

В *первом* подразделе задача поиска словаря рассматривается как задача построения аппроксимации с помощью линейного разложения по словарю сигмоидальных функций:

$$\hat{f}(x) = \sum_{j=1}^p \alpha_j \psi_j(\theta_j, x) + \alpha_0 = \psi(\Theta, x) \alpha,$$

где $\hat{f}(x)$ задается матрицей Θ параметров функций словаря и вектором α коэффициентов линейного разложения.

Во *втором* подразделе рассматриваются подзадачи этапа инициализации модели (выделения валидационной выборки и инициализации параметров функций словаря) и предлагаются алгоритмы решения этих подзадач. Экспериментальные результаты на искусственных многомерных функций, приведенные в приложении, демонстрируют преимущество предложенных процедур над стандартными подходами в смысле уменьшения средней ошибки аппроксимации и ее разброса.

Третий подраздел посвящен обучению модели, основанной на разложению по словарю параметрических функций. Рассмотрим выборочную функцию ошибки как функцию от Θ и $\vec{\alpha}$:

$$\tilde{Q}(\Theta, \alpha) = \left(y - \hat{f}(X, \Theta, \alpha) \right)^T \left(y - \hat{f}(X, \Theta, \alpha) \right) + \lambda \alpha^T \Omega \alpha.$$

Заметим, что зависимость функции ошибки \tilde{Q} от параметров функций словаря Θ нелинейна и довольно сложна. В то же время, за-

висимость \tilde{Q} от коэффициентов разложения α квадратична. При фиксированных Θ оптимальное значение α может быть найдено явно с как решение задачи обобщенной гребневой регрессии:

$$\alpha(\Theta) = \left(\Psi(\Theta)^T \Psi(\Theta) + \lambda \Omega \right)^{-1} \Psi(\Theta)^T y. \quad (3)$$

Рассмотрим новую целевую функцию $\tilde{R}(\Theta) = \tilde{Q}(\Theta, \alpha(\Theta))$, где $\alpha(\Theta)$ вычисляется по явной формуле (3). В этом случае возникает зависимость $\alpha(\Theta)$, которую нужно учитывать при нахождении производных \tilde{R}_Θ и $\tilde{R}_{\Theta\Theta}$ модифицированной функции ошибки. Обозначим вектор невязок как $e = \hat{f}(X, \Theta, \alpha) - y$, а матрицу производных модели по Θ , подсчитанных в точках обучающей выборки, как J .

Предложение 1. Градиент и гессиан функции ошибки \tilde{R} выражаются следующими формулами:

$$\begin{aligned} \tilde{R}_\Theta &= \tilde{Q}_\Theta = e^T J; \\ \tilde{R}_{\Theta\Theta} &= J^T J + \hat{f}_{\Theta\Theta} \otimes_1 e^T - (\Psi_\Theta \otimes_1 e^T + J^T \Psi) \times \\ &\quad \times (\Psi^T \Psi + \lambda \Omega)^{-1} (\Psi_\Theta \otimes_1 e^T + J^T \Psi)^T, \end{aligned}$$

$$где \Psi_\Theta \otimes_1 e^T = \sum_{i=1}^N e_i \psi_\Theta(x_i).$$

Таким образом, учтена нелинейная зависимость $\alpha(\Theta)$ и получены явные формулы для первых и вторых производных функции ошибки $\tilde{R}_{\Theta\Theta}$. Рассмотрим вместо гессиана $\tilde{R}_{\Theta\Theta}$ некоторое его приближение H , которое позволяет строить выпуклые локально-квадратичные модели.

Предложение 2. Матрица $H \stackrel{\text{def}}{=} J^T J - (J^T \Psi) (\Psi^T \Psi + \lambda \Omega)^{-1} (J^T \Psi)^T$ неотрицательно определена для любого $\lambda \geq 0$.

Предложенная функция ошибки позволяет существенно сократить количество параметров модели и, как следствие, уменьшить вычислительную сложность алгоритма. В приложении находится описание ряда вычислительных экспериментов, демонстрирующих повышение точности аппроксимации как при использовании только модифицированной функции ошибки, так и в комбинации с описанными выше процедурами решения подзадач этапа инициализации.

Третья глава посвящена решению задачи аппроксимации для выборок с факторным планом эксперимента. В *первом* разделе обсуждается задача поиска оптимального решения в выбранном классе функций F_{tensor} с штрафом $P_\lambda(f) \in \mathbb{P}_\lambda$. Показано, что задача сводится к минимизации квадратичной функции ошибки по элементам тензора коэффициентов разложения \mathcal{A} . Особенностью этой квадратичной функции является ее высокая размерность и неразреженная структура матрицы гессиана. В общем случае для минимизации функций такого рода требуется $O(P^3 + NP^2)$ операций (напомним, что N — число наблюдений, а P — объем словаря). В некоторых специальных случаях известны более эффективные способы.

В случае полного плана мы предложим явную формулу для оптимального решения, которая в случае $n_k \geq p_k$ имеет сложность $O(N_{full} \sum n_k)$, где n_k — число точек в факторе σ_k , $N_{full} = \prod n_k$ — мощность полного факторного плана, а p_k — объем словаря Δ_k . В случае неполного плана мы воспользуемся итеративным подходом, который позволит получить сложность $O(\min(N_{full} - N + 1, N + 1, P) \times N_{full} \sum n_k)$.

Сформулируем предположение, которым мы будем пользоваться в этой главе.

Предположение (Положительной Определенности). *Пусть все матрицы $\Psi_k^T \Psi_k$ и Ω_k ($k = 1, \dots, K$) положительно определены.*

Это техническое условие, которое фактически означает, что словари Δ_k не вырождены.

Во *втором* разделе исследуется задача нахождения оптимального решения в случае полного факторного плана. Следующий результат дает достаточное условие на существование и единственность решения.

Предложение 3. *Если если выполняется предположение ПО, то решение задачи аппроксимации по выборке с полным факторным планом в классе F_{tensor} со штрафом $P_\lambda(f) \in \mathbb{P}_\lambda$ существует и единственно.*

Как было указано выше, решение задачи аппроксимации соответствует минимизации квадратичной формы. В случае полного

плана она будет записана как

$$R(\mathcal{A}) = \left[\text{vec}(\mathcal{A})^T \Omega \text{vec}(\mathcal{A}) + (\text{vec}(\mathcal{Y}) - \Psi \text{vec}(\mathcal{A}))^T (\text{vec}(\mathcal{Y}) - \Psi \text{vec}(\mathcal{A})) \right],$$

где матрица Ω определяется из (2) по аналогии с (1), то есть является взвешенной суммой произведений Кронекера.

Предлагается следующий способ нахождения решения задачи. С помощью методов линейной алгебры строится преобразование гессиана квадратичной функции ошибки $R(\mathcal{A})$.

Лемма 1. *Если справедливо предположение ПО, то гессиан $\Omega + \Psi^T \Psi$ представим в виде*

$$\Omega + \Psi^T \Psi = V D V^T,$$

где $V = \bigotimes_{k=1}^K V_k$, а D — положительно определенная диагональная матрица.

Следующая теорема дает оптимальное решение задачи аппроксимации для полного факторного плана в явном виде.

Теорема 1. *Если справедливо предположение ПО, то оптимальный тензор коэффициентов разложения по словарю Δ_{tensor} вычисляется по формуле:*

$$\mathcal{A}^* = \left((\mathcal{Y} \otimes_1 \Psi_1 V_1^{-T} \dots \otimes_K \Psi_K V_K^{-T}) \odot D^{-1} \right) \otimes_1 V_1^{-1} \dots \otimes_K V_K^{-1}. \quad (4)$$

где D определяется с помощью равенства $\text{vec}(D) = \text{diag}(D)$, а символ \odot подразумевает поэлементное умножение тензоров или матриц.

Предложение 4. *Сложность вычисления оптимального тензора коэффициентов разложения \mathcal{A} по формуле (4) равна*

$$O \left(\sum_{k=1}^K n_k p_k^2 + N \sum_{k=1}^K n_k \prod_{l=1}^k p_l / n_l \right).$$

Введем обозначение $T = N \sum_{k=1}^K n_k \prod_{l=1}^k p_l / n_l$ для второго слагаемого, которое в большинстве случаев и вносит основной вклад

в вычислительную сложность. T операций необходимо для последовательного умножении тензора \mathcal{Y} на матрицы $\Psi_k V_k^{-T}$ вдоль всех направлений.

Сравним вычислительную сложность полученной формулы со сложностью вычисления оптимальных коэффициентов для произвольной квадратичной задачи.

Пример 1. Рассмотрим случай интерполирующего словаря и плана эксперимента с одинаковым числом уровней во всех факторах ($p_k = n_k = N^{1/K} = P^{1/K}$). Тогда сложность (4) составит $O(P \sum p_k) = O(P^{1+1/K})$ вместо $O(P^3)$ для стандартного подхода. Для типичной ситуации $n = 10$, $K = 6$ эти оценки составят 10^7 и 10^{18} соответственно.

В *третьем* разделе рассматривается задача поиска оптимального решения для выборок с неполных факторным планом. Для сохранения возможности использовать эффективный аппарат тензорной алгебры расширим неполный факторный план до полного. Дополним множество значений $\{y_i\}_{i=1}^N$ нулями для всех точек из $\Sigma_{full} \setminus \Sigma$. Это позволит нам переписать расширенное множество значений в виде тензора. Обозначим этот тензор как $\tilde{\mathcal{Y}}$, а расширенную матрицу регрессоров, посчитанную во всех точках из Σ_{full} , — как $\tilde{\Psi}$. Теперь введем модифицированную взвешенную функцию ошибки $\tilde{R}(\mathcal{A})$:

$$\tilde{R}(\mathcal{A}) = \text{vec}(\mathcal{A})^T \Omega \text{vec}(\mathcal{A}) + (\text{vec}(\tilde{\mathcal{Y}}) - \tilde{\Psi} \text{vec}(\mathcal{A}))^T W (\text{vec}(\tilde{\mathcal{Y}}) - \tilde{\Psi} \text{vec}(\mathcal{A})).$$

где W — диагональная матрица весов, которая формируется следующим образом:

- $W_{i,i} = 1$, если $x_i \in \Sigma$ (i -я точка полного плана Σ_{full} представлена в обучающей выборке);
- $W_{i,i} = 0$, если $x_i \in (\Sigma_{full} \setminus \Sigma)$ (i -я точка была пропущена).

При таком выборе весовой матрицы W справедливо $\tilde{R}(\mathcal{A}) \equiv R(\mathcal{A})$ при любом способе заполнения пропущенных значений в тензоре \mathcal{Y} . Таким образом, задача минимизации $R(\mathcal{A})$ может быть заменена эквивалентной задачей минимизации функции $\tilde{R}(\mathcal{A})$, которая также является квадратичной по коэффициентам разложения \mathcal{A} .

Достаточное условие на существование и единственность минимума функции $\tilde{R}(\mathcal{A})$ эквивалентно случаю полного факторного плана.

Предложение 5. *Если справедливо предположение ПО, то решение задачи аппроксимации по выборке с неполным факторным планом в классе F_{tensor} со штрафом $P_\lambda(f) \in \mathbb{P}_\lambda$ существует и единственно.*

Ниже приведены три леммы, которые позволяют ввести такую замену переменных, что спектр гессиана функции ошибки будет содержать много совпадающих собственных чисел.

Лемма 2. *Гессиан H_A функции $\tilde{R}(\mathcal{A})$ может быть представлен в виде*

$$H_A = V \left(\tilde{D} + U^T W U \right) V^T,$$

где \tilde{D} — положительно определенная диагональная матрица, U раскладывается в произведение Кронекера ортогональных матриц размеров $n_k \times p_k$ $U = \bigotimes_{k=1}^K U_k$, а $V = \bigotimes_{k=1}^K V_k$.

Лемма 3. *Пусть $\dot{\mathcal{A}} = (\mathcal{A} \otimes_1 V_1 \dots \otimes_K V_K) \odot \tilde{D}^{0.5}$, где тензор \tilde{D} задается с помощью равенства $\text{vec}(\tilde{D}) = \text{diag}(\tilde{D})$.*

Определим квадратичную функцию $\dot{R}(\dot{\mathcal{A}})$ с помощью соотношения $\dot{R}(\dot{\mathcal{A}}) \equiv \tilde{R}(\mathcal{A})$. Тогда гессиан функции $\dot{R}(\dot{\mathcal{A}})$ имеет не более чем $N + 1$ различных собственных чисел.

Лемма 4. *Пусть $\ddot{\mathcal{A}} = (\mathcal{A} \otimes_1 V_1 \dots \otimes_K V_K) \odot D^{0.5}$.*

Определим квадратичную функцию $\ddot{R}(\ddot{\mathcal{A}})$ с помощью соотношения $\ddot{R}(\ddot{\mathcal{A}}) \equiv \tilde{R}(\mathcal{A})$. Тогда гессиан функции $\ddot{R}(\ddot{\mathcal{A}})$ будет иметь не более чем $N_{full} - N + 1$ различных собственных чисел.

Леммы 2, 3 и 4 позволяют найти оптимальное решение.

Теорема 2. *Если справедливо предположение ПО, то решение задачи минимизации $\tilde{R}(\mathcal{A})$ будет найдено с помощью метода сопряженных градиентов*

- не более чем за $\min(N_{full} - N + 1, N + 1)$ шагов;

- вычислительная сложность каждого шага не будет превосходить $O(T)$, где $T = N \sum_{k=1}^K n_k \prod_{l=1}^k p_l / n_l$;
- сложность дополнительных вычислений, которые выполняются однократно, составит $O(T + \sum p_k^2(n_k + p_k))$.

Отметим, что структура спектра гессиана позволяет получить меньшее число шагов метода сопряженных градиентов, а использование тензорной алгебры — снизить вычислительную сложность каждого шага.

Следствие. Общая вычислительная сложность алгоритма поиска оптимальных коэффициентов разложения \mathcal{A} составляет

$$O\left(\sum p_k^2(n_k + p_k) + T \times \min(N_{full} - N + 1, N + 1, P)\right).$$

Очевидно, что основным слагаемым в оценке сложности является $O(T \times \min(N_{full} - N + 1, N + 1, P))$. Напомним, что основная часть сложности вычисления оптимального решения для полного факторного плана составляет $O(T)$. Таким образом, сложность возрастает почти “непрерывно” при переходе от полного плана к неполному и затем увеличивается пропорционально количеству пропущенных точек.

Заметим, что любое множество Σ может быть представлено в виде неполного факторного плана (с огромным числом пропущенных точек). Естественно, что для произвольной выборки предложенный подход не может превосходить по эффективности стандартную формулу. При постановке задачи аппроксимации для неполного факторного плана мы ограничились формулировкой “ Σ составляет достаточно большое подмножество Σ_{full} ”. Получив точные оценки для вычислительной сложности формализуем понятие “достаточно большое подмножество”. Пусть α — это доля точек обучающей выборки в полном факторном плане, $N = \alpha N_{full}$. Рассмотрим распространенный вариант формирования словаря, при котором $P \approx N$, а $p_k \sim n_k$.

Предложение 6. Пусть объем словаря равен числу точек выборки $P = N = \alpha N_{full}$. Тогда отношение сложности предложенного

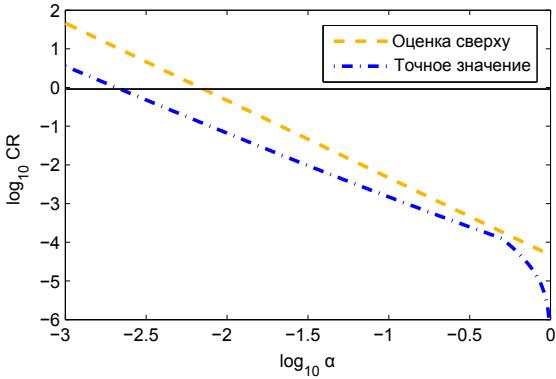


Рис. 1: Отношение вычислительных сложностей (CR) предложенного подхода и явной формулы

метода к сложности стандартного подхода будет равно

$$CR = \frac{\min(\alpha, 1 - \alpha) \sum_{k=1}^K n_k \alpha^{1/k}}{\alpha^3 \prod_{k=1}^K n_k} < \frac{\sum_{k=1}^K n_k}{\alpha^2 \prod_{k=1}^K n_k}.$$

Назовем критической долей точек α_{cr} такой объем обучающей выборки, при котором отношение вычислительных сложностей равно 1.

Пример 2. Рассмотрим случай $n_k = n$, подставим константы из примера 1 ($n = 10, K = 6$) и построим график зависимости отношения сложностей от α (см. рисунок) и ее верхней оценки, приведенной выше. Приведем некоторые характеристические значения, полученные из графика:

- критическая доля точек равна $\alpha_{cr} = 0.0024$;
- при $\alpha = 0.1$ (90% точек пропущено) сложность предложенного метода в 520 раз меньше сложности стандартного;
- при $\alpha = 0.9$ (10% точек пропущено) сложность предложенного метода в 129 000 раз меньше сложности стандартного.

Если же выбирать объем словаря $P = N_{full}$ (что необходимо в ситуации, когда Σ распределено существенно неравномерно в Σ_{full}), то предложенный способ будет эффективнее стандартного способа вычисления независимо от доли точек α .

Четвертый раздел посвящен выбору параметров регуляризации λ . Для полного факторного плана описывается вычислительно эффективный алгоритм выбора вектора λ , основанный на минимизации ошибки скользящего контроля¹⁰.

Предложение 7. *Ошибка скользящего контроля в случае полного факторного плана выражается в виде*

$$Q_{loo} = \left\langle \mathcal{Y} - \hat{\mathcal{Y}}, (\mathcal{Y} - \hat{\mathcal{Y}}) \odot \mathcal{L}^2 \right\rangle, \\ \mathcal{L} = \left(\mathcal{I} - \mathcal{D}^{-1} \otimes_1 (U_1^T \odot U_1^T) \dots \otimes_K (U_K^T \odot U_K^T) \right)^{-1}; \quad (5)$$

где $\hat{\mathcal{Y}} = \mathcal{A} \otimes_1 \Psi_1^T \dots \otimes_K \Psi_K^T$ – значения модели $f(\mathcal{A})$ в точках плана Σ_{full} , а \mathcal{I} – тензор размера $n_1 \times \dots \times n_K$.

Сложность формулы (5) равна $O(T)$, $T = N_{full} \sum_{k=1}^K n_k \prod_{l=1}^k p_l / n_l$.

Для случая неполного факторного плана предложен итеративный алгоритм, основанный на заполнении пропущенных значений с помощью решения задачи аппроксимации по неполному плану и оценке ошибки скользящего контроля по модифицированной формуле:

$$\tilde{Q}_{loo} = \left\langle (\mathcal{Y} - \hat{\mathcal{Y}}) \odot \mathcal{W}, (\mathcal{Y} - \hat{\mathcal{Y}}) \odot \mathcal{L}^2 \right\rangle.$$

Низкая сложность подсчета ошибки скользящего контроля позволяет использовать ее для выбора оптимального вектора регуляризации λ как для случая полного факторного плана, так и для случая неполного.

Пятый раздел посвящен задаче построения аппроксимации по двум выборкам разноточных данных, каждая из которых имеет факторный план. Кроме того, предполагается, что заданы априорные скалярные веса, определяющие степень доверия к каждому из

¹⁰Fukunaga K., Hummels D. M. Leave-one-out procedures for nonparametric error estimates // Pattern Analysis and Machine Intelligence, IEEE Transactions on. 1989. V. 11, no. 4. P. 421-423.

источников данных. В разделе показано, что решение этой задачи может быть сведено к задаче построения аппроксимации по неполному факторному плану, но с использованием весов $W_{i,i}$, отличных от {0, 1}.

В **заключении** приведены основные результаты работы.

Приложение состоит из двух разделов. *Первый* раздел содержит описание тех элементов теории тензорной алгебры, которые используются в работе. Во *втором* разделе изложены результаты экспериментального сравнения предложенного подхода с другими методами решения задачи аппроксимации на выборках данных, как построенных с помощью искусственных функций, так и полученных из прикладных задач инженерного проектирования.

Публикации по теме диссертации

1. Беляев М.Г. Аппроксимация данных, порожденных декартовым произведением // Труды МФТИ. 2013. Т. 5, № 3. С. 11–23.
2. Беляев М.Г. Аппроксимация многомерных зависимостей по структурированным выборкам // Искусственный интеллект и принятие решений. 2013. № 3. С. 24–39.
3. Belyaev M., Burnaev E. Approximation of a multidimensional dependency based on a linear expansion in a dictionary of parametric functions // Informatics and its Applications. 2013. Vol. 7, no. 3. P. 114–125.
4. Беляев М.Г. Аппроксимация и интерполяция на основе тензорного произведения параметрических словарей // Труды конференции Информационные Технологии и Системы. 2012. С. 32–40.
5. Surrogate Modeling of Stability Constraints for Optimization of Composite Structures / S. Grihon, E. Burnaev, M. Belyaev et al. // Surrogate-Based Modeling and Optimization. Engineering applications / Ed. by S. Koziel, L. Leifsson. 2013. P. 359–391.
6. Беляев М.Г. Аппроксимация зашумленных данных, имеющих структуру декартова произведения // Сборник докладов конференции ИОИ-9. 2012. С. 188–192.

Беляев Михаил Геннадьевич

Построение аппроксимации многомерных зависимостей по
выборкам с факторным планом

АВТОРЕФЕРАТ

Подписано в печать 14.11.2013.
Формат 60 × 84 1/16. Усл. печ. л. 1,0.
Тираж 100 экз. Заказ 360.

Федеральное государственное автономное образовательное
учреждение высшего профессионального образования
«Московский физико-технический институт
(государственный университет)»

Отдел оперативной полиграфии «Физтех-полиграф» 141700,
Московская обл., г. Долгопрудный, Институтский пер., 9.