На правах рукописи

Беляев Михаил Геннадьевич

Моделирование анизотропных зависимостей по выборкам с факторным планом эксперимента

05.13.18 — Математическое моделирование, численные методы и комплексы программ

АВТОРЕФЕРАТ
диссертации на соискание учёной степени
кандидата физико-математических наук

Москва — 2015

Работа выполнена в ФГБУ науки Институте проблем передачи информации им. А.А. Харкевича Российской академии наук

Научный руководитель:

Кулешов Александр Петрович,
доктор технических наук, академик РАН,
директор Института проблем передачи
информации им. А.А. Харкевича РАН

Официальные оппоненты:

Назин Александр Викторович,
доктор физико-математических наук, профессор,
ведущий научный сотрудник лаборатории № 7
Института проблем управления им. В.А. Трапезникова

Конушин Антон Сергеевич,
кандидат физико-математических наук, доцент,
доцент кафедры автоматизации систем вычислительных
комплексов факультета ВМК Московского государственного
университета имени М.В. Ломоносова

Ведущая организация: Федеральный исследовательский центр «Информатика и управление» Российской академии наук.

Защита состоится 14 декабря 2015 г. в 11:00 на заседании диссертационного совета Д 002.077.05 при Федеральном государственном бюджетном учреждении науки Институте проблем передачи информации им. А.А. Харкевича РАН, расположенном по адресу: 127051, г. Москва, Большой Каретный переулок, д.19 стр. 1.

С диссертацией можно ознакомиться в библиотеке ФГБУ науки Института проблем передачи информации им. А.А. Харкевича РАН и на сайте www.iitp.ru.

Автореферат разослан «___» ______ 2015 г.

Ученый секретарь диссертационного совета Д 002.077.05 при ИППИ РАН, доктор физико-математических наук	Цитович Иван Иванович

Общая характеристика работы

Актуальность темы.

В задачах инженерного проектирования получил распространение подход к построению моделей физических явлений по выборке данных, состоящей из пар вида (описание объекта и условий функционирования, входной вектор) – (характеристики объекта, выходной вектор). В силу особенностей генерации наборов подобных пар в приложениях часто используют полный факторный план эксперимента (множество входных векторов образует декартово произведение некоторых множеств, называемых факторами). Часть значений выходного вектора может быть невычислима, что делает план эксперимента неполным факторным. Для выборок с факторным планом эксперимента характерен высокий объем, который растет экспоненциально по размерности входного вектора. Еще одна типичная особенность подобных задач — это пространственная неоднородность породившей данные зависимости, бóльшую часть вариабельности которой обуславливает только несколько компонент входного вектора. Как правило, это априорное знание учитывается во время планирования эксперимента и для таких компонент используются бóльшие мощности факторов.

В ряде работ было показано, что использование тензорного произведения сплайнов позволяет строить вычислительно эффективные алгоритмы для полных факторных планов эксперимента. Среди таких работ отметим C. de Boor, 1979[1], где была рассмотрена задача интерполяции по полному факторному плану и предложен эффективный итеративный алгоритм нахождения решения, и P. Dierckx, 1982[2], где модель строилась с помощью бикубических сплайнов с явным контролем гладкости модели при помощи специального штрафа. Результатов по неполному факторному плану существенно меньше. В частности, в работе G. Pisinger, 2007[3], была рассмотрена задача интерполяции по неполному двумерному факторному плану и предложено решение, основанное на использовании техники псевдообращения.

Основным ограничением этих и многих других работ является неявное предположение о пространственной однородности моделируе-

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

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

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

мой зависимости. Важным шагом стала работа B. Marx, 2005[4], в которой для задач с двумерным полным факторным планом был рассмотрен новый вид штрафа, позволяющий учитывать изменчивость модели по каждой из переменных с различными весами. Вместе с тем, использование подобных методов не было обосновано теоретически.

Параллельно с использованием тензорного произведение сплайнов для задач с факторным планом эксперимента шло развитие многомерных сглаживающих сплайнов (например, thin-plate сплайнов) для решения задач с произвольным планом и изотропной моделируемой функцией. Два ключевых отличия этих двух направлений заключались в следующем: теоретические свойства thin-plate сплайнов были хорошо изучены (так, их асимптотическая оптимальность была показана в работе Utreras, 1988[5]), но высокая вычислительная сложность не позволяла использовать их для выборок большого объема.

Работа L. Xiao, 2013[6], объединила результаты из этих двух направлений. В работе был предложен метод моделирования анизотропных зависимостей по выборке данных с полным факторным планом эксперимента, доказана его асимптотическая оптимальность и разработан вычислительно эффективный алгоритм построения модели. Вместе с тем, оценка ошибки была получена только в двумерном случае, а предложенный алгоритм и использованная при доказательстве техника базируются на том факте, что план эксперимента полный факторный.

Таким образом, **актуально** исследование методов моделирования анизотропных зависимостей по выборкам с факторным планом эксперимента (в том числе неполным), изучение теоретических свойств полученных методов и создание вычислительно эффективных алгоритмов построения моделей по таким данным.

Целью диссертационной работы работы является создание метода моделирования пространственно неоднородных зависимостей по выборкам данных с факторным планом эксперимента. Для достижения поставленной цели необходимо было решить следующие **задачи**.

1. Предложить математическую постановку задачи и метод построения модели, учитывающие пространственную неоднородность

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

[5]Utreras F. Convergence rates for multivariate smoothing spline functions // Journal of approximation theory. 1988. V. 52, no. 1. P. 1–27.

[6]Xiao L., Li Y., Ruppert D. Fast bivariate P-splines: the sandwich smoother // J. of the Royal Statistical Society: Series B. 2013. V. 75, no. 3. P. 577–599.

моделируемых зависимостей.
2. Обосновать предложенный метод построения модели в рамках сформулированной математической постановки задачи.
3. Разработать вычислительно эффективный алгоритм построения модели по данным с факторным планом эксперимента.
4. Реализовать предложенный алгоритм в виде комплекса программ.

Научная новизна работы состоит в том, что в ней впервые получены следующие результаты.

1. Для размерности входного вектора более двух найдена оценка ошибки приближения пространственно неоднородных зависимостей с помощью тензорного произведения сплайнов с использованием штрафа на изменчивость модели.
2. Предложен вычислительно эффективный алгоритм построения модели с помощью тензорного произведения сплайнов по выборкам с неполным факторным планом эксперимента.
3. Построен вычислительно эффективный алгоритм выбора коэффициентов регуляризации для выборок с факторным планом эксперимента.
4. Решена задача построения модели по двум выборкам данных различной точности, план эксперимента каждой из которых является неполным факторным.

Практическая значимость диссертационной работы определяется широким применением разработанных алгоритмов для решения ряда задач моделирования в аэрокосмической отрасли. Разработанные алгоритмы также были включены в программные продукты компании Datadvance, которые используются в ряде крупнейших компаний аэрокосмической, машиностроительной и других высокотехнологичных отраслей.

Общая методика исследования. Для решения поставленных задач в работе используются методы функционального анализа, матричной и тензорной алгебры, теории оптимизации, теории приближения функций.

Основные положения, выносимые на защиту:

1. Разработанный метод является асимптотически оптимальным при надлежащем выборе параметров регуляризации.
2. Предложенный итеративный алгоритм построения модели обладает следующими свойствами:
   (a) Необходимо не более $min(n + 1, N - n + 1)$ итераций для

нахождения решения (n — объем выборки, N — объем соответствующего полного факторного плана эксперимента);

   (b) Вычислительная сложность одной итерации не превышает $O(N \sum n_k)$, где n_k — мощности факторов плана.

3. Вычислительная сложность предложенного метода подсчета ошибки скользящего контроля и компонент ее градиента не превышает $O(N \sum n_k)$.
4. Разработанный метод успешно использован для решения ряда задач моделирования в инженерном проектировании, в том числе задачи моделирования прочностных характеристик элементов обшивки самолета A350 – 900 и задачи моделирования аэродинамических характеристик суборбитального космического летательного аппарата.

Достоверность изложенных в работе результатов обеспечивается использованием строгих доказательств, основанных на хорошо изученных методах функционального анализа, теории оптимизации; совпадением элементов полученных скоростей сходимости с известными результатами в частном случае одномерных задач; совпадением вычислительной сложности предложенного алгоритма в частном случае полного факторного плана со сложностью известных алгоритмов для полного факторного плана; успешным использованием результатов исследования в задачах инженерного проектирования.

Апробация работы. Результаты работы докладывались и обсуждались на следующих российских и международных конференциях: International workshop on advances in predictive modeling and optimization (2013, Германия); 9-я международная конференция "Интеллектуализация Обработки Информации" (2012, Черногория); конференции молодых ученых "Информационные Технологии и Системы" (2012, 2013); 55-я научная конференции Московского физико-технического института (2012); 5th European Conference on Computational Mechanics (2014, Испания); 5th International Conference on Mechanical and Aerospace Engineering (2014, Испания). Кроме того, полученные в работе результаты докладывались и обсуждались на научных семинарах лаборатории структурных методов анализа данных в предсказательном моделировании МФТИ (2012 - 2014), семинарах сектора интеллектуального анализа данных и моделирования ИППИ РАН (2014-2015), семинаре "Теория автоматического управления и оптимизации" лаборатории 7 ИПУ РАН (2015).

Публикации. Результаты по теме диссертации изложены в 7 публикациях, 3 из которых изданы в журналах, рекомендованных ВАК ([1–3]), 4 — в тезисах докладов (включая индексируемые Web of Science тезисы [4], и индексируемые Scopus тезисы [5]). Основные результаты представлены в работах [1–3], написанных без соавторов.

Объем и структура работы. Диссертация состоит из введения, пяти глав и заключения. Полный объем диссертации составляет **120** стр., включая **10** рисунков и **4** таблицы. Список литературы содержит **75** наименований.

Содержание работы

Во **введении** описывается концепция использования моделей, построенных по данным, и формулируются характерные особенности планов эксперимента в задачах инженерного проектирования. В соответствии с описанными особенностями ставятся цели и формулируются задачи исследования.

Первая глава посвящена формальной постановке задачи. Рассмотрим построение оценки $\hat{f}$ функции многих переменных f_0 по заданной выборке данных $\{x_i \in \Omega, y_i = f_0(x_i) + \xi_i\}_{i=1}^n$, где Ω — некоторая область в R^d, $\Omega \subset H, H = [0,1]^d$, а ξ_i — независимые одинаково распределенные случайные величины с $\mathrm{E}\xi_i = 0$, $\mathrm{E}\xi_i^2 = \sigma^2$. Будем считать, что $f_0 \in W_2^{\mathbf{m}}$, где $W_2^{\mathbf{m}}$ — анизотропное пространство Соболева с $\mathbf{m} = (m_1, \ldots, m_d)$, m_k — натуральные числа. По определению $W_2^{\mathbf{m}}(\Omega)$ — это пространство локально суммируемых на Ω функций, имеющих на Ω обобщенные производные $D_k^{m_k} f(x)$ $(k = 1, \ldots, d)$, $D_k^{m_k} \equiv \partial^{m_k} / (\partial x^k)^{m_k}$, и конечную норму

$$\|f\|_{W_2^{\mathbf{m}}(\Omega)} = \|f\|_{L_2(\Omega)} + \sum_{k=1}^{d} \left\| D_k^{m_k} f \right\|_{L_2(\Omega)}.$$

Для упрощения обозначений далее в записи вида $W_2^{\mathbf{m}}(\Omega)$ будем опускать (Ω).

Для $k = 1, \ldots, d$ определим множества $\theta^k = \{\theta_1^k, \ldots, \theta_{n_k}^k\}$, где $0 < \theta_1^k < \cdots < \theta_{n_k}^k < 1$. Будем называть конечное множество $\Theta \subset [0,1]^d$ полным факторным планом с числами уровней $n_1, \ldots, n_d$, если

$$\Theta = \theta^1 \times \cdots \times \theta^d = \left\{ z_{i_1,\ldots,i_d} = \left(\theta_{i_1}^1, \ldots, \theta_{i_d}^d \right) \right\}_{i_1,\ldots,i_d=1}^{n_1,\ldots,n_d}.$$

Будем предполагать, что план эксперимента $X \equiv \{x_i\}_{i=1}^n$ заданной выборки данных есть пересечение полного факторного плана Θ с областью Ω. Если $\exists\, x \in \Theta : x \notin X$, то будем называть такой план неполным факторным.

Оценка $\hat{f}$ будет строиться с помощью тензорного произведения B-сплайнов. B-сплайны порядка m — это базис в пространстве кусочно-полиномиальных функций из $C^{m-1}([0,1])$, степень каждого полинома не выше m, а C^{m-1} — это пространство непрерывно дифференцируемых функций до порядка $m-1$ включительно. B-сплайны строятся с помощью m последовательных сверток индикаторной функции и обладают рядом удобным для анализа свойств, в частности обеспечивая наименьший среди сплайнов носитель.

Пусть $\eta_k = \{(j_k - 1)h_k\}_{j_k=1}^{p_k+1}$ — это равномерное разбиение отрезка $[0,1]$ с шагом $h_k = p_k^{-1}$ для всех $k = 1,\dots,d$. Будем переходить от одного полиномиального участка сплайнов к другому в точках множества η_k (p_k — число таких участков). Пусть $S_k([0,1])$ — это пространство сплайнов порядка m_k+1, заданных с помощью соответствующего разбиения η_k. Обозначим тензорное произведение сплайнов как $S_{\mathbf{m}}\big([0,1]^d\big)$, $S_{\mathbf{m}} = S_1 \otimes S_2 \cdots \otimes S_d$.

Введем d дискретно-непрерывных полунорм на $S_{\mathbf{m}}$, задаваемых следующим образом:

$$\|f\|_{\Theta,k} = \left(\frac{n_k}{N}\sum_{i_l, l\neq k}\int_0^1 \big(f(\theta^1_{i_1},\dots,x^k,\dots,\theta^d_{i_d})\big)^2 dx^k\right)^{1/2},$$

где $N = \prod_k n_k$ — мощность полного факторного плана Θ.

Будем строить оценку $\hat{f}$ с помощью минимизации на $S_{\mathbf{m}}$ функционала, включающего среднеквадратичную ошибку, вычисленную в точках выборки, и штрафное слагаемое:

$$\hat{f} = \operatorname*{argmin}_{f\in S_{\mathbf{m}}}\left(\frac{1}{n}\sum_{i=1}^{n}(y_i - f(x_i))^2 + \sum_{k=1}^{d}\lambda_k \|D_k^{m_k} f\|^2_{\Theta,k}\right). \tag{1}$$

При $d=1$ решение (1) совпадает с классическими сглаживающими сплайнами. В изотропных Соболевских пространствах стандартным обобщением сглаживающих сплайнов на случай $d>1$ являются thin-plate сплайны

$$\operatorname*{argmin}_{f\in W_2^m}\left(\frac{1}{n}\sum_{i=1}^{n}(y_i - f(x_i))^2 + \lambda\sum_{\|\alpha\|_1 = m}\|D^\alpha f\|_2^2\right),$$

где $\alpha = (\alpha_1,\dots,\alpha_d)$, $D^\alpha \equiv \partial^{\|\alpha\|_1}\big/\big((\partial x^1)^{\alpha_1}\dots(\partial x^d)^{\alpha_d}\big)$.

Предложенный функционал (1) является одним из возможных обобщений одномерных сглаживающих сплайнов на анизотропный многомерный случай. Как было отмечено во введении, в ряде задач инженерного проектирования моделируемые функции f_0 априори являются пространственно неоднородными, в силу чего при генерации выборки данных часто используется полный факторный план эксперимента с большим числом уровней n_k для менее гладких факторов. Предположение о принадлежности f_0 анизотропному классу $W_2^{\mathbf{m}}$ — это один из возможных способов формализовать подобные априорные инженерные знания о природе функции f_0.

Выбор способа ограничения изменчивости модели и использование тензорного произведения сплайнов обусловлены требованием к вычислительной сложности метода.

Вторая глава посвящена обоснованию предложенного метода с помощью исследования некоторых теоретических свойств. Для непараметрических оценок (в том числе оценок вида (1)) известна оптимальная скорость сходимости ошибки $\mathrm{E}\|\hat{f} - f_0\|_2^2$, которая равна $n^{-2m_h \big/ (2m_h+d)}$. В данной главе проводится исследование асимптотического поведения ошибки $\mathrm{E}\|\hat{f} - f_0\|_2^2$ предложенной модели (1) .

В рамках данной главы мы будем рассматривать не зафиксированный факторный план эксперимента Θ и соответствующую выборку данных, а семейство полных факторных планов эксперимента $\{\Theta_\mu\}_{\mu=1}^{\infty}$, причем число точек $n_\mu = \left|\Theta_\mu \cap \Omega\right| \to \infty$ монотонно при $\mu \to \infty$ (где $\left|\cdot\right|$ — мощность множества). Ниже в целях упрощения обозначений мы будем опускать индекс μ в зависящих от него параметрах, а символ C далее будет использоваться для обозначения положительных констант, зависящих только от, быть может, m, d и Ω, но не зависящих от f_0, n, λ_k, h_k.

Введем предположения, в рамках которых мы будем работать.

Предположение 1. *Справедлива следующая модель данных*

1. *Область Ω удовлетворяет сильному условию l-рога*[7], $l_k = 1/m_k$.
2. $f_0 \in W_2^{\mathbf{m}}(\Omega)$, $2m_h > d$, *где m_h — это гармоническое среднее чисел $m_1, \ldots, m_d$, $m_h = d\big/\sum_{k=1}^{d} m_k^{-1}$.*

Введем некоторые ограничение на параметры модели.

Предположение 2. *Будем полагать, что*

1. $h_k = n^{-\beta/m_k}$, *а параметр β может лежать в интервале $-m_h/d < \beta \leq -m_h\big/(2m_h + d)$;*
2. $Cn^{-2m_h/d} < \lambda_k \leq \lambda_0$, *где λ_0 — некоторая константа, достаточ-*

[7] *Бесов О. В., Ильин В. П., Никольский С. М.* Интегральные представления функций и теоремы вложения. – НАУКА-Физматлит, 1996.

но большая для любого практического применения (например, $\lambda_0 = 1$).

Предположение 3. *План эксперимента (множество точек $\{x_i\}_{i=1}^N$) является асимптотически равномерным:*

1. *$N \le Cn$;*
2. *Ω может быть разбита на такие множества Ω_i, что*
$$x_i \in \Omega_i, \quad x_i \notin \Omega_j (j \ne i);$$
$$\Omega = \bigcup_{i=1}^N \Omega_i, \quad |\Omega_i| \le Cn^{-1};$$
Ω_i удовлетворяет условию l-рога, $l_k = 1/m_k$.

   где $|\Omega_i|$ — объем Ω_i.
3. *Точки распределены достаточно равномерно по интервалам разбиения сплайнов, а именно*
$$\forall j_k = 1, \ldots, p_k \quad 0 < \left| \left\{ \theta^k_{i_k} : \theta^k_{i_k} \in \big((j_k - 1) h_k, j_k h_k \big] \right\} \right| < C h_k n_k.$$

Для построения оценки ошибки предложенного метода последовательно рассматриваются следующие задачи:

1. Приближение известной функции f_0 с помощью тензорного произведения сплайнов:
$$\hat{f}_f = \underset{f \in S_{\mathbf{m}}}{\operatorname{argmin}} \left(\|f_0 - f\|^2_{L_2,\Omega} + \sum_{k=1}^d \lambda_k \|D_k^{m_k} f\|^2_{L_2,H} \right).$$
2. Приближение функции f_0 по известным значениям $f_0(x_i)$:
$$\hat{f}_s = \underset{f \in S_{\mathbf{m}}}{\operatorname{argmin}} \left(\frac{1}{n} \sum_{i=1}^n (f_0(x_i) - f(x_i))^2 + \sum_{k=1}^d \lambda_k \|D_k^{m_k} f\|^2_{L_2,H} \right).$$
3. Приближение функции f_0 по зашумленным значениям $y_i = f_0(x_i) + \xi_i$:
$$\hat{f}_n = \underset{f \in S_{\mathbf{m}}}{\operatorname{argmin}} \left(\frac{1}{n} \sum_{i=1}^n (y_i - f(x_i))^2 + \sum_{k=1}^d \lambda_k \|D_k^{m_k} f\|^2_{L_2,H} \right).$$
4. Приближение функции f_0 по зашумленным значениям $y_i = f_0(x_i) + \xi_i$ с использованием дискретно-непрерывной нормы производных в штрафном слагаемом:
$$\hat{f} = \underset{f \in S_{\mathbf{m}}}{\operatorname{argmin}} \left(\frac{1}{n} \sum_{i=1}^n (y_i - f(x_i))^2 + \sum_{k=1}^d \lambda_k \|D_k^{m_k} f\|^2_{\Theta,k} \right).$$

Опустим результаты для промежуточных моделей и рассмотрим финальную модель $\hat{f}$.

Теорема 1. *Пусть выполнены предположения 1, 2, 3. Тогда существует такой объем выборки n_0, что для всех $n > n_0$ справедлива следующая оценка ошибки приближения f_0*

$$\mathrm{E}\|f_0 - \hat{f}\|^2_{L_2} \leq \ C\left[max_k \lambda_k \big|f_0\big|^2_{W_2^{\mathbf{m}},\Omega} + n^{-1}\lambda^{-1}\sigma^2\right],$$

где $\lambda = \prod \lambda_k^{1/2m_k}$.

Выбирая $\lambda_k \asymp n^{-(2m_h)/(2m_h+d)}$ (такой выбор удовлетворяет ограничениям, введенным в предположении 2) мы получим

$$\mathrm{E}\|\hat{f} - f_0\|^2_{L_2} \asymp n^{-2m_h/(2m_h+d)}.$$

Полученная скорость сходимости является оптимальной.

Третья глава посвящена построению вычислительно эффективного алгоритма нахождения решения задачи (1) для случая неполных факторных планов эксперимента. Легко показать, что задача (1) поиска оптимального решения в выбранном классе функций $S_{\mathbf{m}}$ сводится к минимизации квадратичной функции ошибки по коэффициентам разложения по базису в $S_{\mathbf{m}}$, причем эти коэффициенты могут быть выписаны в виде тензора $\mathcal{A}$. Особенностью этой квадратичной функции является ее высокая размерность и неразреженная структура матрицы гессиана. В общем случае для минимизации таких функций требуется $O(P^3 + nP^2)$ операций (где $P = \prod p_k$ — число коэффициентов).

Для построения алгоритма нам потребуются базовые операции с тензорами. Сложение, скалярное произведение, поэлементное умножение тензоров определены стандартным образом. Будем говорить, что тензор $\mathcal{A}$ с размерами $u_1 \dots \times t_k \dots \times u_d$ — это результат умножение тензора $\mathcal{B}$ (с размерами $u_1 \dots \times u_k \dots \times u_d$) по измерению k на $u_k \times t_k$-матрицу C и обозначать $\mathcal{A} = \mathcal{B} \otimes_k \mathrm{C}$, если

$$\mathcal{A}_{i_1,\dots,i_{k-1},j,i_{k+1},\dots,i_d} = \sum_{i_k=1}^{u_k} \mathcal{B}_{i_1,\dots,i_k,\dots,i_d} \mathrm{C}_{i_k\, j}.$$

Также будем использовать следующее обозначение для умножения тензора $\mathcal{A}$ на матрицы $\mathrm{C}_k, k = 1, \dots, d$:

$$\mathcal{A} \overset{d}{\underset{1}{\otimes}}_k \mathrm{C}_k \equiv \mathcal{A} \otimes_1 \mathrm{C}_1 \dots \otimes_d \mathrm{C}_d.$$

Построим индикатор попадания точек полного факторного плана $\Theta = \left\{z_{i_1,\dots,i_d}\right\}_{i_1,\dots,i_d=1}^{n_1,\dots,n_d}$ в план эксперимента X заданной выборки данных. Пусть $\mathcal{W}$ — тензор с размерами $n_1 \times n_2 \cdots \times n_d$, причем

$$\mathcal{W}_{i_1,\ldots,i_d} = \begin{cases} 1, & z_{i_1,\ldots,i_d} \in X, \\ 0, & z_{i_1,\ldots,i_d} \in \Theta \setminus X. \end{cases}$$

Построим тензор наблюдений в точках полного плана $\mathcal{Y}$ с размерами $n_1 \times n_2 \cdots \times n_d$, дополнив заданные значения $\{y_i\}_{i=1}^n$ нулями, то есть:

$$\mathcal{Y}_{i_1,\ldots,i_d} = \begin{cases} y_i, & \exists\, x_i : x_i = z_{i_1,\ldots,i_d}, \\ 0, & z_{i_1,\ldots,i_d} \in \Theta \setminus X. \end{cases}$$

Пусть $\Delta_k = \{\psi^k_{j_k}\}_{j_k=1}^{p_k}$ — это базисы в пространстве одномерных сплайнов S_k для $k = 1,\ldots,d$. Тогда произвольная функция $f \in S_{\mathbf{m}}$ может быть представлена в виде

$$f(x) = \sum_{j_1,\ldots,j_d=1}^{p_1,\ldots,p_d} \mathcal{A}_{j_1,\ldots,j_d} \psi^1_{j_1}(x^1) \otimes \psi^2_{j_2}(x^2) \cdots \otimes \psi^d_{j_d}(x^d),$$

где $\mathcal{A}$ — тензор коэффициентов разложения с размерами $p_1 \times p_2 \cdots \times p_d$.

Для всех $k = 1,\ldots,d$ обозначим матрицу значений функций из базиса Δ_k в точках из множества θ^k как R_k,

$$\mathrm{R}_k = \big\{\psi^k_{j_k}(\theta^k_{i_k})\big\}_{i_k,j_k=1}^{n_k,p_k},$$

а матрицу скалярного произведения производных $\psi^k_{j_k} \in \Delta_k$ как H_k,

$$\mathrm{H}_k = \left\{ \int_{0,1} D^{m_k}\psi^k_{j_k}(x^k) D^{m_k}\psi^k_{g_k}(x^k) dx^k \right\}_{j_k,g_k=1}^{p_k,p_k}.$$

Матрицы R_k имеют размеры $n_k \times p_k$, а матрицы H_k — размеры $p_k \times p_k$.

Минимизируемый функционал в (1) может быть переписан как функция от тензора коэффициентов разложения $\mathcal{A}$

$$Q_A(\mathcal{A}) = \frac{1}{n}\Big\langle \big(\mathcal{Y} - \mathcal{A} \overset{d}{\underset{1}{\otimes}}_k \mathrm{R}_k\big) \odot \mathcal{W}, \mathcal{Y} - \mathcal{A} \overset{d}{\underset{1}{\otimes}}_k \mathrm{R}_k \Big\rangle_2$$

$$+ \sum_{k=1}^{d} \lambda_k \frac{n_k}{N} \Big\langle \mathcal{A} \overset{d}{\underset{1}{\otimes}}_l \big(\delta_{kl}\mathrm{H}_l + (1-\delta_{kl})\mathrm{R}_l^{\mathrm{T}}\mathrm{R}_l\big), \mathcal{A} \Big\rangle_2,$$

где $\odot$ — поэлементное умножение тензоров, а δ_{kl} — символ Кронекера. $Q_A(\mathcal{A})$ является квадратичной функцией от коэффициентов $\mathcal{A}$.

С помощью обобщенного сингулярного разложения представим H_k и R_k в следующем виде:

$$\mathrm{H}_k \;=\; \mathrm{V}_k\,\mathrm{D}_k\,\mathrm{V}_k^{\mathrm{T}}; \quad \mathrm{R}_k \;=\; \mathrm{U}_k\,\mathrm{V}_k^{\mathrm{T}},$$

где D_k — это диагональные $p_k \times p_k$-матрицы, V_k — невырожденные $p_k \times p_k$-матрицы, а U_k $n_k \times p_k$-матрицы с ортонормальными столбцами.

Определим тензор $\mathcal{D}$ с размерами $p_1 \times p_2 \cdots \times p_d$ как

$$\mathcal{D} = \sum_{k=1}^{d} \lambda_k \frac{n_k}{N} \mathcal{I} \underset{1}{\overset{d}{\otimes}}_l \big(\delta_{kl} \mathrm{D}_l + (1 - \delta_{kl}) \mathrm{I}_{p_l}\big),$$

где $\mathcal{I}$ — тензор с размерами $p_1 \times p_2 \cdots \times p_d$, состоящий из 1, а I_{p_l} — единичная $p_l \times p_l$-матрица. Так же введем обозначение $\widetilde{\mathcal{D}} = n^{-1}\mathcal{I}_p + \mathcal{D}$.

Введем замену переменных

$$\mathcal{B} = \begin{cases} \big(\mathcal{A} \underset{1}{\overset{d}{\otimes}}_k \mathrm{V}_k\big) \odot \mathcal{D}^{\,1/2},\ n < N/2 \\ \big(\mathcal{A} \underset{1}{\overset{d}{\otimes}}_k \mathrm{V}_k\big) \odot \widetilde{\mathcal{D}}^{1/2},\ n \geq N/2, \end{cases}$$

где возведение тензора в степень выполняется поэлементно. Пусть Q_B — это квадратичная форма Q_A после замены переменных, $Q_B\big(\mathcal{B}\big) \equiv Q_A\big(\mathcal{A}\big)$.

Лемма 1. *Гессиан квадратичной формы Q_B не вырожден и имеет не более $min(N - n + 1, n + 1)$ различных собственных чисел.*

Теорема 2. *Минимизация квадратичной формы Q_B с помощью метода сопряженных градиентов требует не более* $\min(N - n + 1, n + 1)$ *шагов, а вычислительная сложность каждого шага не превосходит $O\big(T\big)$, где $T = N \sum_{k=1}^{d} n_k \prod_{l=1}^{k} p_l/n_l$.Общая вычислительная сложность составляет*

$$\sum p_k^2(n_k + p_k) + 3T \times \min\big(N - n + 1, n + 1, P\big).$$

В случае малого числа „пропусков“ в полном факторном плане (т.е. при $n \approx N$), и $p_k \sim n_k = N^{1/d}$ вычислительная сложность составит $O\Big((N - n + 1)P^{1+1/d}\Big)$ вместо $O(P^3)$. Более того, если план эксперимента полный факторный, то $N = n$ и, согласно теореме 2, оптимальный тензор коэффициентов $\mathcal{A}^*$ будет найден за одну итерацию. В этом частном случае также можно выписать явную вычислительно эффективную формулу для оптимальных коэффициентов разложения:

$$\mathcal{A}^* = \left(\big(\mathcal{Y} \underset{1}{\overset{d}{\otimes}}_k \mathrm{U_k}\big) \odot \widetilde{\mathcal{D}}^{-1}\right) \underset{1}{\overset{d}{\otimes}}_k \mathrm{V_k}^{-T}. \tag{2}$$

Рассмотрим два примера, иллюстрирующих выигрыш в вычислительной сложности, достигаемый с помощью предложенного подхода.

Пример 1. *Рассмотрим случай $p_k = n_k = N^{1/d} = P^{1/d}$ (полный факторный план эксперимента с одинаковым числом уровней во всех факторах). Тогда сложность* (2) *составит $O(P \sum p_k) = O(P^{1+1/d})$ вместо $O(P^3)$ для подходов, не использующих структурные особен-*

ности данных. Для типичной ситуации $n_k = 10$, $d = 6$ эти оценки составят 10^7 и 10^{18} соответственно.

Утверждение 1. *Пусть $\alpha = n/N$ — это доля точек выборки в полном факторном плане, а объем базиса равен числу точек выборки $P = n$. Тогда отношение сложности предложенного метода к сложности не учитывающих структуру подходов будет равно*

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

Пример 2. *Некоторые значения для случая $d = 6, n_k = 10$:*

- *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.*

Вторая часть главы посвящена методу выбора вектора параметров регуляризации Λ по данным. Воспользуемся классическим подходом, основанном на минимизации ошибки скользящего контроля Q_{loo}, и предложим вычислительно эффективный способ ее подсчета.

Утверждение 2. *Ошибка скользящего контроля может быть подсчитана по следующей вычислительно эффективной формуле*

$$Q_{loo} = \left\langle \mathcal{Y} - \widehat{\mathcal{Y}},\ (\mathcal{Y} - \widehat{\mathcal{Y}}) \odot \left(\mathcal{I} - \widetilde{\mathcal{D}}^{-1} \overset{d}{\underset{1}{\otimes}}_k \big(\mathrm{U}_k^{\mathrm{T}} \odot \mathrm{U}_k^{\mathrm{T}}\big)\right)^{-2} \right\rangle, \qquad (3)$$

где $\widehat{\mathcal{Y}} = \mathcal{A} \otimes_1 \Psi_1^{\mathrm{T}} \ldots \otimes_d \Psi_d^{\mathrm{T}}$ — значения модели $f(\mathcal{A})$ в точках полного плана Θ, а $\mathcal{I}$ — тензор размера $n_1 \times \cdots \times n_d$.

Утверждение 3. *Компоненты градиента Q_{loo} по λ могут быть вычислены по формуле*

$$\begin{aligned}\frac{\partial Q_{loo}}{\lambda_k} = &- 2\left\langle (\widehat{\mathcal{Y}} - \mathcal{Y}) \odot \widetilde{\mathcal{L}}^2, \left(\Big(\mathcal{Y} \overset{d}{\underset{1}{\otimes}}_l \mathrm{U}_l\Big) \odot \widetilde{\mathcal{D}}^{-2} \odot \frac{\partial \mathcal{D}}{\partial \lambda_k} \right) \overset{d}{\underset{1}{\otimes}}_l \mathrm{U}_l^{\mathrm{T}} \right\rangle \\ &- 2\left\langle (\widehat{\mathcal{Y}} - \mathcal{Y})^2,\ \widetilde{\mathcal{L}}^3 \odot \left[\left(\widetilde{\mathcal{D}}^{-2} \odot \frac{\partial \mathcal{D}}{\partial \lambda_k} \right) \overset{d}{\underset{1}{\otimes}}_l \big(\mathrm{U}_l^{\mathrm{T}} \odot \mathrm{U}_l^{\mathrm{T}}\big) \right] \right\rangle.\end{aligned} \qquad (4)$$

$$\widetilde{\mathcal{L}} = \left(\mathcal{I} - \widetilde{\mathcal{D}}^{-1} \otimes_1 \left(\mathrm{U}_1^{\mathrm{T}} \odot \mathrm{U}_1^{\mathrm{T}}\right) \ldots \otimes_d \left(\mathrm{U}_d^{\mathrm{T}} \odot \mathrm{U}_d^{\mathrm{T}}\right)\right)^{-1};$$

$$\frac{\partial \mathcal{D}}{\partial \lambda_k} = \frac{n_k}{N} \mathcal{I} \overset{d}{\underset{1}{\otimes}}_l \big(\delta_{kl} \mathrm{D}_l + (1 - \delta_{kl}) \mathrm{I}_{p_l}\big).$$

Утверждение 4. *Вычислительная сложность формул (3), (4) составляет $O(T)$, где $T = N \sum_{k=1}^{d} n_k \prod_{l=1}^{k} p_l / n_l$.*

В случая неполного факторного плана используется модифицированная формула ошибки скользящего контроля, в которой учитываются лишь точки с известными значениями y_i:

$$\widetilde{Q}_{loo} = \Big\langle \big(\mathcal{Y} - \widehat{\mathcal{Y}}\big) \odot \mathcal{W},\ \big(\mathcal{Y} - \widehat{\mathcal{Y}}\big) \odot \mathcal{L}^2 \Big\rangle.$$

Вычислительно эффективная формула для подсчета градиента этой функции элементарным образом выводятся из полученной формулы (4) поэлементным домножением одного из аргументов в каждом скалярном произведении на тензор $\mathcal{W}$.

Если план эксперимента полный факторный, то оптимальное значение вектора параметров регуляризации Λ находится с помощью оптимизации Q_{loo} методом сопряженных градиентов с использованием построенных вычислительно эффективных формул для подсчета значения и градиента этой функции. Если же план неполный факторный, то используется итеративная процедура, каждая итерация которой состоит из двух шагов. На первом шаге для зафиксированного вектора параметров регуляризации Λ вычисляются оптимальные коэффициенты разложения и с их помощью оцениваются значения в точках полного плана, не вошедших в выборку. Затем на втором шаге эти значения используются при минимизации $\widetilde{Q}_{loo}$ по параметрам регуляризации Λ методом сопряженных градиентов. Найденное значение Λ фиксируется и начинается новая итерация. В экспериментах процедура сходится до приемлемой точности (в смысле вариабельности построенной модели) за $5 - 10$ итераций.

В **четвертой главе** описан разработанный автором программный комплекс на языке Matlab, в котором реализованы созданные в работе алгоритмы. Помимо непосредственно построения модели по заданной выборке с факторным планом эксперимента, разработанный комплекс имеет некоторые дополнительные функциональные возможности.

- Построенная модель может быть дополнительно сглажена (т.е. параметры регуляризации Λ могут быть установлены выше значений, подобранных по данным). Требуемый уровень гладкости задается с помощью параметра относительной гладкости, который изменяется от 0 (исходная модель) до 1 (линейная модель).
- В силу использования базисных функций с компактным носителем построенная модель стремится к константе при выходе за границы гиперкуба $H = [0,1]^d$. В программном комплексе реализована возможность гладкого продолжения моделируемой зависимости на куб $[-\delta, 1+\delta]^d$, заданный с помощью параметра δ,

при этом вычислительная сложность алгоритма не меняется.

- В некоторых задачах структура данных такова, что в полном факторном плане есть один или несколько многомерных факторов. Разработана и реализована эвристическая вычислительно-эффективная процедура построения словаря параметрических многомерных функций в каждом из таких факторов. Полученный словарь затем используется при формировании класса функций, в котором ищется решение задачи, наравне с базисами B-сплайнов.

Также в компании Datadvance разработанный программный комплекс был переписаны на языке C++ и включен в состав двух программных продуктов компании: MACROS (библиотека методов интеллектуального анализа данных и многокритериальной оптимизации) и pSeven (платформа для автоматизации инженерных расчетов, анализа данных и оптимизации).

Пятая глава содержит результаты некоторых вычислительных экспериментов и состоит из трех разделов. Первый раздел посвящен исследованию свойств предложенного метода на искусственных данных, имитирующих типичные аэродинамические зависимости, и данных из реальных инженерных задач. Размерность задач варьировалась от 2 до 7, объем выборки от $2e3$ до $5e5$. План эксперимента в задачах 1, 2 полный факторный, в задачах 3, 4 — неполный факторный. Корень из среднеквадратичной ошибки на тестовых данных, нормированный на аналогичную ошибку аппроксимации константой, и время построения модели в секундах для трех алгоритмов (нейронная сеть, регрессия на основе гауссовских процессов и предложенный подход) представлены в таблице 1. Следует отметить, что приведенные задачи имеют существенно неравномерный план эксперимента, что заметно ухудшает качество моделей, построенных не учитывающими структуру методами.

	Задача 1		Задача 2		Задача 3		Задача 4	
	СКО	Вр.	СКО	Вр.	СКО	Вр.	СКО	Вр.
Нейр. сеть	0.38	12.5	0.003	3800	0.107	333	0.124	255
Гаусс. пр.	0.42	28.0	-	-	0.098	98	0.213	72
Предл. подх.	0.12	0.11	0.0009	2.96	0.011	0.62	0.035	2.1

Таблица 1: Ошибки аппроксимации и время обучения

Следующий раздел посвящен краткому описанию результатов

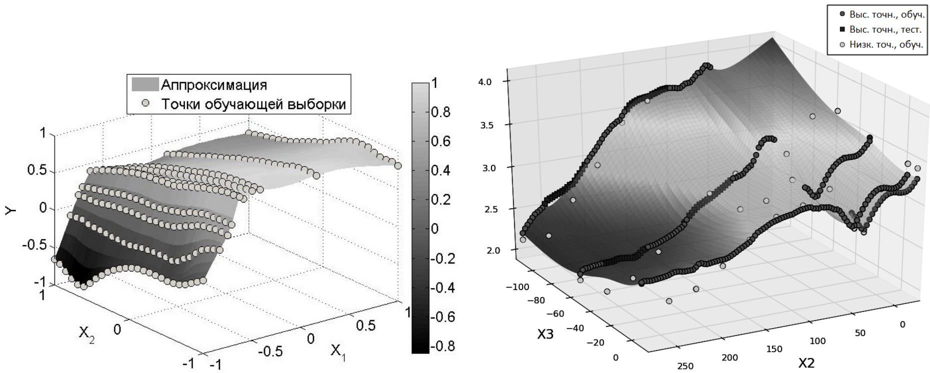

Рис. 1: Примеры срезов моделей: задача 1 (слева), моделирование аэродинамических характеристик космического летательного аппарата по данным низкой и высокой точности (справа).

двух прикладных задач инженерного проектирования: моделирования прочностных характеристик элементов обшивки самолета $A350-900$ и моделирования аэродинамических характеристик суборбитального космического летательного аппарата. Во второй из указанных задач модель было необходимо строить с помощью консолидации двух выборок разноточных данных, каждая из которых имеет факторный план. В разделе описываются несколько подходов к решению этой задачи, основанные на предложенном в работе алгоритме.

В **заключении** приведены основные результаты работы.

1. Предложен новый метод построения модели по данным с факторным планом эксперимента, позволяющий формализовать априорные знания о структуре моделируемой зависимости с помощью введения штрафа на взвешенную сумму дискретно-непрерывных норм различных частных производных модели.
2. Найдена оценка средней ошибки приближения и доказано, что предложенный метод является асимптотически оптимальным при надлежащем выборе параметров регуляризации.
3. Предложен алгоритм, позволяющий вычислительно эффективно строить модель по данным с факторным планом эксперимента.
4. Разработан вычислительно эффективный алгоритм выбора параметров регуляризации по данным.
5. Создан программный комплекс, реализующий разработанные алгоритмы.

6. С помощью разработанного программного комплекса решен ряд задач в аэрокосмической отрасли, в том числе задача построения модели по двум выборкам разноточных данных, план эксперимента каждой из которых неполный факторный.

Публикации по теме диссертации

1. Беляев М. Г. Анизотропные сглаживающие сплайны в задачах с факторным планом эксперимента // Доклады Академии Наук. 2015. Т. 5, № 461. С. 521–524.

2. Беляев М.Г. Аппроксимация многомерных зависимостей по структурированным выборкам // Искусственный интеллект и принятие решений. 2013. № 3. С. 24–39.

3. Беляев М.Г. Аппроксимация данных, порожденных декартовым произведением // Труды МФТИ. 2013. Т. 5, № 3. С. 11–23.

4. Surrogate models for spacecraft aerodynamic problems / M. Belyaev, E. Burnaev, Y. Kapushev et al. // In proc. of 5th European Conference on Computational Mechanics. 2014. P. 418–422.

5. On Approximation of Reserve Factors Dependency on Loads for Composite Stiffened Panels / G. Sterling, P. Prikhodko, E. Burnaev et al. // Advanced Materials Research. 2014. Vol. 1016. P. 85–89.

6. Беляев М.Г. Аппроксимация зашумленных данных, имеющих структуру декартова произведения // Сборник докладов конференции ИОИ-9. 2012. С. 188–192.

7. Беляев М.Г. Аппроксимация и интерполяция на основе тензорного произведения параметрических словарей // Труды конференции Информационные Технологии и Системы. 2012. С. 32–40.

Беляев Михаил Геннадьевич

Моделирование анизотропных зависимостей по выборкам с факторным планом эксперимента

АВТОРЕФЕРАТ