Федеральное государственное бюджетное учреждение науки
Институт проблем передачи информации им. А.А. Харкевича
Российской академии наук (ИППИ РАН)

На правах рукописи

Зайцев Алексей Алексеевич

МЕТОДЫ ПОСТРОЕНИЯ РЕГРЕССИОННЫХ МОДЕЛЕЙ РАЗНОРОДНЫХ ИСТОЧНИКОВ ДАННЫХ ДЛЯ ИНДУСТРИАЛЬНОЙ ИНЖЕНЕРИИ

Специальность 05.13.18 —
математическое моделирование, численные методы и
комплексы программ

Диссертация на соискание учёной степени
кандидата физико-математических наук

Научный руководитель:
кандидат физико-математических наук, доцент
Бурнаев Евгений Владимирович

Москва – 2017

Оглавление

Введение

Общая характеристика работы

Актуальность темы. В индустриальной инженерии одной из основных задач является задача проектирования изделия, характеристики которого удовлетворяют заданным требованиям. Часто в задачах индустриальной инженерии применяют подход, основанный на использовании быстро вычислимой регрессионной модели, построенной по выборке пар «параметры изделия (входной вектор, точка) — его характеристики (выходной вектор)», где характеристики изделия получаются в результате ресурсоемкого численного моделирования или натурных экспериментов.

В задачах индустриальной инженерии используемые данные могут быть разными по точности и стоимости получения, разнородными: часть данных может быть порождена источником данных высокой точности, а другая часть — источником данных низкой точности, при этом ресурсоемкость использования источника данных высокой точности обычно существенно выше ресурсоемкости использования источника данных низкой точности. Например, в задаче построения модели зависимости подъемной силы крыла самолета от его формы данные высокой точности могут быть получены из экспериментов в аэродинамической трубе, а данные низкой точности — из расчетов с помощью численного моделирования. При наличии разнородных источников данных можно выбирать для каких изделий использовать источник данных высокой точности, а для каких — низкой, чтобы для заданного общего ресурсного ограничения построить по полученным данным как можно более точную регрессионную модель.

Не существует универсального алгоритма для построения регрессионных моделей. Часто применяют метод, основанный на предположении о том, что моделируемая функция есть реализация гауссовского процесса. Такой метод называют регрессией на основе гауссовских процессов. Он широко используется для построения нелинейных регрессионных моделей, в том числе и по выборкам разнородных данных. Для однородных данных исследования регрессии на основе гауссовских процессов проводились в работах А.Н. Колмогорова, М. Штайна, А. Ван Дер Ваарта, и других, минимаксная ошибка — ошибка для наилучшей аппроксимации для наихудшей целевой функции заданной гладкости — была получена в работе Г.К. Голубева и Е.А. Крымовой. Однако, для разнородных данных существующие результаты либо опираются на эвристики, либо получены в предположениях, не позволяющих использовать такие результаты на практике. Часто эффективный план экспериментов для разнородных источников данных таков, что требуется использование чрезмерных вычислительных ресурсов для построения регрессионной модели. Однако, на сегодняшний день вычислительно эффективные подходы к регрессии на основе гауссовских процессов разработаны только для однородных данных. Таким образом, **актуальны** разработка вычислительно эффективных методов построения регрессионных моделей разнородных данных на основе гауссовских процессов, проведение исследования таких моделей данных и разработка метода выбора эффективного плана экспериментов для разнородных источников данных для подхода на основе гауссовских процессов в условиях заданного ресурсного ограничения.

Объектом исследования являются регрессионные модели индустриальной инженерии на основе гауссовских процессов, параметры которых оцениваются по разнородным данным. **Предметом** исследования являются методы построения регрессионных моделей разнородных данных и метод выбора эффективного плана экспериментов, предназначенного для построения таких моделей.

Целями данной работы является разработка вычислительно эффективных методов построения регрессионных моделей разнородных данных, оценка качества таких регрессионных моделей, и разработка методов выбора эффективного плана экспериментов для таких моделей.

Поставленные цели определили следующие **задачи** исследования:

1. Разработать вычислительно эффективные методы построения регрессионных моделей разнородных данных, которые учитывают типичные особенности таких данных, и создать их программную реализацию.

2. Получить оценку качества регрессионных моделей данных на основе гауссовских процессов.

3. Разработать метод выбора эффективного плана экспериментов — соотношения между размерами выборок разнородных данных при заданном ресурсном ограничении, которое максимизирует качество получаемой регрессионной модели.

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

1. Разработан новый метод построения регрессионных моделей на основе гауссовских процессов по выборкам разнородных данных, основанный на численных методах для низкоранговой аппроксимации.

2. В многомерном случае получены минимаксные ошибки интерполяции для моделей нелинейной регрессии на основе гауссовских процессов, построенных по выборкам как однородных, так и разнородных данных.

3. Разработан новый метод выбора соотношения размеров выборок разнородных данных, минимизирующего минимаксную ошибку интерполяции при заданном ресурсном ограничении.

Теоретическая и практическая значимость представленной диссертационной работы определяется строгостью полученных математических результатов и широким использованием рассмотренных методов для моделирования по выборкам разнородных данных. Предложенные в работе методы используются для решения прикладных задач, возникающих в инженерной практике.

Общая методика исследования. Для решения поставленных задач в работе используются методы математической статистики, теории случайных процессов, аппарата анализа Фурье, статистической теории машинного обучения, матричной алгебры.

Основные положения, выносимые на защиту:

1. Разработанный метод построения нелинейных регрессионных моделей для выборок разнородных данных на основе низкоранговой аппроксимации имеет трудоемкость $O(\phi(n)^2 n)$ вместо $O(n^3)$ для стандартного подхода. Значение $\phi(n)$ обычно выбирают порядка $\min(c, n)$, где c — константа, задаваемая требованием к качеству модели.

2. Полученная теоретическая оценка качества регрессионной модели многомерных нелинейных зависимостей, в том числе в случае наличия разнородных источников данных, позволяет определить целесообразность использования разнородных источников данных.

3. Разработанный метод выбора соотношения между размерами выборок разнородных данных является теоретически оптимальным и обеспечивает высокое качество регрессионных моделей на практике.

4. Разработанные методы вошли в состав программного комплекса, предназначенного для решения задач анализа данных в индустриальной инженерии.

5. С помощью разработанного программного комплекса решен ряд задач индустриальной инженерии.

Достоверность изложенных в работе результатов определяется использованием корректных математических методов, основанных на хорошо изученных подходах из теории математической статистики; результатами проведенных численных экспериментов; согласованностью полученных результатов с ранее известными; а также успешным использованием предложенных подходов для решения реальных задач индустриальной инженерии.

Апробация работы. Результаты диссертации докладывались и обсуждались на следующих конференциях: международная конференция молодых ученых «Информационные Технологии и Системы» (2012, Петрозаводск; 2016, Репино), 9-ая Международная конференция «Интеллектуализация обработки информации» (2012, Будва, Черногория), Conference on Structural Inference in Statistics (2013, Потсдам, Германия), конференции 5th Symposium on Conformal and Probabilistic Prediction with Applications (2016, Мадрид, Испания). Также результаты работы обсуждались на семинарах лаборатории структурных методов анализа данных в предсказательном моделировании МФТИ (2013, 2015, 2016, Москва), «Математические модели информационных технологий» НИУ ВШЭ (2015, Москва), отдела Интеллектуальных систем ВЦ РАН (2015, Москва), «Байесовские методы машинного обучения» ВМК МГУ (2015, Москва), И. Оселедца Сколтеха (2016, Москва); различных лабораторий ИППИ РАН (2016, Москва).

Публикации. Основные результаты по теме диссертации изложены в 7 печатных работах, из которых 6 [1–6] изданы в журналах, рекомендованных ВАК.

Личный вклад автора. Содержание диссертации и основные положения, выносимые на защиту, отражают персональный вклад автора в опубликованные работы. Подготовка к публикации выносимых на защиту результатов проводилась совместно с соавторами.

В цикле работ [1, 2, 3] постановки задач принадлежат соавторам, доказательство результатов получено лично диссертантом. Идеи некоторых вычислительных экспериментов принадлежат Е.В. Бурнаеву, постановка экспериментов и анализ их результатов были сделаны автором.

В работе [4] основные результаты получены автором, вычислительные эксперименты проведены автором. Е.В. Бурнаевым предложены постановки задач.

Структура и объем диссертации. Диссертация состоит из введения, 5 глав, заключения и библиографии. Общий объем диссертации 148 стр., включая 46 рисунков. Библиография включает 114 наименований.

Благодарности. Автор благодарен своему научному руководителю Е.В. Бурнаеву за постановки задач, плодотворные обсуждения, помощь в подготовке работ по теме диссертации к публикации, постоянную поддержку. Автор также благодарен Г.К. Голубеву за профессиональную экспертизу полученных результатов и ценные обсуждения.

Глава 1

Постановка задачи моделирования по выборкам разнородных данных

1.1 Введение

В индустриальной инженерии одной из основных задач является задача проектирования изделия, характеристики которого удовлетворяют заданным требованиям. Зачастую для одного изделия получение его характеристик является ресурсоемкой задачей, так как сегодня как проведение натурных экспериментов, так и использование вычислительных кодов требуют существенных временных и материальных затрат [33]. В то же время в процессе проектирования изделия часто необходимо провести большое количество экспериментов.

Одним из способов преодоления этой проблемы является использование регрессионных моделей, построенных по выборкам данных, или суррогатных моделей [33]. В таком случае модели строятся на основе выборки пар "параметры изделия (входной вектор) — его характеристики (выходной вектор)". Стоимость использования построенных по данным моделей существенно ниже стоимости использования вычислительных кодов и тем более проведения натурных экспериментов. Кроме того, большая свобода в выборе метода моделирования позволяет учесть априорные требования инженера к модели.

Часто встречающейся особенностью задач в индустриальной инженерии является наличие разнородных источников данных, отличающихся по стоимости получения ответа и его точности. Например, дорогим, но достаточно точным способом измерения аэродинамических характеристик крыла самолета являются эксперименты в аэродинамической трубе, другим, менее точным и более дешевым способом измерения аэродинамических характеристик является использование кодов вычислительной гидродинамики.

Не существует универсального алгоритма регрессионного моделирования. В диссертации рассматривается популярный в индустриальной инженерии подход, основанный на предположении о том, что целевая зависимость есть реализация гауссовского процесса [33] (обычно, в том числе и в многомерном случае, в литературе используют термин «гауссовские процессы» вместо математически более правильного термина «гауссовское случайное поле», в исследовании так же используются принятая в литературе терминология). Такой подход получил название регрессии на основе гауссовских процессов или кригинга.

Приведем сперва несколько примеров задач индустриальной инженерии, в которых используются модели, построенные по выборкам данных, в том числе полученных из разных источников, с использованием регрессии на основе гауссовских процессов. Затем дадим формальную постановку задачи, мотивированную рассмотренными примерами.

1.2 Задачи индустриальной инженерии, в которых доступны разнородные источники данных

1.2.1 Задача оптимизации формы вращающегося диска

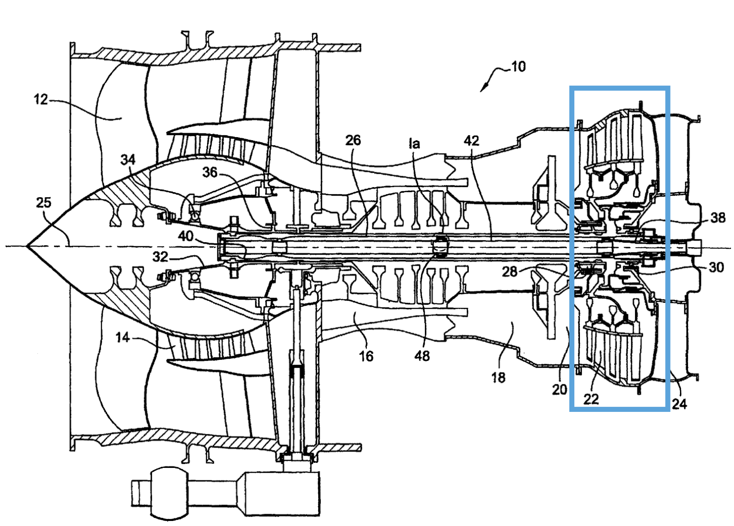
Описание модели вращающегося диска

Вращающийся диск является важной частью двигателя самолета (см. Рисунок 1.1а). Три характеристики определяют качество диска: масса, максимальное радиальное смещение $u_{\max}$ при заданной нагрузке и максимальная нагрузка $s_{\max}$ при вращении [6, 66, 12]. Массу диска легко подсчитать явно, зная геометрические параметры диска. Зависимости максимального радиального смещения и нагрузки от формы диска сложнее [66, 44], для подсчета $u_{\max}$ и $s_{\max}$ необходимо использование вычислительных кодов.

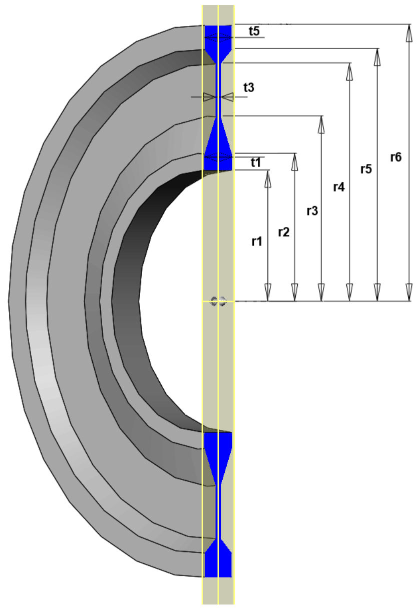
В данной задаче параметризация геометрии диска — вектор из 8 параметров: радиусов r_i, $i = 1, \dots, 6$, которые задают места, в которых меняется толщина диска, и значения t_1, t_3, t_5, которые задают соответствующие толщины. В рассматриваемой задаче фиксированы значения радиусов r_4, r_5 и толщины t_3. Таким образом, размерность пространства параметров изделия равна шести. Геометрия и параметризация вращающегося диска приведены на рисунке 1.1б.

Сравнение вычислительных кодов низкой и высокой точности

Рассмотрим два вычислительных кода для подсчета $u_{\max}$ и $s_{\max}$. Источник данных низкой точности — вычислительный код, решающий систему обыкновенных дифференциальных уравнений, используя метод Рунге-Кутты. Источник данных высокой точности — вычислительного кода на основе метода конечных элементов. Для обычного персонального компьютера использование источника данных низкой точности занимает примерно 0.01 секунду, а использование источника данных высокой точности — примерно 300 секунд.



(а) Двигатель самолета. Вращающийся диск выделен голубым прямоугольником в правой части рисунка.



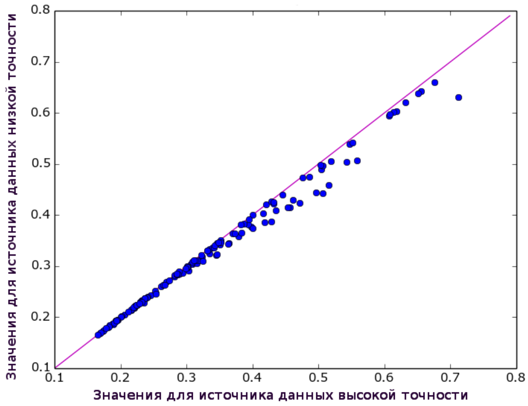
(б) Геометрия вращающегося диска

Рис. 1.1 – Задача о вращающемся диске

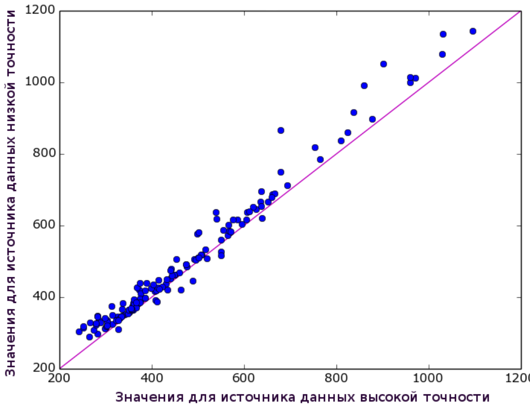
Для сравнения источников данных разной точности приведем диаграммы рассеяния значений разной точности, полученных в одинаковых точках, и срезы, полученные с помощью разнородных источников данных.

Используя метод оптимальных латинских гиперкубов, предназначенный для генерации близкого к равномерному случайного плана экспериментов, получаем выборку случайных точек из заданного диапазона входных переменных. Для такого плана экспериментов вычислим значения, полученные после применения разнородных источников данных. Полученные значения изображены на рисунке 1.2. Видно, что для больших значений целевых функций разница между результатами работы вычислительных кодов возрастает.

Для центральной точки области плана экспериментов ($r_1 = 0.06, r_2 = 0.13, r_3 = 0.16, r_4 = 0.185, t_1 = 0.027, t_3 = 0.027$) были построены одномерные срезы по каждому из входных параметров. То есть, все переменные кроме одной фиксировалось, а оставшаяся варьировалась в заданном диапазоне. Характерные срезы для выхода $u_{\max}$ приведены на рисунке 1.3, а для выхода $s_{\max}$ — на рисунке 1.4. Для $u_{\max}$ у разнородных источников данных поведение схожее. Для $s_{\max}$ поведение разнородных источников данных не всегда совпадает: нелинейное поведение отличается для входа r_1, а локальные максимумы отли-

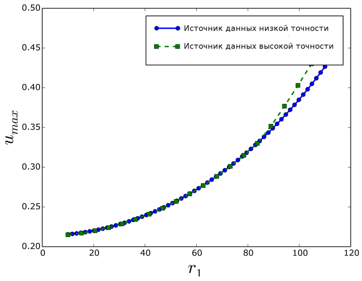


(а) Диаграмма рассеяния для $u_{\max}$

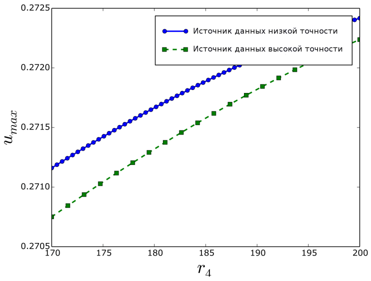


(б) Диаграмма рассеяния для $s_{\max}$

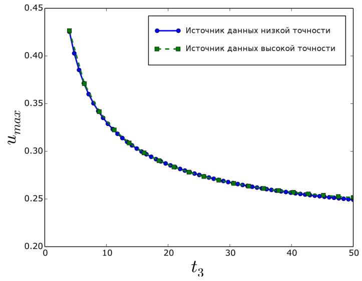
Рис. 1.2 – Сравнение вычислительных кодов разной точности с помощью диаграмм рассеяния



(а) Срез для $u_{\max}$ относительно входа r_1



(б) Срез для $u_{\max}$ относительно входа r_4



(в) Срез для $u_{\max}$ относительно входа t_3

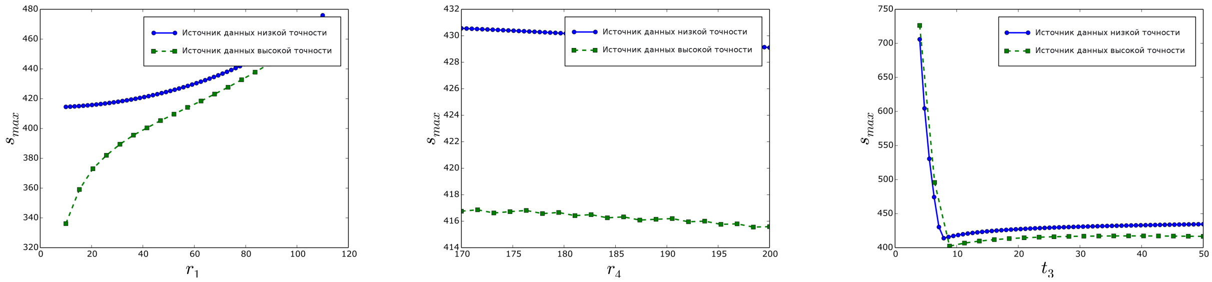
Рис. 1.3 – Сравнение срезов для разнородных источников данных для выхода $u_{\max}$

чаются для среза по входу t_3.

Так как поведение источников данных достаточно регулярное и разнородные источники данных похожи друг на друга, можно сгенерировать начальный план экспериментов, подсчитать для него выходы с помощью разнородных источников данных, а затем построить на основе полученных выборок регрессионную модель, которая бы достаточно точно описывала разнородные источники данных.

Оптимизация формы вращающегося диска

Важной является задача поиск вращающегося диска, масса которого была бы как можно меньше, но который бы вместе с тем был бы еще достаточно проч-



(а) Срез для $s_{\max}$ относительно входа r_1

(б) Срез для $s_{\max}$ относительно входа r_4

(в) Срез для $s_{\max}$ относительно входа t_3

Рис. 1.4 – Сравнение срезов для разнородных источников данных для выхода $s_{\max}$

ным:

$$
\begin{aligned}
& m, u_{\max} \to \min_{r_1,\ldots,r_6,t_1,t_3,t_5}, \qquad (1.1)\\
& u_{\max} \leq 0.3, s_{max} \leq 600,\\
& 10 \leq r_1 \leq 110, 120 \leq r_2 \leq 140,\\
& 150 \leq r_3 \leq 168, 170 \leq r_4 \leq 200,\\
& 4 \leq t_1 \leq 50, 4 \leq t_3 \leq 50,\\
& r_5 = 210, r_6 = 230, t_5 = 32.
\end{aligned}
$$

Для численного решения такой оптимизационной задачи можно использовать регрессионную модель зависимости характеристик диска от его формы вместо исходного ресурсоемкого вычислительного кода.

1.2.2 Регрессионная модель зависимости характеристик крыла самолета от его геометрии и режима полета

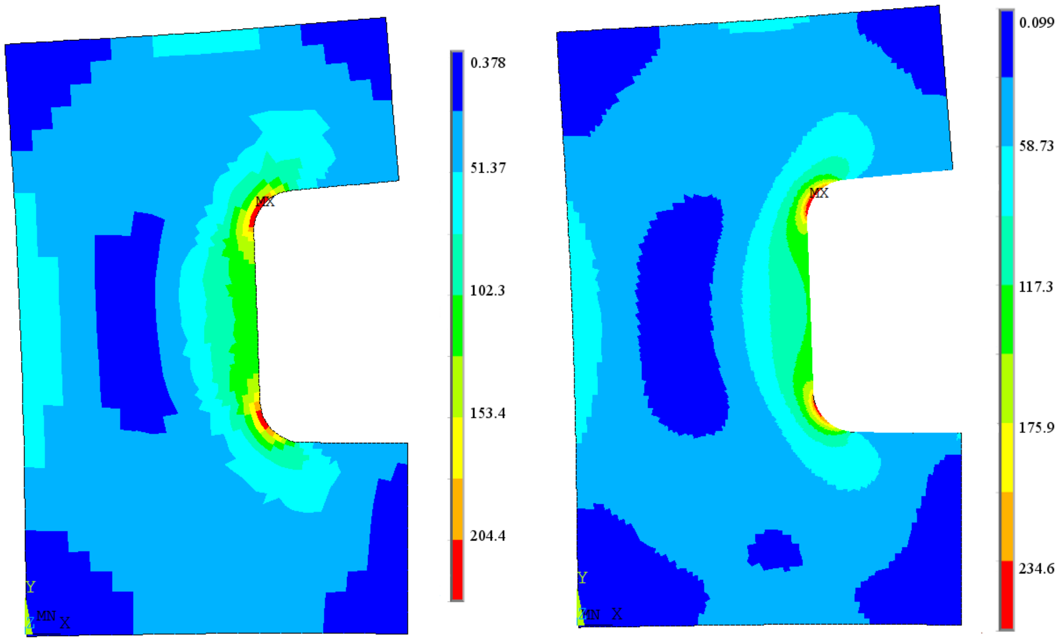
Во многих работах рассматривается построение регрессионных моделей для зависимости аэродинамического качества крыла самолета от его геометрии и режима полета [39, 34].

Будем рассматривать три различных сетки для вычислительного кода, решающего уравнения Эйлера [11], и, таким образом, получим источники данных трех различных точностей. 60 входных параметров включают описание режима полета (число Маха и угол атаки) и геометрии крыла самолета (58 параметров). Обычно для регрессионного моделирования используют более компактную параметризацию крыла самолета [13]. В данном случае использовалась процедура, обобщающая метод главных компонент на случай нелинейных моделей [14], что позволило снизить размерность пространства параметров с 60 до шести. Две целевые характеристики — коэффициент подъемной силы C_l и коэффициент сопротивления C_d — обычно используют для оценки аэродинамического качества крыла самолета. Таким образом, задача состоит в построении по выборке разнородных данных модели зависимости аэродинамического качества крыла самолета от его геометрии и режима полета.

1.2.3 Задача построения модели зависимости характеристик C-образного пресса от его геометрии

Прочность C-образного пресса для заданной нагрузки зависит от его геометрии. В данном случае используется параметризация геометрии, изображенная на рисунке 1.5. Подсчет целевых выходов — максимального смещения и максимальной нагрузки — производится с помощью вычислительно трудоемкого метода конечных элементов. Таким образом, для ускорения процесса инженерного проектирования необходимо построить регрессионную модель зависимости прочности пресса от его геометрии.

Единственный параметр сетки — размер элемента. Меняя размер элемента, мы модифицируем время вычислений и точность полученных значений прочности. Рисунок 1.5 представляет результат использования вычислительных кодов для различных размеров элемента. Время получения одного значения на персональном компьютере Intel-Core $i7$ с 4 физическими ядрами, 3.4 GHz, 8гигабайт RAM составило для кода с малым размером элемента порядка 27.8 секунд, для кода со средним размером элемента порядка 3.0 секунд, для кода с большим размером элемента порядка 2.3 секунд.



(а) Большой размер элемента, $\approx 160\,\text{mm}$

(б) Малый размер элемента, $\approx 20\,\text{mm}$

Рис. 1.5 – Результаты использования вычислительного кода для различных размеров элемента сетки

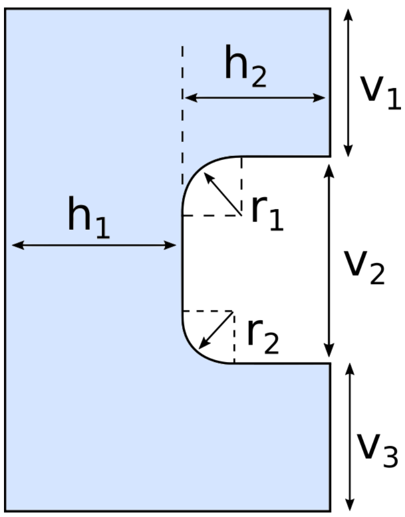


Рис. 1.6 – Параметризация геометрии С-образного пресса

1.3 Формальная постановка задачи

Таким образом, для большого количества задач индустриальной инженерии имеет смысл использовать регрессионные модели разнородных данных вместо исходных вычислительно тяжелых разнородных источников данных. При этом, обычно используют методологию регрессии на основе гауссовских процессов, которая может быть использована в том числе и в случае наличия разнородных источников данных.

Существует большое разнообразие моделей разнородных данных. Однако, из-за ограниченного объема данных и инженерной направленности таких моделей обычно используют линейную модель связи между источниками разнородных данных. Ряд прикладных исследований подтверждает, что линейная модель оказывается лучше прочих на целом ряде задач [51, 33, 76, 107]. Кроме того, она предоставляет простую интерпретацию искомой зависимости [51], что является преимуществом для модели, используемой широким кругом инженеров, для которых математическая статистики или анализ данных не являются основной специальностью [33]. Более подробный обзор современных моделей разнородных данных приведен в разделе 3.1.

В исследовании рассматривается линейную модель корегионализации или кокригинг как модель разнородных данных [51]. Такая модель имеет следующий вид:

$$u(\mathbf{x}) = \rho f(\mathbf{x}) + g(\mathbf{x})\,, \tag{1.2}$$

где ρ — фиксированная константа, а $f(\mathbf{x})$ и $g(\mathbf{x})$ — реализации двух независимых стационарных гауссовских процессов определенных на $\mathbb{R}^d$.

Будем говорить о $u(\mathbf{x})$ как об источнике данных высокой точности, а о $f(\mathbf{x})$ как об источнике данных низкой точности. Тогда $g(\mathbf{x})$ — поправка к $f(\mathbf{x})$, возникающая из-за низкой точности источника данных, соответствующего $f(\mathbf{x})$. Параметр ρ содержит информацию о силе связи между $f(\mathbf{x})$ и $u(\mathbf{x})$, и используется в модели, так как обычно $f(\mathbf{x})$ и $u(\mathbf{x})$ нормированы.

Наблюдаются значения $u(\mathbf{x})$ и $f(\mathbf{x})$, и задача состоит в построении интерполяции $\widetilde{u}(\mathbf{x})$ функции, порожденной источником данных высокой точности $u(\mathbf{x})$ на основе наблюдений, порожденных разнородными источниками данных.

1.4 Выводы

Оказывается, что в большом числе задач индустриальной инженерии для построения регрессионных моделей, служащих заменой ресурсоемким моделям, основанным на первых принципах, могут использоваться разнородные источники данных. Данная глава описывает такой подход и приводит примеры задач индустриальной инженерии, в которых он может использоваться.

Для построения регрессионных моделей разнородных данных используют подход на основе гауссовских процессов. В диссертационном исследовании рассматривается линейная модель связи между разнородными источниками данных, формально представленная в этой главе. Исследованию таких моделей как с теоретической, так и с практической точки зрения и посвящено диссертационное исследование.

Глава 2

Регрессия на основе гауссовских процессов

Пусть задана выборка $\mathbf{S} = (D, \mathbf{u}) = \{\mathbf{x}_i, u_i\}_{i=1}^n$, размера n состоящая из точек $\mathbf{x}_i \in \mathbb{X} \subseteq \mathbb{R}^d$ и значений целевой функции $u_i = u(\mathbf{x}_i)$ в этих точках. Задача состоит в построении регрессионной модели $\widetilde{u}(\mathbf{x})$ целевой функции $u(\mathbf{x})$ с использованием выборки $\mathbf{S}$.

Предположим, что $u(\mathbf{x})$ является реализацией гауссовского процесса. Для ясности и упрощения выкладок будет считать среднее значение гауссовского процесса равным нулю. Ковариационная функция для гауссовского процесса, задающего $u(\mathbf{x})$ в точках $\mathbf{x}$ и $\mathbf{x}'$ имеет вид

$$\operatorname{cov}(u(\mathbf{x}), u(\mathbf{x}')) = c(\mathbf{x}, \mathbf{x}').$$

Обозначим такие предположения как $u(\mathbf{x}) \sim \mathcal{GP}\left(0, c(\mathbf{x}, \mathbf{x}')\right)$. В таком случае совместное распределение вектора $\mathbf{u}$ — нормальное, $\mathbf{u} \propto \mathcal{N}(\mathbf{0}, \mathbf{C})$, где $\mathbf{C}$ — ковариационная матрица вида $\mathbf{C} = \{c(\mathbf{x}_i, \mathbf{x}_j)\}_{i,j=1}^n$.

Для гауссовского процесса апостериорное распределение $u(\mathbf{x})$ при заданной выборке $\mathbf{S}$ в точке $\mathbf{x} \in I\!\!R^d$ будет нормальным:

$$u(\mathbf{x})|\mathbf{S} \propto \mathcal{N}(\mu(\mathbf{x}), v^2(\mathbf{x})).$$

Выражения для апостериорного среднего $\mu(\mathbf{x})$ и дисперсии $v^2(\mathbf{x})$ по выборке записываются аналитически:

$$\mu(\mathbf{x}) = \mathbf{c}^{\mathsf{T}}(\mathbf{x})\mathbf{C}^{-1}\mathbf{u},$$
$$v^2(\mathbf{x}) = c(\mathbf{x},\mathbf{x}) - \mathbf{c}^{\mathsf{T}}(\mathbf{x})\mathbf{C}^{-1}\mathbf{c}(\mathbf{x}).$$

Здесь $\mathbf{c}(\mathbf{x}) = (c(\mathbf{x},\mathbf{x}_1),\dots,c(\mathbf{x},\mathbf{x}_n))^{\mathsf{T}}$ — вектор ковариаций между $u(\mathbf{x})$ и значениями $(u(\mathbf{x}_1),\dots,u(\mathbf{x}_n))$. Апостериорное среднее $\mu(\mathbf{x})$ используется как прогноз $\widetilde{u}(\mathbf{x})$ целевой функции $u(\mathbf{x})$, а апостериорная дисперсия $v^2(\mathbf{x})$ используется как оценка неопределенности прогноза.

Введем параметрическое предположение, состоящее в том, что ковариационная функция принадлежит семейтву

$$c(\mathbf{x},\mathbf{x}') \in \{c_{\boldsymbol{\theta}}(\mathbf{x},\mathbf{x}'), \boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^p\},$$

где Θ — компактное множество. Тогда для построения прогноза нужно оценить вектор параметров ковариационной функции $\boldsymbol{\theta}$. Параметрическое предположение может быть неправильно специфицированным, то есть возможно, что истинная ковариационная функция $c(\mathbf{x},\mathbf{x}') \notin \{c_{\boldsymbol{\theta}}(\mathbf{x},\mathbf{x}'),\ \boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^p\}$, и настоящая ковариационная матрица $\mathbf{C}$ отличается от $\mathbf{C}_{\boldsymbol{\theta}}$ для любых $\boldsymbol{\theta} \in \Theta$.

Совместное распределение вектора $\mathbf{u}$ — многомерное нормальное с нулевым средним. Ковариационная матрица распределения для фиксированной ковариационной функции $c_{\boldsymbol{\theta}}(\mathbf{x},\mathbf{x}')$ имеет вид $\mathbf{C}_{\boldsymbol{\theta}} = \{c_{\boldsymbol{\theta}}(\mathbf{x}_i,\mathbf{x}_j)\}_{i,j=1}^n$. Следовательно, логарифм (квази)правдоподобия имеет вид

$$L(\boldsymbol{\theta}) = -\frac{1}{2}\left[n\ln 2\pi + \ln|C_{\boldsymbol{\theta}}| + \mathbf{u}^{\mathsf{T}}\mathbf{C}_{\boldsymbol{\theta}}^{-1}\mathbf{u}\right]. \tag{2.1}$$

Оценка максимума правдоподобия $\widetilde{\boldsymbol{\theta}} = \operatorname{argmax}_{\boldsymbol{\theta}\in\Theta} L(\boldsymbol{\theta})$ часто используется для оценки параметров $\boldsymbol{\theta}$ ковариационной функции гауссовского процесса. Если задано априорное распределение параметров ковариационной функции $\Pi(d\boldsymbol{\theta})$, то их апостериорное распределение $\boldsymbol{\theta}$ при заданной выборке $\mathbf{S}$ имеет вид

$$\boldsymbol{\theta} \,\big|\, \mathbf{S} \propto \exp\big\{L(\boldsymbol{\theta})\big\}\,\Pi(d\boldsymbol{\theta}).$$

Математическое ожидание апостериорного распределения $\boldsymbol{\theta}$ будет *байесовской оценкой параметров* $\overline{\boldsymbol{\theta}}$.

2.1 Оценка параметров в регрессии на основе гауссовских процессов

Рассмотрим подробнее процедуру оценки параметров $\boldsymbol{\theta}$. Пусть так же наблюдаются только зашумелнные нормальным белым шумом с дисперсией $\widetilde{sigma}^2$ значения. Логарифм правдоподобия гауссовского процесса запишем в виде:

$$\log\ p(\mathbf{u}|D, \boldsymbol{\theta}, \widetilde{\sigma}) = -\frac{1}{2}\left(n \log 2\pi + \log|\mathbf{C}_{\boldsymbol{\theta},\widetilde{\sigma}}| + \mathbf{u}^{\mathsf{T}}\mathbf{C}_{\boldsymbol{\theta},\widetilde{\sigma}}^{-1}\mathbf{u}\right), \tag{2.2}$$

здесь $\mathbf{C}_{\boldsymbol{\theta},\widetilde{\sigma}} = \mathbf{C}_{\boldsymbol{\theta}} + \widetilde{\sigma}^2 I_n$ — матрица ковариаций для обучающей выборки, а $\|\mathbf{C}_{\boldsymbol{\theta},\widetilde{\sigma}}\|$ — ее определитель.

Логарифм правдоподобия зависит от параметров $\boldsymbol{\theta}$ ковариационной функции и дисперсии шума $\widetilde{\sigma}^2$. Максимизируя (2.2), получаем оценку максимума правдоподобия для $\boldsymbol{\theta}$ и $\widetilde{\sigma}^2$:

$$\log\ p(\mathbf{u}|D, \boldsymbol{\theta}, \widetilde{\sigma}) \to \max_{\boldsymbol{\theta},\widetilde{\sigma}}. \tag{2.3}$$

В этом разделе будем рассматривать коварионную функцию из *семейства квадратичных экспоненциальных ковариационных функций*:

$$c_{\boldsymbol{\theta}}(\mathbf{x}, \widetilde{\mathbf{x}}) = \theta_0^2 \exp\left(-\sum_{i=1}^{p} \theta_i^2 (x_i - \widetilde{x}_i)^2\right). \tag{2.4}$$

При оптимизации обычно используют параметризацию $\boldsymbol{\theta}$: $\theta_i = \exp(\tau_i), i = 1, \ldots, p$ [67, 63].

Приведем так же для сравнения семейство коварионных функций на основе расстояния Махалонобиса:

$$c_{\Lambda,\sigma}(\mathbf{x} - \widetilde{\mathbf{x}}) = \sigma^2 \exp\left((\mathbf{x} - \widetilde{\mathbf{x}})^T \Lambda^2 (\mathbf{x} - \widetilde{\mathbf{x}})\right),$$

где Λ^2 — некоторая положительно определенная матрица. Семейство ковариационных функций на основе расстояния Махаланобиса богаче, чем семейство квадратичных экспоненциальных ковариационных функций. Однако, количество параметров ковариационной функции, которые нужно оценить получается квадратичным от размерности входной размерности задачи.

2.1.1 Использование байесовского подхода для оценки параметров модели регрессии на основе гауссовского процесса

Недостатки стандартного подхода к оценке параметров модели

В разделе 2.1 описан базовый алгоритм построения модели на основе гауссовских процессов. На большом классе задач этот алгоритм обеспечивает хорошее качество аппроксимации моделируемой зависимости, однако в некоторых случаях приводит к вырожденной модели.

Рассмотрим, например, ковариационную функцию из семейства (2.4). При оптимизации правдоподобия по параметрам ковариационной функции $\boldsymbol{\theta}' = \{\theta_0, \ldots, \theta_p\}, \boldsymbol{\theta} = \{\theta_1, \ldots, \theta_p\}$ и дисперсии шума $\widetilde{\sigma}$ могут наблюдаться следующие эффекты:

- Оптимум может лежать в области, где $\|\boldsymbol{\theta}\| \sim 0$, при этом матрица $\mathbf{C}$ в этой области плохо обусловлена,
- При $\|\boldsymbol{\theta}\| \gg 1$ матрица $\mathbf{C} \to I_n$, что приводит к вырожденной аппроксимации,
- Правдоподобие $\log p(\mathbf{u}|D, \boldsymbol{\theta}, \widetilde{\sigma})$ может слабо зависеть от значений параметров $\boldsymbol{\theta}$ и $\widetilde{\sigma}$.

В случае, когда максимум правдоподобия лежит в области малых значений параметров $\|\boldsymbol{\theta}\| \sim 0$, матрица ковариаций $\mathbf{C}$ плохо обусловлена. Из формулы (2.1) следует, что аппроксимация $\widetilde{u}(\mathbf{x})$ может быть представлена в виде

$\widetilde{u}(\mathbf{x}) = \sum_{i=1}^{n} \alpha_i c(\mathbf{x}, \mathbf{x}_i)$, где $\boldsymbol{\alpha} = \left(\mathbf{C}_{\boldsymbol{\theta}} + \widetilde{\sigma}^2 I_n\right)^{-1} \mathbf{u}$. Если значение $\widetilde{\sigma}$ недостаточно большое, то из-за плохой обусловленности матрицы $\mathbf{C}_{\boldsymbol{\theta}}$ норма вектора $\boldsymbol{\alpha}$ принимает большие значения, и в результате в аппроксимации $\widetilde{u}(\mathbf{x})$ становится существенным численный шум. В случае, когда максимум правдоподобия лежит в области больших значений параметров $\|\boldsymbol{\theta}\| \gg 1$ также происходит вырождение аппроксимационной модели: при $\widetilde{\sigma} \approx 0$ для $(\mathbf{x}_i, u_i) \in \mathbf{S}$ выполнено $\widetilde{u}(\mathbf{x}_i) \approx u_i$, при этом $\widetilde{u}(\mathbf{x}) \approx \frac{1}{n} \sum_{i=1}^{n} u_i$ для $\mathbf{x} \notin D$.

Байесовский вывод

Описанные выше проблемы можно решить с помощью байесовской регуляризации параметров ковариационной функции. Предположим, что параметры $\boldsymbol{\theta}$ распределены с некоторой априорной плотностью $\overline{p}(\boldsymbol{\theta}|\gamma)$, где γ — гиперпараметры априорного распределения параметров ковариационной функции. Тогда логарифм совместной плотности распределения данных и параметров модели имеет вид:

$$\log P(\mathbf{u}, \boldsymbol{\theta} | D, \sigma, \widetilde{\sigma}, \gamma) = \log\ p(\mathbf{u} | D, \boldsymbol{\theta}, \sigma, \widetilde{\sigma}) + \log\ \overline{p}(\boldsymbol{\theta}|\gamma). \tag{2.5}$$

Слагаемое $\log\ \overline{p}(\boldsymbol{\theta}|\gamma)$ выполняет роль регуляризации значений параметров ковариационной функции $\boldsymbol{\theta}$. Максимизация совместной апостериорной плотности параметров ковариационной функции и выборки (2.5) по $\widetilde{\sigma}, \gamma, \boldsymbol{\theta}'$ дает MAP-оценку (Maximum Aposteriori Probability) параметров модели $\widetilde{\sigma}, \boldsymbol{\theta}'$ и параметров априорного распределения γ. В данной работе в качестве априорного распределения $\overline{p}(\boldsymbol{\theta}|\gamma)$ используются нормальное и гамма- распределения.

Для нормального априорного распределения на параметры ковариационной функции рассмотрим следующее преобразование вектора $\boldsymbol{\theta}$:

$$\boldsymbol{\theta} = \boldsymbol{\theta}_{\parallel} + \boldsymbol{\theta}_{\perp},$$

где $\boldsymbol{\theta}_{\parallel} = |\boldsymbol{\theta}_{\parallel}| \cdot \mathbf{e}$, $\mathbf{e} = \left(\frac{1}{\sqrt{p}}, \ldots, \frac{1}{\sqrt{p}}\right)$, а вектор $\boldsymbol{\theta}_{\perp}$ ортогонален вектору $\mathbf{e}$. Предположим, что величина $|\boldsymbol{\theta}_{\parallel}|$ распределена нормально со средним $\mu_{\parallel} > 0$ и дисперсией $d_{\parallel}$, а вектор $\boldsymbol{\theta}_{\perp}$ не зависит от $|\boldsymbol{\theta}_{\parallel}|$ и имеет нормальное распределение с нулевым средним и диагональной ковариационной матрицей $d_\perp I_\perp$, где $I_\perp$ — единичная матрица размера $(p-1)$, $d_\perp$ — дисперсия компонент вектора $\boldsymbol{\theta}_\perp$. Такое предположение основано на априорном предположении о параметрах ковариационной функции, которое состоит в том, что все компоненты вектора $\boldsymbol{\theta}$ не слишком велики и не слишком малы по величине и не сильно отличаются друг от друга. Таким образом, в данном случае $\gamma = \{\mu_\parallel, d_\parallel, d_\perp\}$, а $\log \overline{p}(\boldsymbol{\theta}|\gamma) = \log p_{\mathcal{N}(\mu_\parallel, d_\parallel)}(\boldsymbol{\theta}) + \log p_{\mathcal{N}(\mathbf{0}, d_\perp I_\perp)}(\boldsymbol{\theta})$, где $p_{\mathcal{N}(\mu,\Sigma)}(\cdot)$ — функция плотности нормального распределения со средним μ и ковариационной матрицей Σ.

Опишем теперь априорное гамма распределение параметров ковариационной функции. Пусть все компоненты вектора $\boldsymbol{\theta}$ независимы и распределены с одним и тем же гамма-распределением с параметрами α и β. В таком случае

$$\log\ \overline{p}(\boldsymbol{\theta}|\gamma) = \log\ \overline{p}(\boldsymbol{\theta}|\alpha,\beta) = \sum_{i=1}^{d} \big[\alpha \log \beta + (\alpha - 1)\log \theta_i - \beta\theta_i - \log \Gamma(\alpha)\big], \tag{2.6}$$

где $\alpha > 1, \beta > 0$, а $\Gamma(\alpha)$ — гамма-функция. Такой выбор априорного распределения параметров штрафует слишком большие и слишком маленькие значения θ_i. Отметим, что аналогичным образом можно проводить регуляризацию параметра θ_0.

Раздел 2.1.2 содержит ряд экспериментальных результатов, которые показывают что оба предложенных метода байесовской регуляризации приводят к заметному улучшению надежности работы алгоритма и качеству полученной модели.

Многоуровневый байесовский вывод

Для того, чтобы процедуры, основанные на байесовском подходе, могли использоваться для решения прикладных задач, необходимо обеспечить разумный выбор параметров априорного распределения [80]. Часто исследователь не может выбрать параметры априорного распределения для каждой задачи — если, например, число задач велико, или разработанный алгоритм будет использоваться как черный ящик, то есть, пользователь не будет обладать необходимыми знаниями для выбора параметров априорного распределения в байесовском подходе.

В данном разделе предложен многоуровневый байесовский вывод [17] для модели регрессии на основе гауссовских процессов, позволяющий избежать описанной выше проблемы. Идея подхода состоит во введении дополнительного гиперaприорного распределения гиперпараметров априорного распределения параметров исходной модели. Процедура выбора модели в таком случае состоит в выборе параметров модели, гиперпараметров априорного распределения и параметров гипераприорного распределения исходя из принципа максимума апостериорной вероятности. При этом, выбор параметров гипераприорного распределения меньше влияет на функционал правдоподобия, а вводимая этим распределением регуляризация позволяет избежать получения вырожденных значений гиперпараметров априорного распределения при оптимизации (2.5) и, как следствие, параметры ковариационной функции не будут вырожденными.

После введения дополнительного априорного распределения получим следующую вероятностную модель:

- априорное гамма-распределение $\mathit{\Gamma}(\alpha, \beta)$ на параметры ковариационной функции $\boldsymbol{\theta}$,
- априорное гамма-распределение $\mathit{\Gamma}(\alpha_\sigma, \beta_\sigma)$ на параметр ковариационной функции σ,
- гипераприорное лог-нормальное распределение с заранее фиксированными параметрами μ, σ^2_{prior} на каждый из параметров $\alpha, \beta, \alpha_\sigma, \beta_\sigma$ байесовской регуляризации.

Тогда совместная плотность распределения будет пропорциональна:

$$p(\mathbf{u}|D, \boldsymbol{\theta}, \sigma^2, \alpha, \beta, \alpha_\sigma, \beta_\sigma, \mu, \sigma^2_{prior}) \propto p(\mathbf{u}|D, \boldsymbol{\theta}, \sigma^2) \cdot \mathit{\Gamma}(\boldsymbol{\theta}|\alpha, \beta) \cdot \mathit{\Gamma}(\sigma^2|\alpha_\sigma, \beta_\sigma) \times$$

$$\times \log\mathcal{N}(\alpha|\mu,\sigma^2_{prior})\cdot\log\mathcal{N}(\beta|\mu,\sigma^2_{prior})\cdot\log\mathcal{N}(\alpha_\sigma|\mu,\sigma^2_{prior})\cdot\log\mathcal{N}(\beta_\sigma|\mu,\sigma^2_{prior}),$$

где $\mathit{\Gamma}(\cdot)$ — плотность гамма-распределения, а $\log\mathcal{N}(\cdot)$ — плотность лог-нормального распределения. Подставляя явные выражения для распределений получаем явное выражение для максимизации:

$$\begin{aligned}
&\log p(\mathbf{u}|D,\boldsymbol{\theta},\sigma^2,\alpha,\beta,\alpha_\sigma,\beta_\sigma,\mu,\sigma^2_{prior}) \propto \\
&\propto -\frac{1}{2}\left[n\log 2\pi + \log|\mathbf{C}_{\boldsymbol{\theta},\sigma}| + \mathbf{u}^T\mathbf{C}^{-1}_{\boldsymbol{\theta},\sigma}\mathbf{u}\right] - \\
&-\log\mathit{\Gamma}(\alpha) + \alpha\log\beta + (\alpha-1)\sum_{i=1}^{d}\log\theta_i - \beta\sum_{i=1}^{d}\theta_i - \\
&-\log\mathit{\Gamma}(\alpha_\sigma) + \alpha_\sigma\log\beta_\sigma + (\alpha_\sigma-1)\log\sigma^2 - \beta_\sigma\sigma^2 - \\
&-\frac{1}{2}\left[2\log\alpha + \frac{1}{\sigma^2_{prior}}(\log\alpha-\mu)^2 + 2\log\beta + \frac{1}{\sigma^2_{prior}}(\log\beta-\mu)^2\right] - \\
&-\frac{1}{2}\left[2\log\alpha_\sigma + \frac{1}{\sigma^2_{prior}}(\log\alpha_\sigma-\mu)^2 + 2\log\beta_\sigma + \frac{1}{\sigma^2_{prior}}(\log\beta_\sigma-\mu)^2\right].
\end{aligned}$$

2.1.2 Вычислительные эксперименты

Ниже представлены полученные результаты вычислительных экспериментов при использовании байесовской регуляризации и многоуровневой байесовской регуляризации.

Байесовская регуляризация

В данном разделе приведено сравнение байесовского подхода с подходом на основе оценки максимума правдоподобия для задачи аппроксимации большого набора тестовых функций.

В экспериментах сравнивается работа трех алгоритмов оценки параметров модели регрессии на основе гауссовских процессов:

- MLE GP — метод максимума правдоподобия для оценки параметров;
- Gamma GP — байесовский метод оценки параметров с априорным гамма-

распределением;

- Normal GP — байесовский метод оценки параметров с априорным нормальным распределением.

Сравнение алгоритмов проводилось с помощью кривых Долана-Мора [30], которые позволяют наглядно представить результаты тестирования нескольких алгоритмов на большом наборе задач. Пусть у нас есть набор алгоритмов $a_i, i = 1, \ldots, A$ и набор задач $t_j, j = 1, \ldots, T$. Тогда для каждого алгоритма мы можем измерить качество его работы (например, среднюю абсолютную ошибку на тестовой выборке) на каждой задаче и получить набор чисел q_{ij}. Пусть лучший алгоритм для рассматриваемой задачи имеет качество $q_j, j = 1, \ldots, T$, такое что $\forall i = 1, \ldots, A$: $q_{ij} \geq q_j$. Определим $r_{ij} = \frac{q_{ij}}{q_j} \geq 1$. Чем ближе r_{ij} к единице — тем лучше сработал алгоритм a_i на задаче t_j. Будем менять параметр τ от 1 до ∞. Для каждого τ определим точку $c_i(\tau)$ на кривой Долана-Мора для алгоритма a_i как долю задач, на которых r_{ij} меньше, чем τ. С увеличением τ величина $c_i(\tau)$ стремится к единице, и чем выше лежит кривая $c_i(\tau)$ — тем лучше работает алгоритм по сравнению с другими.

Для демонстрации экспериментальных результатов использовался набор из 30 тестовых функций, которые применяются для тестирования задач оптимизации [85, 84, 46]. Для каждой функции равномерно случайно генерировалось по 2 обучающие выборки с размерами 10, 20, 40, 80 и 160 точек. 95% квантиль абсолютной ошибки на контрольных выборках размером 10000 точек служила мерой качества алгоритмов.

Результаты представлены в Таблице 2.1. В качестве меры качества работы алгоритма так же использовалась площадь под кривой Долана-Мора. Оба алгоритма с байесовской регуляризацией показали результаты лучше, чем алгоритм без регуляризации. Таким образом, использование байесовской регуляризации позволяет улучшить качество модели и устранить случаи появления вырожденных аппроксимаций.

Таблица 2.1 – Площадь под кривой Долана-Море для трех различных алгоритмов оценки параметров модели регрессии на основе гауссовских процессов

	Площадь под кривой Долана-Море
MLE GP	1.926
Gamma GP	1.930
Normal GP	1.951

Многоуровневая байесовская регуляризация

При оптимизации (2.5) по параметрам $\gamma = (\alpha, \beta)$ априорного распределения могут возникать ситуации, в которых, например, $\beta \approx 0$. Для такого β регуляризация не штрафует большие значения параметров $\boldsymbol{\theta} = \{\theta_i\}_{i=1}^p$, и возможно вырождение итоговой аппроксимации в константу с "пиками" в точках обучающей выборки. Приведем результаты тестирования многоуровневого байесовского вывода, предложенного в разделе 2.1.1, проводимого с целью понять, наблюдается ли описанное выше поведение на практике.

Тестирование проводилось на большом наборе функций разных входных размерностей. Параметр μ логнормального распределения был выбран равным 1.5. Сравнивались две меры качества аппроксимации: среднеквадратичная ошибка и мера изменчивости аппроксимации, позволяющая детектировать ее вырождение. Мера изменчивости аппроксимации оценивалась по значениям $\widetilde{\mathbf{u}} = \widetilde{u}(D_*)$ и $\mathbf{u}_*$ согласно формуле

$$\text{variability} = \frac{\max(\widetilde{\mathbf{u}}) - \min(\widetilde{\mathbf{u}})}{\max(\mathbf{u}_*) - \min(\mathbf{u}_*)},$$

где $\max(\mathbf{u})$ — максимальное значение вектора $\mathbf{u}$. Для каждого размера выборки подсчитывалось количество задач, для которых изменчивость меньше 10^{-4} (в этом случае полагаем, что аппроксимация выродилась в константу).

В экспериментах параметр σ^2_{prior} брался равным $1.5, 1, 0.5, 0.3, 0.2$. Кривая Долана-Море, построенная по значению среднеквадратичной ошибки, приведена на рисунке 2.1. Видно, что при уменьшении значения σ^2_{prior} точность аппроксимации уменьшается.

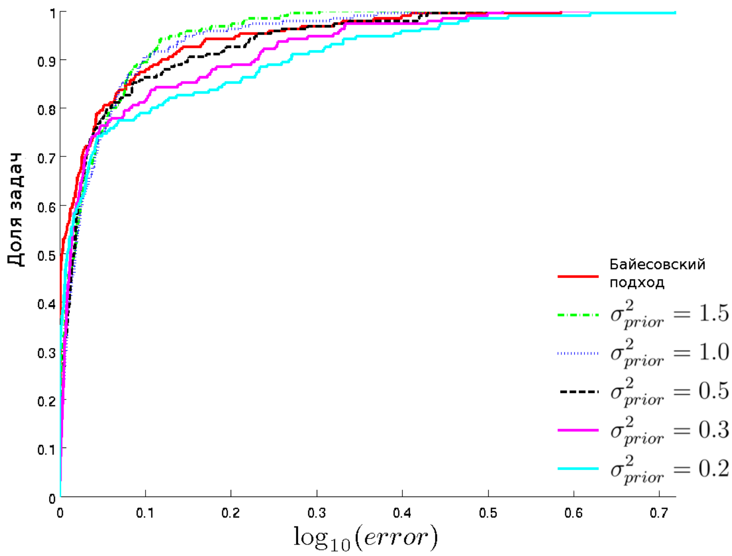


Рис. 2.1 – Кривая Долана-Мора для байесовской регуляризации и многоуровневой байесовской регуляризации при различных значениях σ^2_{prior}

В таблице 2.2 приведены результаты статистики по количеству случаев построения моделей с ошибками аппроксимации, близкими к ошибкам аппроксимации константой. Уже для значения $\sigma^2_{prior} = 0.5$ вырожденных регрессионных моделей нет, в то же время среднеквадратичная ошибка модели остается примерно такой же, как и при использовании обычного байесовского подхода.

Таблица 2.2 – Выбор параметра σ^2_{prior} в многоуровневой байесовской регуляризации (МБР): число задач, на которых изменчивость аппроксимации меньше 10^{-4}, то есть произошло вырождение итоговой аппроксимации. Всего рассматривалось 190 тестовых задач для каждого размера обучающей выборки n, который варьировался в пределах от 10 и до 320 точек.

Размер выборки	Байесовская регуляризация	Значение σ^2_{prior} для МБР				
		1.5	1	0.5	0.3	0.2
10	16	0	0	0	0	0
20	2	0	0	0	0	0
40	31	0	0	0	0	0
80	37	14	0	0	0	0
160	24	15	1	0	0	0
320	3	1	0	0	0	0

Таким образом, предложенный подход позволяет избежать вырожденных моделей регрессии на основе гауссовских процессов, не ухудшая качество модели в среднем.

2.1.3 Выводы

Регрессия на основе гауссовских процессов широко используется для построения регрессионных моделей в прикладных задачах. Однако, метод максимума правдоподобия обладает рядом недостатков, связанных с тем, что он не учитывает наши априорные представления о регрессионной модели. Использование байесовской регуляризации позволяет на практике учесть наши априорные представления о модели и улучшить качество полученных суррогатных моделей.

В отличие от других пакетов для построения регрессионных моделей на основе гауссовских процессов, таких как DACE [62], предложенный метод полностью автоматизирован. Экспериментальные результаты показывают высокое качество работы предложенного алгоритма по сравнению с обычными методами, особенно для пространственно неоднородных функций.

2.2 Теорема Бернштейна-фон Мизеса для регрессии на основе гауссовских процессов

В регрессии на основе гауссовских процессов предполагается, что заданная выборка значений целевой функции является реализацией гауссовского процесса, и ковариационная функция процесса зависит только от взаимного расположения точек в пространстве. Апостериорное среднее гауссовского процесса в новой точке используется для прогноза целевой функции в этой точке, а апостериорная дисперсия используется в качестве оценки неопределенности прогноза.

Обычно предполагают, что ковариационная функция гауссовского процесса лежит в параметрическом семействе. Тогда задание регрессионной модели эквивалентно выбору параметров ковариационной функции. Для оценки параметров используют метод максимального правдоподобия и байесовские методы, особенно эффективные в случае наличия априорных знаний о модели [68, 80].

Теорема Бершнтейна-фон Мизеса [100] обосновывает использование байесовского подхода: теорема утверждает, что апостериорное распределение асимптотически нормальное, а математическое ожидание и ковариационная матрица близки к оценке максимума правдоподобия и ее ковариационной матрице соответственно. Классическая версия теоремы Бернштейна-фон Мизеса, изложенная, например, Л. Ле Камом [59], использует типичные для математической статистики условия: предполагается, что параметрическое предположение выполняется (введенное параметрическое семейство содержит истинную модель) и размер выборки стремится к бесконечности (такие результаты верны только асимптотически).

Подобные условия использовались и для получения теоретических результатов для регрессии на основе гауссовских процессов в работах К. Мардиа[65], Б. Шаби[86, 49] и Т. Чу[26], при этом некоторые результаты были получены лишь для случая ограничения ковариации [86]. Однако, в реальных задачах параметрическое предположение как правило нарушено, а размер выборки конечен. Более того, для размера выборки стремящегося к бесконечности оценка функции в новой точке будет достаточно точной, даже если ковариационная функция изначально была выбрана неправильно [93].

Отметим, что для некоторых классов моделей современные версии теоремы Бернштейна-фон Мизеса учитывают возможность ошибочности параметрического предположения и содержат неасимптотические результаты [70, 69, 90]. Как правило, рассматриваются выборки с независимыми одинаково распределенными наблюдениями. Однако, модель регрессии на основе гауссовских процессов предполагает, что корреляции между наблюдениями существенны, таким образом, существующие подходы к доказательству теоремы Бернштейна-фон Мизеса для других классов параметрических моделей могут быть использованы лишь частично.

Данный раздел диссертации содержит теорему Бернштейна-фон Мизеса для регрессии на основе гауссовских процессов. Полученный результат выполнен для конечного размера выборки и возможной неверной спецификации параметрического предположения. В разделе 2.2.1 содержится утверждение теоремы и введенные предположения. Раздел 2.2.2 описывает примеры ковариационных функций, которые удовлетворяет введенным предположениям. В разделе 2.2.3 приведены результаты проведенных вычислительных экспериментов. Раздел 2.2.4 содержит доказательство сформулированных в этой части диссертации утверждений.

2.2.1 Утверждение теоремы

Обозначим $\|\cdot\|_2$ спектральную норму матрицы, $I_n \in \mathbb{R}^{n\times n}$ обозначим единичную матрицу, $I\!E\{\cdot\}$ обозначим математическое ожидание, $\mathrm{Var}\{\cdot\}$ обозначим оператор ковариации относительно распределения $\mathcal{GP}\left(0, c(\mathbf{x}, \mathbf{x}')\right)$ рассматриваемого гауссовского процесса.

Далее будем рассматривать неинформативное априорное распределение параметров $\Pi(d\boldsymbol{\theta})$ и возможную неправильную спецификацию параметрической модели $c(\mathbf{x}, \mathbf{x}') \notin \{c_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{x}'), \, \boldsymbol{\theta} \in \Theta\}$, то есть реальное распределение данных может не соответствовать выбранному параметрическому семейству моделей.

Пусть выполнены следующие предположения:

(A1) Существует центральная точка $\boldsymbol{\theta}^* = \mathrm{argmax}_{\boldsymbol{\theta}\in\Theta} I\!EL(\boldsymbol{\theta})$.

(A2) Ковариационная функция $c_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{x}')$ три раза непрерывно дифференцируема относительно $\boldsymbol{\theta} \in \Theta$ для $\mathbf{x}, \mathbf{x}' \in \mathbb{X}$,

(A3) Существуют такие константы $0 < \overline{\lambda} < \infty$ и $0 < \lambda_0 < \infty$, что $\|\mathbf{C}\|_2 \le \overline{\lambda}, \|\mathbf{C}_{\boldsymbol{\theta}}\|_2 \le \overline{\lambda}$, $\|\mathbf{C}^{-1}\|_2 \le \lambda_0, \|\mathbf{C}_{\boldsymbol{\theta}}^{-1}\|_2 \le \lambda_0^{-1}$. для $\boldsymbol{\theta} \in \Theta$,

(A4) Существуют такие константы $0 < \lambda_1, \lambda_2, \lambda_3 < \infty$, что $\left\|\frac{\partial \mathbf{C}_{\boldsymbol{\theta}}}{\partial \theta_i}\right\|_2 \le \lambda_1$, $\left\|\frac{\partial^2 \mathbf{C}_{\boldsymbol{\theta}}}{\partial \theta_i \partial \theta_j}\right\|_2 \le \lambda_2$, $\left\|\frac{\partial^3 \mathbf{C}_{\boldsymbol{\theta}}}{\partial \theta_i \partial \theta_j \partial \theta_m}\right\|_2 \le \lambda_3$ для $\boldsymbol{\theta} \in \Theta$ и для всех $i, j, m \in \{1, \ldots, p\}$.

(A5) Существует достаточно маленькая константа $C > 0$, такая что $\frac{\lambda_i}{\lambda_0} < C$ для $i = 1, 2, 3$.

(A6) Минимальное собственное значение матрицы $\frac{1}{n} D_0^2$ больше чем $\mathsf{d}_0 > 0$, и минимальное собственное значение матрицы $\frac{1}{n} V_0^2$ больше чем $\mathsf{v}_0 > 0$, где

$$D_0^2 = -\nabla^2 I\!\!E L(\boldsymbol{\theta})\big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \text{ и } V_0^2 = \operatorname{Var}\bigl\{\nabla L(\boldsymbol{\theta})\bigr\}\big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}.$$

Эти предположения близки к предположениям, которые используются для получение асимптотических результатов для процедур оценки параметров гауссовских процессов.

Мы рассматриваем свойства апостериорного среднего и апостериорной ковариационной матрицы для вектора параметров модели. Апостериорное среднее $\overline{\boldsymbol{\theta}}$ является байесовской оценкой для центральной точки $\boldsymbol{\theta}^*$:

$$\overline{\boldsymbol{\theta}} \stackrel{\text{def}}{=} I\!\!E\bigl\{\boldsymbol{\theta} \,\big|\, \mathbf{S}\bigr\}.$$

Апостериорная ковариационная матрица имеет вид:

$$S^2 \stackrel{\text{def}}{=} I\!\!E\left\{\left(\boldsymbol{\theta}-\overline{\boldsymbol{\theta}}\right)\left(\boldsymbol{\theta}-\overline{\boldsymbol{\theta}}\right)^{\top} \,\big|\, \mathbf{S}\right\}.$$

Ниже мы рассматриваем некоторую окрестность центральной точки $\boldsymbol{\theta}^* = \operatorname{argmax}_{\boldsymbol{\theta}\in\Theta} I\!\!E L(\boldsymbol{\theta})$:

$$\Theta_0(\mathtt{r}_0) = \{\boldsymbol{\theta}\in\Theta : \|D_0(\boldsymbol{\theta}-\boldsymbol{\theta}^*)\| \le \mathtt{r}_0\}.$$

Теорема 2.1. *Пусть предположения (A1)—(A6) выполняются, и для размера выборки n выполнено:*

$$n \ge 4C\mathtt{r}_0^2 p^3$$

для фиксированной константы $C > 0$, зависящей только от констант в предположениях (A1)—(A6). Тогда существует $\diamondsuit(\mathtt{r}_0,\mathtt{x}) \le \frac{\diamondsuit_0(\mathtt{r}_0,\mathtt{x})}{\sqrt{n}}$ и случайное множество $\Omega(\mathtt{x})$ с вероятностью по меньшей мере $1-5e^{-\mathtt{x}}$ такие, что на $\Omega(\mathtt{x})$

$$\mathbb{P}(\widetilde{\boldsymbol{\theta}} \notin \Theta_0(\mathtt{r}_0)) \le 3e^{-\mathtt{x}}, \tag{2.7}$$

$$\mathbb{P}(\overline{\boldsymbol{\theta}} \notin \Theta_0(\mathtt{r}_0)) \le 5e^{-\mathtt{x}}. \tag{2.8}$$

Так же выполнено для $\Delta_\circ = \mathtt{r}_0\diamondsuit(\mathtt{r}_0,\mathtt{x})$

$$\|D_0(\overline{\boldsymbol{\theta}}-\widetilde{\boldsymbol{\theta}})\|^2 \le 4\Delta_\circ(\mathtt{x}) + \diamondsuit^2(\mathtt{r}_0,\mathtt{x}) + 4e^{-\mathtt{x}}, \tag{2.9}$$

$$\|I_p - D_0 S^2 D_0\|_2 \le 4\mathit{\Delta}_\circ(\mathtt{x}) + 4e^{-\mathtt{x}}. \tag{2.10}$$

Более того, для любого измеримого множества $A \subset I\!\!R^p$ *и вектора* $\boldsymbol{\gamma} \propto \mathcal{N}(\mathbf{0}, I_p)$ *выполнено*

$$I\!\!P\left(D_0(\boldsymbol{\theta} - \widetilde{\boldsymbol{\theta}}) \in A|\mathbf{S}\right) \ge \tag{2.11}$$
$$\ge \exp\left\{-2\mathit{\Delta}_\circ(\mathtt{x}) - 3e^{-\mathtt{x}}\right\}\left(I\!\!P(\boldsymbol{\gamma} \in A) - \diamondsuit(\mathtt{r}_0, \mathtt{x})\right) - e^{-\mathtt{x}},$$
$$I\!\!P\left(D_0(\boldsymbol{\theta} - \widetilde{\boldsymbol{\theta}}) \in A|\mathbf{S}\right) \le \tag{2.12}$$
$$\le \quad \exp\left\{2\mathit{\Delta}_\circ(\mathtt{x}) + 2e^{-\mathtt{x}}\right\}\left(I\!\!P(\boldsymbol{\gamma} \in A) + \diamondsuit(\mathtt{r}_0, \mathtt{x})\right) + e^{-\mathtt{x}}.$$

Полученная теорема типа теоремы Бернштейна-фон Мизеса отличается от доказанных ранее:

- Результаты выполняются для конечных выборок — мы не требуем стремления размера выборки к бесконечности,
- Параметрическое предположение может быть неправильно специфицировано — мы не требуем корректной спецификации параметрической модели.

Неравенства (2.7) и (2.8) утверждают, что оценка максимума правдоподобия и байесовская оценка близки к центральной точке, и расстояние до центральной точки убывает со скоростью $\frac{1}{\sqrt{n}}$, так как мы можем выбрать $\mathtt{r}_0$ так, что с ростом выборки $\mathtt{r}_0$ будет убывать как $\frac{1}{\sqrt{n}}$. (для правильно специфицированной модели центральная точка совпадает с вектором истинных значений ковариационной функции, для неправильно специфицированной модели центральная точка — точка, которая определяет ковариационную функцию, ближайшую к истинной ковариационной функции). Неравенства (2.9) и (2.10) показывают, что байесовская оценка $\overline{\boldsymbol{\theta}}$ близка к оценке максимального правдоподобия $\widetilde{\boldsymbol{\theta}}$, и ковариационная матрица апостериорного распределения S^2 близка к матрице D_0^{-2}. Неравенства (2.11) и (2.12) утверждают, что апостериорное распределение $\mathrm{Law}(\boldsymbol{\theta} \mid \mathbf{S})$ близко к нормальному, а именно расстояние полной вариации между апостериорным распределением и нормальным мало.

Приведем значения констант в теореме:

$$\begin{aligned}
&\mathtt{g} \stackrel{\text{def}}{=} \frac{\sqrt{n\mathsf{v}_0}}{2\sqrt{p}\frac{1}{\lambda_0^2}\overline{\lambda}\lambda_1}, B \stackrel{\text{def}}{=} D_0^{-1}V_0^2D_0^{-1},\\
&\mathsf{v} \geq \frac{1}{n}\|V_0^2\|_2 \text{ и } \mathsf{d} \geq \frac{1}{n}\|D_0^2\|_2,\\
&\mathtt{p}_B \stackrel{\text{def}}{=} \operatorname{tr}(B) \leq p\left(\mathsf{v}/\mathsf{d}_0\right)^2, \mathtt{v}_B^2 \stackrel{\text{def}}{=} 2\operatorname{tr}(B^2) \leq 2\left(\mathsf{v}/\mathsf{d}_0\right)^4,\\
&\lambda_B \stackrel{\text{def}}{=} \lambda_{\max}(B) \leq p\left(\mathsf{v}/\mathsf{d}_0\right)^2, \mathtt{g}_c \stackrel{\text{def}}{=} \sqrt{\mathtt{g}^2 - 2\mathtt{p}_B/3}, \mathtt{y}_c^2 \leq \mathtt{p}_B + 6\lambda_B\mathtt{x}_c,\\
&2\mathtt{x}_c \stackrel{\text{def}}{=} \left(3/2\mathtt{g}^2 - \mathtt{p}_B\right)/\lambda_B + \log\|I_p - 2B/(3\lambda_B)\|_2,
\end{aligned}$$

$$z(B,\mathtt{x}) \stackrel{\text{def}}{=} \begin{cases} \mathtt{p}_B + 2\mathtt{v}_B\mathtt{x}^{1/2}, & \mathtt{x} \leq \mathtt{v}_B/(18\lambda_B),\\ \mathtt{p}_B + 6\lambda_B\mathtt{x}, & \mathtt{v}_B/(18\lambda_B) < \mathtt{x} \leq \mathtt{x}_c,\\ |\mathtt{y}_c + 2\lambda_B(\mathtt{x}-\mathtt{x}_c)/\mathtt{g}_c|, & \mathtt{x} > \mathtt{x}_c, \end{cases}$$

$$\begin{aligned}
&z_{\mathbb{H}}(\mathtt{x}) \stackrel{\text{def}}{=} \begin{cases} \sqrt{4p + 2\mathtt{x}}, & 4p + 2\mathtt{x} \leq \mathtt{g}^2,\\ \mathtt{g}^{-1}\mathtt{x} + \frac{1}{2}(\mathtt{g}^{-1}4p + \mathtt{g}), & 4p + 2\mathtt{x} > \mathtt{g}^2, \end{cases}\\
&C_3 \stackrel{\text{def}}{=} 4\frac{1}{\lambda_0^6}\lambda_1^3\overline{\lambda}^3 + 5.5\frac{1}{\lambda_0^4}\overline{\lambda}^2\,\lambda_1\lambda_2 + \frac{1}{\lambda_0^2}\overline{\lambda}\lambda_3,\\
&\delta(\mathtt{r}) \stackrel{\text{def}}{=} \frac{C_3p^{\frac{3}{2}}}{\sqrt{n}\mathsf{d}_0^{\frac{3}{2}}}, \nu_0^2 \stackrel{\text{def}}{=} \max\left(1, 2\left(\frac{\sqrt{p}}{\sqrt{\mathsf{v}_0}}\frac{1}{\lambda_0^2}\overline{\lambda}\lambda_1\right)^2\right), \mathtt{b} = \frac{\mathsf{d}_0}{2\mathsf{d}},\\
&\diamondsuit(\mathtt{r},\mathtt{x}) \stackrel{\text{def}}{=} \left(\delta(\mathtt{r}) + 6\frac{1}{n}\nu_0 z_{\mathbb{H}}(\mathtt{x})\right)\mathtt{r}.
\end{aligned}$$

2.2.2 Примеры ковариационных функций, удовлетворяющих предположениям теоремы

Рассмотрим широко используемое параметрическое семейство квадратичных экспоненциальных ковариационных функций:

$$c_{\boldsymbol{\theta}}(\mathbf{x},\mathbf{x}') = \theta_1^2 \exp\left(-\theta_2^2 \sum_{i=1}^{d}(x_i - x_i')^2\right) + \sigma^2\delta(\mathbf{x}-\mathbf{x}'), \tag{2.13}$$

где $\delta(\cdot)$ обозначает функцию Кронекера. Первое слагаемое в (2.13) определяет ковариацию между наблюдениями гауссовского процесса, второе слагаемое задает дополнительный белый шум с дисперсией σ^2.

Такая ковариационная функция удовлетворяет предположениям (A1)—(A6), а именно выполнена следующая теорема.

Теорема 2.2. *Для плана экспериментов* $D = \{\mathbf{x}_i\}_{i=1}^n$, *такого что* $x_{ij} \in \{\sigma m, m \in \mathbb{Z}\}$, *и достаточно большой дисперсии шума* $\sigma^2 > 0$ *для параметрического семейств квадратичных экспоненциальных ковариационных функций* (2.13) *выполнены предположения (A1)—(A6) и выполнено условие* (2.35) *на собственные числа ковариационных матриц для правильно специфицированной модели.*

Теорема 2.3. *Если ковариационная функция имеет вид* (2.38), *то существует такой дизайн экспериментов, что выполнены предположения (A1)—(A6) без необходимости выполнения условия типа* (2.35).

Подробное доказательство приведенных утверждений содержится в разделе 2.2.4.

2.2.3 Вычислительные эксперименты

Мы проводим вычислительные эксперименты для различных размерностей пространства параметров. Используются следующие ковариационные функции: изотропная квадратичная экспоненциальная ковариационная функция (2.13) и анизотропная квадратичная экспоненциальная ковариационная функция

$$c_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{x}') = \exp\left(-\sum_{i=1}^{d} \theta_i^2 (x_i - x_i')^2\right) + \sigma^2 \delta(\mathbf{x} - \mathbf{x}'). \tag{2.14}$$

Мы рассматриваем случай известной дисперсии шума $\sigma^2 = 0.001$. В дисертации в качестве неинформативного априорного распределения выступает многомерное равномерное распределение на гиперкубе $\Theta = [0, \theta_1^{\max}] \times \ldots \times [0, \theta_p^{\max}]$.

Мы генерируем выборку $\mathbf{S}$ для параметров ковариационной функции $\boldsymbol{\theta} \in \Theta$ используя, что совместное распределение $\mathbf{u}$ является нормальным с нулевым средним и заданной ковариационной матрицей $\mathbf{C}_{\boldsymbol{\theta}} = \{c_{\boldsymbol{\theta}}(\mathbf{x}_i, \mathbf{x}_j)\}_{i,j=1}^n$: для точек $D = \{\mathbf{x}_i\}_{i=1}^n$, $\mathbf{x}_i \in \mathbb{X} = [0,1]^d$ из равномерного распределения на гиперкубе $\mathbb{X} = [0,1]^d$ мы подсчитываем ковариационную матрицу и генерируем нормальный случайный вектор для заданной ковариационной матрицы.

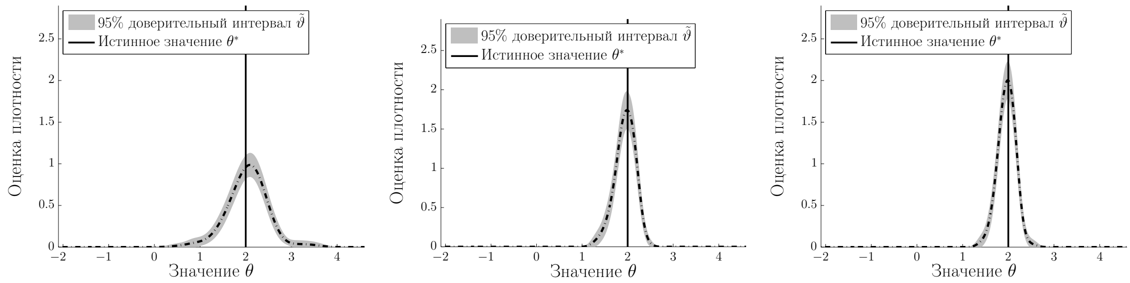
Рассмотрим как распределение полученных оценок максимума правдоподобия $\widetilde{\boldsymbol{\theta}}$ и байесовских оценок $\overline{\boldsymbol{\theta}}$ зависит от размера выборки; неравенства (2.7) и (2.8) утверждают, что ОМП $\widetilde{\boldsymbol{\theta}}$ и байесовские оценки $\overline{\boldsymbol{\theta}}$ концентрируются вокруг центральной точки с возрастанием размера выборки.

Для выбранного размера выборки сгенерируем 200 различных выборок по заданному правилу и получим $\widetilde{\boldsymbol{\theta}}$ и $\overline{\boldsymbol{\theta}}$ для каждой сгенерированной выборки, то есть получим всего 200 различных $\widetilde{\boldsymbol{\theta}}$ и $\overline{\boldsymbol{\theta}}$. Используя такие $\widetilde{\boldsymbol{\theta}}$ и $\overline{\boldsymbol{\theta}}$, сделаем непараметрическую оценку плотности (используется ядерная оценка плотности с гауссовским ядром [104], ширина ядра подбирается с помощью процедуры скользящего контроля).

Рассмотрим одномерный случай $p = 1$ для ковариационной функции (2.14) и двумерный случай $p = 2$ для ковариационной функции (2.13). Для одномерного случая приведем оценки плотности и 95% доверительные интервалы оценки плотности для $\overline{\boldsymbol{\theta}}$ на рисунке 2.2. Для двумерного случая приведем оценки плотности для $\widetilde{\boldsymbol{\theta}}$ и $\overline{\boldsymbol{\theta}}$ на рисунке 2.3.

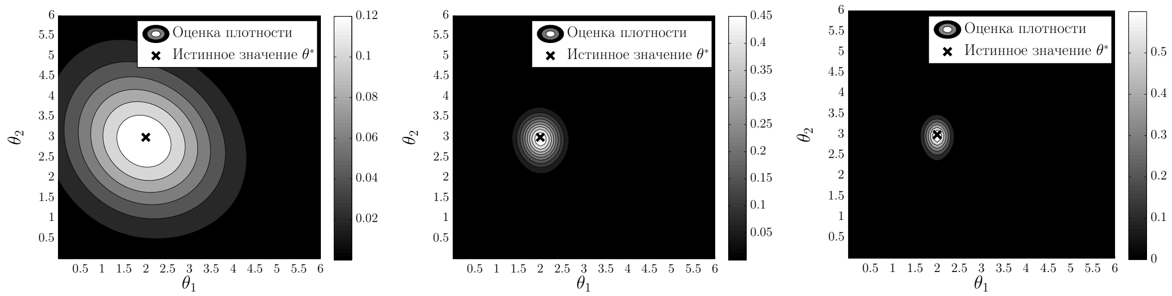
В обоих случаях с увеличением размера выборки распределения оценок $\widetilde{\boldsymbol{\theta}}$ и $\overline{\boldsymbol{\theta}}$ концентрируются около настоящего значения $\boldsymbol{\theta}^*$. Кроме того на рисунке 2.3 видно, что форма распределений ОМП $\widetilde{\boldsymbol{\theta}}$ и байесовских оценок $\overline{\boldsymbol{\theta}}$ сходная.

Рассмотрим верхние оценки уклонений апостериорного среднего и ковариационной матрицы. Будем проводить эксперимент для одномерной ($p = 1$) и двумерной ($p = 2$) ковариационных функций. В Теореме 2.1 показано, что нормы $\left\| D_0 \left(\overline{\boldsymbol{\theta}} - \widetilde{\boldsymbol{\theta}} \right) \right\|_2^2$ и $\left\| I_p - D_0 S^2 D_0 \right\|_2$ ограничены сверху величинами, которые убывают со скоростью $\frac{1}{\sqrt{n}}$ с увеличением размера выборки n. Рисунки 2.4б и 2.4г показывают как ведут себя эти уклонения для одномерной квадратичной экс-

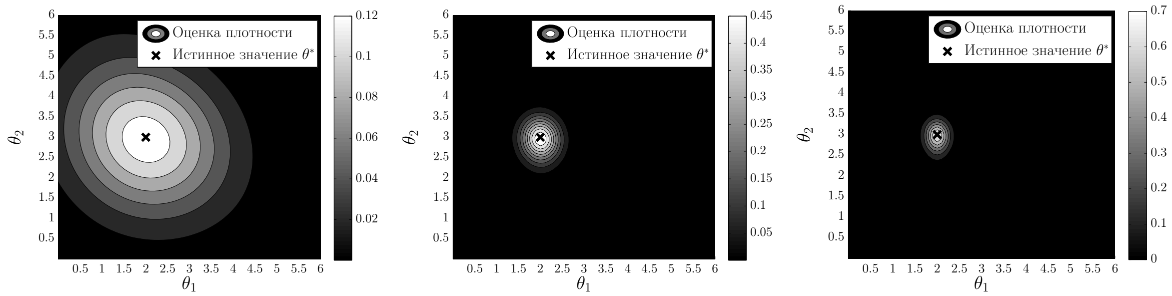


(а) Размер выборки $n = 10$ (б) Размер выборки $n = 200$ (в) Размер выборки $n = 1000$

Рис. 2.2 – Ядерная оценка плотности для $\overline{\boldsymbol{\theta}}$, одномерное пространство параметров $p = 1$



(а) Размер выборки $n = 10$, ЯОП для $\widetilde{\boldsymbol{\theta}}$ (б) Размер выборки $n = 200$, ЯОП для $\widetilde{\boldsymbol{\theta}}$ (в) Размер выборки $n = 1000$, ЯОП для $\widetilde{\boldsymbol{\theta}}$

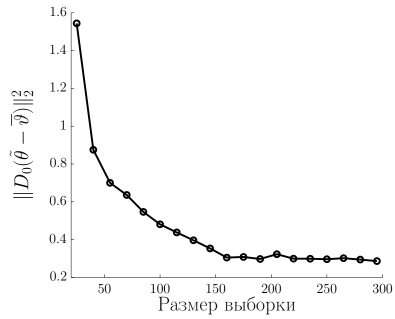


(г) Размер выборки $n = 10$, ЯОП для $\overline{\boldsymbol{\theta}}$ (д) Размер выборки $n = 200$, ЯОП для $\overline{\boldsymbol{\theta}}$ (е) Размер выборки $n = 1000$, ЯОП для $\overline{\boldsymbol{\theta}}$

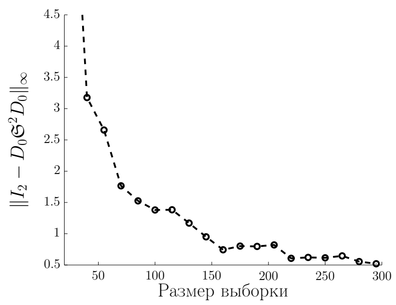
Рис. 2.3 – Ядерная оценка плотности (ЯОП) для $\widetilde{\boldsymbol{\theta}}$ и $\overline{\boldsymbol{\theta}}$, двумерное пространство параметров $p = 2$

поненциальной ковариационной функции (2.14) и двумерной изотропной квадратичной экспоненциальной ковариационной функции (2.13) соответственно. И в одномерном случае, и в двумерном случае уклонения уменьшаются с увеличением размера выборки n.

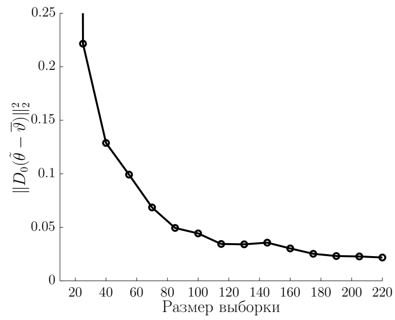
Рассмотрим так же шестимерное пространство параметров $p = 6$ и ковариационную функцию (2.14). Из рисунка 2.6 видно, как величина $\left\| D_0 \left(\overline{\boldsymbol{\theta}} - \widetilde{\boldsymbol{\theta}} \right) \right\|_2^2$ убывает с увеличением размера выборки n.



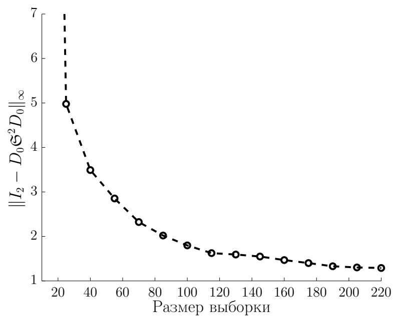
(а) Зависимость $\left\| D_0 \left(\overline{\boldsymbol{\theta}} - \widetilde{\boldsymbol{\theta}} \right) \right\|_2^2$ от размера выборки, $p = 1$



(б) Зависимость $\left\| I_p - D_0 S^2 D_0 \right\|_2$ от размера выборки, $p = 1$



(в) Зависимость $\left\| D_0 \left(\overline{\boldsymbol{\theta}} - \widetilde{\boldsymbol{\theta}} \right) \right\|_2^2$ от размера выборки, $p = 2$



(г) Зависимость $\left\| I_p - D_0 S^2 D_0 \right\|_2$ от размера выборки, $p = 2$

Рис. 2.4 – Зависимость (2.9) and (2.10) от размера выборки n для размерности пространства параметров p равной 1 и 2

Теорема Бернштейна-фон-Мизеса утверждает, что апостериорное распределение вектора параметров $\boldsymbol{\theta}$ близко к нормальному распределению: расстояние полной вариации между этими распределениями уменьшается с увеличением размера выборки n. Мы сравниваем апостериорное распределение параметров ковариационной функции и нормальное распределение с тем же средним значением $\overline{\boldsymbol{\theta}}$ и такой же ковариационной матрицей S^2. Мы приводим результаты

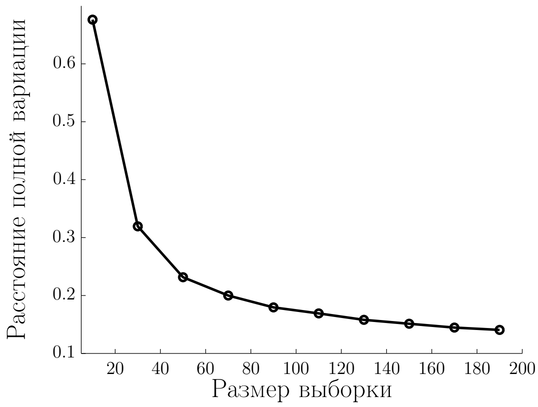


Рис. 2.5 – Зависимость расстояния полной вариации между апостериорным распределением вектора параметров ковариационной функции и соответствующим нормальным распределением, размерность пространства параметров равна двум, $p = 2$

для ковариационной функции (2.13), для которой выполнены предположения теоремы БвМ.

На рисунке 2.5 показано, что расстояние полной вариации между апостериорным распределением и ближайшим нормальным распределением уменьшается с увеличением размера выборки n. Следовательно, оценки (2.11) и (2.12) действительно имеют место.

Неравенство (2.1) определяет минимальный необходимый размер выборки n, который гарантирует выполнение теоремы БвМ. Рассмотрим риск оценок параметров ковариационной функции. На рисунке 2.7 показана зависимость качества оце$\|\boldsymbol{\theta}^* - \boldsymbol{\theta}^*\|$ от размера выборки n и размерности пространства параметров p. Для размерности p если размер выборки n больше критического значения, байесовские оценки параметров достаточно точны. Следовательно, можно оценить размер выборки, необходимый для получения достаточной точной байесовской оценки параметров.

2.2.4 Доказательство теоремы

Доказательство приведенной теоремы построено следующим образом. Показано, что если выполнены предположения из раздела 2.2.1, то выполнены глобальные условия $(\boldsymbol{ED_2})$, $(\mathscr{L})$ и локальные условия $(\boldsymbol{ED_0})$, $(\mathscr{L}_\mathbf{0})$, $(\boldsymbol{\mathcal{I}})$. Эти

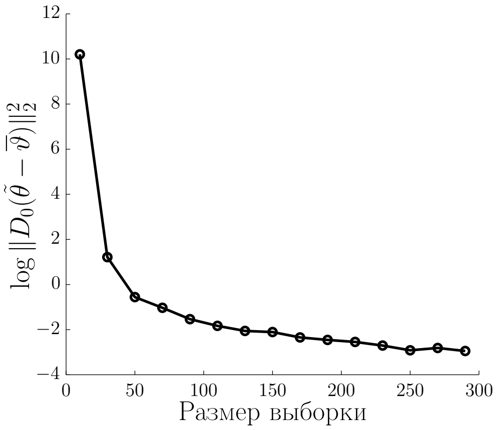


Рис. 2.6 – Зависимость $\left\| D_0 \left(\overline{\boldsymbol{\theta}} - \widetilde{\boldsymbol{\theta}} \right) \right\|_2^2$ от размера выборки, размерность пространства параметров $p = 6$

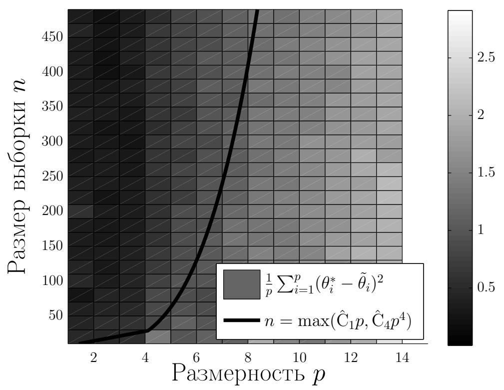


Рис. 2.7 – Зависимость риска оценки параметров ковариационной функции от размер выборки n и размерности пространства параметров p

оценки экспоненциальных моментов квазиправдоподобия $L(\boldsymbol{\theta})$ и его производных позволяют нам использовать современную теорию эмпирических процессов [97, 19] для получения для конечной выборки гауссовской аппроксимации логарифма квази-правдоподобия и соответствующего апостериорного распределения. Следовательно, выполнена теорема Бернштейна-фон Мизеса [97, 89, 90].

Вспомогательные утверждения

В этом разделе собраны вспомогательные результаты, которые необходимы для доказательства основных результатов. Так же введены необходимые обозначения.

Логарифм правдоподобия выборки $\mathbf{S}$ зависит от параметров ковариационной функции $\boldsymbol{\theta}$

$$L(\boldsymbol{\theta}) = -\frac{1}{2}\left[n \log 2\pi + \ln |\mathbf{C}_{\boldsymbol{\theta}}| + \mathbf{u}^{\mathsf{T}} \mathbf{C}_{\boldsymbol{\theta}}^{-1} \mathbf{u}\right],$$

где $\mathbf{C}_{\boldsymbol{\theta}} = \{c_{\boldsymbol{\theta}}(\mathbf{x}_i, \mathbf{x}_j)\}_{i,j=1}^{n}$, $\mathbf{u} \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$, $\mathbf{C} \in \mathbb{R}^{n \times n}$ — истинная ковариационная матрица.

Обозначим $\zeta(\boldsymbol{\theta}) = L(\boldsymbol{\theta}) - I\!\!E L(\boldsymbol{\theta})$ и обозначим $L(\boldsymbol{\theta}, \boldsymbol{\theta}^*) = L(\boldsymbol{\theta}) - L(\boldsymbol{\theta}^*)$

эксцесс.

Лемма 1. *Математическое ожидание логарифма правдоподобия имеет вид*

$$\mathbb{E}L(\boldsymbol{\theta}) = -\frac{1}{2}\left[n\log 2\pi + \ln|\mathbf{C}_{\boldsymbol{\theta}}| + \operatorname{tr}\left(\mathbf{C}_{\boldsymbol{\theta}}^{-1}\mathbf{C}\right)\right]. \tag{2.15}$$

Доказательство. Так как $\mathbf{u}^\mathsf{T}\mathbf{C}_{\boldsymbol{\theta}}^{-1}\mathbf{u} = \varepsilon^\mathsf{T}\Omega_{\boldsymbol{\theta}}\varepsilon$, где $\varepsilon \sim \mathcal{N}(\mathbf{0}, I)$, $I \in \mathbb{R}^{n\times n}$ — единичная матрица, $S_{\boldsymbol{\theta}} = \mathbf{C}^{1/2}\mathbf{C}_{\boldsymbol{\theta}}^{-1}\mathbf{C}^{1/2}$. Используя Лемму B.1 из [91], получаем $\mathbb{E}\mathbf{u}^\mathsf{T}\mathbf{C}_{\boldsymbol{\theta}}^{-1}\mathbf{u} = \operatorname{tr}(S_{\boldsymbol{\theta}}) = \operatorname{tr}\left(\mathbf{C}_{\boldsymbol{\theta}}^{-1}\mathbf{C}\right)$. □

Используем здесь стандартные формулы для матричных производных (они приведены, например, в [75]). Тогда матрица $\frac{\partial \mathbf{C}_{\boldsymbol{\theta}}}{\partial\theta_i}$ может быть вычислена по формуле

$$\frac{\partial \mathbf{C}_{\boldsymbol{\theta}}^{-1}}{\partial\theta_i} = -U_i.$$

Частные производные логарифма детерминанта ковариационной матрицы равны

$$\frac{\partial \ln|\mathbf{C}_{\boldsymbol{\theta}}|}{\partial\theta_i} = \operatorname{tr}\left(\mathbf{C}_{\boldsymbol{\theta}}^{-1}\frac{\partial \mathbf{C}_{\boldsymbol{\theta}}}{\partial\theta_i}\right) = \operatorname{tr}(U_i\mathbf{C}_{\boldsymbol{\theta}}).$$

Частные производные следа матрица равны

$$\frac{\partial \operatorname{tr}\mathbf{C}_{\boldsymbol{\theta}}}{\partial\theta_i} = \operatorname{tr}\left[\frac{\partial \mathbf{C}_{\boldsymbol{\theta}}}{\partial\theta_i}\right].$$

Частная производная логарифма правдоподобия имеет вид

$$\nabla_i L(\boldsymbol{\theta}) = -\frac{1}{2}\left[\operatorname{tr}\left(\mathbf{C}_{\boldsymbol{\theta}}^{-1}\frac{\partial \mathbf{C}_{\boldsymbol{\theta}}}{\partial\theta_i}\right) - \mathbf{u}^\mathsf{T}\mathbf{C}_{\boldsymbol{\theta}}^{-1}\frac{\partial \mathbf{C}_{\boldsymbol{\theta}}}{\partial\theta_i}\mathbf{C}_{\boldsymbol{\theta}}^{-1}\mathbf{u}\right].$$

Частная производная математического ожидания логарифма правдоподобия равна

$$\nabla_i \mathbb{E}L(\boldsymbol{\theta}) = -\frac{1}{2}\left[\operatorname{tr}\left(\mathbf{C}_{\boldsymbol{\theta}}^{-1}\frac{\partial \mathbf{C}_{\boldsymbol{\theta}}}{\partial\theta_i}\right) - \operatorname{tr}\left(\mathbf{C}_{\boldsymbol{\theta}}^{-1}\frac{\partial \mathbf{C}_{\boldsymbol{\theta}}}{\partial\theta_i}\mathbf{C}_{\boldsymbol{\theta}}^{-1}\mathbf{C}\right)\right]. \tag{2.16}$$

Для того, чтобы записать явное выражение для матрицы V_0^2 нужно проинтегрировать функционал вида $(\mathbf{u}^\mathsf{T}A\mathbf{u})(\mathbf{u}^\mathsf{T}B\mathbf{u})$ по многомерной нормальной плотности. Выполнена следующая лемма [75]

Лемма 2. *Пусть* $A = \{a_{ij}\}_{i,j=1}^n$ *и* $B = \{b_{ij}\}_{i,j=1}^n$ *— симметричные матрицы, и* $\mathbf{C} = \{c_{ij}\}_{i,j=1}^n$ *— положительно определенная матрица. Тогда для* $\mathbf{u} \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$

выполнено, что

$$\mathbb{E}(\mathbf{u}^\top A\mathbf{u})(\mathbf{u}^\top B\mathbf{u}) = 2\operatorname{tr}(A\mathbf{C}B\mathbf{C}) + \operatorname{tr}(A\mathbf{C})\operatorname{tr}(B\mathbf{C}).$$

Лемма 3. *Матрица* $V_0^2 = \operatorname{Var}\bigl\{\nabla\zeta(\boldsymbol{\theta}^*)\bigr\}$ *имеет вид*

$$V_0^2 = \left\{\frac{1}{2}\operatorname{tr}(U_i\mathbf{C}U_j\mathbf{C})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\right\}_{i,j=1}^p . \tag{2.17}$$

Доказательство. Используя Лемму 2 с $A = U_i$, $B = U_j$ и $\mathbf{u} \sim \mathcal{N}(\mathbf{0},\mathbf{C}), \mathbf{C} \in \mathbb{R}^{n\times n}$, получаем

$$\mathbb{E}(\mathbf{u}^\top U_i\mathbf{u})(\mathbf{u}^\top U_j\mathbf{u}) = 2\operatorname{tr}\left(U_i\mathbf{C}U_j\mathbf{C}\right) + \operatorname{tr}\left(U_i\mathbf{C}\right)\operatorname{tr}(U_j\mathbf{C})\,. \tag{2.18}$$

Так как

$$V_0^2 = \left\{\frac{1}{4}\left(\mathbb{E}(\mathbf{u}^\top U_i\mathbf{u})(\mathbf{u}^\top U_j\mathbf{u}) - \mathbb{E}(\mathbf{u}^\top U_i\mathbf{u})\mathbb{E}(\mathbf{u}^\top U_j\mathbf{u})\right)\big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\right\}_{i,j=1}^p ,$$

то используя (2.18), мы получаем утверждение леммы. □

Лемма 4. *Матрица* $D_0^2 = -\nabla^2\, \mathbb{E}L(\boldsymbol{\theta})|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}$ *имеет вид:*

$$D_0^2 = \left\{\frac{1}{2}\operatorname{tr}\Big[U_i\mathbf{C}_{\boldsymbol{\theta}}U_j\mathbf{C} + U_j\mathbf{C}_{\boldsymbol{\theta}}U_i\mathbf{C} - U_i\mathbf{C}_{\boldsymbol{\theta}}U_j\mathbf{C}_{\boldsymbol{\theta}} + U_{ij}(\mathbf{C}_{\boldsymbol{\theta}} - \mathbf{C})\Big]\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\right\}_{i,j=1}^p .$$

Доказательство. Мы получаем матрицу $D_0^2 = -\nabla^2\, \mathbb{E}L(\boldsymbol{\theta})|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}$ непосредственным вычислением из (2.16):

$$\begin{aligned}\frac{\partial^2\mathbb{E}L(\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j} &= -\frac{1}{2}\frac{\partial}{\partial\theta_j}\left[\operatorname{tr}\left(\mathbf{C}_{\boldsymbol{\theta}}^{-1}\frac{\partial\mathbf{C}_{\boldsymbol{\theta}}}{\partial\theta_i}\right) - \operatorname{tr}\left(\mathbf{C}_{\boldsymbol{\theta}}^{-1}\frac{\partial\mathbf{C}_{\boldsymbol{\theta}}}{\partial\theta_i}\mathbf{C}_{\boldsymbol{\theta}}^{-1}\mathbf{C}\right)\right] = \\ &= -\frac{1}{2}\operatorname{tr}\left[-U_j\frac{\partial\mathbf{C}_{\boldsymbol{\theta}}}{\partial\theta_i} + \mathbf{C}_{\boldsymbol{\theta}}^{-1}\frac{\partial^2\mathbf{C}_{\boldsymbol{\theta}}}{\partial\theta_i\partial\theta_j} + U_j\frac{\partial\mathbf{C}_{\boldsymbol{\theta}}}{\partial\theta_i}\mathbf{C}_{\boldsymbol{\theta}}^{-1}\mathbf{C} - U_{ij}\mathbf{C} + \mathbf{C}_{\boldsymbol{\theta}}^{-1}\frac{\partial\mathbf{C}_{\boldsymbol{\theta}}}{\partial\theta_i}U_j\mathbf{C}\right].\end{aligned}$$

□

Лемма 5. *Для симметричных матриц* $A, B \in \mathbb{R}^{n\times n}$ *выполнено, что*

$$\operatorname{tr}(AB) \le \frac{1}{2}\left(\|A\|_F^2 + \|B\|_F^2\right).$$

Доказательство.

$$\frac{1}{2}\left(\|A\|_F^2 + \|B\|_F^2\right) = \frac{1}{2}\sum_{i,j=1}^{n} a_{ij}a_{ji} + b_{ij}b_{ji} \geq \sum_{i,j=1}^{n} a_{ij}b_{ji} = \operatorname{tr}(AB).$$

□

Лемма 6. *Для матрицы* $A \in \mathbb{R}^{n\times n}$ *выполнено, что* $|\operatorname{tr}(A)| \leq n\|A\|_2$.

Доказательство. Выполнено

$$|\operatorname{tr}(A)|^2 = \left|\sum_{i=1}^{n} \lambda_i(A)\right|^2 \leq n\sum_{i=1}^{n} \lambda_i^2(A) = n\operatorname{tr}(A^2) \leq$$
$$\leq n\operatorname{tr}(A^T A) = n\sum_{i=1}^{n} \lambda_i(A^\mathsf{T} A) \leq n^2\lambda_{\max}(A^\mathsf{T} A) = n^2\|A\|_2^2.$$

Следовательно,

$$|\operatorname{tr}(A)| \leq n\|A\|_2.$$

□

Мы получаем ниже верхние оценки для норм V_0^2 и D_0^2.

Лемма 7. *Существует* $0 < \mathsf{v}^2 < \infty$*, которое не зависит от размера выборки, такое что*

$$\|V_0^2\|_2^2 \leq n^2\mathsf{v}^2.$$

Доказательство. Лемма 3 дает точный вид V_0^2:

$$V_0^2 = \left\{\frac{1}{2}\operatorname{tr}(U_i\mathbf{C}U_j\mathbf{C})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\right\}_{i,j=1}^{p}.$$

Получим верхние оценки для элементов $\operatorname{tr}(U_i\mathbf{C}U_j\mathbf{C})$ матрицы V_0^2.

$$|\operatorname{tr}(U_i\mathbf{C}U_j\mathbf{C})| \leq \frac{1}{2}\left(\|U_i\mathbf{C}\|_F^2 + \|U_j\mathbf{C}\|_F^2\right) \leq n\max_{i=\overline{1,p}}\|U_i\mathbf{C}\|_2^2.$$

Для $\|U_i\mathbf{C}\|_2$ выполнено, что:

$$\|U_i\mathbf{C}\|_2 = \left\|\mathbf{C}_{\boldsymbol{\theta}}^{-1}\frac{\partial\mathbf{C}_{\boldsymbol{\theta}}}{\partial\theta_i}\mathbf{C}_{\boldsymbol{\theta}}^{-1}\mathbf{C}\right\|_2 \leq \|\mathbf{C}_{\boldsymbol{\theta}}^{-1}\|_2\left\|\frac{\partial\mathbf{C}_{\boldsymbol{\theta}}}{\partial\theta_i}\right\|_2\|\mathbf{C}_{\boldsymbol{\theta}}^{-1}\|_2\|\mathbf{C}\|_2 \leq \frac{1}{\lambda_0^2}\overline{\lambda}\lambda_1.$$

Следовательно,

$$\|V_0^2\|_{\max} = \max_{i,j=\overline{1,p}} \frac{1}{2} \left|\mathrm{tr}(U_i \mathbf{C} U_j \mathbf{C})\right| \le \frac{1}{2} n \frac{1}{\lambda_0^4} \overline{\lambda}^2 \lambda_1^2.$$

Тогда спектральная норма матрицы V_0^2 ограничена сверху

$$\|V_0^2\|_2 \le p \|V_0^2\|_{\max} \le \frac{1}{2} p n \frac{1}{\lambda_0^4} \overline{\lambda}^2 \lambda_1^2 = n \mathsf{v},$$

где v имеет вид

$$\mathsf{v} = \frac{1}{2} p \frac{1}{\lambda_0^4} \overline{\lambda}^2 \lambda_1^2.$$

$\square$

Аналогичным способом получим оценку сверху для спектральной нормы D_0^2.

Лемма 8. *Существует* $0 < \mathsf{d}^2 < \infty$, *которое не зависит от размера выборки, такое что*

$$\|D_0^2\|_2^2 < n^2 \mathsf{d}^2.$$

Доказательство. Лемма 4 дает точный вид матрицы D_0^2:

$$D_0^2 = \left\{ \frac{1}{2} \mathrm{tr} \Big[U_i \mathbf{C}_{\boldsymbol{\theta}} U_j \mathbf{C} + U_j \mathbf{C}_{\boldsymbol{\theta}} U_i \mathbf{C} - U_i \mathbf{C}_{\boldsymbol{\theta}} U_j \mathbf{C}_{\boldsymbol{\theta}} + U_{ij} (\mathbf{C}_{\boldsymbol{\theta}} - \mathbf{C}) \Big] \Big|_{\boldsymbol{\theta} = \boldsymbol{\theta}^*} \right\}_{i,j=1}^{p}.$$

Теперь ограничим все компоненты матрицы D_0^2. Из Леммы 5:

$$\|\mathrm{tr}\left(U_i \mathbf{C}_{\boldsymbol{\theta}} U_j \mathbf{C}_{\boldsymbol{\theta}}\right)\| \le \frac{1}{2} \left(\|U_i \mathbf{C}_{\boldsymbol{\theta}}\|_F^2 + \|U_j \mathbf{C}_{\boldsymbol{\theta}}\|_F^2 \right) \le n \max_{i \in \{1,2,\ldots,p\}} \|U_i \mathbf{C}_{\boldsymbol{\theta}}\|_2^2 \le n \frac{1}{{\lambda_0}^4} \overline{\lambda}^2 \lambda_1^2.$$

Аналогично получаем:

$$\mathrm{tr}\left(U_i \mathbf{C} U_j \mathbf{C}_{\boldsymbol{\theta}}\right) \le n \frac{1}{{\lambda_0}^4} \overline{\lambda}^2 \lambda_1^2,$$
$$\mathrm{tr}\left(U_i \mathbf{C} U_j \mathbf{C}\right) \le n \frac{1}{{\lambda_0}^4} \overline{\lambda}^2 \lambda_1^2.$$

Так же нам нужна верхняя оценка для $\|\mathrm{tr}(U_{ij}(\mathbf{C}_{\boldsymbol{\theta}} - \mathbf{C}))\|$:

$$\begin{aligned} \|\mathrm{tr}(U_{ij}(\mathbf{C}_{\boldsymbol{\theta}} - \mathbf{C}))\| &\le \frac{1}{2} \left(\|\mathbf{C}_{\boldsymbol{\theta}}^{-1} \frac{\partial^2 \mathbf{C}_{\boldsymbol{\theta}}}{\partial \theta_i \partial \theta_j} \mathbf{C}_{\boldsymbol{\theta}}^{-1}\|_F^2 + \|\mathbf{C} - \mathbf{C}_{\boldsymbol{\theta}}\|_F^2 \right) \le \\ &\le \frac{1}{2} \left(\|\mathbf{C}_{\boldsymbol{\theta}}^{-1} \frac{\partial^2 \mathbf{C}_{\boldsymbol{\theta}}}{\partial \theta_i \partial \theta_j} \mathbf{C}_{\boldsymbol{\theta}}^{-1}\|_F^2 + 2\|\mathbf{C}\|_F^2 + 2\|\mathbf{C}_{\boldsymbol{\theta}}\|_F^2 \right) \le \end{aligned}$$

$$\leq \frac{1}{2} n \left(\|\mathbf{C}_{\boldsymbol{\theta}}^{-1}\|_2^2 \|\frac{\partial^2 \mathbf{C}_{\boldsymbol{\theta}}}{\partial\theta_i \partial\theta_j}\|_2^2 \|\mathbf{C}_{\boldsymbol{\theta}}^{-1}\|_2^2 + 2\|\mathbf{C}\|_2^2 + 2\|\mathbf{C}_{\boldsymbol{\theta}}\|_2^2 \right) \leq$$
$$\leq \frac{1}{2} n \left(\frac{1}{\lambda_0^4} \lambda_2^2 + 4\overline{\lambda}^2 \right)$$

Используя полученные выше результаты, запишем верхнюю оценку для $\|D_0^2\|_{\max}$:

$$\|D_0^2\|_{\max} = \frac{1}{2} \max_{i,j=\overline{1,p}} \left| \operatorname{tr}\Big[U_i \mathbf{C}_{\boldsymbol{\theta}} U_j \mathbf{C} + U_j \mathbf{C}_{\boldsymbol{\theta}} U_i \mathbf{C} - U_i \mathbf{C}_{\boldsymbol{\theta}} U_j \mathbf{C}_{\boldsymbol{\theta}} + U_{ij}(\mathbf{C}_{\boldsymbol{\theta}} - \mathbf{C}) \Big] \right|\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \leq$$
$$\leq \frac{1}{2} n \left(3 \frac{1}{{\lambda_0}^4} \overline{\lambda}^2 \lambda_1^2 + \frac{1}{2} \frac{1}{\lambda_0^4} \lambda_2^2 + 2\overline{\lambda}^2 \right).$$

Тогда для $\|D_0^2\|_2$ выполнено, что:

$$\|D_0^2\|_2 \leq p \|D_0^2\|_{\max} \leq pn \frac{1}{2} \left(3 \frac{1}{{\lambda_0}^4} \overline{\lambda}^2 \lambda_1^2 + \frac{1}{2} \frac{1}{\lambda_0^4} \lambda_2^2 + 2\overline{\lambda}^2 \right).$$

Наконец,

$$\|D_0^2\|_2^2 < n^2 \mathsf{d}^2,$$

для

$$\mathsf{d} = p \frac{1}{2} \left(3 \frac{1}{{\lambda_0}^4} \overline{\lambda}^2 \lambda_1^2 + \frac{1}{2} \frac{1}{\lambda_0^4} \lambda_2^2 + 2\overline{\lambda}^2 \right).$$

□

Лемма 9. *Для $\boldsymbol{\theta} \in \Theta$ выполнено, что*

$$\left| \frac{\partial^3 \mathbb{E} L(\boldsymbol{\theta})}{\partial\theta_i \partial\theta_j \partial\theta_k} \right| \leq n C_3, \tag{2.19}$$

с

$$C_3 = 4 \frac{1}{\lambda_0^6} \lambda_1^3 \overline{\lambda}^3 + 5.5 \frac{1}{\lambda_0^4} \overline{\lambda}^2 \ \lambda_1 \lambda_2 + \frac{1}{\lambda_0^2} \overline{\lambda} \lambda_3. \tag{2.20}$$

Доказательство. Заметим, что

$$\frac{\partial^2 \mathbb{E} L(\boldsymbol{\theta})}{\partial\theta_i \partial\theta_j} = -\frac{1}{2} \operatorname{tr} \left[U_j \mathbf{C}_{\boldsymbol{\theta}} U_i \mathbf{C} + U_i \mathbf{C}_{\boldsymbol{\theta}} U_j \mathbf{C} - U_i \mathbf{C}_{\boldsymbol{\theta}} U_j \mathbf{C}_{\boldsymbol{\theta}} + U_{ij}(\mathbf{C}_{\boldsymbol{\theta}} - \mathbf{C}) \right]$$

и

$$\frac{\partial U_i}{\partial \theta_k} = -U_k \mathbf{C}_{\boldsymbol{\theta}} U_i - U_i \mathbf{C}_{\boldsymbol{\theta}} U_k + U_{ij}.$$

Тогда выполнено, что

$$\frac{\partial^3 \mathbb{E}L(\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j\partial\theta_k} =$$

$$-\frac{1}{2}\operatorname{tr}\Big[-\big(U_k\mathbf{C}_{\boldsymbol{\theta}}U_j\mathbf{C}_{\boldsymbol{\theta}}U_i + U_j\mathbf{C}_{\boldsymbol{\theta}}U_k\mathbf{C}_{\boldsymbol{\theta}}U_i + U_j\mathbf{C}_{\boldsymbol{\theta}}U_i\mathbf{C}_{\boldsymbol{\theta}}U_k +$$
$$+ U_k\mathbf{C}_{\boldsymbol{\theta}}U_i\mathbf{C}_{\boldsymbol{\theta}}U_j + U_i\mathbf{C}_{\boldsymbol{\theta}}U_k\mathbf{C}_{\boldsymbol{\theta}}U_j + U_i\mathbf{C}_{\boldsymbol{\theta}}U_j\mathbf{C}_{\boldsymbol{\theta}}U_k\big)\mathbf{C}+$$
$$+ \left(U_i\mathbf{C}_{\boldsymbol{\theta}}U_k\mathbf{C}_{\boldsymbol{\theta}}U_j + U_j\mathbf{C}_{\boldsymbol{\theta}}U_k\mathbf{C}_{\boldsymbol{\theta}}U_i\right)\mathbf{C}_{\boldsymbol{\theta}}+$$
$$+ \left(U_{kj}\mathbf{C}_{\boldsymbol{\theta}}U_i + U_j\mathbf{C}_{\boldsymbol{\theta}}U_{ik} + U_{ik}\mathbf{C}_{\boldsymbol{\theta}}U_j + U_i\mathbf{C}_{\boldsymbol{\theta}}U_{kj}\right)\mathbf{C}-$$
$$- \left(U_{ik}\mathbf{C}_{\boldsymbol{\theta}}U_j - U_i\mathbf{C}_{\boldsymbol{\theta}}U_{kj}\right)\mathbf{C}_{\boldsymbol{\theta}} - U_k\mathbf{C}_{\boldsymbol{\theta}}U_{ij}(\mathbf{C}_{\boldsymbol{\theta}} - \mathbf{C})-$$
$$- U_{ij}\mathbf{C}_{\boldsymbol{\theta}}U_k(\mathbf{C}_{\boldsymbol{\theta}} - \mathbf{C}) + U_{ij}\mathbf{C}_{\boldsymbol{\theta}}U_k\mathbf{C}_{\boldsymbol{\theta}} + U_{ijk}(\mathbf{C}_{\boldsymbol{\theta}} - \mathbf{C})\Big].$$

Выполнены следующие оценки для спектральных норм

$$\|U_i\|_2 \le \frac{1}{\lambda_0^2}\lambda_1,$$
$$\|U_{ij}\|_2 \le \frac{1}{\lambda_0^2}\lambda_2,$$
$$\|U_{ijk}\|_2 \le \frac{1}{\lambda_0^2}\lambda_3.$$

Используя Лемму 6, мы получаем верхнюю оценку для $|\operatorname{tr}[U_i\mathbf{C}_{\boldsymbol{\theta}}U_j\mathbf{C}_{\boldsymbol{\theta}}U_k\mathbf{C}_{\boldsymbol{\theta}}]|$:

$$|\operatorname{tr}[U_i\mathbf{C}_{\boldsymbol{\theta}}U_j\mathbf{C}_{\boldsymbol{\theta}}U_k\mathbf{C}_{\boldsymbol{\theta}}]| \le n\|U_i\mathbf{C}_{\boldsymbol{\theta}}U_j\mathbf{C}_{\boldsymbol{\theta}}U_k\mathbf{C}_{\boldsymbol{\theta}}\|_2 \le n\|U_i\|_2\|U_j\|_2\|U_k\|_2\|\mathbf{C}_{\boldsymbol{\theta}}\|_2^3 \le n\frac{1}{\lambda_0^6}\lambda_1^3\overline{\lambda}^3.$$

Верхние оценки для $|\operatorname{tr}[U_i\mathbf{C}_{\boldsymbol{\theta}}U_j\mathbf{C}_{\boldsymbol{\theta}}U_k\mathbf{C}]|$, $|\operatorname{tr}[U_{ij}\mathbf{C}_{\boldsymbol{\theta}}U_k\mathbf{C}_{\boldsymbol{\theta}}]|$ и $|\operatorname{tr}[U_{ijk}(\mathbf{C}_{\boldsymbol{\theta}} - \mathbf{C})]|$ получаем аналогично:

$$|\operatorname{tr}[U_i\mathbf{C}_{\boldsymbol{\theta}}U_j\mathbf{C}_{\boldsymbol{\theta}}U_k\mathbf{C}_{\boldsymbol{\theta}}]| \le n\frac{1}{\lambda_0^6}\lambda_1^3\overline{\lambda}^3,$$
$$|\operatorname{tr}[U_i\mathbf{C}_{\boldsymbol{\theta}}U_j\mathbf{C}_{\boldsymbol{\theta}}U_k\mathbf{C}]| \le n\frac{1}{\lambda_0^6}\lambda_1^3\overline{\lambda}^3,$$
$$|\operatorname{tr}[U_{ij}\mathbf{C}_{\boldsymbol{\theta}}U_k\mathbf{C}_{\boldsymbol{\theta}}]| \le n\frac{1}{\lambda_0^4}\lambda_2\overline{\lambda}^2\lambda_1,$$
$$|\operatorname{tr}[U_{ijk}(\mathbf{C}_{\boldsymbol{\theta}} - \mathbf{C})]| \le 2n\frac{1}{\lambda_0^2}\overline{\lambda}\lambda_3.$$

Используя полученные верхние оценки, получаем утверждение Леммы. $\square$

Глобальные условия

В этом разделе доказаны условия о глобальных свойствах эмпирического процесса, необходимые для выполнения утверждения теоремы.

($\boldsymbol{ED_2}$) *Существует величина* $\omega > 0$, *и для каждого* $\mathtt{r} > 0$ *существует константа* $\mathtt{g}(\mathtt{r})$, *такие что для всех* $\boldsymbol{\theta} \in \Theta_0(\mathtt{r})$ *выполнено следующее неравенство*

$$\sup_{\boldsymbol{\beta},\boldsymbol{\gamma}\in I\!\!R^p} \log I\!\!E \exp\left\{\frac{\lambda}{\omega}\frac{\boldsymbol{\beta}^{\mathsf{T}}\nabla^2\zeta(\boldsymbol{\theta})\boldsymbol{\gamma}}{\|D_0\boldsymbol{\beta}\|_2\|D_0\boldsymbol{\gamma}\|_2}\right\} \le \nu_0^2\lambda^2/2, \qquad |\lambda| \le \mathtt{g}(\mathtt{r}). \tag{2.21}$$

Доказательство. Верно, что

$$\nabla_i\zeta(\boldsymbol{\theta}) = \frac{1}{2}\left[\mathbf{u}^{\mathsf{T}}U_i\mathbf{u} - \operatorname{tr}(U_i\mathbf{C})\right].$$

Тогда

$$\nabla_{ij}^2\zeta(\boldsymbol{\theta}) = -\frac{1}{2}\left[\mathbf{u}^{\mathsf{T}}A_{ij}\mathbf{u} - \operatorname{tr}(A_{ij}\mathbf{C})\right],$$

где

$$A_{ij} = A_{ij}(\boldsymbol{\theta}) = U_i\mathbf{C}_{\boldsymbol{\theta}}U_j - U_{ij} + U_j\mathbf{C}_{\boldsymbol{\theta}}U_i.$$

Выполнена следующая верхняя оценка для $\|A_{ij}\|_2$:

$$\begin{aligned}\|A_{ij}\|_2 &\le \sup_{i,j\in\{1,2,\dots,p\}} 2\|U_i\mathbf{C}_{\boldsymbol{\theta}}U_j\|_2 + \|U_{ij}\|_2 \le \\ &\le \sup_{i,j\in\{1,2,\dots,p\}} 2\|\mathbf{C}_{\boldsymbol{\theta}}^{-1}\frac{\partial\mathbf{C}_{\boldsymbol{\theta}}}{\partial\boldsymbol{\theta}_i}\mathbf{C}_{\boldsymbol{\theta}}^{-1}\frac{\partial\mathbf{C}_{\boldsymbol{\theta}}}{\partial\boldsymbol{\theta}_j}\mathbf{C}_{\boldsymbol{\theta}}^{-1}\|_2 + \|\mathbf{C}_{\boldsymbol{\theta}}^{-1}\frac{\partial^2\mathbf{C}_{\boldsymbol{\theta}}}{\partial\theta_i\partial\theta_j}\mathbf{C}_{\boldsymbol{\theta}}^{-1}\|_2 \le \\ &\le \sup_{i,j\in\{1,2,\dots,p\}} 2\|\mathbf{C}_{\boldsymbol{\theta}}^{-1}\|_2^3\|\frac{\partial\mathbf{C}_{\boldsymbol{\theta}}}{\partial\boldsymbol{\theta}_i}\|_2\|\frac{\partial\mathbf{C}_{\boldsymbol{\theta}}}{\partial\boldsymbol{\theta}_j}\|_2 + \|\mathbf{C}_{\boldsymbol{\theta}}^{-1}\|_2^2\|\frac{\partial^2\mathbf{C}_{\boldsymbol{\theta}}}{\partial\theta_i\partial\theta_j}\|_2 \le 2\frac{1}{\lambda_0^3}\lambda_1^2 + \frac{1}{\lambda_0^2}\lambda_2.\end{aligned} \tag{2.22}$$

Обозначим Z

$$Z = Z(\boldsymbol{\theta}) = \frac{1}{\|D_0\boldsymbol{\beta}\|_2\|D_0\boldsymbol{\gamma}\|_2}\sum_{i,j=1}^{p}\beta_i\gamma_jA_{ij}.$$

Для каждого $\boldsymbol{\gamma}$, такого что $\|\boldsymbol{\gamma}\| = 1$, выполнено

$$\|D_0\boldsymbol{\gamma}\| \geq \sqrt{\lambda_{\min}(D_0^2)} \geq \sqrt{n\mathsf{d}_0}.$$

Тогда используя (2.22), получаем:

$$\begin{aligned} \|Z\mathbf{C}\|_2 &= \left\| \frac{1}{\|D_0\boldsymbol{\gamma}\|\|D_0\boldsymbol{\beta}\|} \sum_{i,j=1}^{p} \beta_i \gamma_j A_{ij}\mathbf{C} \right\|_2 \leq \\ &\leq p^2 \frac{1}{\|D_0\boldsymbol{\gamma}\|\|D_0\boldsymbol{\beta}\|} \sup_{i,j=\overline{1,p}} \|A_{ij}\mathbf{C}\|_2 \leq \\ &\leq p^2 \frac{1}{n\mathsf{d}_0} \sup_{i,j=\overline{1,p}} \|A_{ij}\mathbf{C}\|_2 \\ &\leq p^2 \frac{1}{n\mathsf{d}_0} \left(2\frac{1}{\lambda_0^3}\lambda_1^2 + \frac{1}{\lambda_0^2}\lambda_2 \right). \end{aligned} \tag{2.23}$$

Мы выбираем ω как $\frac{1}{n}$ и $\mathtt{g}(\mathtt{r})$ как

$$\mathtt{g}(\mathtt{r}) = \frac{\mathsf{d}_0}{2p^2 \left(2\frac{\lambda_0^3}{\overline{\lambda}}^2 + \frac{1}{\lambda_0^2}\lambda_2 \right)}, \tag{2.24}$$

тогда для любого $|\lambda| \leq \mathtt{g}(\mathtt{r})$ выполнено, что

$$\|\lambda Z\mathbf{C}\|_2 \leq \frac{1}{2}.$$

Следовательно, матрица $I - \lambda Z\mathbf{C}$ положительно определена для $|\lambda| \leq \mathtt{g}(\mathtt{r})$.

Левая часть выражения (2.21) может быть переписана как

$$\sup_{\boldsymbol{\theta}\in\Theta_0(\mathtt{r})} \sup_{\boldsymbol{\beta},\boldsymbol{\gamma}\in I\!R^p} \log \left[\exp\left(-\frac{\lambda}{2}\operatorname{tr}(Z\mathbf{C}) \right) \int_{I\!R^n} \frac{1}{(2\pi)^{\frac{n}{2}}|\mathbf{C}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\mathbf{u}^T(\mathbf{C}^{-1} - \lambda Z)\mathbf{u} \right) d\mathbf{u} \right]. \tag{2.25}$$

Математическое ожидание в (2.25) существует, так как матрица $I - \lambda Z\mathbf{C}$ положительно определена для $|\lambda| \leq \mathtt{g}(\mathtt{r})$. Мы интегрируем математическое ожидание, и получаем, что (2.21) эквивалентно

$$\sup_{\theta\in\Theta_0(\mathtt{r})} \sup_{\boldsymbol{\beta},\boldsymbol{\gamma}\in I\!R^p} -\frac{1}{2} \left[\operatorname{tr} \lambda Z\mathbf{C} + \log|I - \lambda Z\mathbf{C}| \right] \leq \frac{\nu_0^2\lambda^2}{2}, |\lambda| \leq \mathtt{g}(\mathtt{r}).$$

Для положительно определенной $I - \lambda Z\mathbf{C}$ выполнено

$$\log|I - \lambda Z\mathbf{C}| = -\operatorname{tr}\sum_{i=1}^{\infty}\frac{1}{i}(\lambda Z\mathbf{C})^i.$$

Тогда

$$-\frac{1}{2}\left[\operatorname{tr}\lambda Z\mathbf{C} + \log|I - \lambda Z\mathbf{C}|\right] = \frac{1}{2}\operatorname{tr}\left(\sum_{i=2}^{\infty}\frac{1}{i}(\lambda Z\mathbf{C})^i\right).$$

Следовательно, достаточно показать

$$\sup_{\boldsymbol{\theta}\in\Theta_0(\mathtt{r})}\ \sup_{\boldsymbol{\beta},\boldsymbol{\gamma}\in I\!\!R^p}\frac{1}{2}\operatorname{tr}\left(\sum_{i=2}^{\infty}\frac{1}{i}(\lambda Z\mathbf{C})^i\right) \le \frac{\nu_0^2\lambda^2}{2}, |\lambda| < \mathtt{g}(\mathtt{r}).$$

Заметим, что

$$|\operatorname{tr}\left((\lambda Z\mathbf{C})^i\right)| = |\lambda^i\operatorname{tr}((Z\mathbf{C})^i)| = \left|\lambda^i\sum_{j=1}^{n}\lambda_j^i(Z\mathbf{C})\right| \le n|\lambda|^i\|Z\mathbf{C}\|_2^i.$$

Следовательно, достаточно показать

$$\sup_{\boldsymbol{\theta}\in\Theta_0(\mathtt{r})}\ \sup_{\boldsymbol{\beta},\boldsymbol{\gamma}\in I\!\!R^p} n\sum_{i=2}^{\infty}|\lambda|^i\|Z\mathbf{C}\|_2^i \le \nu_0^2\lambda^2, |\lambda| < \mathtt{g}.$$

Для $i \ge 3$ нужно доказать выполнение неравенства

$$n\,|\lambda|^{i-2}\,\|Z\mathbf{C}\|_2^i \le \frac{1}{2^{i-1}}\nu_0^2.$$

Используя верхнюю оценку (2.23) для $\|Z\mathbf{C}\|_2$, получаем

$$n\,|\lambda|^{i-2}\,\|Z\mathbf{C}\|_2^i \le n|\lambda|^{i-2}\left(p^2\frac{1}{n\mathsf{d}_0}\left(2\frac{1}{\lambda_0^3}\lambda_1^2 + \frac{1}{\lambda_0^2}\lambda_2\right)\right)^i.$$

Используя $\mathtt{g}(\mathtt{r})$ из (2.24) и ν_0^2 вида

$$\nu_0^2 = \max\left(1, 2\left(p^2\frac{1}{n\mathsf{d}_0}\left(2\frac{1}{\lambda_0^3}\lambda_1^2 + \frac{1}{\lambda_0^2}\lambda_2\right)\right)^2\right),$$

получаем, что для любого $|\lambda| \le \mathtt{g}(\mathtt{r})$ выполнено

$$n\,|\lambda|^{i-2}\,\|Z\mathbf{C}\|_2^i \le \left(p^2\frac{1}{n\mathsf{d}_0}\left(2\frac{1}{\lambda_0^3}\lambda_1^2 + \frac{1}{\lambda_0^2}\lambda_2\right)\right)^2\frac{1}{2^{i-2}} \le \frac{\nu_0^2}{2^{i-1}}.$$

Следовательно, так как $n\|Z\mathbf{C}\|_2^2 \le \frac{\nu_0^2}{2}$, то выполнено

$$\sup_{\boldsymbol{\theta}\in\Theta_0(\mathtt{r})} \sup_{\boldsymbol{\beta},\boldsymbol{\gamma}\in\mathbb{R}^p} n\sum_{i=2}^{\infty} |\lambda|^i \|Z\mathbf{C}\|_2^i \le \lambda^2\nu_0^2 \sum_{i=2}^{\infty} \frac{1}{2^{i-1}} \le \nu_0^2\lambda^2, |\lambda| < \mathtt{g}(\mathtt{r}).$$

Таким образом, для выбранных $\nu_0^2 \ge 1$ и $\mathtt{g}(\mathtt{r})$ утверждение леммы выполняется.

□

($\mathscr{L}$) *Существует функция* $\mathtt{b}(\mathtt{r}) > 0$, *такая что* $\mathtt{r}\mathtt{b}(\mathtt{r})$ *— неубывающая функция* $\mathtt{r}$ *и*

$$\frac{-2\mathbb{E}L(\boldsymbol{\theta},\boldsymbol{\theta}^*)}{\|D_0(\boldsymbol{\theta}-\boldsymbol{\theta}^*)\|_2^2} \ge \mathtt{b}(\mathtt{r}), \forall \mathtt{r} \ge \mathtt{r}_0, \boldsymbol{\theta}\in\Theta_0(\mathtt{r}).$$

Доказательство. Утверждение леммы эквивалентно утверждению

$$-2\mathbb{E}L(\boldsymbol{\theta},\boldsymbol{\theta}^*) \ge \mathtt{b}(\mathtt{r})\|D_0(\boldsymbol{\theta}-\boldsymbol{\theta}^*)\|_2^2.$$

Разложение в ряд Тейлора $\mathbb{E}L(\boldsymbol{\theta})$ имеет вид:

$$\mathbb{E}L(\boldsymbol{\theta}) = \mathbb{E}L(\boldsymbol{\theta}^*) + \nabla\mathbb{E}L(\boldsymbol{\theta}^*)(\boldsymbol{\theta}-\boldsymbol{\theta}^*) - \frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\theta}^*)^\mathsf{T} D_0^2(\widetilde{\boldsymbol{\theta}})(\boldsymbol{\theta}-\boldsymbol{\theta}^*)$$

для некоторых $\widetilde{\boldsymbol{\theta}} \in [\boldsymbol{\theta}^*, \boldsymbol{\theta}]$.

Так как $\boldsymbol{\theta}^*$ — локальный минимум функции $\mathbb{E}L(\boldsymbol{\theta})$, то выполнено, что $\nabla\mathbb{E}L(\boldsymbol{\theta}^*) = 0$. Для $\mathbb{E}L(\boldsymbol{\theta},\boldsymbol{\theta}^*) = \mathbb{E}L(\boldsymbol{\theta}) - \mathbb{E}L(\boldsymbol{\theta}^*)$ неравенство из утверждения имеет вид:

$$\|D_0(\widetilde{\boldsymbol{\theta}})(\boldsymbol{\theta}-\boldsymbol{\theta}^*)\|_2^2 \ge \mathtt{b}(\mathtt{r})\|D_0(\boldsymbol{\theta}-\boldsymbol{\theta}^*)\|_2^2.$$

Используя Лемму 8, получаем верхнюю границу для $\|D_0(\boldsymbol{\theta}-\boldsymbol{\theta}^*)\|_2^2$:

$$\|D_0(\boldsymbol{\theta}-\boldsymbol{\theta}^*)\|_2^2 \le nd\|\boldsymbol{\theta}-\boldsymbol{\theta}^*\|_2^2.$$

Давайте теперь получим нижнюю границу для $\|D_0(\widetilde{\boldsymbol{\theta}})(\boldsymbol{\theta}-\boldsymbol{\theta}^*)\|_2^2$:

$$\|D_0(\widetilde{\boldsymbol{\theta}})(\boldsymbol{\theta}-\boldsymbol{\theta}^*)\|_2^2 \ge \lambda_{\min}(D_0^2(\widetilde{\boldsymbol{\theta}}))\|\boldsymbol{\theta}-\boldsymbol{\theta}^*\|_2^2.$$

Для элементов матрицы $D_0^2(\widetilde{\boldsymbol{\theta}})$ выполнено:

$$(D_0^2(\widetilde{\boldsymbol{\theta}}))_{ij} = (D_0^2)_{ij} + \sum_{k=1}^{p} (\widetilde{\theta}_k - \theta_k^*) \frac{\partial^3 I\!\!E L(\breve{\boldsymbol{\theta}})}{\partial\theta_i \partial\theta_j \partial\theta_k}$$

для некоторого $\breve{\boldsymbol{\theta}} \in [\boldsymbol{\theta}^*, \boldsymbol{\theta}]$. Обозначим матрицу, которая состоит из элементов $\sum_{k=1}^{p} (\widetilde{\theta}_k - \theta_k^*) \frac{\partial^3 I\!\!E L(\breve{\boldsymbol{\theta}})}{\partial\theta_i \partial\theta_j \partial\theta_k}$ как A. Используя Лемму 9, получаем следующую оценку:

$$\|A\|_2 \le p\|A\|_{\max} \le p^{\frac{3}{2}} \sup_{i,j,k} \left| \frac{\partial^3 I\!\!E L(\breve{\boldsymbol{\theta}})}{\partial\theta_i \partial\theta_j \partial\theta_k} \right| \|\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|_2 \le p^{\frac{3}{2}} n C_3 \|\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|_2.$$

Нам нужно получить оценку минимального собственного значения матрицы $D_0(\widetilde{\boldsymbol{\theta}})$, и мы получаем:

$$\begin{aligned}
\lambda_{\min}(D_0(\widetilde{\boldsymbol{\theta}})) &= \inf_{\gamma \in I\!\!R^p, \|\gamma\|=1} \boldsymbol{\gamma}^\mathsf{T} D_0^2(\widetilde{\boldsymbol{\theta}}) \boldsymbol{\gamma} = \\
&= \inf_{\gamma \in I\!\!R^p, \|\gamma\|=1} \boldsymbol{\gamma}^\mathsf{T} (D_0^2 + A) \boldsymbol{\gamma} \ge n\mathsf{d}_0 - \sup_{\|\gamma\|=1} \|\boldsymbol{\gamma}^\mathsf{T} A \boldsymbol{\gamma}\| \ge \\
&\ge n\mathsf{d}_0 - \|A\|_2 \ge n\mathsf{d}_0 - p^{\frac{3}{2}} n C_3 \|\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|_2 \ge \\
&\ge n\mathsf{d}_0 - p^{\frac{3}{2}} n C_3 \sqrt{\frac{\mathsf{d}}{\mathsf{d}_0}} \delta\boldsymbol{\theta},
\end{aligned}$$

здесь $\delta\boldsymbol{\theta} = \sup_{\boldsymbol{\theta} \in \Theta} \|\boldsymbol{\theta} - \boldsymbol{\theta}^*\|_2$.

Таким образом, достаточно показать, что:

$$n\mathsf{d}_0 - p^{\frac{3}{2}} n C_3 \sqrt{\frac{\mathsf{d}}{\mathsf{d}_0}} \delta\boldsymbol{\theta} \ge \mathtt{b}(\mathtt{r}) n\mathsf{d} \text{ или}$$

$$\mathsf{d}_0 - \mathtt{b}(\mathtt{r})\mathsf{d} \ge p^{\frac{3}{2}} C_3 \sqrt{\frac{\mathsf{d}}{\mathsf{d}_0}} \delta\boldsymbol{\theta}. \tag{2.26}$$

Предположение (A5) дает (2.26) для достаточно малого $\frac{\lambda_i}{\lambda_0} < C, i = 1, 2, 3$, таким образом выполнено утверждение выше, и Лемма $(\mathscr{L})$ выполнена. Отметим, что мы можем выбрать $\mathtt{b}(\mathtt{r})$ как

$$\mathtt{b}(\mathtt{r}) = \mathtt{b} \stackrel{\text{def}}{=} \frac{\mathsf{d}_0}{2\mathsf{d}}.$$

$\square$

Локальные условия

Покажем, что для введенных в разделе 2.2.1 предположений выполнены следующие локальные условия.

$(\boldsymbol{ED_0})$ *Существует матрица* $V_0^2 = \operatorname{Var}\{\nabla\zeta(\boldsymbol{\theta}^*)\}$ *и константы* $\mathsf{g} > 0$, $\nu_0 \ge 1$, *такие что для всех* $|\lambda| \le \mathsf{g}$ *выполнено следующее неравенство*

$$\sup_{\boldsymbol{\gamma}\in I\!\!R^p} \log I\!\!E \exp\left\{\lambda \frac{\boldsymbol{\gamma}^{\mathsf{T}}\nabla\zeta(\boldsymbol{\theta}^*)}{\|V_0\boldsymbol{\gamma}\|_2}\right\} \le \nu_0^2\lambda^2/2. \tag{2.27}$$

Доказательство. Обозначим

$$Z = \frac{1}{\|V_0\boldsymbol{\gamma}\|_2}\sum_{i=1}^{p}\gamma_i U_i(\boldsymbol{\theta}^*).$$

Мы получим верхнюю оценку для спектральной нормы $\|Z\mathbf{C}\|_2$. Здесь и далее мы предполагаем без ограничения общности, что $\|\boldsymbol{\gamma}\| = 1$. Заметим, что $\|\mathbf{C}_{\boldsymbol{\theta}}^{-1}\|_2 < \frac{1}{\lambda_0}, \|\mathbf{C}\|_2 < \overline{\lambda}$ из-за предположения (A3), $\|\frac{\partial\mathbf{C}_{\boldsymbol{\theta}}}{\partial\theta_i}\|_2 < \lambda_1$ из-за предположения (A4). Следовательно, $\|U_i\mathbf{C}\|_2 < \frac{1}{\lambda_0^2}\overline{\lambda}\lambda_1$ и $\|\sum_{i=1}^{p}\gamma_i U_i\mathbf{C}\|_2 < p\frac{1}{\lambda_0^2}\overline{\lambda}\lambda_1$. Из предположения (A6) следует $\|V_0\boldsymbol{\gamma}\|_2 \ge \sqrt{\lambda_{\min}(V_0^2)} \ge \sqrt{n\mathsf{v}_0}$.

Матрица $Z\mathbf{C} = \frac{1}{\|V_0\boldsymbol{\gamma}\|}\sum_{i=1}^{p}\gamma_i U_i\mathbf{C}$. Следовательно, используя верхние оценки выше, получаем

$$\|Z\mathbf{C}\|_2 = \left\|\frac{1}{\|V_0\boldsymbol{\gamma}\|}\sum_{i=1}^{p}\gamma_i U_i\mathbf{C}\right\|_2 \le \frac{\sqrt{p}}{\sqrt{n\mathsf{v}_0}}\frac{1}{\lambda_0^2}\overline{\lambda}\lambda_1. \tag{2.28}$$

Мы выбираем g как

$$\mathsf{g} = \frac{1}{2}n^{\frac{1}{2}}\frac{\sqrt{\mathsf{v}_0}}{\sqrt{p}\frac{1}{\lambda_0^2}\overline{\lambda}\lambda_1}, \tag{2.29}$$

таким образом для любого $|\lambda| \le \mathsf{g}$ выполнено, что

$$\|\lambda Z\mathbf{C}\|_2 \le \frac{1}{2}.$$

Напомним, что $\zeta(\boldsymbol{\theta}) = L(\boldsymbol{\theta}) - I\!\!E L(\boldsymbol{\theta})$. Градиент $\zeta(\boldsymbol{\theta})$ имеет вид

$$\nabla_i\zeta(\boldsymbol{\theta}) = -\frac{1}{2}\left[-\mathbf{u}^T\mathbf{C}_{\boldsymbol{\theta}}^{-1}\frac{\partial\mathbf{C}_{\boldsymbol{\theta}}}{\partial\theta_i}\mathbf{C}_{\boldsymbol{\theta}}^{-1}\mathbf{u} + \operatorname{tr}\left(\mathbf{C}_{\boldsymbol{\theta}}^{-1}\frac{\partial\mathbf{C}_{\boldsymbol{\theta}}}{\partial\theta_i}\mathbf{C}_{\boldsymbol{\theta}}^{-1}\mathbf{C}\right)\right].$$

Тогда (2.27) эквивалентно

$$\sup_{\boldsymbol{\gamma}\in I\!\!R^p} \log I\!\!E \exp\left\{\lambda \frac{1}{2\|V_0\boldsymbol{\gamma}\|} \sum_{i=1}^{p} \gamma_i \left(\mathbf{u}^T U_i \mathbf{u} - \operatorname{tr}(U_i \mathbf{C})\right)\right\} \le \frac{\nu_0^2\lambda^2}{2}. \tag{2.30}$$

Левая часть (2.30) может быть переписана как

$$\sup_{\boldsymbol{\gamma}\in I\!\!R^p} \log \left[\exp\left(-\frac{\lambda}{2}\operatorname{tr}(Z\mathbf{C})\right) \int_{I\!\!R^n} \frac{1}{(2\pi)^{\frac{n}{2}}|\mathbf{C}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\mathbf{u}^T(\mathbf{C}^{-1} - \lambda Z)\mathbf{u}\right) d\mathbf{u}\right]. \tag{2.31}$$

Математическое ожидание в (2.31) существует, так как матрица $I - \lambda Z\mathbf{C}$ положительно определена для $|\lambda| \le \mathsf{g}$. Интегрируя математическое ожидание, получаем, что (2.21) эквивалентно

$$\sup_{\boldsymbol{\gamma}\in I\!\!R^p} -\frac{1}{2}\left[\operatorname{tr}\lambda Z\mathbf{C} + \log|I - \lambda Z\mathbf{C}|\right] \le \frac{\nu_0^2\lambda^2}{2}, |\lambda| < \mathsf{g}.$$

Для положительно определенной $I - \lambda Z\mathbf{C}$ выполнено

$$\log|I - \lambda Z\mathbf{C}| = -\operatorname{tr}\sum_{i=1}^{\infty}\frac{1}{i}(\lambda Z\mathbf{C})^i.$$

Тогда

$$-\frac{1}{2}\left[\operatorname{tr}\lambda Z\mathbf{C} + \log|I - \lambda Z\mathbf{C}|\right] = \frac{1}{2}\operatorname{tr}\left(\sum_{i=2}^{\infty}\frac{1}{i}(\lambda Z\mathbf{C})^i\right).$$

Таким образом, необходимо доказать, что

$$\sup_{\boldsymbol{\gamma}\in I\!\!R^p} \frac{1}{2}\operatorname{tr}\left(\sum_{i=2}^{\infty}\frac{1}{i}(\lambda Z\mathbf{C})^i\right) \le \frac{\nu_0^2\lambda^2}{2}, |\lambda| < \mathsf{g}.$$

Заметим, что

$$|\operatorname{tr}\left((\lambda Z\mathbf{C})^i\right)| = |\lambda^i \operatorname{tr}((Z\mathbf{C})^i)| = \left|\lambda^i \sum_{j=1}^{n} \lambda_j^i(Z\mathbf{C})\right| \le n|\lambda|^i \|Z\mathbf{C}\|_2^i.$$

Следовательно, необходимо доказать, что

$$n\sum_{i=2}^{\infty}|\lambda|^i\|Z\mathbf{C}\|_2^i \le \nu_0^2\lambda^2, |\lambda| < \mathsf{g}.$$

Для $i \ge 3$ нужно получить неравенство вида

$$n\,|\lambda|^{i-2}\,\|Z\mathbf{C}\|_2^i \le \frac{1}{2^{i-1}}\nu_0^2.$$

Так как выполнена верхняя оценка (2.28) для $\|Z\mathbf{C}\|_2$, то имеет место

$$n\,|\lambda|^{i-2}\,\|Z\mathbf{C}\|_2^i \le n|\lambda|^{i-2}\left(\frac{\sqrt{p}}{\sqrt{n\mathtt{v}_0}}\frac{1}{\lambda_0^2}\overline{\lambda}\lambda_1\right)^i.$$

Используя $\mathtt{g}$ из (2.29) и ν_0^2 вида

$$\nu_0^2 = \max\left(1, 2\left(\frac{\sqrt{p}}{\sqrt{\mathtt{v}_0}}\frac{1}{\lambda_0^2}\overline{\lambda}\lambda_1\right)^2\right), \tag{2.32}$$

получаем, что для любых $|\lambda| \le \mathtt{g}$ выполнено

$$n\,|\lambda|^{i-2}\,\|Z\mathbf{C}\|_2^i \le \left(\frac{\sqrt{p}}{\sqrt{\mathtt{v}_0}}\frac{1}{\lambda_0^2}\overline{\lambda}\lambda_1\right)^2\frac{1}{2^{i-2}} \le \frac{\nu_0^2}{2^{i-1}}.$$

Следовательно, так как $n\|Z\mathbf{C}\|_2^2 \le \frac{\nu_0^2}{2}$, то выполнено, что

$$\sup_{\gamma\in I\!\!R^p} n\sum_{i=2}^{\infty}|\lambda|^i\|Z\mathbf{C}\|_2^i \le \lambda^2\nu_0^2\sum_{i=2}^{\infty}\frac{1}{2^{i-1}} \le \nu_0^2\lambda^2, |\lambda| < \mathtt{g}.$$

Таким образом, для выбранного в (2.32) $\nu_0^2 \ge 1$ и выбранного в (2.29) $\mathtt{g}$ утверждение Леммы выполнено.

□

($\mathscr{L}_\mathbf{0}$) *Для каждого $\mathtt{r} \le \mathtt{r}_0$, существует константа $\delta(\mathtt{r}) \le \frac{1}{2}$, такая, что для любого $\boldsymbol{\theta} \in \Theta_0(\mathtt{r})$ и для $D_0 = -\nabla^2 I\!\!E L(\boldsymbol{\theta})$ выполнено, что:*

$$\left\|D_0^{-1}D^2(\boldsymbol{\theta})D_0^{-1} - \mathbf{I}_p\right\|_\infty \le \delta(\mathtt{r}). \tag{2.33}$$

Доказательство. Для того, чтобы доказать (2.33), достаточно показать, что

$$\left\|D^2(\boldsymbol{\theta}) - D_0^2\right\|_2 \le \|D_0^2\|_2\frac{1}{\sqrt{p}}\delta(\mathtt{r}).$$

Переходя от норм к их явным выражениям, получаем:

$$\max_{i,j\in\{1,2,\dots,p\}} \left| D^2(\boldsymbol{\theta}) - D_0^2 \right|_{ij} \le \|D_0^2\|_2 \frac{1}{p^{\frac{3}{2}}} \delta(\mathtt{r}).$$

Для $\max_{i,j\in\{1,2,\dots,p\}} \left| D^2(\boldsymbol{\theta}) - D_0^2 \right|_{ij}$ имеет место следующее разложение по Тейлору:

$$\max_{i,j\in\{1,2,\dots,p\}} \left| D^2(\boldsymbol{\theta}) - D_0^2 \right|_{ij} = \max_{i,j\in\{1,2,\dots,p\}} \left| \sum_{k=1}^{p} \frac{\partial^3 \mathbb{E} L(\widetilde{\boldsymbol{\theta}})}{\partial\theta_i \partial\theta_j \partial\theta_k} (\theta_k - \theta_k^*) \right| \le$$
$$\le C_3 n \|\boldsymbol{\theta} - \boldsymbol{\theta}^*\|_2 \le C_3 n \frac{\mathtt{r}_0}{\sqrt{n \mathsf{d}_0}},$$

здесь C_3 имеет вид из (2.20), $\widetilde{\boldsymbol{\theta}} \in [\boldsymbol{\theta}, \boldsymbol{\theta}^*]$

Теперь нужно доказать, что:

$$C_3 \sqrt{n} \frac{\mathtt{r}_0}{\sqrt{\mathsf{d}_0}} \le n \frac{\mathsf{d}_0 \delta(\mathtt{r})}{p^{\frac{3}{2}}}.$$

Это эквивалентно:

$$C_3 \mathtt{r}_0 \le \sqrt{n} \frac{d_0^{\frac{3}{2}} \delta(\mathtt{r})}{p^{\frac{3}{2}}}.$$

Так как $\delta(\mathtt{r}) \le \frac{1}{2}$, нужно, чтобы размер выборки был по крайней мере

$$n \ge 4 \mathsf{d}_0^3 C_3^2 \mathtt{r}_0^2 p^3. \tag{2.34}$$

Следовательно, взяв значение $\delta(\mathtt{r})$ меньшее, чем $\frac{1}{2}$ и вида $\frac{\mathtt{C}}{\sqrt{n}}$, мы можем контролировать значение константы C_3, используя предположение (A5).

$\square$

$(\boldsymbol{\mathcal{I}})$ Существует константа $\mathfrak{a} > 0$, такая что $\mathfrak{a}^2 D_0^2 \ge V_0^2$.

Доказательство. В силу введенных предположений для D_0^2 и V_0^2 Лемма выполнена с $\mathfrak{a}^2 = \frac{\mathsf{v}}{\mathsf{d}_0}$. $\square$

Примеры ковариационной функций, которые удовлетворяют предположениям

Квадратичная экспоненциальная ковариационная функция Проверим теперь введенные предположения для квадратичной экспоненциальной ковариационной функции:

$$c_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{x}') = \theta_2^2 \exp\left(-\theta_1^2 \|\mathbf{x} - \mathbf{x}'\|_2^2\right) + \sigma^2 \delta(\mathbf{x}, \mathbf{x}').$$

Ниже мы предполагаем, что

- Дисперсия шума $0 < \sigma^2 < \infty$ известна и фиксирована. Более того, она достаточно велика.
- Выполняется параметрическое предположение, то есть $c(\mathbf{x}, \mathbf{x}') = c_{\boldsymbol{\theta}^*}(\mathbf{x}, \mathbf{x}')$ для некоторого $\boldsymbol{\theta} = \boldsymbol{\theta}^*$.
- Точки дизайна $D = \{\mathbf{x}_i\}_{i=1}^n, \mathbf{x}_i \in I\!R^d$ лежат на сетке $\{\delta m, m \in \mathbb{Z}^d\}$.
- Параметрическое множество имеет вид $\Theta \in [\underline{\theta}_1, \overline{\theta}_1] \times [\underline{\theta}_2, \overline{\theta}_2]$ с $0 < \underline{\theta}_1 < \overline{\theta}_1 < \infty$, $0 < \underline{\theta}_2 < \overline{\theta}_2 < \infty$.

Матрица $\frac{\partial \mathbf{C}_{\boldsymbol{\theta}}}{\partial \theta_1} = -2\theta_1 \mathbf{C}_0 \circ H$, матрица $\frac{\partial \mathbf{C}_{\boldsymbol{\theta}}}{\partial \theta_2} = \frac{2}{\theta_2} \mathbf{C}_0$, где $\mathbf{C}_0 = \{\theta_2^2 \exp\left(-\theta_1^2 \|\mathbf{x}_i - \mathbf{x}_j\|_2^2\right)\}_{i,j=1}^n$, $\circ$ — Адамарово произведение двух матриц, $H = \{h_{ij}\}_{i,j=1}^n = \{\|\mathbf{x}_i - \mathbf{x}_j\|_2^2\}_{i,j=1}^n$.

Так как выполнено параметрическое предположение, то $D_0^2 = V_0^2$ и D_0^2 имеет вид

$$\frac{1}{n} D_0^2 = \frac{4}{n} \left.\begin{pmatrix} \theta_1^2 \operatorname{tr}\left(K_{\boldsymbol{\theta}}^{-1}(K_{\boldsymbol{\theta}}^0 \circ H) K_{\boldsymbol{\theta}}^{-1}(K_{\boldsymbol{\theta}}^0 \circ H)\right) & -\frac{\theta_1}{\theta_2} \operatorname{tr}\left(K_{\boldsymbol{\theta}}^{-1}(K_{\boldsymbol{\theta}}^0 \circ H) \mathbf{C}_{\boldsymbol{\theta}}^{-1} \mathbf{C}_0\right) \\ -\frac{\theta_1}{\theta_2} \operatorname{tr}\left(K_{\boldsymbol{\theta}}^{-1}(K_{\boldsymbol{\theta}}^0 \circ H) \mathbf{C}_{\boldsymbol{\theta}}^{-1} \mathbf{C}_0\right) & \frac{1}{\theta_2^2} \operatorname{tr}\left(\mathbf{C}_{\boldsymbol{\theta}}^{-1} \mathbf{C}_{\boldsymbol{\theta}} \mathbf{C}_{\boldsymbol{\theta}}^{-1} \mathbf{C}_0\right) \end{pmatrix}\right|_{\boldsymbol{\theta} = \boldsymbol{\theta}^*}.$$

Лемма 10. *Предположение (A2) выполняется, более того, ковариационная функция бесконечно дифференцируема.*

Лемма 11. *Предположения (A3)—(A5) выполняются.*

Доказательство. Доказательство следует подходу, предложенному в [86]. Так как спектральная норма матрицы ограничена сверху нормой по строкам, то

$$\|\mathbf{C}_{\boldsymbol\theta}\|_2 \le \|\mathbf{C}_{\boldsymbol\theta}\|_\infty = \sup_{1\le i\le n} \theta_2^2 \sum_{j=1}^{n} \exp(-\theta_1^2 h_{ij}) + \sigma_{ij}^2 \le$$
$$\le \overline{\theta}_2^2 \sum_{s\in\delta\mathbb{Z}^d} \exp(-\theta_1^2\|s\|_2^2) + \sigma^2 \le \overline{\theta}_2^2 \sum_{s\in\delta\mathbb{Z}^d} \exp(-\underline{\theta}_1^2\|s\|_2^2) + \sigma^2 \le$$
$$\le \overline{\theta}_2^2 \left(1 + \left(\frac{2}{\delta}\right)^d \int_{I\!R^d} \exp(-\underline{\theta}_1^2\|\mathbf{x}\|_2^2) d\mathbf{x}\right) + \sigma^2 \le \overline{\theta}_2^2 \left(1 + \left(\frac{2\sqrt{2\pi}}{\delta\underline{\theta}_1^2}\right)^d\right) + \sigma^2.$$

Следовательно, для любого n выполнено, что $\|\mathbf{C}_{\boldsymbol\theta}\|_2 \le \overline{\theta}_2^2 \left(1 + \left(\frac{2\sqrt{2\pi}}{\delta\underline{\theta}_1}\right)^d\right) + \sigma^2$. Нижняя оценка для $\|\mathbf{C}_{\boldsymbol\theta}\|_2$ выполнена, так как матрица $\mathbf{C}_0$ неотрицательно определена. Таким образом, выполнено (A3).

Аналогичные оценки выполнены для $\frac{\partial \mathbf{C}_{\boldsymbol\theta}}{\partial\theta_i}$, $\frac{\partial^2 \mathbf{C}_{\boldsymbol\theta}}{\partial\theta_i\partial\theta_j}$, $\frac{\partial^3 \mathbf{C}_{\boldsymbol\theta}}{\partial\theta_i\partial\theta_j\partial\theta_k}$, $i,j,k = \overline{1,p}$, следовательно (A4) выполнено. Более того, оценки $\frac{\partial \mathbf{C}_{\boldsymbol\theta}}{\partial\theta_i}$, $\frac{\partial^2 \mathbf{C}_{\boldsymbol\theta}}{\partial\theta_i\partial\theta_j}$, $\frac{\partial^3 \mathbf{C}_{\boldsymbol\theta}}{\partial\theta_i\partial\theta_j\partial\theta_k}$, $i,j,k = \overline{1,p}$ не зависят от дисперсии шума σ^2. Например, аналогично мы получаем, что $\|\frac{\partial \mathbf{C}_{\boldsymbol\theta}}{\partial\theta_2}\|_2 \le \overline{\theta}_2^2\overline{\theta}_1 \left(\frac{2\sqrt{2\pi}}{\delta\underline{\theta}_1^3}\right)^d$ и $\|\frac{\partial \mathbf{C}_{\boldsymbol\theta}}{\partial\theta_1}\|_2 \le 2\overline{\theta}_2 \left(\frac{2\sqrt{2\pi}}{\delta\underline{\theta}_1}\right)^d$ таким образом, для достаточно большого σ^2 мы гарантируем выполнение (A5), так как $\lambda_0 \ge \sigma^2$.

$\square$

Обозначим $\{a_i\}_{i=1}^n$ собственные значения матрицы $2\theta_1 K_{\boldsymbol\theta}^{-1}(K_{\boldsymbol\theta}^0 \circ H)\big|_{\boldsymbol\theta=\boldsymbol\theta^*}$ и $\{b_i\}_{i=1}^n$ собственные значения матрицы $\frac{2}{\theta_2}\mathbf{C}_{\boldsymbol\theta}^{-1}\mathbf{C}_0$. Пусть так же $\{a_i\}_{i=1}^n$ и $\{b_i\}_{i=1}^n$ упорядочены по возрастанию. Предположим, что $\{a_i\}_{i=1}^n$ и $\{b_i\}_{i=1}^n$ достаточно разные, то есть существует константа $\mu > 0$, такая что

$$\frac{1}{n^2}\sum_{i,j=1}^{n}(a_i b_j - a_j b_i)^2 \ge \mu. \tag{2.35}$$

Проверим теперь предположение (A6).

Лемма 12. *Пусть выполнено неравенство* (2.35). *Тогда предположение (A6) выполнено.*

Доказательство. Обозначим как A матрицу $-2\theta_1 K_{\boldsymbol\theta}^{-1}(K_{\boldsymbol\theta}^0 \circ H)$ и как B матрицу

$\frac{2}{\theta_2}\mathbf{C}_{\boldsymbol{\theta}}^{-1}\mathbf{C}_0$. Тогда D_0^2 имеет вид:

$$D_0^2 = \begin{pmatrix} \operatorname{tr}\left(A^2\right) & \operatorname{tr}(AB) \\ \operatorname{tr}(AB) & \operatorname{tr}\left(B^2\right) \end{pmatrix}.$$

Наименьшее собственно значение матрицы D_0^2 имеет вид

$$\frac{1}{n}\lambda_{\min}(D_0^2) = \frac{1}{n}\left(\operatorname{tr} A^2 + \operatorname{tr} B^2 - \sqrt{\left(\operatorname{tr} A^2 - \operatorname{tr} B^2\right)^2 + 4\operatorname{tr}(AB)^2}\right).$$

Посмотрим на выражение

$$\left(\operatorname{tr}(A^2) + \operatorname{tr}(B^2)\right)^2 - \left(\operatorname{tr}(A^2) - \operatorname{tr}(B^2)\right)^2 - 4\operatorname{tr}(AB)^2 = 4\left(\operatorname{tr} A^2 \operatorname{tr} B^2 - \operatorname{tr}(AB)^2\right).$$

Заметим, что

$$4\left(\operatorname{tr} A^2 \operatorname{tr} B^2 - \operatorname{tr}(AB)^2\right) \geq 4n^2\mu, \tag{2.36}$$

так как выполнено (2.35).

Теперь ограничим

$$\begin{aligned} &\operatorname{tr}(A^2) + \operatorname{tr}(B^2) + \sqrt{\left(\operatorname{tr} A^2 - \operatorname{tr} B^2\right)^2 + 4\operatorname{tr}(AB)^2} \leq \\ &\leq \|A|_F^2 + \|B\|_F^2 + \sqrt{\|A\|_F^4 + 2\|A\|_F^2\|B\|_F^2 + \|B\|_F^4 + 4\|AB\|_F^2}. \end{aligned}$$

Выполнено, что

$$\|A\|_F^2 \leq n\|A\|_2^2 \leq n4\overline{\theta}_1^2\|\mathbf{C}_{\boldsymbol{\theta}}^{-1}\|_2^2\|\mathbf{C}_{\boldsymbol{\theta}}^0 \circ H\|_2^2 \leq n\frac{1}{\lambda_0^2}\left.\left\|\frac{\partial \mathbf{C}_{\boldsymbol{\theta}}}{\partial \theta_1}\right\|_2^2\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \leq n\frac{1}{\lambda_0^2}\lambda_1^2.$$

Аналогично получаем для $\|B\|_F^2$:

$$\|B\|_F^2 \leq n\|B\|_2^2 \leq n4\frac{1}{\overline{\theta}_1^2}\|\mathbf{C}_{\boldsymbol{\theta}}^{-1}\|_2^2\|\mathbf{C}_{\boldsymbol{\theta}}^0\|_2^2 \leq n\frac{1}{\lambda_0^2}\left.\left\|\frac{\partial \mathbf{C}_{\boldsymbol{\theta}}}{\partial \theta_2}\right\|_2^2\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \leq n\frac{1}{\lambda_0^2}\lambda_1^2.$$

Так же выполнено

$$\|AB\|_F^2 \leq n\|AB\|_2^2 \leq n\|A\|_2^2\|B\|_2^2 \leq n\left(\frac{1}{\lambda_0^2}\lambda_1^2\right)^2.$$

Следовательно,

$$\operatorname{tr}(A^2) + \operatorname{tr}(B^2) + \sqrt{\left(\operatorname{tr} A^2 - \operatorname{tr} B^2\right)^2 + 4\operatorname{tr}(AB)^2} \le 2(1+\sqrt{2})n\frac{1}{\lambda_0^2}\lambda_1^2. \qquad (2.37)$$

Тогда

$$\frac{1}{n}\lambda_{\min}(D_0^2) \ge \frac{4n^2\mu}{2(1+\sqrt{2})n^2\frac{1}{\lambda_0^2}\lambda_1^2} = \frac{2}{1+\sqrt{2}}\frac{\lambda_0^2\mu}{\lambda_1^2} > 0.$$

Здесь мы использовали неравенства (2.36), (2.37) и утверждение, состоящее в том, что если $x^2 - y^2 \ge z$ и $x + y \le t$, то выполнено $x - y \ge \frac{z}{t}$. $\square$

Лемма 13. *Предположение (A1) выполнено.*

Доказательство. Параметрическое предположение выполнено, и, следовательно, расстояние Кульбака-Лейблера $\mathrm{KL}(\mathcal{N}(\mathbf{0}, \mathbf{C}_{\boldsymbol{\theta}^*}) \| \mathcal{N}(\mathbf{0}, \mathbf{C}_{\boldsymbol{\theta}}))$ равно нулю тогда и только тогда, когда $\boldsymbol{\theta} = \boldsymbol{\theta}^*$. Следовательно, предположение (A1) выполнено. $\square$

Произвдение ковариационных функций Рассмотрим теперь пример ковариационной функции, которая удовлетворяет (A1)–(A8) без дополнительного предположения 2.35. Пусть теперь ковариационная функция имеет вид:

$$c_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{x}') = \theta_2 \exp\left(-\theta_1 \|\mathbf{x} - \mathbf{x}'\|_2^2\right) c(\mathbf{x}, \mathbf{x}') + \sigma^2 \delta(\mathbf{x}, \mathbf{x}'), \qquad (2.38)$$

здесь $\theta_{11} > \theta_1 > \theta_{10}, \theta_{21} > \theta_2 > \theta_{20}$. Функция $c(\mathbf{x}, \mathbf{x}') = c(\|\mathbf{x} - \mathbf{x}'\|) = c(d)$ кусочно полиномиальная ковариационная функция, такая что

$$c(d) = v(1-ld)_+^{j+3}((j^3+9j^2+23j+15)(ld)^3+(6j^2+36j+45)(ld)^2+(15j+45)ld+15),$$

здесь $j = \lfloor \frac{d}{2} \rfloor + q + 1$, v, l — параметры ковариационной функции.

Функция $c(d)$ является ковариационной функцией [80]. Исследуемая ковариационная функция — произведение квадратичной экспоненциальной ковариационной функции и кусочно полиномиальной ковариационной функции, сложенная с ковариационной функцией, соответствующей белому шуму. Следовательно, $c_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{x}')$ — корректная ковариационная функция.

Ниже так же предполагаем, что

- Дисперсия шума $0 < \sigma^2 < \infty$ известна и фиксирована. Более того, она достаточно велика.
- Выполняется параметрическое предположение, то есть $c(\mathbf{x}, \mathbf{x}') = c_{\boldsymbol{\theta}^*}(\mathbf{x}, \mathbf{x}')$ для некоторого $\boldsymbol{\theta} = \boldsymbol{\theta}^*$.
- Параметрическое множество имеет вид $\Theta \in [\underline{\theta}_1, \overline{\theta}_1] \times [\underline{\theta}_2, \overline{\theta}_2]$ с $0 < \underline{\theta}_1 < \overline{\theta}_1 < \infty$, $0 < \underline{\theta}_2 < \overline{\theta}_2 < \infty$.

Будем рассматривать такой дизайн $\{\mathbf{x}_i\}_{i=1}^n$, что расстояние $\|\mathbf{x}_i - \mathbf{x}_{i+1}\| = d$ постоянно, и расстояние $\|\mathbf{x}_i - \mathbf{x}_j\| > 1.5d$ для $|i-j| > 1$. Возьмем такой набор точек на горизонтальной прямой в $\mathbb{R}^d$. Тогда для такого дизайна так же будут выполнены свойства, полученные для рассмотренной выше ковариационной функции. Выберем такие v, l, что $c(0) = 1 > c(d) = c > 0$, $c(1.5d) = 0$. Очевидно, что так сделать можно.

Для такого плана экспериментов ковариационная матрица $\mathbf{C} = \{c_{ij}\}_{i,j=1}^n$ — трехдиагональная:

$$c_{ij} = \begin{cases} \theta_2(1+\sigma^2), i = j, \\ \theta_2(c\exp(-\theta_1 d^2) + \sigma^2), |i-j| = 1, \\ 0, |i-j| > 1. \end{cases} \tag{2.39}$$

Обозначим λ_i собственные числа матрицы $\mathbf{C}_{\boldsymbol{\theta}}$.

Матрица $\frac{\partial \mathbf{C}_{\boldsymbol{\theta}}}{\partial \theta_1} = -2\theta_1 \mathbf{C}_0 \circ H$, матрица $\frac{\partial \mathbf{C}_{\boldsymbol{\theta}}}{\partial \theta_2} = \frac{2}{\theta_2}\mathbf{C}_0$, где $\mathbf{C}_0 = \{\theta_2^2 \exp\left(-\theta_1^2 \|\mathbf{x}_i - \mathbf{x}_j\|_2^2\right)\}_{i,j=1}^n$, $\circ$ — Адамарово произведение двух матриц, $H = \{h_{ij}\}_{i,j=1}^n = \{\|\mathbf{x}_i - \mathbf{x}_j\|_2^2\}_{i,j=1}^n$.

Так как выполнено параметрическое предположение, то $D_0^2 = V_0^2$ и D_0^2 имеет вид

$$\frac{1}{n} D_0^2 = \frac{4}{n} \left. \begin{pmatrix} \theta_1^2 \operatorname{tr}\left(K_{\boldsymbol{\theta}}^{-1}(K_{\boldsymbol{\theta}}^0 \circ H) K_{\boldsymbol{\theta}}^{-1}(K_{\boldsymbol{\theta}}^0 \circ H)\right) & -\frac{\theta_1}{\theta_2} \operatorname{tr}\left(K_{\boldsymbol{\theta}}^{-1}(K_{\boldsymbol{\theta}}^0 \circ H)\mathbf{C}_{\boldsymbol{\theta}}^{-1}\mathbf{C}_0\right) \\ -\frac{\theta_1}{\theta_2} \operatorname{tr}\left(K_{\boldsymbol{\theta}}^{-1}(K_{\boldsymbol{\theta}}^0 \circ H)\mathbf{C}_{\boldsymbol{\theta}}^{-1}\mathbf{C}_0\right) & \frac{1}{\theta_2^2} \operatorname{tr}\left(\mathbf{C}_{\boldsymbol{\theta}}^{-1}\mathbf{C}_{\boldsymbol{\theta}}\mathbf{C}_{\boldsymbol{\theta}}^{-1}\mathbf{C}_0\right) \end{pmatrix} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}.$$

Доказательство предположений (A1-A6) не отличается от доказательств, приведенных выше для квадратичной экпоненциальной ковариационной функции.

Лемма 14. *Предположение (A8) выполнено.*

Доказательство. Параметрическое предположение выполнено. Аналогично предыдущей ковариационной функции достаточно показать, что выполнено (2.36). А именно:

$$\begin{aligned} Z = &\operatorname{tr}(\mathbf{C}_{\boldsymbol\theta}^{-1}\mathbf{C}_{\boldsymbol\theta}^{0}\mathbf{C}_{\boldsymbol\theta}^{-1}\mathbf{C}_{\boldsymbol\theta}^{0})\operatorname{tr}(\mathbf{C}_{\boldsymbol\theta}^{-1}(\mathbf{C}_{\boldsymbol\theta}\circ H)\mathbf{C}_{\boldsymbol\theta}^{-1}(\mathbf{C}_{\boldsymbol\theta}\circ H))- \\ &-\operatorname{tr}(\mathbf{C}_{\boldsymbol\theta}^{-1}(\mathbf{C}_{\boldsymbol\theta}\circ H)\mathbf{C}_{\boldsymbol\theta}^{-1}\mathbf{C}_{\boldsymbol\theta}^{0})\operatorname{tr}(\mathbf{C}_{\boldsymbol\theta}^{-1}(\mathbf{C}_{\boldsymbol\theta}\circ H)\mathbf{C}_{\boldsymbol\theta}^{-1}\mathbf{C}_{\boldsymbol\theta}^{0}) > n^2\mu \end{aligned} \tag{2.40}$$

для некоторого $\mu > 0$.

Получим $\operatorname{tr}(\mathbf{C}_{\boldsymbol\theta}^{-1}\mathbf{C}_{\boldsymbol\theta}^{0}\mathbf{C}_{\boldsymbol\theta}^{-1}\mathbf{C}_{\boldsymbol\theta}^{0})$. Для этого воспользуемся тем, что $\mathbf{C}_{\boldsymbol\theta}^{0} = \mathbf{C}_{\boldsymbol\theta} - \theta_1\sigma^2 I$.

$$\begin{aligned} &\operatorname{tr}(\mathbf{C}_{\boldsymbol\theta}^{-1}\mathbf{C}_{\boldsymbol\theta}^{0}\mathbf{C}_{\boldsymbol\theta}^{-1}\mathbf{C}_{\boldsymbol\theta}^{0}) = \operatorname{tr}(\mathbf{C}_{\boldsymbol\theta}^{-1}\mathbf{C}_{\boldsymbol\theta}^{0}\mathbf{C}_{\boldsymbol\theta}^{-1}(\mathbf{C}_{\boldsymbol\theta} - \theta_1\sigma^2 I)) = \\ &= \operatorname{tr}(\mathbf{C}_{\boldsymbol\theta}^{-1}\mathbf{C}_{\boldsymbol\theta}^{0}\mathbf{C}_{\boldsymbol\theta}^{-1}\mathbf{C}_{\boldsymbol\theta}) - \operatorname{tr}(\mathbf{C}_{\boldsymbol\theta}^{-1}\mathbf{C}_{\boldsymbol\theta}^{0}\mathbf{C}_{\boldsymbol\theta}^{-1}\theta_1\sigma^2 I) = \operatorname{tr}(\mathbf{C}_{\boldsymbol\theta}^{-1}\mathbf{C}_{\boldsymbol\theta}^{0}) - \theta_1\sigma^2\operatorname{tr}(\mathbf{C}_{\boldsymbol\theta}^{-1}\mathbf{C}_{\boldsymbol\theta}^{0}\mathbf{C}_{\boldsymbol\theta}^{-1}). \end{aligned}$$

Рассмотрим сначала $\operatorname{tr}(\mathbf{C}_{\boldsymbol\theta}^{-1}\mathbf{C}_{\boldsymbol\theta}^{0})$:

$$\begin{aligned} \operatorname{tr}(\mathbf{C}_{\boldsymbol\theta}^{-1}\mathbf{C}_{\boldsymbol\theta}^{0}) &= \operatorname{tr}(\mathbf{C}_{\boldsymbol\theta}^{-1}(\mathbf{C}_{\boldsymbol\theta} - \theta_1\sigma^2 I)) = \\ &= \operatorname{tr}(\mathbf{C}_{\boldsymbol\theta}^{-1}\mathbf{C}_{\boldsymbol\theta}) - \operatorname{tr}(\mathbf{C}_{\boldsymbol\theta}^{-1}\theta_1\sigma^2 I) = \operatorname{tr}(I) - \theta_1\sigma^2\operatorname{tr}(\mathbf{C}_{\boldsymbol\theta}^{-1}) = \\ &= n - \theta_1\sigma^2\sum_{i=1}^{n}\frac{1}{\lambda_i}, \end{aligned}$$

λ_i — собственные числа матрицы $\mathbf{C}_{\boldsymbol\theta}$.

Теперь рассмотрим $\operatorname{tr}(\mathbf{C}_{\boldsymbol\theta}^{-1}\mathbf{C}_{\boldsymbol\theta}^{0}\mathbf{C}_{\boldsymbol\theta}^{-1})$:

$$\begin{aligned} \operatorname{tr}(\mathbf{C}_{\boldsymbol\theta}^{-1}\mathbf{C}_{\boldsymbol\theta}^{0}\mathbf{C}_{\boldsymbol\theta}^{-1}) &= \operatorname{tr}(\mathbf{C}_{\boldsymbol\theta}^{-1}\mathbf{C}_{\boldsymbol\theta}^{-1}\mathbf{C}_{\boldsymbol\theta}^{0}) = \\ &= \operatorname{tr}(\mathbf{C}_{\boldsymbol\theta}^{-1}\mathbf{C}_{\boldsymbol\theta}^{-1}(\mathbf{C}_{\boldsymbol\theta} - \theta_1\sigma^2 I)) \\ &= \operatorname{tr}(\mathbf{C}_{\boldsymbol\theta}^{-1}\mathbf{C}_{\boldsymbol\theta}^{-1}\mathbf{C}_{\boldsymbol\theta}) - \operatorname{tr}(\mathbf{C}_{\boldsymbol\theta}^{-1}\mathbf{C}_{\boldsymbol\theta}^{-1}\theta_1\sigma^2 I)) = \\ &= \operatorname{tr}(\mathbf{C}_{\boldsymbol\theta}^{-1}) - \theta_1\sigma^2\operatorname{tr}(\mathbf{C}_{\boldsymbol\theta}^{-1}\mathbf{C}_{\boldsymbol\theta}^{-1}) = \sum_{i=1}^{n}\frac{1}{\lambda_i} - \theta_1\sigma^2\sum_{i=1}^{n}\frac{1}{\lambda_i^2}. \end{aligned}$$

Следовательно, подставляя полученные выражения для $\operatorname{tr}(\mathbf{C}_{\boldsymbol{\theta}}^{-1}\mathbf{C}_{\boldsymbol{\theta}}^{0})$ и $\operatorname{tr}(\mathbf{C}_{\boldsymbol{\theta}}^{-1}\mathbf{C}_{\boldsymbol{\theta}}^{0}\mathbf{C}_{\boldsymbol{\theta}}^{-1})$ в выражением выше, получаем:

$$\operatorname{tr}(\mathbf{C}_{\boldsymbol{\theta}}^{-1}\mathbf{C}_{\boldsymbol{\theta}}^{0}\mathbf{C}_{\boldsymbol{\theta}}^{-1}\mathbf{C}_{\boldsymbol{\theta}}^{0}) = \sum_{i=1}^{n}\left(1 - 2\theta_1\sigma^2\frac{1}{\lambda_i} + \theta_1^2\sigma^4\frac{1}{\lambda_i^2}\right).$$

Получим $\operatorname{tr}(\mathbf{C}_{\boldsymbol{\theta}}^{-1}(\mathbf{C}_{\boldsymbol{\theta}} \circ H)\mathbf{C}_{\boldsymbol{\theta}}^{-1}(\mathbf{C}_{\boldsymbol{\theta}} \circ H))$. Так как матрица $\mathbf{C}_{\boldsymbol{\theta}}$ трехдиагональная и имеет вид (2.39), то

$$\mathbf{C}_{\boldsymbol{\theta}} \circ H = d^2(\mathbf{C}_{\boldsymbol{\theta}} - \theta_1(1+\sigma^2)I).$$

Следовательно,

$$\begin{aligned}
&\operatorname{tr}(\mathbf{C}_{\boldsymbol{\theta}}^{-1}(\mathbf{C}_{\boldsymbol{\theta}} \circ H)\mathbf{C}_{\boldsymbol{\theta}}^{-1}(\mathbf{C}_{\boldsymbol{\theta}} \circ H)) = \\
=&\operatorname{tr}(\mathbf{C}_{\boldsymbol{\theta}}^{-1}(\mathbf{C}_{\boldsymbol{\theta}} \circ H)\mathbf{C}_{\boldsymbol{\theta}}^{-1}d^2(\mathbf{C}_{\boldsymbol{\theta}} - \theta_1(1+\sigma^2)I)) = \\
=&d^2\left[\operatorname{tr}(\mathbf{C}_{\boldsymbol{\theta}}^{-1}(\mathbf{C}_{\boldsymbol{\theta}} \circ H)\mathbf{C}_{\boldsymbol{\theta}}^{-1}\mathbf{C}_{\boldsymbol{\theta}}) - \operatorname{tr}(\mathbf{C}_{\boldsymbol{\theta}}^{-1}\mathbf{C}_{\boldsymbol{\theta}} \circ H)\mathbf{C}_{\boldsymbol{\theta}}^{-1}\theta_1(1+\sigma^2)I)\right] = \\
=&d^2\left[\operatorname{tr}(\mathbf{C}_{\boldsymbol{\theta}}^{-1}(\mathbf{C}_{\boldsymbol{\theta}} \circ H)) - \theta_1(1+\sigma^2)\operatorname{tr}(\mathbf{C}_{\boldsymbol{\theta}}^{-1}(\mathbf{C}_{\boldsymbol{\theta}} \circ H)\mathbf{C}_{\boldsymbol{\theta}}^{-1})\right] =
\end{aligned}$$

Рассмотрим слагаемое $\operatorname{tr}(\mathbf{C}_{\boldsymbol{\theta}}^{-1}(\mathbf{C}_{\boldsymbol{\theta}} \circ H))$:

$$\begin{aligned}
&\operatorname{tr}(\mathbf{C}_{\boldsymbol{\theta}}^{-1}(\mathbf{C}_{\boldsymbol{\theta}} \circ H)) = \\
=&\operatorname{tr}(\mathbf{C}_{\boldsymbol{\theta}}^{-1}d^2(\mathbf{C}_{\boldsymbol{\theta}} - \theta_1(1+\sigma^2)I)) = \\
=&d^2\operatorname{tr}(\mathbf{C}_{\boldsymbol{\theta}}^{-1}(\mathbf{C}_{\boldsymbol{\theta}} - \theta_1(1+\sigma^2)I)) = \\
=&d^2\left[\operatorname{tr}(\mathbf{C}_{\boldsymbol{\theta}}^{-1}\mathbf{C}_{\boldsymbol{\theta}}) - \operatorname{tr}(\mathbf{C}_{\boldsymbol{\theta}}^{-1}\theta_1(1+\sigma^2)I)\right] = \\
=&d^2\left[\operatorname{tr}(I) - \theta_1(1+\sigma^2)\operatorname{tr}(\mathbf{C}_{\boldsymbol{\theta}}^{-1})\right] = \\
=&d^2\left[n - \theta_1(1+\sigma^2)\sum_{i=1}^{n}\frac{1}{\lambda_i}\right].
\end{aligned}$$

Рассмотрим слагаемое $\operatorname{tr}(\mathbf{C}_{\boldsymbol{\theta}}^{-1}(\mathbf{C}_{\boldsymbol{\theta}} \circ H)\mathbf{C}_{\boldsymbol{\theta}}^{-1})$:

$$\begin{aligned}
&\operatorname{tr}(\mathbf{C}_{\boldsymbol{\theta}}^{-1}(\mathbf{C}_{\boldsymbol{\theta}} \circ H)\mathbf{C}_{\boldsymbol{\theta}}^{-1}) = \\
=&\operatorname{tr}(\mathbf{C}_{\boldsymbol{\theta}}^{-1}\mathbf{C}_{\boldsymbol{\theta}}^{-1}d^2(\mathbf{C}_{\boldsymbol{\theta}} - \theta_1(1+\sigma^2)I)) =
\end{aligned}$$

$$
\begin{aligned}
=&d^2 \operatorname{tr}(\mathbf{C}_{\boldsymbol{\theta}}^{-1}\mathbf{C}_{\boldsymbol{\theta}}^{-1}(\mathbf{C}_{\boldsymbol{\theta}} - \theta_1(1+\sigma^2)I)) = \\
=&d^2 \left[\operatorname{tr}(\mathbf{C}_{\boldsymbol{\theta}}^{-1}\mathbf{C}_{\boldsymbol{\theta}}^{-1}\mathbf{C}_{\boldsymbol{\theta}}) - \operatorname{tr}(\mathbf{C}_{\boldsymbol{\theta}}^{-1}\theta_1(1+\sigma^2)I)\right] = \\
=&d^2 \left[\operatorname{tr}(\mathbf{C}_{\boldsymbol{\theta}}^{-1}) - \theta_1(1+\sigma^2)\operatorname{tr}(\mathbf{C}_{\boldsymbol{\theta}}^{-1}\mathbf{C}_{\boldsymbol{\theta}}^{-1})\right] = \\
=&d^2 \left[\sum_{i=1}^{n}\frac{1}{\lambda_i} - \theta_1(1+\sigma^2)\sum_{i=1}^{n}\frac{1}{\lambda_i^2}\right].
\end{aligned}
$$

Тогда

$$
\begin{aligned}
&\operatorname{tr}(\mathbf{C}_{\boldsymbol{\theta}}^{-1}(\mathbf{C}_{\boldsymbol{\theta}}\circ H)\mathbf{C}_{\boldsymbol{\theta}}^{-1}(\mathbf{C}_{\boldsymbol{\theta}}\circ H)) = \\
&= d^4\left[n - 2\theta_1(1+\sigma^2)\sum_{i=1}^{n}\frac{1}{\lambda_i} + \theta_1^2(1+\sigma^2)^2\sum_{i=1}^{n}\frac{1}{\lambda_i}\right].
\end{aligned}
$$

Получим таким же способом $\operatorname{tr}(\mathbf{C}_{\boldsymbol{\theta}}^{-1}(\mathbf{C}_{\boldsymbol{\theta}}\circ H)\mathbf{C}_{\boldsymbol{\theta}}^{-1}\mathbf{C}_{\boldsymbol{\theta}}^{0})$:

$$
\operatorname{tr}(\mathbf{C}_{\boldsymbol{\theta}}^{-1}(\mathbf{C}_{\boldsymbol{\theta}}\circ H)\mathbf{C}_{\boldsymbol{\theta}}^{-1}\mathbf{C}_{\boldsymbol{\theta}}^{0}) = d^2\sum_{i=1}^{n}\left(1 - \theta_1(1+2\sigma^2)\frac{1}{\lambda_i} + \theta_1^2\sigma^2(1+\sigma^2)\frac{1}{\lambda_i^2}\right)
$$

Введем обозначения $a_i = \theta_1\sigma^2\frac{1}{\lambda_i}$, $b_i = \theta_1(1+\sigma^2)\frac{1}{\lambda_i}$, $c_i = b_i - a_i = \theta_1\frac{1}{\lambda_i}$. Тогда

$$
\operatorname{tr}(\mathbf{C}_{\boldsymbol{\theta}}^{-1}\mathbf{C}_{\boldsymbol{\theta}}^{0}\mathbf{C}_{\boldsymbol{\theta}}^{-1}\mathbf{C}_{\boldsymbol{\theta}}^{0}) = \sum_{i=1}^{n}(1 - 2a_i + a_i^2) = \sum_{i=1}^{n}(1-a_i)^2.
$$

$$
\operatorname{tr}(\mathbf{C}_{\boldsymbol{\theta}}^{-1}(\mathbf{C}_{\boldsymbol{\theta}}\circ H)\mathbf{C}_{\boldsymbol{\theta}}^{-1}(\mathbf{C}_{\boldsymbol{\theta}}\circ H)) = \sum_{i=1}^{n}(1 - 2(a_i+c_i) + (a_i+c_i)^2) = \sum_{i=1}^{n}(1-a_i-c_i)^2.
$$

$$
\operatorname{tr}(\mathbf{C}_{\boldsymbol{\theta}}^{-1}(\mathbf{C}_{\boldsymbol{\theta}}\circ H)\mathbf{C}_{\boldsymbol{\theta}}^{-1}\mathbf{C}_{\boldsymbol{\theta}}^{0}) = \sum_{i=1}^{n}(1 - (2a_i+c_i) + a_i(a_i+c_i)) = \sum_{i=1}^{n}(1-a_i)(1-a_i-c_i).
$$

Обозначим $\alpha_i = 1 - a_i$, $\beta_i = 1 - a_i - c_i$. Тогда Z из (2.40) имеет вид:

$$
\begin{aligned}
\frac{1}{d^4}Z &= \left(\sum_{i=1}^{n}\alpha_i^2\right)\left(\sum_{i=1}^{n}\beta_i^2\right) - \left(\sum_{i=1}^{n}\alpha_i\beta_i\right)^2 = \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n}(\alpha_i\beta_j - \alpha_j\beta_i)^2 =
\end{aligned}
$$

$$
\begin{aligned}
&=\sum_{i=1}^{n}\sum_{j=1}^{n}((1-a_i)(1-a_j-c_j)-(1-a_j)(1-a_i-c_i))^2=\\
&=\sum_{i=1}^{n}\sum_{j=1}^{n}((1-a_i)(-c_j)-(1-a_j)(-c_i))^2=\\
&=\sum_{i=1}^{n}\sum_{j=1}^{n}(c_i-c_j+c_ja_i-c_ia_j)^2=\\
&=\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\frac{1}{\lambda_i}\theta_1-\frac{1}{\lambda_j}\theta_1+\frac{1}{\lambda_i\lambda_j}\theta_1^2\sigma^2-\frac{1}{\lambda_i\lambda_j}\theta_1^2\sigma^2\right)^2=\\
&=\theta_1\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\frac{1}{\lambda_i}-\frac{1}{\lambda_j}\right)^2.
\end{aligned}
$$

Так как выполнено условие (A3), то

$$
\left(\frac{1}{\lambda_i}-\frac{1}{\lambda_j}\right)^2=\frac{(\lambda_i-\lambda_j)^2}{\lambda_i^2\lambda_j^2}\geq\frac{1}{\lambda_{max}^4}(\lambda_i-\lambda_j)^2.
$$

Следовательно,

$$
\begin{aligned}
\frac{1}{d^4}Z&\geq\frac{\theta_1}{\lambda_{max}^4}\sum_{j=1}^{n}(\lambda_i-\lambda_j)^2=\frac{\theta_1}{\lambda_{max}^4}(n\,\mathrm{tr}(\mathbf{C}^2)-\mathrm{tr}(\mathbf{C})^2)=\\
&=\frac{\theta_1}{\lambda_{max}^4}\left(n\sum_{i,j=1}^{n}c_{ij}^2-\left(\sum_{i=1}^{n}c_{ii}\right)^2\right)=\\
&=\frac{\theta_1}{\lambda_{max}^4}(n^2c(0)^2+2n(n-1)c(d)^2-n^2c(0)^2)=2n(n-1)\frac{\theta_1c(d)^2}{\lambda_{max}^4}.
\end{aligned}
$$

Следовательно, $Z\geq n^2\mu$ с $\mu=\frac{\theta_1c(0)^2d^4}{2\lambda_{max}^4}$, что и требовалось доказать. □

Лемма 15. *Предположение (A1) выполнено.*

Доказательство. Параметрическое предположение выполнено, и, следовательно, расстояние Кульбака-Лейблера $\mathrm{KL}(\mathcal{N}(\mathbf{0},\mathbf{C}_{\boldsymbol{\theta}^*})\|\mathcal{N}(\mathbf{0},\mathbf{C}_{\boldsymbol{\theta}}))$ равно нулю тогда и только тогда, когда $\boldsymbol{\theta}=\boldsymbol{\theta}^*$. Следовательно, предположение (A1) выполнено. □

2.3 Выводы

Рассматривалась модель регрессии на основе гауссовских процессов. В первой части главы введенны основные определения для регрессии на основе гауссовских процессов, описаны процедуры оценки параметров для такой регрессионной модели. В том числе описана байесовская процедура с использованием гипераприорного распределения, которая оказывается менее подвержена вычислительным проблемам, характерным для данного метода, и, кроме того, позволяет получать более точные регрессионные модели.

Во второй части главы рассмотрено использование байесовского подхода для оценки параметров ковариационной функции гауссовского процесса для неинформативного априорного распределения. Для случая возможной неправильной спецификации модели в неассимптотическом случае доказана теорема тиаа Бернштейна-фон-Мизеса: апостериорное распределение параметров ковариационной функции близко к нормальному со средним близким к оценке максимума правдоподобия (ОМП) и ковариационной матрицей близкой к ковариационной матрице ОМП. В работе показано, что предположения, сделанные в теореме, выполняются для широко используемого класса ковариационных функций. Полученные теоретические результаты проиллюстрированы вычислительными экспериментами.

Глава 3

Регрессия на основе гауссовских процессов для разнородных данных

Часто инженер может варьировать качество полученного решения, и данные для построения модели состоят не только из точных значений целевой функции, но и из приближенных значений целевой функции, полученных с помощью более грубого, но менее затратного источника данных [50, 33].

Для моделирования таких разнородных данных удобной оказывается модель регрессии на основе гауссовских процессов[51], с помощью которой можно эффективно восстанавливать нелинейные зависимости и оценивать точность прогноза зависимости в заданной точке. Использование такой модели описано, например, М. Кеннеди[51], и в работах по геостатистике [28]. После этого похожая модель широко использовалась для решения различных прикладных задач в индустриальной инженерии [34, 77, 40]. Такая модель близка к модели, которая используется для суррогатного моделирования с помощью регрессии на основе гауссовских процессов для многомерного выхода [5, 20]. Размер выборки, которая может использоваться для построения регрессионной модели с использованием регрессии на основе гауссовских процессов, ограничен несколькими тысячами точек, так как в процессе оценки параметров регрессии необходимо обращать матрицу ковариаций точек выборки[35].

Поэтому если выборка однородных данных большого размера, то для построения модели регрессии на основе гауссовских процессов используют приближенные подходы. Приведем их краткий обзор. Аппроксимация Нистерма [31] исходной ковариационной матрицы выборки на основе подмножества базовых точек существенно сокращает сложность вычислений. Использование аппроксимации Нистрема для регрессии на основе гауссовских процессов дано в работе Л. Фостера[35], так же похожие подходы описаны в книге К. Расмуссена[80], вариации подхода и обзор литературы приведены в работах [78, 35, 95]. Другим возможным подходом для работы с выборками большого размера в регрессии на основе гауссовских процессов является ограничение ковариации: за счет того, что значения ковариации меньше заданного порога считаются нулевыми, удается получать разреженные ковариационные матрицы, и использовать математический аппарат для работы с такими матрицами — Р. Фуррер[36] предложил такую процедуру в 2006 году, и затем в серии работ К. Кауфманн, Б. Шаби, Т. Чу[26, 86, 49] провели ее теоретический анализ. Еще один тип приближений основан на использования разных вариантов приближенного байесовского вывода, так в работах Дж. Хенсмана [42] и Титсиаса [99] используется вариационный вывод. Иной способ увеличить допустимый размер выборки — разбиение пространства дизайна на отдельные области, в каждой из которых для оценки значений модели используется только подвыборка исходной выборки. Существует ряд работ [72, 87, 9], в которых продемонстрирована эффективность такого подхода. Использование специальной структуры данных так же позволяет работать с выборками большого размера для построения регрессионной модели на основе гауссовских процессов. Например, Д. Зиммерман[114] предложил эффективный алгоритм для данных, которые представляются в виде тензорного произведения одномерных факторов, а М. Беляев[12] предложил эффективный алгоритм для тензорного произведения многомерных факторов, в которых часть значений пропущена; но в большинстве случаев такую специальную структуру данных выделить не удается.

Однако, для построения регрессии на основе гауссовских процессов по данным разной точности без специальной структуры до настоящего времени не было предложено способа работы с выборками размера больше нескольких тысяч точек. В то же время, в случае использования данных разной точности выборки большого размера встречаются чаще, поскольку "стоимость" получения одного значения из источника данных низкой точности обычно значительно ниже "стоимости" вычисления одного значения из источника данных высокой точности, и выборка данных низкой точности обычно большого размера.

В диссертации предложено два подхода к моделированию разнородных данных в случае наличия возможности получить выборку большого размера из сточника данных низкой точности. Первым подход использует аппроксимацию Нистрема для приближения ковариационных матриц больших размеров, второй подход предполагает наличия черного ящика для источника данных низкой точности.

3.1 Построение регрессионных моделей разнородных данных

В данном описаны основные подходы к построению регрессионных моделей разнородных данных, предложена классификация работ по использованным подходам и решенным прикладным задачам.

3.1.1 Постановки задач

Пусть для описания характеристики некоторой системы существуют две модели. Модель, порожденная *источником данных высокой точности*, $u(\mathbf{x})$ позволяет получать точные оценки характеристики, однако, стоимость использования такой модели высока. Качество модели, порожденной *источником данных низкой точности*, $f(\mathbf{x}_f)$ ниже, но стоимость ее вычисления невелика.

В работах встречаются различные вариации такого описания моделей. Вместо двух моделей может быть задана иерархия моделей. То есть, задан ряд моделей, отсортированных по точности $u(\mathbf{x}_u), f^1(\mathbf{x}_f), f^2(\mathbf{x}_f), \ldots, f^L(\mathbf{x}_f)$ [38]. Так же может использоваться информация о градиентах модели[40].

Точки, в которых задано значение функции высокой точности принадлежат пространству дизайна $\mathbb{X}^u \subseteq \mathbb{R}^{d_u}$, точки, в которых задано значение функции низкой точности принадлежат пространству дизайна $\mathbb{X}^f \subseteq \mathbb{R}^{d_f}$. Пространства дизайна для функций низкой и высокой точности могут отличаться. То есть, для точек из пространства дизайна $\mathbf{x}_u \in \mathbb{X}^u \subseteq \mathbb{R}^{d_u}$, $\mathbf{x}_f \in \mathbb{X}^f \subseteq \mathbb{R}^{d_f}$ размерности d_f и d_u могут не совпадать [82].

Рассмотрим теперь возможные способы решения поставленных задач.

3.1.2 Эвристические модели

Многие подходы к построению суррогатных моделей переменной точности используют эвристики, которые явно задают вид связи между источниками данных разной точности [52, 2, 64, 94, 109, 40]. Как правило, используемая модель — частный случай модели, представленной ниже:

$$\widetilde{u}(\mathbf{x}) = \rho(\mathbf{x})\widetilde{f}(\mathbf{x}) + g(\mathbf{x}),$$

где $\widetilde{u}(\mathbf{x})$ — суррогатная модель $u(\mathbf{x})$, $\widetilde{f}(\mathbf{x})$ — суррогатная модель $f(\mathbf{x})$, $\rho(\mathbf{x})$, $g(\mathbf{x})$ — функции из некоторых заданных семейств.

В разных работах используются различные семейства в качестве $\rho(\mathbf{x})$ и $\delta(\mathbf{x})$. Кроме того, используются различные суррогатные модели для $\widetilde{f}(\mathbf{x})$. Во многих случаях оправдано вычисление значения источника данных низкой точности $f(\mathbf{x})$ в каждой точке оценки $\widetilde{u}(\mathbf{x})$, и, таким образом, суррогатная модель $f(\mathbf{x})$ не строится. Используются квадратичные и линейные регрессионные модели, локальные линейные модели, модели регрессии на основе гауссовских процессов (GP) и модели на основе радиальных базисных функций (RBF). Конкретные примеры различных семейств функций $\rho(\mathbf{x}), \widetilde{f}(\mathbf{x}), \delta(\mathbf{x})$ приведены в таблице 3.1.

Статья	Год	Модель $\widetilde{f}(\mathbf{x})$	Модель $\rho(\mathbf{x})$	Модель $g(\mathbf{x})$
[2]	2000	$f(\mathbf{x})$	Локальная линейная	0
[64]	2001	$f(\mathbf{x})$	Линейная	0
[94]	2010	$f(\mathbf{x})$ или RBF	Квадратичная	0
			0	Квадратичная
[109]	2012	$f(\mathbf{x})$	0	GP
[40]	2012	GP	0	GP

Таблица 3.1 – Используемые семейства функций $\rho(\mathbf{x}), \widetilde{f}(\mathbf{x}), g(\mathbf{x})$. Пояснения к таблице приведены в тексте.

3.1.3 Кокригинг

Предложенный довольно давно в геостатистике [27], кокригинг в его текущем виде приведен в работах [51, 50]. Предполагается, что точная функция $u(\mathbf{x})$ — реализация гауссовского процесса, который является суммой двух независимых гауссовских процессов, один из которых есть $f(\mathbf{x})$, умноженное на заданный коэффициент ρ, а $g(\mathbf{x})$ моделирует разность между источниками данных разной точности:

$$u(\mathbf{x}) = \rho f(\mathbf{x}) + g(\mathbf{x}).$$

Кокригинг обладает рядом преимуществ по сравнению с эвристическими моделями, которые были описаны выше:

- не требует дополнительной обработки ситуаций, когда план экспериментов для точной функции не является подмножеством плана экспериментов для источника данных низкой точности функции,
- учитывает корреляции между источниками данных низкой и высокой точности,
- позволяет получать оценку неопределенности модели в точке.

Кокригинг используется для решения различных задач:

- аппроксимации[51, 50, 77, 76, 103, 107],
- оптимизации и адаптивного планирования экспериментов [54, 106],

- построения поверхности, на которой функция имеет заданное значение [25].

Отметим, что для суррогатной оптимизации и адаптивного дизайна эксперимента используются простые эвристические подходы. Например, в работе [34] предлагается использовать ожидаемое улучшение для получения новых точек, в которых нужно вычислить точную функцию, в работе [43] предлагается взвешивать ожидаемое улучшение для источника данных низкой и высокой точности в соответствии с относительной стоимостью вычисления этих функций.

Так же отметим, что в работах[76, 107] используется модель вида

$$u(\mathbf{x}) = \rho(\mathbf{x}) f(\mathbf{x}) + g(\mathbf{x}), \tag{3.1}$$

где $\rho(\mathbf{x})$ — функция из некоторого семейства (например, линейная функция).

Кокригинг используется и для построения аппроксимации точной функции в случае наличия множества моделей, упорядоченных по точности $u(\mathbf{x}_h)$, $f^1(\mathbf{x}_l)$, $f^2(\mathbf{x}_l), \ldots, f^L(\mathbf{x}_l)$ [38]. В этом случае используется следующее предположение о зависимости между моделями:

$$\begin{aligned} f^{L-1}(\mathbf{x}) &= \rho_L f^L(\mathbf{x}) + g^L(\mathbf{x}), \\ \cdots &= \cdots \\ f^1(\mathbf{x}) &= \rho_2 f^2(\mathbf{x}) + g^2(\mathbf{x}), \\ u(\mathbf{x}) &= \rho_1 f^1(\mathbf{x}) + g^1(\mathbf{x}). \end{aligned}$$

Предполагается, что $u(\mathbf{x}), f^i(\mathbf{x}), g^i(\mathbf{x})$ — реализации гауссовских процессов.

3.1.4 Отображение пространства

Пусть теперь существует отображение из пространства дизайна для источника данных высокой точности в пространство дизайна для источника данных низкой точности $\mathbf{x}_f = t(\mathbf{x}_u)$ такое, что

$$u(\mathbf{x}_u) \approx f(t(\mathbf{x}_u)).$$

Тогда, восстановив такое отображение, мы получим способ оценки выхода источника данных низкой точности с использованием источника данных низкой точности или его суррогатной модели. Так же, используя якобиан такого отображения, можно оценить градиенты функции, порожденной источником данных высокой точности.

Подход был, видимо, впервые предложен в работе [7] в 1994 году. Была показана тесная связь такого подхода с методом доверительной области [8]. Обзор работ, опубликованных до 2004 года, приведен в статье [8]. Работы отличаются семейством отображений $T = \{t(\mathbf{x})\}$, способом выбора отображений и методом оптимизации, который затем используется. В некоторых работах было получено описание теоретических свойств метода оптимизации, основанного на таком подходе [82, 83].

3.1.5 Другие подходы

Для решения некоторых прикладных задач используются подходы, которые не укладываются в перечисленные выше модели.

В работе [29] предложен следующий подход. Для выборки данных низкой точности строится аппроксимация вида:

$$\widetilde{f}(\mathbf{x}) = \sum_{i=1}^{H} h_i(\mathbf{x}),$$

где $h_i(\mathbf{x})$ — некоторый набор базисных функций. Затем для источника данных высокой точности строится регрессионная модель вида:

$$\widetilde{u}(\mathbf{x}) = \sum_{i=1}^{H} w_i h_i(\mathbf{x}).$$

Таким образом, для источника данных высокой точности необходимо настроить лишь H параметров, решив задачу построения модели линейной регрессии. Другим достоинством такого подхода является возможность проведения регуляризации весов базисных функций w_i.

В работе [10] источник данных низкой точности использовался для восстановления по профилю давления геометрии крыла. Для вычисления точной функции использовалась следующая процедура:

- Для заданного профиля давления восстановить геометрию, используя источник данных низкой точности.
- По геометрии крыла вычислить значение источника данных высокой точности.

Затем проводилась оптимизация профиля давления. Мотивацией для использования такого подхода являлась меньшая размерность параметризации профиля давления по сравнению с размерностью параметризации геометрии крыла (6 и 13 соответственно).

В некоторых работах [32, 81] источник низкой точности или его суррогатная модель используются в качестве одного из признаков, по которым строится суррогатная модель для источника данных высокой точности.

Так же отметим задачу, решаемую в работе [4]. Предполагается, что заданы две модели, на основании которых нужно оценить значение функции в точке. Предложенный подход к решению состоит в семплировании обоих моделей в окрестности искомой точки и восстановлении матрицы ковариации таких моделей. Тогда можно явно минимизировать дисперсию оценки функции в точке в классе оценок, линейных по известным значениям моделей в точке.

3.1.6 Выводы

Во всех представленных работах делается явное предположение о модели данных, то есть, о связи между источниками данных низкой и высокой точности. Попытки учета особенностей задачи ограничиваются наложением специальных байесовских предположений о параметрах модели [50, 76]. Например, в работе [50] дисперсия априорного распределения параметров была задана исходя из того, что реальные значения не могли отличаться больше чем на два порядка от априорных. Как правило, в статьях предлагается развитие методов уже представленных в литературе, таких как кокригинг и отображение пространства, или рассматриваются простые [2, 64, 94, 109, 40, 32, 81] или сложные [10] эвристики.

3.2 Регрессия на основе гауссовских процессов для разнородных данных

Рассмотрим случай разнородных данных: задана выборка данных низкой точности $\mathbf{S}_f = (D_f, \mathbf{f}) = \left\{\mathbf{x}_i^f, f(\mathbf{x}_i^f)\right\}_{i=1}^{n_f}$ и выборка данных высокой точности $\mathbf{S}_u = (D_u, \mathbf{u}) = \{\mathbf{x}_i^u, u(\mathbf{x}_i^u)\}_{i=1}^{n_u}$ с $\mathbf{x}_i^f, \mathbf{x}_i^u \in \mathbb{R}^d$, $f(\mathbf{x}), u(\mathbf{x}) \in \mathbb{R}$. Источники данных низкой точности $f(\mathbf{x})$ и высокой точности $u(\mathbf{x})$ моделируют одно и то же физическое явление, но с разной точностью.

Используя выборки значений данных разной точности, необходимо построить суррогатную модель $\widetilde{u}(\mathbf{x}) \approx u(\mathbf{x})$ источника данных высокой точности. Кроме того, мы хотим получить оценку неопределенности прогноза значений источника данных высокой точности в новых точках.

Для моделирования данных разной точности мы используем широко распространенную модель кокригинга, описанную А. Форрестером [34]:

$$f(\mathbf{x}) = f_0(\mathbf{x}) + \varepsilon_f, \;\; u(\mathbf{x}) = \rho f_0(\mathbf{x}) + g(\mathbf{x}),$$

где $g(\mathbf{x}) = g_0(\mathbf{x}) + \varepsilon_g$, $f_0(\mathbf{x})$, $g_0(\mathbf{x})$ — независимые гауссовские процессы с нулевыми средними и ковариационными функциями $c_f(\mathbf{x}, \mathbf{x}')$ и $c_g(\mathbf{x}, \mathbf{x}')$ соответственно, и ε_f, ε_g — гауссовский белый шум с дисперсиями σ_l^2 и σ_d^2, соответственно. Обозначим $D = (D_f, D_u)^T$, $\mathbf{y} = (\mathbf{f}, \mathbf{u})^T$.

Тогда апостериорное среднее для значений источника данных высокой точности в новых точках D имеет вид

$$\widetilde{\mathbf{u}}(D^*) = \mathbf{C}(D^*, D)\mathbf{C}^{-1}\mathbf{y},$$

где

$$\mathbf{C}(D^*, D) = \begin{pmatrix} \rho\mathbf{C}_f(D^*, D_f) \\ \rho^2\mathbf{C}_f(D^*, D_u) + \mathbf{C}_g(D^*, D_u) \end{pmatrix},$$

$$\mathbf{C}(D, D) = \begin{pmatrix} \mathbf{C}_f(D_f, D_f) & \rho\mathbf{C}_f(D_f, D_u) \\ \rho\mathbf{C}_f(D_u, D_f) & \rho^2\mathbf{C}_f(D_u, D_u) + \mathbf{C}_g(D_u, D_u) \end{pmatrix},$$

$\mathbf{C}_f(D_a, D_b)$, $\mathbf{C}_g(D_a, D_b)$ — матрицы попарных ковариаций гауссовских процессов $f(\mathbf{x})$ и $g(\mathbf{x})$ в точках из некоторых множеств D_a и D_b соответственно. Апостериорная ковариационная матрица имеет вид:

$$\begin{aligned} \mathbb{V}\left(D^*\right) = \rho^2\mathbf{C}_f(D^*, D^*) + \mathbf{C}_g(D^*, D^*) - \\ -\mathbf{C}(D^*, D)\mathbf{C}^{-1}\left(\mathbf{C}(D^*, D)\right)^T. \end{aligned}$$

Для оценки параметров ковариационных функций гауссовских процессов $f(\mathbf{x})$ и $g(\mathbf{x})$ используют следующий алгоритм:

1. Оценить параметры ковариационной функции $c_f(\mathbf{x}, \mathbf{x})$, используя алгоритм для регрессии на основе гауссовских процессов с данными одной точности и выборкой $\mathbf{S} = \mathbf{S}_f$,

2. Подсчитать оценки апостериорного среднего $\widetilde{f}(\mathbf{x})$ для гауссовского процесса $f_0(\mathbf{x})$ и $\mathbf{x} \in D_h$,

3. Оценить параметр ρ и параметры гауссовского процесса $g(\mathbf{x})$ с ковариационной функцией $c_g(\mathbf{x}, \mathbf{x}')$, максимизируя правдоподобие (2.1) с выборкой $\mathbf{S} = \mathbf{S}_{\text{diff}} = (D_u, \mathbf{u} - \rho\widetilde{\mathbf{f}}(D_h))$ и $c(\mathbf{x}, \mathbf{x}') = c_g(\mathbf{x}, \mathbf{x}')$.

3.3 Разреженная регрессия на основе гауссовских процессов для разнородных данных

Эффективное использование регрессии на основе гауссовских процессов для разнородных данных возможно только для выборок с размером не больше нескольких тысяч точек, так как в процессе обучения модели и ее использования приходится обращать ковариационную матрицу размера $n \times n$, где $n = n_u + n_f$ — общий размер выборки. Вычислительная сложность такой процедуры составляет $O(n^3)$.

В исследовании предлагается новый приближенный метод для регрессии на основе гауссовских процессов для выборок разнородных данных, который может быть использован в том числе для разнородных данных. Подход основан на использовании аппроксимации Нистрема матриц $\mathbf{C}(D^*, D)$, $\mathbf{C}$ и $\mathbf{C}(D^*, D^*)$ для подвыборки базовых точек исходной выборки. Представленные результаты являются обобщением результатов Л. Фостера [35] на случай разнородных данных.

Зададим подвыборку $\mathbf{S}_1 = (D^1, \mathbf{y}^1)$, $D^1 = \left(D_f^1, D_u^1\right)^T, \mathbf{y}^1 = \left(\mathbf{f}(D_f^1), \mathbf{u}(D_u^1)\right)^T$ *базовых* точек из исходной выборки такого размера $n_1 = n_u^1 + n_f^1$, что для заданной подвыборки имеющиеся вычислительные ресурсы позволяют обратить соответствующие ковариационные матрицы и оценить параметры гауссовскго процесса за разумное время. Достаточно надежный метод задания подвыборки базовых точек состоит в случайном выборе без повторения точек из исходной выборки [78].

Тогда используя подвыборку базовых точек и положив

$$\mathbf{C}_{11} = \begin{pmatrix} \mathbf{C}_f(D_f^1, D_f^1) & \rho \mathbf{C}_f(D_f^1, D_u^1) \\ \rho \mathbf{C}_f(D_u^1, D_f^1) & \rho^2 \mathbf{C}_f(D_u^1, D_u^1) + \mathbf{C}_g(D_u^1, D_u^1) \end{pmatrix},$$

$$\mathbf{C}_1 = \begin{pmatrix} \mathbf{C}_f(D_f^1, D_f) & \rho \mathbf{C}_f(D_f^1, D_u) \\ \rho \mathbf{C}_f(D_u^1, D_f) & \rho^2 \mathbf{C}_f(D_u^1, D_u) + \mathbf{C}_g(D_u^1, D_u) \end{pmatrix},$$

$$\mathbf{C}_1^* = \begin{pmatrix} \rho \mathbf{C}_f(D^*, D_f^1) \\ \rho^2 \mathbf{C}_f(D^*, D_u^1) + \mathbf{C}_g(D^*, D_u^1) \end{pmatrix}$$

для новых точек $D^* = \{\mathbf{x}_i^*\}_{i=1}^{n^*}$, получаем аппроксимации матриц $\mathbf{C}(D^*, D)$, $\mathbf{C}$ и $\mathbf{C}(D^*, D^*)$ соответственно:

$$\widehat{\mathbf{C}}(D^*, D) = \mathbf{C}_1^* \mathbf{C}_{11}^{-1} \mathbf{C}_1,$$
$$\widehat{\mathbf{C}} = (\mathbf{C}_1)^T \mathbf{C}_{11}^{-1} \mathbf{C}_1,$$
$$\widehat{\mathbf{C}}(D^*, D^*) = \mathbf{C}_1^* \mathbf{C}_{11}^{-1} (\mathbf{C}_1^*)^T.$$

Определим

$$R = \begin{pmatrix} \frac{1}{\sigma_l} I_{n_l} & 0 \\ 0 & \frac{1}{\sqrt{\rho^2 \sigma_f^2 + \sigma_g^2}} I_{n_u} \end{pmatrix},$$

где I_k — единичная матрица размера k, $\mathbf{C}_1 = R\mathbf{C}_1$, и $V = \mathbf{C}_1 V_{11}^{-T}$, V_{11} — разложение Холецкого матрицы $\mathbf{C}_{11}$.

Утверждение 3.1. *Для апостериорного среднего аппроксимация Нистрема имеет вид:*

$$\widetilde{\mathbf{u}}(D^*) = \mathbf{C}_1^* V_{11} (I_{n_1} + V^T V)^{-1} V^T \mathbf{y},$$

Для оценки неопределенности прогноза результат будет зависеть от используемого приближения ковариационных матриц. Рассмотрим следующий вариант приближения ковариационных матриц: $c(\mathbf{x}^*, D) \approx \mathbf{C}_1^* \mathbf{C}_{11}^{-1} \mathbf{C}_1^T$, $c(D, D) \approx R^{-2} + \mathbf{C}_1 \mathbf{C}_{11}^{-1} \mathbf{C}_1^T$.

Утверждение 3.2. *Для апостериорной дисперсии в случае использования описанного выше приближения для ковариационных матриц аппроксимации Нистрема будут иметь следующий вид:*

$$\widehat{\sigma}^2(\mathbf{x}^*) = v(\mathbf{x}^*, \mathbf{x}^*) - \mathbf{C}_1^* V_{11}^{-1} (I + V^T V)^{-1} (V^T V) V_{11}^{-T} \mathbf{C}_1^{*T}.$$

Отметим, что прямое применение результата, полученного в работе [57] для данных одной точности дает оценку точности используемой аппроксимации Нистрема для разнородных данных. А именно, с вероятностью $1-\delta$ выполнено:

$$\frac{\|\mathbf{C}(D^*,D^*)-\widehat{\mathbf{C}}(D^*,D^*)\|_2}{\|\mathbf{C}(D^*,D^*)\|_2}\le\frac{\|\mathbf{C}(D^*,D^*)-\widehat{\mathbf{C}}_{n_1}(D^*,D^*)\|_2}{\|\mathbf{C}(D^*,D^*)\|_2}+\varDelta,$$
$$\frac{\|\mathbf{C}(D^*,D)-\widehat{\mathbf{C}}(D^*,D)\|_2}{\|\mathbf{C}(D^*,D)\|_2}\le\frac{\|\mathbf{C}(D^*,D)-\widehat{\mathbf{C}}_{n_1}(D^*,D)\|_2}{\|\mathbf{C}(D^*,D)\|_2}+\varDelta,$$

где $\varDelta$ порядка $O\left(\frac{1}{\sqrt{n}}\right)O\left(\sqrt{\log\frac{1}{\delta}}\right)$, $\|\cdot\|_2$ — l_2 матричная норма, а $\widehat{\mathbf{C}}_{n_1}(D^*,D^*)$ — наилучшая в смысле l_2 нормы аппроксимация ранга n_1.

Полученные теоретические оценки качества аппроксимации мотивируют использованние Алгоритма для построения регрессионных моделей разнородных данных

Алгоритм 1 Построение регрессионной модели разнородных данных с использование низкоранговой аппроксимации ковариационных матриц

Вход: Выборки разнородных данных $\mathbf{S}_u$, $\mathbf{S}_f$, размеры подвыборок базовых точек n_f^1, n_u^1

1: Выбрать случайным образом подвыборки базовых точек $\mathbf{S}_f^1$, $\mathbf{S}_u^1$ размером n_f^1, n_u^1 из выборок $\mathbf{S}_f$, $\mathbf{S}_u$ соответственно.
2: Используя подвыборки $\mathbf{S}_f^1$, $\mathbf{S}_u^1$ оценить параметры ковариационных функций для регрессионной модели разнородных данных на основе гауссовских процессов.
3: Подсчитать матрицы $\boldsymbol{\alpha}=V_{11}(I_{n_1}+V^TV)^{-1}V^T\mathbf{y}$, $M_1=V_{11}^{-1}(I+V^TV)^{-1}(V^TV)V_{11}^{-T}$, необходимые для вычисления регрессионной модели и ее неопределенности.
4: **Вернуть** $\boldsymbol{\alpha}, M_1$.

3.4 Регрессия на основе гауссовских процессов для разнородных данных при наличии черного ящика для источника данных низкой точности

Часто инженеру доступен черный ящик для источника данных низкой точности $f(\mathbf{x})$; черный ящик вычисляет значение источника данных низкой точности в любой точке из пространства дизайна $\mathbb{X} \subseteq \mathbb{R}^d$ в режиме реального времени. Пусть мы построили регрессионную модель для гауссовских процессов с использованием данных разной точности. Для того чтобы улучшить качество модели обновим апостериорное среднее и апостериорную дисперсию для прогноза $u(\mathbf{x})$ в новой точке $\mathbf{x}$ используя значение источника данных низкой точности в этой точке $f(\mathbf{x})$, подсчитанное с помощью черного ящика. Опишем вычислительно эффективную процедуру обновления модели.

Зададим

$$\mathbf{c}_f(\mathbf{x}, D) = \begin{pmatrix} \mathbf{C}_f(\mathbf{x}, D_f) \\ \rho \mathbf{C}_f(\mathbf{x}, D_u) \end{pmatrix},$$

где $\mathbf{x}$ — некоторая новая точка, в который мы хотим оценить значение источника данных высокой точности. Для выборки с включенной дополнительной точкой $\mathbf{x}$ мы получим обновленную ковариационную матрицу:

$$\mathbf{C}_{\text{exp}} = \begin{pmatrix} \mathbf{C} & \mathbf{c}_f \\ \mathbf{c}_f^T & c_f(\mathbf{x}, \mathbf{x}) \end{pmatrix}.$$

Обновленный вектор ковариаций между обучающей выборкой и значением источника данных высокой точности в точке $\mathbf{x}$ имеет вид

$$\mathbf{c}_{\text{exp}} = \begin{pmatrix} \rho \mathbf{C}_f(\mathbf{x}, D_f) \\ \rho^2 \mathbf{C}_f(\mathbf{x}, D_u) + \mathbf{C}_g(\mathbf{x}, D_u) \\ \rho c_f(\mathbf{x}, \mathbf{x}) \end{pmatrix}.$$

Используя значение $f(\mathbf{x})$, подсчитанное с помощью черного ящика, зададим $\mathbf{y}_{\text{exp}} = \begin{pmatrix} \mathbf{y} \\ f(\mathbf{x}) \end{pmatrix}$. Тогда обновленные значения для апостериорного среднего и дисперсии имеют вид:

$$\widetilde{u}^{\text{exp}}(\mathbf{x}) = \mathbf{k}_{\text{exp}} \mathbf{C}_{\text{exp}}^{-1} \mathbf{y}_{\text{exp}}, \tag{3.2}$$

$$\mathbb{V}_{\text{exp}}(\mathbf{x}) = \rho^2 \mathbf{C}_f(\mathbf{x}, \mathbf{x}) + \mathbf{C}_g(\mathbf{x}, \mathbf{x}) - \mathbf{c}_{\text{exp}}^T \mathbf{C}_{\text{exp}}^{-1} \mathbf{c}_{\text{exp}}. \tag{3.3}$$

Утверждение 3.3. *Пусть мы знаем разложение Холецкого L и его обратную матрицу L^{-1} для ковариационной матрицы исходной обучающей выборки $\mathbf{C}$.*

Тогда подсчет обновленного апостериорного среднего $\widetilde{u}^{\exp}(\mathbf{x})$ *с использованием* (3.2) *и обновленной апостериорной дисперсии* $\mathbb{V}_{\exp}(\mathbf{x})$ *с использованием* (3.3) *требует* $O(n^2)$ *операций, где* $n = n_u + n_f$.

Такая вычислительно эффективная процедура обновления регрессионной модели путем добавления к выборке значения функции низкой точности может быть описана в виде алгоритма 2.

Алгоритм 2 Процедура эффективного использования черного ящика для источника данных низкой точности для подсчета аппроксимации и неопределенности аппроксимации

Вход: Выборки разнородных данных $\mathbf{S}_u$, $\mathbf{S}_f$, точка $\mathbf{x}^*$, в которой нужно оценить $\widetilde{u}(\mathbf{x})$

1: Вычислить значение функции низкой точности $f(\mathbf{x}^*)$ в точке $\mathbf{x}^*$.
2: Подсчитать разложение Холецкого ковариационной матрицы расширенной выборки, в которую добавили $f(\mathbf{x}^*)$.
3: Получить на основе разложения Холецкого для расширенной выборки прогноз $\widetilde{u}^{\exp}(\mathbf{x}^*)$ и его неопределенность $\mathbb{V}_{\exp}(\mathbf{x}^*)$.
4: **Вернуть** $\widetilde{u}^{\exp}(\mathbf{x}^*), \mathbb{V}_{\exp}(\mathbf{x}^*)$.

3.5 Комплекс программ

Разработанные алгоритмы были переписаны на языке `C++` и вошли в состав комплекса программ MACROS и pSeven компании DATADVANCE. Компонента Generic Tool for Data Fusion в библиотеке MACROS получила третий уровень готовности технологии, что соответствует возможности решения прикладных задач с помощью такой технологии [18, 47]. Перечислим способы моделирования данных разной точности, представленную в этой компоненте:

- на основе построения модели разности между функцией высокой точности и функцией низкой точности;
- на основе гауссовских процессов;
- на основе гауссовских процессов для больших выборок;

- на основе гауссовских процессов для случая наличия черного ящика для функции низкой точности;
- на основе гауссовских процессов при наличии данных более чем двух различных упорядоченных точностей.

Отметим дополнительные функциональные особенности модуля:

- Для каждого способа моделирования использовалась байесовская процедура оценки параметров, аналогичная представленной в разделе 2 настоящей диссертации. Использование такой процедуры позволило избежать ухудшения вычислительных свойств задачи, связанных с плохой обусловленностью ковариационной матрицы данных [68, 74].
- Предложен механизм обеспечения близости модели, построенной с использование данных разной точности, к данным высокой точности, то есть для модели и точек из выборки значений функции высокой точности будет выполнено, что $|\widetilde{u}(\mathbf{x}_i) - u(\mathbf{x}_i)|$ мало, при этом гарантируются достаточно хорошие вычислительные свойства задачи [98, 80].
- Разработана процедура скользящего контроля, позволяющая проводить его в случае наличия данных разной точности.
- Предложен алгоритм, позволяющий эффективно строить модели с размерностью выхода больше одного для данных разной точности.
- Предложена эвристика для выбора способа моделирования данных разной точности, основанная на размере выборки и размерности данных.
- Предложена эвристика для выбора дизайна функции низкой точности в случае наличия черного ящика для функции низкой точности.
- Разработана опция, позволяющая пользователю выбирать разумный с его точки зрения компромисс между временем обучения модели и ее точностью.

Благодаря разработанному набору опций инструмент с одной стороны оказывается достаточно гибким, с другой — является доступным для неспециалиста в анализе данных. Таким образом, разработанный модуль позволяет решать задачи инженерного проектирования тем, кто является специалистами в инженерном проектировании, а не в анализе данных.

3.6 Выводы

В работе рассмотрено два подхода к построению и использованию регрессионных моделей разнородных данных, которые позволяют работать с выборками большого размера в случае использования регрессии на основе гауссовских процессов: первый из предложенных подходов аппроксимирует ковариационную матрицу выборки и ее обратную матрицу, второй подход использует возможность использования источника данных низкой точности в режиме реального времени для обновления модели, добавляя в выборку значение источника данных низкой точности в той точке, в которой аппроксимируется значение источника данных высокой точности. Предложенные здесь подходы позволяют увеличить размер выборки, для которой оказывается применима регрессия на основе гауссовских процессов для разнородных данных.

Предложенные подходы вошли в состав библиотеки MACROS, содержащей функциональность для интеллектуального анализа данных и многокритериальной оптимизации, и платформе pSeven, предназначенной для автоматизации инженерного анализа, многодисциплинарной оптимизации и анализа данных.

Глава 4

Выбор соотношения между размерами выборок разнородных данных, минимизирующего минимаксную ошибку интерполяции

4.1 Введение

В некоторых случаях данные для регрессионного моделирования могут быть разнородными: некотрые данные поступают из источника данных высокой точности, а другие — из источника данных низкой точности [34]. В этой главе будет говорить о подходах на основе гауссовских процессов [111, 80] для построения регрессионных моделей разнородных источников данных [108, 40, 55].

Первыми работами, дающими оценку точности методов регрессионного моделирования на основе гауссовских процессов, видимо, следует считать работы Винера и Колмогорова [105, 53]. В этих работах была получена ошибка интерполяции в точке для одномерного пространства признаков и бесконечной выборки на сетке. Дальнейшее развитие подходов к оценке точности таких методов дано в книге Штайна [92], работы которого вдохновлены фундаментальными результатами Ибрагимова и Розанова [45].

В последнее десятилетие удалось наконец получить результаты для ошибки интерполяции в области — интегралу математического ожидания квадрата ошибки регрессионной модели по заданной области. Достаточно представительно современные результаты отражают работы Ван Дер Ваарта [101] и Голубева и Крымовой [37]. В работе Ван Дер Ваарта результаты получены для конечной выборки, но для сложных предположениях об истинной функции. В работе Голубева и Крымовой результаты получены для выборки на бесконечной решетке, в том числе и в минимаксном смысле, то есть когда известна гладкость гауссовского процесса, но неизвестна его ковариационная функция. Отметим так же, что этими двумя работами современный прогресс в решении задачи оценки интерполяции не исчерпывается, смотрите, например, работы [113, 96, 16].

В то время как для однородных данных полученные результаты в достаточно многих случаях позволяют получить ошибку интерполяции, для разнородных данных рассмотрение ошибки интерполяции для регрессии на основе гауссовских процессов с теоретической точки зрения удалось найти лишь в работе [112]. В ней для квадратичной экспоненциальной ковариационной функции и квадратичной ошибки в заданной одной точке получен критерий целесообразности использования разнородных данных. Другие работы, посвященные регрессионному моделированию разнородных данных с теоретической точки зрения [102, 15, 73] фокусируются на результатах, связанных с качеством оценки параметров в таких моделях, и не освещают вопрос применимости и оптимальности подхода на основе использования разнородных данных.

Из-за отсутствия теоретических результатов инженерам приходится использовать различные эвристики для выбора плана экспериментов, позволяющего получить высокие по качеству модели [3, 88, 58]; или использовать адаптивные планы экспериментов [79, 48, 21, 60], которые предполагают наличие достаточно точной регрессионной модели данных.

В данной главе диссертационной работы получены следующие результаты:

- **Минимаксная ошибка интерполяции для многомерного случая.** Получена ошибка интерполяции для регрессии на основе гауссовских процессов для многомерного случая и известной ковариационной функции. Затем получена минимаксная ошибка интерполяции для функций из заданного класса гладкости для многомерного пространства признаков. Этот результат является нетривиальным обобщением результата, полученного в [37].

- **Оптимальное соотношение размеров выборок разнородных данных.** Получена ошибка интерполяции для заданной ковариационной функции для разнородных данных для аддитивной модели кокригинга [51]. Затем получена в тех же условиях минимаксная ошибка интерполяции. С использованием полученной минимаксной ошибки становится возможным выявить, в каких случаях использование разнородных данных позволяет улучшить модель по сравнению со случаем однородных данных для заданного бюджета вычислений. Так же получено оптимальное соотношение между размерами выборок разнородных данных. Использованная теоретическая постановка отличается от того, что происходит в реальных задачах: рассматривается выборка на бесконечной решетке и для получения минимаксно оптимальной ошибки требуется знание относительных гладкостей случайных процессов, соответствующих источникам данных высокой и низкой точности. Однако, полученные теоретические результаты позволяют обосновать предложенный в исследовании алгоритм для решения реальных задач.

- **Алгоритм для выбора соотношения между размерами выборок разнородных данных.** Предложен алгоритм для выбора соотношения между размерами выборок разнородных данных, мотивированный полученными теоретическими результатами. В то время как обычно подход к выбору плана эксперимента предполагает наличие достаточно точной модели данных [79], предложенный в исследовании метод требует лишь знания коэффициента корреляции между источниками разнородных данных и относительной стоимости использования таких источников даных. Применимость предложенного алгоритма исследуется путем сравнения с рядом естественных эвристик на искусственных и реальных данных в разделе 5.3 следующей главы.

Так же глава содержит доказательства полученных теорем.

4.2 Минимаксная ошибка интерполяции для регрессии на основе гауссовских процессов

Для регрессии на основе гауссовских процессов существует зазор между теоретическими результатами и приложениями. А именно, так как основной подход к теоретическому исследованию задачи включает преобразования Фурье, обычно рассматривают план экспериментов на бесконечной решетке [37, 92]. Однако, ясно, что несмотря на это, во многих случаях схожие результаты имеют место на практике. Поэтому в этой главе рассматриваются планы экспериментов на бесконечных решетках, а в следующей главе, посвященной вычислительным экспериментам, показано, как соотносятся полученные теоретические выводы с тем, что происходит на практике, а именно для случайных планов экспериментов конечного размера.

4.2.1 Ошибка интерполяции

Пусть $f(\mathbf{x})$ — стационарный гауссовский процесс на $\mathbb{R}^d$ с ковариационной функцией $c(\mathbf{x}) = \mathbb{E}(f(\mathbf{x}_0 + \mathbf{x}) - \mathbb{E}f(\mathbf{x}_0 + \mathbf{x}))(f(\mathbf{x}_0) - \mathbb{E}f(\mathbf{x}_0))$ и спектральной плотно-

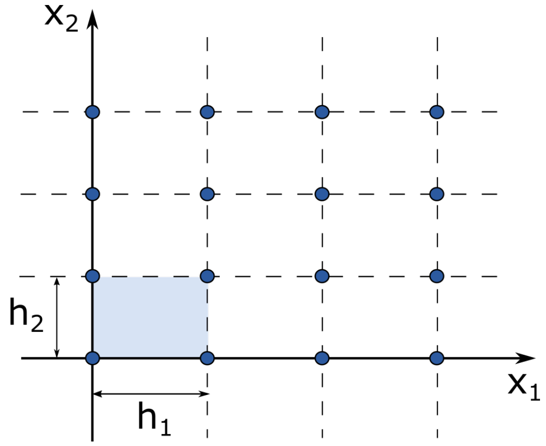


Рис. 4.1 – План экспериментов D_H для $d = 2$.

стью $F(\boldsymbol{\omega})$

$$F(\boldsymbol{\omega}) = \int_{\mathbb{R}^d} e^{2\pi \mathrm{i} \boldsymbol{\omega}^{\mathrm{T}} \mathbf{x}} c(\mathbf{x}) d\mathbf{x} \,.$$

Пусть мы знаем значения реализации $f(\cdot)$ на бесконечной прямоугольной сетке $D_H = \{\mathbf{x_k} : \mathbf{x_k} = H\mathbf{k}, \mathbf{k} \in \mathbb{Z}^d\}$, где H — диагональная матрица с элементами $h_1, \ldots, h_d$ на диагонали. Пример такого плана эксперимента для размерности пространства признаков $d = 2$ приведен на рисунке 4.1.

Ошибка интерполяции для области $\Omega_H = [0, h_1] \times \ldots \times [0, h_d]$ определяется как:

$$\sigma_H^2(\widetilde{f}, F) = \frac{1}{\mu(\Omega_H)} \int_{\Omega_H} \mathbb{E} \left[\widetilde{f}(\mathbf{x}) - f(\mathbf{x}) \right]^2 d\mathbf{x} \,, \tag{4.1}$$

где $\mu(\Omega_H) = \prod_{i=1}^d h_i$ — мера Лебега Ω_H, и $\widetilde{f}(\mathbf{x})$ — интерполяция $f(\mathbf{x})$. Мы будем рассматривать $\widetilde{f}(\mathbf{x})$ вида

$$\widetilde{f}(\mathbf{x}) = \mu(\Omega_H) \sum_{\mathbf{x}' \in D_H} K(\mathbf{x} - \mathbf{x}') f(\mathbf{x_k}) \,, \tag{4.2}$$

где $K(\cdot)$ — симметричное ядро.

Теорема 4.1. *Ошибка интерполяции для $\widetilde{f}(\mathbf{x})$ из (4.2) и наблюдений на D_H для стационарного гауссовского процесса $f(\mathbf{x})$ со спектральной плотностью $F(\boldsymbol{\omega})$ имеет вид*

$$\sigma_H^2(\widetilde{f}, F) = \int_{\mathbb{R}^d} F(\boldsymbol{\omega}) \left[\left(1 - \widehat{K}(\boldsymbol{\omega})\right)^2 + \sum_{\mathbf{x} \in D_{H^{-1}} \setminus \{\mathbf{0}\}} \widehat{K}^2 \left(\boldsymbol{\omega} + \mathbf{x}\right) \right] d\boldsymbol{\omega} \,,$$

где $\widehat{K}(\boldsymbol{\omega})$ — преобразование Фурье $K(\boldsymbol{\omega})$. Более того, оптимальное $\widehat{K}(\boldsymbol{\omega})$, минимизирующее ошибку интерполяции, имеет вид:

$$\widehat{K}(\boldsymbol{\omega}) = \frac{F(\boldsymbol{\omega})}{\sum_{\mathbf{x}\in D_{H^{-1}}} F\left(\boldsymbol{\omega}+\mathbf{x}\right)}\,.$$

Следствие 1. *Функция $\widetilde{f}(\mathbf{x})$, которая минимизирует квадратичную ошибку $\mathbb{E}(\widetilde{f}(\mathbf{x}) - f(\mathbf{x}))^2$ имеет вид (4.2), где $K(\cdot)$— симметричное ядро. Это обосновывает использование $\widetilde{f}(\mathbf{x})$ вида (4.2) для интерполяции.*

Следствие 2. *Легко видеть, что для $\widetilde{f}(\mathbf{x})$ из (4.2) выполнено, что*

$$\sigma^2_H(\widetilde{f}, F) = \sigma^2_{SH}(\widetilde{f}, F)\,,$$

где $S = \mathrm{diag}(s_1, \ldots, s_d)$, и $s_i \in \mathbb{Z}^+, i = 1, \ldots, d$.

Используя теорему 4.1 можно получать ошибки интерполяции для различных ковариационных функций. Например,

Следствие 3. *Для гауссовского процесса на $\mathbb{R}$ с экспоненциальной ковариационной функцией, которой соответствует спектральная плотность $F_\theta(\omega) = \frac{\theta}{\theta^2+\omega^2}$ ошибка интерполяции (4.1) для наилучшей интерполяции имеет вид:*

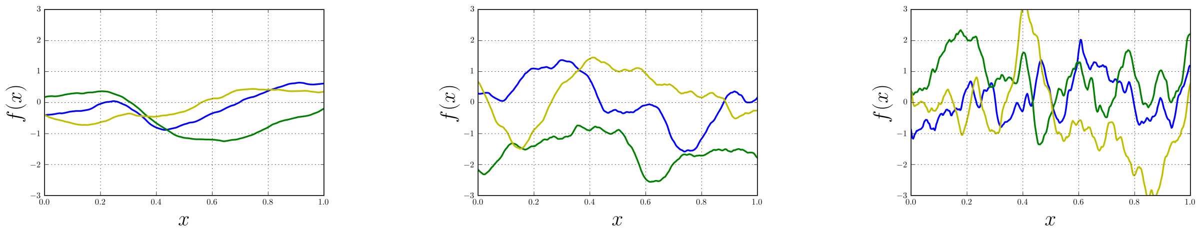
$$\sigma^2_h(\widetilde{f}, F_\theta) \approx \frac{2}{3}\pi^2\theta h + O((\theta h)^2),\ \theta h \to 0\,.$$

Следствие 4. *Для гауссовского процесса на $\mathbb{R}$ с квадратичной экспоненциальной ковариационной функцией, которой соответствует спектральная плотность $F_\theta(\omega) = \frac{1}{\sqrt{\theta}}\exp\left(-\frac{\omega^2}{2\theta}\right)$ ошибка интерполяции (4.1) для наилучшей интерполяции ограничена сверху и снизу:*

$$\frac{4}{3}h\sqrt{\theta}\exp\left(-\frac{1}{8h^2\theta}\right) \le \sigma^2_h(\widetilde{f}, F_\theta) \le 7h\sqrt{\theta}\exp\left(-\frac{1}{8h^2\theta}\right),\ \theta h^2 \to 0\,.$$

4.2.2 Минимаксная ошибка интерполяции

Для многих ковариационных функций явное вычисление ошибки интерполяции в наилучшем случае не может быть сделано аналитически. Кроме того,



(а) $\mathcal{F}(10,1)$ (б) $\mathcal{F}(100,1)$ (в) $\mathcal{F}(1000,1)$

Рис. 4.2 – Реализации гауссовских процессов для ковариационной функции Матерна $c(x) = (1+\sqrt{3}\theta|x|)\exp(-\sqrt{3}\theta|x|)$ ($\nu = \frac{3}{2}$) и различных значений L в $\mathcal{F}(L,1)$ и $d = 1$.

во многих случаях истинная ковариационная функция неизвестна, а получение ошибок интерполяции для неверно заданной ковариационной функции — еще более технически сложная задача.

Вместо этого рассмотрим минимаксную ошибку интерполяции, которая соответствует ответу в наихудшем случае. Определим множество $\mathcal{F}(L,\boldsymbol{\lambda})$ спектральных плотностей $F(\boldsymbol{\omega})$ для заданного $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_d) \in \mathbb{R}^d$ и $L > 0$ как

$$\mathcal{F}(L,\boldsymbol{\lambda}) = \left\{ F \,:\, \mathbb{E}\sum_{i=1}^{d} \lambda_i^2 \left(\frac{\partial f_F(\mathbf{x})}{\partial x_i}\right)^2 \leq L\,, \mathbf{x} \in \mathbb{R}^d \right\}, \tag{4.3}$$

где $f(\mathbf{x}) = f_F(\mathbf{x})$ — гауссовский процесс со спектральной плотностью $F(\boldsymbol{\omega})$, наблюдаемый в точке $\mathbf{x} \in \mathbb{R}^d$. Примеры реализаций гауссовских процессов для различных L для $d = 1$ и ковариационной функции Матерна [80] показаны на рисунке 4.2.

Определим минимаксную ошибку интерполяции (или минимаксный риск) как

$$R^H(L,\boldsymbol{\lambda}) = \inf_{\widetilde{f}} \sup_{F \in \mathcal{F}(L,\boldsymbol{\lambda})} \sigma_H^2(\widetilde{f}, F)\,.$$

Ошибка описывает насколько большой будет ошибка интерполяции в наихудшем случае. Выполнена следующая теорема:

Теорема 4.2. *Для гауссовского процесса $f(\mathbf{x})$ определенного на $\mathbb{R}^d$ и известного для точек из D_H, такого что его спектральная плотность принадлежит*

$\mathcal{F}(L, \boldsymbol{\lambda})$, минимаксная ошибка интерполяции имеет вид

$$R^H(L, \boldsymbol{\lambda}) = \frac{L}{2\pi^2} \max_{i \in \{1,\dots,d\}} \left(\frac{h_i}{\lambda_i}\right)^2 .$$

Более того, $\widetilde{f}(\mathbf{x})$, минимизирующая минимаксную ошибку интерполяции, имеет вид

$$\widetilde{f}(\mathbf{x}) = \mu(\Omega_H) \sum_{\mathbf{x}' \in D_H} K(\mathbf{x} - \mathbf{x}') f(\mathbf{x}') ,$$

где $K(\mathbf{x})$ — симметричное ядро, Фурье преобразование которого $\widehat{K}(\boldsymbol{\omega})$ имеет вид

$$\widehat{K}(\boldsymbol{\omega}) = \begin{cases} 1 - \sqrt{\sum_{i=1}^d \omega_i^2 \cdot h_i^2} & \textit{если } \sum_{i=1}^d \omega_i^2 \cdot h_i^2 \leq 1 , \\ 0, & \textit{иначе} . \end{cases}$$

Отметим, что можно минимизировать минимаксную ошибку интерполяции по матрице H для заданного среднего количества точек в гиперкубе со стороной один: $\prod_{i=1}^d \frac{1}{h_i} = n$. Соответствующая оптимальная матрица $H^* = \mathrm{diag}(h_1^*, \dots, h_d^*)$ имеет вид:

$$h_i^* = \sqrt[d]{\frac{n\lambda_i^d}{\prod_{j=1}^d \lambda_j}} .$$

И для нее минимаксная ошибка интерполяции: $R^{H^*}(L, \boldsymbol{\lambda}) = \frac{L}{2\pi^2} \sqrt[\frac{d}{2}]{\frac{n}{\prod_{i=1}^d \lambda_i}}$.

4.3 Минимаксная ошибка интерполяции для модели разнородных источников данных

4.3.1 Модель разнородных источников данных

Приведем еще раз определение модели, идентичное введенному в разделе 1.3. Пусть истинная функция имеет вид:

$$u(\mathbf{x}) = \rho f(\mathbf{x}) + g(\mathbf{x}) , \tag{4.4}$$

где ρ — фиксированная константа, и $f(\mathbf{x})$ и $g(\mathbf{x})$ — реализации двух независимых стационарных гауссовских процессов определенных на $\mathbb{R}^d$. Спектральная плотность случайного процесса, соответствующего $f(\mathbf{x})$ — $F(\boldsymbol{\omega})$, а спектральная плотность случайного процесса, соответствующего $g(\mathbf{x})$ — $G(\boldsymbol{\omega})$.

Будем говорить о реализации $u(\mathbf{x})$ как об источнике данных высокой точности, и о реализации $f(\mathbf{x})$ как об источнике данных низкой точности. Тогда $g(\mathbf{x})$ — поправка к $f(\mathbf{x})$, возникающая из-за низкой точности источника данных, соответствующего $f(\mathbf{x})$.

Наблюдаются значения $u(\mathbf{x})$ и $f(\mathbf{x})$, и задача состоит в построении аппроксимации $\widetilde{u}(\mathbf{x})$ функции, порожденной источником данных высокой точности $u(\mathbf{x})$ на основе наблюдений, порожденных разнородными источниками данных.

4.3.2 Ошибка интерполяции

Естественно предположить, что мы наблюдаем дешевый источник данных низкой точности $f(\mathbf{x})$ на более плотной сетке, чем дорогой источник данных высокой точности $u(\mathbf{x})$. А именно, мы наблюдаем $u(\mathbf{x})$ на $D_u = D_H$, и $f(\mathbf{x})$ на $D_f = D_{\frac{H}{m}}$ с $m \in \mathbb{Z}^+$. Таким образом, $D_u \subseteq D_f$.

Используя эти наблюдения, мы строим интерполяцию $u(\mathbf{x})$ на гиперкубе Ω_H. Функция $\widetilde{u}(\mathbf{x})$ минимизирует ошибку интерполяции $\sigma^2_{H,m}(\widetilde{u}, F, G, \rho)$ для наблюдений $u(\mathbf{x})$ на D_H и $f(\mathbf{x})$ на $D_{\frac{H}{m}}$:

$$\sigma^2_{H,m}(\widetilde{u}, F, G, \rho) = \frac{1}{\mu(\Omega_H)} \int_{\Omega_H} \mathbb{E}\left[\widetilde{u}(\mathbf{x}) - u(\mathbf{x})\right]^2 d\mathbf{x}\,. \tag{4.5}$$

Теорема 4.3. *Минимум ошибки интерполяции* (4.5) *для модели разнородных источников данных* (4.4) *и наблюдений* $u(\mathbf{x})$ *на* D_H *и* $f(\mathbf{x})$ *на* $D_{\frac{H}{m}}$ *имеет вид:*

$$\sigma^2_{H,m}(\widetilde{u}, F, G, \rho) = \sigma^2_H(\widetilde{g}, G) + \rho^2 \sigma^2_{\frac{H}{m}}(\widetilde{f}, F)\,, \tag{4.6}$$

где $\widetilde{g}(\mathbf{x})$ *и* $\widetilde{f}(\mathbf{x})$ *минимизируют* $\sigma^2_H(\widetilde{g}, G)$ *и* $\sigma^2_{\frac{H}{m}}(\widetilde{f}, F)$ *соответственно.*

4.3.3 Минимаксная ошибка интерполяции

Получим минимаксную ошибку интерполяции для разнородных данных также, как для случая однородных данных. Предположим, что настоящие спектральные плотности процессов $f(\cdot)$ и $g(\cdot)$ неизвестны, но процессы достаточно гладкие и принадлежат классам $\mathcal{F}(L_f) = \mathcal{F}(L_f, \mathbf{1})$ и $\mathcal{F}(L_g) = \mathcal{F}(L_g, \mathbf{1})$ соответственно. Для ясности ограничим изложение случаем $\boldsymbol{\lambda} = \mathbf{1} \in \mathbb{R}^d$ и $H = h\mathrm{I}$ для некоторого $h > 0$, где $\mathrm{I} \in \mathbb{R}^{d \times d}$. Результаты, приведенные ниже, выполнены в более общей постановке, описанной в разделе 4.2 и определенных там $\boldsymbol{\lambda} \in \mathbb{R}^d$ и H. Однако, изложение для такой более широкой постановки делает полученные выводы более громоздкими и не добавляет глубины полученным результатам.

Задача состоит в получении минимаксной ошибки интерполяции для $u(\mathbf{x})$. В частности хотим получить минимаксную ошибку интерполяции для разнородных данных:

$$R^{h,m}(L_f, L_g) = \inf_{f,g} \sup_{\substack{F \in \mathcal{F}(L_f), \\ G \in \mathcal{F}(L_g)}} \sigma^2_{h\mathrm{I},m}(\widetilde{u}, F, G, \rho)\,. \tag{4.7}$$

Теорема 4.4. *Минимаксная ошибка интерполяции* (4.7) *для модели* (4.4) *и наблюдений $u(\mathbf{x})$ в точках из D_H, а $f(\mathbf{x})$ — в точках из $D_{\frac{H}{m}}$, имеет вид*

$$R^{h,m}(L_f, L_g) = \rho^2 \frac{L_f}{2} \left(\frac{h}{m\pi}\right)^2 + \frac{L_g}{2} \left(\frac{h}{\pi}\right)^2 . \tag{4.8}$$

4.4 Оптимальное отношение между размерами выборок разнородных данных

Полученные результаты позволяют получить оптимальное отношение m размеров сеток для разнородных данных. Рассмотрим следующую постановку задачи: одно вычисление $u(\mathbf{x})$ стоит w, а одно вычисление $f(\mathbf{x})$ — только 1; общий бюджет вычислений равен числу точек в единичном гиперкубе $\frac{1}{h^d}$, умноженному на стоимость вычислений; бюджет ограничен сверху B.

Для такой постановки общий бюджет равен $B = \mathrm{w}\frac{1}{h^d} + \delta\frac{1}{h^d}$, где $\delta = m^d$

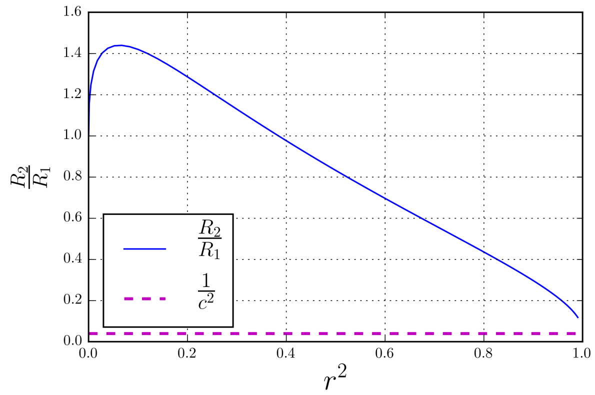


Рис. 4.3 – Зависимость отношения $\frac{R_2}{R_1}$ минимаксных ошибок интерполяции от коэффициента корреляции r для $L_f = 3, L_g = 1, \mathrm{w} = 5,\ d = 1$.

отношение размеров выборок разнородных данных.

Используя Теорему 4.4, получаем:

Теорема 4.5. *Минимум минимаксной ошибки интерполяции* (4.8) *для заданного B имеет вид*

$$\min_{\substack{h,\delta:\\ Bh^d = \mathrm{w}+\delta}} R^{h,m}(L_f, L_g) = \rho^2 \frac{L_f}{2} \left(\frac{\mathrm{w}+\delta^*}{\pi B \delta^*}\right)^{\frac{2}{d}} + \frac{L_g}{2} \left(\frac{\mathrm{w}+\delta^*}{\pi B}\right)^{\frac{2}{d}},$$

и оптимальное отношение размеров выборок $\delta^* = \left(\frac{L_f}{L_g}\mathrm{w}\rho^2\right)^{\frac{d}{d+2}}$.

Оптимальное отношение δ^* зависит от отношения стоимости вычислений w, коэффициента ρ и гладкостей L_f и L_g гауссовских процессов $f(\mathbf{x})$ и $g(\mathbf{x})$ соответственно.

Если использовать только источник данных высокой точности $u(\mathbf{x})$, то получаем следующую минимаксную ошибку интерполяции для заданного бюджета вычислений B:

$$\min_{h:\, Bh^d = \mathrm{w}} R^h(L_f, L_g) = \rho^2 \frac{L_f}{2} \left(\frac{\mathrm{w}}{\pi B}\right)^{\frac{2}{d}} + \frac{L_g}{2} \left(\frac{\mathrm{w}}{\pi B}\right)^{\frac{2}{d}}.$$

Отметим, что мы можем получить схожие результаты для заданной ковариационной функции, а не только в минимаксном случае, используя Теорему 4.3 и следствие 3 или 4.

4.4.1 Сравнение минимаксных ошибок интерполяции для различных значений параметров

Рассмотрим теперь при каких условиях и до какой степени использование разнородных источников данных уменьшает ошибку интерполяции в сравнении с использованием только источника точных данных для фиксированного общего бюджета вычислений. Обозначим как $R_2 = R^{h,\delta^*}(L_f, L_g, \rho)$ — минимаксную ошибку интерполяции, полученную при использовании разнородных источников данных, и как $R_1 = R^h(L_f, L_g, \rho)$ — минимаксную ошибку интерполяции, полученную при использовании данных только одной точности для такого же общего бюджета вычислений. Отношение $\frac{R_2}{R_1}$ характеризует насколько выгодно использовать разнородные источники данных: $\frac{R_2}{R_1} \geq 1$ означает, что никакого дополнительного преимущества нет, а $\frac{R_2}{R_1} < 1$ означает, что удалось уменьшить минимаксную ошибку интерполяции.

Отношение $\frac{R_2}{R_1}$ имеет вид:

$$\frac{R_2}{R_1} = \frac{\left(1 + \left(\frac{L_f^d \rho^{2d}}{L_g^d \mathrm{w}^2}\right)^{\frac{1}{d+2}}\right)^{\frac{d+2}{d}}}{1 + \rho^2 \frac{L_f}{L_g}}.$$

Положим $V_f = \mathbb{E} f^2(\mathbf{x})$ и $V_g = \mathbb{E} g^2(\mathbf{x})$. Тогда коэффициент корреляции r между $u(\mathbf{x})$ и $f(\mathbf{x})$ есть $r = \frac{1}{\sqrt{1 + \frac{V_g}{V_f}\frac{1}{\rho^2}}}$. Тогда выполнено

$$r \to 0: \frac{R_2}{R_1} \approx 1 + \frac{d+2}{d}\left(\frac{L_f V_f}{L_g V_g}\right)^{\frac{d}{d+2}} \frac{r^{\frac{2d}{d+2}}}{\mathrm{w}^{\frac{2}{d+2}}},$$
$$r \to 1: \frac{R_2}{R_1} \approx \frac{1}{\mathrm{w}^{\frac{2}{d}}} + \frac{2+d}{d}\left(\frac{L_g V_f}{L_f V_g}\right)^{\frac{d}{d+2}} \frac{(1-r^2)^{\frac{d}{d+2}}}{\mathrm{w}^{\frac{4}{d(d+2)}}}.$$

Если $r \to 0$, то использование разнородных источников данных не позволяет уменьшить ошибку, в то время как для $r \to 1$ отношение $\frac{R_2}{R_1}$ стремится к $\frac{1}{\mathrm{w}^{\frac{2}{d}}}$, и обычно $\mathrm{w} \gg 1$. Скорость сходимости к такому значению увеличивается при увеличении w и уменьшении $\frac{L_g}{L_f}$.

Рисунок 4.3 показывает как отношение $\frac{R_2}{R_1}$ зависит от $r = r(\rho)$ для $d = 1$.

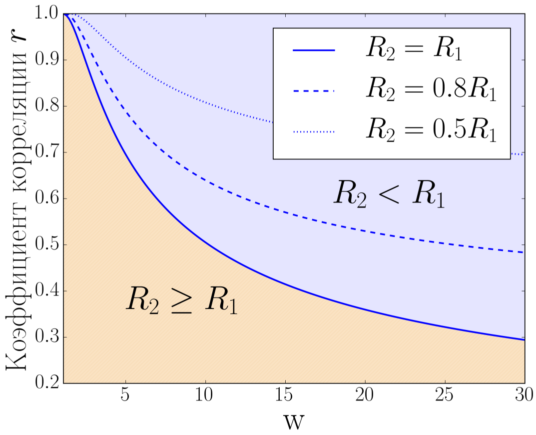


Рис. 4.4 – Кривые $R_2 = kR_1$ for $L_f = 2$, $L_g = 1$, $d = 1$. Если $R_2 < R_1$ имеет смысл использовать разнородные данные

Для малых r выполнено, что $R_2 > R_1$, в то время как для больших r значение $\frac{R_2}{R_1}$ сходится к $\frac{1}{\mathrm{w}^2}$, $\mathrm{w} \gg 1$.

Рисунок 4.4 показывает наименьшее значение r для $d = 1$, для которого при заданном $\mathrm{w} > 1$ выполнено $R_2 \leq R_1$.

Для $d > 1$ и $c \gg 1$ минимальное значение r, такое что $R_2 \leq R_1$ — порядка $\frac{1}{\sqrt{\mathrm{w}}}$:

$$r \approx \frac{1}{\sqrt{\mathrm{w}}} \left(\frac{V_f}{V_g}\right)^{\frac{1}{2}} \left(\frac{L_g}{L_f}\right)^{\frac{1}{2}}.$$

4.4.2 Алгоритма выбора оптимального соотношения размеров выборок разнородных данных

Если известны параметры ковариационной функции, то оценку параметров L_f и L_g легко получить, используя производные ковариационной функции $\frac{\partial^2 \mathrm{w}(\mathbf{x})}{\partial x_i \partial x_j}$ в точке $\mathbf{x} = \mathbf{0}$ по аргументам. Однако, для маленьких выборок оценка параметров ковариационных функций может быть достаточно плохой [111], так же достаточно сложно оценить L_f и L_g непосредственно по данным [56].

Таким образом, предположим, что $L_f = L_g$ и, используя Теорему 4.4, получим **Алгоритм 3**, который может быть использован для оценки оптимального отношения размеров выборок разнородных данных δ^* и создания плана экспериментов, для которого такое отношение имеет место. Преимущество предложенной техники состоит в том, что она может быть использована для создания планов экспериментов даже для подход, не основанных на методологии гауссовских процессов.; кроме того она практически не требует знаний о природе зависимости между источниками данных низкой и высокой точности, а именно, нужна лишь оценка коэффициента корреляции r.

Алгоритм 3 Создание планов экспериментов D_f и D_u для источников данных низкой и высокой точности соответственно

Вход: Корреляция r между источниками данных, бюджет B, стоимость использования источника данных высокой точности w (стоимость использования источника данных низкой точности 1)

1: $\rho^2 \leftarrow 1/(1 - \frac{1}{r^2})$
2: $\delta^* \leftarrow \left(\mathrm{w}\rho^2\right)^{\frac{d}{d+2}}$
3: $n_f \leftarrow \frac{B\delta^*}{\mathrm{w}+\delta^*}$
4: $n_u \leftarrow \frac{B}{\mathrm{w}+\delta^*}$
5: Сгенерировать план экспериментов D_f, D_u такой, что $D_u \subseteq D_f$, с $|D_f| = n_f$, $|D_u| = n_u$.
6: **Вернуть** D_f, D_u.

4.5 Доказательства

В этом разделе собраны доказательства утверждений, приведенных ранее в этой главе диссертационного исследования.

4.5.1 Доказательства для раздела 4.2.1

Доказательство теоремы 4.1. Выполнено, что

$$\mathbb{E}[f(\mathbf{x}) - \widetilde{f}(\mathbf{x})]^2 =$$
$$\int_{\mathbb{R}^d} F(\boldsymbol{\omega}) \left|1 - |H| \sum_{\mathbf{k}\in\mathbb{Z}^d} K(\mathbf{x} - \mathbf{x_k}) e^{-2\pi i \boldsymbol{\omega}^T(\mathbf{x_k} - \mathbf{x})}\right|^2 d\boldsymbol{\omega} =$$
$$\int_{\mathbb{R}^d} F(\boldsymbol{\omega}) \left|1 - |H| \sum_{\mathbf{k}\in\mathbb{Z}^d} \left(\int_{\mathbb{R}^d} \widehat{K}(\mathbf{u}) e^{-2\pi i \mathbf{u}^T(\mathbf{x} - \mathbf{x_k})} d\mathbf{u}\right) e^{-2\pi i \boldsymbol{\omega}^T(\mathbf{x_k} - \mathbf{x})}\right|^2 d\boldsymbol{\omega},$$

где $\widehat{K}(\mathbf{u})$ — преобразование Фурье $K(\mathbf{x})$. В соответствии с формулой суммирования Пуассона:

$$\sum_{\mathbf{k}\in\mathbb{Z}^d} \exp(2\pi \mathrm{i} \mathbf{k}^T \boldsymbol{\omega}) = \sum_{\mathbf{k}\in\mathbb{Z}^d} \delta(\boldsymbol{\omega} + \mathbf{k}),$$

где $\delta(\boldsymbol{\omega})$ — дельта функция Дирака, получаем

$$\mathbb{E}[f(\mathbf{x}) - \widetilde{f}(\mathbf{x})]^2 =$$

$$\int_{\mathbb{R}^d} F(\boldsymbol{\omega}) \left| 1 - |H| \sum_{\mathbf{k}\in\mathbb{Z}^d} \int_{\mathbb{R}^d} \widehat{K}(\mathbf{u}) \exp(2\pi \mathrm{i}(\boldsymbol{\omega} - \mathbf{u})^T \mathbf{x}) \delta(\mathbf{u} - \boldsymbol{\omega} + H^{-1}\mathbf{k}) d\mathbf{u} \right|^2 d\boldsymbol{\omega} =$$

$$= \int_{\mathbb{R}^d} \left| 1 - \sum_{\mathbf{k}\in\mathbb{R}^d} \widehat{K}(\boldsymbol{\omega} - H^{-1}\mathbf{k}) \exp(2\pi \mathrm{i} H^{-1} \mathbf{x}^T \mathbf{k}) \right|^2 d\boldsymbol{\omega}.$$

Используя ортогональность системы функций $\exp(2\pi \mathrm{i} H^{-1} \mathbf{x}^T \mathbf{k})$ on $\mathbf{x} \in [0, h_1] \times \ldots \times [0, h_d]$ интегрируем равенство, чтобы получить:

$$\sigma_H^2(\widetilde{f}, F) = \int_{\mathbb{R}^d} F(\boldsymbol{\omega}) \left| [1 - \widehat{K}(\boldsymbol{\omega})]^2 + \sum_{\mathbf{k}\in\mathbb{Z}^d\setminus\{\mathbf{0}\}} \widehat{K}^2(\boldsymbol{\omega} + H^{-1}\mathbf{k}) \right|^2 d\boldsymbol{\omega}.$$

Для того чтобы получить $\widehat{K}(\boldsymbol{\omega})$, которая минимизирует ошибку интерполяции, перепишем ее как

$$\sigma_H^2(\widetilde{f}, F) = \int_{\mathbb{R}^d} \left| [1 - \widehat{K}(\boldsymbol{\omega})]^2 F(\boldsymbol{\omega}) + \widehat{K}(\boldsymbol{\omega})^2 \sum_{\mathbf{k}\in\mathbb{Z}^d\setminus\{\mathbf{0}\}} \widehat{F}(\boldsymbol{\omega} + H^{-1}\mathbf{k}) \right|^2 d\boldsymbol{\omega}.$$

Решая квадратичную задачу оптимизации для каждого $\boldsymbol{\omega}$, получаем:

$$\widehat{K}(\boldsymbol{\omega}) = \frac{\widehat{F}(\boldsymbol{\omega})}{\sum_{\mathbf{k}\in\mathbb{Z}^d} \widehat{F}(\boldsymbol{\omega} + H^{-1}\mathbf{k})}.$$

Тогда

$$\sigma_H^2(\widetilde{f}, F) = \int_{\mathbb{R}^d} F(\boldsymbol{\omega}) \frac{\sum_{\mathbf{k}\in\mathbb{Z}^d\setminus\{\mathbf{0}\}} \widehat{F}(\boldsymbol{\omega} + H^{-1}\mathbf{k})}{\sum_{\mathbf{k}\in\mathbb{Z}^d} \widehat{F}(\boldsymbol{\omega} + H^{-1}\mathbf{k})} d\boldsymbol{\omega}. \tag{4.9}$$

□

Доказательство следствия 1. Выполнено, что наилучшая интерполяция имеет вид

$$\widetilde{f}(\mathbf{x}) = \mu(\Omega_H) \sum_{\mathbf{k}\in\mathbb{Z}^d} \phi(\mathbf{x}, \mathbf{x_k}) f(\mathbf{x_k})$$

для некоторых $\phi(\mathbf{x}, \mathbf{x}')$. Записывая уравнения Колмогорова-Винера-Хопфа для ковариационной функции $c(\mathbf{x})$, получаем

$$\sum_{\mathbf{k}\in\mathbb{Z}^d} \phi(\mathbf{x}, \mathbf{x_k}) c(\mathbf{x_k} - \mathbf{x_m}) = c(\mathbf{x} - \mathbf{x_m}) \tag{4.10}$$

для всех $\mathbf{m} \in \mathbb{Z}^d$. Докажем, что $\phi(\mathbf{x}, \mathbf{x_k}) = \phi(\mathbf{x} - \mathbf{x_k})$.

Рассмотрим две суммы из (4.10):

$$\sum_{\mathbf{k}\in\mathbb{Z}^d} \phi(\mathbf{x}, \mathbf{x_k}) c(\mathbf{x_k} - \mathbf{x_m}) = c(\mathbf{x} - \mathbf{x_m}),$$

$$\sum_{\mathbf{k}\in\mathbb{Z}^d} \phi(\mathbf{x} - \mathbf{x_s}, \mathbf{x_k}) c(\mathbf{x_k} - \mathbf{x_{m-s}}) = c(\mathbf{x} - \mathbf{x_s} - \mathbf{x_{m-s}}).$$

В силу $\mathbf{x_{m-s}} = H\mathbf{m} - H\mathbf{s} = \mathbf{x_m} - \mathbf{x_s}$ выполнено

$$\sum_{\mathbf{k}\in\mathbb{Z}^d} \phi(\mathbf{x}, \mathbf{x_k}) c(\mathbf{x_k} - \mathbf{x_m}) = c(\mathbf{x} - \mathbf{x_m}) = c(\mathbf{x} - \mathbf{x_s} - \mathbf{x_{m-s}}) =$$
$$\sum_{\mathbf{k}\in\mathbb{Z}^d} \phi(\mathbf{x} - \mathbf{x_s}, \mathbf{x_k}) c(\mathbf{x_k} - \mathbf{x_{m-s}}).$$

Следовательно,

$$\sum_{\mathbf{k}\in\mathbb{Z}^d} \left[\phi(\mathbf{x}, \mathbf{x_k}) - \phi(\mathbf{x} - \mathbf{x_s}, \mathbf{x_k} - \mathbf{x_s})\right] c(\mathbf{x_k} - \mathbf{x_m}) = 0.$$

Положительная определенность ковариационной функции $c(\mathbf{x})$ дает

$$\phi(\mathbf{x}, \mathbf{x_k}) = \phi(\mathbf{x} - \mathbf{x_s}, \mathbf{x_k} - \mathbf{x_s}).$$

Для $\mathbf{x_s} = \mathbf{x_k}$ получаем

$$\phi(\mathbf{x}, \mathbf{x_k}) = \phi(\mathbf{x} - \mathbf{x_k}, \mathbf{0}) = K(\mathbf{x} - \mathbf{x_k}).$$

Используя формулу суммирования Пуассона,

$$\frac{1}{\mu(\Omega_H)}\Phi(\boldsymbol{\omega})\sum_{\mathbf{k}\in\mathbb{Z}^d}F(\boldsymbol{\omega}-\boldsymbol{\omega}_{\mathbf{k}})=F(\boldsymbol{\omega}),$$

где $\Phi(\boldsymbol{\omega})$ — преобразование Фурье $\phi(\mathbf{x})$. Тогда

$$\Phi(\boldsymbol{\omega})=\frac{\mu(\Omega_H)F(\boldsymbol{\omega})}{\sum_{\mathbf{k}\in\mathbb{Z}^d}F(\boldsymbol{\omega}-\boldsymbol{\omega}_{\mathbf{k}})}.$$

Таким образом, оптимальная интерполяция имеет вид:

$$\widetilde{f}(\mathbf{x})=\mu(\Omega_H)\sum_{\mathbf{k}\in\mathbb{Z}^d}K(\mathbf{x}-\mathbf{x}_{\mathbf{k}})f(\mathbf{x}_{\mathbf{k}}).$$

И

$$\widehat{K}(\boldsymbol{\omega})=\frac{\Phi(\boldsymbol{\omega})}{\mu(\Omega_H)}.$$

□

Доказательство следствия 3. Получим ошибку интерполяции для экспоненциальной ковариационной функции вида $c(x)=\sqrt{\frac{\pi}{2}}\exp\left(-\theta|x|\right)$ для $x\in\mathbb{R}$. Спектральная плотность для такой ковариационной функции есть $F(\omega)=\frac{\theta}{\theta^2+\omega^2}$.

Подсчитаем ошибку интерполяции

$$\sigma_h^2\left(\widetilde{f},F\right)=\int_{-\infty}^{\infty}F(\omega)\frac{\sum_{k\neq 0}F(\omega+\frac{k}{h})}{\sum_k F(\omega+\frac{k}{h})}d\omega.$$

Выполнено, что

$$\sum_{k=-\infty}^{\infty}F(\omega+\frac{k}{h})=\sum_{k=-\infty}^{\infty}\frac{\theta}{(\omega+\frac{k}{h})^2+\theta^2}=h\sum_{k=-\infty}^{\infty}\frac{h\theta}{(h\omega+k)^2+h^2\theta^2}=$$
$$=\pi h\coth(\pi\theta h)\frac{1}{1+\sin^2(\pi h\omega)(\coth^2(\pi\theta h)-1)}.$$

Тогда

$$\int_{-\infty}^{\infty}F(\omega)\frac{\sum_{k\neq 0}F(\omega+\frac{k}{h})}{\sum_k F(\omega+\frac{k}{h})}d\omega=$$

$$\int_{-\infty}^{\infty} \frac{\theta}{\theta^2+\omega^2}\left(1-\frac{\theta}{\theta^2+\omega^2}\frac{1+\sin^2(\pi h\omega)(\coth^2(\pi\theta h)-1)}{\pi h\coth(\pi\theta h)}\right)d\omega.$$

Возьмем интегралы аналитически. А именно,

$$\int_{-\infty}^{\infty} \frac{\theta}{\theta^2+\omega^2}d\omega = \pi.$$

Также

$$\int_{-\infty}^{\infty} \frac{\theta^2}{(\theta^2+\omega^2)^2}d\omega = \frac{\pi}{2\theta}.$$

Наконец

$$\int_{-\infty}^{\infty} \frac{\theta^2}{(\theta^2+\omega^2)^2}\sin^2(\pi\omega h)d\omega =$$
$$-\frac{\pi^2 h}{2}\left(\cosh(\pi\theta h)-\sinh(\pi\theta h)\right)\left(\cosh(\pi\theta h)-\left(\frac{1}{\pi\theta h}+1\right)\sinh(\pi\theta h)\right).$$

Следовательно,

$$\int_{-\infty}^{\infty} \frac{\theta}{\theta^2+\omega^2}\left(1-\frac{\theta}{\theta^2+\omega^2}\frac{1+\sin^2(\pi h\omega)(\coth^2(\pi\theta h)-1)}{\pi h\coth(\pi\theta h)}\right)d\omega =$$
$$\pi-\frac{\pi}{2\pi\theta h\coth(\pi\theta h)}+$$
$$+\frac{\pi^2}{2}\exp(-\pi\theta h)\left(\exp(-\pi\theta h)-\frac{1}{\pi\theta h}\sinh(\pi\theta h)\right)\frac{\coth^2(\pi\theta h)-1}{\coth(\pi\theta h)}.$$

Для $h\to 0$ раскладывая по Тейлору получаем:

$$\sigma_h^2(\widetilde{f},F) = \frac{2\pi^2}{3}\theta h + O((\theta h)^2).$$

$\square$

Доказательство следствия 4. Заметим, что ошибка интерполяции имеет вид

$$\sigma_h^2(\widetilde{f},F) = \int_{-\infty}^{\infty} F(\omega)\frac{\sum_{k\neq 0}F(\omega+\frac{k}{h})}{\sum_s F(\omega+\frac{s}{h})}d\omega.$$

Получим верхнюю и нижнюю оценку для такого выражения. Обозначим $v=\frac{1}{h}$.

Получим верхнюю оценку, разбив область интегрирования $(-\infty,\infty)$ на три части $(-\infty,-v/2]$, $(-v/2,v/2]$, $(v/2,+\infty)$ и получив верхнюю оценку для каждой части.

Заметим, что

$$0 \leq \frac{\sum_{k \neq 0} F(\omega + kv)}{\sum_s F(\omega + sv)} \leq 1.$$

Следовательно, используя оценки типа Чернова [24], мы получаем

$$\int_{v/2}^{\infty} F(\omega) \frac{\sum_{k \neq 0} F(\omega + kv)}{\sum_s F(\omega + sv)} d\omega \leq \int_{v/2}^{\infty} F(\omega) d\omega = \\ = \int_{v/2}^{\infty} \frac{1}{\sqrt{\theta}} \exp\left(-\frac{\omega^2}{2\theta}\right) d\omega \leq \sqrt{2} \exp\left(-\frac{v^2}{8\theta}\right). \tag{4.11}$$

Аналогично для $(-\infty, -v/2)$.

Теперь получим верхнюю оценку для $(-v/2, v/2)$. Получим сперва нижнюю и верхнюю оценку для суммы ряда $\sum_{s \neq 0} F(\omega + sv)$. Спектральная плотность для квадратичной экспоненциальной плотности убывает на $[0, +\infty)$ относительно ω. Таким образом,

$$\int_{\Delta + u}^{+\infty} F(x) dx \leq \sum_{s=1}^{\infty} \Delta F(\Delta s + u) \leq \Delta F(s + u) + \int_{\Delta + u}^{+\infty} F(x) dx.$$

Используя [1], формула 7.1.13, получаем для ω таких, что $|\omega| \leq \frac{v}{2}$:

$$\frac{4\sqrt{\theta}}{v + \sqrt{v^2 + 16\theta}} \exp\left(-\frac{v^2}{8\theta}\right) \leq \int_{\frac{v}{2}}^{\infty} F(\omega) d\omega \leq \frac{4\sqrt{\theta}}{v + \sqrt{v^2 + \frac{32}{\pi}\theta}} \exp\left(-\frac{v^2}{8\theta}\right).$$

Так же

$$v \sum_{k \in \mathbb{Z}^+} F(\omega + kv) \leq vF(\omega + v) + \int_{\frac{v}{2}}^{\infty} F(\omega) d\omega \leq \\ \frac{v}{\sqrt{\theta}} \exp\left(-\frac{v^2}{8\theta}\right) + \frac{4\sqrt{\theta}}{v + \sqrt{v^2 + \frac{32}{\pi}\theta}} \exp\left(-\frac{v^2}{8\theta}\right).$$

Теперь получим искомую верхнюю оценку для интеграла по интервалу $(-v/2, v/2)$ для достаточно больших v:

$$\int_{-v/2}^{v/2} F(\omega) \frac{\sum_{k \neq 0} F(\omega + kv)}{\sum_s F(\omega + sv)} d\omega \leq$$

$$\leq \int_{-v/2}^{v/2} F(\omega) \frac{F(\omega+v)+F(\omega-v)+\frac{4\sqrt{\theta}}{v\left(v+\sqrt{v^2+\frac{32}{\pi}\theta}\right)}\exp\left(-\frac{v^2}{8\theta}\right)}{F(\omega)+F(\omega+v)+F(\omega-v)+\frac{4\sqrt{\theta}}{v\left(v+\sqrt{v^2+\frac{32}{\pi}\theta}\right)}\exp\left(-\frac{v^2}{8\theta}\right)} d\omega \leq$$

$$\leq \int_{-v/2}^{v/2} F(\omega) \frac{F(\omega+v)+F(\omega-v)}{F(\omega)+F(\omega+v)+F(\omega-v)} d\omega +$$

$$+ \int_{-v/2}^{v/2} F(\omega) \frac{\frac{4\sqrt{\theta}}{v\left(v+\sqrt{v^2+\frac{32}{\pi}\theta}\right)}\exp\left(-\frac{v^2}{8\theta}\right)}{F(\omega)+\frac{4\sqrt{\theta}}{v\left(v+\sqrt{v^2+\frac{32}{\pi}\theta}\right)}\exp\left(-\frac{v^2}{8\theta}\right)} d\omega \leq$$

$$\leq \int_{-v/2}^{v/2} F(\omega+v)+F(\omega-v) d\omega + \frac{4\sqrt{\theta}}{v+\sqrt{v^2+\frac{32}{\pi}\theta}}\exp\left(-\frac{v^2}{8\theta}\right) \leq$$

$$\leq \frac{12\sqrt{\theta}}{v+\sqrt{v^2+\frac{32}{\pi}\theta}}\exp\left(-\frac{v^2}{8\theta}\right) \leq \frac{7\sqrt{\theta}}{v}\exp\left(-\frac{v^2}{8\theta}\right).$$

Выполнено, что

$$\frac{\sum_{k\neq 0} F(\omega+kv)}{\sum_s F(\omega+sv)} \geq \frac{F(\omega+v)+F(\omega-v)}{F(\omega)+F(\omega+v)+F(\omega-v)}.$$

Для ω таких что $|\omega| \leq \frac{v}{2}$ имеет место:

$$1+\frac{F(\omega+v)}{F(\omega)}+\frac{F(\omega-v)}{F(\omega)} \leq 3.$$

Тогда для достаточно больших v выполнена следующая нижняя оценка:

$$\int_{-\infty}^{\infty} F(\omega)\frac{\sum_{k\neq 0} F(\omega+kv)}{\sum_s F(\omega+sv)} d\omega \geq \int_{-v/2}^{v/2} F(\omega)\frac{\sum_{k\neq 0} F(\omega+kv)}{\sum_s F(\omega+sv)} d\omega \geq$$

$$\geq \int_{-\frac{v}{2}}^{\frac{v}{2}} F(\omega)\frac{F(\omega+v)+F(\omega-v)}{F(\omega)+F(\omega+v)+F(\omega-v)} d\omega \geq \int_{-\frac{v}{2}}^{\frac{v}{2}} \frac{F(\omega+v)+F(\omega-v)}{3} d\omega =$$

$$= \frac{2}{3}\int_{\frac{v}{2}}^{\frac{3v}{2}} F(\omega) d\omega \geq\geq \frac{4}{3}\frac{\sqrt{\theta}}{v}\exp\left(-\frac{v^2}{8\theta}\right).$$

$\square$

4.5.2 Доказательства для раздела 4.2.2

Для доказательства основного результата потребуется следующая лемма

Лемма 16. *Пусть $c \geq 0$ и $\omega \geq 0$ такое что $c^2 + \omega^2 \leq 1$, $c^2 + (1 - \omega^2) \leq 1$. Тогда*

$$\left(1 - \sqrt{c^2 + \omega^2}\right)^2 + \left(1 - \sqrt{c^2 + (1 - \omega)^2}\right)^2 \leq \left(1 - \sqrt{c^2}\right)^2 = (1 - c)^2. \tag{4.12}$$

Доказательство. Приведем сперва схему доказательства. Докажем, что для ω, которое максимизирует левую часть неравенства (4.12), это неравенство выполнено. Для этого докажем, что для допустимых $\omega \in [1 - \sqrt{1 - c^2}, \frac{1}{2}]$ производная левой части по ω меньше нуля для всех допустимых c, так что $\omega = 1 - \sqrt{1 - c^2}$ доставляет максимум левой части. Затем покажем, что для такого ω неравенство выполняется.

Частные производные левой и правой части относительно ω равны

$$\begin{aligned} g(\omega, c) &= \frac{\partial}{\partial \omega}\left(\left(1 - \sqrt{c^2 + \omega^2}\right)^2 + \left(1 - \sqrt{c^2 + (1 - \omega)^2}\right)^2\right) = \\ &= -2\frac{\left(1 - \sqrt{c^2 + \omega^2}\right)\omega}{\sqrt{c^2 + \omega^2}} + 2\frac{\left(1 - \sqrt{c^2 + (1 - \omega)^2}\right)(1 - \omega)}{\sqrt{c^2 + (1 - \omega)^2}} = \\ &= -2\left(\frac{1}{\sqrt{c^2 + \omega^2}} - 1\right)\omega + 2\left(\frac{1}{\sqrt{c^2 + (1 - \omega)^2}} - 1\right)(1 - \omega). \end{aligned}$$

Если $\omega = \frac{1}{2}$, частная производная равна нулю. Покажем, что для таких $\omega < \frac{1}{2}$, что $c^2 + \omega^2 < 1$, $c^2 + (1 - \omega)^2 < 1$, выполнено $g(\omega, c) < 0$. Это будет означать, что исходная функция убывает для $\omega \in [1 - \sqrt{1 - c^2}, \frac{1}{2}]$.

Промаксимизируем $g(\omega, c)$ по c. Функция $g(\omega, c)$ достигает максимума на краю области допустимости или в локальном оптимуме по c. Найдем локальный оптимум по c, то есть такое c, что частная производная $g(\omega, c)$ по c равна нулю:

$$\frac{c(1 - \omega)}{((1 - \omega)^2 + c^2)^{\frac{3}{2}}} - \frac{c\omega}{(\omega^2 + c^2)^{\frac{3}{2}}} = 0.$$

Следовательно,

$$\frac{1 - \omega}{\omega} = \frac{((1 - \omega)^2 + c^2)^{\frac{3}{2}}}{(\omega^2 + c^2)^{\frac{3}{2}}}. \tag{4.13}$$

И

$$c^2 = \omega^{\frac{2}{3}}(1 - \omega)^{\frac{2}{3}}(\omega^{\frac{2}{3}} + (1 - \omega)^{\frac{2}{3}}).$$

Покажем, что это локальный максимум. А именно, докажем, что вторая производная $g(\omega, c)$ по c меньше 0:

$$-\frac{(1-\omega)}{((1-\omega)^2+c^2)^{\frac{3}{2}}}+\frac{\omega}{(\omega^2+c^2)^{\frac{3}{2}}}+\frac{3c^2(1-\omega)}{((1-\omega)^2+c^2)^{\frac{5}{2}}}-\frac{3c^2\omega}{(\omega^2+c^2)^{\frac{5}{2}}}\leq 0.$$

Или:

$$\frac{\omega}{(\omega^2+c^2)^{\frac{5}{2}}}((\omega^2+c^2)-3c^2)-\frac{(1-\omega)}{((1-\omega)^2+c^2)^{\frac{5}{2}}}(((1-\omega)^2+c^2)-3c^2)\leq 0.$$

В локальном оптимуме выполнено (4.13), следовательно:

$$\frac{(1-\omega)}{(\omega^2+c^2)((1-\omega)^2+c^2)^{\frac{3}{2}}}(\omega^2-2c^2)-\frac{(1-\omega)}{((1-\omega)^2+c^2)^{\frac{5}{2}}}((1-\omega)^2-2c^2)\leq 0.$$

Тогда,

$$\frac{(1-\omega)}{((1-\omega)^2+c^2)^{\frac{5}{2}}(\omega^2+c^2)}\Big(((1-\omega)^2+c^2)(\omega^2-2c^2)-(\omega^2+c^2)((1-\omega)^2-2c^2)\Big)\leq 0.$$

Из-за ограничения на значения ω это неравенство эквивалентно:

$$((1-\omega)^2+c^2)(\omega^2-2c^2)-(\omega^2+c^2)((1-\omega)^2-2c^2)\leq 0,$$

или

$$2c^2\omega^2-c^2(1-\omega)^2-2c^2(1-\omega)^2+c^2\omega^2\leq 0.$$

Это неравенство выполнено, так как $\omega\leq\frac{1}{2}$ и $(1-\omega)^2\geq\omega^2$.

Таким образом, локальный экстремум есть локальный минимум, и функция достигает максимальных значений на границе допустимой области. А именно, $c^2=1-(1-\omega)^2$ или $c^2=0$ дают максимальные значения.

Для таких значений c производная меньше нуля. Используя $c^2=1-(1-\omega)^2$, получаем

$$-2\left(\frac{1}{\sqrt{1-(1-\omega)^2+\omega^2}}-1\right)\omega\leq 0.$$

Аналогично для $c^2=0$

$$-2(1-\omega)+2\omega=4\omega-2\leq 0.$$

Следовательно, целевая функция убывает по ω на $[1-\sqrt{1-c^2}, \frac{1}{2}]$, и $\omega = \frac{1}{2}$ доставляет локальный минимум. Таким образом, максимум левой части — в $\omega = 1-\sqrt{1-c^2}$. Легко видеть, что в таком случае левая часть неравенства (4.12) всегда меньше правой. $\square$

Докажем теперь основную теорему.

Доказательство теоремы 4.2. Докажем верхние и нижние оценки для $R^H(L, \boldsymbol{\lambda})$ равные $\frac{L}{2\pi^2} \max_{i \in \{1,\dots,d\}} \left(\frac{h_i}{\lambda_i}\right)^2$. Сначала докажем нижнюю оценку, потом верхнюю.

Рассмотрим функционал

$$\varPhi(F, \widehat{K}) = \int_{\mathbb{R}^d} F(\boldsymbol{\omega}) \left[(1-\widehat{K}(\boldsymbol{\omega}))^2 + \sum_{\mathbf{x} \in D_{H-1} \setminus \{\mathbf{0}\}} \widehat{K}^2(\boldsymbol{\omega} + \mathbf{x})\right] d\boldsymbol{\omega},$$

который равен ошибке интерполяции $\sigma_H^2(\widetilde{f}, F)$ для

$$\widetilde{f}(\mathbf{x}) = \mu(\varOmega_H) \sum_{\mathbf{x}' \in D_H} K(\mathbf{x} - \mathbf{x}') f(\mathbf{x}'),$$

такого что $\widehat{K}(\boldsymbol{\omega})$ — преобразования Фурье $K(\mathbf{x})$.

Функционал линеен по $F(\boldsymbol{\omega})$ и квадратичен по $\widehat{K}(\boldsymbol{\omega})$, а нам нужна седловая точка $R^H(L, \boldsymbol{\lambda})$, такая что:

$$R^H(L, \boldsymbol{\lambda}) = \inf_{\widetilde{f}} \sup_{F \in \mathcal{F}(L, \boldsymbol{\lambda})} \sigma_H^2(\widetilde{f}, F) = \sup_{F \in \mathcal{F}(L, \boldsymbol{\lambda})} \inf_{\widetilde{f}} \sigma_H^2(\widetilde{f}, F).$$

Выполнено, что (4.9)

$$\min_{\widehat{K}} \varPhi(F, \widehat{K}) = \int_{\mathbb{R}^d} F(\boldsymbol{\omega}) \frac{\sum_{\mathbf{x} \in D_{H-1} \setminus \{\mathbf{0}\}} F(\boldsymbol{\omega} + \mathbf{x})}{\sum_{\mathbf{x} \in D_{H-1}} F(\boldsymbol{\omega} + \mathbf{x})} d\boldsymbol{\omega}.$$

Рассмотрим класс спектральных плотностей $F_\varepsilon(\boldsymbol{\omega})$:

$$F_\varepsilon(\boldsymbol{\omega}) = \begin{cases} \frac{A_\varepsilon}{(2\varepsilon)^d}, & \exists \mathbf{s} \in U_h : \|\boldsymbol{\omega} - \mathbf{s}\|_\infty \le \varepsilon, \\ 0, & \text{иначе}, \end{cases}$$

здесь $U_h = \left\{ \left(0, 0, \dots, \frac{1}{2h_j}, \dots, 0\right), \left(0, 0, \dots, -\frac{1}{2h_j}, \dots, 0\right) \right\}$, и индекс jтакой, что

$$j = \arg \max_{i \in \{1, \dots, d\}} \left(\frac{h_i}{\lambda_i} \right)^2 .$$

В силу (4.3)

$$(2\pi)^2 \int_{\mathbb{R}^d} F(\boldsymbol{\omega}) \sum_{i=1}^{d} \lambda_i^2 \omega_i^2 d\boldsymbol{\omega} \le L,$$

и для $\varepsilon \to 0$

$$A_\varepsilon \to \frac{L}{2\pi^2} \left(\frac{h_j}{\lambda_j} \right)^2 .$$

Действительно, для $\varepsilon \to 0$:

$$(2\pi)^2 \int_{\mathbb{R}^d} F(\boldsymbol{\omega}) \sum_{i=1}^{d} \lambda_i^2 \omega_i^2 d\boldsymbol{\omega} \to 2(2\pi)^2 \frac{A_\varepsilon}{(2\varepsilon)^d} (2\varepsilon)^d \left(\frac{\lambda_j}{h_j} \right)^2 = 2A_\varepsilon \left(\frac{\pi \lambda_j}{h_j} \right)^2 = L.$$

Для $\varepsilon \to 0$ выполнено что

$$\min_{\widehat{K}} \mathit{\Phi}(F_\varepsilon, \widehat{K}) \to 2 \frac{1}{2} \frac{A_\varepsilon}{(2\varepsilon)^d} (2\varepsilon)^d = A_\varepsilon = \frac{L}{2\pi^2} \left(\frac{h_j}{\lambda_j} \right)^2 .$$

Следовательно, получили нижнюю оценку, равную $\frac{L}{2\pi^2} \left(\frac{h_j}{w_j} \right)^2$.

Получим теперь верхнюю оценку.

Для произвольного $\widehat{K}(\boldsymbol{\omega})$ выполнено, что

$$\begin{aligned} & R^H(L, \boldsymbol{\lambda}) \le \max_{F \in \mathcal{F}(L, \boldsymbol{\lambda})} \mathit{\Phi}(F, \widehat{K}) \le \\ & \le L \left(\frac{1}{2\pi} \right)^2 \max_{\boldsymbol{\omega}} \left\{ \frac{1}{\sum_{i=1}^{d} \lambda_i^2 \omega_i^2} \left[(1 - \widehat{K}(\boldsymbol{\omega}))^2 + \sum_{\mathbf{x} \in D_{H-1} \setminus \{\mathbf{0}\}} \widehat{K}^2(\boldsymbol{\omega} + \mathbf{x}) \right] \right\} . \end{aligned}$$

Рассмотрим теперь

$$\widehat{K}(\boldsymbol{\omega}) = \begin{cases} 1 - \|\boldsymbol{\omega}\|, \|\boldsymbol{\omega}\|^2 \le 1, \\ 0, \text{иначе.} \end{cases}$$

Теперь докажем, что для такого $\widehat{K}(\boldsymbol{\omega})$ выполнено

$$\left[(1-\widehat{K}(\boldsymbol{\omega}))^2 + \sum_{\mathbf{x}\in\mathbb{Z}^d\setminus\{\mathbf{0}\}} \widehat{K}^2(\boldsymbol{\omega}+\mathbf{x})\right] \leq 2\|\boldsymbol{\omega}\|^2. \tag{4.14}$$

Выполнено $(1-\widehat{K}(\boldsymbol{\omega}))^2 \leq \|\boldsymbol{\omega}\|^2$. Осталось показать, что

$$\sum_{\mathbf{x}\in\mathbb{Z}^d\setminus\{\mathbf{0}\}} \widehat{K}^2(\boldsymbol{\omega}+\mathbf{x}) \leq \|\boldsymbol{\omega}\|^2. \tag{4.15}$$

Докажем, что для $\|\boldsymbol{\omega}\|_\infty < 1$ и $c^2 \geq 0$:

$$\sum_{\substack{\mathbf{x}\in\mathbb{Z}^d\setminus\{\mathbf{0}\},\\ \|\boldsymbol{\omega}+\mathbf{x}\|^2+c^2\leq 1}} \left(1-\sqrt{c^2+\sum_{i=1}^d(\omega_i+x_i)^2}\right)^2 \leq \sum_{\substack{i\in\{1,\cdots,d\},\\ c^2+(1-\omega_i)^2\leq 1}} \left(1-\sqrt{c^2+(1-\omega_i)^2}\right)^2.$$

Будем проводить индукцию по d. Для $d=1$ левая часть совпадает с правой. Пусть теперь для $(d-1)$ утверждение индукции выполнено. Покажем, что оно выполняется и для d.

Для $\boldsymbol{\omega} = (\omega_1, \omega_2, \ldots, \omega_d)$ такого что $\|\boldsymbol{\omega}\|_\infty < 1$, i-ая компонента вектора $\boldsymbol{\omega}+\mathbf{x}, \mathbf{x}\in\mathbb{Z}^d\setminus\{\mathbf{0}\}$ такого что $\|\boldsymbol{\omega}+\mathbf{x}\| \leq 1$ либо ω_i, либо $(1-\omega_i)$. Следовательно, все такие $\boldsymbol{\omega}+\mathbf{x}$ имеют вид $\mathbf{s}\boldsymbol{\omega}+(\mathbf{1}-\mathbf{s})(\mathbf{1}-\boldsymbol{\omega})$, где $\mathbf{s}$ — вектор, компоненты которого либо ноль, либо один.

Выполнено, что $(1-\sqrt{c^2+(1-\omega_1)^2+\omega_2^2+\ldots+\omega_d^2})^2 \leq (1-\sqrt{c^2+(1-\omega_1)^2})^2$, if $c^2+(1-\omega_1)^2+\omega_2^2+\ldots+\omega_d^2 \leq 1$.

Теперь рассмотрим все члены вида $(1-\sqrt{c^2+(1-\omega_1)^2+\ldots})^2$, для которых существует $j\neq 1$, такой что $(1-\omega_j)^2$ находится в сумме под квадратным корнем. В силу утверждения индукции сумма таких членов ограничена сверху следующей суммой:

$$\sum_{\substack{i\in\{2,\ldots,d\},\\ c^2+(1-\omega_1)^2+(1-\omega_i)^2\leq 1}} (1-\sqrt{c^2+(1-\omega_1)^2+(1-\omega_i)^2})^2.$$

Аналогично получим оценку сверху для суммы членов $(1-\sqrt{c^2+\omega_1^2+\ldots})^2$ с

$(1-\omega_j)^2$ в сумме под корнем:

$$\sum_{\substack{i\in\{2,\dots,d\},\\ c^2+\omega_1^2+(1-\omega_i)^2\le 1}} (1-\sqrt{c^2+\omega_1^2+(1-\omega_i)^2})^2.$$

Используя Лемму 16 для пары слагаемых $(1-\sqrt{c^2+(1-\omega_1)^2+(1-\omega_i)^2})^2+(1-\sqrt{c^2+\omega_1^2+(1-\omega_i)^2})^2$ получаем:

$$(1-\sqrt{c^2+(1-\omega_1)^2+(1-\omega_i)^2})^2+(1-\sqrt{c^2+\omega_1^2+(1-\omega_i)^2})^2\le$$
$$\le (1-\sqrt{c^2+(1-\omega_i)^2})^2.$$

Эта верхняя оценка так же выполнена, если есть только один член или нет ни одного для i-го индекса. Следовательно, выполнено утверждение индукции: целевая сумма ограничена сверху $\sum_{i=1}^d (1-\sqrt{c^2+(1-\omega_i)^2})^2$.

Подставляя $c^2=0$, получаем (4.15).

Рассмотрим теперь случай $\|\boldsymbol{\omega}\|_\infty \ge 1$. А именно $\boldsymbol{\omega}=\{\widehat{\omega}_1+1,\omega_2,\dots,\omega_d\}$, причем $\|(\widehat{\omega}_1,\omega_2,\dots,\omega_d)\|_\infty<1$, и $\widehat{\omega}_1\ge 0$, $\omega_i\ge 0, i=\overline{2,d}$. Тогда $\|\boldsymbol{\omega}\|^2=1+2\widehat{\omega}_1+\widehat{\omega}_1^2+\sum_{i=2}^d \omega_i^2$. Для вектора $(\widehat{\omega}_1,\omega_2,\dots,\omega_d)$ выполнено (4.15). Для исходного вектора $\boldsymbol{\omega}$ появляется дополнительное слагаемое $\widehat{K}^2((\widehat{\omega}_1,\omega_2,\dots,\omega_d))$, если Евклидова норма такого вектора меньше либо равна 1 — но это новое слагаемое меньше либо равно 1, так как иначе в сумме такого слагаемого нет. Таким образом, выполнена целевая оценка для $\boldsymbol{\omega}$. Остальные случаи для $\|\boldsymbol{\omega}\|_\infty>1$ рассматриваются так же. Следовательно, для всех $\boldsymbol{\omega}$ оценка (4.14) имеет место.

Выполнено что

$$\max_{\boldsymbol{\omega}} \frac{\sum_{i=1}^d \omega_i^2}{\sum_{i=1}^d \left(\frac{\lambda_i}{h_i}\right)^2 \omega_i^2} = \max_{i\in\{1,\dots,d\}} \left(\frac{h_i}{\lambda_i}\right)^2.$$

Следовательно, выполнена следующая верхняя оценка минимаксной ошибки интерполяции

$$R^H(L,\boldsymbol{\lambda}) \le \frac{L}{2\pi^2} \max_{i\in\{1,\dots,d\}} \left(\frac{h_i}{\lambda_i}\right)^2.$$

Верхняя оценка совпадает с нижней. Теорема доказана. □

4.5.3 Доказательства для раздела 4.3.2

Доказательство теоремы 4.3. Для удобства обозначим все точки из $D_H = \{\mathbf{x}_i\}$ и аналогично все точки из $D_{\frac{H}{m}} = \{\widetilde{\mathbf{x}}_j\}$. Тогда для регрессии на основе гауссовских процессов наилучшая оценка в точке имеет вид:

$$\widetilde{u}(\mathbf{x}) = \sum_i k_i u(\mathbf{x}_i) + \sum_j \widetilde{k}_j f(\widetilde{\mathbf{x}}_j).$$

для некоторых $k_i, \widetilde{k}_j$. Задача состоит в поиске коэффициентов k_i, $\widetilde{k}_j$, минимизирующих $\mathbb{E}(u(\mathbf{x}) - \widetilde{u}(\mathbf{x}))^2$. Используя независимость случайных процессов $f(\mathbf{x})$ и $g(\mathbf{x})$, получаем:

$$\begin{aligned}
\mathbb{E}(u(\mathbf{x}) - \widetilde{u}(\mathbf{x}))^2 &= \mathbb{E}\left[\rho f(\mathbf{x}) + g(\mathbf{x}) - \sum_i k_i(\rho f(\mathbf{x}_i) + g(\mathbf{x}_i)) - \sum_j \widetilde{k}_j f(\widetilde{\mathbf{x}}_j)\right]^2 = \\
&= \mathbb{E}\left[\rho f(\mathbf{x}) - \sum_i \rho k_i f(\mathbf{x}_i) - \sum_j \widetilde{k}_j f(\widetilde{\mathbf{x}}_j)\right]^2 + \\
&= \mathbb{E}\left[g(\mathbf{x}) - \sum_i k_i g(\mathbf{x}_i)\right]^2.
\end{aligned}$$

Для всех i существует такой индекс j, что $\mathbf{x}_i = \widetilde{\mathbf{x}}_j$. Обозначим

$$\widetilde{k}'_j = \begin{cases} \frac{1}{\rho}\widetilde{k}_j, & \forall i, \widetilde{\mathbf{x}}_j \neq \mathbf{x}_i, \\ \frac{1}{\rho}\widetilde{k}_j + k_i, & \exists i, \widetilde{\mathbf{x}}_j = \mathbf{x}_i. \end{cases}$$

Тогда существует взаимнооднозначное соответствие между $(\{k_i\}, \{\widetilde{k}_j\})$ и $(\{k_i\}, \{\widetilde{k}'_j\})$, так что минимизация $\mathbb{E}(u(\mathbf{x}) - \widetilde{u}(\mathbf{x}))^2$ по $k_i, \widetilde{k}_j$ эквивалента минимизации этой функции по $k_i, \widetilde{k}'_j$. Тогда

$$\begin{aligned}
&\mathbb{E}\left[\rho f(\mathbf{x}) - \sum_i k_i \rho f(\mathbf{x}_i) - \sum_j \widetilde{k}_j f(\widetilde{\mathbf{x}}_j)\right]^2 + \mathbb{E}\left[g(\mathbf{x}) - \sum_i k_i g(\mathbf{x}_i)\right]^2 = \\
=&\rho^2 \mathbb{E}\left[f(\mathbf{x}) - \sum_j \widetilde{k}'_j f(\widetilde{\mathbf{x}}_j)\right]^2 + \mathbb{E}\left[g(\mathbf{x}) - \sum_i k_i g(\mathbf{x}_i)\right]^2.
\end{aligned}$$

Для слагаемых $\mathbb{E}\left[f(\mathbf{x})-\sum_j \widetilde{k}'_j f(\widetilde{\mathbf{x}}_j)\right]^2$ и $\mathbb{E}\left[g(\mathbf{x})-\sum_i k_i g(\mathbf{x}_i)\right]^2$ получили задачу минимизации эквивалентную задаче для однородных данных, причем в первой задаче минимизируем по коэффициентам $\widetilde{k}'_j$, а во второй — по коэффициентам k_i.

Для k_i и $\widetilde{k}'_j$, которые минимизируют ошибку интерполяции в точке для однородных данных выполнено, что $\widetilde{k}'_j = K_f(\mathbf{x}-\mathbf{x}_j)$, $k_i = K_g(\mathbf{x}-\mathbf{x}_i)$ для некоторых симметричных ядер $K_f(\mathbf{x}-\mathbf{x}_j)$, $K_g(\mathbf{x}-\mathbf{x}_i)$.

Далее продолжим доказательство для $f(\mathbf{x})$ и $g(\mathbf{x})$ аналогично случаю однородных данных Для $\mathbb{E}\left[g(\mathbf{x})-\sum_i K_g(\mathbf{x}-\mathbf{x}_i)g(\mathbf{x}_i)\right]^2$ выполнено, что

$$\frac{1}{|H|}\int_{\substack{x_i\in[0,h_i],\\ i=\overline{1,d}}} \mathbb{E}\left[g(\mathbf{x})-\sum_i K_g(\mathbf{x}-\mathbf{x}_i)g(\mathbf{x}_i)\right]^2 d\mathbf{x} =$$

$$= \int_{\mathbb{R}^d} G(\boldsymbol{\omega})\left[\left[1-\widehat{K}_g(\boldsymbol{\omega})\right]^2 + \sum_{\mathbf{k}\in\mathbb{Z}^d\setminus\{\mathbf{0}\}} \widehat{K}_g^2\left(\boldsymbol{\omega}+H^{-1}\mathbf{k}\right)\right] d\boldsymbol{\omega}$$

Аналогично для интервала $[0,\frac{h_1]}{m}]\cdot\ldots\cdot[0,\frac{h_d}{m}]$ for $\mathbb{E}\left[f(\mathbf{x})-\sum_j K_f(\mathbf{x}-\widetilde{\mathbf{x}}_j)f(\widetilde{\mathbf{x}}_j)\right]^2$:

$$\frac{m^d}{|H|}\int_{\substack{x_i\in[0,\frac{h_i}{m}],\\ i=\overline{1,d}}} \mathbb{E}\left[f(\mathbf{x})-\sum_j K_f(\mathbf{x}-\widetilde{\mathbf{x}}_j)f(\widetilde{\mathbf{x}}_j)\right]^2 d\mathbf{x} =$$

$$= \int_{\mathbb{R}^d} F(\boldsymbol{\omega})\left[\left[1-\widehat{K}_f(\boldsymbol{\omega})\right]^2 + \sum_{\mathbf{k}\in\mathbb{Z}^d\setminus\{\mathbf{0}\}} \widehat{K}_f^2\left(\boldsymbol{\omega}+H^{-1}\mathbf{k}\right)\right] d\boldsymbol{\omega}.$$

Следовательно,

$$\frac{1}{|H|}\int_{\substack{x_i\in[0,h_i],\\ i=\overline{1,d}}} \mathbb{E}\left[f(\mathbf{x})-\sum_j K_f(\mathbf{x}-\widetilde{\mathbf{x}}_j)f(\widetilde{\mathbf{x}}_j)\right]^2 d\mathbf{x} =$$

$$= \int_{\mathbb{R}^d} F(\boldsymbol{\omega})\left[\left[1-\widehat{K}_f(\boldsymbol{\omega})\right]^2 + \sum_{\mathbf{k}\in\mathbb{Z}^d\setminus\{\mathbf{0}\}} \widehat{K}_f^2\left(\boldsymbol{\omega}+mH^{-1}\mathbf{k}\right)\right] d\boldsymbol{\omega}.$$

Таким образом, целевая ошибка интерполяции (4.5) имеет вид:

$$\sigma^2_{H,m}(\widetilde{u}, F, G, \rho) = \int_{\mathbb{R}^d} G(\boldsymbol{\omega}) \left[\left[1 - \widehat{K}_g(\boldsymbol{\omega})\right]^2 + \sum_{\mathbf{k} \in \mathbb{Z}^d \setminus \{\mathbf{0}\}} \widehat{K}^2_g \left(\boldsymbol{\omega} + H^{-1}\mathbf{k}\right) \right] d\boldsymbol{\omega} +$$
$$+ \rho^2 \int_{\mathbb{R}^d} F(\boldsymbol{\omega}) \left[\left[1 - \widehat{K}_f(\boldsymbol{\omega})\right]^2 + \sum_{\mathbf{k} \in \mathbb{Z}^d \setminus \{\mathbf{0}\}} \widehat{K}^2_f \left(\boldsymbol{\omega} + mH^{-1}\mathbf{k}\right) \right] d\boldsymbol{\omega}.$$

Наконец,

$$\sigma^2_{H,m}(\widetilde{u}, F, G, \rho) = \sigma^2_H(\widetilde{g}, G) + \rho^2 \sigma^2_{\frac{H}{m}}(\widetilde{f}, F).$$

□

4.5.4 Доказательства для раздела 4.4

В этом разделе приведем доказательство теоремы 4.5.

Доказательство Теоремы 4.5. Минимаксная ошибка интерполяции имеет вид:

$$R_2 = \frac{L_g}{2} \frac{1}{\pi^2} \left(\frac{\mathrm{w} + (m^*)^d}{B} \right)^{\frac{2}{d}} + \rho^2 \frac{L_f}{2} \frac{1}{\pi^2} \left(\frac{\mathrm{w} + (m^*)^d}{(m^*)^d B} \right)^{\frac{2}{d}}.$$

Обозначим $\delta = (m^*)^d$. Тогда необходимо минимизировать по a следующее выражение

$$\frac{L_g}{2} (\mathrm{w} + \delta)^{\frac{2}{d}} + \rho^2 \frac{L_f}{2} \left(\frac{\mathrm{w} + \delta}{\delta} \right)^{\frac{2}{d}}.$$

Частные производные по δ должны равняться 0:

$$\frac{L_g}{2} (\mathrm{w} + \delta)^{\frac{2}{d} - 1} \frac{2}{d} + \rho^2 \frac{L_f}{2} \left(\frac{\mathrm{w} + \delta}{\delta} \right)^{\frac{2}{d} - 1} \frac{2}{d} \frac{-c}{\delta^2} = 0.$$

Следовательно,

$$m^* = \sqrt[d+2]{\mathrm{w} \rho^2 \frac{L_f}{L_g}}.$$

□

4.6 Выводы

В этом разделе получена минимаксная ошибка интерполяции для регрессии на основе гауссовских процессов для случая многомерного входа, в том числе и для разнородных источников данных. Полученные результаты использованы для оценки ошибки интерполяции в случае наличия источников разнородных данных. Такая ошибка позволяет охарактеризовать случаи, в которых точность регрессионной модели может быть улучшена в случае использования разнородных данных.

Более того, получено оптимальное соотношение между размерами выборок разнородных данных. Предложен мотивированный полученными теоретическими результатами алгоритм выбора соотношения между размерами выборок разнородных данных.

Глава 5

Приложения разработанных методов построения регрессионных моделей разнородных источников данных

В данном разделе рассматривается использование разработанных методов для построения регрессионных моделей разнородных данных для решения прикладных задач индустриальной инженерии. Рассматривается, в каких случаях могут использоваться подходы, разработанные в исследовании, как полученные регрессионные модели могут использоваться для оптимизации изделий, каким образом может быть выбран оптимальным образом план экспериментов.

Для краткости в этой главе будем говорить об источнике данных низкой точности как о грубой функции, а об источнике данных высокой точности как о точной функции.

5.1 Использование предложенных в диссертации методов для построения регрессионных моделей разнородных данных

В этом разделе рассматриваются несколько искусственных задач и реальную задачу построения регрессионной модели для характеристик вращающегося диска в двигателе самолета и оптимизацию формы этого вращающегося диска.

5.1.1 Методология вычислительных экспериментов

Сравниваются четыре перечисленных ниже подхода для построения регрессионных моделей разнородных данных, которые затем используются для решения оптимизационной задачи; последние два подхода впервые предложены в диссертационном исследовании в разделе 3:

- GP — регрессия на основе гауссовских процессов с использованием только данных от источника данных высокой точности,
- VFGP — регрессия на основе гауссовских процессов для данных разной точности,
- SVFGP — регрессия на основе гауссовских процессов для данных разной точности для больших выборок, описанная в разделе 3.3,
- BB VFGP — регрессия на основе гауссовских процессов для данных разной точности с использованием черного ящика для источника данных низкой точности, описанная в разделе 3.4. В экспериментах используется исходный план эксперимента такой же, как для VFGP, а для обновления модели в каждой новой точке подсчитывается значение функции низкой точности в этой точке.

В качестве ковариационной функции для описанных выше алгоритмов используется широко распространенная квадратичная экспоненциальная ковариационная функция, описанная в разделе 2.1 и в [80]. Для того, чтобы избежать обращения плохо обусловленных матриц [68, 74], вводится байесовское априорное распределение для параметров ковариационной функции, аналогичное 2.1.1 и затем используются байесовские оценки параметров, полученные из принципа максимума апостериорной вероятности. Кроме того, для повышения устойчивости процедуры оценки параметров, в реализованном комплексе программ используется нормализациях исходных входных и выходных данных.

Для оценки параметров SVFGP используется только подвыборка базовых точек, в то время как для предсказания значений в новых точках используется вся выборка.

Оценка качества полученных регрессионных моделей для искусственных данных проводится с использованием отложенной тестовой выборки, так как искусственную аналитическую функцию мы можем мгновенно подсчитать в любом числе точек. Оценка качества моделей для реальных данных проводится с помощью скользящего контроля [41]. Отметим, что значение грубой функции в точке используется для построения модели, только если такая точка не принадлежит к выбранной тестовой выборке. Для одномерной целевой функции и тестовой выборки значений точной функции $\mathbf{S}_{\text{test}} = \{\mathbf{x}_i^{\text{test}}, u_i^{\text{test}} = u(\mathbf{x}_i^{\text{test}})\}_{i=1}^{n_t}$ значение ошибки RRMS для регрессионной модели $\widetilde{u}(\mathbf{x})$ равно

$$RRMS(D_{\text{test}}, \widetilde{u}) = \sqrt{\frac{\sum_{i=1}^{n_t}(\widehat{u}(\mathbf{x}_i^{\text{test}}) - u_i^{\text{test}})^2}{\sum_{i=1}^{n_t}(\overline{u} - u_i^{\text{test}})^2}}, \tag{5.1}$$

здесь $\overline{u} = \frac{1}{n_t}\sum_{i=1}^{n_t} u_i^{\text{test}}$. Значение RRMS обычно лежит от 0 до 1. Для регрессионных моделей высокого качества значение RRMS близко к 0, а для регрессионных моделей низкого качества значения RRMS близки или больше чем 1.

5.1.2 Задача с искусственными данными

Мы рассматриваем задачу построения регрессионной модели для разнородных данных для широко используемой в литературе искусственной задачи[34]. В этой задаче точная функция $u(x)$ и грубая функция $f(x)$ заданы следующим

Таблица 5.1 – RRMS ошибки для различных размеров выборки значений точной функции n_u

n_u	6	15	30
GP	0.7102	0.0159	$3.83e-04$
VFGP	0.3036	$7.42e-04$	$1.38e-04$
BB VFGP	0.1610	$6.90e-07$	$1.67e-07$

образом:

$$
\begin{aligned}
u(x) &= (6x-2)^2 \sin(12x-4),\\
f(x) &= 0.5u(x) + 10(x-1).
\end{aligned}
$$

Для того, чтобы оценить точность построенных регрессионных моделей используется следующая процедура:

- Сгенерировать выборку значений точной функции размера $n_u \leq 100$ с точками, равномерно распределенными на $[0, 1]$. Мы рассматриваем $n_f = 6, 15$ и 30,
- Сгенерировать выборку значений грубой функции с точками из выборки для точной функции и дополнительными $(100 - n_u)$ точками, равномерно распределенными на $[0, 1]$,
- Построить регрессионные модели, используя алгоритмы GP, VFGP и BB VFGP.
- Оценить точность построенных регрессионных моделей с помощью тестовой выборки размера 1000 значений точной функции.

Усредняя результаты по 50 запускам для каждого значения n_u, получаем ошибки RRMS, представленные в Таблице 5.1. Видно, что использование черного ящика для источника данных низкой точности существенно улучшает качество моделей для всех рассмотренных значений n_u.

5.1.3 Задача с искусственными данными для обучающей выборки большого размера

Для того, чтобы оценить качество предложенных методов, будем использовать искусственную функцию со множеством локальных особенностей и размерностью входа $d = 6$. Для построения достаточно точной регрессионной модели в таком случае действительно нужна большая обучающая выборка. В качестве точной функции $u(\mathbf{x})$ и грубой функции $f(\mathbf{x})$ используются

$$u(\mathbf{x}) = 20 + \sum_{i=1}^{d}(x_i^2 - 10\cos(2\pi x_i)) + \varepsilon_u,\ \mathbf{x} \in [0,1]^d,$$

$$f(\mathbf{x}) = u(\mathbf{x}) + 0.2\sum_{i=1}^{d}(x_i + 1)^2 + \varepsilon_f,\ \mathbf{x} \in [0,1]^d.$$

Точная функция зашумлена гауссовскм белым шумом ε_u с дисперсией 0.001, и грубая функция зашумлена гауссовским белым шумом ε_f с дисперсией 0.002. При генерации выборок используются точки из $[0,1]^d$ полученные с помощью метода оптимальных латинских гиперкубов OLHS [71]. Для тестирования свойств экстраполяции ограничивается область генерации точек обучающей выборки до $[0, 0.5]$ вместо $[0, 1]$ по очереди для каждой входной размерности из 6, в то время как точки тестовой выборки равномерно распределены на всем гиперкубе $[0,1]^6$. Размер выборки для точной функции равнялся $n_u = 100$, размер выборок, которые использовались в качестве базовых для SVFGP, равнялся $n^1 = 1000$ во всех экспериментах.

Результаты усреднены по 5 запускам для каждого рассмотренного значения n_f.

- Таблица 5.2 содержит RRMS ошибки для VFGP, SVFGP и BB VFGP,
- Таблица 5.3 содержит RRMS ошибки для VFGP, SVFGP и BB VFGP в случае использования модели для экстраполяции,
- Таблица 5.4 содержит время обучения для алгоритмов VFGP, SVFGP и BB VFGP.

Таблица 5.2 – Сравнение ошибок RRMS для VFGP, SVFGP и BB VFGP. Приведены так же стандартные отклонения ошибок, подсчитанные на основе проведенных 5 запусках для различных обучающих выборок

n_f	1000	3000	5000
VFGP	0.0502 ± 0.0018	0.0170 ± 0.0005	0.0058 ± 0.0004
SVFGP	0.0502 ± 0.0018	0.0305 ± 0.0010	0.0260 ± 0.0017
BB VFGP	$0.0010 \pm 6 \cdot 10^{-5}$	$0.00029 \pm 8 \cdot 10^{-5}$	$0.00017 \pm 8 \cdot 10^{-6}$

Таблица 5.3 – Сравнение ошибок RRMS для VFGP, SVFGP и BB VFGP. Экстраполяция. Приведены так же стандартные отклонения ошибок, подсчитанные на основе проведенных 5 запусках для различных обучающих выборок

n_f	1000	3000	5000
VFGP	0.3636 ± 0.1315	0.1351 ± 0.0075	0.1028 ± 0.0177
SVFGP	0.3636 ± 0.1315	0.3281 ± 0.0186	0.3586 ± 0.0987
BB VFGP	0.00099 ± 0.00058	0.00113 ± 0.00042	0.00034 ± 0.00010

Ошибки RRMS для SVFGP сравнимы с RRMS ошибкам VFGP для одинакового размера выборки, но время обучения для SVFGP существенно меньше, если общий размер обучающей выборки равен 5000. Для BB VFGP время обучения в этом эксперименте совпадает со временем обучения для VFGP, но для размера выборки 1000 для BB VFGP получаются ошибки меньше, чем при использовании выборки размера 5000 и алгоритма VFGP. В случае экстраполяции модели, результаты существенно лучше для BB VFGP.

5.1.4 Задача о вращающемся диске

Сравним работу методов построения регрессионных моделей в реальной задаче построения модели зависимости характеристик вращающегося диска в двигателе самолета от его формы, описанной в разделе 1.2.1.

Таблица 5.4 – Сравнение времени обучения в секундах для VFGP, SVFGP и BB VFGP

n_f	1000	3000	5000
VFGP	30.46	852.70	7283.27
SVFGP	30.46	33.42	37.50
BB VFGP	30.38	842.97	7672.60

Таблица 5.5 – RRMS ошибки для четырех предложенных подходов. Выход $u_{\max}$

n_u	20	40	60	80	100
GP	0.3368	0.1826	0.1305	0.1091	0.0756
VFGP	0.1679	0.0998	0.0822	0.0564	0.0435
SVFGP	0.1018	0.0658	0.0494	0.0427	0.0339
BB VFGP	0.0964	0.0717	0.0503	0.0434	0.0347

Таблица 5.6 – RRMS ошибки для четырех предложенных подходов. Выход $s_{\max}$

n_u	20	40	60	80	100
GP	0.5261	0.3181	0.2164	0.2095	0.1643
VFGP	0.2336	0.2326	0.2058	0.1321	0.1088
SFGP	0.1674	0.1095	0.1023	0.0939	0.0812
BB VFGP	0.1583	0.1283	0.1295	0.0899	0.0793

Построение регрессионных моделей

В данном разделе сравниваются подходы SVFGP (регрессия на основе разреженных гауссовских процессов для данных разной точности) и BB VFGP (регрессия на основе гауссовских процессов для данных разной точности с использованием черного ящика для грубой функции), предложенные в данной работе, и широко используемые подходы GP (регрессия на основе гауссовских процессов с использованием только точных данных) и VFGP.

Выборка была получена с помощью техники OLHS. Для построения регрессионной модели использовались выборка размером n_u значений точной функции и выборка размером 1000 для грубой функции; n_u изменялось от 20 до 100. Для оценки качества модели использовался скользящий контроль для выборки значений точной функции размером 140 точек. Для SVFGP мы использовали выборку размером $n_f = 5000$ значений грубой функции, из которой случайным образом выбиралось $n_f^1 = 1000$ базовых точек.

Результаты представлены в таблице 5.5 для выхода $u_{\max}$ и в таблице 5.6 для выхода $s_{\max}$. VFGP работает лучше, чем GP, и оба предложенных в работе подхода SVFGP и BB VFGP работают лучше VFGP и GP. В рассмотренной задаче SVFGP и BB VFGP дают примерно одинаковые значения ошибок скользящего контроля; таким образом, выбор техники из двух предложенных может быть задан дополнительными сведениями о задаче. Например, доступен ли черный ящик для грубой функции при использовании регрессионной модели, будет ли использоваться модель для экстраполяции данных и так далее.

5.1.5 Оптимизация формы вращающегося диска

Решается задача оптимизации формы вращающегося диска, описанная в разделе 1.2.1. В представленной задаче несколько целевых функций, поэтому результатом оптимизации будет Парето фронт, а не единственная точка.

Один запуск алгоритма оптимизации состоит в данном случае из следующих шагов:

- Сгенерировать начальную выборку размера 30, используя LHS.
- Построить регрессионные модели, используя GP, VFGP, SVFGP и BB VFGP.
- Решить задачу мультикритериальной оптимизации, используя в качестве целевых функций и ограничений их регрессионные модели, полученные на предыдущем шаге.
- Для оценки качества модели подсчитать истинные значения целевых функций и ограничений.

Для выбранных настроек алгоритма оптимизации размер Парето фронта был около 30, таким образом, для каждого запуска нам понадобилось около 60 запусков вычислительного кода для точной функции. Так как количество точек, полученных в результате каждого запуска оптимизации, порядка 300 и всего было сделано 10 запусков, вместо истинных значений целевых функций и ограничений использовались их регрессионные модели, полученные с использованием всей доступной выборки значений точной функции размеров 500 точек, расположенных равномерно по всей области дизайна с дополнительными точками в окрестности целевого Парето фронта.

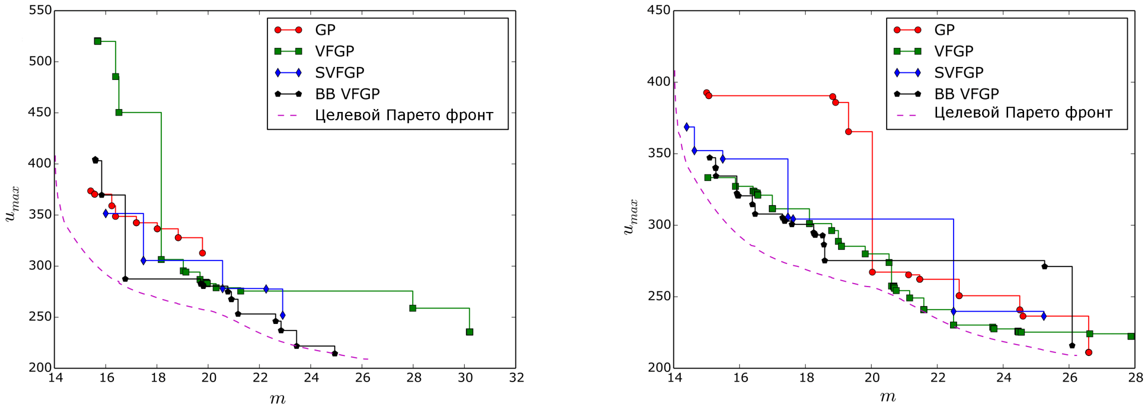


Рис. 5.1 – Сравнение Парето фронтов, полученных оптимизацией регрессионных моделей, построенных с помощью GP, VFGP и BB VFGP с настоящим Парето фронтом

Примеры полученных Парето фронтов — на рисунке 5.1. Для проведенных запусков SVFGP и BB VFGP работают лучше, чем GP и VFGP.

Результаты оптимизации представлены в таблице 5.7. Сравниваются минимальные значения для различных взвешенных сумм двух целевых функций m и $u_{\max}$, усредненные по 10 запускам оптимизации для различных начальных выборок. Мы получаем наилучшее значения массы m, используя алгоритм SVFGP и наилучшее значение $u_{\max}$, используя алгоритм BB VFGP, вто время как алгоритмы GP и VFGP работают хуже. Так же используя BB VFGP мы получаем существенно большую долю точек в Парето фронте, которые удовлетворяют ограничениям оптимизационной задачи, чем для алгоритмов GP, VFGP и SVFGP. Так как необходимо вычислять значения точной функции во всех точках Парето фронта, большая доля точек, которые удовлетворяют ограничениям снижает количество лишних вычислений точной функции.

5.2 Построение регрессионных моделей для крыла самолета и для С-образного пресса

Рассматривались задачи построения регрессионных моделей зависимости характеристик изделия от его параметров, описанные в разделах 1.2.2 и 1.2.3. В этом случае использовалась предложенная в диссертационной работе техника

Таблица 5.7 – Результаты оптимизации для 3 различных алгоритмов построения регрессионной модели. Представлена доля полученных точек, которые удовлетворяют ограничениям.

Целевая функция	GP	VFGP	SVFGP	BB VFGP
m	16.62	15.69	**15.09**	15.63
$0.8m + 0.2u_{\max}$	73.65	70.74	70.71	**68.10**
$0.6m + 0.4u_{\max}$	125.10	117.37	116.21	**112.55**
$0.4m + 0.6u_{\max}$	176.55	163.89	161.18	**156.99**
$0.2m + 0.8u_{\max}$	228.00	210.33	206.12	**201.44**
$u_{\max}$	279.44	256.77	251.05	**245.89**
Доля допустимых точек	0.54	0.57	0.55	**0.75**

MFGP, предназначенная для построения моделей в случае наличия двух и более различных источников данных, упорядоченных по точности. В силу требований к задаче вместо ошибки RRMS использовалась ошибка RMAE, которая для тестовой выборки значений точной функции $\mathbf{S}_{\text{test}} = \{\mathbf{x}_i^{\text{test}}, u_i^{\text{test}} = u(\mathbf{x}_i^{\text{test}})\}_{i=1}^{n_t}$ и регрессионной модели $\widetilde{u}(\mathbf{x})$ имеет вид:

$$RMAE(D_{\text{test}}, \widetilde{u}) = \sqrt{\frac{\sum_{i=1}^{n_t} |\widehat{u}(\mathbf{x}_i^{\text{test}}) - u_i^{\text{test}}|^2}{\sum_{i=1}^{n_t} |\overline{u} - u_i^{\text{test}}|}},$$

Результаты для задач построения регрессионных моделей для крыла самолета и для C-образного пресса представлены в таблице 5.8. Для задачи построения регрессионной модели зависимости аэродинамического качества крыла самолета от геометрии крыла и режим полета представленные результаты демонстрируют, что для построения модели коэффициента сопротивления использование дополнительных источников данных низкой точности позволяет улучшить качество модели, улучшение есть также и при добавлении еще одного источника данных более низкой точности. Для моделей коэффициента подъемной силы улучшение не так заметно, но значимо в большинстве проведенных экспериментов.

Для задачи построения модели зависимости прочности C-образного диска приведенные результаты показывают, что использование разнородных источников данных значимо улучшает качество для модели максимального смещения, но слабо улучшает качество модели максимального стресса, так как получен-

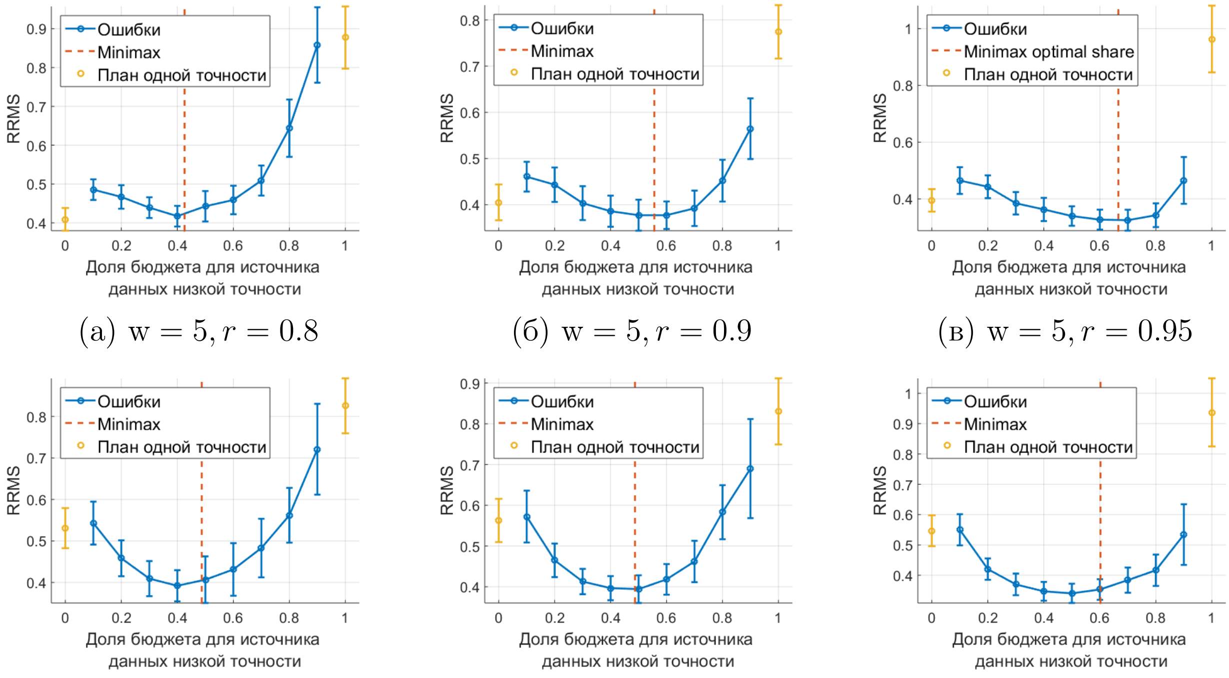
Таблица 5.8 – RMAE ошибки со стандартными отклонениями для реальных задач. Отмеченные знаком $*$ ошибки с $p < 0.05$ меньше ошибок для других техник (например, в случае если ошибки для техники MFGP существенно меньше ошибок для техник GP и VFGP)

Выход	Размер выборки $\lvert D_1\rvert$	$\lvert D_2\rvert$	$\lvert D_3\rvert$	GP	VFGP	MFGP
Задача моделирования крыла самолета						
Коэффициент подъемной силы	20	40	80	0.8162	0.7874	0.7406*
	50	100	200	0.678	0.616	0.559*
	100	200	400	0.578	0.523	0.483*
	300	300	300	0.396	0.375	0.364*
Коэффициент сопротивления	20	40	80	0.2107	0.2055	0.2043
	50	100	200	0.150	0.147	0.133*
	100	200	400	0.129	0.124	0.103*
	300	300	300	0.111	0.100	0.093*
Задача моделирования С-образного диска						
Максимальный стресс	20	40	80	0.9338	0.9568	0.9239*
	50	100	200	0.9178	0.8217	0.8007*
	100	200	400	0.9155	0.7770	0.7668*
	300	300	300	0.8564	0.7222	0.7223
Максимальное смещение	20	40	80	0.6653	0.5803	0.4532*
	50	100	200	0.4606	0.3417	0.2565*
	100	200	400	0.3087	0.2218	0.1693*
	300	300	300	0.1694	0.1504	0.1385*

ные модели максимального стресса в целом не очень высоко качества даже для достаточно больших выборок размером $|D_1| = |D_2| = |D_3| = 300$.

5.3 Выбор соотношения между размерами выборок разнородных данных

Получим качество работы предложенного в главе 4 алгоритма для выбора оптимального отношения размеров выборок данных разной точности для заданного бюджета. Будем использовать искусственные данные, сгенерированные как реализации гауссовских процессов, и реальные данные, взятые, как правило, из инженерных приложений.



(а) $\mathrm{w} = 5, r = 0.8$ (б) $\mathrm{w} = 5, r = 0.9$ (в) $\mathrm{w} = 5, r = 0.95$

(г) $\mathrm{w} = 10, r = 0.8$ (д) $\mathrm{w} = 10, r = 0.9$ (е) $\mathrm{w} = 10, r = 0.95$

Рис. 5.2 – Искусственные данные. Зависимость ошибки RRMS от доли бюджета, предназначенной для использования источником данных низкой точности. Мы рассматриваем случай $d = 3$, различных коэффициентов корреляции r и стоимостей использования данных высокой точности w. Крайние слева и справа точки соответствуют планам экспериментов, для которых мы используем только данные низкой точности либо только данные высокой точности. Результаты усреднены по 20 запускам. Пунктирная вертикальная линия соответствует минимаксно оптимальному отношению размеров выборок, сплошная линия — реальным полученным ошибкам.

Будем использовать ковариационную функцию Матерна с $\nu = \frac{3}{2}$ $c_{\boldsymbol{\theta}}(\mathbf{x} - \mathbf{x}')$, которая обеспечивает дифференцируемые аппроксимации целевой функции [80]:

$$c_{\boldsymbol{\theta}}(\mathbf{x} - \mathbf{x}') = (1 + \sqrt{3}\mathrm{d}_{\boldsymbol{\theta}}(\mathbf{x} - \mathbf{x}')) \exp(-\sqrt{3}\mathrm{d}_{\boldsymbol{\theta}}(\mathbf{x} - \mathbf{x}')) \,,$$

где $\mathrm{d}_{\boldsymbol{\theta}}(\mathbf{x} - \mathbf{x}') = \sqrt{\sum_{i=1}^{d} \theta_i (x_i - x_i')^2}$. Для построение модели регрессии на основе гауссовских процессов будем использовать байесовские оценки параметров ковариационной функции [22] полученные с использованием подходов [23], так как открытые решения как правило требуют ручной подстройки для каждой конкретной задачи [61].

Качество модели будем измерять оценивая ошибку RRMS (5.1), полученную на тестовой выборке для искусственных данных или с использованием скольззящего контроля для реальных данных.

5.3.1 Эксперименты на искусственных данных

В этом разделе будем генерировать искусственные данные как реализации гауссовских процессов с заданной ковариационной функцией. Мы следуем модели $u(\mathbf{x}) = \rho f(\mathbf{x}) + g(\mathbf{x})$, и генерируем вложенные планы экспериментов, то есть $D_u \subseteq D_f$, точки обучающей выборки равномерно случайно выбираются из $[0, 1]^d$. Общий бюджет вычислений 300, и стоимость использования $u(\mathbf{x})$ либо $\mathrm{w} = 5$, либо $\mathrm{w} = 10$. Так как точные значения ρ и r известны, мы используем их для вычисления δ^*.

Рисунки 5.2 показывают зависимость ошибки RRMS от доли бюджета, использованной для источника данных низкой точности. Видно, что наша оценка оптимального значения δ^* близка к оптимальному на практике значению несмотря на то, что используются не сеточные планы экспериментов, и параметры модели регрессии оцениваются по обучающей выборке.

5.3.2 Алгоритмы для сравнения

Мы сравниваем наш алгоритм для оценки оптимального отношения размеров выборок **MinMinimax** с четырьмя разными эвристиками:

- **High** — используются только данные высокой точности,
- **Low** — используются только данные низкой точности,
- **EqualSize** — размеры выборок высокой и низкой точности равны,
- **EqualBudget** — доли бюджета, использованные источниками данных разной точности, равны.

Относительные размеры выборок приведены в таблице 5.9.

Таблица 5.9 – Размеры выборки данных высокой точности n_u и данных низкой точности n_f для заданного общего бюджета B

Алгоритм	n_{u}	n_{f}
High	B/w	0
EqualSize	$B/(\mathrm{w}+1)$	$B/(\mathrm{w}+1)$
EqualBudget	$B/(2\mathrm{w})$	$B/2$
Low	0	B
MinMinimax	$B/(\mathrm{w}+\delta^*)$	$\delta^* B/(\mathrm{w}+\delta^*)$

Таблица 5.10 – RRMS ошибки усредненные по 20 запускам скользящего контроля для реальных данных

Задача	Номер выхода	d	High	EqualSize	EqualBudget	MinMinimax	Low
Euler	1	11	0.767	0.892	0.846	**0.742**	0.913
Euler	2	11	**0.066**	0.077	0.269	0.380	0.397
Airfoil	1	6	0.546	0.594	0.539	0.5221	**0.4852**
Airfoil	2	6	**0.120**	0.142	0.130	0.138	0.296
MachAngle	1	2	**0.088**	0.106	0.195	0.195	0.405
MachAngle	2	2	**0.093**	0.114	0.171	0.179	0.365
Press12	1	6	0.559	0.601	0.358	**0.277**	0.284
Press12	2	6	0.443	0.491	0.271	**0.176**	**0.176**
Press13	1	6	0.559	0.575	0.386	**0.348**	0.543
Press13	2	6	0.449	0.485	0.278	**0.179**	**0.179**
Disk	1	6	0.299	0.340	0.192	0.193	**0.163**
Disk	2	6	0.446	0.457	0.299	0.299	**0.272**
SVM	1	2	**0.148**	0.149	0.184	0.164	0.608
Supernova	1	3	0.039	0.048	**0.018**	0.057	0.057

5.3.3 Эксперименты на реальных данных

Рассматривается набор реальных задач. Первые три задачи (**Euler**, **Airfoil** [14], **MachAngle**) соответствуют подсчету коэффициентов подъемной силы и сопротивления для крыла самолета в зависимости от геометрии крыла и условий полета. Следующие две задачи (**Press** [23], **Disk** [110]) рассматривают зависимость максимального напряжения и максимального смещения в зависимости от геометрии изделия. Хотя доступны данные трех различных точностей для задачи **Press**, в каждом эксперименте мы используем только две точности. Последние две задачи ([48], **SVM**, **Supernova**) связаны с моделированием зависимости качества модели от ее параметров. Входные размерности для задач даны в таблице 5.10.

Бюджет B равен 300 для всех задач кроме **Euler**, так как для этой задачи доступная выборка данных недостаточного размера. Для того, чтобы сравнение было представительным, для всех задач w = 5 . Если алгоритм **MinMinimax** выдавал $n_u < 1$, то для **MinMinimax** использовался только источник данных низкой точности. Для **MinMinimax** использовался коэффициент корреляции r, подсчитанный по всей выборке. Кроме того, перед построением регрессионных моделей проводилась нормализация данных, а именно использовались нормированные выходы с нулевым средним и единичной дисперсией.

Полученные ошибки представлены в таблице 5.10. Модели, построенные с помощью выборок, план экспериментов для которых получен с помощью метода **MinMinimax**, в проведенных экспериментах близки к наилучшим. Однако, оказалось, что данные могут обладать двумя особенностями: часто невозможно улучшить модель с использованием данных низкой точности для заданного бюджета; или наоборот необходимо использовать слишком маленькую выборку данных высокой точности. Например, для набора данных **Supernova** подход **MinMinimax** сработал не очень хорошо, так как предложенный им размер выборки данных высокой точности — 4, что очевидно недостаточно для построения точной регрессионной модели кокригинга.

5.4 Выводы

Предложенные в работе алгоритмы были разработаны на языке Matlab. Затем они были переписаны на языке C++ и вошли в состав разработанного в компании DATADVANCE программного комплекса, использующегося в двух программных продуктах компании: pSeven Core (библиотека, содержащая функциональность для интеллектуального анализа данных и многокритериальной оптимизации) и pSeven (платформа, предназначенная для автоматизации инженерного анализа, многодисциплинарной оптимизации и анализа данных).

Проведенные вычислительные эксперименты на реальных и модельных данных показывают, что использование предложенных методов приводит к получению более точных регрессионных моделей за меньшее время. Например, использование регрессионной модели, построенной с помощью предложенных в исследовании методов, для оптимизации формы вращающегося диска позволило на 20% улучшить качество модели для того же ресурсного ограничения.

Заключение

В данной диссертации проведены разработка и исследование методов построение регрессионных моделей разнородных данных в индустриальной инженерии. В частности:

1. Разработан алгоритм построения нелинейных регрессионных моделей для выборок разнородных на основе низкоранговой аппроксимации с трудоемкостью $O(\phi(n)^2 n)$ вместо $O(n^3)$ для стандартного подхода, для которого значение $\phi(n)$ обычно выбирают порядка $\min(c, n)$, где c — константа, задаваемая требованием к качеству модели, причем вычислительные эксперименты, в том числе и для реальных задач, демонстрируют, что такой алгоритм позволяет улучшать качество получаемых регрессионных моделей разнородных данных и время на их построение.

2. Получена оценка качества нелинейной регрессионной модели на основе гауссовских процессов, в том числе и для источников данных разной точности, позволяющая оценивать целесообразность использования источников данных разной точности и оценивать качество модели как для случая, когда ковариационная функция задана, так и когда лишь известна гладкость рассматриваемой целевой функции.

3. Впервые предложен метод выбора соотношения между размерами выборок разнородных данных, максимизирующего качество регрессионной модели, построенной на основе плана экспериментов с выбранным соотношением размеров выборок разнородных данных, причем такой метод обеспечивает теоретически оптимальное соотношение между размерами

выборок и достаточно хорошее качество моделей для широкого класса задач регрессионного моделирования в индустриальной инженерии.

4. Разработанные методы вошли в состав программного комплекса, предназначенного для решения задач анализа данных в индустриальной инженерии, причем особое внимание уделено качеству построенных моделей и наличию всей необходимой функциональности для построения регрессионных моделей разнородных данных и решению оптимизационных задач с помощью предложенного подхода.

5. С помощью разработанного программного комплекса решен ряд задач индустриальной инженерии, причем во многих случаях с помощью предложенных в диссертации методов построения регрессионных моделей разнородных данных и методов выбора плана экспериментов разнородных данных, в частности для задачи построения регрессионной модели зависимости характеристик вращающегося диска в двигателе самолета от его формы, крыла самолета от его геометрии и режима полета, C-образного пресса от его формы.

Список литературы

1. M. Abramowitz and I. Stegun. *Handbook of mathematical functions: with formulas, graphs, and mathematical tables.* English. Vol. 55. Courier Corporation, 1964.

2. N. Alexandrov et al. *First-order model management with variable-fidelity physics applied to multi-element airfoil optimization.* English. Tech. rep. NASA, 2000.

3. N. Alexandrov et al. *Optimization with variable-fidelity models applied to wing design.* English. Tech. rep. DTIC Document, 1999.

4. D. Allaire and K. Willcox. "Fusing information from multifidelity computer models of physical systems". English. In: *Information Fusion (FUSION), 2012 15th International Conference on.* IEEE. 2012, pp. 2458–2465.

5. M.A. Álvarez and N.D. Lawrence. "Computationally efficient convolved multiple output Gaussian processes". English. In: *Journal of Machine Learning Research* 12 (2011), pp. 1425–1466.

6. S.C. Armand. *Structural optimization methodology for rotating disks of aircraft engines.* English. Tech. rep. National Aeronautics, Space Administration, Office of Management, Scientific, and Technical Information Program, 1995.

7. J. Bandler et al. "Space mapping technique for electromagnetic optimization". English. In: *Microwave Theory and Techniques, IEEE Transactions on* 42.12 (1994), pp. 2536–2544.

8. John W Bandler et al. “Space mapping: the state of the art”. English. In: *Microwave Theory and Techniques, IEEE Transactions on* 52.1 (2004), pp. 337–361.
9. S. Banerjee et al. “Gaussian predictive process models for large spatial data sets”. English. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 70.4 (2008), pp. 825–848.
10. T.R. Barrett, N.W. Bressloff, and A.J. Keane. “Airfoil shape design and optimization using multifidelity analysis and embedded inverse design”. English. In: *AIAA journal* 44.9 (2006), pp. 2051–2060.
11. J. Batina. “Unsteady Euler airfoil solutions using unstructured dynamic meshes”. English. In: *AIAA journal* 28.8 (1990), pp. 1381–1388.
12. M. Belyaev, E. Burnaev, and Y. Kapushev. “Gaussian Process Regression for Structured Data Sets”. English. In: *Lecture Notes in Artificial Intelligence. Proceedings of SLDS 2015. A. Gammerman et al. (Eds.)* Vol. 9047. London, UK: Springer, 2015, pp. 106–115.
13. A. Bernstein, A. Kuleshov, and Y. Yanovich. “Information preserving and locally isometric&conformal embedding via Tangent Manifold Learning”. English. In: *Data Science and Advanced Analytics (DSAA), 2015. 36678 2015. IEEE International Conference on*. IEEE. 2015, pp. 1–9.
14. A. Bernstein et al. “Comparison of three geometric parameterization methods and their effect on aerodynamic optimization”. English. In: *Proceedings of International Conference on Evolutionary and Deterministic Methods for Design, Optimization and Control with Applications to Industrial and Societal Problems, Eurogen*. 2011, pp. 14–16.
15. M. Bevilacqua et al. “Covariance tapering for multivariate Gaussian random fields estimation”. English. In: *Statistical Methods & Applications* (2015), pp. 1–17.

16. A. Bhattacharya, D. Pati, and D. Dunson. “Anisotropic function estimation using multi-bandwidth Gaussian processes”. English. In: *Annals of statistics* 42.1 (2014), pp. 352–381.

17. C.M. Bishop. *Pattern recognition and machine learning*. English. Vol. 4. 4. Springer New York, 2006.

18. C. Bockenheimer. “The Airbus SHM Development Process”. English. In: *2nd International Symposium on NDT in Aerospace*. 2010.

19. S. Boucheron, G. Lugosi, and P. Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. English. OUP Oxford, 2013.

20. P. Boyle and M. Frean. “Dependent Gaussian processes”. English. In: *Advances in Neural Information Processing Systems* 17 (2005), pp. 217–224.

21. E. Burnaev and M. Panov. “Adaptive design of experiments based on Gaussian processes”. English. In: *Statistical Learning and Data Sciences*. Springer, 2015, pp. 116–125.

22. E. Burnaev, M. Panov, and A. Zaytsev. “Regression on the basis of nonstationary Gaussian processes with Bayesian regularization”. English. In: *Journal of Communications Technology and Electronics* 61.6 (2016), pp. 661–671.

23. E. Burnaev and A. Zaytsev. “Surrogate modeling of multifidelity data for large samples”. English. In: *Journal of Communications Technology and Electronics* 60.12 (2015), pp. 1348–1355.

24. S.-H. Chang, P.C. Cosman, and L.B. Milstein. “Chernoff-type bounds for the Gaussian error function”. English. In: *Communications, IEEE Transactions on* 59.11 (2011), pp. 2939–2944.

25. R.B. Chen et al. “Contour estimation via two fidelity computer simulators under limited resources”. English. In: *Computational Statistics* 1 (2012), pp. 1–22.

26. T. Chu, J. Zhu, and H. Wang. "Penalized maximum likelihood estimation and variable selection in geostatistics". English. In: *The Annals of Statistics* 39.5 (2011), pp. 2607–2625.

27. N. Cressie. *Geostatistics.* English. Vol. 43. 4. Taylor & Francis, 1989, pp. 197–202.

28. N. Cressie. *Statistics for Spatial Data: Wiley Series in Probability and Statistics.* English. Wiley: New York, NY, USA, 1993.

29. J.A. Cumming and M. Goldstein. "Small sample Bayesian designs for complex high-dimensional models based on information gained using fast approximations". English. In: *Technometrics* 51.4 (2009), pp. 377–388.

30. E.D. Dolan and J.J. More. "Benchmarking optimization software with performance profiles". English. In: *Mathematical Programmming* 91.2 (2002), pp. 201–213.

31. P. Drineas and M.W. Mahoney. "On the Nyström method for approximating a Gram matrix for improved kernel-based learning". English. In: *The Journal of Machine Learning Research* 6 (2005), pp. 2153–2175.

32. M. El-Beltagy and W. Wright. "Gaussian processes for model fusion". English. In: *Artificial Neural Networks—ICANN 2001* 1 (2001), pp. 376–383.

33. A. Forrester, A. Sobester, and A. Keane. *Engineering design via surrogate modelling: a practical guide.* English. Wiley, 2008.

34. A.I.J. Forrester, A. Sóbester, and A.J. Keane. "Multi-fidelity optimization via surrogate modelling". English. In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science* 463.2088 (2007), pp. 3251–3269.

35. L. Foster et al. "Stable and efficient Gaussian process calculations". English. In: *The Journal of Machine Learning Research* 10 (2009), pp. 857–882.

36. R. Furrer, M.G. Genton, and D. Nychka. “Covariance tapering for interpolation of large spatial datasets”. English. In: *Journal of Computational and Graphical Statistics* 15.3 (2006).

37. G.K. Golubev and E.A. Krymova. “On interpolation of smooth processes and functions”. English. In: *Problems of Information Transmission* 49.2 (2013), pp. 127–148.

38. L.L. Gratiet. “Bayesian analysis of hierarchical multi-fidelity codes”. English. In: *arXiv preprint arXiv:1112.5389* 1 (2011), p. 1.

39. Z. Han and S. Görtz. “Hierarchical kriging model for variable-fidelity surrogate modeling”. English. In: *AIAA journal* 50.9 (2012), pp. 1885–1896.

40. Z. Han, S. Görtz, and R. Zimmermann. “Improving variable-fidelity surrogate modeling via gradient-enhanced kriging and a generalized hybrid bridge function”. English. In: *Aerospace Science and Technology* 25.1 (2013), pp. 177–189.

41. T. Hastie et al. “The elements of statistical learning: data mining, inference and prediction”. English. In: *The Mathematical Intelligencer* 27.2 (2005), pp. 83–85.

42. J. Hensman, N. Fusi, and N.D. Lawrence. “Gaussian processes for big data”. English. In: *UAI-13*. 2013, pp. 282–290.

43. D. Huang et al. “Sequential kriging optimization using multiple-fidelity evaluations”. English. In: *Structural and Multidisciplinary Optimization* 32.5 (2006), pp. 369–382.

44. Z. Huang et al. “Optimal design of aeroengine turbine disc based on kriging surrogate models”. English. In: *Computers & structures* 89.1 (2011), pp. 27–37.

45. I.A. Ibragimov and Y.A. Rozanov. *Gaussian random processes*. English. Vol. 9. Springer Science & Business Media, 2012.

46. T. Ishigami and T. Homma. “An importance qualification technique in uncertainty analysis for computer models”. English. In: *Proceedings of the Isuma’90, First International Symposium on Uncertainty Modelling and Analysis*. 1990.

47. M. Jones et al. “A manufacturing technology readiness impact assessment transitional framework”. English. In: *Aerospace Conference, 2012 IEEE*. IEEE. 2012, pp. 1–9.

48. K. Kandasamy et al. “Gaussian Process Optimisation with Multi-fidelity Evaluations”. English. In: *NIPS*. 2016, pp. 992–1000.

49. C.G. Kaufman and B.A. Shaby. “The role of the range parameter for estimation and prediction in geostatistics”. English. In: *Biometrika* 100.2 (2013), pp. 473–484.

50. M.C. Kennedy and A. O’Hagan. “Bayesian calibration of computer models”. English. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63.3 (2001), pp. 425–464.

51. M.C. Kennedy and A. O’Hagan. “Predicting the output from a complex computer code when fast approximations are available”. English. In: *Biometrika* 87.1 (2000), pp. 1–13.

52. H.S. Kim, M. Koc, and J. Ni. “A hybrid multi-fidelity approach to the optimal design of warm forming processes using a knowledge-based artificial neural network”. English. In: *International Journal of Machine Tools and Manufacture* 47.2 (2007), pp. 211–222.

53. A.N. Kolmogorov. “Interpolation and Extrapolation of Stationary Random Sequences”. English. In: *Izv. Akad. Nauk SSSR, Ser. Mat.* 5.1 (1941), pp. 3–14.

54. S. Koziel et al. “Cost-efficient electromagnetic-simulation-driven antenna design using co-Kriging”. English. In: *Microwaves, Antennas & Propagation, IET* 6.14 (2012), pp. 1521–1528.

55. S. Koziel et al. “Efficient multi-objective simulation-driven antenna design using co-kriging”. English. In: *Antennas and Propagation, IEEE Transactions on* 62.11 (2014), pp. 5900–5905.

56. S. Kucherenko et al. “Derivative based global sensitivity measures and their link with global sensitivity indices”. English. In: *Mathematics and Computers in Simulation* 79.10 (2009), pp. 3009–3017.

57. S. Kumar, M. Mohri, and A. Talwalkar. “Sampling methods for the Nyström method”. English. In: *Journal of Machine Learning Research* 13 (2012), pp. 981–1006.

58. Y. Kuya et al. “Multifidelity surrogate modeling of experimental and computational aerodynamic data sets”. English. In: *AIAA journal* 49.2 (2011), pp. 289–298.

59. L. Le Cam and G.L. Yang. *Asymptotics in statistics: some basic concepts.* English. Springer Science and Business Media, 2000.

60. L. Le Gratiet and C. Cannamela. “Cokriging-based sequential design strategies using fast cross-validation techniques for multi-fidelity computer codes”. English. In: *Technometrics* 57.3 (2015), pp. 418–427.

61. L. Le Gratiet and J. Garnier. “Recursive co-kriging model for design of computer experiments with multiple levels of fidelity”. English. In: ().

62. S.N. Lophaven, H.B. Nielsen, and J. Sondergaard. *Aspects of the Matlab Toolbox DACE.* English. Tech. rep. Technical University of Denmark, 2002.

63. B. MacDonald, P. Ranjan, and H. Chipman. “GPfit: An R Package for Fitting a Gaussian Process Model to Deterministic Simulator Outputs”. English. In: *Journal of Statistical Software* 64.12 (2015).

64. J.I. Madsen and M. Langthjem. “Multifidelity response surface approximations for the optimum design of diffuser flows”. English. In: *Optimization and Engineering* 2.4 (2001), pp. 453–468.

65. K.V. Mardia and R.J. Marshall. "Maximum likelihood estimation of models for residual covariance in spatial regression". English. In: *Biometrika* 71.1 (1984), pp. 135–146.

66. S.C. Mohan and D.K. Maiti. "Structural optimization of rotating disk using response surface equation and genetic algorithm". English. In: *International Journal for Computational Methods in Engineering Science and Mechanics* 14.2 (2013), pp. 124–132.

67. B. Nagy, J.L. Loeppky, and W.J. Welch. *Correlation parameterization in random function models to improve normal approximation of the likelihood or posterior*. English. Tech. rep. 229. Department of Statistics, The University of British Columbia, 2007.

68. R.M. Neal. "Monte Carlo implementation of Gaussian process models for Bayesian regression and classification". English. In: *arXiv preprint physics/9701026* (1997).

69. M. Panov and V. Spokoiny. "Finite Sample Bernstein–von Mises Theorem for Semiparametric Problems". English. In: *Bayesian Analysis* 10.3 (2015), pp. 665–710.

70. M.E. Panov and V.G. Spokoiny. "Critical dimension in the semiparametric Bernstein—von Mises theorem". English. In: *Proceedings of the Steklov Institute of Mathematics* 287.1 (2013), pp. 232–255.

71. J.-S. Park. "Optimal Latin-hypercube designs for computer experiments". English. In: *Journal of statistical planning and inference* 39.1 (1994), pp. 95–111.

72. S. Park and S. Choi. "Hierarchical Gaussian Process Regression." English. In: *ACML*. 2010, pp. 95–110.

73. F. Pascual and H. Zhang. "Estimation of linear correlation coefficient of two correlated spatial processes". English. In: *Sankhyā: The Indian Journal of Statistics* (2006), pp. 307–325.

74. A. Pepelyshev. "The role of the nugget term in the Gaussian process method". English. In: *mODa 9–Advances in Model-Oriented Design and Analysis*. Springer, 2010, pp. 149–156.

75. K.B. Petersen and M.S. Pedersen. *The Matrix Cookbook*. English. Technical University of Denmark, 2006.

76. P.Z.G. Qian and C.F.J. Wu. "Bayesian hierarchical modeling for integrating low-accuracy and high-accuracy experiments". English. In: *Technometrics* 50.2 (2008), pp. 192–204.

77. Z. Qian et al. "Building surrogate models based on detailed and approximate simulations". English. In: *Journal of Mechanical Design* 128.4 (2006), pp. 668–677.

78. J. Quinonero-Candela and C.E. Rasmussen. "A unifying view of sparse approximate Gaussian process regression". English. In: *The Journal of Machine Learning Research* 6 (2005), pp. 1939–1959.

79. P. Ranjan et al. "Follow-up experimental designs for computer models and physical processes". English. In: *Journal of Statistical Theory and Practice* 5.1 (2011), pp. 119–136.

80. C.E. Rasmussen and C.K.I. Williams. *Gaussian processes for machine learning*. English. Vol. 1. MIT press Cambridge, MA, 2006.

81. E.L. Ratkova et al. "An accurate prediction of hydration free energies by combination of molecular integral equations theory with structural descriptors". English. In: *The Journal of Physical Chemistry B* 114.37 (2010), pp. 12068–12079.

82. T.D. Robinson. "Surrogate-based optimization using multifidelity models with variable parameterization". English. PhD thesis. Massachusetts Institute of Technology, 2007.

83. T.D. Robinson et al. "Surrogate-based optimization using multifidelity models with variable parameterization and corrected space mapping". English. In: *AIAA Journal* 46.11 (2008), pp. 2814–2822.

84. J. Ronkkonen and J. Lampinen. "An extended mutation concept for the local selection based differential evolution algorithm". English. In: *Genetic and Evolutionary Computation Conference*. 2007.

85. A. Saltelli and I.M. Sobol. "About the use of rank transformation in sensitivity analysis of model output". English. In: *Reliab. Eng. Syst. Safety* 50.3 (1995), pp. 225–239.

86. B. Shaby and D. Ruppert. "Tapered covariance: Bayesian estimation and asymptotics". English. In: *Journal of Computational and Graphical Statistics* 21 (2012), pp. 433–452.

87. J.Q. Shi, R. Murray-Smith, and D.M. Titterington. "Hierarchical Gaussian process mixtures for regression". English. In: *Statistics and Computing* 15.1 (2005), pp. 31–41.

88. T. Simpson et al. "Design and analysis of computer experiments in multidisciplinary design optimization: a review of how far we have come or not". English. In: *12th AIAA/ISSMO multidisciplinary analysis and optimization conference*. Vol. 5. 2008, pp. 10–12.

89. V. Spokoiny. "Bernstein-von Mises theorem for growing parameter dimension". English. In: *arxive.org* 1302.3430 (2013), p. 1.

90. V. Spokoiny. "Parametric estimation. Finite sample theory". English. In: *Annals of statistics* 6 (2012), pp. 2877–2909.

91. V. Spokoiny and M. Zhilova. "Sharp deviation bounds for quadratic forms". English. In: *Mathematical Methods of Statistics* 22.2 (2013), pp. 100–113.

92. M. Stein. *Interpolation of spatial data: some theory for kriging*. English. Springer Science & Business Media, 2012.

93. M.L. Stein et al. "Asymptotically efficient prediction of a random field with a misspecified covariance function". English. In: *The Annals of Statistics* 16.1 (1988), pp. 55–63.

94. G. Sun et al. “A two-stage multi-fidelity optimization procedure for honeycomb-type cellular materials”. English. In: *Computational Materials Science* 49.3 (2010), pp. 500–511.

95. Sh. Sun, J. Zhao, and J. Zhu. “A review of Nyström methods for large-scale machine learning”. English. In: *Information Fusion* 26 (2015), pp. 36–48.

96. T. Suzuki. “PAC-Bayesian bound for Gaussian process regression and multiple kernel additive model”. English. In: *Proceedings of the 25th Annual conference on Computational Learning Theory*. Vol. 22. 2012, pp. 39–60.

97. M. Talagrand. *The generic chaining: upper and lower bounds of stochastic processes*. English. Vol. 154. Springer, 2005.

98. M.E. Tipping. “Sparse Bayesian learning and the relevance vector machine”. English. In: *The journal of machine learning research* 1 (2001), pp. 211–244.

99. M.E. Titsias. “Variational learning of inducing variables in sparse Gaussian processes”. English. In: *International Conference on Artificial Intelligence and Statistics*. 2009, pp. 567–574.

100. A.W. Van der Vaart. *Asymptotic statistics*. English. Vol. 3. Cambridge university press, 2000.

101. A.W. Van der Vaart and J.H. Van Zanten. “Rates of contraction of posterior distributions based on Gaussian process priors”. English. In: *The Annals of Statistics* 36.3 (2008), pp. 1435–1463.

102. D. Velandia et al. “Maximum likelihood estimation for a bivariate Gaussian process under fixed domain asymptotics”. English. In: *arXiv preprint arXiv:1603.09059* (2016).

103. S. Wang, W. Chen, and K.L. Tsui. “Bayesian validation of computer models”. English. In: *Technometrics* 51.4 (2009), pp. 439–451.

104. L. Wasserman. *All of statistics: a concise course in statistical inference*. English. Springer, 2004.

105. N. Wiener. *Extrapolation, interpolation, and smoothing of stationary time series*. English. Vol. 2. MIT press Cambridge, MA, 1949.

106. Y. Xiong, W. Chen, and K.L. Tsui. "A new variable-fidelity optimization framework based on model fusion and objective-oriented sequential sampling". English. In: *Journal of Mechanical Design* 130.11 (2008), p. 111401.

107. Y. Xiong et al. "A better understanding of model updating strategies in validating engineering models". English. In: *Computer methods in applied mechanics and engineering* 198.15 (2009), pp. 1327–1337.

108. W. Xu et al. "Integrating seismic data in reservoir modeling: the collocated cokriging alternative". English. In: *SPE annual technical conference and exhibition*. Society of Petroleum Engineers. 1992.

109. M.K. Zahir and Z. Gao. "Variable Fidelity Surrogate Assisted Optimization Using A Suite of Low Fidelity Solvers". English. In: *Open Journal of Optimization* 1.1 (2012), pp. 8–14.

110. A. Zaytsev. "Variable Fidelity Regression Using Low Fidelity Function Blackbox and Sparsification". English. In: *Conformal and Probabilistic Prediction with Applications*. Springer, 2016, pp. 147–164.

111. A.A. Zaytsev, E.V. Burnaev, and V.G. Spokoiny. "Properties of the Bayesian Parameter Estimation of a Regression Based on Gaussian Processes". English. In: *Journal of Mathematical Sciences* 203.6 (2014), pp. 789–798.

112. H. Zhang, W. Cai, et al. "When Doesn't Cokriging Outperform Kriging?" English. In: *Statistical Science* 30.2 (2015), pp. 176–180.

113. Y. Zhang, J. Duchi, and M. Wainwright. "Divide and conquer kernel ridge regression: A distributed algorithm with minimax optimal rates". English. In: *Journal of Machine Learning Research* 16 (2015), pp. 3299–3340.

114. R. Zimmermann. "Asymptotic behavior of the likelihood function of covariance matrices of spatial Gaussian processes". English. In: *Journal of Applied Mathematics* 2010 (2011), p. 1.

Публикации автора по теме диссертации

З1. Зайцев А. А., Бурнаев Е. В., Спокойный В. Г. Свойства байесовской оценки параметров регрессии на основе гауссовских процессов // Фундаментальная и прикладная математика. 2013. Т. 18, № 2. С. 53–65.

З2. Зайцев А. А., Бурнаев Е. В., Спокойный В. Г. Свойства апостериорного распределения модели зависимости на основе гауссовских случайных полей // Автоматика и телемеханика. 2013, Т. 74, № 10, С. 55–67.

З3. Бурнаев Е. В., Зайцев А. А., Спокойный В. Г. Теорема Бернштейна–фон Мизеса для регрессии на основе гауссовских процессов // Успехи математических наук. 2013. Т. 68, № 5. С. 179–180.

З4. Burnaev E.V., Zaytsev A.A. Surrogate modeling of multifidelity data for large samples // Journal of Communications Technology and Electronics, 2015. V. 60, № 12. P. 1348–1355.

З5. Zaytsev A. Variable fidelity regression using low fidelity function blackbox and sparsification // Lecture Notes in Computer Science, 2016. V. 9653, P. 147–164.

З6. Zaytsev A. Reliable surrogate modeling of engineering data with more than two levels of fidelity // IEEE ICMAE conference, 2016. V. 9653, P. 341–345.

З7. Зайцев А. Ошибка интерполяции для регрессии на основе данных разной точности // ИТИС 2016, 40-я междисциплинарная школа-конференция, 15-20 сентября, 2016. С. 289–294.