

Барсуков Егор Вячеславович

Выпускная квалификационная работа

**ОПТИМАЛЬНЫЕ ПЛАНЫ ДЛЯ ОЦЕНИВАНИЯ ПРОИЗВОДНЫХ В
ПОЛИНОМИАЛЬНОЙ РЕГРЕССИОННОЙ МОДЕЛИ БЕЗ СВОБОДНОГО
ЧЛЕНА**

Уровень образования: бакалавриат

Направление 01.03.02 «Прикладная математика и информатика»

Основная образовательная программа СВ.5004.2016 «Прикладная математика и
информатика»

Профиль «Вычислительная стохастика и статистические модели»

Научный руководитель:

Профессор кафедры статистического
моделирования

д. ф.-м. н., профессор В. Б. Мелас

Рецензент:

Профессор кафедры математического
обеспечения и применения ЭВМ,

Санкт-Петербургский государственный
электротехнический университет «ЛЭТИ»

д. т. н., профессор Ю. Д. Григорьев

Санкт-Петербург

2020

Оглавление

Введение	3
Глава 1. С-оптимальные планы эксперимента	5
1.1. Определения	5
1.2. Теорема Элвинга	6
1.3. Явная формула для весов оптимального плана	6
Глава 2. Оптимальный план на отрезке с началом в нуле для оценки производной	7
2.1. Промежуток $[0, 1]$	7
2.2. Произвольный промежуток с началом в нуле	11
2.3. Пример для полиномиальной модели третьей степени	11
Глава 3. Численное нахождение с-оптимальных планов	15
3.1. Описание алгоритма	15
3.1.1. Решение задачи оптимизации	16
3.1.2. Проверка критерия оптимальности	16
3.2. Пример нахождения оптимальных планов для оценки производной	17
Заключение	17
Список литературы	20

Введение

Рассмотрим регрессионную модель

$$y_j = \theta^\top f(x_j) + \varepsilon_j, \quad j = 1 \dots N, \quad x_j \in \mathcal{X}, \quad (1)$$

где N — количество экспериментов, $\mathcal{X} \subset \mathbb{R}$, $f(x) = (f_1(x), \dots, f_n(x))^\top$ — регрессионная функция, $\theta = (\theta_1, \dots, \theta_n)^\top$ — неизвестные параметры, ε_i — некоррелированные ошибки наблюдения. При этом $E[\varepsilon_i] = 0$, $D[\varepsilon_i] = \sigma^2$.

Для того, чтобы при фиксированном количестве наблюдений N получить наилучшую в каком-либо смысле оценку θ строят *планы эксперимента* т.е. наборы точек $x_i \in \mathcal{X}$, в каждой из которых должно быть произведено n_i экспериментов так, что $\sum_i^N n_i = N$. В каком именно смысле будет улучшена оценка параметров модели зависит от выбора одного из нескольких критериев оптимальности.

В работе будут рассматриваться полиномиальные регрессионные модели без нулевого члена, при этом регрессионная функция имеет вид $f(x) = (x, \dots, x^n)^\top$. Для таких моделей во многих случаях явным образом были описаны оптимальные планы. Несколько работ были посвящены нахождению D-оптимальных планов [1–4]. Также существуют для такой модели явные решения для нахождения E-оптимальных планов [5–8].

В этой работе будут рассматриваться c -оптимальные планы эксперимента. Ими называются планы, минимизирующие дисперсию значения скалярного произведения θ и c для заданного $c \in \mathbb{R}^n$ [8]. В общем случае нахождение c -оптимальных планов может быть достаточно сложным: для случаев малой размерности решение можно найти используя теорему Элвинга [9], однако явного решения для произвольного c не существует.

В практических приложениях важны частные случаи: $c = f(z)$ для некоторого $z \notin \mathcal{X}$ — задача экстраполяции в точке z и $c = f'(z)$ — задача оценки производной в точке z . Для обычной полиномиальной модели оптимальный план экстраполяции был получен достаточно давно [10], также существует несколько явных решений для задачи оценки производной в некоторых случаях [11, 12].

В этой работе рассмотрен случай нахождения планов для оценки производной при полиномиальной модели без свободного члена при $\mathcal{X} = [0, d]$. Такая модель, например, может быть использована в тех случаях, когда, исходя из практической задачи,

z может принимать только положительные значения и существует априорное знание о значении функции в точке 0. Простым примером такой функции является зависимость расстояния от начальной точки от времени.

Также в главе 3 построен алгоритм нахождения s -оптимальных планов в общем случае, который будет работать и для произвольных s и для произвольных регрессионных функций, что позволяет применять s -оптимальные планы в различных практических задачах, для которых аналитические результаты либо недоступны, либо трудно применимы.

Глава 1

С-оптимальные планы эксперимента

1.1. Определения

Определение 1. Согласно [13] непрерывным планом эксперимента в регрессионной модели (1) будем называть дискретную вероятностную меру

$$\xi = \begin{pmatrix} x_1 & \dots & x_m \\ \omega_1 & \dots & \omega_m \end{pmatrix}, \quad x_i \in \mathcal{X},$$

где $\omega_i \geq 0$, $\sum_i^n \omega_i = 1$, $i = 1, \dots, m$.

Для проведения N измерений с мерой ξ необходимо провести $n_i \approx N\omega_i$ измерений в точке x_i таким образом, чтобы $\sum_i^m n_i = N$.

Определение 2. Информационной матрицей для непрерывного плана эксперимента, заданного мерой ξ , является

$$M(\xi) = \int_{\mathcal{X}} f(x)f^{\top}(x)\xi(dt).$$

Определение 3. *c-оптимальным планом* для некоторого вектора c называется план эксперимента ξ , минимизирующий функцию

$$\Phi(\xi) = \begin{cases} c^{\top} M(\xi)^{-} c, & \text{если существует } v, \text{ такой, что } c = M(\xi)v \\ +\infty, & \text{иначе} \end{cases}, \quad (1.1)$$

где $M(\xi)^{-}$ — обобщенно обратная матрица к информационной матрице плана ξ . План называется допустимым, если существует такой v , что $c = M(\xi)v$.

Как было отмечено во введении, *c-оптимальный план* минимизирует дисперсию несмещенной оценки по методу наименьших квадратов $c^{\top} \hat{\theta}$ линейной комбинации $c^{\top} \theta$ [8].

Определение 4. Если $c = f'(z)$ для некоторого $z \in \mathbb{R}$, то соответствующий *c-оптимальный план* называется *оптимальным планом для оценки производной* в точке z .

1.2. Теорема Элвинга

Для решения задачи нахождения c -оптимальных планов во множестве случаев (в том числе в данной работе) используется теорема Элвинга, являющаяся критерием c -оптимальности плана эксперимента.

Теорема 1. (Элвинга) [11] Допустимый план ξ^* с носителем $x_1, \dots, x_m \in \mathcal{X}$ и весами $\omega_1, \dots, \omega_m$ является c -оптимальным тогда и только тогда, когда существует $p \in \mathbb{R}^n$ и константа h такие, что выполняются следующие условия:

$$|p^\top f(x_i)| = 1, \quad i = 1..m \leq n, \quad (1.2a)$$

$$|p^\top f(x)| \leq 1, \quad x \in \mathcal{X}, \quad (1.2b)$$

$$c = h \sum_{i=1}^m \omega_i f(x_i) p^\top f(x_i). \quad (1.2c)$$

Кроме того

$$h^2 = c^\top M^-(\xi^*)c.$$

Функция $p^\top f(x)$ в обозначениях теоремы Элвинга в этой работе также будет называться *экстремальным многочленом*.

1.3. Явная формула для весов оптимального плана

Теорема 2. [12] Оптимальный план для оценивания производной полиномиальной модели без свободного члена с опорными точками t_1^*, \dots, t_m^* , где $m = n$ или $m = n - 1$ имеет веса, вычисленные по следующей формуле:

$$\omega_i = \frac{|L'_i(z)|}{\sum_{j=1}^m |L'_j(z)|}, \quad (1.3)$$

где L_i задается следующим образом

$$L_i(x) = \frac{x \prod_{l \neq i} (x - t_l^*)}{t_i^* \prod_{l \neq i} (t_i^* - t_l^*)}, \quad (1.4)$$

то есть является многочленом без нулевого члена степени m , который равен 1 в точке t_i^* и нулю в остальных опорных точках. Такой многочлен далее в этой работе будет называться i -ым базисным многочленом Лагранжа без нулевого члена построенным по точкам t_1^*, \dots, t_m^* .

Глава 2

Оптимальный план на отрезке с началом в нуле для оценки производной

В этой главе описаны оптимальные планы на отрезке $[0, d]$ для оценки производной в модели без нулевого члена. Как было отмечено ранее, необходимость в таких планах может возникать в некоторых практических задачах. С точки зрения решения эта задача отличается от случая носителя $[-1, 1]$, которая была рассмотрена в [12], тем, что экстремальный многочлен должен обладать другими свойствами. В разделе 2.1 рассмотрен случай отрезка $[0, 1]$, а в разделе 2.2 рассмотрены отличия для промежутка общего вида $[0, d]$, в конце главы приведен пример применения доказанной теоремы.

Везде в этой главе считается, что $f(x) = (x, x^2, \dots, x^n)^\top$, то есть рассматривается модель без нулевого члена.

2.1. Промежуток $[0, 1]$

При доказательстве основной теоремы этой главы будет использована следующая лемма.

Лемма 1. Пусть $P_1(x)$ и $P_2(x)$ — многочлены степени n с корнями $t_1^1 < \dots < t_n^1$ и $t_1^2 < \dots < t_n^2$ соответственно. При этом корни располагаются следующим образом:

$$t_1^1 \leq t_1^2 \leq t_2^1 \leq t_2^2 \leq \dots \leq t_n^1 \leq t_n^2,$$

где хотя бы одно из неравенств $t_l^1 \leq t_l^2$ ($l = 1, \dots, n$) является строгим. Также обозначим корни многочленов $P_1'(x)$ и $P_2'(x)$ как v_1^1, \dots, v_{n-1}^1 и v_1^2, \dots, v_{n-1}^2 . Тогда справедливо следующее утверждение

$$v_1^1 < v_1^2 < v_2^1 < v_2^2 < \dots < v_{n-1}^1 < v_{n-1}^2.$$

Доказательство этого утверждения можно найти в [14] или в приложении к [12].

Следующая лемма дает вид экстремального многочлена, необходимого для проверки критерия s -оптимальности для промежутка $[0, 1]$ в полиномиальной модели без нулевого члена.

Лемма 2. Пусть $T_n(x)$ — многочлен Чебышева первого рода степени n . Тогда

$$S_n(x) = T_n\left(x\left(1 + \cos \frac{\pi}{2n}\right) - \cos \frac{\pi}{2n}\right) \quad (2.1)$$

является многочленом степени n без нулевого члена, который на промежутке $[0, 1]$ не превосходит по модулю единицу, при этом равенство достигается ровно в n точках x_i^* :

$$x_i^* = \frac{\cos \frac{(n-i)\pi}{n} + \cos \frac{\pi}{2n}}{1 + \cos \frac{\pi}{2n}}, \quad i = 1, \dots, n. \quad (2.2)$$

Доказательство. Известно, что многочлены Чебышева первого рода не превосходят по модулю единицу на отрезке $[-1, 1]$ и при этом равенство достигается в $n + 1$ точках. По построению $S_n(x)$ удовлетворяет этим же свойствам, но на отрезке $[-\cos \frac{\pi}{2n}, 1]$. Так как $-\cos \frac{\pi}{2n}$ — наименьший корень $T_n(x)$, и он больше всего одной его точки экстремума $\tilde{x}_0^* = -1$, то $S_n(x)$ — многочлен без нулевого члена, не превосходящий по модулю единицу на промежутке $[0, 1]$ и достигающий ее ровно n раз. Из известных экстремальных точек многочленов Чебышева и построения $S_n(x)$ следует, что его равенство единице по модулю достигается в (2.2). \square

Следующая теорема формулирует условия, когда многочлен (2.1) действительно определяет план для оценки производной.

Теорема 3. Пусть для $i = 1, \dots, n$ корни производной многочлена L_i из (1.4), построенного по точкам (2.2), равны u_1^i, \dots, u_{n-1}^i , причем $u_k^i < u_l^i$ при $k < l$. Тогда план с опорными точками (2.2) и весами (1.3) является оптимальным планом для оценки производной в точке z в полиномиальной модели степени n без свободного члена на промежутке $[0, 1]$, если выполнено одно из следующих условий:

- $z \in (-\infty, u_1^n)$,
- $z \in (u_{n-1}^1, +\infty)$,
- $z \in (u_j^1, u_{j+1}^n)$ для $j = 1, \dots, n - 2$.

Доказательство. Доказательство теоремы сводится к проверке условий теоремы Элвинга. Выполнение условий (1.2a) и (1.2b) следует из леммы 2, если за p взять вектор коэффициентов многочлена S_n . Поэтому для доказательства теоремы достаточно проверить условия (1.2c) при $c = f'(z)$.

Так как теорема 2 о весах оптимального плана для оценки производной в случае полиномиальной модели выполняется для любых промежутков, то веса имеют вид (1.3).

Введем обозначения $F = ((x_j^*)^i)_{i,j=1}^n$ и $\beta = (\omega_i(-1)^{i+n})_{i=1}^n$. Так как x_j^* при $i = 1, \dots, n$ являются экстремальными точками многочлена S_n и при этом $S_n(x_j^*) = (-1)^{j+n}$, то выполнение равенства для некоторого h

$$f'(z) = hF\beta \quad (2.3)$$

эквивалентно выполнению условия (1.2с) теоремы Элвинга для нахождения оптимального плана оценки производной в точке z .

Утверждение $F^{-1}F = I_n$, где I_n — единичная матрица размера n , в силу того, что i -ый столбец матрицы F на самом деле равен $f(x_i^*)$ ($i = 1, \dots, n$), можно переписать, как систему равенств

$$e_i^\top F^{-1}f(x_j^*) = \delta_{ij}, \quad i, j = 1, \dots, n, \quad (2.4)$$

где δ_{ij} — дельта Кронекера, а e_i — i -ый единичный вектор. Поскольку в левой части (2.4) содержатся многочлены без нулевого коэффициента степени не больше n , вычисленные в точках x_j^* , $j = 1, \dots, n$, а для каждого i существует только одно j , такое, что $\delta_{ij} \neq 0$, то они определяют все базисные многочлены Лагранжа без нулевого члена степени n вычисленные в точках x_j^* , $j = 1, \dots, n$, таким образом

$$e_i^\top F^{-1}f(z) = L_i(z), \quad i, j = 1, \dots, n.$$

Если в предыдущем выражении вычислить производную по z , переписать полученное выражение в векторной форме и домножить обе части равенства слева на F получим

$$f'(z) = F(L'_1(z), \dots, L'_n(z))^\top. \quad (2.5)$$

Приравняв правые части (2.5) и (2.3) и домножив обе части равенства на F^{-1} слева, получаем, что

$$h\beta = (L'_1(z), \dots, L'_n(z))^\top,$$

что с учетом известного вектора β и положив $h = \sum_{i=1}^n |L'_i(z)|$ эквивалентно тому, что $\text{sign}(L'_i(z)) = \text{sign}((-1)^{i+n})$, $i = 1, \dots, n$ или, так как экстремальным многочленом также может быть $-S(x)$, $\text{sign}(L'_i(z)) = \text{sign}((-1)^{i+n+1})$, $i = 1, \dots, n$.

Таким образом для того, чтобы доказать, что оптимальный план находится в точках $(x_i^*)_{i=1}^n$, $i = 1, \dots, n$ с указанными весами, осталось доказать равенство (или противоположность) знаков $L'_i(z)$ и $S_n(x_i^*)$ для $i = 1, \dots, n$. Но так как знаки значений

многочлена S_n в экстремальных точках чередуются, достаточно проверить при каких z выражения $(-1)^i L'_i(z)$ имеет одинаковый знак.

Так как корни многочленов L_i и L_j для любых i и j таких, что $i < j$ удовлетворяют условию леммы 1, то, последовательно её применяя ко всем базисным многочленам, получаем, что корни их производных, обозначения для которых были описаны в условии этой теоремы, удовлетворяют следующему соотношению:

$$u_1^n < u_1^{n-1} < \dots < u_1^1 < u_2^n < u_2^{n-1} < \dots < u_{n-1}^1.$$

Так как все узловые точки больше нуля, знак многочлена $L_i(z)$ при $z \rightarrow -\infty$ будет равен $(-1)^{n+i+1}$. В то же время знак $L'_i(z)$ будет противоположным $L_i(z)$, так как меняется четность многочлена и при этом не меняется знак при старшем коэффициенте, то есть $\text{sign}(L'_i(z)) = \text{sign}((-1)^{n+i})$ при $z \rightarrow -\infty$. И, следовательно, $\text{sign}((-1)^i L'_i(z)) = \text{sign}((-1)^{n+2i}) = \text{sign}((-1)^n)$ при достаточно малом z , то есть $\text{sign}((-1)^i L'_i(z))$ имеет постоянный знак для любых i , что означает, что при $z \in (-\infty, u_1^n)$ третье условие теоремы Элвинга выполняется и план является оптимальным.

На промежутках $[u_j^n, u_j^1]$ ($j = 1, \dots, n-1$) каждый базисный многочлен меняет свой знак ровно один раз и на этих промежутках знаки производных не совпадают со знаками экстремального многочлена, а на промежутках (u_j^1, u_{j+1}^n) ($j = 1, \dots, n-2$) нет ни одного корня и поэтому $\text{sign}((-1)^i L'_i(z)) = \text{sign}((-1)^{n+j})$ при $z \in (u_j^1, u_{j+1}^n)$ ($j = 1, \dots, n-2$), это подтверждает третье условие теоремы Элвинга в этих промежутках и показывает, что рассматриваемый план оптимален.

На промежутке $(-\infty, u_{n-1}^1]$ каждый базисный многочлен поменял свой знак одинаковое количество раз, а так как при $z \rightarrow -\infty$ условие выполнялось, то при $z \in (u_{n-1}^1, +\infty)$ план также является оптимальным.

Таким образом план эксперимента с опорными точками (2.2) и соответствующими им весами (1.3) является оптимальным планом для оценки производной в точке z в модели без нулевого члена тогда и только тогда, когда $z \in (-\infty, u_1^n) \cup (u_{n-1}^1, +\infty)$ или $z \in (u_j^1, u_{j+1}^n)$ для $j = 1, \dots, n-2$.

□

2.2. Произвольный промежуток с началом в нуле

В общем случае полученные условия оптимальности для всех положительных промежутков, начинающихся в нуле, существенно не отличается от единичного отрезка. Пусть требуется найти условия оптимальности плана на промежутке $[0, d]$. Возьмем многочлен $S_n^d(x) = S_n(\frac{x}{d})$, пусть \widehat{x}_i^* — его экстремальные точки. Тогда

$$\widehat{x}_i^* = dx_i^*, \quad i = 1, \dots, n \quad (2.6)$$

где x_i^* — экстремальные точки S_n из (2.2). $S_n^d(x)$ по построению по модулю не превышает единицу на промежутке $[0, d]$ и достигает единицы ровно n раз, что необходимо для выполнения первых двух условий теоремы Элвинга. Построив базисные многочлены по точкам (2.6) вместо (2.2) условия и дальнейшее доказательство теоремы 3 полностью переносятся на случай промежутка $[0, d]$.

2.3. Пример для полиномиальной модели третьей степени

С помощью доказанной теоремы найдем некоторые планы для оценки производной в модели $f(x) = (x, x^2, x^3)^\top$ с областью планирования $\mathcal{X} = [0, 1]$. Экстремальный многочлен согласно (2.1) в этом случае равен

$$\begin{aligned} S_3(x) &= T_3 \left(x \left(1 + \frac{\sqrt{3}}{2} \right) - \frac{\sqrt{3}}{2} \right) \\ &= \left(\frac{\sqrt{3}}{2} + 1 \right)^3 x^3 - 6\sqrt{3} \left(\frac{\sqrt{3}}{2} + 1 \right)^2 x^2 + 6 \left(\frac{\sqrt{3}}{2} + 1 \right) x, \end{aligned}$$

а его экстремальные точки по (2.2) равны

$$x_1^* = 3\sqrt{3} - 5, \approx 0.1962 \quad x_2^* = \sqrt{3} - 1 \approx 0.7362, \quad x_3^* = 1. \quad (2.7)$$

Тогда базисные многочлены (1.4), построенные по x_1^* , x_2^* , x_3^* , выглядят следующим образом:

$$\begin{aligned} L_1(x) &= \frac{x(x-1)(x-\sqrt{3}+1)}{6(41\sqrt{3}-71)}, \\ L_2(x) &= \frac{x(x-1)(x-3\sqrt{3}+5)}{38-22\sqrt{3}}, \\ L_3(x) &= \frac{1}{3}x \left((4\sqrt{3}+7)x^2 - 2(2\sqrt{3}+3)x + 2 \right), \end{aligned}$$

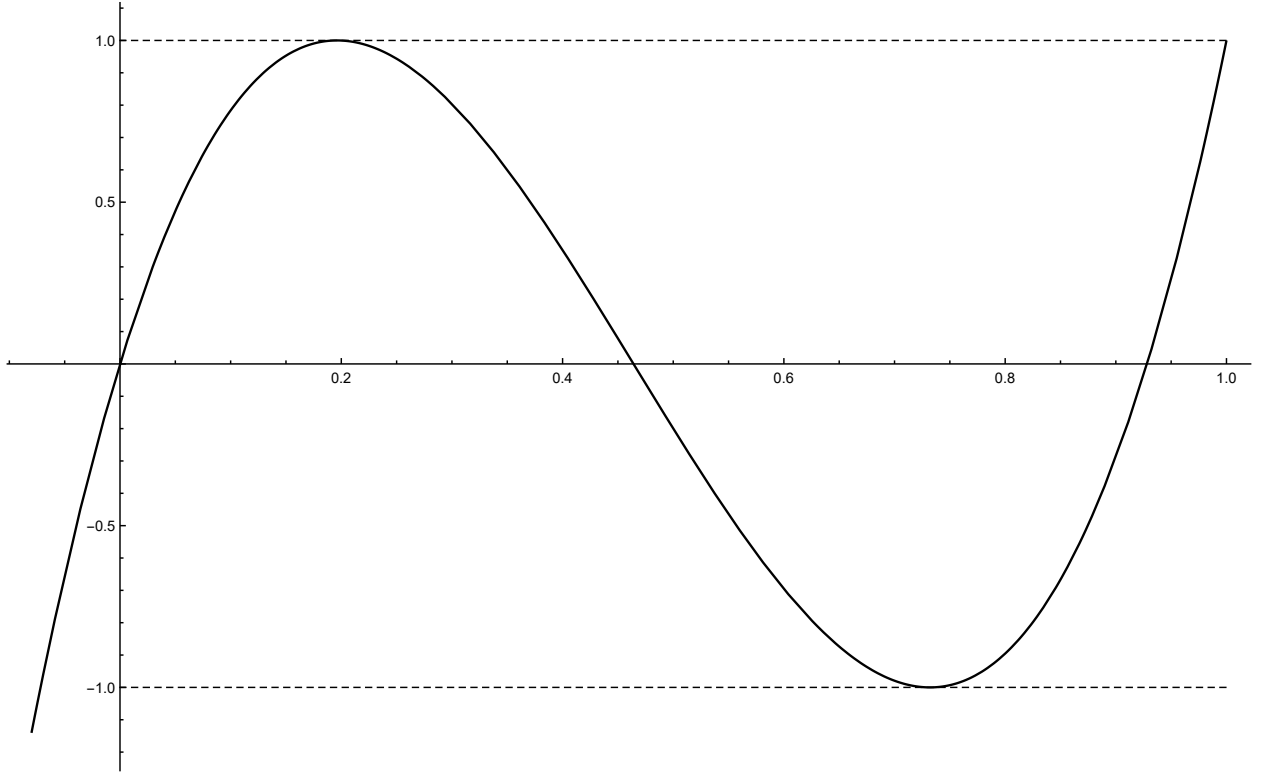


Рис. 2.1. Экстремальный многочлен для $n = 3$.

и, соответственно, их производные:

$$L_1'(x) = \frac{3x^2 - 2\sqrt{3}x + \sqrt{3} - 1}{6(41\sqrt{3} - 71)},$$

$$L_2'(x) = \frac{-3x^2 + (6\sqrt{3} - 8)x - 3\sqrt{3} + 5}{22\sqrt{3} - 38},$$

$$L_3'(x) = \left(4\sqrt{3} + 7\right)x^2 + \left(-4 - \frac{8}{\sqrt{3}}\right)x + \frac{2}{3}.$$

Из производных базисных многочленов по (1.3) можно вычислить веса плана. Поскольку из доказательства теоремы известно, что на промежутках, где выполняется условие теоремы, знаки производных базисных многочленов чередуются, то для более простого явного вида формулу для весов можно переписать как

$$\omega_i = \left| \frac{L_i'(z)}{\sum_{j=1}^m (-1)^j L_j'(z)} \right|.$$

С учетом этого, веса можно явно выписать:

$$\begin{aligned}\omega_1 &= \frac{1}{12} (41\sqrt{3} + 71) \left| \frac{3z^2 - 2\sqrt{3}z + \sqrt{3} - 1}{\frac{3}{2}z (14\sqrt{3} - (15\sqrt{3} + 26)z + 24) - 3(\sqrt{3} + 2)} \right|, \\ \omega_2 &= \frac{1}{4} (11\sqrt{3} + 19) \left| \frac{z(-3z + 6\sqrt{3} - 8) - 3\sqrt{3} + 5}{\frac{3}{2}z (14\sqrt{3} - (15\sqrt{3} + 26)z + 24) - 3(\sqrt{3} + 2)} \right|, \\ \omega_3 &= \left| \frac{z \left((4\sqrt{3} + 7)z - \frac{8}{\sqrt{3}} - 4 \right) + \frac{2}{3}}{\frac{3}{2}z (- (15\sqrt{3} + 26)z + 14\sqrt{3} + 24) - 3(\sqrt{3} + 2)} \right|.\end{aligned}\quad (2.8)$$

Теперь осталось показать, при каких z план с опорными точками (2.7) и весами (2.8) является оптимальным планом для оценки производной в точке z . Для этого найдем корни производных базисных многочленов (здесь для них используются обозначения из теоремы).

$$\begin{aligned}u_1^1 &= \frac{1}{3} \left(\sqrt{3} - \sqrt{3(2 - \sqrt{3})} \right), & u_2^1 &= \frac{1}{3} \left(\sqrt{3} + \sqrt{3(2 - \sqrt{3})} \right), \\ u_1^2 &= \frac{1}{3} \left(3\sqrt{3} - \sqrt{58 - 33\sqrt{3}} - 4 \right), & u_2^2 &= \frac{1}{3} \left(3\sqrt{3} + \sqrt{58 - 33\sqrt{3}} - 4 \right), \\ u_1^3 &= \frac{1}{3} \left(4\sqrt{3} - \sqrt{2(21 - 12\sqrt{3})} - 6 \right), & u_2^3 &= \frac{1}{3} \left(4\sqrt{3} + \sqrt{2(21 - 12\sqrt{3})} - 6 \right).\end{aligned}$$

Их приближенные значения:

$$\begin{aligned}u_1^1 &\approx 0.2784, & u_2^1 &\approx 0.8762, \\ u_1^2 &\approx 0.0927, & u_2^2 &\approx 0.7046, \\ u_1^3 &\approx 0.0906, & u_2^3 &\approx 0.5281.\end{aligned}$$

По теореме 3 план с опорными точками (2.7) и весами (2.8) является оптимальным планом для оценки производной если $z < u_1^3$, $z > u_2^1$ или $u_1^1 < z < u_2^3$.

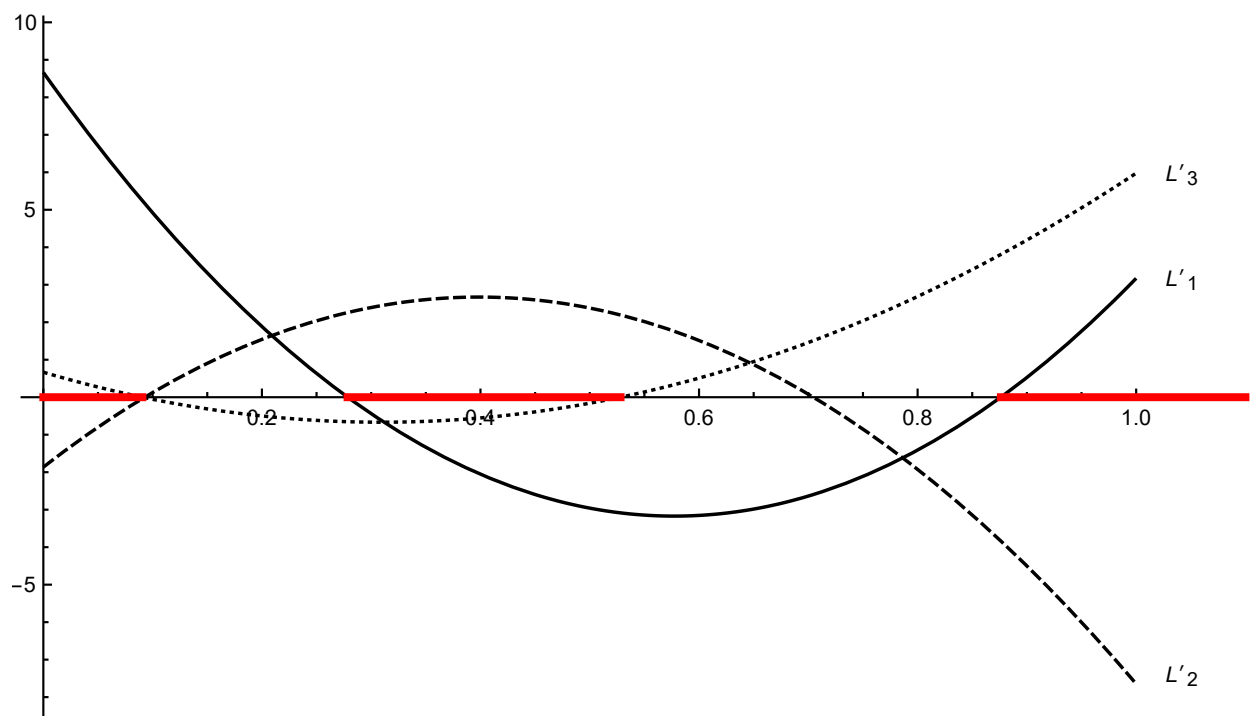


Рис. 2.2. Производные базисных многочленов. Красным выделены промежутки, на которых выполняется теорема.

Глава 3

Численное нахождение s -оптимальных планов

Как было показано в предыдущей главе, задача нахождения s -оптимальных планов далеко не всегда является тривиальной. Для многих случаев нет явных аналитических решений, а, даже там, где они существуют, не всегда просто их применять. В этой главе описан алгоритм численного нахождения и проверки s -оптимальных планов для произвольного s , работающий для модели общего вида.

3.1. Описание алгоритма

Пусть стоит задача построить s -оптимальный план для некоторого вектора s , в модели, заданной вектор-функцией $f(x) = (f_1, \dots, f_n)^\top$, и область планирования $\mathcal{X} = [a, b]$ является конечным отрезком.

Определение s -оптимального плана уже содержит в себе задачу оптимизации, поэтому для нахождения кандидата в оптимальные планы можно просто минимизировать функцию (1.1). Однако нет гарантии, что численная процедура сойдется к нужному плану, поэтому для финального ответа необходимо удостовериться, что найденный план является s -оптимальным. Для этого нужно приспособить критерий s -оптимальности — теорему Элвинга — для численной проверки. В алгоритме 4 дается схема такого подхода: в цикле будет стартовать алгоритм численной оптимизации со случайной начальной точкой, пока он не сойдется к решению.

Алгоритм 1: Общая схема алгоритма

Результат: ξ **1** повторять**2** $\xi \leftarrow$ случайный план в области планирования;**3** $\xi \leftarrow$ результат оптимизации (1.1) с начальной точкой ξ **4** до тех пор, пока ξ не s -оптимальный;

Остаются вопросы: как оптимизировать такую функцию и как численно проверять условия теоремы Элвинга?

3.1.1. Решение задачи оптимизации

Высокой скоростью сходимости обладают методы первого и второго порядка, однако для их корректной работы и сходимости к локальному минимуму необходимо наличие производных соответствующего порядка. Известно [15], что если ранг матрицы $A(x)$ не зависит от x , то существует явное выражение для производной $A^-(x)$, причем оно является непрерывным:

$$\frac{d}{dx}A^-(x) = -A^- \left(\frac{d}{dx}A \right) A^- + A^- A^{-T} \left(\frac{d}{dx}A^T \right) (1 - AA^-) + (1 - A^-A) \left(\frac{d}{dx}A^T \right) A^{-T} A^-.$$

Поэтому для плана фиксированного размера (при $x_i^* \neq x_j^*, i \neq j$) можно воспользоваться методами выпуклой оптимизации с ограничениями. При сравнении нескольких разных алгоритмов наилучшую скорость сходимости показал квазиньютоновский алгоритм оптимизации с ограничениями L-BFGS-B [16].

Важным ограничением при таком подходе является постоянный размер плана. Поэтому для того, чтобы алгоритм мог сходиться к решениям меньших размеров, инициализацию случайных планов из алгоритма 4 необходимо сделать случайной не только по размещению опорных точек и весам, но и по размеру его носителя.

Алгоритм L-BFGS-B имеет ограничения типа “коробка”, что подходит для ограничения области планирования, однако не вполне подходит для весов, которые должны равняться в сумме единице. Чтобы разрешить это, на веса были ограничены снизу нулем, а сверху единицей, но в целевой функции перед ее вычислением осуществляется их нормирование.

При построении таблицы планов для последовательно возрастающих точек, в которых оценивается производная, можно в качестве начального приближения использовать план, построенный для предыдущей точки. Этот прием позволяет находить оптимальные планы значительно быстрее и не требует многократных вычислений для различных начальных планов.

3.1.2. Проверка критерия оптимальности

Для проверки s -оптимальности плана ξ с помощью теоремы 1 необходимо сначала найти вектор p такой, что выполняются условия (1.2a) и (1.2b). Из этих условий, в частности, следует, что опорные точки, находящиеся во внутренности области \mathcal{X} являются экстремумами функции $f(x)^\top p$. Добавив эти условия к (1.2a) можно получить систему

уравнений, определяющую p :

$$\begin{pmatrix} f_1(x_1) & \dots & f_n(x_1) \\ f_1(x_2) & \dots & f_n(x_2) \\ \vdots & \ddots & \vdots \\ f_1(x_n) & \dots & f_n(x_n) \\ f'_1(x_{i_1}) & \dots & f'_n(x_{i_1}) \\ \vdots & \ddots & \vdots \\ f'_1(x_{i_m}) & \dots & f'_n(x_{i_m}) \end{pmatrix} p = \begin{pmatrix} 1 \\ s_2 \\ \vdots \\ s_n \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad (3.1)$$

где i_1, \dots, i_m — индексы опорных точек, лежащих во внутренности \mathcal{X} , а $s_j = \pm 1$. Обозначим P множество всех векторов, являющихся решениями системы (3.1) для $s_j \in \{\pm 1\}$. По построению каждый элемент P удовлетворяет условиям (1.2a) и (1.2b).

Тогда, согласно теореме 1, если среди элементов P найдется такой p , и будет существовать такая константа h , что выполнится условие (1.2c), то план ξ будет являться c -оптимальным. Несмотря на то, что для задачи нахождения h существует явное решение, для упрощения вычислений его значение можно не считать, а просто проверять коллинеарность вектора в правой части (1.2c) и вектора c .

3.2. Пример нахождения оптимальных планов для оценки производной

Реализация алгоритма была протестирована на уже рассмотренном в разделе 2.3 примере: $f(x) = (x, x^2, x^3)^\top$, $\mathcal{X} = [0, 1]$, $c = f'(z)$ для $z \in [0, 1]$.

В таблице 3.1 представлены результаты работы алгоритма. Выделенные строки обозначают те планы, которые были численно найдены для случаев, когда оптимальные планы уже были аналитически построены. Можно заметить, что численные результаты совпадают с описанными ранее аналитическими. Также видно, что на промежутках, где не было найдено аналитическое решение, оптимальный план состоит из двух точек.

Алгоритм численной оптимизации сходится за 5-8 итераций. В случае трехточечного оптимального плана алгоритм сходится с первого раза к c -оптимальному плану. Если количество точек в плане не равно трем, то алгоритм первые 2-10 попыток попадает в локальный минимум, который не является оптимальным планом.

Таблица 3.1. Параметры планов для оценивания производных в точке z в полиномиальной модели третьей степени без нулевого члена с областью планирования $[0, 1]$

z	x_1	x_2	x_3	w_1	w_2	w_3
0.05	0.196	0.732	1.000	0.862	0.103	0.035
0.10	0.215	0.804		0.993	0.007	
0.15	0.332	1.000		0.985	0.015	
0.20	0.467	1.000		0.951	0.049	
0.25	0.625	1.000		0.865	0.135	
0.30	0.196	0.732	1.000	0.126	0.684	0.190
0.35	0.196	0.732	1.000	0.292	0.568	0.141
0.40	0.196	0.732	1.000	0.389	0.506	0.105
0.45	0.196	0.732	1.000	0.465	0.465	0.070
0.50	0.196	0.732	1.000	0.538	0.423	0.092
0.55	0.204	0.762		0.585	0.415	
0.60	0.223	0.832		0.585	0.415	
0.65	0.241	0.901		0.585	0.415	
0.70	0.260	0.970		0.585	0.415	
0.75	0.375	1.000		0.542	0.458	
0.80	0.533	1.000		0.513	0.487	
0.85	0.668	1.000		0.504	0.496	
0.90	0.196	0.732	1.000	0.057	0.488	0.455
0.95	0.196	0.732	1.000	0.137	0.469	0.394
1.00	0.196	0.732	1.000	0.189	0.455	0.355

Заключение

В работе был получен результат о явном виде некоторых оптимальных планов для оценки производной в полиномиальной модели без нулевого члена для интервала планирования на положительной полуоси с началом в нуле. Был построен пример применения этого результата в случае полинома третьей степени.

Также был построен алгоритм численного нахождения s -оптимальных планов в общем виде и показаны результаты его работы в случае, который до этого был рассмотрен с помощью аналитических методов.

Список литературы

1. Hoel P. G. Efficiency problems in polynomial estimation // Annals of Mathematical Statistics. — 1958. — Vol. 29, no. 4. — P. 1134–1145.
2. Studden W. J. D_s -optimal designs for polynomial regression using continued fractions // Annals of Statistics. — 1980. — Vol. 8, no. 5. — P. 1132–1141.
3. Dette H. A generalization of D- and D_1 -optimal designs in polynomial regression // Annals of Statistics. — 1990. — Vol. 18. — P. 1784–1805.
4. Dette H., Franke T. Robust designs for polynomial regression by maximizing a minimum of D- and D_1 -efficiencies // Annals of Statistics. — 1990. — Vol. 29, no. 4. — P. 1024–1049.
5. Pukelsheim F., Studden W. J. E-optimal designs for polynomial regression // Annals of Statistics. — 1990. — Vol. 21, no. 1. — P. 402–415.
6. Dette H. A note on E-optimal designs for weighted polynomial regression // Annals of Statistics. — 1990. — Vol. 21, no. 2. — P. 767–771.
7. Heiligers B. E-optimal designs in weighted polynomial regression // Annals of Statistics. — 1994. — Vol. 22, no. 2. — P. 917–929.
8. Dette H., Studden W. J. Geometry of E-optimality // Annals of Statistics. — 1993. — Vol. 21, no. 1. — P. 416–433.
9. Elfving G. Optimal allocation in linear regression theory // The Annals of Mathematical Statistics. — 1952. — Vol. 23. — P. 255–262.
10. Hoel P. G., Levine A. Optimal spacing and weighting in polynomial prediction // The Annals of Statistics. — 1964. — Vol. 35, no. 4. — P. 1553–1560.
11. Dette H., Melas V., A. Pepelyshev. Optimal designs for estimating the slope of a regression // Statistics. — 2010. — Vol. 44, no. 6. — P. 617–628.
12. Dette H., Melas V., Shpilev P. Some explicit solutions of c-optimal design problems for polynomial regression with no intercept // Annals of the Institute of Statistical Mathematics. — 2019.
13. Kiefer J. General Equivalence Theory for Optimum Designs (Approximate Theory) // The Annals of Statistics. — 1974. — Vol. 2, no. 5. — P. 416–433.
14. Sahm M. Optimal designs for estimating individual coefficients in polynomial regression : Ph.D. thesis / M. Sahm ; Fakultät für Mathematik, Ruhr-Universität Bochum, Germany. — 1998.

15. Golub Gene H, Pereyra Victor. The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate // SIAM Journal on numerical analysis. — 1973. — Vol. 10, no. 2. — P. 413–432.
16. L-BFGS-B - Fortran Subroutines for Large-Scale Bound Constrained Optimization : Rep. / ACM Trans. Math. Software ; Executor: Ciyou Zhu, Richard H. Byrd, Pei-huang Lu, Jorge Nocedal : 1994.