Краткие сообщения

УДК 517.977.58 DOI: 10.14529/ctcr210314

ИССЛЕДОВАНИЕ ПАРАМЕТРОВ ЗАДАЧИ СПЛАЙНОВОЙ АППРОКСИМАЦИИ ЗАШУМЛЕННЫХ ДАННЫХ ЧИСЛЕННЫМИ МЕТОДАМИ ОПТИМАЛЬНОГО УПРАВЛЕНИЯ

И.П. Болодурина, Л.С. Гришина, Л.М. Анциферова
Оренбургский государственный университет, г. Оренбург, Россия

В настоящее время проблемы искажения данных измерений шумом и появления неопределенностей в критериях качества послужили причиной повышенного интереса к исследованиям в области сплайновой аппроксимации. При этом существующие методы минимизации эмпирического риска, предполагая, что шум является равномерным распределением с более тяжелыми хвостами, чем гауссов, ограничивают области применения данных исследований. Проблема оценки искаженных шумом данных, как правило, основывается на решении оптимизационной задачи с функцией, содержащей неопределенность, возникающей на основе задачи поиска оптимальных параметров. В связи с этим оценка искаженных шумов не может быть разрешена классическими методами. **Цель исследования.** Данное исследование направлено на решение и анализ задачи сплайновой аппроксимации данных в условиях неопределённости на основе параметризации управления и алгоритма проекции градиента. **Методы.** Исследование задачи сплайновой аппроксимации зашумленных данных проведено методом приближения кусочно-постоянной функции управления. При этом параметризация управления возможна только при конечном числе точек разрыва первого рода. В рамках экспериментального исследования применен алгоритм метода проекции градиента для численного решения задачи сплайновой аппроксимации. Предложенные методы применены для исследования параметров задачи сплайновой аппроксимации данных в условиях неопределённости. **Результаты.** Численное исследование подхода к параметризации управления и алгоритма проекции градиента проведено на основе разработанного программно-алгоритмического средства для решения задачи сплайновой модели аппроксимации в условиях неопределенности. Для оценки искаженных шумом данных проведены численные эксперименты по исследованию параметров модели и установлено, что повышение значения параметра α ведёт к увеличению точности, но к потере гладкости. Кроме того, проведенный анализ показал, что рассмотренные законы распределения не изменили точность и скорость сходимости алгоритма. **Заключение.** Предложенный подход для решения задачи сплайновой аппроксимации в условиях неопределенности позволяет определить проблемы искажения данных измерений шумом и появления неопределенностей в критериях качества. Исследование параметров модели показало, что построенная система устойчива к ошибке начального приближения, а законы распределения не оказывают существенного влияния на точность и сходимость метода проекции градиента.

Ключевые слова: цитирования, наукометрические методы, агрегирование библиографической информации, модификация метода Winnowing, метод Левенштейна, метод шинглов.

Введение

В настоящее время исследования в области сплайновой аппроксимации стали наиболее актуальными в прикладных задачах. Например, сплайны играют ведущую роль в качестве генераторов контуров или кривых, а сглаживающие сплайны достаточно часто применяются в статистике. Классические сплайны представляют интерполяционные кривые, в то время как сплайны сглаживания удовлетворяют условию аппроксимации достаточно «близко» к точкам исходных данных. В рамках данной работы сглаживающие сплайны представлены как естественная часть теории оптимального управления (ОУ). При этом теоретические концепции управления позволяют находить и интерпретировать сплайны наиболее эффективно.

Исходная форма полиномиальной интерполяции содержит недостатки, которые ограничивают ее применение во многих сферах деятельности, в частности, для многоразмерных задач.

Основные недостатки исправляет теория сплайнов, в связи с этим рассмотрение вопросов применения сплайнов в различных областях является актуальным вопросом.

В рамках данного исследования рассмотрена линейная квадратичная задача ОУ, для которой не задано начальное состояние системы, однако известны априорные сведения о данной системе. При решении задачи о поиске данных параметров с наиболее эффективным значением функционала качества по набору измеренных данных, необходимо учитывать шум и искажения данных. Данная работа рассматривает частный пример задачи ОУ – задачу аппроксимации сплайновой. Функционал качества вычисляется на основе входных данных и, следовательно, содержит искаженную информацию. В связи с этим данная задача относится к задаче с неопределенностью.

1. Обзор исследований

Исследованиями и разработкой методов аппроксимации сплайн-функций с целью моделирования взаимосвязи между целевым откликом и несколькими переменными предикторами занимаются по всему миру.

Сплайн широко используется в обработке сигналов, дискретных вычислениях, статистике и, в частности, сплайн сглаживания позволят получить гладкую кривую, которая наилучшим образом подходит для аппроксимации данных с шумом [1, 2]. Теоретический сглаживающий сплайн в работе [3] является обобщением сглаживающего сплайна с использованием новых подходов управления, с помощью которых кривая сплайна определяется выходом линейной динамической системы.

В исследовании [4] показано, что управляющие теоретические сплайны относительно полиномиальных позволяют построить более широкий набор сглаживающих кривых. Кроме того, теоретический сплайн управления доказал свою полезность для планирования траектории мобильных роботов [5], контурного моделирования изображений [6], оценки распределения вероятностей [7] и других. Дополнительные приложения и наиболее обширные сведения о теории управления сплайнами представлены в исследовании [8].

Традиционный вид теоретических сплайнов управления основан на оптимизации L_2 и имеет два основных недостатка. Во-первых, данный подход требует количество параметров, равное размерности данных, чтобы представить подобранную кривую.

Оператор наименьшей абсолютной усадки и выбора LASSO представлен в работах [9, 10]. Методы минимизации эмпирического риска, предполагая, что шум является равномерным распределением, представлены в исследовании [11]. Затем проблема описывается в выпуклой оптимизации, которая может быть эффективно решена с помощью численных методов.

В исследовании [12] рассматривается подход к решению задачи ОУ с априорными знаниями и предположениями о распределении шума на входных данных. Данную проблему оценки параметров авторы сформулировали в виде стохастической задачи выбора оптимального параметра.

Таким образом, анализ современных исследований показал, что оценку шума исходных данных можно проводить статистическими методами. Тогда для перевода функционала качества с неопределенностью в детерминированную функцию можно использовать выпуклую комбинацию среднего значения и дисперсии исходной функции. В результате применения данного подхода и формирования детерминированной задачи выбора оптимального параметра возможно использование градиентных алгоритмов оптимизации.

Данная работа направлена на исследование задачи сплайновой аппроксимации данных в условиях неопределённости на основе параметризации управления и алгоритма метода проекции градиента.

2. Постановка задачи сплайновой аппроксимации данных

Рассмотрим задачу сплайновой аппроксимации данных, если входные данные представляют собой зафиксированные значения функции в конкретные моменты времени с применением метода интерполяции сплайнами [13].

Пусть $D=\{(t_i,\alpha_i):i=1,\ldots,n\}$ – набор данных в момент времени t_i и значением α_i, а $F=\{f\in C^2[0,T] \mid f(t_i)=\alpha_i\}$ – множество дважды непрерывно дифференцируемых функций, которые интерполируют данные.

Рассмотрим задачу сплайновой аппроксимации данных:

$$\min_{f \in L_2[0,T]} \int_0^T u(t)^2\,dt + \lambda \sum_{i=1}^{N} \left(y(t_i) - \eta_i \right))^2, \tag{1}$$

$$\frac{dx(t)}{dt} = Ax(t) + Bu(t), \tag{2}$$

$$x(0) = x_0, \tag{3}$$

$$y(t) = Cx(t). \tag{4}$$

Пусть $\boldsymbol{f} : [0,T] \to R^m$ – функция, которую необходимо аппроксимировать. После проведения эксперимента данные измерений $\boldsymbol{f}(t)$ при $t_1, \ldots, t_N$ могут отличаться от реальных значений из-за шума или погрешности измерения. Тогда значения измерений в каждый момент времени имеют вид

$$\beta_i = \boldsymbol{f}(\boldsymbol{t_i}) + \boldsymbol{\varepsilon_i},\ i = 1, \ldots, N. \tag{5}$$

где $\boldsymbol{\varepsilon_i} = \left[\boldsymbol{\varepsilon}_1^{\boldsymbol{i}}, \ldots, \boldsymbol{\varepsilon}_{\boldsymbol{m}}^{\boldsymbol{i}} \right]$ – случайные отклонения, генерируемые наложенным шумом и соответствующей ошибкой измерения. Предположим, что средние значения, дисперсии и третьи моменты известны.

Оптимизируемая функция примет вид

$$\boldsymbol{g}(\boldsymbol{x}_0, \boldsymbol{u}) = \sum_{i=1}^{N} w_i \left\| \boldsymbol{y}(t_i) - \beta_i \right\|^2 + \lambda \int_0^T \left\| \boldsymbol{u}(t) \right\|^2 dt, \tag{6}$$

где $w_i,\ i = 1, \ldots, N$ – весовые коэффициенты, λ – константа для регулирования гладкости кривой аппроксимации $\boldsymbol{x}(t \mid \boldsymbol{x}_0, \boldsymbol{u})$, $\| \,.\, \|$ – Евклидова норма.

Так как функция $g(\boldsymbol{x}_0, \boldsymbol{u})$ содержит случайные значения, то преобразуем ее к детерминированной с помощью выпуклой комбинации среднего значения и дисперсии данной функции.

Таким образом, задача состоит в отыскании начального условия $\boldsymbol{x}_0 \in X$ и управляющей функции $\boldsymbol{u}(t) \in \boldsymbol{U}$ для целевого функционала

$$\hat{G}(\boldsymbol{x}_0, \boldsymbol{u}) = \alpha \mathrm{M}\left\{ g(\boldsymbol{x}_0, \boldsymbol{u}) \right\} + (1 - \alpha) Var\left\{ g(\boldsymbol{x}_0, \boldsymbol{u}) \right\}, \tag{7}$$

минимизируемого по $\boldsymbol{x}_0 \in X$ и $\boldsymbol{u}(t) \in \boldsymbol{U}$, где $\mathrm{M}\{.\}$ – математическое ожидание, $Var\{.\}$ – дисперсия, а $\alpha \in [0,1]$ – весовой коэффициент.

Функцию $\hat{G}(\boldsymbol{x}_0, \boldsymbol{u})$ можно преобразовать в детерминированную и задача ОУ будет иметь вид:

$$\begin{aligned} \hat{G}(\boldsymbol{x}_0, \boldsymbol{u}) = {} & \alpha\{ \sum_{i=1}^{N} w_i (\boldsymbol{y}(\boldsymbol{t_i}))^{\mathsf{T}}\, \boldsymbol{y}(t_i) - 2 \sum_{i=1}^{N} w_i (\boldsymbol{y}(\boldsymbol{t_i}))^{\mathsf{T}}\, \mathrm{M}\{\beta_i\} + \sum_{i=1}^{N} w_i \mathrm{M}\left\{ (\beta_i)^{\mathsf{T}} \beta_i \right\} + \lambda \int_0^T \left\| \boldsymbol{u}(t) \right\|^2 dt \} + \\ & + (1 - \alpha)\{ 4 \sum_{i=1}^{N} \sum_{l=1}^{N} (\boldsymbol{y}(\boldsymbol{t_i}))^{\mathsf{T}}\, Cov\left\{ (\beta_i)^{\mathsf{T}} \beta_l \right\} \boldsymbol{y}(\boldsymbol{t_l}) + Var\left\{ \sum_{i=1}^{N} w_i (\beta_i)^{\mathsf{T}} \beta_i \right\} - \\ & - 4 \sum_{i=1}^{N} \sum_{l=1}^{N} 1^m\, Cov\left\{ (\beta_i)^2, \beta_l \right\} \boldsymbol{y}(\boldsymbol{t_l}) \}, \end{aligned} \tag{8}$$

где $\boldsymbol{u}(t) : [0,T] \to U$ – функция управления, где U – компактное выпуклое множество в R^r, $y(t) \in R^m$ – результирующая функция, а $\boldsymbol{x}_0 \in X$ – неизвестное начальное состояние, где X является компактным подмножеством R^n.

Подход к параметризации управления является наиболее распространенным для решения задачи ОУ общего вида. В этом случае временной интервал разбивается на несколько частей и на каждом строится кусочно-постоянная функция управления с возможными точками разрыва первого рода в точках разбиения интервала.

Используя данную схему аппроксимации, формируют конечномерную задачу аппроксимации исходной ОУ, которая соответствует задаче выбора оптимального параметра.

3. Методы решения задачи сплайновой аппроксимации данных

Для численного решения задачи (2)–(4), (8) представим управление кусочно-постоянной функцией с конечным числом переключений методом параметризации управления.

Для каждого $i=1,\ldots,N+1$ разобьем интервал $\left[t_{i-1},t_i\right]$ на p_i подинтервалов, а граничные точки обозначим как τ_j^i, где $j=1,\ldots,p_i-1$.

Последовательность $\left[\tau_{j-1}^i,\tau_j^i\right)$, $i=1,\ldots,N+1;\ j=1,\ldots,p_i$ называется разделом временного горизонта, а $\tau^{M-1}=\left[\tau_1^1,\ldots,\tau_{p_{N+1}-1}^{N+1}\right]\in R^{M-1}$ – вектором допустимого времени переключения при $M=\sum_{i=1}^{N+1}p_i$.

Для некоторого управления $\boldsymbol{u}(t)\in\boldsymbol{U}$ определим конечномерное приближение

$$\boldsymbol{u}(t)\approx u^M(t)=\sum_{i=1}^{N+1}\sum_{j=1}^{p_i}\gamma^{\zeta_j^i}\chi_{\left[\tau_{j-1}^i,\tau_j^i\right)}(t),\ t\in[0,T], \tag{9}$$

где $\zeta_j^i=\sum_{l=1}^{i-1}p_l+j$ – метка подинтервала, $\gamma^i=\left[\gamma_1^i,\ldots,\gamma_r^i\right]^{\mathrm{T}}\in\boldsymbol{R}^r,\gamma_l^i\in\boldsymbol{U},\ l=1,\ldots,r$ – допустимый вектор управления и $\chi_I(t)$ – индикаторная функция вида

$$\chi_I(t)=\begin{cases}1, \text{ если } t\in I,\\ 0, \text{ иначе.}\end{cases}$$

При подстановке приближения (9) в уравнение (2) получим систему, которая определена на подинтервале $\left[\tau_{j-1}^i,\tau_j^i\right)$:

$$\frac{d\boldsymbol{x}(t)}{dt}=A\boldsymbol{x}(t)+B\boldsymbol{\gamma}^{\zeta_j^i}, \tag{10}$$

$$\boldsymbol{y}(t)=C\boldsymbol{x}(t). \tag{11}$$

Детерминированная задача выбора оптимального параметра

Для динамической системы (10)–(11) найти $\boldsymbol{x}_0\in X$ и $\boldsymbol{\gamma}\in\Gamma$, соответствующие минимуму функции:

$$\hat{G}(\boldsymbol{x}_0,\boldsymbol{u})=\alpha\left\{\sum_{i=1}^{N}w_i(\boldsymbol{y}(\boldsymbol{t_i}))^{\mathrm{T}}\,\boldsymbol{y}(t_i)-2\sum_{i=1}^{N}w_i(\boldsymbol{y}(\boldsymbol{t_i}))^{\mathrm{T}}\,\mathrm{M}\{\beta_i\}+\sum_{i=1}^{N}w_i\mathrm{M}\left\{(\beta_i)^{\mathrm{T}}\beta_i\right\}+\lambda\int_0^T\left\|\boldsymbol{u}(t)\right\|^2dt\right\}+$$
$$+(1-\alpha)\{4\sum_{i=1}^{N}\sum_{l=1}^{N}(\boldsymbol{y}(\boldsymbol{t_i}))^{\mathrm{T}}\,Cov\left\{(\beta_i)^{\mathrm{T}}\beta_l\right\}\boldsymbol{y}(\boldsymbol{t_l})+Var\left\{\sum_{i=1}^{N}w_i(\beta_i)^{\mathrm{T}}\beta_i\right\}-$$
$$\left.-4\sum_{i=1}^{N}\sum_{l=1}^{N}1^m\,Cov\left\{(\beta_i)^2,\beta_l\right\}\boldsymbol{y}(\boldsymbol{t_l})\right\}. \tag{12}$$

Метод проекции градиента

Для решения поставленной задачи (12) в рамках данной работы рассмотрен алгоритм метода проекции градиента [14, 15], основная идея которого состоит в поиске экстремума функции в направлении антиградиента функции. В данной работе антиградиент функции сформирован таким образом, чтобы выполнялось свойство допустимости значений траектории движения.

4. Вычислительные эксперименты

В рамках данной статьи рассмотрено исследование параметров численного решения задачи сплайновой аппроксимации данных при различных весовых коэффициентах между математическим ожиданием и дисперсией исходной функции, а также различных подходах к наложению шума на аппроксимируемые данные.

Для проведения численных экспериментов были зафиксированы следующие параметры задачи:

$A = 1{,}8,\ B = 0{,}2,\ C = 2,\ q = 10\,000,\ x_0 = -0{,}5,\ u = 0{,}5,\ M(\beta_i) = -0{,}528,$

$Var(\beta_i) = 3{,}4483,\ \alpha = 0{,}5,\ p_i = 100.$

Исходные данные содержат 10^6 измерений, имеющие нормальный закон распределения, и с наложенным шумом в виде функции $f(t)$:

$$f(t) = \begin{cases} -0{,}2074x^2 + 0{,}5584x - 0{,}2321,\ if\ t \in [0;0{,}5713], \\ 0{,}02x^3 - 0{,}224x^2 + 0{,}562x - 0{,}2323,\ if\ t \in [0{,}5713;7{,}8013], \\ 0{,}2815x^2 - 3{,}7729x + 12{,}137,\ if\ t \in [7{,}8013;10]. \end{cases} \quad (13)$$

Отметим, что погрешности полученных значений $\max\left(y(t_i) - C * f(t_i)\right) = 0{,}074$, $\max\left(x(t_i) - f(t_i)\right) = 0{,}037$ первого порядка и соответствуют устойчивой модели с достаточно высокой точностью.

Проведём исследование влияния параметра весового коэффициента α между математическим ожиданием и дисперсией исходной функции при следующих значениях $\alpha = 0{,}2;\ 0{,}5;\ 0{,}7$ (рис. 1, 2).

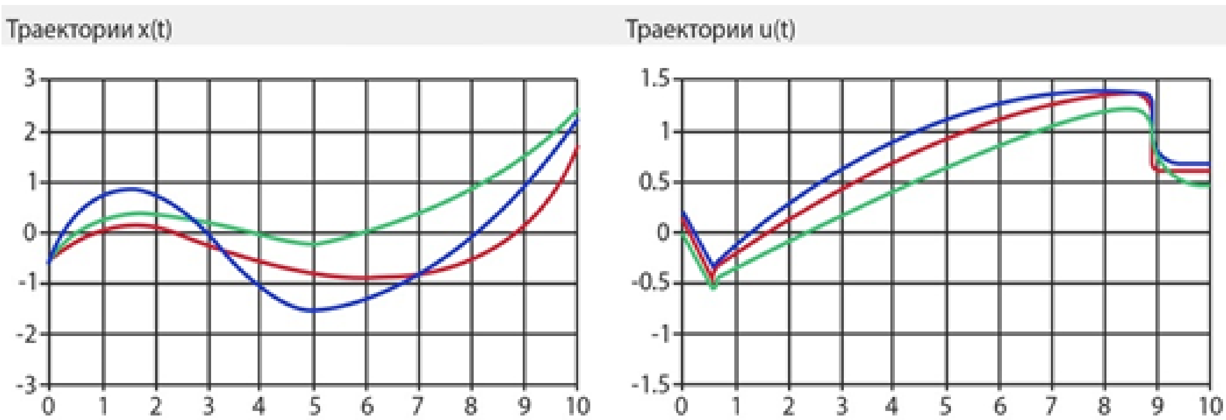

Рис. 1. Траектория $x(t)$ и управление $u(t)$ при различных значениях весового коэффициента α
Fig. 1. Trajectory $x(t)$ and control $u(t)$ at different values of the weight coefficient α

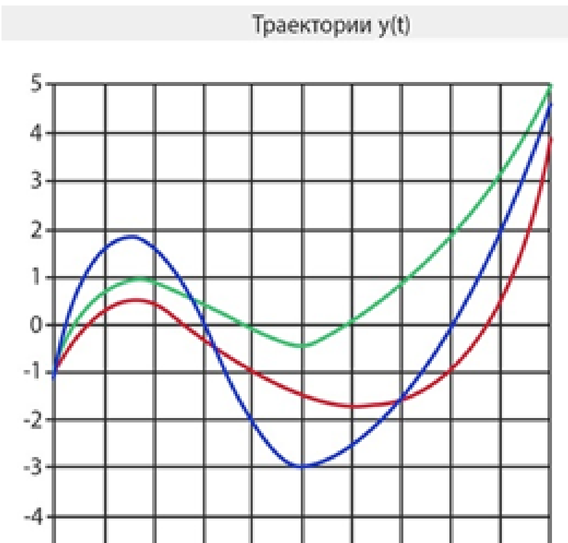

Рис. 2. Траектория $y(t)$ при различных значениях весового коэффициента α
Fig. 2. Trajectory $y(t)$ at different values of the weight coefficient α

Результаты исследования влияния параметра весового коэффициента α на аппроксимационную функцию представлены в табл. 1 с указанием количества итераций и погрешностей.

Таблица 1

Результаты исследования параметра α

Table 1

Results of the study of the parameter α

α	0,2	0,5	0,8
J	34,02	1,2	0,34
Count	11 235	13 101	12 540
$\dfrac{\sum_{i=1}^{N}\left(x(t_i)-f(t_i)\right)}{N}$	0,068	0,037	0,021
$\dfrac{\sum_{i=1}^{N}\left(y(t_i)-f(t_i)\right)}{N}$	0,079	0,044	0,027

При увеличении α значения $x(t_i)$ и $y(t_i)$ имеют большее отклонение к реальным значениям набора данных. Однако данное поведение сопровождается сильным искривлением графиков, что соответствует уровню гладкости второй производной функции $y(t_i)$. При этом погрешность снижается, а точность решения значительно увеличивается.

На основании полученных результатов можно сделать вывод о том, что при $\alpha = 0,5$ решение задачи соответствует достаточно высокой точности и не имеет больших потерь в гладкости функций.

Исследование влияния закона распределения шумов

Исследуем влияние закона распределения шумов, накладываемых на исходную функцию, на точность и скорость сходимости данного метода. Рассмотрим следующие законы распределения накладываемых шумов: нормальный, экспоненциальный и равномерный. Зафиксируем параметры для каждого закона распределения.

Нормальный закон распределения шумов:

$$\beta_i = (\boldsymbol{t_i}) + \frac{1}{\sigma\sqrt{2\pi}} e^{\left(-\frac{(\boldsymbol{t_i}-\mu)^2}{2\sigma^2}\right)}, \; i = 1, \ldots, N, \; \mu = 5, \; \sigma^2 = 0,0001. \tag{14}$$

Равномерный закон распределения шумов:

$$\beta_i = \boldsymbol{f}(\boldsymbol{t_i}) + \begin{cases} 0, \; \boldsymbol{t_i} < a, \\ \dfrac{\boldsymbol{t_i} - a}{b - a} * 3, \; a \le \boldsymbol{t_i} \le b, \\ 0, \; \boldsymbol{t_i} > b. \end{cases} \; i = 1, \ldots, N, \; a = 0, \; b = 10. \tag{15}$$

Экспоненциальный закон распределения шумов:

$$\beta_i = \boldsymbol{f}(\boldsymbol{t_i}) + \begin{cases} 0, \; \boldsymbol{t_i} < 0, \\ 1 - e^{-\lambda \boldsymbol{t_i}}, \; \boldsymbol{t_i} \ge 0, \end{cases} \; i = 1, \ldots, N, \; \lambda = 3. \tag{16}$$

На рис. 3, 4 представлены графики $x(t)$ и $u(t)$, полученные при наложении шумов с разными законами распределения.

Результаты исследования влияния закона распределения шумов на аппроксимационную функцию представлены в табл. 2 с указанием количества итераций и погрешностей.

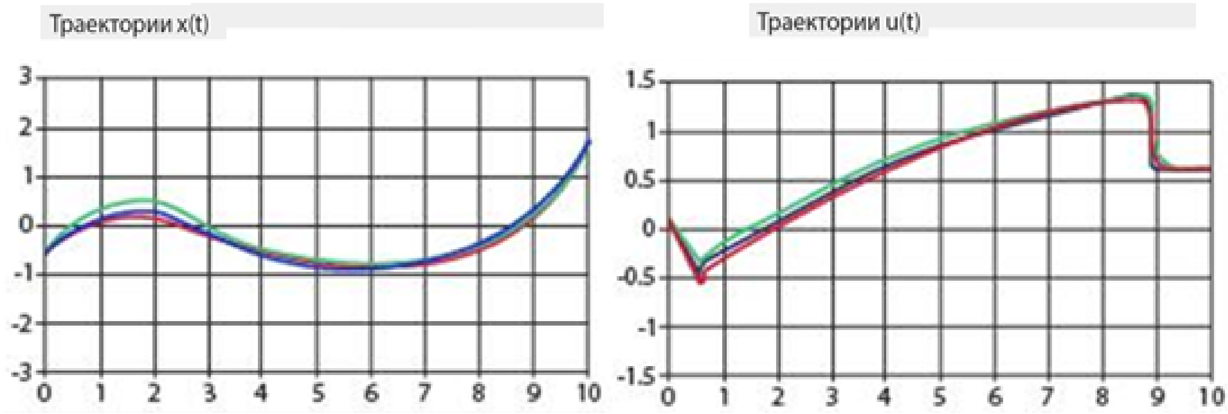

Рис. 3. Траектория $x(t)$ и управление $u(t)$ при различных законах распределения шумов
Fig. 3. Trajectory $x(t)$ and control $u(t)$ under different noise distribution laws

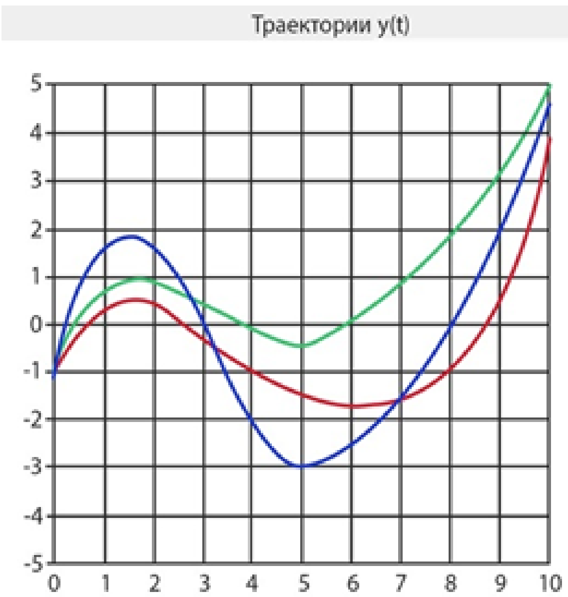

Рис. 4. Траектория $y(t)$ при различных законах распределения шумов
Fig. 4. Trajectory $y(t)$ under different noise distribution laws

Таблица 2

Результаты анализа изменения закона распределения шумов

Table 2

Results of the analysis of changes in the noise distribution law

Закон распределения	Экспоненциальный	Нормальный	Равномерный
J	0,61	0,58	0,563
Count	12 351	11 927	11 468
$\frac{\sum_{i=1}^{N}\left(x(t_i)-f(t_i)\right)}{N}$	0,031	0,024	0,021
$\frac{\sum_{i=1}^{N}\left(y(t_i)-f(t_i)\right)}{N}$	0,06	0,049	0,044

Результаты исследования показали, что различные законы распределения шумов не оказывают сильного воздействия на сходимость метода и количество итераций варьируется не значительно. Таким образом, при равномерном распределении шумов на исходных данных значения погрешностей найденных траекторий сравнительно ниже, чем при иных законах.

При задании параметра α с потерей гладкости, но с повышением точности аппроксимации произойдет изменение порядка ранжирования законов распределения шумов.

Данные результаты свидетельствуют об устойчивости предлагаемого подхода к подбору начальных условий. В частности, весовой коэффициент α регулирует гладкость функции и точность аппроксимации. При этом закон распределения шумов на исходных данных не влияет в должной степени на точность данного подхода и свидетельствует об устойчивости модели и пригодности метода при ограниченных знаниях о шумах.

Заключение

В рамках данного исследования решена задача сплайновой аппроксимации данных в условиях неопределённости. Для применения параметризации управления представлена стохастическая функция, в дальнейшем преобразованная в детерминированную. На основе метода параметризации управления сформирована последовательность приближенного оптимального управления.

С использованием численного метода проекции градиента разработано программное обеспечение для решения задачи сплайновой аппроксимации на зашумленных данных.

Результаты численных экспериментов исследования параметров модели показали, что повышение значений весового коэффициента α при математическом ожидании искомой функции приводит к повышению точности, однако решение в некоторой степени теряет гладкость. Кроме того, проведенный анализ показал, что изученный закон распределения шумов не влияет в должной степени на точность данного подхода и свидетельствует об устойчивости.

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 20-07-01065, а также гранта Президента Российской Федерации для государственной поддержки ведущих научных школ Российской Федерации (НШ-2502.2020.9).

Литература

1. Lai, M.J. Scattered data interpolation with nonnegative preservation using bivariate splines and its application / M.J. Lai, C. Meile // Computer Aided Geometric Design. – 2015. – Vol. 34. – P. 37–49. DOI: 10.1016/j.cagd.2015.02.004

2. A comparative study on application of Chebyshev and spline methods for geometrically nonlinear analysis of truss structures / S.H. Mahdavi, H.A. Razak, S. Shojaee, M.S. Mahdavi // International Journal of Mechanical Sciences. – 2015. – Vol. 101. – P. 241–251. DOI: 10.1016/j.ijmecsci.2015.08.001

3. Inversion of top of atmospheric reflectance values by conic multivariate adaptive regression splines / S. Kuter, G.W. Weber, Z. Akyürek, A. Özmen // Inverse Problems in Science and Engineering. – 2015. – Vol. 23, iss. 4. – P. 651–669. DOI: 10.1080/17415977.2014.933828

4. Agwu, N.N. Optimal control of dynamic systems: application to spline approximations / N.N. Agwu, C.F. Martin // Applied Mathematics and Computation. – 1998. – Vol. 97, iss. 2. – P. 99–138. DOI: 10.1016/S0096-3003(97)10101-1

5. Burt, P.J. A multiresolution spline with application to image mosaics / P.J. Burt, E.H. Adelson // ACM Transactions on Graphics. – 1983. – Vol. 2, iss. 4. – P. 217–236. DOI: 10.1145/245.247

6. A Distributionally robust linear receiver design for multi-access space-time block coded MIMO systems / B. Li, Y. Rong, J. Sun, K.L. Teo // IEEE Transactions on Wireless Communications. – 2016. – Vol. 16, iss. 1. – P. 464–474. DOI: 10.1109/TWC.2016.2625246

7. Some characterizations of robust optimal solutions for uncertain fractional optimization and applications / X.K. Sun, X.J. Long, H.Y. Fu, X.B. Li // Journal of Industrial & Management Optimization. – 2017. – Vol. 13, iss. 2. – pp. 803–824. DOI: 10.3934/jimo.2016047

8. Wang, J. Data-driven tight frame for multi-channel images and its application to joint color-depth image reconstruction / J. Wang, J.F. Cai // Journal of the Operations Research Society of China. – 2015. – Vol. 3. – P. 99–115. DOI: 10.1007/s40305-015-0074-2

9. Yan, H.Y. A linear-quadratic control problem of uncertain discrete-time switched systems / H.Y. Yan, Y. Sun, Y.G. Zhu // Journal of Industrial & Management Optimization. – 2017. – Vol. 13, iss. 1. – P. 267–282. DOI: 10.3934/jimo.2016016

10. A sequential regression model for big data with attributive explanatory variables / Q.T. Zhang, Y. Liu, W. Zhou, Z.W. Yang // Journal of the Operations Research Society of China. – 2015. – Vol. 3. – P. 475–488. DOI: 10.1007/s40305-015-0109-8

11. Friedman, J.H. Multivariate adaptive regression splines / J.H. Friedman // The Annals of Statistics. – 1991. – Vol. 19, iss. 1. – P. 1–141. DOI: 10.1214/aos/ 1176347963

12. Taylan, P. New approaches to regression by generalized additive models and continuous optimization for modern applications in finance, science and technology / P. Taylan, G.W. Weber, A. Beck // Optimization – 2007. – Vol. 56, iss. 6. – P. 675–698. DOI: 10.1080/02331930701618740

13. Исаков, В.Н. Оптимальная регулярная локальная сплайновая интерполяция сигналов / В.Н. Исаков // Вестник Концерна ВКО Алмаз-Антей. – 2016. – Т. 19, № 4. – С. 24–31. DOI: 10.38013/2542-0542-2016-4-24-31

14. Новиков, М. Ю. Конечномерная оптимизация. Алгоритм метода проекции градиента / М.Ю. Новиков // Устойчивое развитие российских регионов: экономическая политика в условиях внешних и внутренних шоков: сб. науч. тр. – Екатеринбург: УрФУ, 2015. – С. 781–785.

15. Голубев, М.О. Метод проекции градиента для сильно выпуклого множества / М.О. Голубев // Известия Саратовского университета. Серия Математика. Механика. Информатика. – 2013. – № 13 (2). – С. 33–38.

Болодурина Ирина Павловна, д-р техн. наук, профессор, заведующий кафедрой прикладной математики, Оренбургский государственный университет, г. Оренбург; prmat@mail.osu.ru.

Гришина Любовь Сергеевна, аспирант кафедры прикладной математики, Оренбургский государственный университет, г. Оренбург; grishina_ls@inbox.ru.

Анциферова Лариса Михайловна, канд. пед. наук, доцент кафедры прикладной математики, Оренбургский государственный университет, г. Оренбург; antsiferova_68@mail.ru.

Поступила в редакцию 21 июня 2021 г.

DOI: 10.14529/ctcr210314

INVESTIGATION OF PARAMETERS OF THE PROBLEM OF SPLINE APPROXIMATION OF NOISY DATA BY NUMERICAL METHODS OF OPTIMAL CONTROL

I.P. Bolodurina, *prmat@mail.osu.ru,*
L.S. Grishina, *grishina_ls@inbox.ru,*
L.M. Antsiferova, *antsiferova_68@mail.ru*
Orenburg State University, Orenburg, Russian Federation

Currently, the problems of distortion of measurement data by noise and the appearance of uncertainties in quality criteria have caused increased interest in research in the field of spline approximation. At the same time, existing methods of minimizing empirical risk, assuming that the noise is a uniform distribution with heavier tails than Gaussian, limit the scope of application of these studies. The problem of estimating noise-distorted data is usually based on solving an optimization problem with a function containing uncertainty arising from the problem of finding optimal parameters. In this regard, the estimation of distorted noise cannot be solved by classical methods. **Aim.** This study is aimed at solving and analyzing the problem of spline approximation of data under uncertainty conditions based on the parametrization of control and the gradient projection algorithm. **Methods.** The study of the problem of spline approximation of noisy data is carried out by the method of approximation of the piecewise constant control function. In this case, parametrization of the control is possible only for a finite number of break points of the first kind. In the framework of the experimental study, the gradient projection algorithm is used for the numerical solution of the spline approximation problem. The proposed methods are used to study the parameters of the problem of spline approximation of data under conditions of uncertainty. **Results.** The numerical study of the control parametrization approach and the gradient projection algorithm is based on the developed software and algorithmic tool for solving the problem of the spline approximation model under uncertainty. To evaluate the noise-distorted data, numerical experiments were conducted to study the model parameters and it was found that increasing the value of the parameter α leads to an increase in accuracy, but a loss of smoothness. In addition, the analysis showed that the considered distribution laws did not change the accuracy and convergence rate of the algorithm. **Conclusion.** The proposed approach for solving the problem of spline approximation under uncertainty conditions allows us to determine the problems of distortion of measurement data by noise and the appearance of uncertainties in the quality criteria. The study of the model parameters showed that the constructed system is stable to the error of the initial approximation, and the distribution laws do not significantly affect the accuracy and convergence of the gradient projection method.

Keywords: citation system, scientometric methods, aggregation of bibliographic information, modification of the Winnowing method, Levenshtein method, shingle method.

References

1. Lai M.J., Meile C. Scattered data interpolation with nonnegative preservation using bivariate splines and its application *Computer Aided Geometric Design*, 2015, vol. 34, pp. 37–49. DOI: 10.1016/j.cagd.2015.02.004

2. Mahdavi S.H., Razak H.A., Shojaee S., Mahdavi M.S. A comparative study on application of Chebyshev and spline methods for geometrically non-linear analysis of truss structures *International Journal of Mechanical Sciences*, 2015, vol. 101, pp. 241–251. DOI: 10.1016/j.ijmecsci.2015.08.001

3. Kuter S., Weber G.W., Akyürek Z., Özmen A. Inversion of top of atmospheric reflectance values by conic multivariate adaptive regression splines *Inverse Problems in Science and Engineering*, 2015, vol. 23, iss. 4, pp. 651–669. DOI: 10.1080/17415977.2014.933828

4. Agwu N.N., Martin C.F. Optimal control of dynamic systems: application to spline approximations *Applied Mathematics and Computation*, 1998, vol. 97, iss. 2, pp. 99–138. DOI: 10.1016/S0096-3003(97)10101-1

5. Burt P.J., Adelson E.H. A multiresolution spline with application to image mosaics *ACM Transactions on Graphics*, 1983, vol. 2, iss. 4, pp. 217–236. DOI: 10.1145/245.247

6. Li B., Rong Y., Sun J., Teo K.L. A Distributionally robust linear receiver design for multi-access space-time block coded MIMO systems *IEEE Transactions on Wireless Communications*, 2016, vol. 16, iss. 1, pp. 464–474. DOI: 10.1109/ TWC.2016.2625246

7. Sun X.K., Long X.J., Fu H.Y., Li X.B. Some characterizations of robust optimal solutions for uncertain fractional optimization and applications *Journal of Industrial & Management Optimization*, 2017, vol. 13, iss. 2, pp. 803–824. DOI: 10.3934/ jimo.2016047

8. Wang J., Cai J.F. Data-driven tight frame for multi-channel images and its application to joint color-depth image reconstruction *Journal of the Operations Research Society of China*, 2015, vol. 3, pp. 99–115. DOI: 10.1007/s40305-015-0074-2

9. Yan H.Y., Sun Y., Zhu Y.G. A linear-quadratic control problem of uncertain discrete-time switched systems *Journal of Industrial & Management Optimization*, 2017, vol. 13, iss. 1, pp. 267–282. DOI: 10.3934/jimo.2016016

10. Zhang Q.T., Liu Y., Zhou W., Yang Z.W. A sequential regression model for big data with attributive explanatory variables *Journal of the Operations Research Society of China*, 2015, vol. 3, pp. 475–488. DOI: 10.1007/s40305-015-0109-8

11. Friedman J.H. Multivariate adaptive regression splines *The Annals of Statistics*, 1991, vol. 19, iss. 1, pp. 1–141. DOI: 10.1214/aos/ 1176347963

12. Taylan P., Weber G.W., Beck A. New approaches to regression by generalized additive models and continuous optimization for modern applications in finance, science and technology *Optimization*, 2007, vol. 56, iss. 6, pp. 675–698. DOI: 10.1080/02331930701618740

13. Isakov V.N. [Optimal regular local spline signal interpolation]. *Bulletin of Koncern VKO Almaz-Antey*, 2016, vol. 19, no. 4. – pp. 24–31. (in Russ.)

14. Novikov M.Yu. [Finite-dimensional optimization. Algorithm of the gradient projection method]. *Ustoychivoye razvitiye rossiyskikh regionov: ekonomicheskaya politika v usloviyakh vneshnikh i vnutrennikh shokov: sb. nauch. tr.* [Sustainable development of Russian Regions: Economic Policy in the conditions of external and internal shocks: collection of scientific papers], Yekaterinburg: Ural Federal University Publ., 2015, pp. 781–785. (in Russ.)

15. Golubev M.O. [The gradient projection method for a strongly convex set]. *Izvestiya Saratovskogo universiteta. Seriya Matematika. Mekhanika. Informatika*, 2013, vol. 13, no. 2, pp. 33–38. (in Russ.)

Received 21 June 2021

ОБРАЗЕЦ ЦИТИРОВАНИЯ

Болодурина, И.П. Исследование параметров задачи сплайновой аппроксимации зашумленных данных численными методами оптимального управления / И.П. Болодурина, Л.С. Гришина, Л.М. Анциферова // Вестник ЮУрГУ. Серия «Компьютерные технологии, управление, радиоэлектроника». – 2021. – Т. 21, № 3. – С. 138–148. DOI: 10.14529/ctcr210314

FOR CITATION

Bolodurina I.P., Grishina L.S., Antsiferova L.M. Investigation of Parameters of the Problem of Spline Approximation of Noisy Data by Numerical Methods of Optimal Control. *Bulletin of the South Ural State University. Ser. Computer Technologies, Automatic Control, Radio Electronics*, 2021, vol. 21, no. 3, pp. 138–148. (in Russ.) DOI: 10.14529/ctcr210314