

Устойчивый алгоритм непараметрической идентификации при наличии аномальных измерений

Ю. Е. Воскобойников^{1,2,†}, В. А. Боева¹

¹Новосибирский государственный архитектурно-строительный университет (Сибстрин), Новосибирск, Россия

²Новосибирский государственный технический университет, Новосибирск, Россия

[†]Контактный автор: Воскобойников Юрий Е., e-mail: voscob@mail.ru

Поступила 10 марта 2020 г., доработана 23 марта 2020 г., принята в печать 14 апреля 2020 г.

Математические модели многих технических систем имеют вид интегрального уравнения Вольтерра I рода с разностным ядром. Для таких систем задача идентификации заключается в построении оценки для импульсной переходной функции системы по измеренным (с шумами) значениям входного и выходного сигналов и является некорректно поставленной. В недавней работе авторов предложен устойчивый алгоритм идентификации, использующий аппарат сглаживающих кубических сплайнов для вычисления первых производных входного и выходного сигналов. К сожалению, сглаживающие кубические сплайны неудовлетворительно фильтруют аномальные измерения. Поэтому предложен двухшаговый алгоритм идентификации, на первом шаге которого аномальные измерения удаляются с использованием пространственно-локального фильтра, а затем строятся сглаживающие сплайны.

Ключевые слова: задача идентификации, алгоритм решения уравнения Вольтерра II рода, фильтрация аномальных измерений, сглаживающие кубические сплайны.

Цитирование: Воскобойников Ю.Е., Боева В.А. Устойчивый алгоритм непараметрической идентификации при наличии аномальных измерений. Вычислительные технологии. 2020; 25(3):46–53.

Введение и цель исследования

Интегральные модели достаточно часто используются для описания и моделирования технических систем [1]. Если система стационарная (т.е. ее параметры не меняются во времени), то для ее описания применяется интегральное уравнение Вольтерра I рода [1, 2]:

$$\int_0^t k(t-\tau)\varphi(\tau)d\tau = f(t), \quad t \in [0, T], \quad (1)$$

где $k(t)$ — импульсная переходная функция (ИПФ) системы; $\varphi(\tau)$, $f(t)$ — ее входной и выходной сигналы. Техническую реализуемость системы определяет условие

$$k(t) = 0 \quad \text{при} \quad t < 0. \quad (2)$$

Для модели (1) задача непараметрической идентификации [2, 3] заключается в построении оценки для $k(t)$ по зарегистрированным значениям входного и выходного сигналов идентифицируемой системы.

Рассматриваемая задача идентификации является некорректно поставленной, поскольку при ее решении может нарушаться одно или несколько требований корректности задачи по Адамару [3]. Наиболее часто нарушается условие устойчивости решения к погрешностям (ошибкам) исходных данных, когда малые погрешности могут вызвать сколь угодно большие ошибки решения. Для построения устойчивых решений некорректных задач используются специальные методы регуляризации, как детерминированные [3], так и статистические [4]. В работе авторов [5] построен эффективный алгоритм, вычисляющий оценку ИПФ из решения интегрального уравнения Вольтерра II рода, что является уже корректно поставленной задачей. Однако при этом необходимо было вычислить первые производные от входного и выходного сигналов, что, в свою очередь, также является неустойчивой задачей. Для устойчивого вычисления первых производных использовались сглаживающие кубические сплайны (СКС), что обусловило устойчивость решения задачи идентификации даже к большим уровням шумов измерения входного и выходного сигналов идентифицируемой системы.

В ряде случаев зарегистрированные значения входного и выходного сигналов могут содержать аномальные измерения, которые значительно (в разы) отличаются от близлежащих измерений, и их появление можно объяснить наличием шумовой составляющей (импульсного шума), дисперсия которой существенно (на два и более порядков) отличается от дисперсии шумов измерений других значений сигналов. К сожалению, СКС очень слабо сглаживают такие импульсные шумы, и это обуславливает большие ошибки в вычислении первых производных, что, в свою очередь, приводит к существенным ошибкам идентификации ИПФ динамической системы.

В данной работе предложен новый двухшаговый алгоритм идентификации, на первом шаге которого эффективно фильтруются аномальные измерения (т.е. удаляется импульсный шум) с использованием пространственно-локальных фильтров, а на втором шаге осуществляется построение СКС по отфильтрованным значениям входного и выходного сигналов и устойчивого решения с использованием алгоритма, изложенного в работе [5].

1. Метод исследования

Для построения устойчивого алгоритма решения задачи идентификации в работе [5] осуществляется переход от уравнения (1) к уравнению задачи идентификации, которое при выполнении условия (2) имеет вид

$$\int_0^t \varphi(t-\tau)k(\tau)d\tau = f(t), \quad t \in [0, T]. \quad (3)$$

Путем дифференцирования этого уравнения по переменной t и несложных преобразований приходим к интегральному уравнению Вольтерра II рода:

$$k(t) + \frac{1}{\varphi(0)} \int_0^t \varphi'(t-\tau)k(\tau)d\tau = \frac{f'(t)}{\varphi(0)}, \quad t \in [0, T]. \quad (4)$$

Для устойчивого дифференцирования зашумленных входного и выходного сигналов идентифицируемой системы в работе [5] использовались СКС с естественными краевыми условиями [4, 6]. Напомним, что функция $S_{\varphi,\alpha}(x)$ называется сглаживающим кубическим сплайном на интервале $[0, T]$ с узлами

$$0 = t_1 < t_2 < \dots < t_N = T, \quad (5)$$

если:

- а) на каждом интервале $[t_i, t_{i+1})$ функция $S_{\varphi,\alpha}(t)$ является кубическим многочленом вида

$$S_{\varphi,\alpha}(x) = a_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3;$$

- б) функция $S_{\varphi,\alpha}(t)$ дважды непрерывно дифференцируема на всем интервале $[0, T]$;
 в) функция $S_{\varphi,\alpha}(t)$ удовлетворяет естественным краевым условиям на вторую производную $S''_{\varphi,\alpha}(t)$ вида $S''_{\varphi,\alpha}(t_1) = S''_{\varphi,\alpha}(t_N) = 0$ и доставляет минимум функционалу [4, 6, 7]:

$$F_\alpha(S) = \alpha \int_{t_1}^{t_N} |S''(t)|^2 dt + \sum_{i=1}^n p_i^{-1} (\tilde{\varphi}_i - S(t_i))^2, \quad (6)$$

где p_i^{-1} , $i = 1, \dots, N$, — весовые множители; $\tilde{\varphi}_i$, $i = 1, \dots, N$, — зашумленные значения сглаживаемой функции в узлах сетки (5), для которой строится СКС. В рассматриваемой задаче идентификации это зашумленные значения входного и выходного сигналов идентифицируемой системы. Параметр сглаживания α “управляет” гладкостью сплайна (а следовательно, и ошибкой сглаживания) и может меняться в интервале $[0, \infty)$:

— при $\alpha = 0$ сглаживающий сплайн становится интерполяционным (т. е. проходит через все заданные точки φ_i и фильтрация шума отсутствует);

— при $\alpha \rightarrow \infty$ сглаживающий сплайн вырождается в прямую линию (эффект пере-сглаживания зашумленных данных).

Алгоритм вычисления коэффициентов a_i, b_i, c_i, d_i сглаживающего сплайна при заданном параметре α подробно изложен в [4]. Заметим только, что первоначально решается СЛАУ с квадратной пятидиагональной матрицей размером $(N - 2) \times (N - 2)$ — находится вектор вторых производных, с использованием которого впоследствии вычисляются коэффициенты полиномов (6).

Основная трудность, возникающая на практике, — это выбор величины параметра сглаживания из условия минимума ошибки сглаживания, т. е. нахождения оптимального параметра α_{opt} , который минимизирует среднеквадратическую ошибку (СКО) сглаживания. В работе [4] рассмотрены несколько алгоритмов выбора параметра сглаживания [6–8]. Показано, что эффективной (наилучшей) оценкой для α_{opt} является величина α_W , вычисленная на основе статистического критерия оптимальности фильтрации зашумленных данных. Обоснование этого критерия и его свойства подробно приведены в [4], критерий кратко изложен в работе [5]. Заметим, что вычисление оценки α_W требует задания дисперсии шума измерений, которая в большинстве практических случаев неизвестна. Тогда целесообразно обратиться к оценке дисперсии, предложенной в работе [9] и позволяющей с приемлемой точностью (3–8 %) оценить дисперсию шума измерений.

Рассмотрим поведение сглаживающего сплайна при наличии аномальных измерений (АИ). В качестве иллюстрации на рис. 1 сплошной линией показаны точные значения входного сигнала идентифицируемой системы, точечной кривой — зашумленные

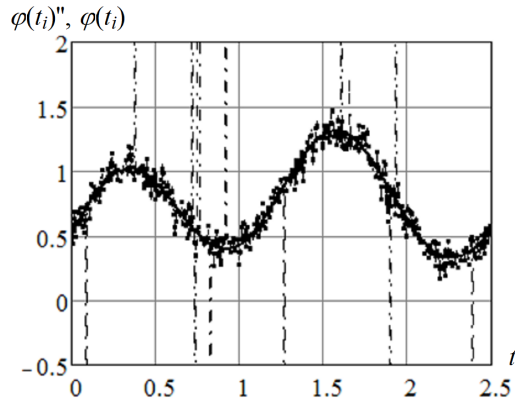


Рис. 1. Измеренные значения входного сигнала
Fig. 1. The measured values of the input signal

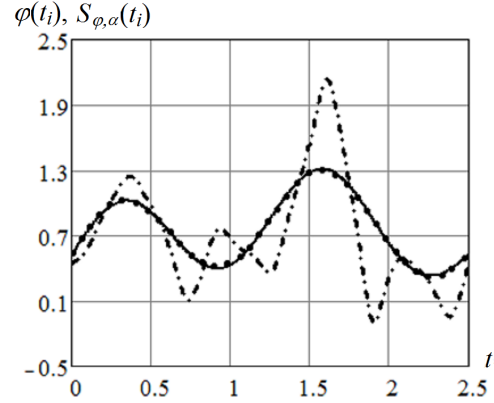


Рис. 2. Значения сглаживающих сплайнов
Fig. 2. Values of smoothing splines

значения (относительный уровень шума 10%), штрихточечной кривой — измеренные значения, содержащие АИ (вероятность появления АИ равна 0.02).

На рис. 2 точечной кривой показаны значения СКС, построенного по измерениям, не содержащим АИ, штрихточечная кривая — значения СКС, построенного по измерениям, содержащим АИ. Параметр сглаживания для этих двух сплайнов был одинаков и равен $\alpha = \alpha_W$, где α_W — оценка оптимального параметра сглаживания (подробнее см. в [6]).

Видно, что второй СКС содержит осцилляции, вызванные АИ, которые дадут существенные ошибки при вычислении первой производной и, соответственно, большую ошибку в решении уравнения (4). Неудовлетворительное сглаживание АИ обусловлено вторым слагаемым $\sum_{i=1}^n p_i^{-1}(\varphi_i - S(t_i))^2$ функционала (6) (слагаемое называется функционалом невязки). Наличие квадрата разности $(\varphi_i - S(t_i))^2$ “заставляет” сплайн “приближаться” к аномальному измерению, поэтому и не удастся эффективно “сгладить” данное АИ.

2. Робастный двухшаговый алгоритм идентификации

Предположим, что зашумленные значения входного и выходного сигналов допускают следующие представления:

$$\tilde{\varphi}_i = \varphi(t_i) + \xi_i, \quad \tilde{f}_i = f(t_i) + \eta_i, \quad i = 1, \dots, N,$$

где ξ_i, η_i — шумы измерения. Для статистического описания АИ введем модели шумов ξ_i, η_i . Шум измерения входного сигнала ξ_i с вероятностью $1 - p_\xi$ имеет нулевое среднее и дисперсию σ_ξ^2 , а с вероятностью p_ξ — нулевое среднее и дисперсию $C_\xi \sigma_\xi^2$, где $C_\xi \gg 1$, т. е. с вероятностью p_ξ появляется импульсная составляющая шума измерения. Если эта вероятность равна нулю, то такой шум (без импульсной составляющей) будем называть однородным. Шум измерения выходного сигнала η_i с вероятностью $1 - p_\eta$ имеет нулевое среднее и дисперсию σ_η^2 , а с вероятностью p_η — нулевое среднее и дисперсию $C_\eta \sigma_\eta^2$, где $C_\eta \gg 1$. На рис. 1 показаны АИ при $p_\xi = 0.02$ и $C_\xi = 400$.

На первом шаге предлагаемого алгоритма идентификации осуществляется нелинейная фильтрация с использованием комбинированного фильтра (КФ) [10]. Его выходной сигнал определяется соотношением (при фильтрации входного сигнала)

$$\hat{\varphi}_j^{\text{K}\Phi} = \text{aver}_K(\hat{\varphi}_i^{\text{M}\Phi} : j - K \leq i \leq j + K, |\hat{\varphi}_i^{\text{M}\Phi} - \hat{\varphi}_j^{\text{M}\Phi}| \leq \Delta_\varphi),$$

где $\hat{\varphi}_i^{\text{M}\Phi}$ — результат фильтрации медианным фильтром (МФ). Результат работы МФ вычисляется как [10]

$$\hat{\varphi}_j^{\text{M}\Phi} = \text{med}_L(\tilde{\varphi}_{j-L}, \tilde{\varphi}_{j-L+1}, \dots, \tilde{\varphi}_j, \dots, \tilde{\varphi}_{j+L}),$$

где med_L — функция, вычисляющая медиану из $2L + 1$ значений, попавших в апертуру фильтра (указаны в скобках). Заметим, что МФ хорошо фильтрует импульсные шумы и сохраняет контрастные составляющие в отфильтрованном сигнале. Усреднение в КФ (сглаживание оставшихся однородных шумов) происходит только для значений $\hat{\varphi}_i^{\text{M}\Phi}$ из интервала $[\hat{\varphi}_i^{\text{M}\Phi} - \Delta_\varphi, \hat{\varphi}_i^{\text{M}\Phi} + \Delta_\varphi]$, что предотвращает сглаживание контрастных составляющих точного сигнала. Таким образом, КФ устраняет импульсные шумы и успешно сглаживает однородные остаточные шумы, при этом размеры апертур удовлетворяют неравенству $K \geq L$. Если величина $K = 0$, то комбинированный фильтр превращается в МФ. Если величина $L = 0$, то КФ превращается в интервальный фильтр скользящего среднего. Величину Δ_φ можно определить соотношением $\Delta_\varphi = 2\sigma_\eta^2$. Аналогичный КФ используется для фильтрации выходного сигнала.

На втором этапе по отфильтрованным значениям входного $\{\hat{\varphi}_i^{\text{K}\Phi}\}$ и выходного $\{\hat{f}_i^{\text{K}\Phi}\}$ сигналов строятся сглаживающие кубические сплайны $S_{\varphi,\alpha}(t)$, $S_{f,\alpha}(t)$ соответственно. После построения сплайнов и вычисления их первых производных $S'_{\varphi,\alpha}(t)$, $S'_{f,\alpha}(t)$ формируется матрица Φ размером $(N - 1) \times (N - 1)$, элементы которой определяются как

$$\Phi_{i,j} = 0 \text{ при } i < j, \quad \Phi_{j,i} = \int_{t_k}^{t_{k+1}} S'_{\varphi,\alpha}(\tau) d\tau = b_k + 2c_k(t_{k+1} - t_k) + 3d_k(t_k + 1 - t_k)^2 \text{ при } i \geq j,$$

где $k = j - i$. Тогда уравнение (4) можно аппроксимировать следующей системой линейных алгебраических уравнений (СЛАУ) вида (см. подробнее в работе [5]):

$$\left(I + \frac{1}{S_{\varphi,\alpha}(t_1)} \Phi \right) \hat{\mathbf{k}} = \frac{1}{S_{\varphi,\alpha}(t_1)} \hat{\mathbf{f}}',$$

где I — единичная матрица размером $(N - 1) \times (N - 1)$, а вектор $\hat{\mathbf{f}}' = \begin{pmatrix} S'_{f,\alpha}(t_1) \\ \vdots \\ S'_{f,\alpha}(t_N) \end{pmatrix}$.

Решая эту систему, получаем вектор $\hat{\mathbf{k}}$, проекции которого являются оценками для значений $k(t_i)$ идентифицируемой ИПФ-системы.

3. Результаты численных исследований

Для проверки работоспособности и эффективности предложенного алгоритма идентификации проведен вычислительный эксперимент. Остановимся только на некоторых результатах этого эксперимента. В качестве $k(t)$ взята ИПФ колебательного звена второго порядка (на рис. 3 — сплошная кривая). Значения входного и выходного сигналов искажались случайными нормально распределенными погрешностями с относительными уровнями $\delta_\varphi = \frac{\|\tilde{\varphi} - \varphi\|}{\|\varphi\|}$, $\delta_f = \frac{\|\tilde{\mathbf{f}} - \mathbf{f}\|}{\|\mathbf{f}\|}$, где φ , \mathbf{f} — векторы, составленные из точных значений сигналов; $\tilde{\varphi}$, $\tilde{\mathbf{f}}$ — векторы зашумленных значений; $\|\cdot\|$ — евклидова норма

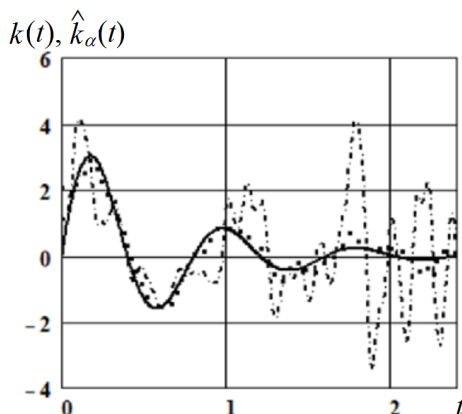


Рис. 3. Импульсная переходная функция и ее оценки

Fig. 3. Pulse transition function and its estimates

вектора. В отсутствие АИ $\delta_\varphi = 0.10$, $\delta_f = 0.10$, при наличии аномальных измерений ($p_\xi = 0.02$, $C_\xi = 400$, $p_\eta = 0.02$, $C_\eta = 400$) — $\delta_\varphi = 0.358$, $\delta_f = 0.421$. На рис. 3 штрихточечной кривой показаны значения ИПФ, вычисленные алгоритмом из работы [6], когда сглаживающие сплайны строились по АИ и имели из-за этого большие ошибки сглаживания (см. штрихточечную кривую на рис. 2). После сглаживания СКС имеем относительные уровни ошибок сглаживания $\delta_\varphi = 0.311$, $\delta_f = 0.591$. В этом случае относительная ошибка идентификации $\delta_k = \frac{\|\hat{\mathbf{k}} - \mathbf{k}\|}{\|\mathbf{k}\|}$ была равна 1.258, что говорит об очень плохой точности идентификации.

Приведем некоторые количественные характеристики предлагаемого двухступенчатого алгоритма идентификации. После выполнения на его первом шаге фильтрации с использованием КФ относительные уровни остаточных шумов были равны: $\delta_\varphi = 0.052$, $\delta_f = 0.061$. Сравнивая уровни шума с уровнями шумов исходных сигналов ($\delta_\varphi = 0.358$, $\delta_f = 0.421$), видим существенное их уменьшение и удаление импульсных составляющих. После сглаживания СКС остаточного шума имеем $\delta_\varphi = 0.038$, $\delta_f = 0.049$. На рис. 3 точечной кривой показаны значения ИПФ, вычисленные двухшаговым алгоритмом идентификации, при этом относительная ошибка идентификации составила $\delta_k = 0.183$. Заметим, что относительная ошибка идентификации алгоритмом работы [5] при отсутствии аномальных измерений равна 0.131.

Заключение

Предложенный двухшаговый алгоритм идентификации позволяет решать задачу идентификации с приемлемой точностью даже при наличии в измеряемых сигналах идентифицируемой системы импульсного шума. Это объясняется отчасти тем, что матрица СЛАУ хорошо обусловлена (число обусловленности зависит от полосы спектра мощности входного сигнала, но не превышает 1000), и поэтому ошибки дифференцирования выходного сигнала гораздо слабее влияют на построенное решение, чем погрешности правой части в регуляризирующем алгоритме решения исходного интегрального уравнения (3). Следует отметить, что общие затраты машинного времени на построение решения возрастают незначительно (из-за пространственно-локальной фильтрации) и составляют примерно 25–30 %.

Список литературы

- [1] **Сидоров Д.Н.** Методы анализа интегральных динамических моделей: теория и приложения. Иркутск: ИГУ; 2013: 293.
- [2] **Бойков И.В.** Аналитические и численные методы идентификации динамических систем. Пенза: ПГУ; 2016: 396.
- [3] **Тихонов А.Н., Арсенин В.Я.** Методы решения некорректных задач. М.: Наука; 1986: 285.
- [4] **Воскобойников Ю.Е.** Математическая обработка эксперимента в молекулярной газодинамике. Новосибирск: Наука; 1984: 238.
- [5] **Воскобойников Ю.Е., Боева В.А.** Новый устойчивый алгоритм непараметрической идентификации технических систем. Современные наукоемкие технологии. 2019; (5):25–29.
- [6] **Завьялов Ю.С., Квасов Б.И., Мирошниченко В.Л.** Методы сплайн-функций. М.: Наука; 1980: 352.
- [7] **Wang Y.** Smoothing splines methods and applications. BOOK SERIES: Chapman & Hall/CRC Monographs on Statistics and Applied Probability. 2011; (121): 347.
- [8] **Wahba G.** Smoothing noisy data with spline functions. Numerische Mathematik. 1975; 24(2):383–393.
- [9] **Воскобойников Ю.Е., Крысов Д.А.** Оценивание характеристик шума измерения в модели “Сигнал + шум”. Автоматика и программная инженерия. 2018; 3(25):54–61.
- [10] **Воскобойников Ю.Е.** Метод устранения артефактов вейвлет-фильтрации. Фундаментальные исследования. 2017; (8-2):246–250.

Stable algorithm of nonparametric identification in case of anomalous measurements

VOSKOBOYNIKOV YURI E.^{1,2,*}, BOEVA VASILISA A.¹

¹Novosibirsk State University of Architecture and Civil Engineering (Sibstrin), 630008, Novosibirsk, Russia

²Novosibirsk State Technical University, 630073, Novosibirsk, Russia

*Corresponding author: Voskoboynikov Yuri E., e-mail: voscob@mail.ru

Received March 10, 2020, revised March 23, 2020, accepted April 14, 2020

Abstract

Volterra integral equation of the first kind often represents stationary dynamic systems. For such a model, the non-parametric identification problem reduces to the estimation of pulse transition characteristics (that is the kernel of integral equation) from the registered noise-contaminated values of input and output signals. To formulate stable solution for identification problem authors propose algorithm that estimates pulse transition characteristics by solving Volterra integral equation of the

second kind and involving first derivatives of input and output signals application that corresponds to non-stable problem. Smoothing cubic splines employed in robust calculation of first derivatives allow finding a stable solution of identification problem even when input and output signals of system identified are essentially noise-contaminated. Unfortunately, measured values of input and output signals also contain anomalous measurements such as pulse noises, glitches, etc. Such measurements are poorly smoothable by splines that cause high levels of first derivatives errors and, conversely, significant pulse transition characteristics identification errors of dynamic system.

For all the reasons aforementioned, in this paper authors present the new stable two-step identification algorithm in case of anomalous measurements. The first step of the algorithm is for non-linear local-spatial combined filtration procedure of input and output signals that helps to effectively remove anomalous measurements. At the second step, smoothing cubic splines are used to calculate stable first derivatives of previously filtered signals. An extensive computational experiment showed the effectiveness of the proposed algorithm, which allows solving the identification problem with acceptable accuracy in practice even at high intensity of anomalous measurements. The experimental results give reason to recommend this algorithm for solving practical problems of identifying stationary systems, the mathematical model of which is the Volterra integral equation of the first kind with a difference kernel.

Keywords: identification problem, algorithm for solving the Volterra II equation, filtration of anomalous measurements, smoothing cubic splines.

Citation: Voskoboynikov Yu.E., Boeva V.A. Stable algorithm of nonparametric identification in case of anomalous measurements. Computational Technologies. 2020; 25(3):46–53. (In Russ.)

References

1. Sidorov D.N. Metody analiza integral'nykh dinamicheskikh modeley: teoriya i prilozheniya [Analysis methods of integral dynamic models: theory and applications]. Irkutsk: IGU; 2013: 293. (In Russ.)
2. Boykov I.V. Analiticheskie i chislennyye metody identifikatsii dinamicheskikh system [Analytical and numerical methods for identification of dynamic systems]. Penza: PGU; 2016: 396. (In Russ.)
3. Tikhonov A.N., Arsenin V.Ya. Metody resheniya nekorrektnykh zadach [Methods for solution of incorrect problems]. Moscow: Nauka; 1986: 285. (In Russ.)
4. Voskoboynikov Yu.E. Matematicheskaya obrabotka eksperimenta v molekulyarnoy gazodinamike [Mathematical processing of experiment in molecular gas dynamics]. Novosibirsk: Nauka; 1984: 238. (In Russ.)
5. Voskoboynikov Yu.E., Boeva V.A. New stable algorithm of nonparametric identification for technical systems. *Sovremennyye naukoemkie tekhnologii*. 2019; (5):25–29. (In Russ.)
6. Zavyalov Yu.S., Kvasov B.I., Miroshnichenko V.L. Metody splayn-funktsiy [Methods of spline functions]. Moscow: Nauka; 1980: 352. (In Russ.)
7. Wang Y. Smoothing splines methods and applications. BOOK SERIES: Chapman & Hall/CRC Monographs on Statistics and Applied Probability. 2011; (121): 347.
8. Wahba G. Smoothing noisy data with spline functions. *Numerische Mathematik*. 1975; 24(2):383–393.
9. Voskoboynikov Yu.E., Krysov D.A. Estimation of the noise measurement characteristics in the model “Signal + Noise”. *Automatics and Software Enginery*. 2018; 3(25):54–61. (In Russ.)
10. Voskoboynikov Yu.E. Method of removal of wavelet filtration artifacts. *Fundamental Research*. 2017; (8-2):246–250. (In Russ.)