

Научная статья

УДК 519.237.5:519.854.33

DOI 10.35266/1999-7604-2023-3-99-104

ИДЕНТИФИКАЦИЯ ПАРАМЕТРОВ КУСОЧНО-ЛИНЕЙНОЙ ФУНКЦИИ РИСКА МЕТОДОМ АНТИРОБАСТНОГО ОЦЕНИВАНИЯ

Сергей Иванович Носков^{1✉}, *Владимир Викторович Тирских*²

^{1, 2} *Иркутский государственный университет путей сообщения, Иркутск, Россия*

¹ *sergey.noskov.57@mail.ru* ✉, <https://orcid.org/0000-0003-4097-2720>

² *tirskikh1963@mail.ru*

Аннотация. В работе дан краткий обзор публикаций по методам идентификации параметров регрессионных моделей, основанных в том числе на метрике Чебышева. Они касаются, в частности: разработки алгоритма однозначного определения чебышевской проекции; нового метода, сочетающего расстояние Минковского с расстоянием Чебышева, которые используются в качестве меры подобия в процессе кластеризации при группировке данных; обобщения частной постановки задачи подгонки кривых или поверхностей к наблюдаемым или измеренным данным, связанной с заменой наименьших квадратов нормой Чебышева; интегральных оценок антропогенной трансформации территории с использованием многомерных статистических методов. Разработан способ оценивания неизвестных параметров регрессионной кусочно-линейной функции риска методом антиробастного оценивания, сводящийся к решению задачи линейно-булевого программирования. Построена функция риска динамики цены одного квадратного метра общей площади квартир на рынке жилья в Российской Федерации с помощью методов наименьших модулей и антиробастного оценивания. В качестве независимых факторов в модели использованы средние цены на блоки стеновые силикатные, плиты перекрытий железобетонные и товарный бетон. Высокие значения критериев адекватности указывают на то, что оба построенных варианта модели достаточно хорошо описывают динамику выходного показателя и могут успешно использоваться для решения задач, связанных с прогнозированием. Установлено, что число максимальных по модулю ошибок аппроксимации модели риска при использовании метода антиробастного оценивания равно трем, т. е. числу независимых переменных.

Ключевые слова: регрессионная модель, кусочно-линейная функция риска, метод антиробастного оценивания, задача линейно-булевого программирования, цены на жилье

Для цитирования: Носков С. И., Тирских В. В. Идентификация параметров кусочно-линейной функции риска методом антиробастного оценивания // Вестник кибернетики. 2023. Т. 22, № 3. С. 99–104. DOI 10.35266/1999-7604-2023-3-99-104.

Original article

IDENTIFYING PARAMETERS OF A PIECEWISE LINEAR RISK FUNCTION WITH A METHOD OF ANTIROBUST ESTIMATION

Sergey I. Noskov^{1✉}, *Vladimir V. Tirskikh*²

^{1, 2} *Irkutsk State Transport University, Irkutsk, Russia*

¹ *sergey.noskov.57@mail.ru* ✉, <https://orcid.org/0000-0003-4097-2720>

² *tirskikh1963@mail.ru*

Abstract. The article reviews literature on identification methods for parameters of regression models, including those based on Chebyshev metrics. The publications contain data on the development of an algorithm for the unambiguous definition of Chebyshev projection; a new method that combines Minkowskian distance and Chebyshev distance, with both being used as a similarity measure in the clustering process when grouping data; generalizations of particular goal setting for curves and surface fitting to the data observed or calculated

as a result of replacing least squares with the Chebyshev norm; and integral estimates of the anthropogenic transformation of the territory using multidimensional statistical methods. Using the anti-robust method of estimation, the authors have developed a method to estimate unknown parameters of a regression piecewise linear risk function, whose aim is to solve a linear Boolean programming problem. The risk function of the dynamics of the price per square meter of living space of dwellings in the housing market in the Russian Federation is built using the least modules method and the anti-robust estimation method. Average pricing for silicate wall blocks, concrete floor slabs, and ready-mixed concrete are used as independent factors in the model. Both versions of the models built describe the dynamics of the output indicator profoundly, as evidenced by the high values of adequacy criteria, and therefore can efficiently solve the forecasting problems. It has been established that the number of maximum module errors of the risk model approximation is equal to three, i. e. the number of independent variables, when applying the anti-robust method.

Keywords: regression model, piecewise linear risk function, method of antirobust estimation, linear Boolean programming problem, housing prices

For citation: Noskov S. I., Tirskikh V. V. Identifying parameters of a piecewise linear risk function with a method of antirobust estimation. *Proceedings in Cybernetics*. 2023;22(3):99–104. DOI 10.35266/1999-7604-2023-3-99-104.

ВВЕДЕНИЕ

Методы регрессионного анализа активно используются при исследовании сложных систем средствами математического моделирования. При этом идентификация параметров регрессионных моделей производится, как правило, посредством минимизации расстояния между вычисленными и фактическими значениями зависимых модельных переменных. В регрессионном анализе существует несколько способов задания этого расстояния: евклидово расстояние, соответствующее методу наименьших квадратов (МНК) [1–3], городское расстояние (метод наименьших модулей – МНМ) [4–6], расстояние Чебышева (метод антиробастного оценивания – МАО) [7]. Следует отметить, что исследованию и применению расстояния Чебышева в рамках прикладной математики посвящено значительное количество работ. В работе [8] представлен алгоритм однозначного определения так называемой чебышевской проекции. В [9] используется новый метод, сочетающий расстояние Минковского с расстоянием Чебышева, которое используется в качестве меры подобия в процессе кластеризации при группировке данных. В статье [10] рассматривается естественное обобщение частной постановки задачи подгонки кривых или поверхностей к наблюдаемым или измеренным данным, связанной с заменой наименьших квадратов нормой Чебышева. Этот критерий может быть более подходящим в контексте решений о принятии/отклонении изготов-

ленных деталей. Получившаяся задача имеет некоторые интересные особенности: она имеет много структур, которые можно использовать, но обычно решение не является уникальным. В работе [11] расстояние Чебышева применяется при получении интегральных оценок антропогенной трансформации территории с использованием многомерных статистических методов. В работе [12] приведено сравнение известных метрик с точки зрения их вычислительной сложности. Сделано предположение о возможном использовании комбинированных метрик расстояния.

МАТЕРИАЛЫ И МЕТОДЫ

Рассмотрим регрессионную кусочно-линейную функцию риска [13]:

$$y_k = \max\{\alpha_1 x_{k1}, \alpha_2 x_{k2}, \dots, \alpha_m x_{km}\} + \varepsilon_k, \quad (1)$$
$$k = \overline{1, n},$$

где k – номер наблюдения;

y и x_i , $i = \overline{1, m}$ – соответственно зависимая (выходная, внутренняя) и независимые (входные, внешние) переменные, значения которых считаются заданными;

α_i , $i = \overline{1, m}$ – подлежащие определению оценки параметров;

ε_k – ошибки приближения;

n – длина выборки.

Отметим, что кусочно-линейные конструкции гораздо реже линейных применяются при

исследовании сложных объектов в силу более высокой вычислительной сложности идентификации неизвестных параметров. В то же время, кусочно-линейные модели весьма содержательны и хорошо интерпретируются. В качестве как входных, так и выходных факторов здесь могут использоваться негативные (вызывающие регресс и/или стагнацию) для объекта исследования показатели, такие как убытки, загрязнения, риски, технические сбои, отказы оборудования, ущерб и т. д. При этом, в соответствии с (1), значение переменной y определяется максимальным значением одного из негативных независимых факторов, а любое уменьшение значений других переменных на это обстоятельство не влияет.

Для определения значений параметров модели (1) необходимо минимизировать функцию потерь, соответствующую выбранному методу идентификации. Для МНМ это:

$$J_1(\alpha) = \sum_{k=1}^n |\varepsilon_k| \rightarrow \min, \quad (2)$$

для МНК:

$$J_2(\alpha) = \sum_{k=1}^n \varepsilon_k^2 \rightarrow \min,$$

для MAO:

$$J_\infty(\alpha) = \lim_{p \rightarrow \infty} \sum_{k=1}^n |\varepsilon_k|^p = \max_{k=1, n} |\varepsilon_k| \rightarrow \min. \quad (3)$$

Три упомянутых выше метода определения оценок параметров регрессионных моделей обладают следующими свойствами по отношению к выбросам (наблюдениям, не согласующимся с выборкой данных в целом): МНМ их игнорирует; MAO, напротив, к ним тяготеет; МНК является в этом отношении своего рода компромиссом между МНМ и MAO.

В работе [13] показано, что задача (2) может быть сведена к следующей задаче линейно-булевого программирования (ЛБП):

$$z_k + u_k - v_k = y_k, k = \overline{1, n}, \quad (4)$$

$$z_k \geq \alpha_i x_{ki}, k = \overline{1, n}, i = \overline{1, m}, \quad (5)$$

$$\alpha_i x_{ki} - z_k \geq (\sigma_{ki} - 1)M, k = \overline{1, n}, i = \overline{1, m}, \quad (6)$$

$$\sum_{i=1}^m \sigma_{ki} = 1, k = \overline{1, n}, \quad (7)$$

$$u_k \geq 0, v_k \geq 0, k = \overline{1, n}, \quad (8)$$

$$\sigma_{ki} \in \{0, 1\}, k = \overline{1, n}, i = \overline{1, m}, \quad (9)$$

$$\sum_{k=1}^n (u_k + v_k) \rightarrow \min, \quad (10)$$

где M – заранее выбранное большое положительное число,

$$u_k + v_k = |\varepsilon_k|, u_k - v_k = \varepsilon_k, \\ z_k = \max\{\alpha_1 x_{k1}, \alpha_2 x_{k2}, \dots, \alpha_m x_{km}\}.$$

Решение задачи ЛБП (4)–(10) не должно вызывать вычислительных трудностей ввиду наличия эффективных компьютерных программ, в частности, LPSolve [14].

Для сведения задачи (3) к задаче ЛБП введем новую неизвестную переменную r :

$$r = \max_{k=1, n} |\varepsilon_k|.$$

Тогда будут справедливы неравенства:

$$u_k + v_k - r \leq 0, k = \overline{1, n}. \quad (11)$$

Таким образом, эквивалентная задаче (3) задача ЛБП будет включать в себя ограничения (4)–(9), (11) и целевую функцию:

$$r \rightarrow \min. \quad (12)$$

РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

Применим полученный способ идентификации неизвестных параметров кусочно-линейной функции риска с помощью MAO для построения модели средней цены одного m^2 общей площади квартир на рынке жилья в Российской Федерации, воспользовавшись поквартальной статистической информацией – с 1 кв. 2017 по 4 кв. 2020 гг. [15].

Введем следующие обозначения:

- y – цена одного m^2 общей площади квартир;
- x_1 – средние цены на блоки стеновые силикатные, руб./ m^3 ;
- x_2 – средние цены на плиты перекрытий железобетонные, руб./ m^3 ;
- x_3 – средние цены на товарный бетон, руб./ m^3 .

Будем строить функцию риска вида:

$$y_k = \max\{\alpha_1 x_{k1}, \alpha_2 x_{k2}, \alpha_3 x_{k3}\} + \varepsilon_k, k = \overline{1, 16}. \quad (13)$$

Применение МНМ путем решения задачи ЛБП (4)–(10) приводит к построению следующей модели:

$$y = \max\{17,14x_1, 7,95x_2, 13,8x_3\}, \quad (14)$$

$$E = 4,35 \%, r = 9\,872,88.$$

Здесь E – средняя относительная ошибка аппроксимации.

Рассчитаем для модели (14) ее вектор срабатываний λ , каждая компонента которого указывает на номер независимой переменной, на которой реализовался максимум в модели (1):

$$\lambda = (2, 2, 3, 2, 2, 2, 2, 2, 3, 3, 2, 2, 2, 2, 2, 2).$$

Использование МАО посредством решения задачи ЛБП (4)–(9), (11)–(12) позволяет получить следующую модель:

$$y = \max\{19,16x_1, 7,65x_2, 15,09x_3\}, \quad (15)$$

$$E = 6,37 \%, r = 6\,949,76,$$

$$\lambda = (2, 3, 1, 3, 3, 3, 3, 3, 3, 3, 3, 3, 1, 3, 3).$$

В таблице представлены результаты моделирования – фактические (y) и расчетные ($y^{\text{МНМ}}$ и $y^{\text{МАО}}$) значения зависимой переменной моделей (14) и (15), а также их ошибки аппроксимации ($\varepsilon^{\text{МНМ}}$ и $\varepsilon^{\text{МАО}}$).

Таблица

Результаты моделирования

Номер наблюдения	y	$y^{\text{МНМ}}$	$\varepsilon^{\text{МНМ}}$	$y^{\text{МАО}}$	$\varepsilon^{\text{МАО}}$
1	56 347,20	64 731,83	–8 384,6	62 257,03	–5 909,8
2	56 516,78	61 060,72	–4 543,9	61 554,50	–5 037,7
3	56 560,78	56 826,36	–265,58	63 510,40	–6 949,6
4	56 882,19	60 257,88	–3 375,6	63 831,94	–6 949,7
5	58 875,59	59 330,75	–455,16	63 816,85	–4 941,2
6	59 969,66	61 022,44	–1 052,7	62 936,66	–2 967,0
7	60 952,83	60 952,98	–0,15	62 658,21	–1 705,3
8	61 831,57	59 679,26	2 152,30	63 860,47	–2 028,9
9	60 705,14	60 741,04	–35,90	66 407,91	–5 702,7
10	61 618,25	58 472,13	3 146,11	63 927,33	–2 309,0
11	62 891,94	63 957,95	–1 066,0	64 207,29	–1 315,3
12	64 059,49	63 671,74	387,74	65 141,36	–1 081,8
13	71 503,24	70 070,36	1 432,87	67 391,49	4 111,7
14	73 438,05	69 645,07	3 792,97	71 468,37	1 969,6
15	76 167,22	70 770,64	5 396,57	72 843,61	3 323,6
16	79 003,00	69 130,11	9 872,88	72 053,23	6 949,7

Примечание: составлено авторами.

Сопоставимо высокие значения критериев адекватности E и r указывают, что модели (14) и (15) достаточно хорошо описывают динамику выходного показателя (этот тезис подтверждает и рисунок) и могут успешно использоваться для решения задач, связанных, в частности, с прогнозированием.

Модель (15) несколько уступает зависимости (14) по критерию E , зато превосходит ее по критерию r . Но что еще более важно, в вектор срабатываний для первой модели

не вошла переменная x_3 , а для второй модели состав этого вектора полный – сработали все переменные.

Обращает на себя внимание еще один весьма интересный факт – число максимальных по модулю ошибок аппроксимации модели (13) при использовании МАО равно трем, т. е. числу независимых переменных. При использовании МАО для обычной линейной регрессии без свободного члена это число равнялось бы четырем [7].

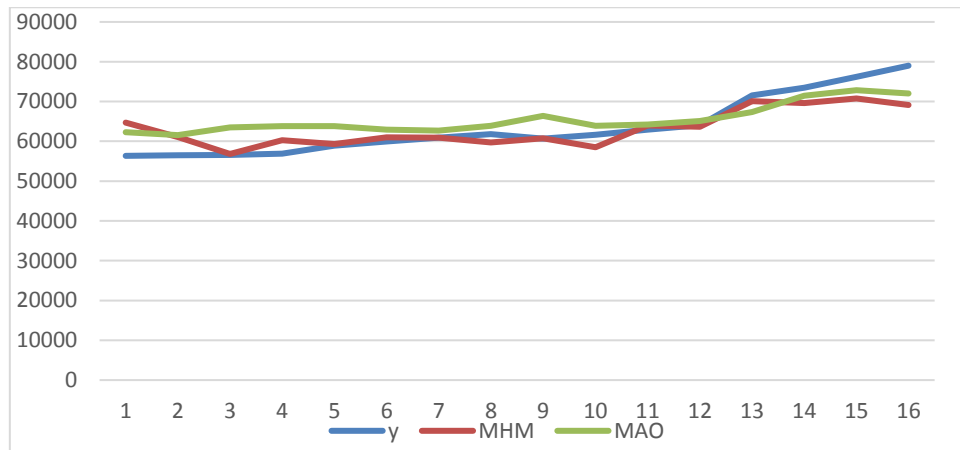


Рисунок. Фактические и расчетные значения зависимой переменной

Примечание: составлено авторами.

ЗАКЛЮЧЕНИЕ

В работе предложен способ оценивания неизвестных параметров регрессионной кусочно-линейной функции риска методом антиробастного оценивания, сводящийся к решению задачи линейно-булевого программирования. Построена функция риска динамики цены одного квадратного метра общей площади квартир на рынке жилья в Российской Федерации с помощью методов наименьших модулей и антиробастного оценивания. В качестве независимых факторов в модель включены средние цены: на блоки стеновые силикатные, плиты перекрытий железобетонные и на

товарный бетон. Сопоставимо высокие значения критериев адекватности указывают на то, что оба построенных варианта модели достаточно хорошо описывают динамику выходного показателя и могут успешно использоваться для решения задач, связанных, в частности, с прогнозированием. Выявлен тот факт, что число максимальных по модулю ошибок аппроксимации модели риска при использовании МАО равно трем, т. е. числу независимых переменных. Для обычной линейной регрессии без свободного члена это число равнялось бы четырем.

Список источников

1. Айвазян С. А. Методы эконометрики. М. : Магистр ; Инфра-М, 2010. 506 с.
2. Орлов А. И. Прикладная статистика. М. : Экзамен, 2004. 656 с.
3. Greene W. H. Econometric analysis. NY: New York University; 2002. 1026 p.
4. Тырсин А. Н., Азарян А. А. Оценивание нелинейных регрессионных зависимостей на основе обобщенного метода наименьших модулей // Обзорные прикладной и промышленной математики. 2018. Т. 25, № 2. С. 185–187.
5. Панюков А. В., Тырсин А. Н. Взаимосвязь взвешенного и обобщенного вариантов метода наименьших модулей // Известия Челябинского научного центра УрО РАН. 2007. № 1. С. 6–11.
6. Миллер Б. М., Колосов К. С. Робастное оценивание на основе метода наименьших модулей и фильтра Калмана // Автоматика и телемеханика. 2020. № 11. С. 72–92. DOI 10.31857/S0005231020110057.
7. Носков С. И. Метод антиробастного оценивания параметров линейной регрессии: число максимальных

References

1. Aivazyan S. A. Metody ekonometriki. Moscow: Magistr; Infra-M; 2010. 506 p. (In Russian).
2. Orlov A. I. Prikladnaia statistika. Moscow: Ekzamen; 2004. 656 p. (In Russian).
3. Greene W. H. Econometric analysis. NY: New York University; 2002. 1026 p.
4. Tyrsin A. N., Azaryan A. A. Otsenivanie nelineinykh regressionnykh zavisimostei na osnove obobshchennogo metoda naimenshikh modulei. *Obozrenie prikladnoi i promyshlennoi matematiki*. 2018;25(2):185–187. (In Russian).
5. Panyukov A. V., Tyrsin A. N. Vzaimosviaz vzveshennogo i obobshchennogo variantov metoda naimenshikh modulei. *Izvestiia Chelyabinskogo nauchnogo tsentra UrO RAN*. 2007;(1):6–11. (In Russian).
6. Miller B. M., Kolosov K. S. Robust estimation based on the least absolute deviations method and the Kalman filter. *Avtomatika i telemekhanika*. 2020;(11):72–92. DOI 10.31857/S0005231020110057. (In Russian).
7. Noskov S. I. Method of antirobast estimation of linear regression parameters: Number of maximum on the

- ных по модулю ошибок аппроксимации // Южно-Сибирский научный вестник. 2020. № 1. С. 51–54.
8. Зоркальцев В. И. Чебышевские приближения могут обходиться без условия Хаара // Динамические системы, оптимальное управление и математическое моделирование : материалы Междунар. симпозиума, посвящ. 100-летию математического образования в Восточной Сибири и 80-летию со дня рождения профессора О. В. Васильева, 07–12 октября 2019 г., г. Иркутск. Иркутск : Иркут. гос. ун-т, 2019. С. 29–33.
 9. Surono S., Putri R. D. A. Optimization of Fuzzy C-Means clustering algorithm with combination of Minkowski and Chebyshev distance using principal component analysis. *Int J Fuzzy Syst.* 2021;23:139–144.
 10. Al-Subaihi I., Watson G. A. Fitting parametric curves and surfaces by l_∞ distance regression. *Bit Numer Math.* 2005;45:443–466.
 11. Шабанов Д. И., Иолин М. М., Борзова А. С. и др. Методика получения интегральных оценок антропогенной трансформации территории с использованием многомерных статистических методов // Геология, география и глобальная энергия. 2014. № 4. С. 176–185.
 12. Сидорович А. С., Сасин Е. А. Сравнительная характеристика основных метрик расстояний // Математические методы в технике и технологиях : материалы XXXI Междунар. науч. конф., 22–26 октября 2018 г., г. Минск. Т. 8. Минск, 2018. С. 82–84.
 13. Носков С. И. Идентификация параметров кусочно-линейной функции риска // Транспортная инфраструктура Сибирского региона : материалы Восьмой Междунар. науч.-практ. конф., 28 марта – 01 апреля 2017 г., г. Иркутск. В 2 т. Т. 1. Иркутск : ИРГУПС, 2017. С. 417–421.
 14. Шипицына Р. Е., Витвицкий Е. Е. Сравнение удобства использования программных продуктов при решении транспортной задачи линейного программирования: LPSolve IDE и Microsoft Excel // Образование. Транспорт. Инновации. Строительство : сб. материалов V Национал. науч.-практич. конф., 28–29 апреля 2022 г., г. Омск. Омск : Сибир. гос. автомобильно-дорожный ун-т (СибАДИ), 2022. С. 250–254.
 15. Носков С. И., Хоняков А. А. Применение функции риска для модельного описания колебания цен на рынке недвижимости // Инженерно-строительный вестник Прикаспия. 2021. № 3. С. 77–82. DOI 10.52684/2312-3702-2021-37-3-77-82.
- module of approximation errors. *South-Siberian Scientific Bulletin.* 2020;(1):51–54. (In Russian).
8. Zorkaltsev V. I. Chebyshevskie priblizheniia mogut obkhoditsia bez usloviia Khaara. In: *Proceedings of the International Symposium devoted to the 100th Anniversary of Mathematical Education in the East Siberia and 80th Anniversary of birth of Prof. Vasilyev O. V. "Dinamicheskie sistemy, optimalnoe upravlenie i matematicheskoe modelirovanie"*, October 7–12, 2019, Irkutsk. Irkutsk: Irkutsk State University; 2019. p. 29–33. (In Russian).
 9. Surono S., Putri R. D. A. Optimization of Fuzzy C-Means clustering algorithm with combination of Minkowski and Chebyshev distance using principal component analysis. *Int J Fuzzy Syst.* 2021;23:139–144.
 10. Al-Subaihi I., Watson G. A. Fitting parametric curves and surfaces by l_∞ distance regression. *Bit Numer Math.* 2005;45:443–466.
 11. Shabanov D. I., Iolin M. M., Borzova A. S. et al. Use of multivariate statistical methods for estimation of the spatial changes in anthropogenic transformation (land use). *Geology, Geography and Global Energy.* 2014;(4):176–185. (In Russian).
 12. Sidorovich A. S., Sasin E. A. Srovnitelnaia kharakteristika osnovnykh metrik rasstoianii. In: *Proceedings of the XXXI International Scientific Conference "Matematicheskie metody v tekhnike i tekhnologiiakh"*, October 22–26, 2018, Minsk. Vol. 8. Minsk; 2018. p. 82–84. (In Russian).
 13. Noskov S. I. Identifikatsiia parametrov kusochno-lineinnoi funktsii riska. In: *Proceedings of the Eighth International Research-to-Practice Conference "Transportnaia infrastruktura Sibirskogo regiona"*, March 28 – April 1, 2017, Irkutsk. In 2 vols. Vol. 1. Irkutsk: IrGUPS; 2017. p. 417–421. (In Russian).
 14. Shipitsyna R. E., Vitvitsky E. E. Comparison of the convenience of using two software products in solving the transport problem of linear programming: LPSolve IDE and Microsoft Excel. In: *Proceedings of the V National Research-to-Practice Conference "Obrzovanie. Transport. Innovatsii. Stroitelstvo"*, April 28–29, 2022, Omsk. Omsk: The Siberian State Automobile and Highway University (SibADI); 2022. p. 250–254. (In Russian).
 15. Noskov S. I., Khonyakov A. A. Application of risk function to model description of price fluctuations in the real estate market. *Engineering and Construction Bulletin of the Caspian Region.* 2021;(3):77–82. DOI 10.52684/2312-3702-2021-37-3-77-82. (In Russian).

Информация об авторах

С. И. Носков – доктор технических наук, профессор.
В. В. Тирских – кандидат физико-математических наук, доцент.

Information about the authors

S. I. Noskov – Doctor of Sciences (Engineering), Professor.
V. V. Tirsikh – Candidate of Sciences (Physics and Mathematics), Docent.