

## КРАТКИЕ СООБЩЕНИЯ

## BRIEF PAPERS

doi: 10.17586/2226-1494-2025-25-1-169-173

УДК 004.852

**Использование метода динамического расширения и смешивания регрессора  
для задач машинного обучения на примере перцептрона**

**Алексей Анатольевич Маргун<sup>1</sup>, Константин Александрович Зименко<sup>2</sup>,  
Алексей Алексеевич Бобцов<sup>3</sup>**

<sup>1,2,3</sup> Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация

<sup>1</sup> Институт проблем машиноведения РАН, Санкт-Петербург, 199178, Российская Федерация

<sup>1</sup> [alexeimargun@gmail.com](mailto:alexeimargun@gmail.com), <https://orcid.org/0000-0002-5333-0594>

<sup>2</sup> [kostyazimenko@gmail.com](mailto:kostyazimenko@gmail.com), <https://orcid.org/0000-0001-6220-7494>

<sup>3</sup> [bobtssov@mail.ru](mailto:bobtssov@mail.ru), <https://orcid.org/0000-0003-1854-6717>

**Аннотация**

Рассмотрено использование метода динамического расширения и смешивания регрессора для повышения скорости обучения в задачах машинного обучения. Предложенный подход продемонстрирован на примере перцептрона, применяемого для задач регрессии и бинарной классификации. Метод позволяет преобразовать мультипараметрическую задачу оптимизации в набор независимых скалярных регрессий, что значительно ускоряет сходимость алгоритма и снижает вычислительные затраты. Результаты компьютерного моделирования, включающие сравнение с методами стохастического градиентного спуска и Adam, подтвердили преимущество предложенного подхода для скорости сходимости и эффективности вычислений.

**Ключевые слова**

перцептрон, регрессия, бинарная классификация, машинное обучение, динамическое расширение и смешивание регрессора

**Благодарности**

Исследование поддержано Министерством науки и высшего образования Российской Федерации (проект № FSER-2025-0002).

**Ссылка для цитирования:** Маргун А.А., Зименко К.А., Бобцов А.А. Использование метода динамического расширения и смешивания регрессора для задач машинного обучения на примере перцептрона // Научно-технический вестник информационных технологий, механики и оптики. 2025. Т. 25, № 1. С. 169–173. doi: 10.17586/2226-1494-2025-25-1-169-173

**Application of the dynamic regressor extension and mixing approach  
in machine learning on the example of perceptron**

**Alexey A. Margun<sup>1</sup>, Konstantin A. Zimenko<sup>2</sup>, Alexey A. Bobtsov<sup>3</sup>**

<sup>1,2,3</sup> ITMO University, Saint Petersburg, 197101, Russian Federation

<sup>1</sup> Institute for Problems in Mechanical Engineering of the Russian Academy of Sciences, Saint Petersburg, 199178, Russian Federation

<sup>1</sup> [alexeimargun@gmail.com](mailto:alexeimargun@gmail.com), <https://orcid.org/0000-0002-5333-0594>

<sup>2</sup> [kostyazimenko@gmail.com](mailto:kostyazimenko@gmail.com), <https://orcid.org/0000-0001-6220-7494>

<sup>3</sup> [bobtssov@mail.ru](mailto:bobtssov@mail.ru), <https://orcid.org/0000-0003-1854-6717>

**Abstract**

This paper explores the application of the Dynamic Regressor Extension and Mixing method to improve the learning speed in machine learning tasks. The proposed approach is demonstrated using a perceptron applied to regression and binary classification problems. The method transforms a multi-parameter optimization problem into a set of independent

scalar regressions, significantly accelerating the convergence of the algorithm and reducing computational costs. Results from computer simulations, including comparisons with stochastic gradient descent and Adam methods, confirm the advantages of the proposed approach in terms of convergence speed and computational efficiency.

#### Keywords

perceptron, regression, binary classification, machine learning, dynamic regressor extension and mixing

#### Acknowledgments

The research was supported by the Ministry of Science and Higher Education of the Russian Federation (project No. FSER-2025-0002).

**For citation:** Margun A.A., Zimenko K.A., Bobtsov A.A. Application of the dynamic regressor extension and mixing approach in machine learning on the example of perceptron. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2025, vol. 25, no. 1, pp. 169–173 (in Russian). doi: 10.17586/2226-1494-2025-25-1-169-173

Перцептрон является одним из фундаментальных элементов машинного обучения. Разработанный в середине прошлого века, он представляет собой простейшую архитектуру нейронной сети [1, 2]. Его концепция, основанная на попытке имитировать биологический нейрон, стала фундаментом для создания современных нейронных сетей. Процесс обучения перцептрона представляет собой оптимизационную задачу поиска весов, обеспечивающих минимизацию некоторого целевого функционала. Для решения данной задачи существует множество подходов, но наиболее распространенными являются: метод наименьших квадратов, пакетный градиентный спуск, стохастический градиентный спуск, мини-пакетный градиентный спуск и его модификации (оптимизатор импульса, Adam и т. д.) [3–7]. В настоящей работе на основе применения мини-пакетного подхода предлагается использование/адаптация метода динамического расширения и смешивания регрессора (ДРСР) [8] для обучения перцептрона при решении двух классических задач машинного обучения, а именно: регрессии и бинарной классификации. Отметим, что метод ДРСР в настоящее время активно используется специалистами в области теории автоматического управления при решении задач параметрической идентификации линейных статических регрессионных моделей.

Цель работы — сравнение с известными подходами метода ДРСР, а также подтверждение эффективности его использования для повышения скорости обучения, т. е. уменьшения количества итераций для приведения целевого функционала к окрестности минимума.

Задача обучения перцептрона для регрессии заключается в построении модели, которая предсказывает непрерывное значение целевой переменной  $y$  на основе набора входных признаков  $\mathbf{X}$ . Предположим, что дана выборка обучающих данных  $(\mathbf{X}_i, y_i)$ ,  $i = \overline{1, N}$ , где  $N$  — число примеров данных;  $\mathbf{X}_i = [1 \ x_{i1} \ x_{i2} \ \dots \ x_{ip}] \in \mathbb{R}^{p+1}$  — вектор признаков размерности  $p + 1$  для  $i$ -го примера;  $y_i \in \mathbb{R}$  — соответствующее значение целевой переменной.

Пусть поставлена задача поиска таких значений весов  $\mathbf{w}^T = [w_0 \ w_1 \ \dots \ w_p] \in \mathbb{R}^{p+1}$ , которые минимизируют разницу между предсказанными  $\hat{y}_i$  и истинными значениями  $y_i$ , где функция предсказания для перцептрона имеет вид:

$$\hat{y}_i = \mathbf{X}_i \mathbf{w}.$$

Традиционно для оценки качества модели в виде функции потерь используется среднеквадратичная

ошибка, которая минимизируется с целью нахождения оптимальных значений параметров  $L_1(\mathbf{w}) = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2$ . В результате обученный перцептрон должен предсказывать значения целевой переменной для новых данных на основе установленной линейной зависимости. Задача обучения перцептрона для классификации заключается в построении модели, которая способна разделять данные на два класса  $C_1$  и  $C_2$ . Модель представляет собой линейную регрессию, к выходу которой добавлена сигмоидная функция активации. Предполагается, что дана обучающая выборка  $(\mathbf{X}_i, y_i)$ ,  $i = \overline{1, N}$ , где  $N$  — число примеров данных;  $y_i \in \{0, 1\}$  — метка класса, где 0 обозначает принадлежность к классу  $C_1$ , а 1 — к классу  $C_2$ . Целью обучения является поиск таких значений весов, которые позволяют классифицировать входные данные, определяя принадлежность к одному из классов. Для этого используется сигмоидная функция активации, которая преобразует линейную комбинацию входных признаков в вероятность принадлежности к классу  $C_2$ . Функция предсказания имеет вид  $\hat{y}_i = \sigma(\mathbf{X}_i \mathbf{w})$ , где  $\sigma(z) = (1 + e^{-z})^{-1}$  — сигмоидная функция, значение которой лежит в интервале  $(0, 1)$  и интерпретируется как вероятность принадлежности к классу  $C_2$ .

Для оценки качества классификации используется логистическая функция потерь (бинарная кросс-энтропия), которая минимизируется с целью нахождения оптимальных значений весов:  $L_2(\mathbf{w}) = -\frac{1}{N} \sum_{i=1}^N (y_i \ln \hat{y}_i + (1 - y_i) \ln(1 - \hat{y}_i))$ .

Рассмотрим предлагаемый для решения данной задачи метод ДРСР. Оригинальная постановка задачи выглядит следующим образом [8]. Имеется линейная регрессия вида

$$y(t) = \mathbf{m}^T(t) \mathbf{w}, \quad (1)$$

где  $y(t) \in \mathbb{R}$  и  $\mathbf{m}(t) \in \mathbb{R}^q$  — известные ограниченные функции времени;  $\mathbf{w}$  — вектор неизвестных параметров, которые требуется оценить. Известно [9], что классический метод градиентного спуска позволяет решить данную задачу, но он имеет ряд существенных недостатков. Один из основных недостатков — допущение о том, что регрессор  $\mathbf{m}(t)$  удовлетворяет условию незатухающего возбуждения [9]. К другим недостаткам относятся немонотонность переходных процессов и неконструктивные алгоритмы улучшения качества параметрической сходимости. Для преодоления данных

ограничений и, в первую очередь, повышения скорости параметрической сходимости в работе [8] предложен следующий метод ДРСР.

1. Применим к уравнению (1)  $q - 1$  разных динамических устойчивых операторов  $H_i$  и получим  $q - 1$  линейных регрессий вида

$$y_{f_i}(t) = \mathbf{m}_{f_i}^T(t)\mathbf{w},$$

где  $y_{f_i} = H_i[y(t)]$ ;  $\mathbf{m}_{f_i}^T(t) = H_i\mathbf{m}_{f_i}^T$ . В качестве операторов могут быть выбраны, например, линейные фильтры первого порядка  $(H_i[\cdot])(t) = \frac{\alpha_i}{s + \alpha_i}[\cdot](t)$ ,  $\alpha_i > 0$ ,  $s = d/dt$  или операторы запаздывания  $(H_i[\cdot])(t) = [\cdot](t - d_i)$ ,  $d_i > 0$ .

2. Сформируем расширенную регрессионную модель

$$\mathbf{Y}_e(t) = \mathbf{M}(t)\mathbf{w}, \quad (2)$$

где  $\mathbf{Y}_e^T(t) = [y(t) \ y_{f_1}(t) \ \dots \ y_{f_{q-1}}(t)]$ ;  $\mathbf{M}^T(t) = [m(t) \ m_{f_1}(t) \ \dots \ m_{f_{q-1}}(t)]$ .

3. Умножим уравнение (2) на союзную матрицу  $\text{adj}\{\mathbf{M}(t)\}$  и получим  $q$  независимых скалярных линейных регрессий

$$Y_i(t) = \Delta(t)w_i, \quad i = 1, \dots, q, \quad (3)$$

где  $Y_i(t)$  —  $i$ -й элемент вектора  $\mathbf{Y}(t) = \text{adj}\{\mathbf{M}(t)\}\mathbf{Y}_e(t)$ ;  $\Delta(t) = \det\{\mathbf{M}(t)\}$ .

4. Далее каждый неизвестный параметр может быть определен с помощью градиентного спуска для соответствующей скалярной линейной регрессии (3). Запишем решение (3) в непрерывном виде в соответствии с [8]

$$\hat{w}_i(t) = \gamma_i \Delta(t)(Y_i(t) - \Delta(t)\hat{w}_i(t)), \quad i = 1, \dots, q,$$

где  $\hat{w}_i(t)$  — оценка параметра  $w_i$ .

Таким образом, применение метода ДРСР позволяет преобразовать задачу мультипараметрической оптимизации к набору скалярных регрессионных задач. Поскольку требуется оптимизировать только одно направление, метод дает возможность получить более высокие показатели качества в сравнении с классическим градиентным спуском, примененным для уравнения (1).

Перейдем непосредственно к использованию метода ДРСР для задач обучения перцептрона. Применение динамических операторов в рассматриваемой постановке задачи не представляет смысла или невозможно, так как признаки и целевая переменная не являются функциями времени. По этой причине вместо расширения регрессора воспользуемся следующим алгоритмом.

Шаг 1. Случайным образом зададим значения весов, шаг обучения  $\alpha$ , количество эпох  $K$ .

Шаг 2. Выполним следующие действия для каждой эпохи обучения  $k$ ,  $k = 1, 2, \dots, K$ .

Шаг 2.1. Случайным образом разобьем данные на мини-пакеты размером, равным числу весов. Обозначим данные пакета  $\mathbf{X}^{bj} \in \mathbb{R}^{(p+1) \times (p+1)}$ ,  $\mathbf{Y}_e^j \in \mathbb{R}^{p+1}$ , где  $j$  — номер мини-пакета.

Шаг 2.2. Для каждого мини-пакета имеем

$$\mathbf{Y}_e^j = \mathbf{X}^{bj}\mathbf{w}. \quad (4)$$

— Умножим слева уравнение (4) на  $\mathbf{X}^{*j} = \text{adj}\{\mathbf{X}^{bj}\}$ . Тогда

$$\mathbf{Y}^j = \text{diag}\{\Delta^j\}\mathbf{w},$$

где  $\mathbf{Y}^j = \mathbf{X}^{*j}\mathbf{Y}_e^j$ ,  $\mathbf{X}^{*j}\mathbf{X}^{bj} = \text{diag}\{\Delta^j\}$ ,  $\Delta^j = \det(\mathbf{X}^{bj})$ .

Для снижения вычислительной сложности вместо использования затратной операции вычисления союзной матрицы воспользуемся правилом Крамера

$$Y_i^j = \det(\mathbf{X}^{bj,i}),$$

где  $\mathbf{X}^{bj,i}$  — матрица  $\mathbf{X}^{bj}$ , в которой  $i$ -й столбец заменен на вектор  $\mathbf{Y}_e^j$  [8];  $Y_i^j$  —  $i$ -й элемент вектора  $\mathbf{Y}^j$ .

— После проведенных преобразований вместо решения задачи для регрессии с  $p + 1$  неизвестными, требуется рассмотреть  $p + 1$  скалярное уравнение

$$Y_i^j = \Delta^j w_i, \quad i = 1, \dots, p + 1.$$

— Далее для определения весов воспользуемся градиентным спуском

$$w_i := w_i - \alpha \Delta^j (Y_i^j - \Delta^j w_i). \quad (5)$$

Заметим, что для повышения устойчивости вместо (5) может быть использована формула обновления весов

$$w_i := w_i - \alpha \Delta^j (Y_i^j - \Delta^j w_i) / (1 + \alpha (\Delta^j)^2).$$

Шаг 3. Процесс обучения завершается, если достигнуто заданное число эпох  $K$  или изменение функции потерь становится меньше некоторого порогового значения.

Представленное решение позволяет производить независимую оценку каждого из весов, что увеличивает скорость сходимости. Заметим, что предложенный алгоритм может быть неустойчив в случае вырожденной матрицы  $\mathbf{X}^{bj}$ , т. е. в случае мультиколлинеарности данных. Для предотвращения данной ситуации применимы различные способы. Например, возможно исключить коллинеарность признаков на этапе предобработки данных.

Для исследования метода ДРСР и его сравнительного анализа с классическим методом стохастического мини-пакетного градиентного спуска (СГС) и Adam используем набор данных, включающий в себя 100 образцов с четырьмя признаками ( $\mathbf{X} \in \mathbb{R}^{100 \times 4}$ ,  $\mathbf{y} \in \mathbb{R}^{100}$ ). Данные получены с помощью функции «make\_regression» библиотеки sklearn. Шум в данных имеет нормальное распределение со среднеквадратическим отклонением, равным 1. Перед обучением проведена нормализация данных. Скорость обучения для всех алгоритмов выбрана  $\alpha = 0,01$ . Начальные веса инициализированы случайными числами в диапазоне от 0 до 1. Результаты компьютерного моделирования методов ДРСР, СГС и Adam представлены на рисунке.

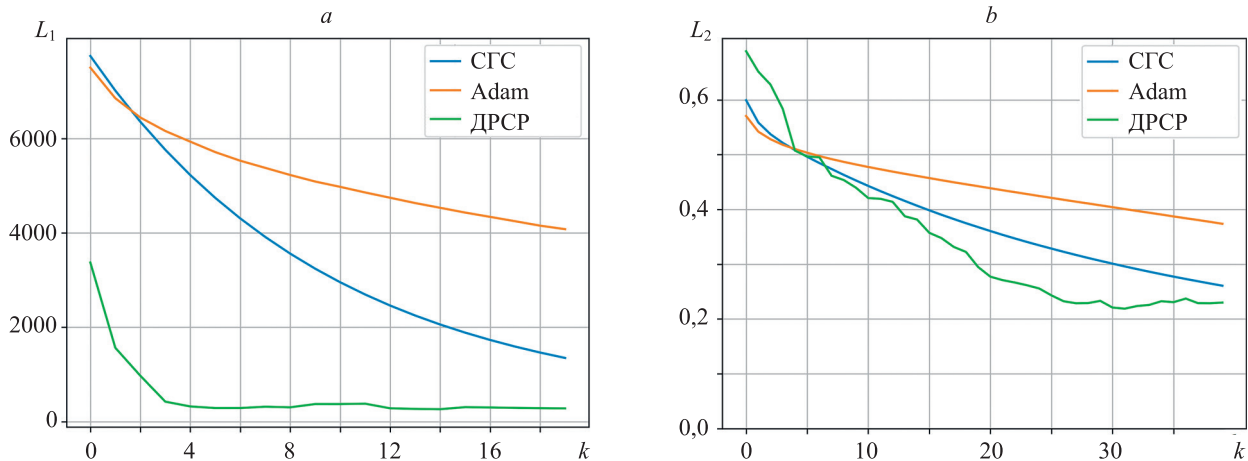


Рисунок. Целевой функционал в процессе обучения для задач регрессии (a) и классификации (b)  
Figure. Loss function in the learning process: for the regression task (a), for the classification task (b)

На рисунке, a приведены графики целевой функции  $L_1$  в процессе обучения для трех сравниваемых итеративных методов. Видно, что решение, основанное на методе ДРСР, имеет существенно более высокую скорость сходимости, особенно на первых итерациях.

Для решения задачи классификации метод ДРСР останется неизменным, за исключением формулы для обновления весов:

$$w_i = w_i - \alpha \Delta(Y_i^j - \sigma(\Delta/w_i)).$$

Для проведения сравнительного анализа был использован набор данных Iris. В нем содержится информация о трех различных видах цветов (всего 150 примеров данных, по 50 на каждый из цветов). Каждый пример имеет четыре признака: длина и ширина чашелистника, длина и ширина лепестка. Для задачи бинарной классификации все примеры разделены на два класса: являющиеся Iris setosa и прочие. На рисунке, b приведены графики целевой функции  $L_2$  в процессе

обучения для трех рассмотренных итеративных методов. Видно, что решение, основанное на методе ДРСР, имеет более высокую скорость сходимости целевого функционала.

Таким образом, в каждом случае предложенное на основе метода ДРСР решение показало более высокую скорость сходимости, особенно на первых итерациях. Также данный метод имеет всего один гиперпараметр, что значительно упрощает его настройку по сравнению с другими методами. Предложенный подход может быть развит на другие классы алгоритмов машинного обучения, в частности метод опорных векторов, полно-связные, сверточные и рекуррентные нейронные сети. В качестве перспективы возможно применение полученного результата для создания гибридных алгоритмов оптимизации: на первых итерациях применяется предложенное решение с высокой скоростью сходимости, на последующих — более устойчивые методы, такие как Adam.

## Литература

1. Rosenblatt F. The perceptron: A probabilistic model for information storage and organization in the brain // *Psychological Review*. 1958. V. 65. N 6. P. 386–408. <https://doi.org/10.1037/h0042519>
2. Хайкин С. Нейронные сети: Полный курс. 2-е изд. М.: «Вильямс», 2006. 1103 с.
3. Karthick K. Comprehensive Overview of Optimization Techniques in Machine Learning Training // *Control Systems and Optimization Letters*. 2024. V. 2. N 1. P. 23–27. <https://doi.org/10.59247/csol.v2i1.69>
4. Reyad M., Sarhan A.M., Arafat M. A modified Adam algorithm for deep neural network optimization // *Neural Computing and Applications*. 2023. V. 35. N 23. P. 17095–17112. <https://doi.org/10.1007/s00521-023-08568-z>
5. Wang Y., Xiao Z., Cao G. A convolutional neural network method based on Adam optimizer with power-exponential learning rate for bearing fault diagnosis // *Journal of Vibroengineering*. 2022. V. 24. N 4. P. 666–678. <https://doi.org/10.21595/jve.2022.22271>
6. Liu M., Yao D., Liu Z., Guo J., Chen J. An improved Adam optimization algorithm combining adaptive coefficients and composite gradients based on randomized block coordinate descent // *Computational Intelligence and Neuroscience*. 2023. V. 10. N 1. P. 4765891. <https://doi.org/10.1155/2023/4765891>
7. Демиденко Е.З. Линейная и нелинейная регрессия. М.: Финансы и статистика, 1981. 302 с.

## References

1. Rosenblatt F. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 1958. vol. 65, no. 6, pp. 386–408. <https://doi.org/10.1037/h0042519>
2. Haykin Simon S. *Neural Networks: A Comprehensive Foundation*. Macmillan, 1994, 696 p.
3. Karthick K. Comprehensive overview of optimization techniques in machine learning training. *Control Systems and Optimization Letters*, 2024, vol. 2, no. 1, pp. 23–27. <https://doi.org/10.59247/csol.v2i1.69>
4. Reyad M., Sarhan A.M., Arafat M. A modified Adam algorithm for deep neural network optimization. *Neural Computing and Applications*, 2023, vol. 35, no. 23, pp. 17095–17112. <https://doi.org/10.1007/s00521-023-08568-z>
5. Wang Y., Xiao Z., Cao G. A convolutional neural network method based on Adam optimizer with power-exponential learning rate for bearing fault diagnosis. *Journal of Vibroengineering*, 2022, vol. 24, no. 4, pp. 666–678. <https://doi.org/10.21595/jve.2022.22271>
6. Liu M., Yao D., Liu Z., Guo J., Chen J. An improved Adam optimization algorithm combining adaptive coefficients and composite gradients based on randomized block coordinate descent. *Computational Intelligence and Neuroscience*, 2023, vol. 10, no. 1, pp. 4765891. <https://doi.org/10.1155/2023/4765891>
7. Demidenko E.Z. *Linear and Nonlinear Regression*. Moscow, Finstat Publ., 1981, 302 p. (in Russian)

8. Aranovskiy S., Bobtsov A., Ortega R., Pyrkin A. Performance enhancement of parameter estimators via dynamic regressor extension and mixing // *IEEE Transactions on Automatic Control*. 2017. V. 62. N 7. <https://doi.org/10.1109/tac.2016.2614889>
9. Ljung L. *System Identification: Theory for the User*. Prentice-Hall, 1987. 519 p.

#### Авторы

**Маргун Алексей Анатольевич** — кандидат технических наук, доцент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация; научный сотрудник, Институт проблем машиноведения РАН, Санкт-Петербург, 199178, Российская Федерация, [sc 55521791600](https://orcid.org/0000-0002-5333-0594), <https://orcid.org/0000-0002-5333-0594>, [alexeimargun@gmail.ru](mailto:alexeimargun@gmail.ru)

**Зименко Константин Александрович** — кандидат технических наук, доцент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, [sc 55355650800](https://orcid.org/0000-0001-6220-7494), <https://orcid.org/0000-0001-6220-7494>, [kostyazimenko@gmail.com](mailto:kostyazimenko@gmail.com)

**Бобцов Алексей Алексеевич** — доктор технических наук, профессор, профессор, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, [sc 8046819200](https://orcid.org/0000-0003-1854-6717), <https://orcid.org/0000-0003-1854-6717>, [bobtsov@mail.ru](mailto:bobtsov@mail.ru)

8. Aranovskiy S., Bobtsov A., Ortega R., Pyrkin A. Performance enhancement of parameter estimators via dynamic regressor extension and mixing. *IEEE Transactions on Automatic Control*, 2017, vol. 62, no. 7. <https://doi.org/10.1109/tac.2016.2614889>
9. Ljung L. *System Identification: Theory for the User*. Prentice-Hall, 1987, 519 p.

#### Authors

**Alexey A. Margun** — PhD, Associate Professor, ITMO University, Saint Petersburg, 197101, Russian Federation; Scientific Researcher, Institute for Problems in Mechanical Engineering of the Russian Academy of Sciences, Saint Petersburg, 199178, Russian Federation [sc 55521791600](https://orcid.org/0000-0002-5333-0594), <https://orcid.org/0000-0002-5333-0594>, [alexeimargun@gmail.com](mailto:alexeimargun@gmail.com)

**Konstantin A. Zimenko** — PhD, Associate Professor, ITMO University, Saint Petersburg, 197101, Russian Federation, [sc 55355650800](https://orcid.org/0000-0001-6220-7494), <https://orcid.org/0000-0001-6220-7494>, [kostyazimenko@gmail.com](mailto:kostyazimenko@gmail.com)

**Alexey A. Bobtsov** — D.Sc., Full Professor, ITMO University, Saint Petersburg, 197101, Russian Federation, [sc 8046819200](https://orcid.org/0000-0003-1854-6717), <https://orcid.org/0000-0003-1854-6717>, [bobtsov@mail.ru](mailto:bobtsov@mail.ru)

*Статья поступила в редакцию 05.11.2024*  
*Одобрена после рецензирования 09.12.2024*  
*Принята к печати 22.01.2025*

*Received 05.11.2024*  
*Approved after reviewing 09.12.2024*  
*Accepted 22.01.2025*



Работа доступна по лицензии  
Creative Commons  
«Attribution-NonCommercial»