



# Методы адаптации параметров в стохастических алгоритмах для оптимизации информационно-вычислительных процессов

Антонов Кирилл Александрович

Научный руководитель: д.т.н., профессор А.А. Шалыто

Специальность:

2.3.8. Информатика и информационные процессы  
(технические науки)

Санкт-Петербург 2025

- Задачи **оптимизации информационно-вычислительных процессов** встречаются в логистике, цифровом управлении и инженерном проектировании<sup>1</sup>.
- Пример такой задачи – **маршрутизация доставки с ограничениями**, где требуется найти наименьший суммарный путь доставщика с ограничением на грузоподъемность за как можно меньшее время<sup>2</sup>.
- Несмотря на неразрешимость этих задач за полиномиальное время, **стохастические алгоритмы** позволяют получать приближённые решения таких задач оптимизации за приемлемое время<sup>3</sup>.
- Быстродействие стохастических алгоритмов существенно зависит от выбора **метода адаптации** их **параметров**<sup>4</sup>.
- **Однако** этот выбор остаётся нетривиальным для практических задач, что зачастую усугубляется отсутствием строгих оценок влияния адаптации параметров на сходимость стохастических алгоритмов<sup>5</sup>.
- **Актуальность работы** обусловлена потребностью в преодолении этих ограничений.

---

<sup>1</sup> *Ilin I., Maydanova S., Lepikhin A., Jahn C., Weigell J., Korablev V.* Digital platforms for the logistics sector of the Russian Federation / Technological Transformation: A New Role For Human, Machines And Management: TT-2020. Springer, 2021. С. 179-188.

<sup>2</sup> *Braekers K., Ramaekers K., Van Nieuwenhuysse I.* The vehicle routing problem: State of the art classification and review / Computers & industrial engineering. 2016. Т. 99. С. 300-313.

<sup>3</sup> *Скобцов Ю. А., Федоров Е. Е.* Метаэвристики. Донецк : Ноулидж, 2013.

<sup>4</sup> *Karafotias G., Hoogendoorn M., Eiben Á. E.* Parameter control in evolutionary algorithms: Trends and challenges / IEEE Transactions on Evolutionary Computation. 2014. Т. 19, № 2. С. 167-187.

<sup>5</sup> *Doerr C., Doerr B.* Theory of parameter control in evolutionary algorithms / ACM Transactions on Evolutionary Learning and Optimization. 2018. Т. 3, № 1. С. 1-37.

# Термины и сокращения

1. **Задача оптимизации** (максимизации): пусть  $\mathcal{X} \subseteq \mathbb{R}^n$ ,  $f : \mathcal{X} \rightarrow \mathbb{R}$ . Найти решение  $x^* \in \mathcal{X}$  такое, что  $f(x^*) = \max_{x \in \mathcal{X}} f(x)$  (если максимум существует).
2. **Стохастический алгоритм** – алгоритм, использующий случайность в шаге поиска для решения задачи оптимизации. Обычно работает в несколько итераций, на каждой из которых генерируются решения. Оператор, который их генерирует, будем называть  $\mathcal{A}$ .
3. **Адаптация параметров** в стохастическом алгоритме<sup>6</sup> – автоматическое обновление параметров алгоритма в процессе работы по наблюдаемым данным поиска. Функцию, задающую адаптацию параметров, будем называть  $\mathcal{C}$ .
4. **Качество решения** – если максимизируется (минимизируется) функция  $f$ , то решение лучшего качества будет иметь большее (меньшее) значение функции  $f$ .
5. **Время оптимизации**  $\tau$  – случайная величина: минимальное число итераций (или вычислений  $f$ ) стохастического алгоритма до достижения требуемого качества решения задачи оптимизации функции  $f$ . За  $\mathbb{E}[\tau]$  обозначим математическое ожидание времени оптимизации.
6. **Псевдобулева оптимизация**: область определения – бинарные строки длины  $n$ :  $\mathcal{X} = \{0, 1\}^n$ .
7. **Комбинаторная оптимизация**: область определения – комбинаторные объекты (например перестановки длины  $n$ :  $\mathcal{X} = S_n$ ).
8. **Вещественнозначная оптимизация**: область определения более чем счетна (например  $\mathcal{X} = \mathbb{R}^n$ ).
9. **МИПО** – задача малоитерационной параллельной оптимизации доставки грузов.

<sup>6</sup>Eiben A. E., Michalewicz Z., Schoenauer M., Smith J. E. Parameter control in evolutionary algorithms / Parameter setting in evolutionary algorithms. Springer, 2007. С. 19-46.

# Зачем нужна адаптация параметров?

- Стохастические алгоритмы обычно имеют параметры  $\theta$ , например, насколько большие случайные модификации вносятся, сколько случайных решений поддерживается, как отбираются решения в процессе поиска и т. д. **Эти параметры определяют** быстродействие алгоритма и качество результата<sup>7</sup>.
- **Фиксированные значения параметров не универсальны:** лучшие значения параметров зависят от задачи, **стадии поиска** и доступного бюджета<sup>8,9</sup>.
- **Ручная настройка параметров дорогая:** требует множества прогонов и экспертного опыта. Однако в практических задачах часто есть **ограниченный вычислительный бюджет** и жёсткие сроки.
- **Адаптация параметров** в стохастическом алгоритме повышает устойчивость к смене задач и ускоряет получение качественных решений, делая стохастические алгоритмы эффективными на практике<sup>10</sup>.

---

<sup>7</sup> Семенкин Е. С., Семенкина М. Е. Самоконфигурируемые эволюционные алгоритмы моделирования и оптимизации. Магнитогорский дом печати, 2014.

<sup>8</sup> Eiben A. E., Michalewicz Z., Schoenauer M., Smith J. E. Parameter control in evolutionary algorithms / Parameter setting in evolutionary algorithms. Springer, 2007. С. 19-46.

<sup>9</sup> Doerr B., Doerr C. Theory of parameter control for discrete black-box optimization: Provable performance gains through dynamic parameter choices / Theory of Evolutionary Computation: Recent Developments in Discrete Optimization. 2020. С. 271-321.

<sup>10</sup> Скобцов Ю. А. От генетических алгоритмов к вычислительному интеллекту / Девятая международная конференция по когнитивной науке. Т. 1. Национальный исследовательский ядерный университет «МИФИ», 2021. С. 649-652.



## Основоположники темы исследования

- **Антамошкин А. Н. и Казаковцев Л. А.** являются одними из основоположников стохастических алгоритмов с адаптацией параметров<sup>11,12</sup>.
- Их метод изменяющихся вероятностей адаптирует множество параметров<sup>13</sup> и имеет приложение для решения задачи комбинаторной оптимизации<sup>14</sup>.
- **Rechenberg I.** предложил классический метод адаптации в эволюционных алгоритмах – правило одной пятой<sup>15</sup>.
- **Schwefel H. P.** развил метод для вещественнозначных функций<sup>16</sup>.
- Впоследствии **Семенкин Е. С.** и его соавторы существенно развили методы адаптации параметров в стохастических алгоритмах<sup>17</sup>.

---

<sup>11</sup> *Antamoshkin A., Saraev V.* On Definition of Informative Subsystem of Signs in the Pattern Recognition Problem / *Computers and Artificial Intelligence*. 1985. Т. 4, № 3. С. 245-252.

<sup>12</sup> *Antamoshkin A. N.* Brainware for Searchal Pseudoboolean Optimization / *Transactions of the Tenth Prague Conference: Information Theory, Statistical Decision Functions, Random Processes held at Prague*. 10A. Springer, 1987. С. 203-206.

<sup>13</sup> *Антамошкин А. Н., Казаковцев Л. А.* Метод изменяющихся вероятностей / *Проблемы случайного поиска*. Вып. 11. Рига : Зинатне, 1988. С. 26-34.

<sup>14</sup> *Антамошкин А. Н., Казаковцев Л. А.* Применение метода изменяющихся вероятностей для задач оптимального размещения на сети / *Вестник Сибирского государственного аэрокосмического университета им. академика М. Ф. Решетнева*. 2014. Вып. 57, № 5. С. 10-19.

<sup>15</sup> *Rechenberg I.* Evolutionsstrategie: Optimierung technischer systeme nach prinzipien der biologischen evolution, frommann–holzboog. 1973.

<sup>16</sup> *Schwefel H. P.* Evolution and Optimum Seeking. 1995.

<sup>17</sup> *Семенкин Е. С., Семенкина О. Э., Коробейников С. П.* Адаптивные поисковые методы оптимизации сложных систем. Сибирский институт бизнеса, управления и психологии, 1996.

# Степень разработанности темы исследования

В настоящее время это направление активно развивают:

- **Антипов Д. С.** и его соавторы<sup>18</sup>, **Doerr B., Doerr C.** и их соавторы<sup>19</sup>: получены теоретические результаты на **простых задачах**;
- **Семенкин Е. С.** и его соавторы<sup>20,21,22</sup>: предложены **новые методы адаптации мутации и скрещивания**;
- **Еремеев А. В.** и его соавторы<sup>23</sup> – расширяется спектр **практических приложений**. Изучены особенности частных случаев задачи МИПО и похожих задач, а также предложены методы их решения.

---

<sup>18</sup> *Antipov D., Buzdalov M., Doerr B.* Lazy parameter tuning and control: choosing all parameters randomly from a power-law distribution / *Proceedings of the Genetic and Evolutionary Computation Conference*. 2021. С. 1115-1123.

<sup>19</sup> *Doerr B., Doerr C.* Theory of parameter control for discrete black-box optimization: Provable performance gains through dynamic parameter choices / *Theory of Evolutionary Computation: Recent Developments in Discrete Optimization*. 2020. С. 271-321.

<sup>20</sup> *Семенкин Е. С., Семенкина М. Е.* Самоконфигурируемые эволюционные алгоритмы моделирования и оптимизации. Магнитогорский дом печати, 2014.

<sup>21</sup> *Stanovov V., Semenkin E.* Success Rate Based Scaling Factor Adaptation in Dual-Population / *Metaheuristics and Nature Inspired Computing (META)*. 2024. С. 142-157.

<sup>22</sup> *Шерстнев П. А., Семенкин Е. С.* SelfCSHAGA: самоконфигурируемый генетический алгоритм оптимизации с адаптацией на основе истории успеха / *Вестник Московского государственного технического университета им. Н. Э. Баумана. Серия «Приборостроение»*. 2025. Вып. 151, № 2. С. 122-139.

<sup>23</sup> *Eremeev A., Reeves C. R.* Non-parametric estimation of properties of combinatorial landscapes / *Workshops on Applications of Evolutionary Computation*. Springer. 2002. С. 31-40; *Eremeev A. V., Kovalenko Y. V.* A memetic algorithm with optimal recombination for the asymmetric travelling salesman problem / *Memetic Computing*. 2020. Т. 12, № 1. С. 23-36.

# Ограничения существующих методов

- **Разрыв между теорией и сложными задачами:** строгие результаты часто получены на упрощённых моделях, а быстроедействие методов адаптации на функциях с множеством локальных оптимумов, которые часто встречаются в прикладных задачах, изучена существенно слабее.
- **Масштабируемость при малом бюджете:** методам адаптации параметров нужны время и данные, чтобы эффективно корректировать значения параметров. Однако при ограниченном числе итераций в высокопараллельной среде эффект может быть слабым, а накладные расходы из-за неоптимальной адаптации становятся критичными.
- **Специфика задач информационно-вычислительных процессов:** ограничения, смешанная природа переменных и требование быстро получать решения высокого качества. В задаче МИПО алгоритм работает в высокопараллельной среде при малом числе итераций стохастического алгоритма. В таком режиме большинство известных методов адаптации не рассчитаны на быструю работу. Поэтому нужны специализированные методы. Компания VeeRoute<sup>24</sup> заинтересована в их разработке и внедрении.

Необходимо разработать новые методы адаптации параметров, которые **работают при ограниченном вычислительном бюджете и переносятся на прикладные задачи информационно-вычислительных процессов.**

<sup>24</sup> Российская компания, позиционирующая себя как платформа комбинаторной оптимизации <https://veeroute.ru/>

# Цель исследования и соответствие паспорту

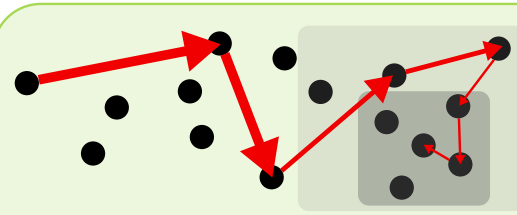
## Цель исследования

Целью исследования является повышение быстродействия **эволюционных алгоритмов** и **алгоритмов Байесовской оптимизации** при решении крупномасштабных задач **комбинаторной** (в том числе псевдобулевой) и **вещественнозначной оптимизации** в условиях **ограниченных вычислительных ресурсов** за счёт использования **адаптации параметров** в указанных алгоритмах.

## Соответствие паспорту специальности

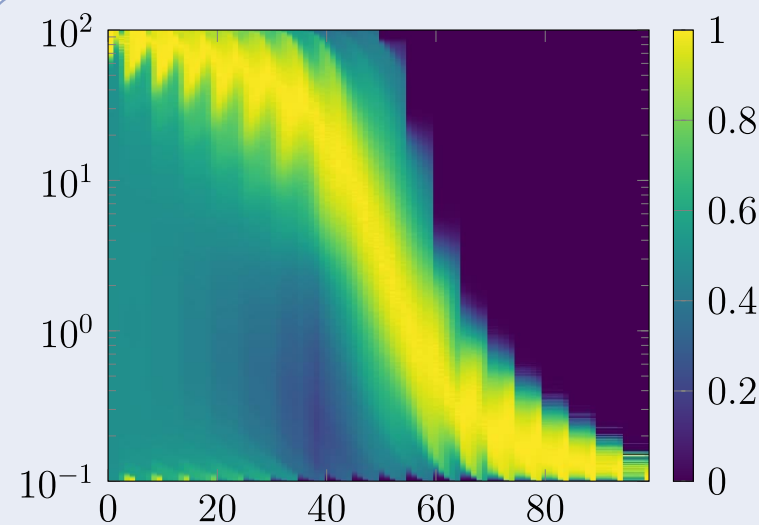
Диссертация соответствует **первому пункту** паспорта специальности 2.3.8: “**Разработка компьютерных методов и моделей описания, оценки и оптимизации информационных процессов** и ресурсов, а также средств анализа и выявления закономерностей на основе обмена информацией пользователями и возможностей используемого программно-аппаратного обеспечения”. Разработанные в диссертации методы направлены на **оптимизацию информационно-вычислительных процессов** по критерию, задаваемому целевой функцией, определённой на множестве возможных параметров этих процессов. Это достигается через применение стохастических алгоритмов с адаптацией параметров.





### Первый метод

Метод адаптации вероятности мутации в  $(1 + \lambda)$ -ЭА, позволяющий повысить быстродействие указанных алгоритмов при наличии  $\lambda$  параллельных вычислителей.

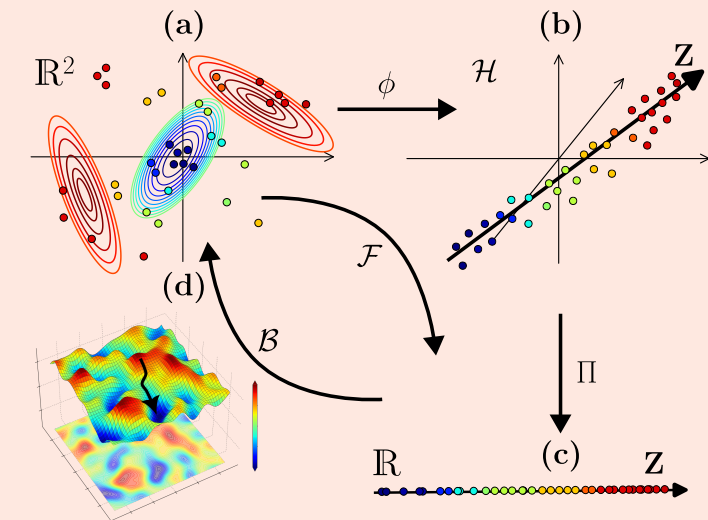


### Второй метод

Метод, позволяющий приблизительно вычислить **максимальное возможное ускорение** в  $(1 + \lambda)$ -ЭА при адаптации вероятности мутации и наличии  $\lambda$  параллельных вычислителей.

### Пример приложения

Адаптация параметров с использованием нейронной сети для решения задачи маршрутизации с ограничениями – МИПО.



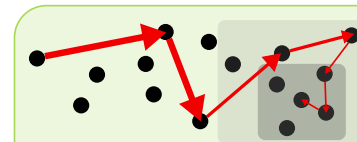
### Третий метод

Метод **Байесовской оптимизации**, позволяющий эффективно решать задачи вещественнозначной оптимизации мультимодальных, многомерных функций при ограниченных вычислительных ресурсах.

# Первое положение, выносимое на защиту (1/12)

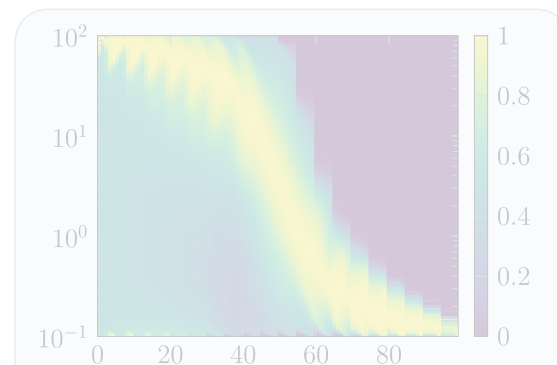
## Формулировка положения

Метод адаптации покомпонентной вероятности мутации в  $(1 + \lambda)$ -эволюционных алгоритмах, не требующий предварительной настройки параметров, отличающийся тем, что с целью повышения быстродействия указанных алгоритмов и обеспечения автоматической корректировки нижней границы вероятности мутации компонент потомков, используется механизм коллективного отбора на основе характеристик полученных потомков.



### Первый метод

Метод адаптации вероятности мутации в  $(1 + \lambda)$ -ЭА, позволяющий повысить быстродействие указанных алгоритмов при наличии  $\lambda$  параллельных вычислителей.

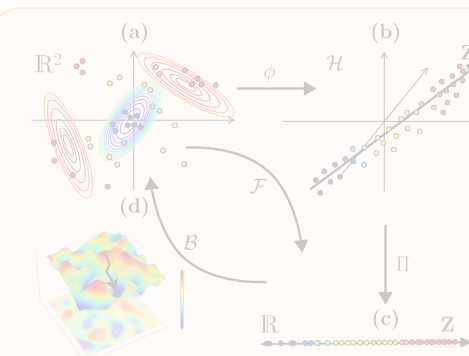


### Второй метод

Метод, позволяющий приближенно вычислить максимальное возможное ускорение в  $(1 + \lambda)$ -ЭА при адаптации вероятности мутации и наличии  $\lambda$  параллельных вычислителей.

### Пример приложения

Адаптация параметров с использованием нейронной сети для решения задачи маршрутизации с ограничениями – МИПО.



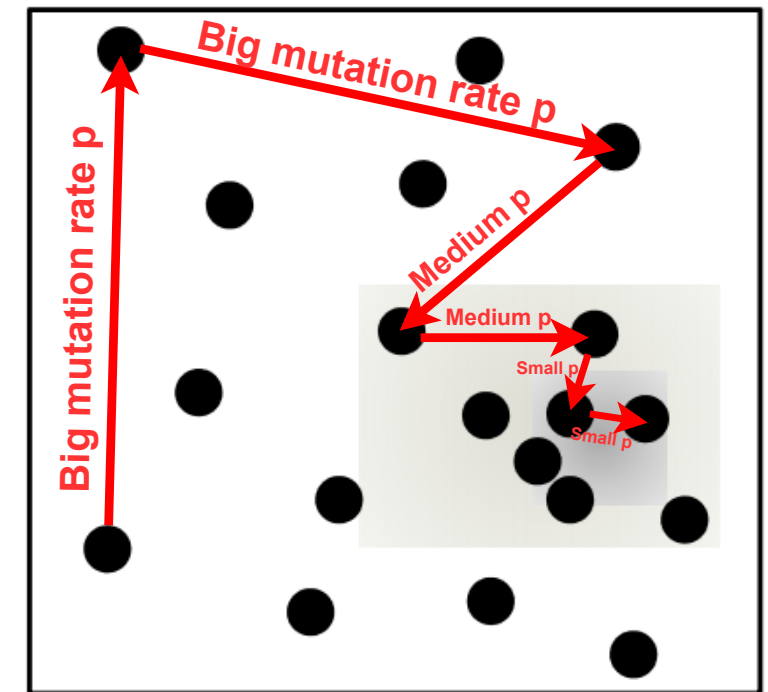
### Третий метод

Метод Байесовской оптимизации, позволяющий эффективно решать задачи вещественнозначной оптимизации мультимодальных, многомерных функций при ограниченных вычислительных ресурсах.

# Первое положение, выносимое на защиту (2/12)

## Цель предлагаемой адаптации параметров

- Рассматриваются  $(1 + \lambda)$ -эволюционные алгоритмы  $((1 + \lambda)$ -ЭА) для псевдобулевых задач  $f : \{0, 1\}^n \rightarrow \mathbb{R}$ .
- Их быстродействие существенно зависит от **вероятности мутации**  $p^{(t)}$  и, особенно, от **нижней границы**  $p_{\min}^{(t)}$ :
  - на ранних стадиях нужна большая мутация (глобальный поиск);
  - ближе к локальному оптимуму выгодна малая мутация (локальный поиск), иногда  $p^{(t)} < 1/n$ .
- **Проблема:** фиксированная нижняя граница  $p_{\min}$  требует ручного выбора (например,  $1/n$  против  $1/n^2$ ). Она по-разному влияет на быстродействие  $(1 + \lambda)$ -ЭА при различных  $\lambda$  и в разных фазах поиска.



Пример адаптации вероятности мутации

### Цель

Повысить быстродействие  $(1 + \lambda)$ -ЭА и обеспечить **адаптацию параметров**  $p^{(t)}, p_{\min}^{(t)}$ .

# Первое положение, выносимое на защиту (3/12)

## Математическая постановка и цель адаптации

- **Задача:** максимизировать  $f : \{0, 1\}^n \rightarrow \mathbb{R}$ , найти  $\mathbf{x}^* \in \arg \max f(\mathbf{x})$ .
- $(1 + \lambda)$ -ЭА хранит родителя  $\mathbf{x}^{(t)}$  и генерирует  $\lambda$  потомков  $\mathbf{y}^{(t,1)}, \dots, \mathbf{y}^{(t,\lambda)}$  мутацией родителя с параметром  $p^{(t)}$ .
- Новый родитель выбирается как

$$\arg \max \left\{ f(\mathbf{x}^{(t)}), f(\mathbf{y}^{(t,1)}), \dots, f(\mathbf{y}^{(t,\lambda)}) \right\}.$$

- **Время оптимизации:**

$$T = \min \left\{ t \geq 0 \mid \mathbf{x}^{(t)} = \mathbf{x}^* \right\}.$$

- **Цель адаптации параметров:**

$$\min \mathbb{E}[T] \quad \text{при} \quad p^{(t)} \in [p_{\min}^{(t)}, p_{\max}],$$

где адаптируются параметры  $p^{(t)}$  и  $p_{\min}^{(t)}$ .

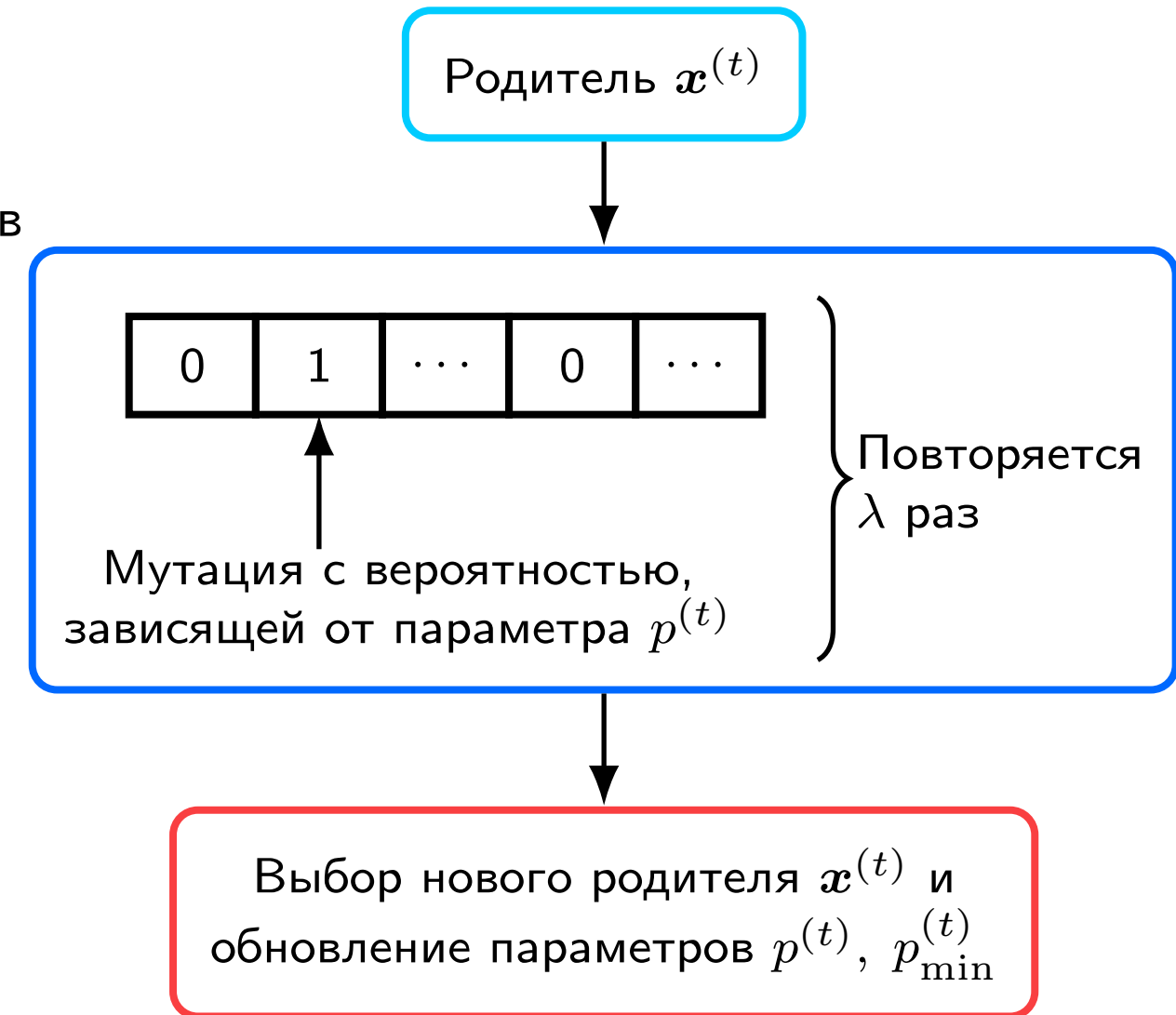


Схема одной итерации в  $(1 + \lambda)$ -ЭА, использующем адаптацию параметров  $p^{(t)}, p_{\min}^{(t)}$



# Первое положение, выносимое на защиту (4/12)

## Аналоги решения задачи адаптации параметров

- **Статическая вероятность мутации:** алгоритм  $(1 + \lambda)$ -ЭА использует фиксированную вероятность мутации  $p^{(t)}$ . Однако это требует ручного выбора  $p^{(t)}$  и не учитывает переход от глобального поиска к локальному<sup>25</sup>.
- **Адаптация правилом "одной пятой":** используется правило "одной пятой" для обновления параметра  $p^{(t)}$  – популярный представитель классических методов адаптации<sup>26</sup>.
- **Адаптация вероятности мутации методом 2-rate:** метод 2-rate использует  $p/2$  и  $2p$  для половин популяции на каждой итерации<sup>27</sup>. Изложен подробнее далее.
- **Адаптация вероятности мутации методом 3-rate и  $(A, b)$ :** расширения метода 2-rate, представленные в одной и той же работе<sup>28</sup>.
- **Адаптация вероятности мутации методом HQEA:** применение обучения с подкреплением для адаптации вероятности мутации<sup>29</sup>.

---

<sup>25</sup> [Doerr B., Doerr C., Ebel F. Tight bounds for mutation-based evolutionary algorithms / Proceedings of the 2015 Annual Conference on Genetic and Evolutionary Computation. ACM, 2015. C. 1439-1446.](#)

<sup>26</sup> [Rechenberg I. Evolutionsstrategie: Optimierung technischer systeme nach prinzipien der biologischen evolution, frommann-holzboog. 1973.](#)

<sup>27</sup> [Doerr B., Gießen C., Witt C., Yang J. The  \$\(1 + \lambda\)\$  evolutionary algorithm with self-adjusting mutation rate / Proceedings of the Genetic and Evolutionary Computation Conference. ACM, 2017. C. 1351-1358.](#)

<sup>28</sup> [Rodionova A., Antonov K., Buzdalova A., Doerr C. Offspring population size matters when comparing evolutionary algorithms with self-adjusting mutation rates / Proceedings of the Genetic and Evolutionary Computation Conference. 2019. C. 855-863.](#)

<sup>29</sup> [Buzdalova A., Doerr C., Rodionova A. Hybridizing the 1/5-th success rule with Q-learning for controlling the mutation rate of an evolutionary algorithm / Parallel Problem Solving from Nature – PPSN XVI. 2020. C. 485-499.](#)

### Метод 2-rate (прототип)<sup>30</sup>

На каждой итерации:

- Половина потомков генерируется с вероятностью  $p^{(t)}/2$ , половина с  $2p^{(t)}$ .
  - Параметр мутации  $p^{(t)}$  обновляется тем параметром, который был у половины с лучшим решением.
- Метод 2-rate  $(1/n)$  использует  $p_{\min} = 1/n$ .
  - Метод 2-rate  $(1/n^2)$  использует  $p_{\min} = 1/n^2$ .
  - **Недостаток:** значение  $p_{\min}$  сильно влияет на быстродействие алгоритма<sup>31</sup>.

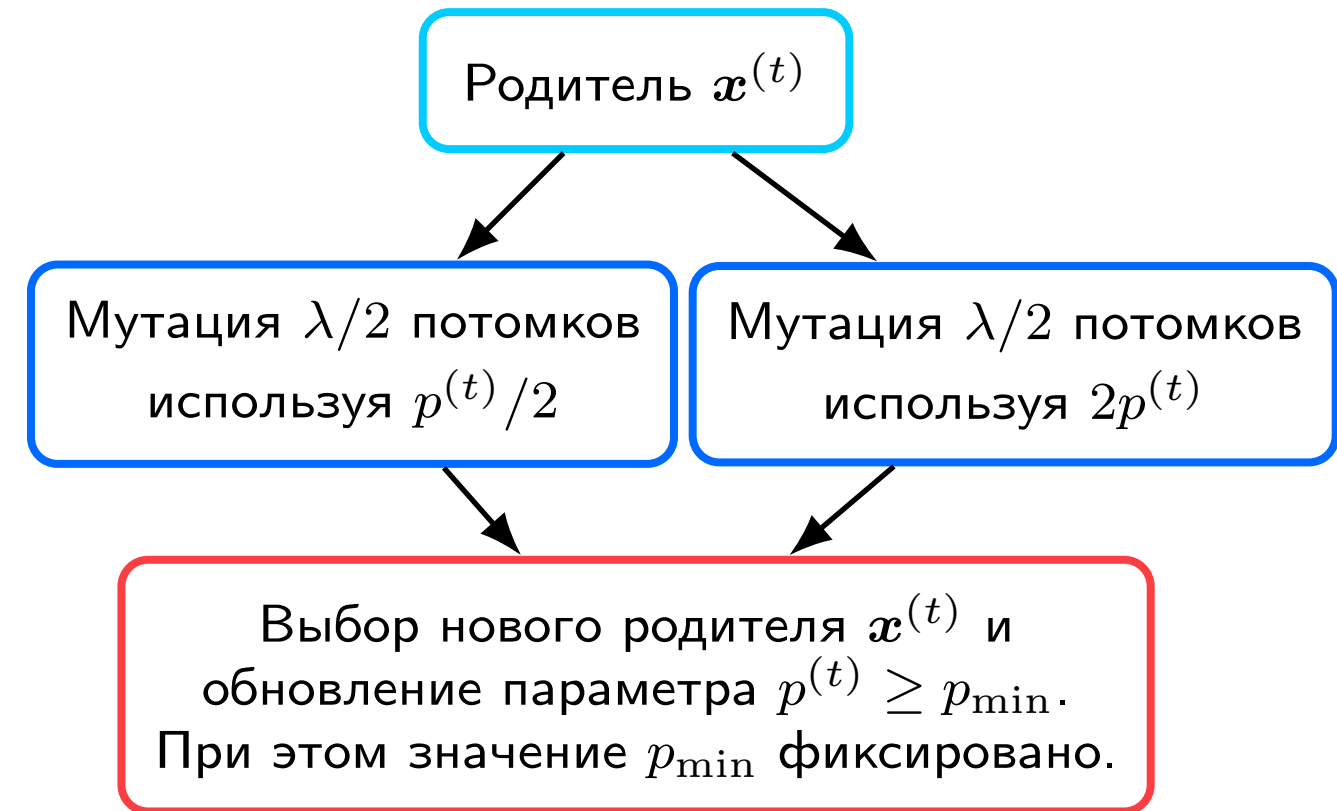


Схема одной итерации  $(1 + \lambda)$ -ЭА, использующего метод 2-rate для адаптации  $p^{(t)}$

<sup>30</sup> [Doerr B., Gießen C., Witt C., Yang J. The  \$\(1 + \lambda\)\$  Evolutionary Algorithm with Self-Adjusting Mutation Rate / \*Algorithmica\*. 2019. T. 81, № 2. С. 593-631.](#)

<sup>31</sup> [Rodionova A., Antonov K., Buzdalova A., Doerr C. Offspring population size matters when comparing evolutionary algorithms with self-adjusting mutation rates / \*Proceedings of the Genetic and Evolutionary Computation Conference\*. 2019. С. 855-863.](#)

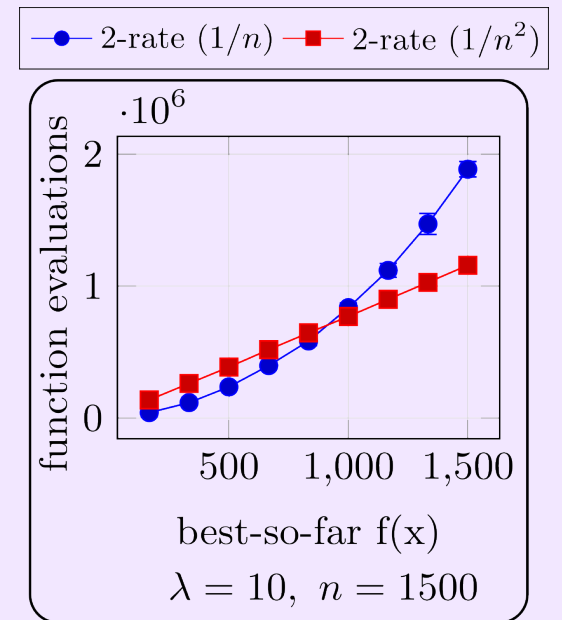
## Недостаток прототипа 2-rate

Даже при удачной адаптации  $p^{(t)}$  остаётся **ручной выбор** нижней границы  $p_{\min}$ , которая:

- должна быть достаточно малой для локального поиска;
- не должна пускать алгоритм в локальный поиск слишком рано.

Например:

- Доказано<sup>32</sup>, что 2-rate  $(1/n)$  асимптотически лучший на ONEMAX.
- Однако 2-rate  $(1/n^2)$  оказывается лучше на LEADINGONES.



Сравнение 2-rate  $(1/n)$  и 2-rate  $(1/n^2)$  на задаче LEADINGONES

## Принцип улучшения

Понижать  $p_{\min}^{(t)}$  от  $1/n$  к  $1/n^2$  по **коллективной информации** о потомках текущих поколений.

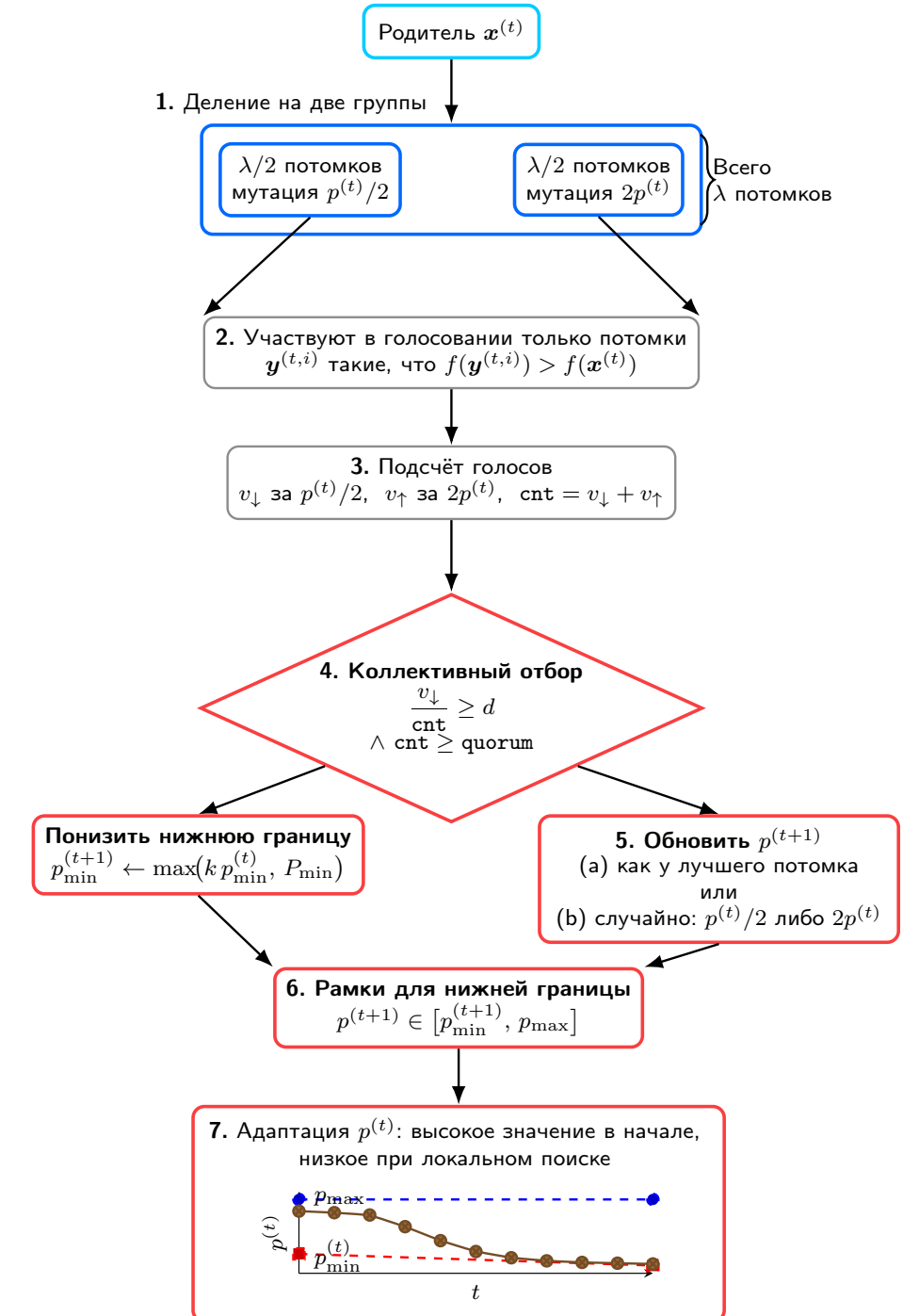
<sup>32</sup> Doerr B., Gießen C., Witt C., Yang J. The  $(1 + \lambda)$  evolutionary algorithm with self-adjusting mutation rate / Proceedings of the Genetic and Evolutionary Computation Conference. ACM, 2017. C. 1351-1358.

# Первое положение, выносимое на защиту (7/12)

## Описание предложенного метода

Предлагаемый метод 2-rate (voting):

1. Потомки делятся на две группы: одна мутируется с вероятностью  $p^{(t)}/2$ , другая с вероятностью  $2p^{(t)}$ .
2. Участвуют в голосовании только те потомки, которые превосходят родителя по значению функции.
3. Для каждой группы подсчитывается число «голосов» за соответствующую вероятность.
4. Если большинство голосов за меньшую вероятность и выполнен порог кворума, то нижняя граница вероятности мутации понижается.
5. Значение  $p^{(t)}$  обновляется по одному из двух сценариев: либо как у лучшего потомка, либо случайно ( $p^{(t)}/2$  или  $2p^{(t)}$ ).
6. Проводится контроль, чтобы  $p^{(t)}$  оставалась в пределах между фиксированными нижней и верхней границами.
7. Это позволяет адаптировать вероятность мутации: высокая в начале, низкая при локальном поиске.





# Первое положение, выносимое на защиту (8/12)

## Определения тестовых функций

1. Использованы классические функции ONEMAX, LEADINGONES<sup>33</sup>.
2. Из W-MODEL<sup>34</sup> выбраны функции PLATEAU и RUGGEDNESS поскольку они содержат свойства задачи МИПО<sup>35</sup>.

- OneMax

$$\text{ONEMAX}(\mathbf{x}) = \sum_{j=1}^n x_j.$$

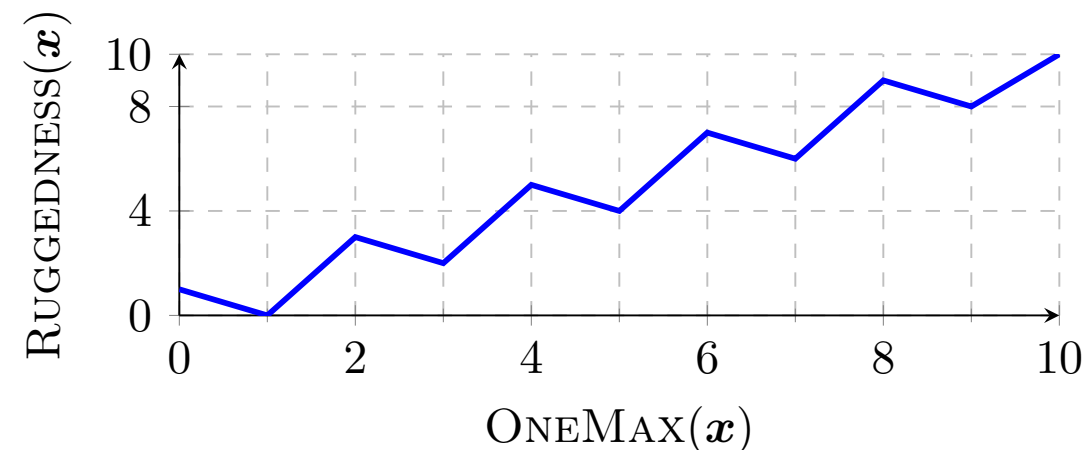
- LeadingOnes

$$\text{LEADINGONES}(\mathbf{x}) = \sum_{i=1}^n \prod_{j=1}^i x_j.$$

- Plateau

$$I \subseteq [n], \quad |I| = \lfloor 0.9n \rfloor, \quad \text{PLATEAU}(\mathbf{x}) = \sum_{j \in I} x_j.$$

- Ruggedness



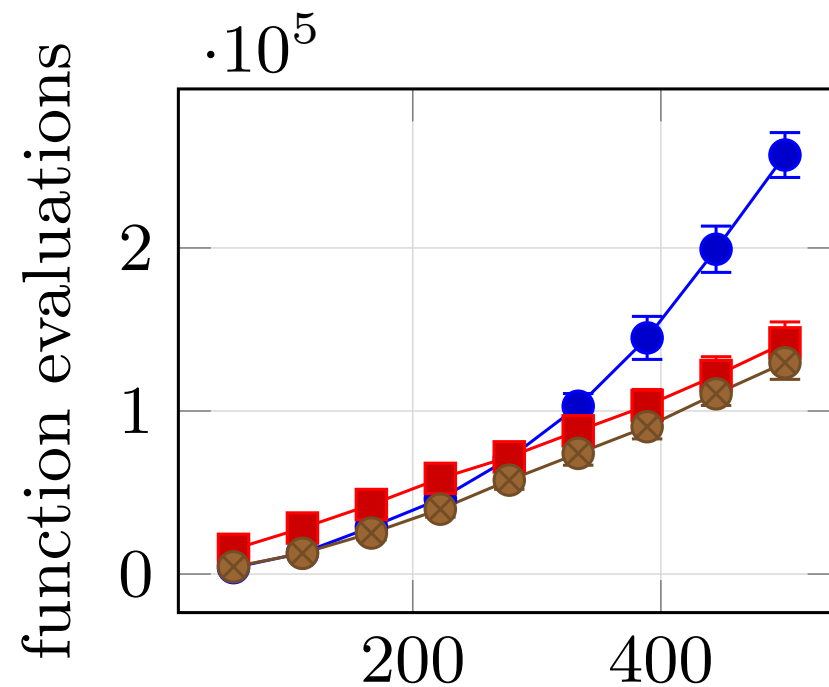
<sup>33</sup> Doerr C., Ye F., Horesh N., Wang H., Shir O. M., Bäck T. Benchmarking discrete optimization heuristics with IOHprofiler / Proceedings of the Genetic and Evolutionary Computation Conference Companion. 2019. C. 1798-1806.

<sup>34</sup> Weise T., Wu Z. Difficult features of combinatorial optimization problems and the tunable w-model benchmark problem for simulating them / Proceedings of the Genetic and Evolutionary Computation Conference Companion. 2018. C. 1769-1776.

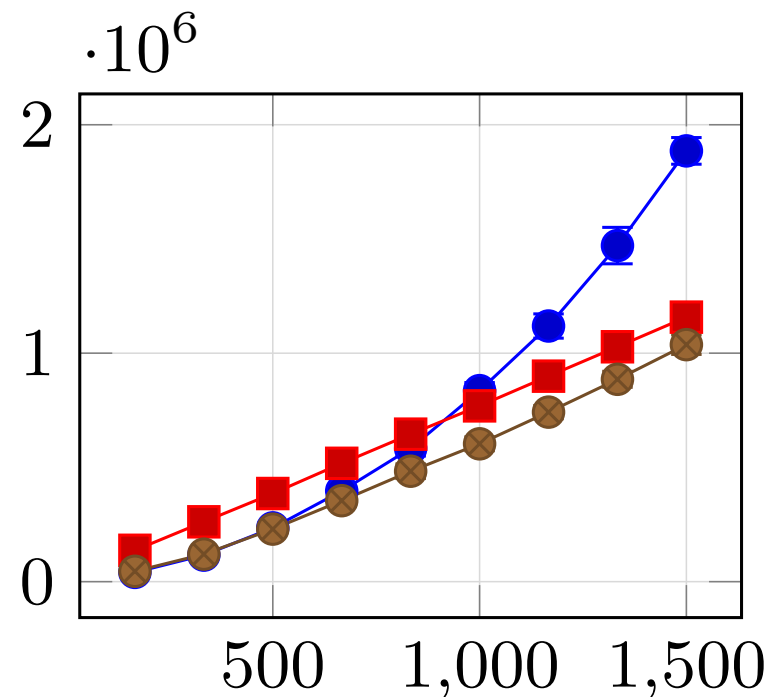
<sup>35</sup> Tayarani-N M.-H., Prügel-Bennett A. An analysis of the fitness landscape of travelling salesman problem / Evolutionary computation. 2016. T. 24, № 2. C. 347-384.

# Первое положение, выносимое на защиту (9/12) Эксперименты на LeadingOnes

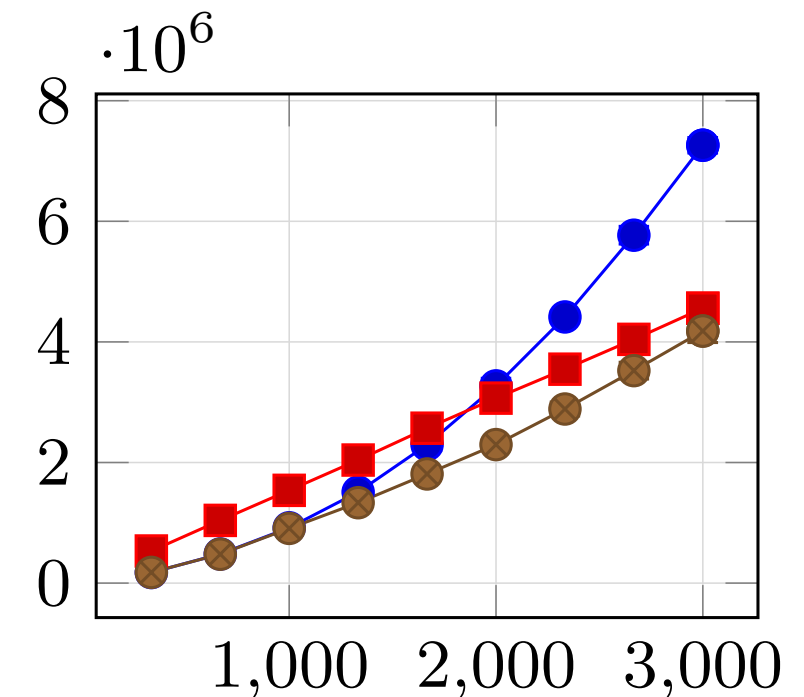
—●— 2-rate ( $1/n$ ) —■— 2-rate ( $1/n^2$ ) —●— 2-rate (voting)



(a)  $\lambda = 5, n = 500$



(b)  $\lambda = 10, n = 1500$

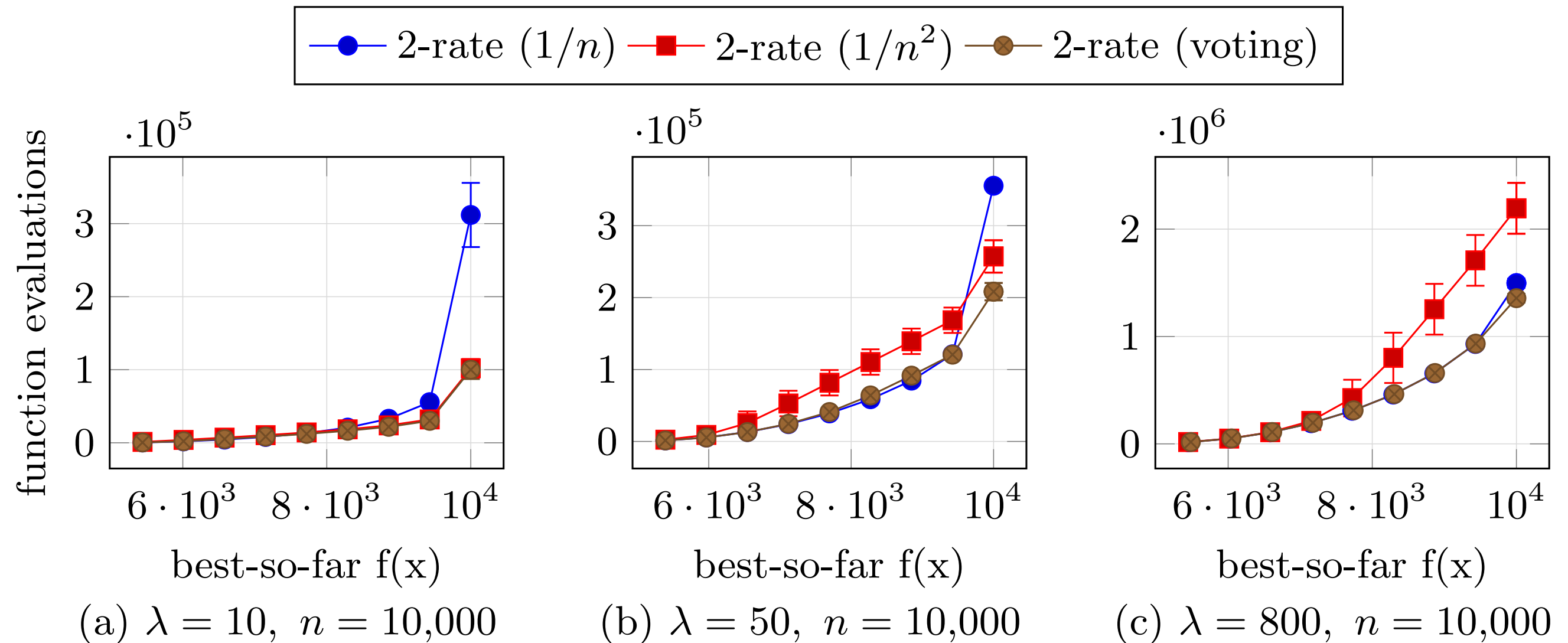


(c)  $\lambda = 16, n = 2000$

Показано, сколько вычислений целевой функции (ось  $Y$ ) требуется для достижения заданного качества решения (ось  $X$ )

Предложенный метод **2-rate (voting)** быстрее обоих прототипов достигает любого уровня качества решения.

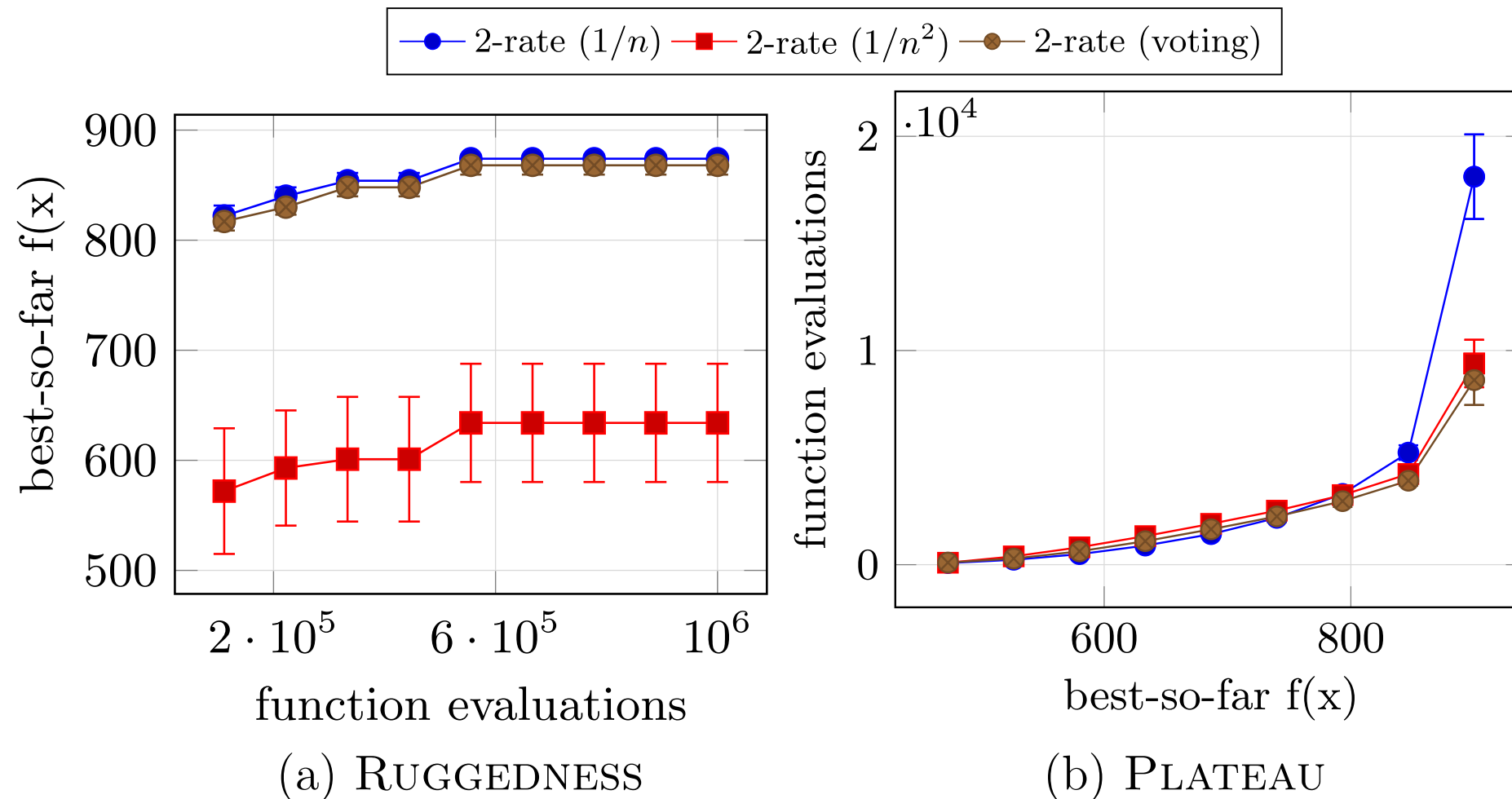
# Первое положение, выносимое на защиту (10/12) Эксперименты на OneMax



Показано, сколько вычислений целевой функции (ось  $Y$ ) требуется для достижения заданного качества решения (ось  $X$ )

Предложенный метод **2-rate (voting)** не ухудшает работу 2-rate ( $1/n$ ) на ONEMAX.

# Первое положение, выносимое на защиту (11/12) Эксперименты на W-Model



Предложенный метод незначительно ухудшает 2-rate ( $1/n$ ) на RUGGEDNESS и незначительно улучшает 2-rate ( $1/n^2$ ) на PLATEAU для  $\lambda = 10, n = 1000$ . Однако для больших  $\lambda$  метод не превосходит 2-rate ( $1/n$ ) на RUGGEDNESS и 2-rate ( $1/n^2$ ) на PLATEAU.



**Разработан метод** адаптации покомпонентной вероятности мутации в  $(1 + \lambda)$ -эволюционных алгоритмах, не требующий предварительной настройки параметров, **позволяющий** повысить быстродействие указанных алгоритмов и обеспечить автоматическую корректировку нижней границы вероятности мутации.

### Открытая реализация.

Репозиторий на GitHub <https://github.com/AntKirill/self-adaptive-mutation-ea>

### Результаты опубликованы в:

1. *Rodionova A., Antonov K., Buzdalova A., Doerr C.* Offspring population size matters when comparing evolutionary algorithms with self-adjusting mutation rates / *Proceedings of the Genetic and Evolutionary Computation Conference*. 2019. С. 855-863.<sup>A</sup>
2. *Antonov K., Buzdalova A., Doerr C.* Mutation Rate Control in the  $(1 + \lambda)$  Evolutionary Algorithm with a Self-adjusting Lower Bound / *International Conference on Mathematical Optimization Theory and Operations Research*. Springer, 2020. С. 305-319.

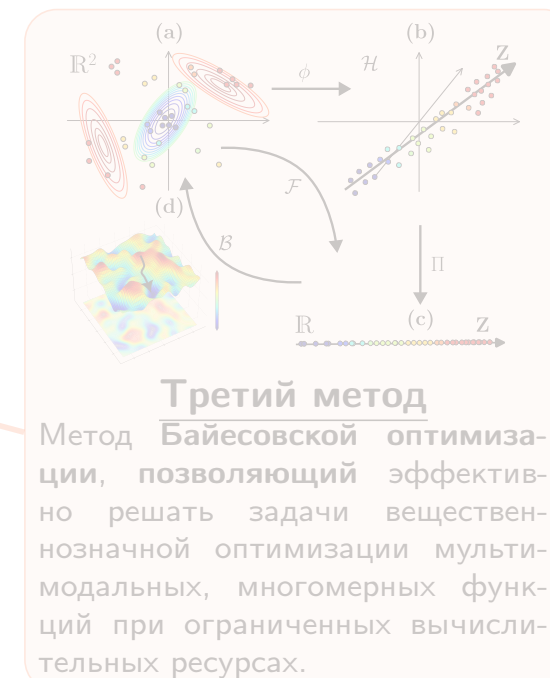
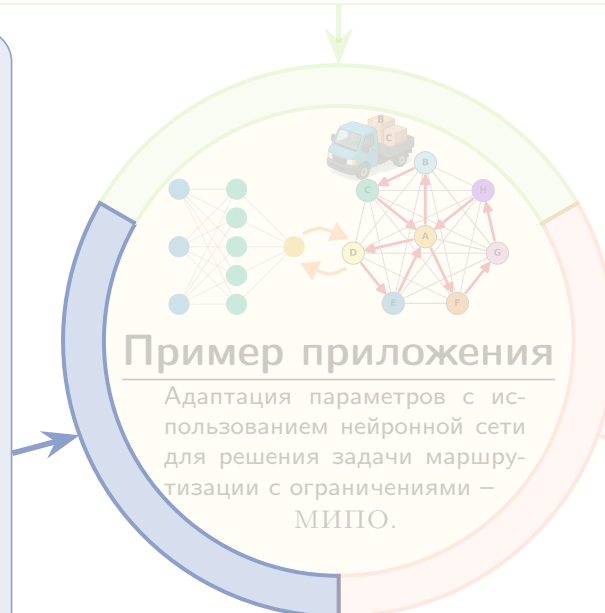
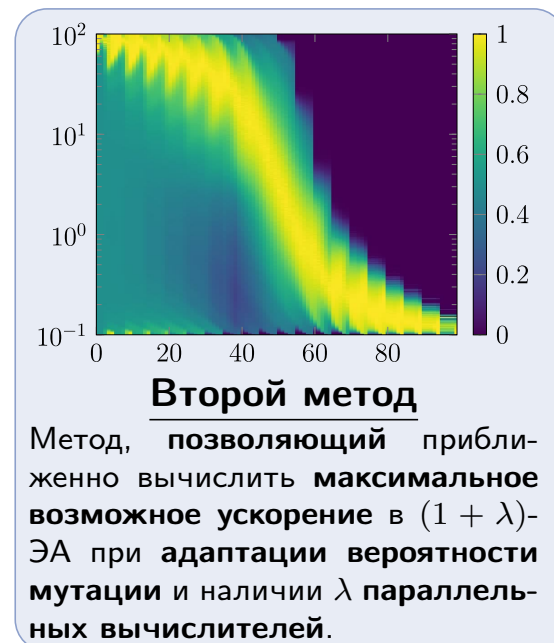
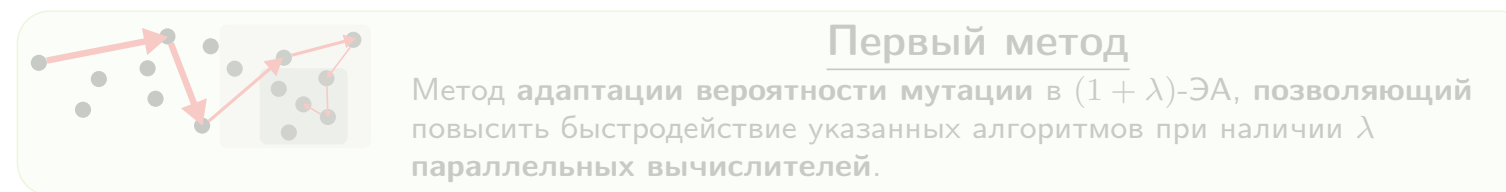
### Личный вклад.

1. В публикации [1] автор разработал и экспериментально исследовал метод 3-rate и влияние нижней границы вероятности мутации (40–50% вклада).
2. В публикации [2] автор разработал предложенный метод с адаптацией нижней границы вероятности мутации, реализовал программные эксперименты, провёл тестирование и выполнил анализ полученных результатов (около 90–95% вклада).

# Второе положение, выносимое на защиту (1/10)

## Формулировка положения

Метод оценки быстродействия  $(1 + \lambda)$ -эволюционных алгоритмов, применимых к псевдобулевым функциям и использующих адаптацию покомпонентной вероятности мутации, отличающийся тем, что с целью получения численных приближений нижних оценок времени работы указанных алгоритмов до достижения оптимума на практически значимых функциях, для которых отсутствуют аналитические нижние оценки, используется сочетание динамического программирования и Монте-Карло-моделирования.



# Второе положение, выносимое на защиту (2/10)

## Математическая постановка и целевой критерий

- **Оптимизация:**  $f : \{0, 1\}^n \rightarrow \mathbb{R}$ , найти  $\mathbf{x}^* \in \arg \max f(\mathbf{x})$ .

- **Время оптимизации:**

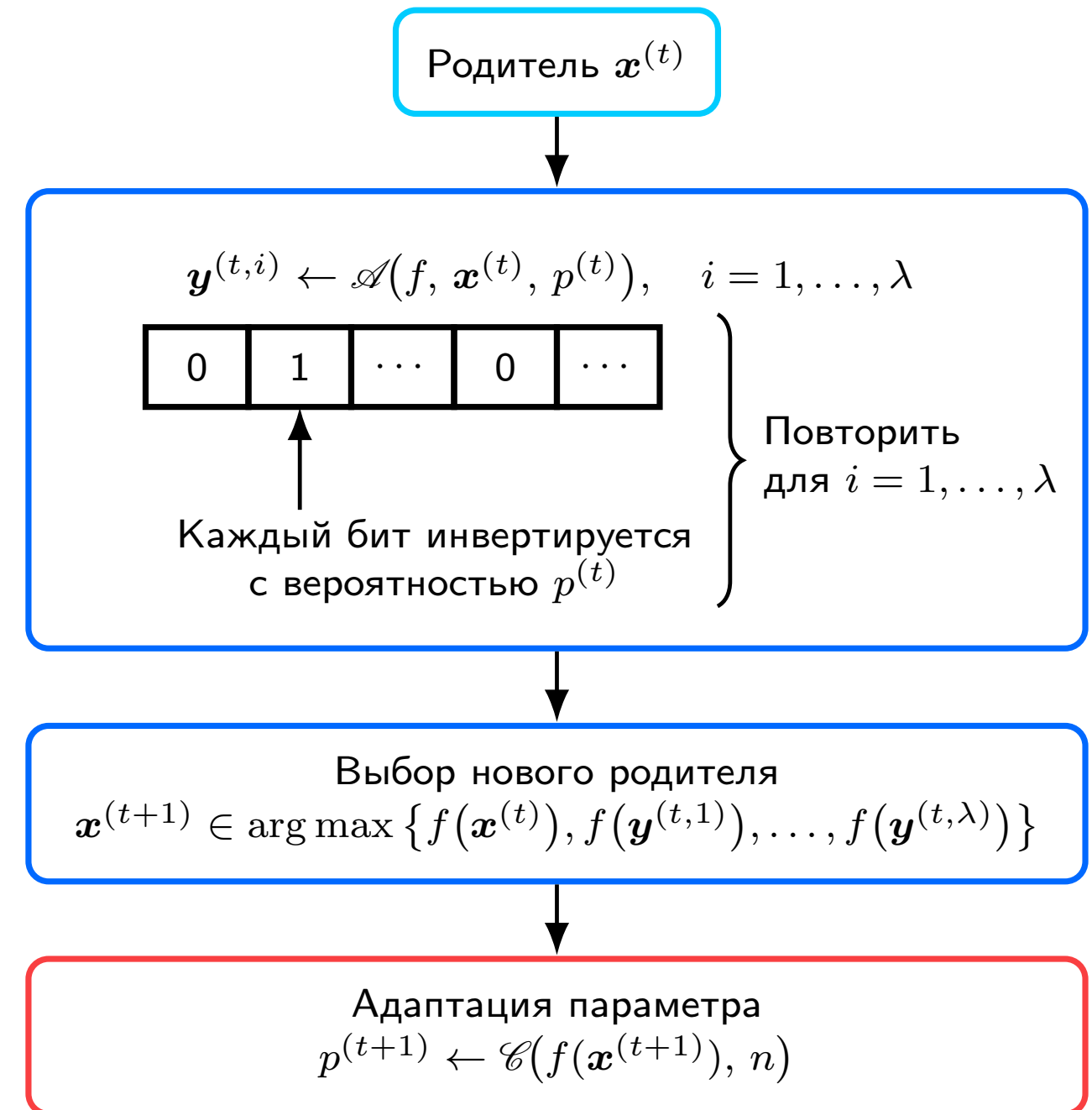
$$T = \min \left\{ t \geq 0 \mid \mathbf{x}^{(t)} = \mathbf{x}^* \right\}.$$

- **Цель:** минимизировать время оптимизации на множестве функций  $\mathcal{C} = \{\mathcal{C}\}$ , задающих адаптацию вероятности мутации:

$$\underline{T}(f, \mathcal{A}) = \min_{\mathcal{C} \in \mathcal{C}} \mathbb{E}[T(f, \mathcal{A}, \mathcal{C})].$$

- **Отношение эквивалентности** на пространстве подмножеств решений  $\{0, 1\}^n$ . Подмножества  $I, J \subset \{0, 1\}^n$  эквивалентны, если найдётся перестановка  $\pi : J \rightarrow J$  такая, что  $\forall \mathbf{x}^{(1)}, \mathbf{x}^{(2)} \in I, \forall \mathbf{y} \in J$ :

$$P[\mathcal{A}(f, \mathbf{x}^{(1)}, p) = \mathbf{y}] = P[\mathcal{A}(f, \mathbf{x}^{(2)}, p) = \pi(\mathbf{y})].$$



# Второе положение, выносимое на защиту (3/10)

## Аналоги: как оценивают время работы ЭА

- **Аналитические оценки.** Методы обеспечивают получение асимптотических оценок величины  $\underline{T}(f, \mathcal{A})$  при помощи математического анализа. Анализ использует свойства конкретных классов задач и редко обобщается между разными классами<sup>36,37</sup>.
- **Применение динамического программирования.** Методы численно приближают оценки величины  $\underline{T}(f, \mathcal{A})$ , используя динамическое программирование. Предполагается, что все вероятности переходов известны<sup>38,39</sup>:  $P[\mathbf{y}^{(t)} = \mathcal{A}(f, \mathbf{x}^{(t)}, p^{(t)})], \forall \mathbf{x}^{(t)}, \mathbf{y}^{(t)} \in \{0, 1\}^n$ .
- **Симуляция работы алгоритма.** Запускаются алгоритм с разными функциями  $\mathcal{C}$ , задающими адаптацию<sup>40,41</sup>. Применимы для всех функций, но получают недостаточно точные приближения  $\underline{T}(f, \mathcal{A})$ .

---

<sup>36</sup> [Böttcher S., Doerr B., Neumann F. Optimal fixed and adaptive mutation rates for the LeadingOnes problem / International Conference on Parallel Problem Solving from Nature. Springer, 2010. C. 1-10.](#)

<sup>37</sup> [Doerr B., Doerr C., Ebel F. Tight bounds for mutation-based evolutionary algorithms / Proceedings of the 2015 Annual Conference on Genetic and Evolutionary Computation. ACM, 2015. C. 1439-1446.](#)

<sup>38</sup> [Buskulic N., Doerr C. Maximizing drift is not optimal for solving OneMax / Proceedings of the Genetic and Evolutionary Computation Conference Companion. 2019. C. 425-426.](#)

<sup>39</sup> [Buzdalov M., Doerr C. Optimal Mutation Rates for the EA on OneMax / International Conference on Parallel Problem Solving from Nature. Springer, 2020. C. 574-587.](#)

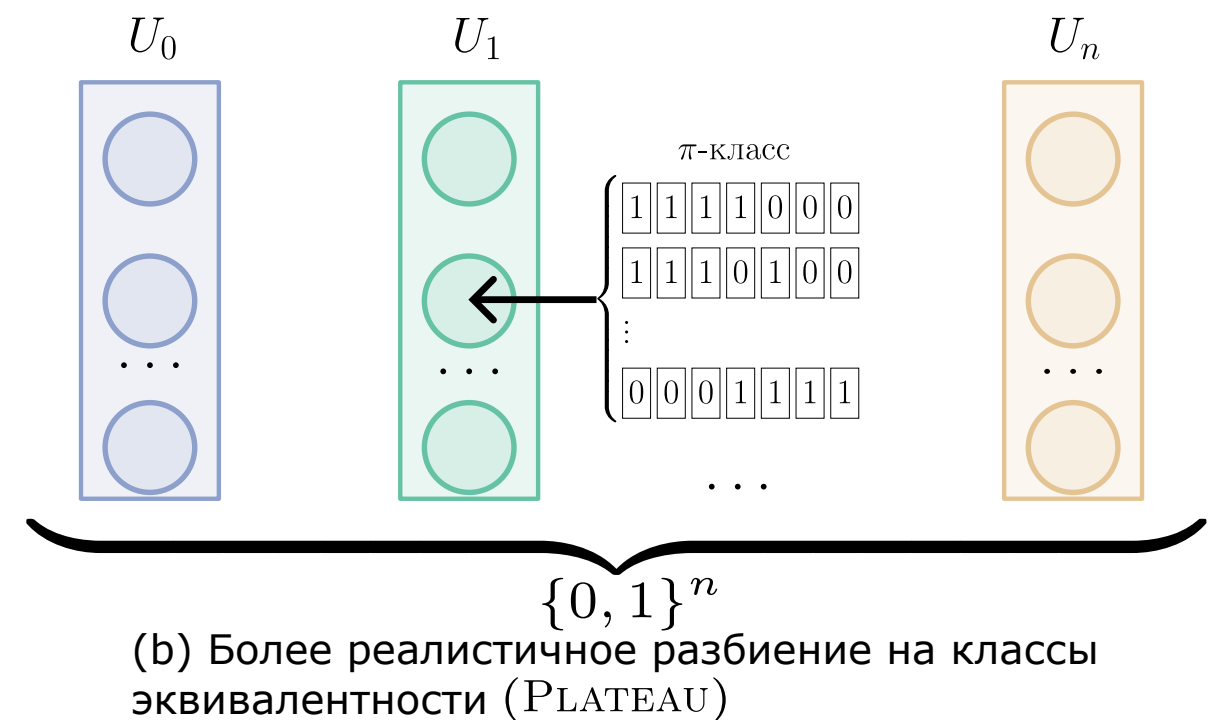
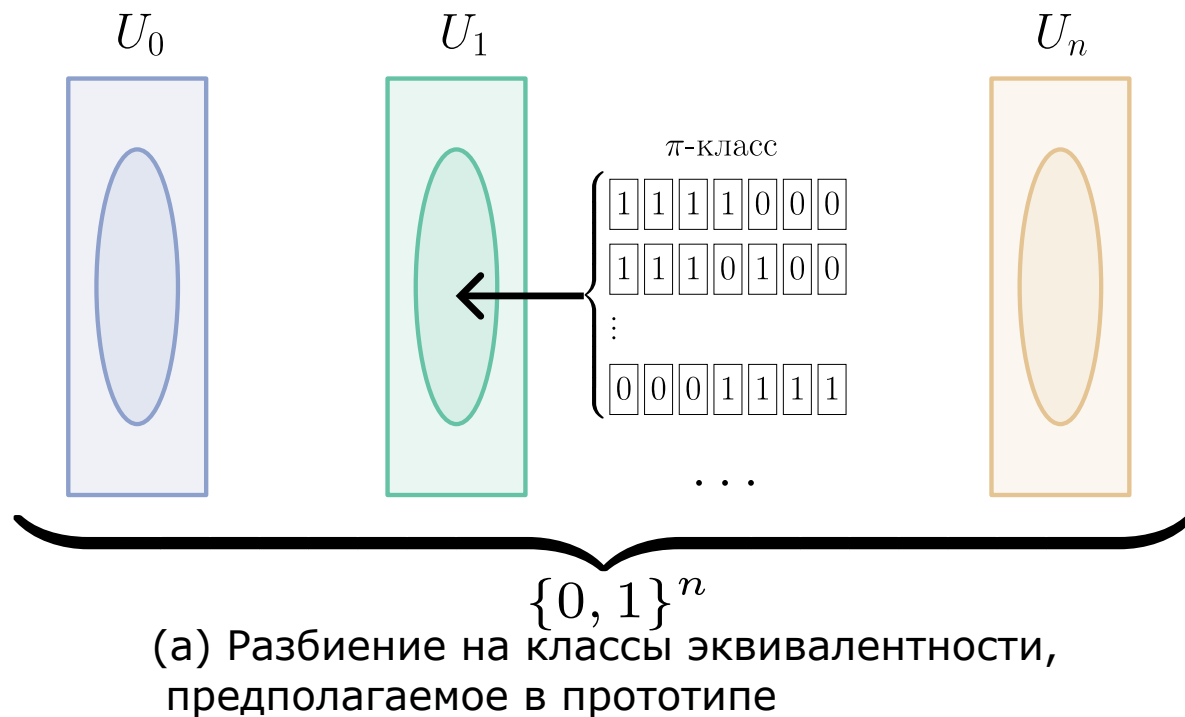
<sup>40</sup> [Hansen N., Auger A., Finck S., Ros R. Real-parameter black-box optimization benchmarking 2010: Experimental setup / Proceedings of the 12th annual conference companion on Genetic and evolutionary computation. 2010. C. 1447-1452.](#)

<sup>41</sup> [Doerr C., Ye F., Horesh N., Wang H., Shir O. M., Bäck T. Benchmarking discrete optimization heuristics with IOHprofiler / Proceedings of the Genetic and Evolutionary Computation Conference Companion. 2019. C. 1798-1806.](#)

# Второе положение, выносимое на защиту (4/10)

## Прототип и его ограничения

- **Прототип:** применение динамического программирования для приближённого вычисления  $\underline{T}(f, \mathcal{A})$ .
- **Недостатки:**
  1. Предположение о том, что вероятности переходов должны быть известны, ограничивает применение метода к задачам, где аналитический вывод этих вероятностей невозможен.
  2. Метод разработан для задачи ONEMAX, где все решения со значением целевой функции  $U_i$  принадлежат одному классу эквивалентности. Это не так, например, на задаче PLATEAU.



Разбиение пространства решений на классы эквивалентности

# Второе положение, выносимое на защиту (5/10)

## Что улучшаем в прототипе и как это достигается

### Численные значения вероятностей переходов между классами эквивалентности

Для приближения вероятности переходов из класса  $i \subset \{0, 1\}^n$  в класс  $j \subset \{0, 1\}^n$ , обозначаемой как  $\hat{p}_{i,j}$ , используются Монте-Карло-симуляции:

$$\mathfrak{P} := \{ \mathcal{A}(f, \mathbf{x}, p; \xi^{(k)}) \mid \mathbf{x} \in i, k = 1, \dots, N \}, \quad \hat{p}_{i,j} := \frac{1}{N} \sum_{y \in \mathfrak{P}} \mathbf{1}\{y \in J\}.$$

### Используется динамическое программирование, учитывающее несколько классов эквивалентности на одном уровне

Для класса  $i$  на уровне  $f$  при фиксированной вероятности мутации  $p$ :

$$T_{i,p} = (\hat{p}_{i,i})T_{i,p} + \sum_{j \in U_f} (\hat{p}_{i,j})T_{j,p} + \sum_{j \in U_{q>f}} (\hat{p}_{i,j})T_j^*,$$

где  $U_f$  – множество классов со значением целевой функции  $f$ ,  $U_{q>f}$  – классы с более высоких уровней (для них  $T_j^*$  уже рассчитан).



## Описание метода

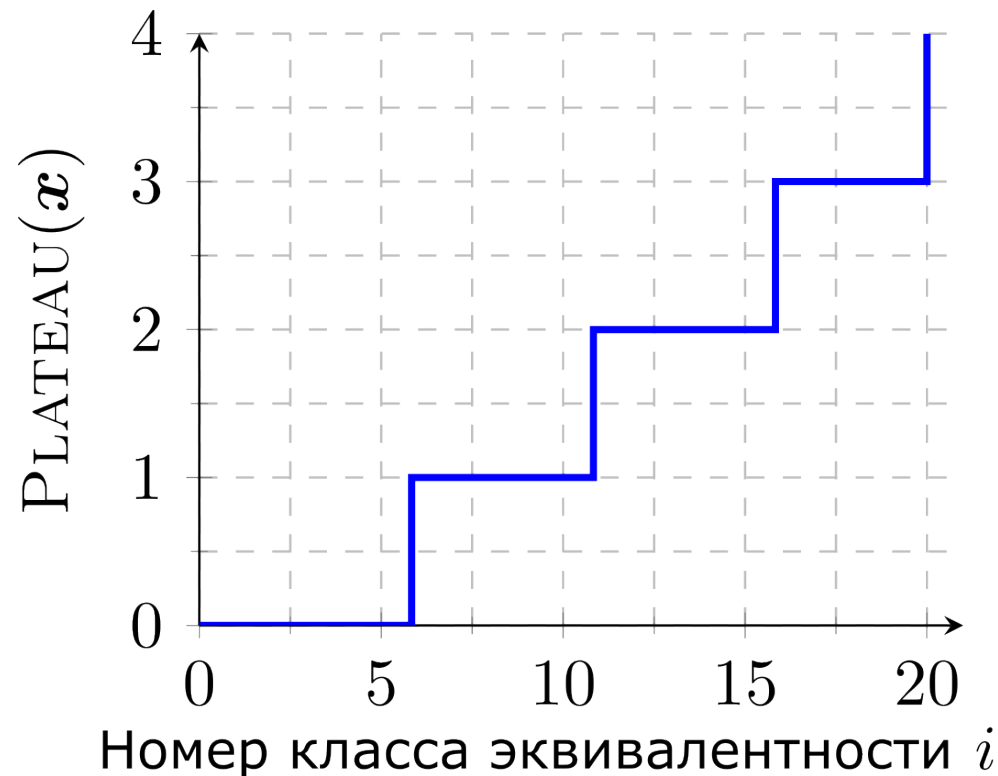
Метод численного приближения нижней границы времени работы для  $(1 + \lambda)$ -ЭА с использованием Монте-Карло-моделирования и динамического программирования:

1. Для каждого уровня целевой функции перебираются возможные значения вероятности мутации  $\hat{\mathcal{C}}(f, n) \in \{p_1^{(f)}, \dots, p_m^{(f)}\}$ .
2. Переходные вероятности  $\hat{p}_{i,j}$  между классами решений аппроксимируются методом Монте-Карло.
3. Используется динамическое программирование, учитывающее несколько классов эквивалентности на одном уровне. Составляется и решается система линейных уравнений для вычисления ожидаемого времени до достижения оптимума  $T_{i,p}$ :

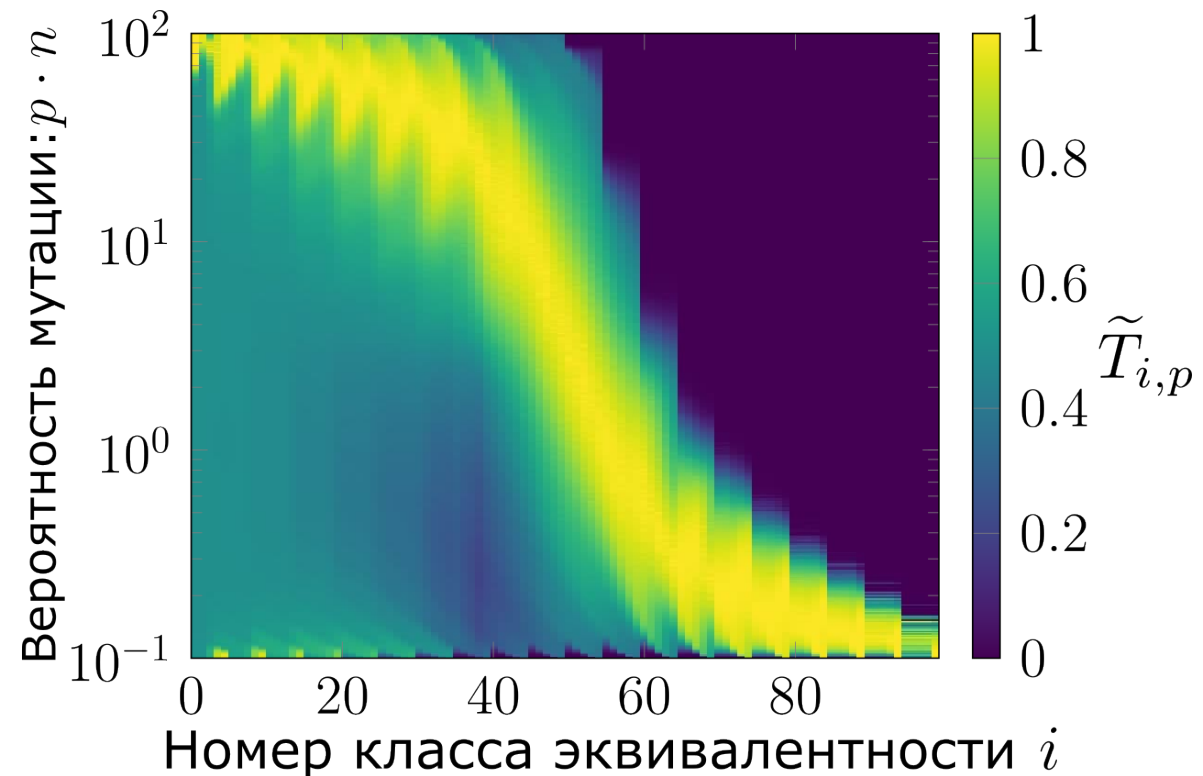
$$\left\{ T_{i,p} = (\hat{p}_{i,i})T_{i,p} + \sum_{j \in U_f} (\hat{p}_{i,j})T_{j,p} + \sum_{j \in U_{q>f}} (\hat{p}_{i,j})T_j^* \right\}_{i \in U_f}$$

4. Для каждого класса выбирается вероятность мутации, минимизирующая ожидаемое время до достижения оптимума:  $T_i^* \leftarrow \min_p(T_{i,p})$ .
5. Метод возвращает множества  $\{T_{j,p}\}, \{T_i^*\}$ .

# Второе положение, выносимое на защиту (7/10) Эксперимент на PLATEAU



(а) Ландшафт PLATEAU



(b) Тепловая карта для PLATEAU

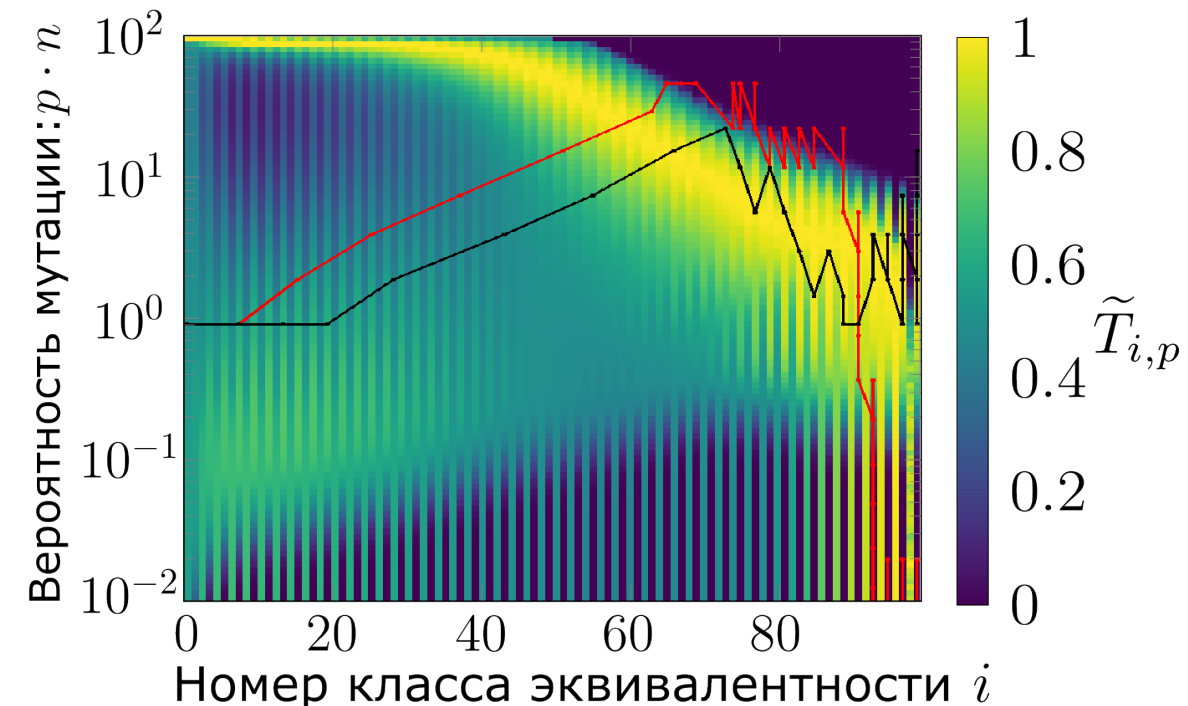
- Визуализировано множество<sup>42</sup>  $\{T_{j,p}\}$ , полученное предложенным методом на задаче PLATEAU.
- Желтым цветом показаны лучшие значения вероятности мутации, для каждого класса эквивалентности.
- По этим тепловым картам стало возможно анализировать методы адаптации вероятности мутации в  $(1 + \lambda)$ -ЭА на задаче PLATEAU.

<sup>42</sup> *Buzdalov M., Doerr C. Optimal Mutation Rates for the EA on OneMax / International Conference on Parallel Problem Solving from Nature. Springer, 2020. С. 574-587.*

# Второе положение, выносимое на защиту (8/10) Эксперимент на Ruggedness



(a) Ландшафт RUGGEDNESS



(b) Тепловая карта для RUGGEDNESS

- Визуализировано множество  $\{T_{j,p}\}$ , полученное предложенным методом на задаче RUGGEDNESS.
- Анализ существующих методов адаптации: красный –  $(A, b)^{43}$ , черный – 2-rate  $(1/n)^{44}$ .

<sup>43</sup> [Rodionova A., Antonov K., Buzdalova A., Doerr C.](#) Offspring population size matters when comparing evolutionary algorithms with self-adjusting mutation rates / [Proceedings of the Genetic and Evolutionary Computation Conference](#). 2019. С. 855-863.

<sup>44</sup> [Doerr B., Gießen C., Witt C., Yang J.](#) The  $(1 + \lambda)$  Evolutionary Algorithm with Self-Adjusting Mutation Rate / [Algorithmica](#). 2019. Т. 81, № 2. С. 593-631.

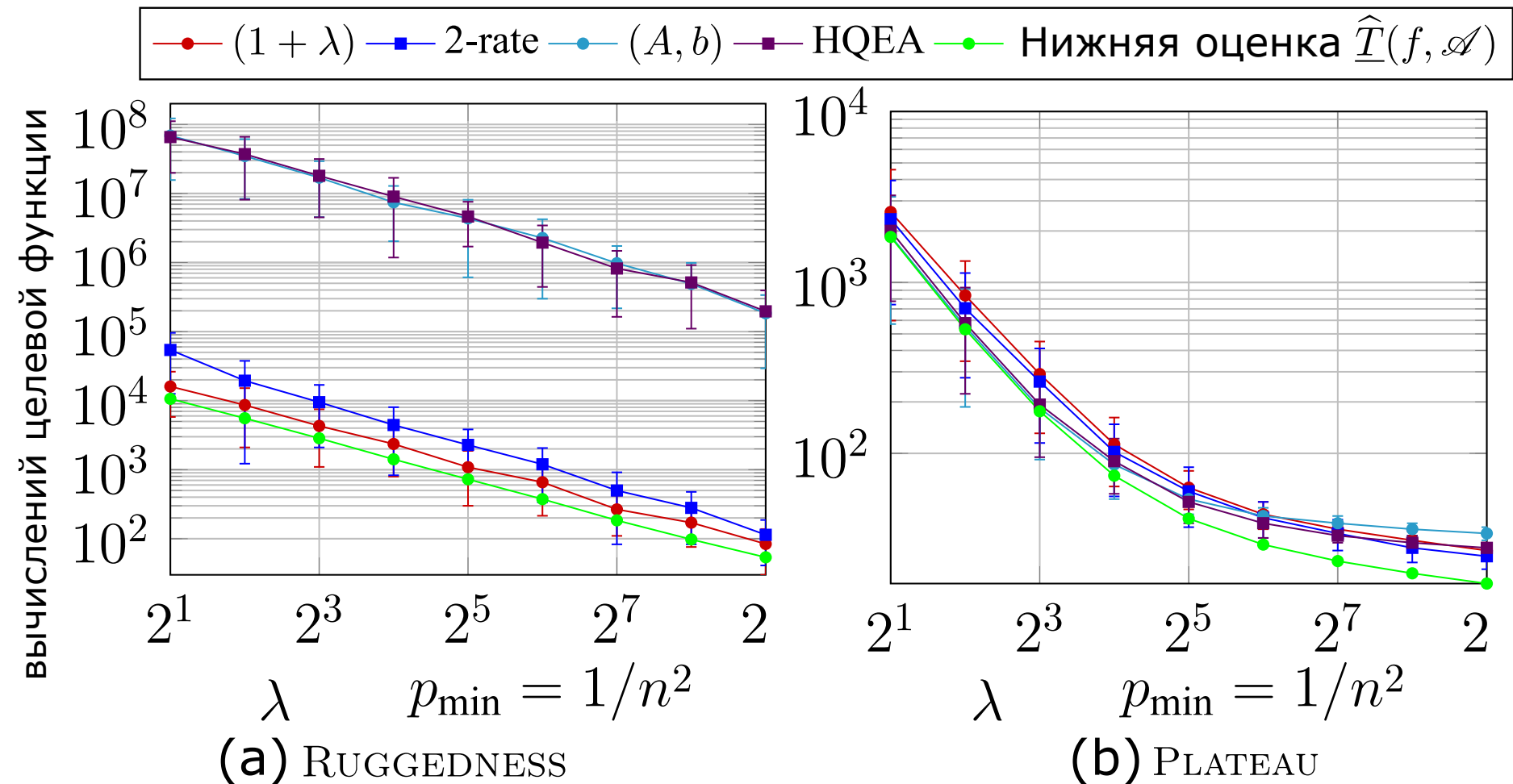
# Второе положение, выносимое на защиту (9/10)

## Численные нижние оценки

Проанализированы аналоги из первого положения, выносимого на защиту.

- $(1 + \lambda)$ -ЭА со статической вероятностью мутации  $p$ ;
- 2-rate – метод адаптации  $p$ ;
- $(A, b)$  – метод адаптации  $p$ ;
- HQEA – метод адаптации  $p$ .

$$\hat{T}(f, \mathcal{A}) = \sum_{i \in U} \frac{|i|}{2^n} T_i^*.$$



Сравнение нижних оценок и времён работы алгоритмов на RUGGEDNESS и PLATEAU при разных  $\lambda$

### Результаты анализа

$(1 + \lambda)$ -ЭА со статическими параметрами работает близко к оптимальному на функции RUGGEDNESS. При больших  $\lambda$  проанализированные методы адаптации оказываются недостаточно быстрыми на функции PLATEAU.

## Выводы, публикации и личный вклад

**Разработан метод** оценки быстродействия  $(1 + \lambda)$ -эволюционных алгоритмов с адаптацией покомпонентной вероятности мутации на основе сочетания **Монте-Карло-моделирования** и **динамического программирования**, который **позволил** получать численные значения приближений нижних оценок времени работы указанных алгоритмов до достижения оптимума на **практически значимых функциях**.

**Открытая реализация.**

Репозиторий на GitHub [https://github.com/AntKirill/ea/tree/cluster\\_experiments/](https://github.com/AntKirill/ea/tree/cluster_experiments/)

**Результаты опубликованы в:**

3. **Antonov K.**, Buzdalov M., Buzdalova A., Doerr C. Blending Dynamic Programming with Monte Carlo Simulation for Bounding the Running Time of Evolutionary Algorithms / IEEE Congress on Evolutionary Computation (CEC). 2021. С. 878-885. 🏆

**Личный вклад.**

В публикации [3] автор провёл исследование, разработал методику смешивания динамического программирования с моделированием Монте-Карло, реализовал программные эксперименты и выполнил анализ результатов (около 80–85% вклада). Соавторы предложили идею применения динамического программирования и формулу перехода (около 15–20% вклада).

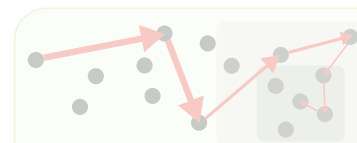


Приз за лучшую статью (Scopus конференция уровня B)

# Третье положение, выносимое на защиту (1/14)

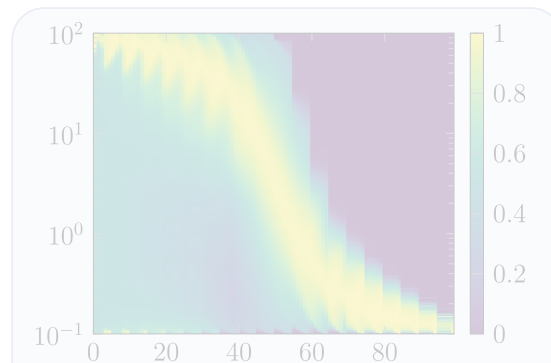
## Третье положение: формулировка

Метод многомерной Байесовской оптимизации с понижением размерности вещественнозначных функций, характеризующихся наличием множества локальных оптимумов, отличающийся тем, что с целью получения решений со значением целевой функции, близким к оптимальному при ограниченных вычислительных ресурсах, осуществляется построение и управление **нелинейным векторным представлением** точек многомерного пространства на основе **ядерного преобразования** с учетом **весов**, определяемых по наблюдаемым значениям целевой функции.



### Первый метод

Метод адаптации вероятности мутации в  $(1 + \lambda)$ -ЭА, позволяющий повысить быстродействие указанных алгоритмов при наличии  $\lambda$  параллельных вычислителей.

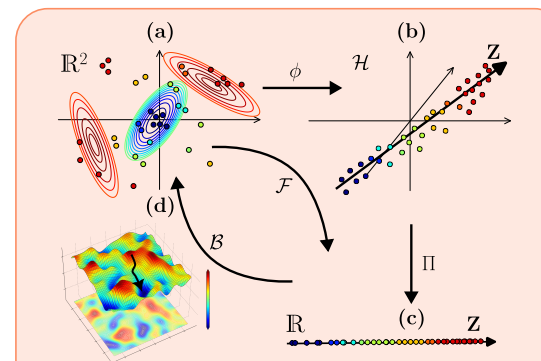


### Второй метод

Метод, позволяющий приближенно вычислить максимальное возможное ускорение в  $(1 + \lambda)$ -ЭА при адаптации вероятности мутации и наличии  $\lambda$  параллельных вычислителей.

### Пример приложения

Адаптация параметров с использованием нейронной сети для решения задачи маршрутизации с ограничениями – МИПО.



### Третий метод

Метод Байесовской оптимизации, позволяющий эффективно решать задачи вещественнозначной оптимизации мультимодальных, многомерных функций при ограниченных вычислительных ресурсах.



# Третье положение, выносимое на защиту (2/14)

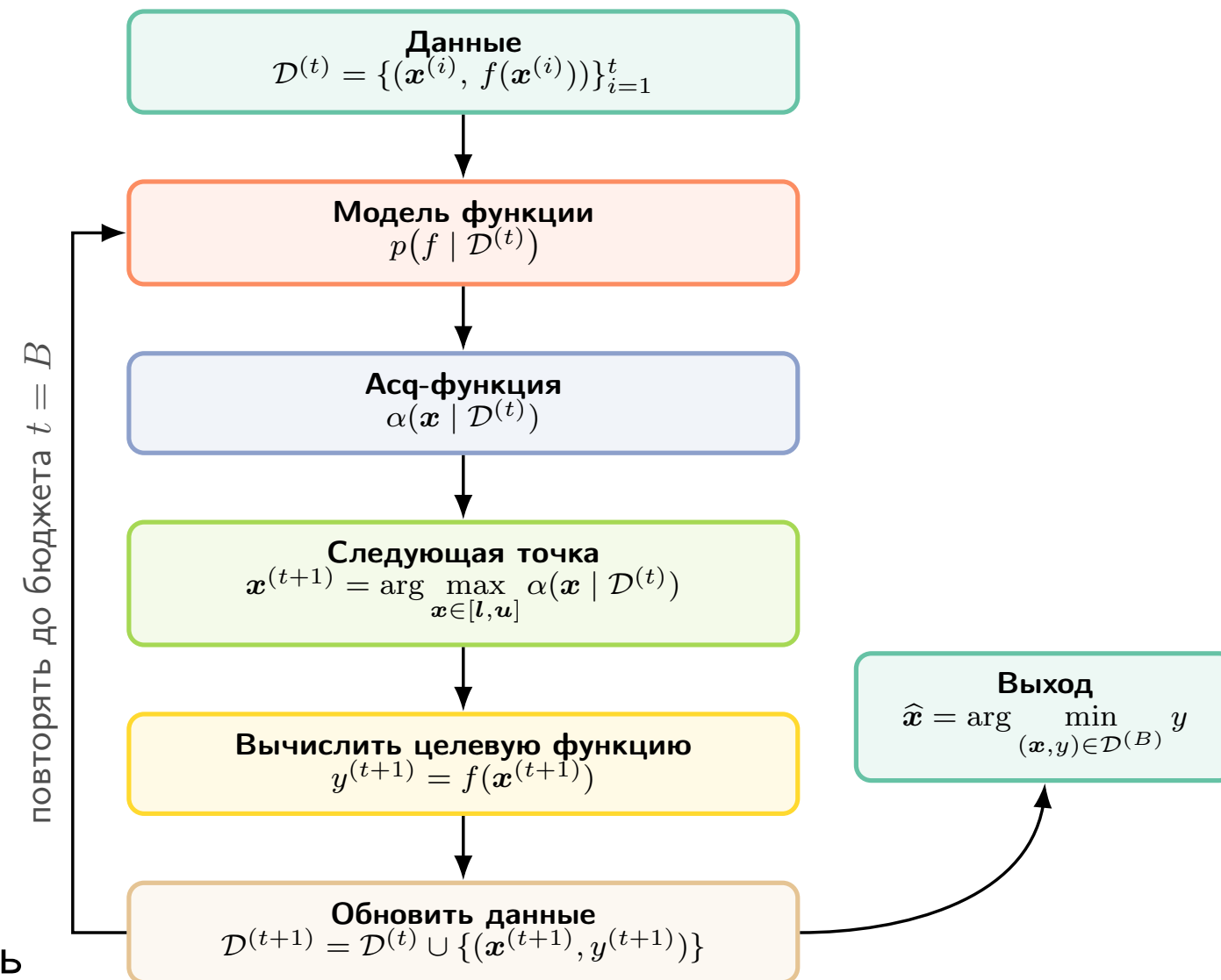
## Как работает Байесовская оптимизация

Байесовская оптимизация (БО)<sup>45</sup>:

1. Инициализация: выбрать несколько точек  $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(t)}$ , посчитать  $f(\mathbf{x}^{(i)})$ .
2. Построить модель функции  $p(f | \mathcal{D}^{(t)})$  по данным  $\mathcal{D}^{(t)} = \{(\mathbf{x}^{(i)}, f(\mathbf{x}^{(i)}))\}_{i=1}^t$ .
3. Построить функцию приобретения (англ. acquisition function)  $\alpha(\mathbf{x} | \mathcal{D}^{(t)})$ . Будем ее называть *асq-функция*.
4. Выбрать следующую точку:

$$\mathbf{x}^{(t+1)} = \arg \max_{\mathbf{x} \in [l, u]} \alpha(\mathbf{x} | \mathcal{D}^{(t)}).$$

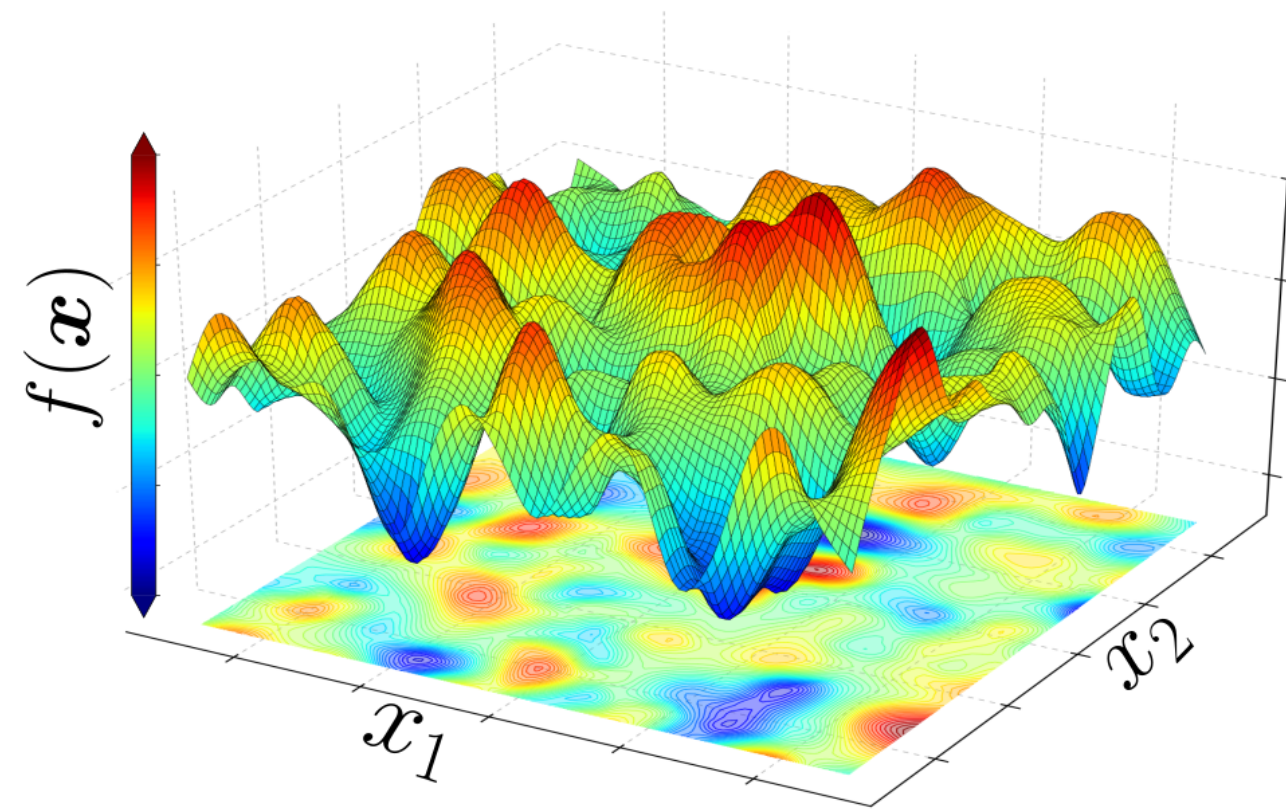
5. Посчитать  $f(\mathbf{x}^{(t+1)})$ , добавить в данные и повторять до бюджета  $t = B$ .



<sup>45</sup> Jones D. R., Schonlau M., Welch W. J. Efficient Global Optimization of Expensive Black-Box Functions / Journal of Global Optimization. 1998. Т. 13, № 4. С. 455-492.

## Цель положения

- БО получает достаточно точные решения при малом бюджете вычислений, но плохо масштабируется по размерности<sup>46</sup>.
- Следовательно, для эффективного применения БО в задачах высокой размерности требуется предварительное снижение размерности задачи.
- Размерности  $d > 20$  уже считаются большими.
- Особенно сложна оптимизация многомерных и мультимодальных функций, т.е. функций с несколькими локальными оптимумами. Такие функции встречаются, например, при обучении нейронных сетей<sup>47</sup>.
- **Цель положения:** сохранить точность БО при заданном числе вычислений целевой функций в высоких размерностях за счёт **адаптации параметров**, задающих **векторное представление** точек в пространстве меньшей размерности.



Пример мультимодальной функции  $f : \mathcal{X} \rightarrow \mathbb{R}$ . На горизонтальной плоскости находится область определения  $\mathcal{X}$ . Каждая точка из этой области имеет две компоненты  $x = (x_1, x_2)$ . По вертикальной оси показано значение функции в каждой из точек. Более холодным цветом показаны меньшие значения, а более теплым – большие значения.

<sup>46</sup> Bull A. D. Convergence rates of efficient global optimization algorithms / *Journal of Machine Learning Research*. 2011. Т. 12, № 10.

<sup>47</sup> Li H., Xu Z., Taylor G., Studer C., Goldstein T. Visualizing the loss landscape of neural nets / *Advances in neural information processing systems*. 2018. Т. 31.

- **Оптимизация:** минимизировать  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  на гиперкубе

$$\boldsymbol{x} \in [\boldsymbol{l}, \boldsymbol{u}] \subset \mathbb{R}^d, \quad \boldsymbol{x}^* \in \arg \min_{\boldsymbol{x} \in [\boldsymbol{l}, \boldsymbol{u}]} f(\boldsymbol{x}).$$

- **Ограничение ресурсов:** бюджет  $B$  вычислений  $f(\boldsymbol{x})$  – допускается не более чем  $B$  вычислений  $f(\boldsymbol{x})$ .
- **Критерий качества при фиксированном бюджете:**

$$f(\hat{\boldsymbol{x}}^*) = \min_{i=1, \dots, B} f(\boldsymbol{x}^{(i)}), \quad \text{минимизировать } \mathbb{E}[f(\hat{\boldsymbol{x}}^*)].$$

- **Цель:** адаптировать параметры, задающие понижение размерности, чтобы уменьшить  $\mathbb{E}[f(\hat{\boldsymbol{x}}^*)]$  и тем самым ускорить достижение малых значений  $f$  при фиксированном  $B$ .

- **Классическая БО (без понижения размерности).** Хорошо работает в малых  $d$ , но деградирует при росте размерности<sup>48</sup>.
- **Понижение размерности через отбор переменных.** Например, присвоение переменным весов в зависимости от их важности<sup>49</sup>.
- **Понижение размерности через аддитивные модели.** Предполагается, что функция может быть представлена как сумма менее сложных функций<sup>50</sup>.
- **Понижение размерности через методы вложения пространства.** Предполагается, что существует низкоразмерное представление исходного пространства, в котором основная изменчивость функции сохраняется. Такие методы позволяют применять БО не в исходном пространстве, а в его компактной проекции<sup>51</sup>.
- **СМА-ES.** Сильный эволюционный алгоритм для вещественной оптимизации, но обычно требует большего бюджета вычислений<sup>52</sup>.

---

<sup>48</sup> *Jones D. R., Schonlau M., Welch W. J.* Efficient Global Optimization of Expensive Black-Box Functions / *Journal of Global Optimization*. 1998. Т. 13, № 4. С. 455-492.

<sup>49</sup> *Ulmasov D., Baroukh C., Chachuat B., Deisenroth M. P., Misener R.* Bayesian optimization with dimension scheduling: Application to biological systems / *Computer Aided Chemical Engineering*. Т. 38. Elsevier, 2016. С. 1051-1056.

<sup>50</sup> *Rolland P., Scarlett J., Bogunovic I., Cevher V.* High-Dimensional Bayesian Optimization via Additive Models with Overlapping Groups / *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics*. PMLR, 03.2018. С. 298-307.

<sup>51</sup> *Raponi E., Wang H., Bujny M., Boria S., Doerr C.* High dimensional Bayesian optimization assisted by principal component analysis / *Parallel Problem Solving from Nature—PPSN XVI: 16th International Conference, PPSN 2020, Leiden, The Netherlands, September 5-9, 2020, Proceedings, Part I* 16. Springer, 2020. С. 169-183.

<sup>52</sup> *Hansen N., Ostermeier A.* Completely Derandomized Self-Adaptation in Evolution Strategies / *Evolutionary Computation*. 2001. Т. 9, № 2. С. 159-195.

- **Прототип:** PCA-BO (метод БО, использующий анализ главных компонент – англ. Principal Component Analysis Bayesian Optimization)<sup>53</sup>.
- PCA-BO строит **линейное** подпространство, на котором затем работает БО.
- **Ограничение прототипа:** линейное представление может плохо описывать область хороших значений  $f(x)$ , если их структура **нелинейна**, что типично для мультимодальных ландшафтов.

---

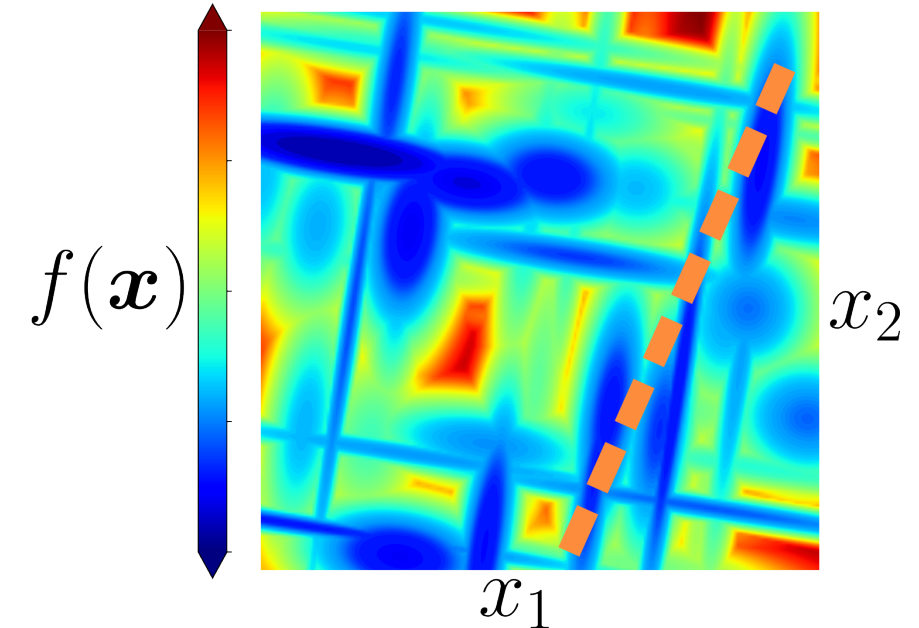
<sup>53</sup> *Raponi E., Wang H., Bujny M., Boria S., Doerr C.* High dimensional Bayesian optimization assisted by principal component analysis / Parallel Problem Solving from Nature–PPSN XVI: 16th International Conference, PPSN 2020, Leiden, The Netherlands, September 5-9, 2020, Proceedings, Part I 16. Springer, 2020. С. 169-183.



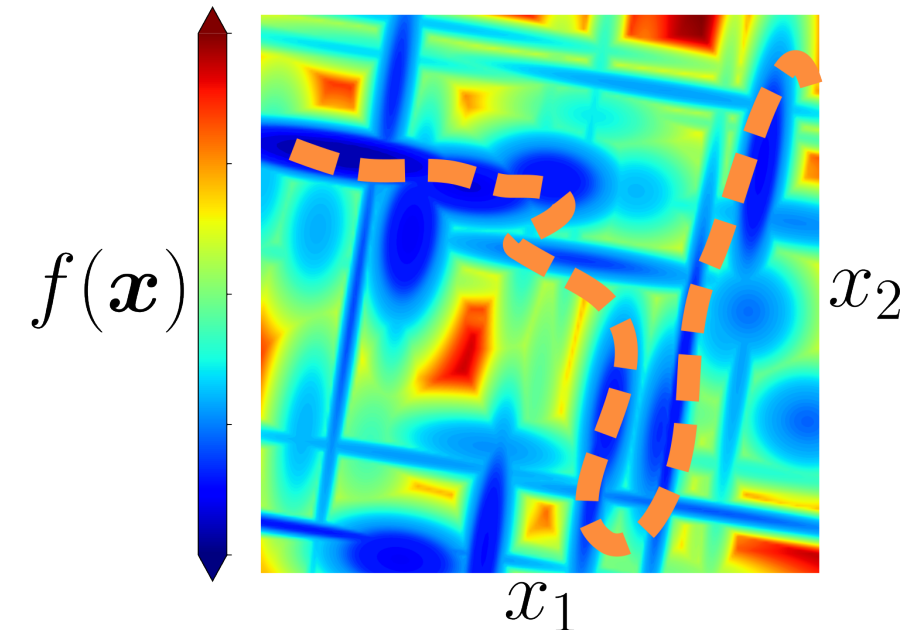
# Третье положение, выносимое на защиту (7/14)

## Линейное и нелинейное подпространства

- На графиках показан ландшафт мультимодальной функции  $f$ . Каждой точке  $x$  на плоскости соответствует значение  $f(x)$ , показанное цветом.
- Решается задача минимизации  $x^* \in \arg \min \{f(x) \mid x \in \mathcal{X}\}$ , поэтому чем более холодный цвет у точки, тем лучше эта точка.
- Пунктиром показаны подпространства. На рисунке (а) подпространство не может проходить через большинство синих областей, содержащих множество достаточно хороших решений задачи оптимизации.
- На рисунке (b) нелинейное подпространство проходит через многие области локальных минимумов, и поэтому оно более информативное.
- Построение модели целевой функции на нелинейном подпространстве позволит получить больше решений высокого качества.



(a) Линейное подпространство



(b) Нелинейное подпространство



### Цель улучшения PCA-ВО

Применить метод **БО** на подпространстве **меньшей размерности**, но заменить линейное подпространство на **нелинейное**: строить и адаптировать нелинейное векторное представление точек  $x \in [l, u]$  так, чтобы оно проходило через области с малыми значениями  $f(x)$ .

- **Что добавляем:** Kernel-PCA<sup>54</sup> (метод ядерного анализа главных компонент – англ. Kernel Principal Component Analysis) вместо PCA.
- Этот метод имеет параметр – функцию ядра. В работе рассматривалось только ядро RBF (радиально-базисная функция – англ. Radial Basis Function, также известная как гауссово ядро)<sup>55</sup>.
- RBF имеет параметр  $\gamma$ , влияющий на получаемое подпространство. Этот параметр адаптируется в процессе работы.

---

<sup>54</sup> *Schölkopf B., Smola A., Müller K.-R.* Nonlinear component analysis as a kernel eigenvalue problem / *Neural Computation*. 1998. Т. 10, № 5. С. 1299-1319.

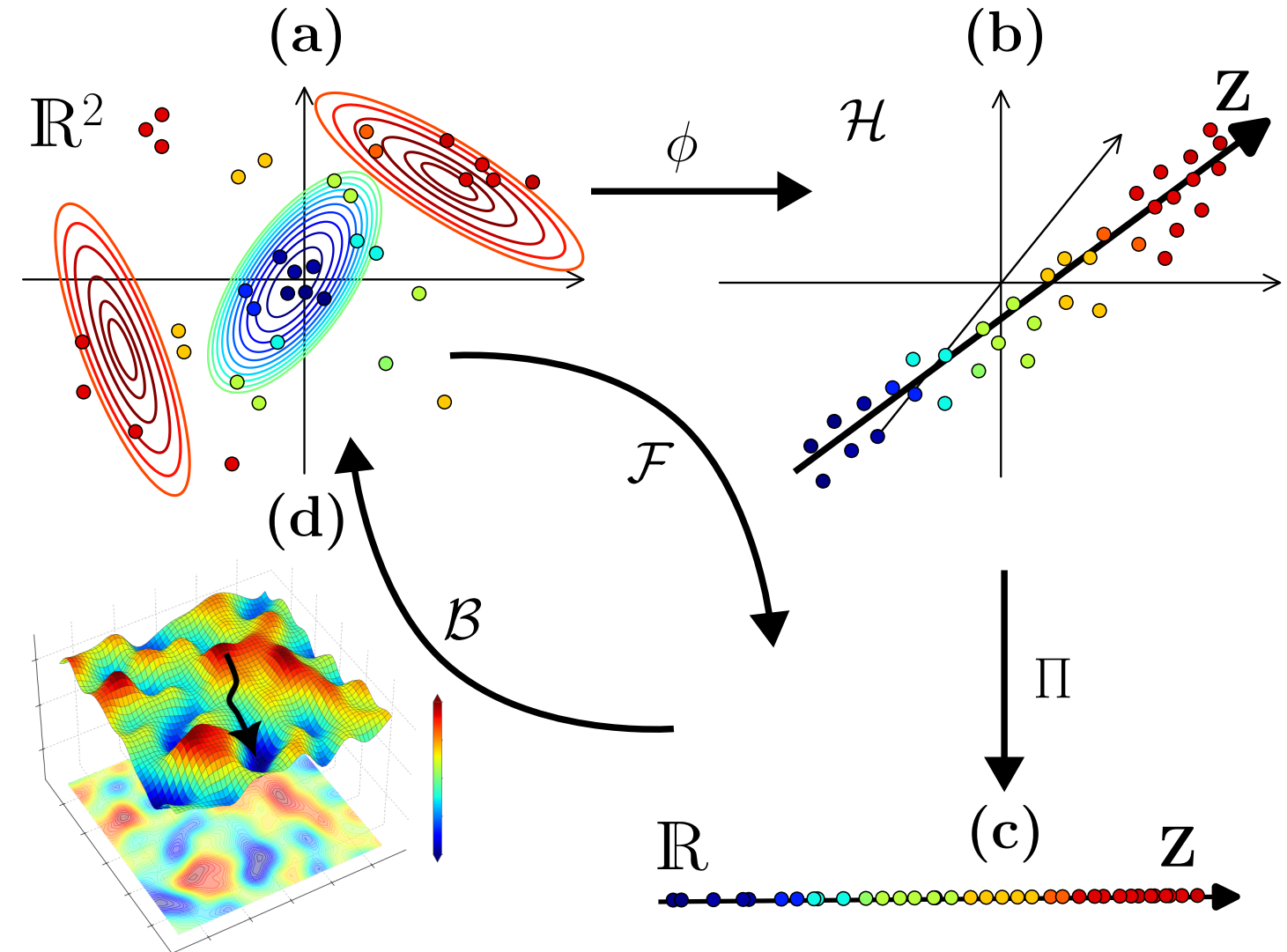
<sup>55</sup> *Rasmussen C. E., Williams C. K. I.* Gaussian processes for machine learning. MIT Press, 2006.

# Третье положение, выносимое на защиту (9/14)

## За счёт чего достигается цель

Построение нелинейного представления точек многомерного пространства на основе ядерного преобразования.

- (a) Масштабированные обучающие точки  $(x, f(x))$  в  $\mathbb{R}^d$  (худшие сдвигаются к центру, лучшие остаются)<sup>56</sup>.
- (b) Ядерное преобразование: отображение через ядро; при удачном  $\gamma$  проявляется скрытая линейная структура.
- (c)  $\mathcal{F}$ : PCA-проекция из  $\mathcal{H}$  в низкую размерность с максимальной сохранённой дисперсией ( $r$  компонент с максимальными собственными числами). Причём  $r \ll d$ .
- (d)  $\mathcal{B}$ : оптимизация предложенной функции для осуществления обратного отображения  $\mathcal{B} : \mathbf{Z} \rightarrow \mathbb{R}^d$ .



<sup>56</sup> *Raponi E., Wang H., Bujny M., Boria S., Doerr C.* High dimensional Bayesian optimization assisted by principal component analysis / Parallel Problem Solving from Nature–PPSN XVI: 16th International Conference, PPSN 2020, Leiden, The Netherlands, September 5-9, 2020, Proceedings, Part I 16. Springer, 2020. C. 169-183.

## Описание предложенного метода

Предложенный метод называется **Kernel-PCA-BO** (метод Байесовской оптимизации, использующий ядерный анализ главных компонент – англ. Kernel Principal Component Analysis Bayesian Optimization).

1. Выбираются начальные точки  $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n_0)}$ , вычисляются значения  $f(\mathbf{x}^{(i)})$ .
2. Точки ранжируются по  $f(\mathbf{x}^{(i)})$  и масштабируются (лучшие точки остаются на месте, худшие сдвигаются в центр).
3. Выбирается параметр  $\gamma$  для RBF-ядра в методе Kernel-PCA. Получаем размерность  $r \ll d$  пространства уменьшенной размерности.
4. Строится прямое отображение  $\mathcal{F} : \mathbf{x} \mapsto \mathbf{z} \in \mathbb{R}^r$  и обратное  $\mathcal{B} : \mathbf{z} \mapsto \mathbf{x} \in \mathbb{R}^d$ .
5. В **подпространстве** выполняется одна итерация БО: строится модель по собранным точкам  $\{(\mathcal{F}(\mathbf{x}^{(t)}), f(\mathbf{x}^{(t)}))\}_{t=1}^n$ , оптимизируется асц-функция<sup>57</sup>. В результате получается новая точка  $\mathbf{z}^{(n+1)}$  в подпространстве.
6. Точка  $\mathbf{z}^{(n+1)}$  восстанавливается в исходное пространство:  $\mathbf{x}^{(n+1)} = \mathcal{B}(\mathbf{z}^{(n+1)})$ . Далее вычисляется значение целевой функции в этой точке  $f(\mathbf{x}^{(n+1)})$ .
7. Когда новая точка попадает в первые 20% решений по значению целевой функции  $f$ , выполняется **адаптация параметров**  $\gamma, r$ .

---

<sup>57</sup> Jones D. R., Schonlau M., Welch W. J. Efficient Global Optimization of Expensive Black-Box Functions / Journal of Global Optimization. 1998. Т. 13, № 4. С. 455-492.

- **Ядро RBF:**

$$k(\mathbf{x}, \mathbf{x}') = \exp(-\gamma \|\mathbf{x} - \mathbf{x}'\|_2^2), \quad \gamma > 0.$$

- Для фиксированного  $\gamma$  метод Kernel-PCA даёт спектр  $\lambda_1(\gamma) \geq \dots \geq \lambda_n(\gamma)$  (по матрице ядра), и выбирается размерность пространства признаков  $\mathbf{Z}$ .
- **Размерность  $r(\gamma)$ :** минимальное число компонент, которое сохраняет долю дисперсии  $\eta$ :

$$r(\gamma) = \min \left\{ k \in [1..n] : \sum_{i=1}^k \lambda_i(\gamma) \geq \eta \sum_{i=1}^n \lambda_i(\gamma) \right\}.$$

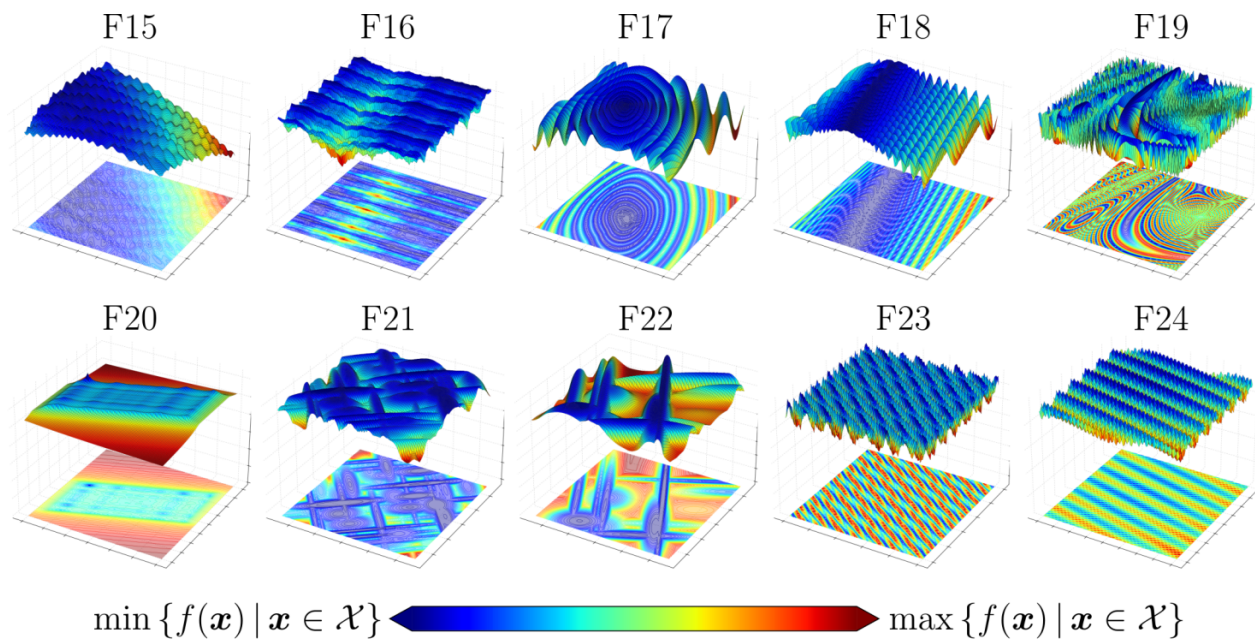
- **Как выбираем  $\gamma$ :** минимизируем “потерю качества” (чем больше, тем хуже):

$$\gamma^* = \arg \min_{\gamma \in (0, \infty)} \underbrace{\left( r(\gamma) - \frac{\sum_{i=1}^{r(\gamma)} \lambda_i(\gamma)}{\sum_{i=1}^n \lambda_i(\gamma)} \right)}_{\mathcal{L}(\gamma)}.$$

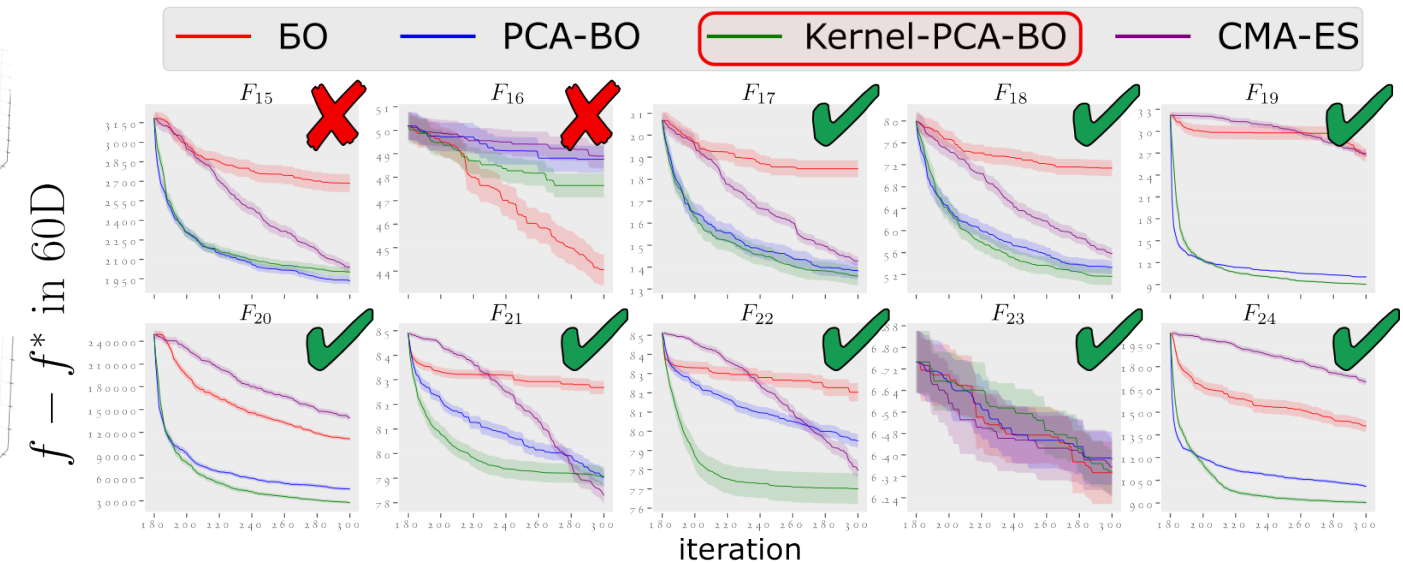
Это одновременно **уменьшает  $r$**  и **увеличивает сохранённую дисперсию**.

- **Численная оптимизация:** ищем  $\gamma^*, r(\gamma^*)$  на заданном отрезке, используя стандартные методы численной оптимизации.

# Третье положение, выносимое на защиту (12/14) Эксперименты на тестовых задачах



(a) Функции VBOB в 2D



(b) Оптимизация функций VBOB в 60D

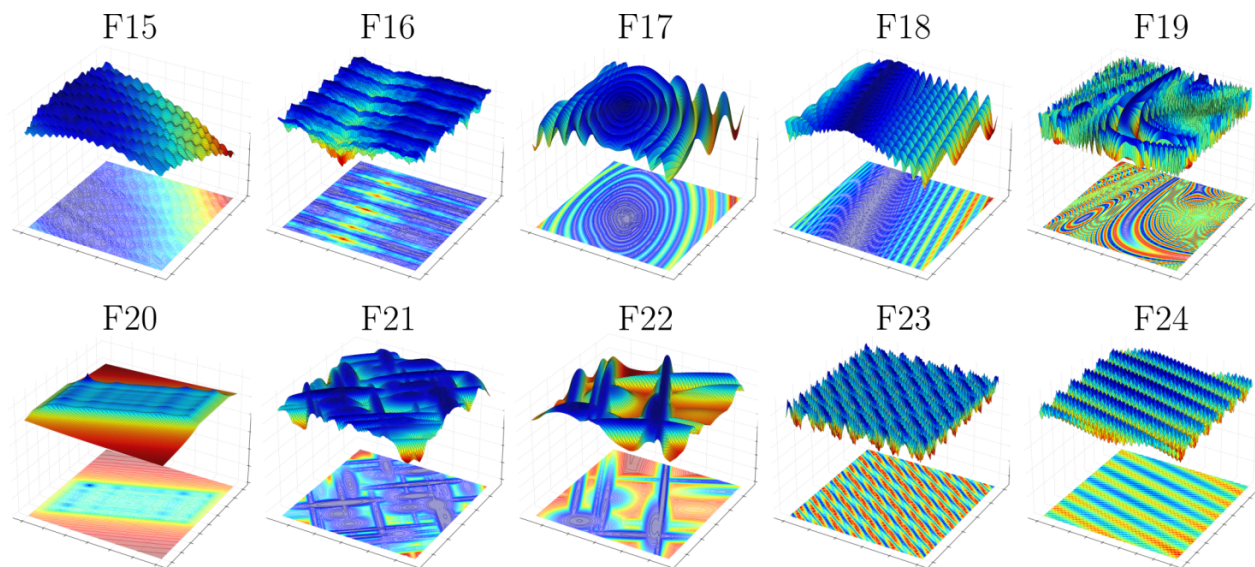
- (a) Сравнение проводилось на функциях VBOB<sup>58</sup> (тестирование оптимизации “черного ящика” – англ. Black-Box Optimization Benchmarking), двумерная версия которых показана на рисунке (a).
- (b) Показаны результаты экспериментов на 60-мерных аналогах этих функций из VBOB. Применялись БО, PCA-BO, CMA-ES и предложенный метод Kernel-PCA-BO. Зеленая галочка означает, что предложенный метод Kernel-PCA-BO либо выигрывает при малом бюджете, либо не уступает лучшему из аналогов. Красный крест означает, что предложенный метод уступает хотя бы одному методу из аналогов.

<sup>58</sup> Hansen N., Auger A., Ros R., Mersmann O., Tušar T., Brockhoff D. COCO: A platform for comparing continuous optimizers in a black-box setting / Optimization Methods and Software. 2020. С. 1-31.



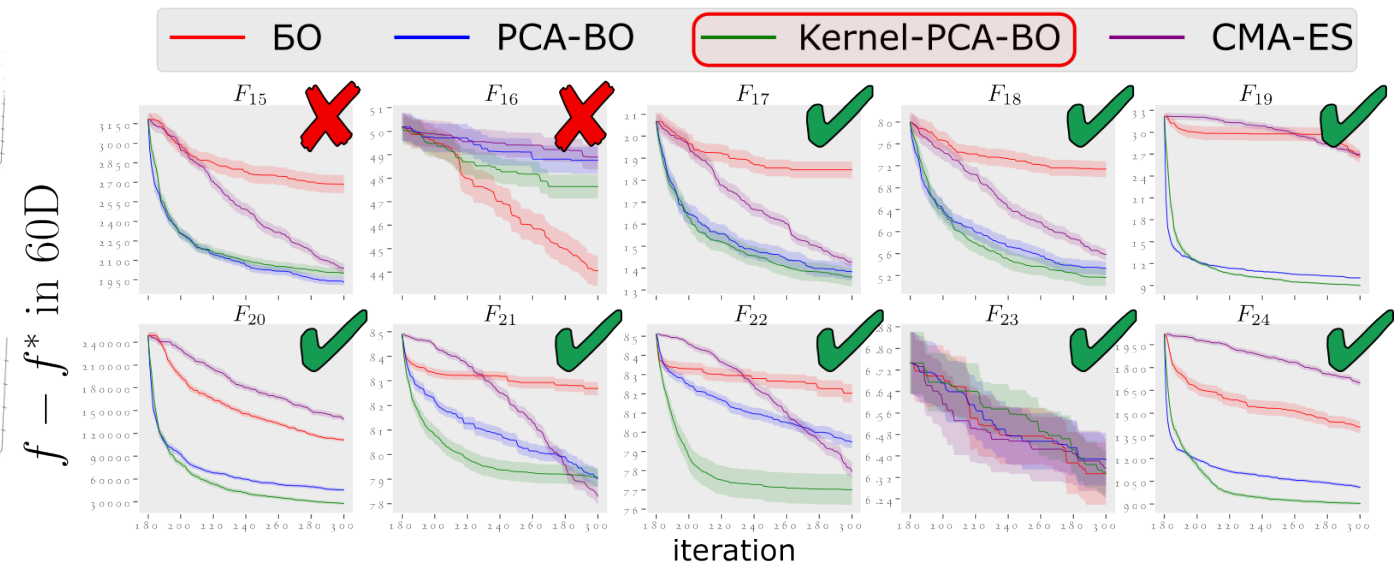
# Третье положение, выносимое на защиту (13/14)

## Выводы по экспериментам



$\min \{f(x) \mid x \in \mathcal{X}\}$    $\max \{f(x) \mid x \in \mathcal{X}\}$

(a) Функции VBOB в 2D



(b) Оптимизация функций VBOB в 60D

- На большинстве функций VBOB предложенный метод Kernel-PCA-BO либо выигрывает при малом бюджете, либо не уступает лучшему из аналогов.
- Метод Kernel-PCA-BO не может изменить векторное представление точек, когда не находятся новые точки с лучшим значением целевой функции.
- Из-за этого наблюдается стагнация при увеличении вычислительного бюджета.



## Выводы, публикации и личный вклад

Разработан метод многомерной Байесовской оптимизации **Kernel-PCA-BO** с понижением размерности вещественнозначных функций, характеризующихся множеством локальных оптимумов, **позволяющий** получать решения со значением целевой функции, близким к оптимальному при ограниченных вычислительных ресурсах на мультимодальных функциях.

### Открытая реализация.

Репозиторий: <https://github.com/wangronin/Bayesian-Optimization/tree/KPCA-BO>

### Результаты опубликованы в:

- 4 **Antonov K.**, *Raponi E., Wang H., Doerr C.* High dimensional Bayesian optimization with kernel principal component analysis / *International Conference on Parallel Problem Solving from Nature (PPSN)*. Springer. 2022. С. 118-131.<sup>A</sup>

### Личный вклад.

В публикации [4] автор предложил и разработал метод Kernel-PCA-BO, реализовал его, организовал и провёл экспериментальное исследование, а также выполнил анализ и интерпретацию результатов (около 85–90% вклада).

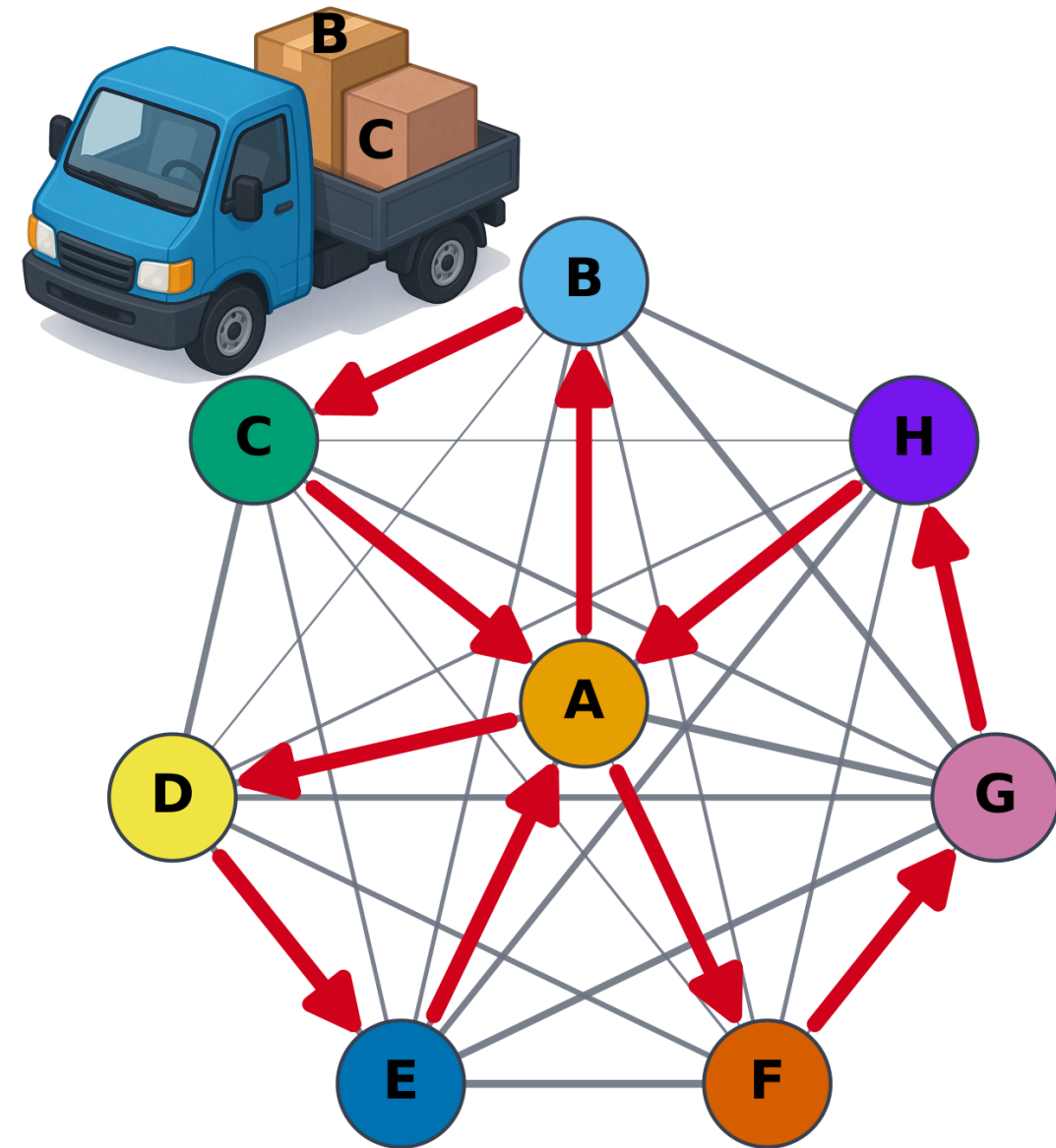
---

<sup>A</sup>Scopus конференция уровня A

# Пример приложения (1/13)

## Решаемая практическая задача

- Маршрутизация доставки грузов с ограничениями на грузоподъёмность<sup>59</sup>.
- Целевая функция: суммарная длина маршрута  $f(x)$  (меньше — лучше). Значение целевой функции решения будем называть *качеством* этого решения.
- Предполагается работа в высокопараллельной среде.
- Цель: минимизировать  $f(x)$  за фиксированное *малое* число итераций стохастического алгоритма.
- Эту постановку обозначим как МИПО – задачу малоитерационной параллельной оптимизации доставки грузов.
- В такой постановке задача, поставленная компанией VeeRoute, ранее не рассматривалась.



<sup>59</sup> *Ralphs T. K., Kopman L., Pulleyblank W. R., Trotter L. E. On the capacitated vehicle routing problem / Mathematical programming. 2003. T. 94. C. 343-359.*

Рассматривается  $(1 + \lambda)$ -ЭА с адаптацией параметра, задающего размер шага мутации  $\theta$ .

1 Инициализация параметров:  $\theta^{(1)} \in \mathbb{R}^m$ ;

2 Инициализация первого поколения:

$$\mathbf{x}^{(1)} \in \mathbf{S}_n;$$

3 **for**  $i \leftarrow 1, 2, \dots, P$  **do**

4     Генерация нового поколения:

$$\mathbf{x}^{(i+1)} \leftarrow \mathcal{A}(f, \theta^{(i)}, \mathbf{x}^{(i)});$$

5     **Адаптация параметров:**

$$\theta^{(i+1)} \leftarrow \mathcal{C}\left(f, \{\theta^{(j)}\}_{j=1}^i, \{\mathbf{x}^{(j)}\}_{j=1}^{i+1}\right);$$

6 Вывод результата оптимизации:

$$\mathbf{x}^* \leftarrow \arg \min \left\{ f(\mathbf{x}) \mid \mathbf{x} \in \bigcup_{j=1}^{P+1} \mathbf{x}^{(j)} \right\};$$

### Цель

Рассмотрим математическое ожидание качества  $\mathbf{x}^*$  наилучшего решения, полученного за это время:

$$\mathcal{L}(f, \mathcal{C}, \mathcal{A}) := \mathbb{E}\left[f(\mathbf{x}^*)\right].$$

Найти функцию  $\mathcal{C}$ , чтобы число  $\mathcal{L}(f, \mathcal{C}, \mathcal{A})$  было как можно меньше.

- **Точные методы:** имеют экспоненциальную сложность и обычно применимы только к малым–средним размерам<sup>60</sup>.
- **Эвристики и метаэвристики:** решают задачу приближенно, требуют меньше времени, чем точные<sup>61,62</sup>.
- **Гибриды:** сочетание эвристик с точными компонентами<sup>63</sup>.
- **Методы машинного обучения и обучения с подкреплением:** нейросети строят маршруты или управляют поиском<sup>64,65</sup>.
- **Однократные и параллельные стратегии:** ускорение получения решений через однократную генерацию и многопоточность<sup>66</sup>.

<sup>60</sup> *Toth P., Vigo D.* Exact solution of the vehicle routing problem / *Fleet management and logistics*. Springer, 1998. С. 1-31.

<sup>61</sup> *Pisinger D., Ropke S.* A general heuristic for vehicle routing problems / *Computers & Operations Research*. 2007. Т. 34, № 8. С. 2403-2435.

<sup>62</sup> *Vidal T., Crainic T. G., Gendreau M., Prins C.* A hybrid genetic algorithm with adaptive diversity management for a large class of vehicle routing problems with time-windows / *Computers & operations research*. 2013. Т. 40, № 1. С. 475-489.

<sup>63</sup> *Helsgaun K.* General k-opt submoves for the Lin–Kernighan TSP heuristic / *Mathematical Programming Computation*. 2009. Т. 1. С. 119-163.

<sup>64</sup> *Kool W., Van Hoof H., Welling M.* Attention, learn to solve routing problems! / *arXiv preprint arXiv:1803.08475*. 2018.

<sup>65</sup> *Ardon L.* Reinforcement Learning to Solve NP-hard Problems: an Application to the CVRP. 2022.

<sup>66</sup> *Bossek J., Doerr C., Kershke P., Neumann A., Neumann F.* Evolving Sampling Strategies for One-Shot Optimization Tasks / *Parallel Problem Solving from Nature — PPSN XVI*. Springer, 2020.

- Точные методы, эвристики, метаэвристики, гибриды и некоторые методы, основанные на машинном обучении, ограничены в своей масштабируемости по числу параллельных потоков и предназначены для получения приближённого решения рассматриваемой задачи оптимизации за значительное время. Это является ограничением в условиях, приближенных к реальному времени.
- Остальные методы машинного обучения и однократной оптимизации предполагают получение решения за одну итерацию. Поэтому возможно улучшение получаемого решения за счёт некоторого увеличения временного интервала.

# Пример приложения (5/13)

## Описание предложенного метода

(a) **Выбор обучающих экземпляров МИПО:** выбираем  $N_{\text{tr.}}$  задач из заданного класса.

(b) **Обучение на каждом экземпляре (независимо):** на **каждом** обучающем экземпляре  $f^{(k)}$ ,  $k \in \{1, 2, \dots, N_{\text{tr.}}\}$  **обучаем свою нейросеть**  $\mathcal{N}_k$ , которая задаёт **адаптацию параметра мутации** по состоянию ЭА  $\theta^{(t)} = \mathcal{N}_k(\mathbf{s}^{(t)}; \phi)$ :

$$\mathbf{s}^{(t)} = \left( \left( \frac{f(\mathbf{y}^{(t)})}{f(\mathbf{x}^{(t)})} \right)^C, \left( \frac{f(\mathbf{y}^{(t-1)})}{f(\mathbf{x}^{(t-1)})} \right)^C, \dots, \frac{\theta^{(t)} - \theta_{\min}}{\theta_{\max} - \theta_{\min}}, \frac{\theta^{(t-1)} - \theta_{\min}}{\theta_{\max} - \theta_{\min}}, \dots \right).$$

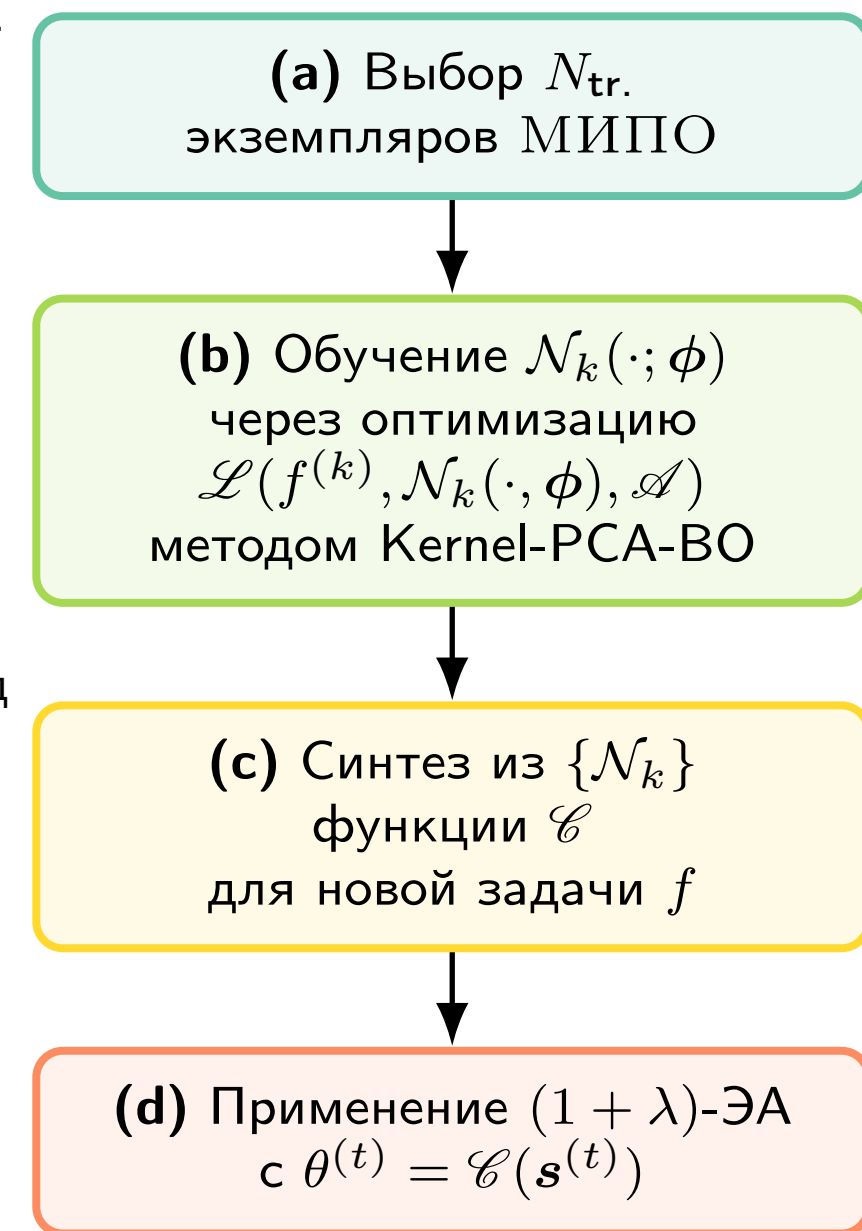
Обучение заключается в подборе  $\phi \in \mathbb{R}^d$  для минимизации  $\mathcal{L}(f^{(k)}, \mathcal{N}_k(\cdot, \phi), \mathcal{A})$ . Параметры  $\phi$  подбираются методом **Kernel-PCA-BO** (метод 3). Гиперпараметры Kernel-PCA-BO выбираются с использованием результатов численного анализа (метод 2) на задачах RUGGEDNESS и PLATEAU.

(c) **Итоговая адаптация для класса:** из множества  $\{\mathcal{N}_k\}_{k=1}^{N_{\text{tr.}}}$  строим **одно правило** адаптации, применимое к любому экземпляру из класса:

$$k^* := \underbrace{\arg \min_{k \in [1, N_{\text{tr.}}]} \left\{ |V(f) - V(f^{(k)})| \right\}}_{\text{ближайшее число вершин к экземпляру } f}, \quad \mathcal{C}(\mathbf{s}) := \frac{V(f^{(k^*)})}{V(f)} \mathcal{N}_{k^*}(\mathbf{s}),$$

где  $V(f)$  – число вершин в экземпляре  $f$ .

(d) **Применение:** используем полученную  $\mathcal{C}$  в заданном  $(1 + \lambda)$ -ЭА для оптимизации  $f$ .





### Проблема

Каждая нейронная сеть обучается для адаптации параметров на своём экземпляре задачи  $f$ . Однако отсутствует набор данных о том, какие параметры являются оптимальными на каждом этапе оптимизации.

### Прототип: обучение с подкреплением<sup>67</sup>

Используется модель ценности  $Q(s, a)$ , оценивающая полезность набора параметров  $\theta$  в состоянии  $s$ :

$$\theta^{(t)} \in \arg \max_{\theta} Q(s^{(t)}, \theta).$$

### Недостатки прототипа

- Оценка  $Q$  опирается на долгосрочные эффекты (зависимость от последующих итераций), что может приводить к неустойчивостям обучения<sup>68</sup>.
- Преимущества чаще демонстрируются на тестовых задачах, а перенос на прикладные области, такие как МИПО, остаётся нетривиальным.

<sup>67</sup> *Nguyen T., Le P., Doerr C., Dang N.* Multi-parameter Control for the  $(1+(\lambda, \lambda))$ -GA on OneMax via Deep Reinforcement Learning / *Proceedings of the 18th ACM/SIGEVO Conference on Foundations of Genetic Algorithms*. 2025. С. 190-201.

<sup>68</sup> *Nguyen T., Le P., Biedenkapp A., Doerr C., Dang N.* On the Importance of Reward Design in Reinforcement Learning-based Dynamic Algorithm Configuration: A Case Study on OneMax with  $(1+(\lambda, \lambda))$ -GA / *Proceedings of the Genetic and Evolutionary Computation Conference*. 2025. С. 1162-1171.

### Идея предложенного метода для МИПО

Вместо отдельной модели  $Q$  предлагается напрямую искать функцию адаптации параметров  $\mathcal{C}$ , которая даёт наилучший результат оптимизации.

### Поиск функции, задающей адаптацию параметров

Пусть качество оптимизации задаётся функционалом потерь  $\mathcal{L}(f, \mathcal{C}, \mathcal{A})$ . Тогда требуется найти

$$\mathcal{C}^* \in \arg \min_{\mathcal{C}} \mathcal{L}(f, \mathcal{C}, \mathcal{A}).$$

### Переход к нейронной сети

По универсальной теореме аппроксимации<sup>69</sup> полагается, что любая из функций  $\{\mathcal{C}\}$  может быть достаточно точно приближена нейронной сетью. Тогда

$$\mathcal{C}^* \approx \mathcal{N}(\cdot, \phi^*), \quad \phi^* \in \arg \min_{\phi} \mathcal{L}(f, \mathcal{N}(\cdot, \phi), \mathcal{A}).$$

Таким образом, задача поиска функции адаптации сводится к задаче обучения нейронной сети, а она, в свою очередь, сводится к задаче оптимизации.

<sup>69</sup> *Hornik K., Stinchcombe M., White H.* Multilayer feedforward networks are universal approximators / *Neural networks*. 1989. Т. 2. № 5. С. 359-366.

### Третий метод, выносимый на защиту

- Метод, **позволяет** получать решения со значением целевой функции, близкими к оптимальному при ограниченных вычислительных ресурсах на мультимодальных функциях в многомерных пространствах.
- Вычисление  $\mathcal{L}$  затратно: требуется полный запуск стохастического алгоритма на  $f$ .
- Пространство весов  $\phi$  многомерно, а ландшафт по  $\phi$  обычно мультимодален.

### Второй метод, выносимый на защиту

- Метод **позволил получить** численные приближения **нижних оценок** времени работы этих алгоритмов до достижения оптимума на **практически значимых функциях**, в том числе на RUGGEDNESS и PLATEAU, обладающих свойствами, близкими к МИПО.
- **Показано**, что на задачах RUGGEDNESS и PLATEAU  $(1 + \lambda)$ -ЭА **со статическими параметрами** часто работают **не хуже** (а иногда и лучше), чем варианты с рассмотренной адаптацией. Поэтому для корректного сравнения с предлагаемым решением МИПО целесообразно включать в базовые аналоги **набор** алгоритмов со статическими параметрами (несколько настроек).
- Метод **позволил выявить** свойства ландшафта функционала потерь  $\mathcal{L}$  в пространстве функций адаптации  $\mathcal{C}$  на экземплярах МИПО: тепловые карты демонстрируют **высокую чувствительность** к выбору параметров, особенно на поздних этапах. Поэтому надёжнее оптимизировать **ожидаемый итоговый результат оптимизации** –  $\mathcal{L}$ , чем опираться на промежуточные прогнозы, например  $Q$ .

### Первый метод, выносимый на защиту

- Метод **обеспечивает** автоматическую адаптацию покомпонентной вероятности мутации в  $(1 + \lambda)$ -ЭА **без предварительной настройки параметров**, включая корректировку нижней границы на основе коллективного отбора потомков.
- Используется как **простой и эффективный аналог** для сравнения с предложенным нейросетевым методом адаптации параметров.
- Применяется при решении задачи МИПО (в вычислительных экспериментах).

Сравнение проводилось со следующими группами  $(1 + \lambda)$ -ЭА.

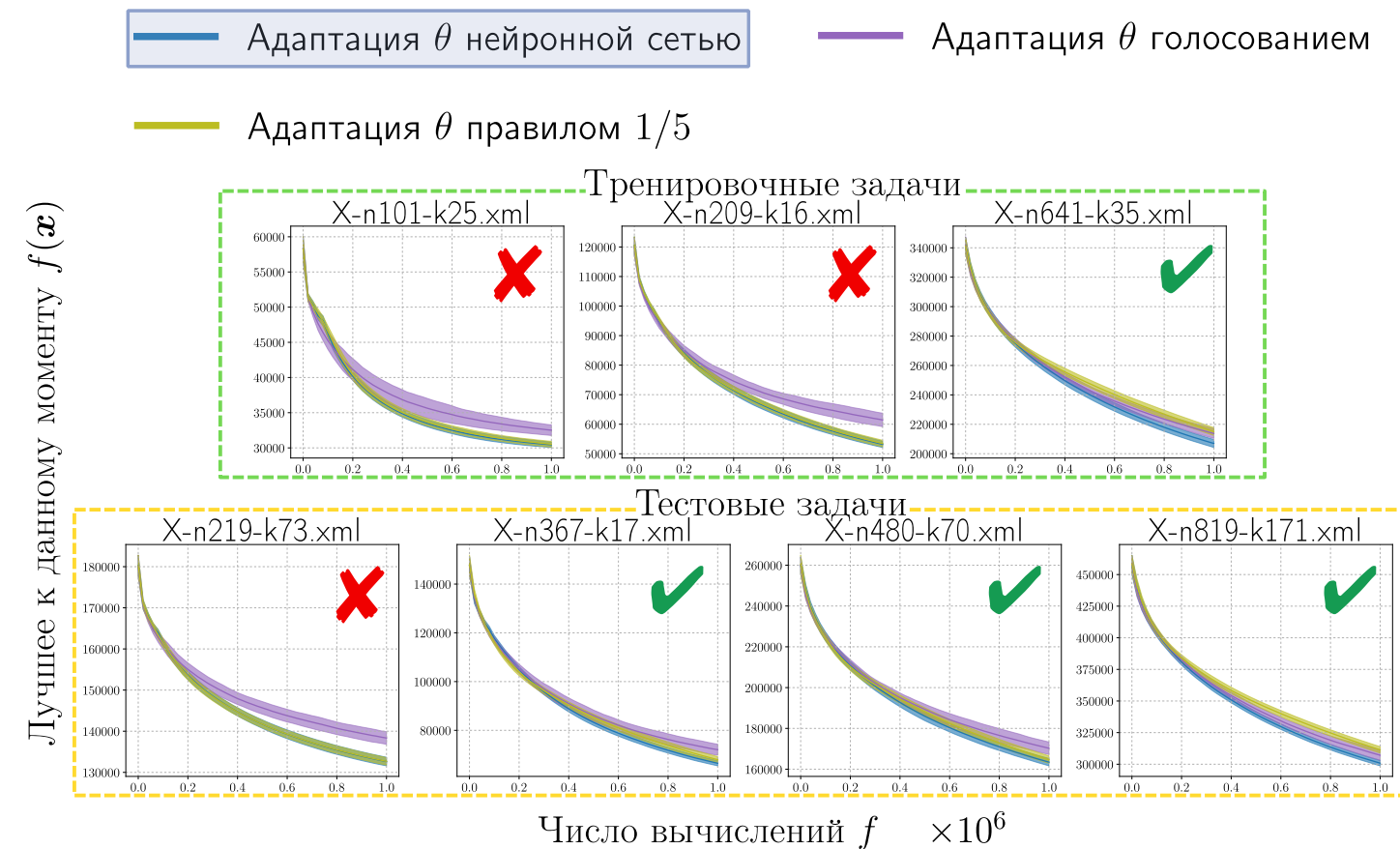
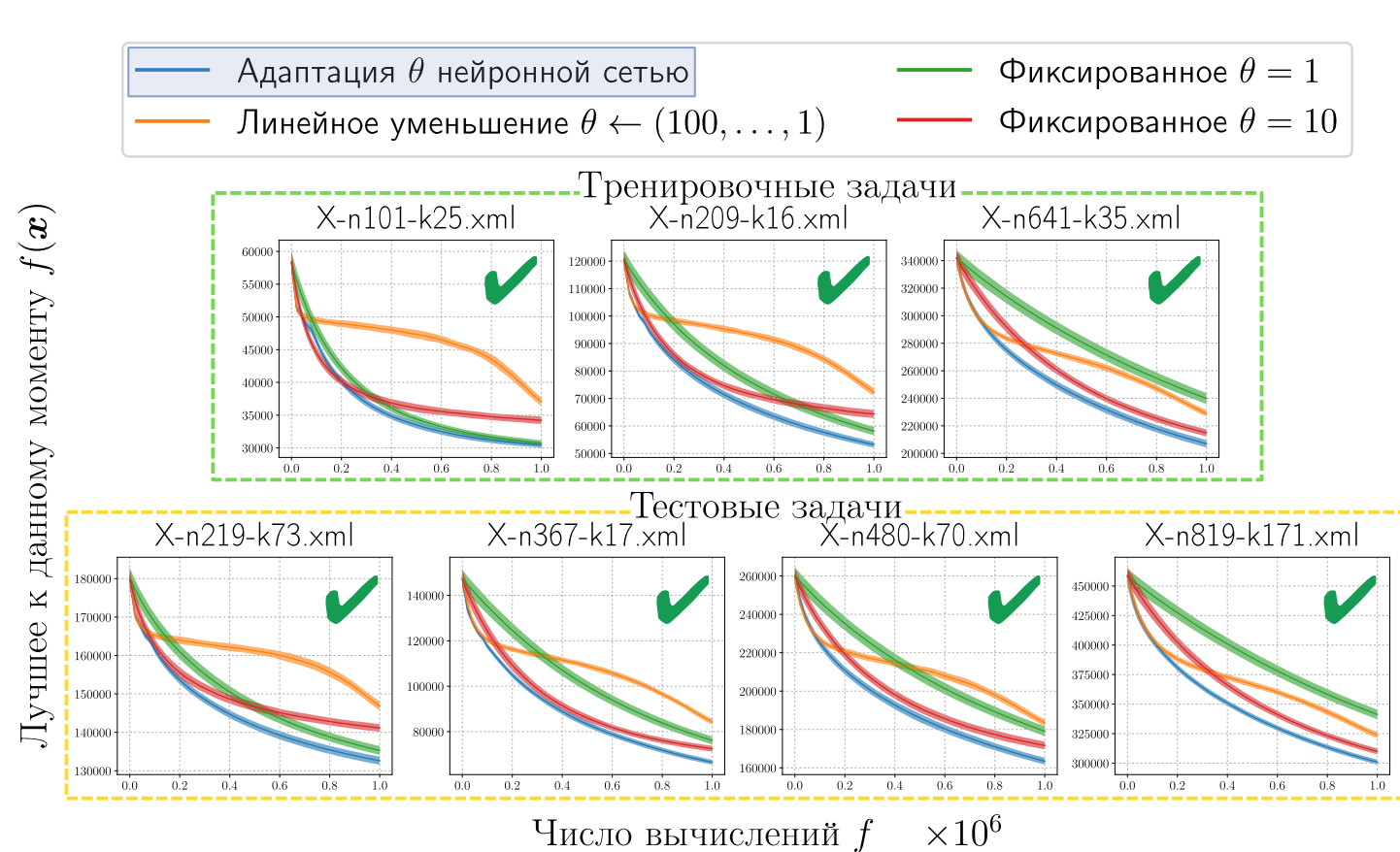
- $(1 + \lambda)$ -ЭА без адаптации параметров. Актуальны, так как с помощью **второго метода, выносимого на защиту**, было установлено, что такие алгоритмы работают за время, близкое к наилучшему возможному, при адаптации размера шага мутации на задачах со свойствами, как у МИПО.
  1. **Линейное уменьшение:**  $\theta$  убывает по номеру итерации от  $\theta_{\max}$  к  $\theta_{\min}$
  2. **Фиксированное  $\theta = 1$ , фиксированное  $\theta = 10$**  – отобраны по результатам предварительных экспериментов.
- $(1 + \lambda)$ -ЭА использующие адаптацию  $\theta$ :
  1. **Адаптация  $\theta$  голосованием**, используется **первый метод, выносимый на защиту**.
  2. **Адаптация  $\theta$  правилом 1/5**, популярный представитель классических методов адаптации<sup>70</sup>.

---

<sup>70</sup> [Doerr B., Doerr C. Optimal parameter choices through self-adjustment: Applying the 1/5-th rule in discrete settings / Proceedings of the 2015 Annual Conference on Genetic and Evolutionary Computation. 2015. C. 1335-1342.](#)

# Пример приложения (10/13)

## Эксперименты



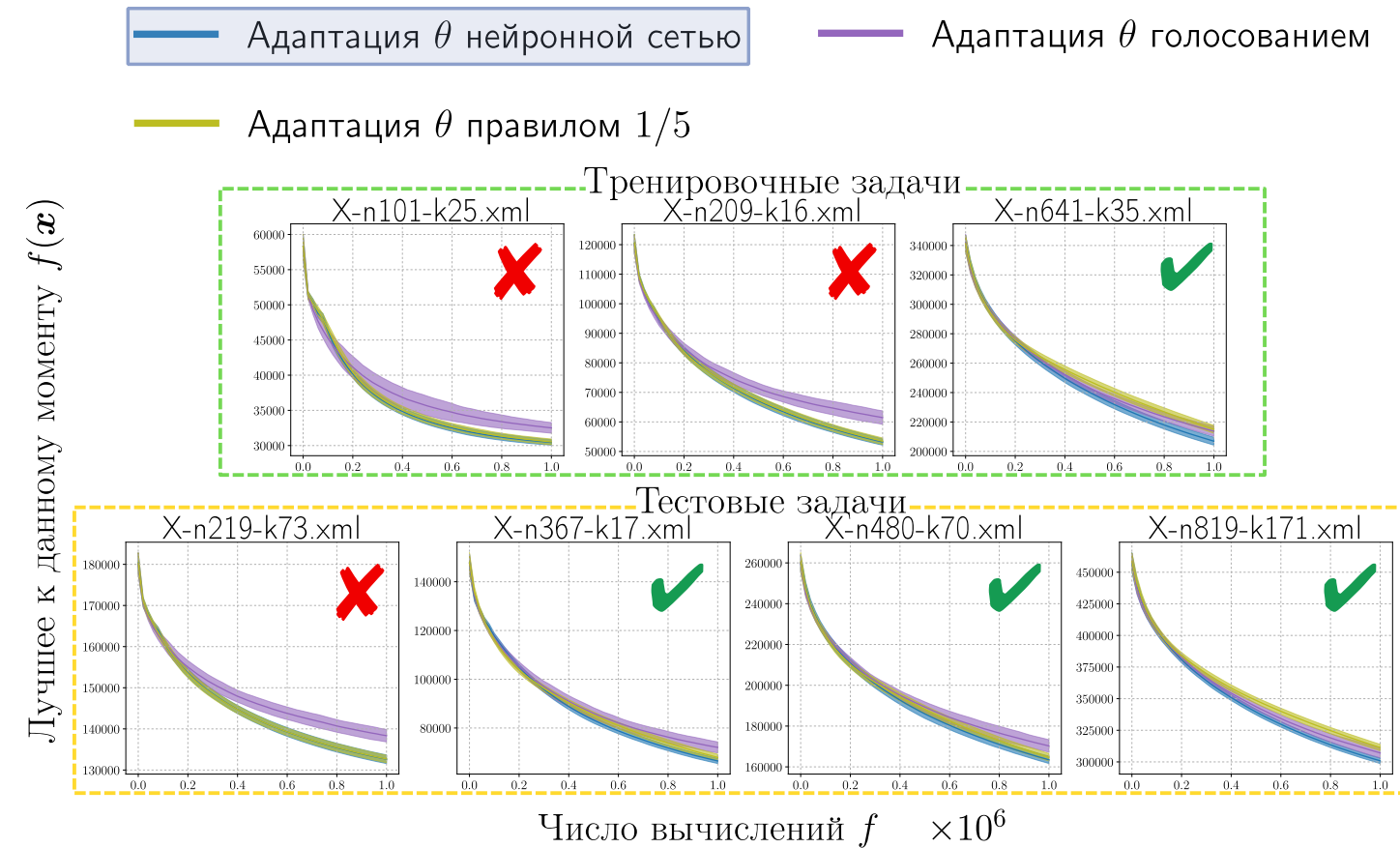
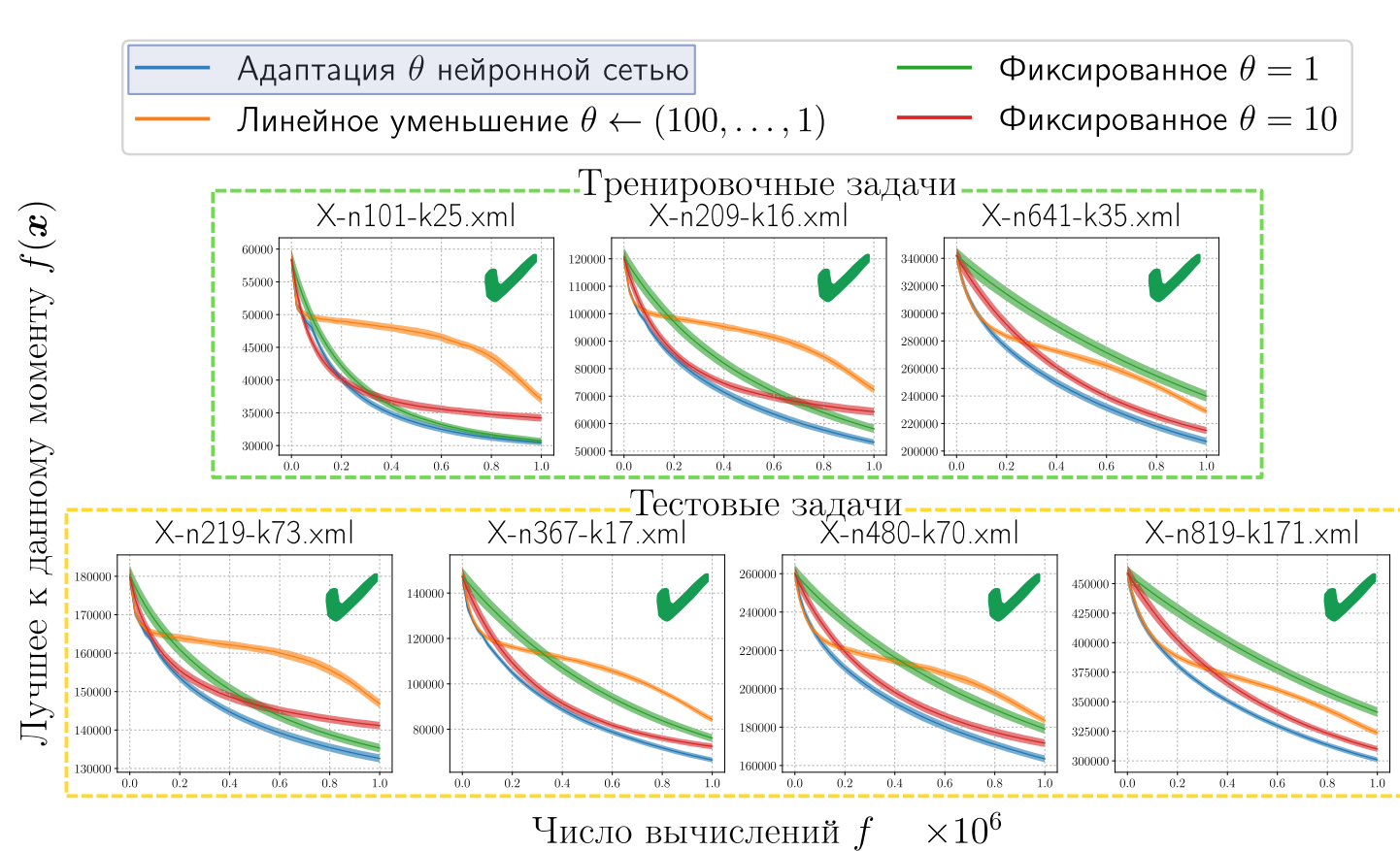
- Для демонстрации метод применён к классу задач из открытого источника<sup>71</sup>.
- В качестве тренировочных задач выбраны те, которые показаны в зелёной рамке. В жёлтой показаны тестовые задачи. Число вершин указано в названии после  $n$ , например, X-n101-k25.xml означает, что в графе  $n = 101$  вершина.
- Зелёная галочка означает, что предложенный метод с адаптацией  $\theta$  превосходит аналоги. Красный крест означает, что предложенный метод такой же, как один из аналогов по качеству результата оптимизации.

<sup>71</sup> Uchoa E., Pecin D., Pessoa A., Poggi M., Vidal T., Subramanian A. New benchmark instances for the capacitated vehicle routing problem / European Journal of Operational Research. 2017. T. 257, № 3. С. 845-858.



# Пример приложения (11/13)

## Анализ результатов



- Предложенный метод на каждой из задач либо лучший, либо такой же, как лучший по качеству результата оптимизации.
- Предложенный метод в среднем на 1.6% превосходит лучший из аналогов по качеству результата оптимизации.
- Лучший из аналогов разный на различных рассмотренных задачах, а предложенный метод всегда либо лучше этого аналога, либо такой же по качеству результата оптимизации.
- Статистическая значимость этого результата подтверждается  $t$ -критерием Уэлча с поправкой Бонферрони.



1. Предложен метод, обучающий по одной нейронной сети на каждой из тренировочных задач МИПО, выбирающий одну нейронную сеть из обученных для новой задачи МИПО и использующий выбранную нейронную сеть для адаптации шага мутации в  $(1 + \lambda)$ -ЭА при решении этой задачи.
2. Проведено сравнение предложенного метода с аналогами. В среднем по всем рассмотренным задачам улучшение решения составляет 1.6% относительно лучшего из аналогов.
3. В главе **использованы все три метода, выносимые на защиту**.
4. Получен акт о предполагаемом внедрении в логистическую платформу комбинаторной оптимизации компании VeeRoute: <https://veeroute.ru/>.
5. Реализация метода доступна на GitHub:  
<https://github.com/kiralexant/pcsi-fewshot-cvrp>.

# Пример приложения (13/13)

## Предполагаемое внедрение

Планируется внедрение предложенного метода, использующего нейронную сеть для адаптации параметров, в компании VeeRoute.

**VEEROUTE**

Общество с ограниченной ответственностью «ВИИРОУТЕ РНД»

ОГРН 1157847178830, ИНН 7801280722, КПП 784001001  
199106, г. Санкт-Петербург, вн.тер.г. муниципальный округ № 7,  
линия 24-я В.О., д. 15/2, литера А, помещ. 1Н, ком. 224  
тел.: (812) 331-15-70, <http://veeroute.com>

АКТ

от 10.10.2025

Составлен о предполагаемом внедрении результатов диссертационной работы Кирилла Александровича Антонова на тему «Методы адаптации параметров в стохастических алгоритмах для оптимизации информационно-вычислительных процессов», представленной на соискание ученой степени кандидата технических наук по специальности 2.3.8. «Информатика и информационные процессы» (технические науки).

Настоящий акт подтверждает то, что метод, основанный на применении многомерной байесовской оптимизации для обучения нейронной сети, использующейся для адаптации параметров в стохастическом алгоритме, который предложен К.А. Антоновым в его диссертационной работе, планируется внедрить в логистическую платформу комбинаторной оптимизации компании «VeeRoute» (<https://veeroute.ru/>) для решения задачи доставки грузов.

Нейронную сеть для адаптации параметров планируется обучить и использовать как часть спроектированного алгоритма компании, работающего в высокопараллельной среде и условиях ограниченного времени до получения ответа.

Генеральный директор



В.А. Кудинов


В рамках работы разработан комплекс методов адаптации параметров в стохастических алгоритмах для оптимизации крупномасштабных информационно-вычислительных процессов при ограниченных вычислительных ресурсах.

- **Метод 1.** Разработан метод адаптации покомпонентной вероятности мутации в  $(1 + \lambda)$ -эволюционных алгоритмах, не требующий предварительной настройки параметров, **позволяющий** повысить быстродействие указанных алгоритмов и обеспечить автоматическую корректировку нижней границы вероятности мутации.
- **Метод 2.** Разработан метод оценки быстродействия  $(1 + \lambda)$ -эволюционных алгоритмов с адаптацией покомпонентной вероятности мутации на основе сочетания **динамического программирования** и **Монте-Карло-моделирования**, который **позволил** получать численные приближения нижних оценок времени работы указанных алгоритмов до достижения оптимума на **практически значимых функциях**.
- **Метод 3.** Разработан метод многомерной Байесовской оптимизации **Kernel-PCA-BO** с понижением размерности вещественнозначных функций, характеризующихся множеством локальных оптимумов, **позволяющий** получать решения со значением целевой функции, близким к оптимальному при ограниченных вычислительных ресурсах.
- **Пример приложения.** Разработанные методы использованы для обучения и анализа нейронных сетей, используемых для адаптации параметров в стохастических алгоритмах, применяемых в высокопараллельной среде для приближённого решения задачи доставки грузов (МИПО).

- A 1. *Rodionova A., Antonov K., Buzdalova A., Doerr C.* Offspring population size matters when comparing evolutionary algorithms with self-adjusting mutation rates / *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO)*. 2019. C. 855-863.
2. *Antonov K., Buzdalova A., Doerr C.* Mutation Rate Control in the Evolutionary Algorithm with a Self-adjusting Lower Bound / *International Conference on Mathematical Optimization Theory and Operations Research (MOTOR)*. Springer. 2020. C. 305-319.
-  3. *Antonov K., Buzdalov M., Buzdalova A., Doerr C.* Blending Dynamic Programming with Monte Carlo Simulation for Bounding the Running Time of Evolutionary Algorithms / *IEEE Congress on Evolutionary Computation (CEC)*. 2021. C. 878-885.
- A 4. *Antonov K., Raponi E., Wang H., Doerr C.* High dimensional Bayesian optimization with kernel principal component analysis / *International Conference on Parallel Problem Solving from Nature (PPSN)*. Springer. 2022. C. 118-131.

---

A Scopus конференция уровня A

 Приз за лучшую статью (Scopus конференция уровня B)

1. The Genetic and Evolutionary Computation Conference (GECCO), Прага, Чехия. 2019.
2. Mathematical Optimization Theory and Operation Research (MOTOR), Новосибирск, Россия. 2020.
3. IEEE Congress on Evolutionary Computation (IEEE CEC), Краков, Польша. 2021.
4. Parallel Problem Solving from Nature (PPSN XVII), Дортмунд, Германия. 2022.
5. Научная и учебно-методическая конференция университета ИТМО, СПб., 2022.
6. XI Конгресс молодых ученых, СПб., ИТМО, 2022.
7. XII Конгресс молодых ученых, СПб., ИТМО, 2023.
8. The Genetic and Evolutionary Computation Conference (GECCO), Мельбурн, Австралия. 2024.



Спасибо за внимание!

**it's** **MO** *re than a*  
**UNIVERSITY**