**© 2025 г. И.А. АКИНФИЕВ (i@iakinfiev.ru)**
**(Санкт-Петербургский государственный университет),**
**О.Н. ГРАНИЧИН, д-р физ.-мат. наук (o.granichin@spbu.ru)**
**(Санкт-Петербургский государственный университет;**
**Институт проблем машиноведения РАН, Санкт-Петербург),**
**Е.Ю. ТАРАСОВА (elizaveta.tarasova@spbu.ru)**
**(Санкт-Петербургский государственный университет)**

# ПОИСКОВЫЙ МЕТОД СТОХАСТИЧЕСКОЙ НЕСТАЦИОНАРНОЙ ОПТИМИЗАЦИИ ФУНКЦИИ С ГЕЛЬДЕРОВСКИМ ГРАДИЕНТОМ[1]

В статье рассматривается поисковый метод стохастической оптимизации с возмущением на входе, предназначенный для отслеживания изменений точки минимума функции (трекинга) с гельдеровским градиентом в условиях наблюдений при почти произвольных неизвестных ограниченных помехах (unknown–but–bounded noise). Подобные методы используются в задачах адаптивного управления (энергетика, логистика, робототехника, трекинг целей), оптимизации зашумленных систем (биомоделирование, физические эксперименты) и онлайн-обучения с дрейфом параметров данных (финансы, потоковая аналитика).

В качестве апробации алгоритма исследуется эффективность его работы в условиях, имитирующих отслеживание эволюции человеческих ожиданий в задачах обучения с подкреплением на основе обратной связи от человека и при отслеживании центра кластера задач в системах массового обслуживания. Поисковые методы с возмущениями на входе активно развивались в работах Б.Т. Поляка с 1990 г.

*Ключевые слова*: трекинг, возмущение на входе, рандомизация, стохастическая оптимизация, безградиентные методы, обучение с подкреплением на основе обратной связи от человека, системы массового обслуживания, неизвестные, но ограниченные помехи.

**DOI:** 10.31857/S0005231025080013, **EDN:** USSKZZ

## 1. Введение

Задача минимизации функции (или функционала) $f(x)$ лежит в основе решения множества практических задач от управления техническими системами до машинного обучения. Аналитические решения часто недоступны из-за высокой размерности, нелинейностей или отсутствия явного вида функции. Даже при аналитическом задании функции их практическая применимость ограничена вычислительными ресурсами, неточностями измерений или ошибками округления. Традиционные итерационные градиентные

[1] Теоретическая часть работы, разделы 1–4, выполнены при финансовой поддержке РНФ в ИПМАШ РАН, (проект № 23-41-00060), практическая часть работы, разделы 5–6, выполнены при финансовой поддержке СПбГУ, шифр проекта 121061000159-6.

методы эффективны для поиска минимума гладких или дифференцируемых функций. Однако в реальных задачах часто возникают ситуации, когда вычисление градиента затруднено или невозможно. Обычно целевая функция подвержена стохастическим возмущениям, либо ее явный вид неизвестен. На практике оптимизируемая функция часто задается некоторым оракулом, обращаясь к которому с запросами (аргументами функции) можно получить возможные реализации. Доступность измерений непосредственно градиента возможна при реализации специальных измерительных устройств для той или иной задачи или за счет конечно-разностных аппроксимаций, которые неработоспособны при высоком уровне помех в получаемых измерениях. В таких случаях требуются альтернативные подходы, не зависящие от информации о градиенте.

Значительный вклад в развитие теории и методов стохастической оптимизации внес Б.Т. Поляк и вся научная группа, с которой он работал. Их исследования охватывают широкий спектр вопросов, включая градиентные методы [1], псевдоградиентные алгоритмы адаптации и обучения [2–4] и методы ускорения сходимости [5–7]. Две статьи [8, 9] дают и сейчас исчерпывающие ответы при анализе сходимости итеративных стохастических алгоритмов в общем случае в терминах среднеквадратических отклонений, а также в линейном случае в терминах матриц ковариаций ошибки. Предложенный в 1990 г. новый поисковый метод стохастической аппроксимации [10] не только развивает общее направление алгоритмов случайного поиска [11], но и существенно продвигает вперед всю общую теорию итеративных алгоритмов оптимизации: в статье показано, что при наблюдении значений оптимизируемой функции с помехами предложенный алгоритм имеет асимптотически оптимальную скорость сходимости в том смысле, что невозможно найти более быстрый алгоритм среди всевозможных итеративных алгоритмов оптимизации для достаточно широкого класса функций. Ранее похожий алгоритм был предложен в [12], и для него была обоснована состоятельность оценок при наблюдении на фоне почти произвольных помех. В англоязычной литературе похожие методы получили название SPSA (Simultaneous Perturbation Stochastic Approximation) [13, 14]. Важной особенностью этих безградиентных методов является то, что вне зависимости от размерности задачи на каждой итерации надо вызывать оракул всего один или два раза с аргументами, выбираемыми на случайной прямой (рандомизация алгоритма), проходящей через точку очередной оценки. Подробный анализ истории развития, а также свойств оценок поисковых алгоритмов стохастической аппроксимации с возмущением на входе, дан в [15–17].

Важным ограничением классических итерационных методов стохастической оптимизации нулевого порядка (не использующих значения градиента), таких как процедура Кифера–Вольфовица [18] в многомерном случае, является необходимость многократного вычисления функции на каждой итерации. Это становится особенно непрактичным в динамических средах, где целевая функция $f_n(x)$ изменяется со временем. Подобная ситуация возника-

ет, например, в задачах оптимизации систем реального времени. Оказалось, что методы типа предложенных ранее поисковых алгоритмов стохастической оптимизации с возмущением на входе в этой ситуации продолжают быть работоспособными при замене уменьшающихся со временем размеров шагов на постоянные [19, 20]. Позже удалось сформулировать и обосновать свойства распределенного алгоритма такого типа, совмещенного с алгоритмом консенсуса [20].

На практике [21, 22] часто сталкиваются со статистическими неопределенностями, которые не имеют второго статистического момента. Например, стабильные распределения, в частности, Леви–Парето лучше описывают цены на акции и товары, чем нормальные распределения [23]. В [24] исследованы свойства оценок алгоритма типа SPSA в таких условиях. В статье эти исследования распространены на случай оптимизации нестационарного функционала среднего риска.

## 2. Постановка задачи

Пусть время является дискретным и определяется номером шага (итерации) $n = 0, 1, \ldots$, $\{F_n(\cdot,\cdot) : \mathbb{R}^d \times \mathbb{R}^q \to \mathbb{R}\}$ – набор функций от двух векторных переменных, дифференцируемых по первому аргументу, на каждом шаге $n$ в известных (выбираемых) точках $x_n$ (план эксперимента) производятся наблюдения (измеряются значения)

$$y_n = F_n(x_n, w_n) + v_n, \tag{1}$$

где $w_n$ – неконтролируемые возмущения, заданные на некотором вероятностном пространстве $\Omega$ и имеющие одинаковое, неизвестное распределение $P_w(\cdot)$, $v_n$ – помехи в наблюдениях (возможно и неслучайные).

Обозначим $\mathcal{F}_{n-1}$ – $\sigma$-алгебру всех вероятностных событий, которые реализовались до момента $n$, $\mathbb{E}$ – символ математического ожидания, $\mathbb{E}_{\mathcal{F}_{n-1}}$ – символ условного математического ожидания при условии $\sigma$-алгебры $\mathcal{F}_{n-1}$.

Рассмотрим задачу минимизации нестационарного функционала среднего риска:

$$f_n(x) = \mathbb{E}_{\mathcal{F}_{n-1}} F_n(x, w) = \int\limits_{\mathbb{R}^q} F_n(x, w) P_w(dw) \to \min_x . \tag{2}$$

Требуется оценить точку минимума $\theta_n$ функции $f_n(x)$, т.е. найти

$$\theta_n = \arg\min_x f_n(x).$$

Точность оценки $x$ точек $\theta_n$ характеризуется с помощью скалярных функций Ляпунова

$$V_n(x) = \|x - \theta_n\|^{\rho+1} = \sum_{i=1}^{n} |x^{(i)} - \theta_n^{(i)}|^{\rho+1},$$

где $\theta_n$ – искомые векторы, $\rho \in (0, 1]$ – показатель Гельдера градиентов функций $V_n(x)$. Далее в статье будут использоваться обозначения $\|\cdot\|_{\rho+1}$ для нормы $l_{\rho+1}$ и $\langle\cdot,\cdot\rangle$ для скалярного произведения в $\mathbb{R}^d$.

Для характеристики поведения оценок точек минимума нестационарного функционала (2) введем два определения.

*Определение 1. Последовательность оценок $\hat{\theta}_n$ точек минимума $\theta_n$ называется $l_{\rho+1}$-стабилизированной, если существует такое $C > 0$, что*

$$\mathbb{E}V_n(\hat{\theta}_n) \leqslant \text{const}, \ \forall\, n.$$

*Определение 2. Число $L$ называется асимптотической верхней границей ошибок оценивания по $l_{\rho+1}$-норме, если для последовательности оценок $\{\hat{\theta}_n\}$ точек минимума $\theta_n$ выполняется:*

$$\overline{\lim_{n\to\infty}} \mathbb{E}V_n(\hat{\theta}_n) \leqslant L < \infty.$$

Далее будем рассматривать задачу о построении последовательности стабилизирующихся оценок $\{\hat{\theta}_n\}$, в рамках определения 2, при следующих условиях, выполняющихся при любом $n > 0$.

(A) *Функции $f_n(\cdot)$ сильно выпуклые по первому аргументу:*

$$\langle \nabla V_n(x), \nabla f_n(x)\rangle \geqslant \mu V_n(x).$$

(B) *При любом $w$ градиенты функций $\nabla F_n(\cdot, w)$ удовлетворяют условию:*

$$\|\nabla F_n(x, w) - \nabla F_n(y, w)\|_1 \leqslant M\|x - y\|_\rho^\rho$$

с некоторой константой $M$.

(C) *Локальное свойство Лебега:* $\forall x \in \mathbb{R}^d$ $\exists$ окрестность $U_x$ точки $x$ и функция $\Phi_x(w)$, такие что : $\mathbb{E}\Phi_x(w) < \infty$ и $\|\nabla F_n(x', w)\|_2 \leqslant \Phi_x(w)$ $\forall x' \in U_x$.

(D) *Скорость дрейфа точки минимума ограничена следующими условиями*

$$a: \ \ \|\theta_n - \theta_{n-1}\|_1 \leqslant A,$$

*либо, если $\{\theta_n\}$ представляют собой последовательность случайных величин, то*

$$\mathbb{E}_{\mathcal{F}_{n-1}}\|\theta_n - \theta_{n-1}\|_{\rho+1}^{\rho+1} \leqslant A^{\rho+1},$$

*а также*

$$\begin{aligned}
&\text{b:} \quad \mathbb{E}_{\mathcal{F}_{n-1}} \|\nabla_x F_n(x, w) - \nabla_x F_{n-1}(x, w)\|_1 \leqslant B\|x - \theta_{n-1}\|_1^\rho,\\
&\text{c:} \quad \mathbb{E}_{\mathcal{F}_{n-1}} \|\nabla_x F_n(\theta_n, w_n)\|_{\rho+1}^{\rho+1} \leqslant C,\\
&\text{d:} \quad \mathbb{E}_{\mathcal{F}_{2n-2}} |F_{2n}(x, w_{2n}) - F_{2n-1}(x, w_{2n-1})|^{\rho+1} \leqslant DV_{2n-2}(x) + E.
\end{aligned}$$

(E) *Для помех наблюдения $v_n$ выполнены условия:*

$$|v_{2n} - v_{2n-1}| \leqslant \sigma_v,$$

*либо, если они представляют собой последовательность случайных величин, то*

$$\mathbb{E}_{\mathcal{F}_{2n-2}}\{|v_{2n} - v_{2n-1}|^{\rho+1}\} \leqslant \sigma_v^{\rho+1}.$$

Заметим, что последнему условию удовлетворяют любые детерминированные ограниченные последовательности $\{v_n\}$. Условие (C) обеспечивает возможность перестановки операций интегрирования и дифференцирования при обосновании стабилизируемости оценок. Ограничения типа (D) включают как дрейф типа случайных блужданий, так и направленный дрейф в определенную сторону. Например, в [1] приводится ограничение на основе (D):

$$\theta_n = \theta_{n-1} + a + \xi_n,$$

где $\xi_n$ – центрированная случайная величина и $a$ – тренд. Стабилизируемость оценок алгоритма поиска минимума в условиях (D) означает применимость его к широкому классу различных задач.

### 3. Поисковый рандомизированный алгоритм оценивания

Пусть $\{\Delta_n\}$ – последовательность пробных одновременных возмущений, подаваемых на вход алгоритма оценивания, является некоторой реализацией последовательности независимых бернуллиевских случайных векторов из $\mathbb{R}^d$, у которых каждая компонента независимо принимает с вероятностью $\frac{1}{2}$ значения, равные $\pm\frac{1}{\sqrt{d}}$. Выберем некоторый начальный вектор $\theta_0 \in \mathbb{R}^d$. Будем оценивать последовательность точек минимума $\{\theta_n\}$ последовательностью $\{\hat{\theta}_n\}$, определяемой алгоритмом стохастической оптимизации с пробным одновременным возмущением на входе, который имеет следующий вид:

$$\begin{cases} \hat{\theta}_{2n-1} = \hat{\theta}_{2n-2}, \\ x_{2n} = \hat{\theta}_{2n-2} + \beta\Delta_n,\ x_{2n-1} = \hat{\theta}_{2n-2} - \beta\Delta_n, \\ \hat{\theta}_{2n} = \hat{\theta}_{2n-2} - \dfrac{\alpha}{2\beta}\Delta_n(y_{2n} - y_{2n-1}), \end{cases} \tag{3}$$

где $\alpha$ и $\beta$ – параметры размеров шага алгоритма. Для обоснования стабилизации оценок алгоритма (3) будем считать, что

(F) *случайные векторы $\Delta_n$ и $w_{2n}, w_{2n-1}$ не зависят между собой, а также от $\mathcal{F}_{n-1}$. Если $\{v_n\}$ предполагаются случайной природы, то $\Delta_n$ не зависят от $v_{2n}, v_{2n-1}$.*

### 4. Стабилизация оценок

Обозначим $H = A + \alpha\beta M$, где $A$ и $M$ – константы ограничения скорости дрейфа точки и ограничение изменения градиентов соответственно.

*Теорема 1. Пусть выполнены условия* (A)–(F) *и параметры алгоритма $\alpha, \beta$ выбраны такими, чтобы константа $K > 0$, определенная далее при доказательстве теоремы, была меньше единицы. Тогда для любого начального приближения $\hat{\theta}_0$ с $E\|\hat{\theta}_0 - \theta_0\|^{\rho+1} < \infty$ оценки алгоритма* (4) *стабили-*

*зируются в следующем смысле:*

$$\overline{\lim_{n\to\infty}}\mathbb{E}\|\hat{\theta}_n - \theta_n\|_{\rho+1} \leqslant \left(\frac{L}{K}\right)^{\frac{1}{\rho+1}},$$

*где $L$ также определена в конце доказательства теоремы.*

Условия (A)–(C), (E)–(F) являются стандартными для доказательства состоятельности оценок алгоритмов стохастической оптимизации с возмущением на входе [18]. Ранее факт среднеквадратичной стабилизации оценок алгоритма (3) был доказан в [19] при более жестких ограничениях.

Доказательство теоремы 1 и определение констант $K$ и $L$ приводится в Приложении 1.

## 5. Моделированиия в RLHF-сценарии

В задачах обучения с подкреплением на основе обратной связи от человека (Reinforcement Learning from Human Feedback, RLHF) ключевой вызов заключается в работе с зашумленными и нестабильными данными [26, 27]. Человеческие оценки часто содержат случайные ошибки и могут изменяться со временем, что усложняет процесс оптимизации. В частности, в задачах, связанных с тонкой настройкой языковых моделей (Large Language Models, LLM), RLHF используется для улучшения качества генерации текста, согласования с предпочтениями пользователей и минимизации нежелательного поведения моделей. Однако субъективность и изменчивость человеческих оценок создают значительные трудности для традиционных методов оптимизации [28–30].

### *5.1. Модель*

В моделировании исследуется эффективность поискового алгоритма в условиях, приближенных к реальным: при наличии шума с тяжелыми хвостами (распределение Парето) и дрейфа предпочтений [28, 30], имитирующего эволюцию человеческих ожиданий. Рассматриваются три сценария: умеренный дрейф, почти стационарные предпочтения и стационарные предпочтения с асимметричным шумом. Это позволяет оценить устойчивость и адаптивность поискового алгоритма в условиях, характерных для RLHF, и определить его применимость для задач, связанных с обучением LLM и других систем, где человеческая обратная связь играет ключевую роль.

Цель моделирования – проверить способность RLHF-агента адаптироваться к модели вознаграждения, которая формируется на основе зашумленных и изменяющихся человеческих оценок. Далее:

- моделируется шум с тяжелыми хвостами (Pareto-распределение), отражающий неопределенность и редкие, но значительные отклонения в оценках;
- вводится модель дрейфа предпочтений, имитирующий постепенное изменение человеческих ожиданий;
- все функции и параметры вводятся в условиях (A)–(F) части 2.

Агент должен минимизировать расхождение между своей оценкой параметра и истинным значением, задаваемым моделью вознаграждения, несмотря на шум и динамику целевых предпочтений, для минимизации используется поисковый алгоритм.

Модель вознаграждения, основанная на RLHF, задается следующим образом:

$$F_n(\mathbf{x}) = -\sum_{i=1}^{m}(x_i - x_n^*)^{1,35}, \tag{4}$$

где $x_n^*$ – целевой параметр, который дрейфует со временем $n$

$$x_n^* = x_{n-1}^* + \delta, \quad x_0^* = 5,$$

отражая изменение предпочтений.

При выборе $x_n$ на основе обратной связи получаем

$$y_n = F_n(\mathbf{x}) + v_n,$$

где $v_n$ – шум, моделирующий неопределенность в обратной связи. Для моделирования использовались два вида шума:

- симметричный шум: $v_i = Z_i \times \text{sign}_i$, где $Z_i \sim \text{Pareto}(\beta, \sigma)$, $\text{sign}_i \sim$ $\sim \text{Uniform}(\{-1, 1\})$;
- асимметричный шум: $v_i = Z_i$, где $Z_i \sim \text{Pareto}(\beta, \sigma)$, что может отражать склонность к завышению оценок.

В табл. 1 представлены основные параметры, использованные при проведении численного моделирования. Они охватывают структуру эксперимента, настройки алгоритма (гиперпараметры), а также характеристики шума и сценарии дрейфа, имитирующие условия нестабильной обратной связи.

**Таблица 1**. Параметры моделирования

| Параметр | Описание | Значение |
|---|---|---|
| Начальная оценка агента | Начальная точка для обучения | $\hat{\theta}_0 = 0$ |
| Число итераций | Количество шагов адаптации | $N = 1000$ |
| Число запусков | Количество независимых экспериментов | $m = 1000$ |
| Гиперпараметры | | |
| Шаг обучения | Консервативный шаг для устойчивости | $\gamma = 0{,}05$ |
| Размер возмущения | Амплитуда для оценки градиента | $c = 0{,}1$ |
| Характеристики шума | | |
| Параметр формы | Определяет тяжесть хвостов | $\beta = 1{,}6$ |
| Масштаб | Интенсивность отклонений | $\sigma = 2{,}0$ |
| Скорость дрейфа | Умеренный дрейф параметров | $\delta = 0{,}01$ |
| | Почти стационарный режим | $\delta = 0{,}0001$ |
| Тип шума | Случайные отклонения | Симметричный |
| | Систематическое смещение | Асимметричный |

### *5.2. Сценарии моделирования*

Для анализа поискового адаптивности рассматриваются три сценария:

1. Умеренный дрейф предпочтений ($\delta = 0{,}01$) и симметричный шум (здесь и далее: шум с симметричным распределением). Имитирует постепенное изменение целевых значений при наличии случайных ошибок в оценках.
2. Почти стационарные предпочтения ($\delta = 0{,}0001$) и симметричный шум. Проверка точности настройки в условиях, близких к стабильным.
3. Стационарные предпочтения с асимметричным шумом ($\delta = 0{,}0001$) и асимметричный шум (здесь и далее: шум с несимметричным распределением). Отражает систематическое искажение обратной связи, например, постоянное завышение оценок.

### *5.3. Процесс адаптации агента*

Агент обновляет свою оценку параметра $\hat{\theta}$ на основе наблюдаемых значений $y$ (вознаграждений), полученных из модели. Алгоритм следует итерационной схеме, описанной в формуле (3):

На каждой четной итерации $k = 2n$ (где $n = 1, 2, \dots$):

1. Используется оценка с предыдущего четного шага, $\hat{\theta}_{2n-2}$ (для $n = 1$ используется $\hat{\theta}_0$).
2. Создается случайный вектор возмущений $\Delta_n$, где каждая компонента независимо принимает значение $+1$ или $-1$ с вероятностью 0,5.
3. Формируются две точки согласно формуле (3):

$$x_{2n} = \hat{\theta}_{2n-2} + \beta\Delta_n,$$
$$x_{2n-1} = \hat{\theta}_{2n-2} - \beta\Delta_n.$$

4. Наблюдаются значения (вознаграждения) в возмущенных точках: $y_{2n}$ (соответствующее $x_{2n}$) и $y_{2n-1}$ (соответствующее $x_{2n-1}$). Эти $y$ включают как истинное значение функции, так и шум ($y_n = F_n(x_n, w_n) + v_n$ в терминах статьи).
5. Обновление оценки: Оценка $\hat{\theta}$ обновляется по формуле, аналогичной третьей строке системы (3), но со знаком "+", так происходит максимизация:

$$\hat{\theta}_{2n} \leftarrow \hat{\theta}_{2n-2} + \frac{\alpha}{2\beta}\Delta_n(y_{2n} - y_{2n-1}).$$

На каждой нечетной итерации $k = 2n - 1$:

1. Копирование оценки: $\hat{\theta}_{2n-1} \leftarrow \hat{\theta}_{2n-2}$.

*5.4. Проверка условий* (A)–(F) *для моделированиия в RLHF-сценарии*

(A) Сильная выпуклость функции $f_n(\mathbf{x})$.

$$\nabla f_n(\mathbf{x}) = -\nabla F_n(\mathbf{x}) = -\left[1{,}35(x_1 - x_n^*)^{0,35}, \ldots, 1{,}35(x_m - x_n^*)^{0,35}\right]^\top,$$

$$\nabla V_n(\mathbf{x}) = [(\rho+1)\mathrm{sign}(x_1 - x_n^*)|x_1 - x_n^*|^\rho, \ldots, (\rho+1)\mathrm{sign}(x_m - x_n^*)|x_m - x_n^*|^\rho]^\top,$$

$$\langle \nabla V_n(\mathbf{x}), \nabla f_n(\mathbf{x})\rangle = -1{,}35(\rho+1)\sum_{i=1}^m |x_i - x_n^*|^{\rho+0,35}.$$

Используя неравенство $|x_i - x_n^*|^{\rho+0,35} \geqslant |x_i - x_n^*|^{\rho+1} a^{-0,65}$, где $a \leqslant |x_i - x_n^*|$, получаем:

$$\sum_{i=1}^m |x_i - x_n^*|^{\rho+0,35} \geqslant a^{-0,65}\sum_{i=1}^m |x_i - x_n^*|^{\rho+1} = a^{-0,65} V_n(\mathbf{x}).$$

Таким образом,

$$\langle \nabla V_n(\mathbf{x}), \nabla f_n(\mathbf{x})\rangle \leqslant -1{,}35(\rho+1)a^{-0,65} V_n(\mathbf{x}),$$

т.е. условие вида $\langle \nabla V_n(\mathbf{x}), \nabla f_n(\mathbf{x})\rangle \geqslant \mu V_n(\mathbf{x})$ выполнено при $\mu = -1{,}35(\rho + +1)a^{-0,65} < 0$. При минимизации $f_n(\mathbf{x})$ условие сильной выпуклости в смысле данного скалярного неравенства выполняется при $\mu < 0$.

(B) Гельдерова непрерывность градиента.

Для функцию вознаграждения $F_n(x)$ градиент имеет вид:

$$\nabla F_n(x) = -1{,}35 \cdot \left[(x_1 - x_n^*)^{0,35}, \ldots, (x_m - x_n^*)^{0,35}\right]^\top.$$

Для любого $i$ компонентная разность градиентов оценивается как:

$$|\partial_i F_n(x) - \partial_i F_n(y)| = 1{,}35\left|(x_i - x_n^*)^{0,35} - (y_i - x_n^*)^{0,35}\right| \leqslant 1{,}35M'|x_i - y_i|^{0,35},$$

где $M'$ – константа Гельдера, зависящая от ограниченной области, на которой определены $x_i$ и $x_n^*$.

Суммируя по всем координатам, получаем:

$$\|\nabla F_n(x) - \nabla F_n(y)\|_1 \leqslant 1{,}35M'\sum_{i=1}^m |x_i - y_i|^{0,35} = M\|x - y\|_{0,35}^{0,35},$$

где $M = 1{,}35M'$.

Заметим, что $M'$ можно оценить явно. $M'$ ограничена сверху на отрезке $s \in [\varepsilon, R]$ при $\varepsilon > 0$. Например, если предполагается, что $|x_i - x_n^*| \geqslant 1$, то

$$M' = \max_{s\in[1,R]} 0{,}35 s^{-0,65} = 0{,}35, \quad \Rightarrow \quad M = 1{,}350{,}35 \approx 0{,}4725.$$

(C) Локальное условие Лебега.

Фиксируем точку $x$ и рассмотрим ее окрестность $U_x = B(x, \varepsilon)$ для некоторого $\varepsilon > 0$. Тогда для любой точки $x' \in U_x$:

$$\|\nabla F_n(x', w)\|_2^2 = 1{,}35^2 \sum_{i=1}^{m} |x'_i - x^*_n|^{0,7} \leqslant 1{,}35^2 m R^{0,7},$$

где $R = \sup_{x' \in U_x} \max_i |x'_i - x^*_n| < \infty$ – конечен по построению окрестности $U_x$.

Тогда можно положить $\Phi_x(w) = 1{,}35\sqrt{m}R^{0,35}$, которая не зависит от $w$ и, следовательно, $\mathbb{E}\Phi_x(w) = \Phi_x(w) < \infty$. Условие (C) выполнено.

(D) Ограниченность скорости дрейфа точки минимума.

(D-a) *Ограниченность изменения минимума.*

Так как $\theta_n = x^*_n \mathbf{1}$, а $x^*_n = x^*_{n-1} + \delta$, имеем $\|\theta_n - \theta_{n-1}\|_1 = \|\delta \mathbf{1}\|_1 = m\delta$. Следовательно, условие (D-a) выполнено при $A = m\delta$.

(D-b) *Ограниченность изменения градиента.*

Пусть $r_i = x_i - x^*_{n-1}$, тогда:

$$|\partial_i F_n(x) - \partial_i F_{n-1}(x)| \leqslant 1{,}35 M' |\delta|^{0,35},$$

где $M'$ – гельдеровская константа функции $s^{0,35}$ на допустимом компакте.

Суммируя по $i$, получаем:

$$\|\nabla_x F_n(x) - \nabla_x F_{n-1}(x)\|_1 \leqslant 1{,}35 M' m |\delta|^{0,35}.$$

Обозначим $R = \inf_{x \neq \theta_{n-1}} \|x - \theta_{n-1}\|_1 > 0$, тогда $\|x - \theta_{n-1}\|_1^\rho \geqslant R^\rho$, и условие (D-b) выполнено при

$$B = \frac{1{,}35 M' m \delta^{\,0,35}}{R^{0,35}}.$$

(D-c) *Ограниченность градиента в точке минимума.*

Так как $\theta_n = x^*_n \mathbf{1}$, получаем $\nabla_x F_n(\theta_n) = \mathbf{0}$, следовательно,

$$\|\nabla_x F_n(\theta_n, w_n)\|_{\rho+1}^{\rho+1} = 0.$$

Таким образом, условие выполнено при $C = 0$.

(D-d) *Ограниченность изменения функции.*

$$|F_n(x) - F_{n-1}(x)| \leqslant \sum_{i=1}^{m} \left| (x_i - x^*_n)^{1,35} - (x_i - x^*_{n-1})^{1,35} \right| \leqslant m M'' \delta^{1,35},$$

где $M''$ – гельдеровская константа.

Шум $v_n$ подчиняется распределению Парето с параметром $\beta = 1{,}6 > 1{,}35$, следовательно:

$$\mathbb{E}|v_n - v_{n-1}|^{1,35} \leqslant \tilde{E} < \infty.$$

Таким образом, условие (D-d) выполнено при $D = 0$ и

$$E = (mM''.\delta^{1,35} + \tilde{E})^{1,35}.$$

(E) Ограниченность изменения наблюдаемого шума.

Рассмотрим помехи наблюдений $v_n$, определяемые через шум Парето:

$$v_n = \begin{cases} Z_n \text{sign}_n, & \text{симметричный шум,} \\ Z_n, & \text{асимметричный шум,} \end{cases}$$

где $Z_n \sim \text{Pareto}(\beta = 1{,}6, \sigma = 2{,}0)$, а $\text{sign}_n \sim \text{Uniform}\{-1, 1\}$.

Условие (E) требует выполнения неравенства:

$$\mathbb{E}_{\mathcal{F}_{2n-2}} |v_{2n} - v_{2n-1}|^{\rho+1} \leqslant \sigma_v^{\rho+1},$$

где $\rho + 1 = 1{,}5 < \beta$, т.е. момент порядка 1,5 существует.

Поскольку $v_{2n}$ и $v_{2n-1}$ независимы, разность $v_{2n} - v_{2n-1}$ также является случайной величиной с конечным моментом порядка $\rho + 1$. Для симметричного случая (с переменными знаками) численное моделирование на $10^6$ реализациях дает $\mathbb{E}|v_{2n} - v_{2n-1}|^{1,5} \approx 53{,}73$, что позволяет положить $\sigma_v^{1,5} = 53{,}73$. Следовательно, условие (E) выполнено с явно определенной константой $\sigma_v^{\rho+1} = 53{,}73$.

(F) Независимость возмущений $\Delta_n$.

По построению поискового алгоритма и экспериментов в RLHF-модели, векторы $\Delta_n$ генерируются независимо от всех внешних факторов. Шум $v_n$ добавляется постфактум и не зависит от выбранного направления возмущения.

### *5.5. Метрики оценки и результаты моделирования*

Для количественной оценки поведения алгоритма в условиях дрейфа оптимума и воздействия шума с тяжелыми хвостами используется система эмпирических метрик, отражающих как точность и устойчивость оценки, так и динамику адаптации к изменяющимся условиям. Метрики подбираются таким образом, чтобы охватывать как установившиеся характеристики алгоритма, так и его поведение на всем протяжении оптимизации. Это позволяет выявить сильные и слабые стороны метода в различных сценариях: от стационарных до резко меняющихся и зашумленных.

Оценка средней точности отслеживания дрейфующего параметра на поздних стадиях работы производится через среднюю абсолютную ошибку по последним итерациям. Стабильность поведения алгоритма при этом определяется по стандартному отклонению этих ошибок. Диапазон колебаний в рамках одного запуска характеризуется через средние минимальные и максимальные ошибки по запускам, что позволяет оценить как достижимый потенциал, так и наихудшие случаи.

Динамические характеристики алгоритма отражаются в метрике среднего времени достижения заданного уровня точности – это дает представление о скорости адаптации при ограничениях на ошибку. Связь с теоретическими определениями устойчивости обеспечивается через два момента ошибки: момент порядка, оценивающий сходимость в среднем, и соответствующую ему асимптотическую границу, которая нормирует ошибку согласно выбранному порядку момента. Используемый порядок выбирается в зависимости от параметров шума, чтобы обеспечить существование соответствующих математических ожиданий.

Формулы метрик представлены в табл. 2, сравнение результатов работы всех метрик представлены в табл. 3, а их динамика – на графиках 1–2.

**Таблица 2**. Основные метрики алгоритма

| Метрика | Формула |
|---|---|
| Среднее абсолютное отклонение за последние 100 итераций | $\mu_{\text{last100}} = \frac{1}{100.m} \sum_{n=N-100}^{N-1} \sum_{i=1}^{m} \lvert x_{n,i} - x_n^* \rvert$ |
| Стандартное отклонение ошибки за последние 100 итераций | $\sigma_{\text{last100}} = \sqrt{\frac{1}{100.m-1} \sum_{n=N-100}^{N-1} \sum_{i=1}^{m} \left( \lvert x_{n,i} - x_n^* \rvert - \mu_{\text{last100}} \right)^2}$ |
| Среднее минимальное отклонение по запускам | $\bar{D}_{\min} = \frac{1}{m} \sum_{i=1}^{m} \min_{0 \le n < N} \lvert x_{n,i} - x_n^* \rvert$ |
| Среднее максимальное отклонение по запускам | $\bar{D}_{\max} = \frac{1}{m} \sum_{i=1}^{m} \max_{0 \le n < N} \lvert x_{n,i} - x_n^* \rvert$ |
| Среднее время сходимости до порога $\epsilon$ | $\bar{T}_\epsilon = \frac{1}{m} \sum_{i=1}^{m} T_{i,\epsilon}$, $T_{i,\epsilon} = \min\{\{n \mid 0 \le n < N, \lvert x_{n,i} - x_n^* \rvert < \epsilon\} \cup \{N\}\}$ |
| $l_{\rho+1}$ метрика оценки ошибки | $\mu_{\text{def2,last100}} = \frac{1}{100} \sum_{n=N-100}^{N-1} \frac{1}{m} \sum_{i=1}^{m} \left( \lvert x_{n,i} - x_n^* \rvert^{\rho+1} \right)^{1/2}$ |

**Таблица 3**. Сравнение результатов для разных условий дрейфа и шума

| Метрика | Умеренный дрейф $\delta = 0{,}01$ (symm) | Почти стационарный $\delta = 0{,}0001$ (symm) | Несимметричный шум $\delta = 0{,}0001$ (asymm) |
|---|---|---|---|
| $l_{\rho+1}$ метрика оценки ошибки | 0,4219 | 0,1954 | 0,1206 |
| Среднее расстояние $E[\lvert x - x^* \rvert]$ | 0,3012 | 0,0520 | 0,0356 |
| Ст. отклонение оценки | 0,1682 | 0,2776 | 0,0989 |
| Минимальное отклонение | 0,0002 | 0,0000 | 0,0000 |
| Максимальное отклонение | 19,7897 | 57,1922 | 10,3462 |
| Время сходимости ($<$1,0) | 20 | 18 | 18 |

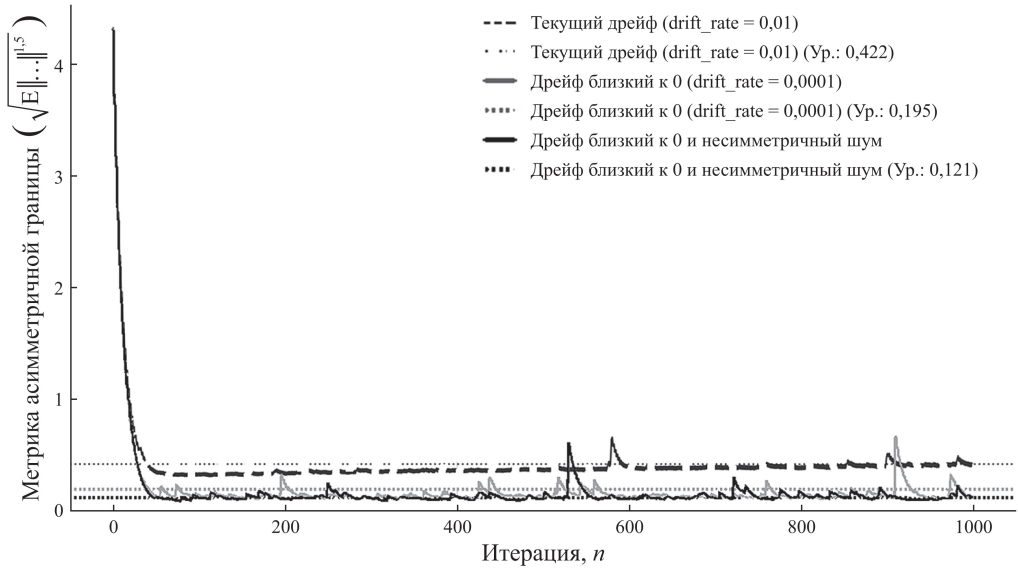


Рис. 1. График зависимости $l_{\rho+1}$ оценки ошибки от номера итерации $n$.

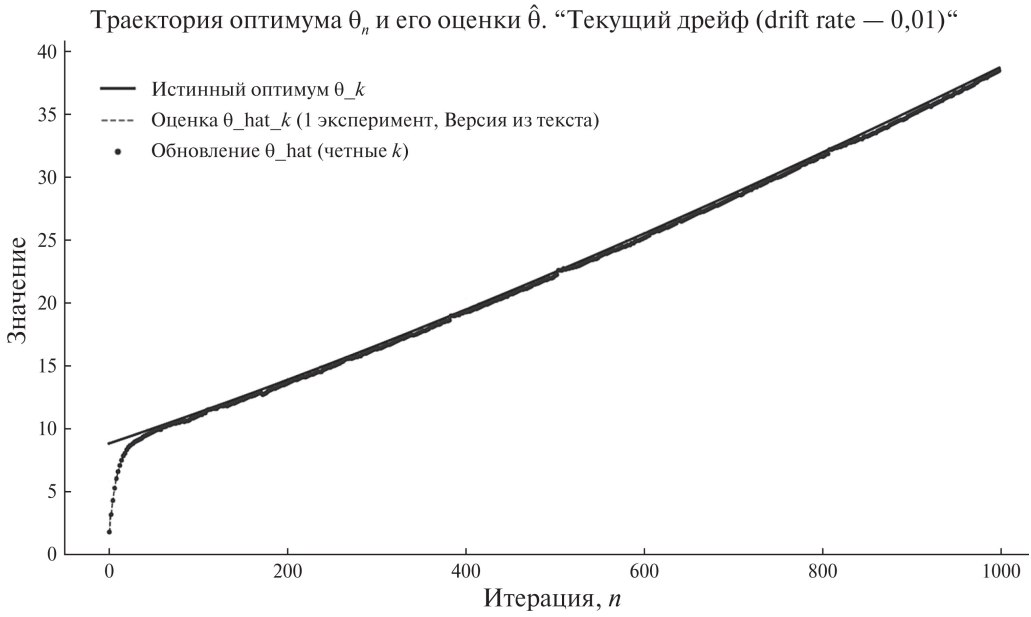


Рис. 2. График истинной траектории оптимума $x_n^*$ и траектории оценки $x_{n,i_0}$.

Результаты моделирования (табл. 3) демонстрируют влияние параметров среды на поведение алгоритма (на основе 1000 экспериментов, $\rho = 0{,}50$, статистика по последним 100 итерациям). При умеренном дрейфе ($\delta = 0{,}01$) и симметричном шуме агент отслеживает цель, однако с заметной средней ошибкой (0,3012), умеренной устойчивостью (Std = 0,1682) и редкими, но значи-

тельными выбросами (максимум – 19,79). Метрики высокого порядка составляют 0,1788 и 0,4219, сходимость достигается за 20 итераций.

Снижение дрейфа до $\delta = 0{,}0001$ (почти стационарная среда) улучшает среднюю ошибку (0,0520), но одновременно увеличивает нестабильность: стандартное отклонение достигает 0,2776, а максимальная ошибка – 57,19. Это указывает на рост чувствительности к шуму с тяжелыми хвостами при ослабленном дрейфе.

Наилучшие результаты достигнуты при асимметричном шуме в условиях слабого дрейфа. Ошибка снижается до 0,0356, вариативность ограничена ($\mathrm{Std} = 0{,}0989$), а максимальные отклонения существенно ниже (10,35). Метрики стабильности (0,0153, 0,1206) и время сходимости (18 итераций) также улучшаются.

Таким образом, снижение дрейфа повышает точность, но устойчивость к шуму зависит от его характера. Асимметричный шум демонстрирует лучший контроль над экстремальными ошибками, вероятно, за счет специфики градиентной оценки. Эффект требует дополнительного анализа.

## 6. Моделирование системы распределения задач в задачах массового обслуживания

Системы массового обслуживания (СМО), такие как современные колл-центры, характеризуются входящим потоком задач, время выполнения которых часто подчиняется распределениям с «тяжелыми хвостами» [31]. Это означает наличие статистически значимой доли задач, требующих несоразмерно большого времени на обработку, что отличает их от систем, описываемых классическими экспоненциальными или нормальными распределениями. Распределение Парето является подходящей моделью для описания таких явлений [32], позволяя учесть влияние редких, но длительных операций на общую производительность системы [33].

Для эффективного управления подобными СМО необходимо адаптивно оценивать характеристики потока и времени обслуживания. Далее описывается пример применения поискового алгоритма (3) стохастической оптимизации для модели динамической подстройки оценок ожидаемого времени обслуживания различных типов задач, подробно описанной в [34]. Метод применяется для итеративной оптимизации параметров $\hat{\theta}_k, \hat{\theta}_m$, представляющих собой адаптивные оценки времени обслуживания для каждого кластера задач $m$ и для системы $(k)$ в целом.

В исследовании [34] представлена симуляционная модель колл-центра. Время обслуживания задач в модели генерируется из распределения Парето, параметры которого для каждого кластера калибруются на основе характеристик логнормальных распределений, аппроксимирующих исторические данные. Поисковый алгоритм (3) используется для уточнения оценок $\hat{\theta}_k$, $\hat{\theta}_m$, которые, в свою очередь, применяются в механизме назначения поступающих задач агентам. Проведенное моделирование показывает работоспособ-

ность метода для рассмотренной задачи в условиях стохастичности и тяжелохвостого характера времени обработки задач.

### *6.1. Модель*

Рассматривается система агентов с одинаковыми ресурсами и производительностью. Нагрузка агента $i$, обозначаемая как $q^i$, соответствует числу задач в его очереди. Каждая задача $x_k$ характеризуется типом $m$ и предсказанным временем выполнения, вычисляемым по формуле:

$$x_{km} = \alpha\hat{\theta}^i_k + (1-\alpha)\hat{\theta}^i_m, \quad \alpha = \frac{\chi|\lambda_m|}{N_m+1},$$
$$\lambda_m(\hat{\theta}_m) = \frac{1}{N_m}\sum_{k\in N_m}\omega_k\frac{\hat{\theta}_m - t_{km}}{\hat{\theta}_m} \to \min,$$

где $\hat{\theta}^i_k$ – индивидуальный прогноз агента $i$ для задачи $k$, $\hat{\theta}^i_m$ – среднее предсказанное время выполнения задач типа $m$ (с учетом локальной истории), $\alpha$ – весовой коэффициент, определяющий вклад индивидуального прогноза и агрегированной статистики, а $\chi$ – коэффициент сходимости. Величина $\lambda_m$ отражает точность предсказания модели для задач типа $m$ и корректируется при поступлении новых наблюдений; здесь $N_m$ – число завершенных задач типа $m$, $\omega_k$ – вес соответствующей ошибки, а $t_{km}$ – фактическое время выполнения $k$-й задачи типа $m$.

Такой механизм расчета предсказания и точности позволяет адаптировать модель к текущему качеству прогнозов, снижая влияние недостоверных данных и усиливая вклад накопленной статистики при высокой уверенности.

На каждом шаге $k$, при поступлении новой задачи $x_k$, проводится назначение $x_k$ такому агенту $i_k$ для балансировки нагрузки агентов:

$$i_k = \arg\min_i \sum_j \left| \frac{q^i_k + x_{km} - q^j_k}{d_{ij}+1} \right|, \tag{5}$$

где $q^j$ – нагрузка агента $j$, $d_{ij}$ – «расстояние» между агентами (например, на основе нагрузки или физического расположения). Агенты соединены в полностью связную топологию, где каждый агент взаимодействует со всеми остальными. Это обеспечивает глобальную коммуникацию при варьирующемся влиянии агентов в зависимости от их относительной близости.

### *6.2. Описание набора данных и первичный анализ*

Для демонстрации эффективности разработанного метода было проведено моделирование системы распределения нагрузки на основе реальных данных операторского колл-центра за сентябрь 2023 г. (более 2,3 млн вызовов). Для каждого обращения регистрировались момент поступления, время ожидания

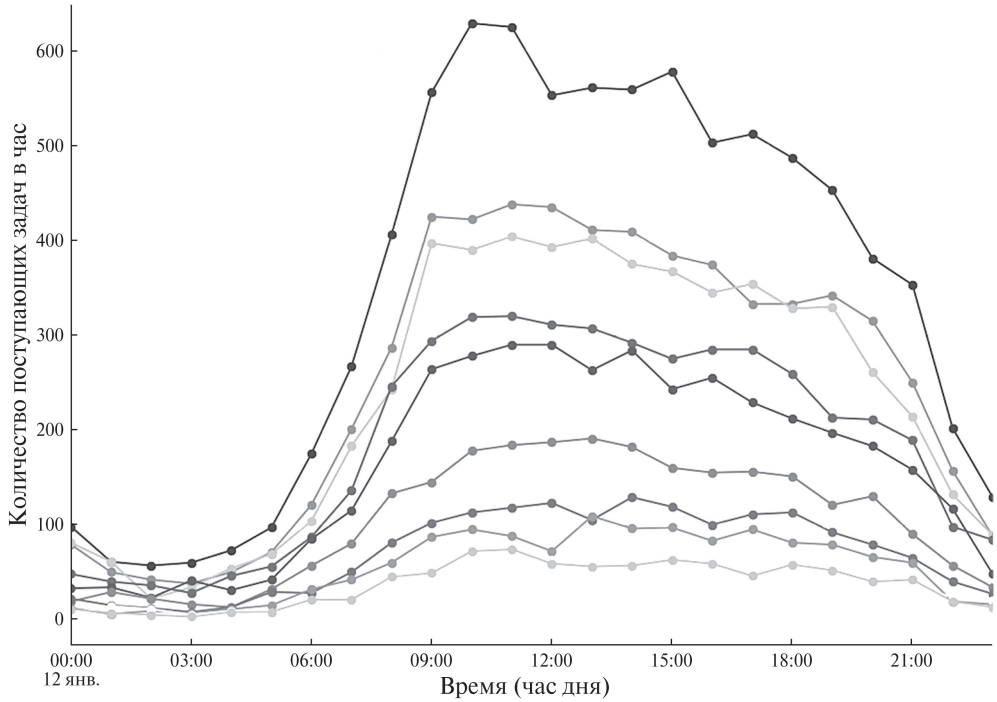


Рис. 3. Пример почасовой интенсивности различных задач, топ 10 кластеров.

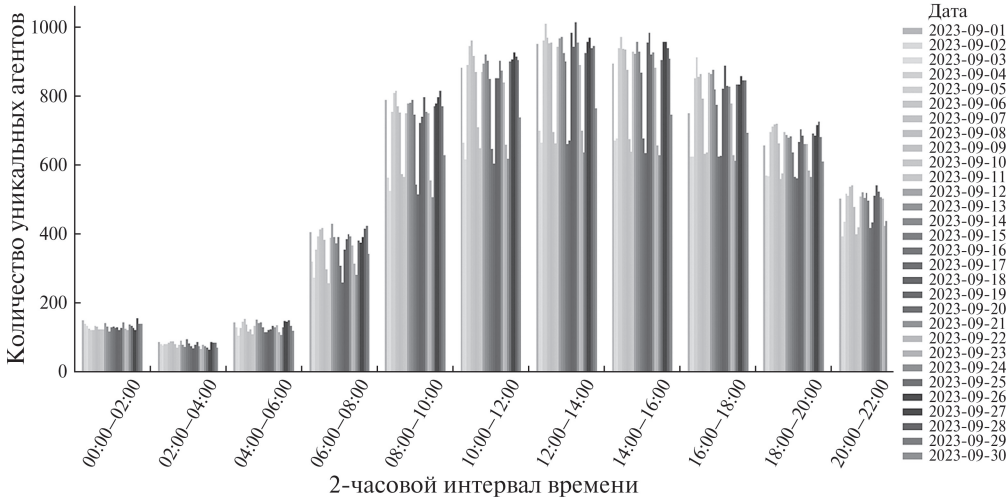


Рис. 4. Число агентов, работающих в одно время, по 2-часовым интервалам.

ответа, фактическая продолжительность разговора (ACD Time) и сегмент клиента.

На рис. 3–4 представлены две взаимодополняющие визуализации, раскрывающие ключевые характеристики входящего потока некоторых кластеров и кадрового потенциала колл-центра. График 3 демонстрирует почасовую ин-

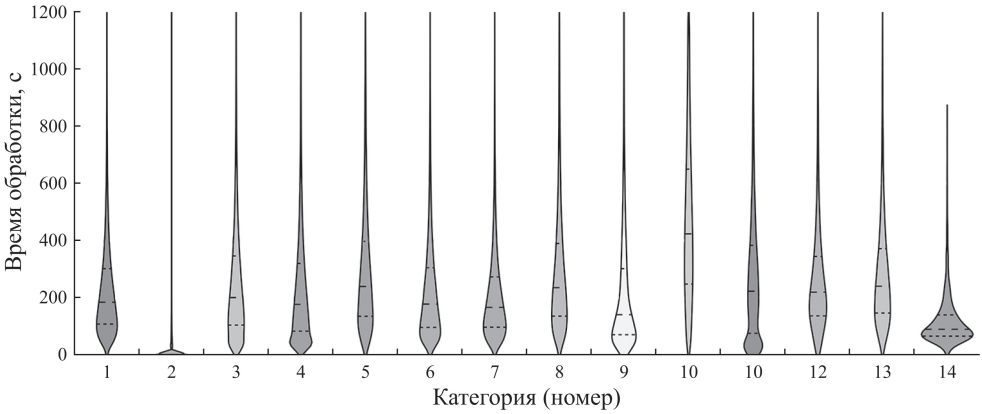


Рис. 5. Длительность разговоров по топ-14 кластерам (макс. ACD = 1200 с).

тенсивность задач по десяти крупнейшим кластерам, при этом пик нагрузки наблюдается в кластере «Young» в интервале 11–12 ч. График 4 показывает распределение активных операторов (принявших более 50 вызовов за двухчасовой интервал), максимальные значения приходятся на период с 8 до 16 ч. При этом кадровые ресурсы не всегда успевают за резкими колебаниями входящего трафика. Симуляционные задержки, агрегированные по времени суток, в целом воспроизводят динамику реального ожидания, включая утренний рост около 8–9 ч и вечерний пик после 17 ч.

Диаграммы на рис. 5 отображают распределение длительности разговоров для 14 клиентских сегментов. В целях обезличивания все сегменты были переименованы в числовые идентификаторы от 1 до 14 (см. табл. 4). Наибольшую вариативность и протяженные хвосты распределения наблюдаются у сегментов 11 и 13, тогда как сегмент 2 характеризуется исключительно коротким диапазоном длительности обработки. Сегменты с номерами 3 и 14 также демонстрируют относительно узкое распределение с короткими медианами.

### *6.3. Результаты моделирования*

Для оценки качества предложенного метода была проведена симуляция работы колл-центра на реальных данных. Результаты позволили оценить как динамику времени ожидания задач в течение суток, так и стабильность распределения нагрузки. На рис. 6–7 приведены результаты одной симуляционной сессии.

График 6 демонстрирует среднее время ожидания задач в 20-минутных интервалах, отражая характерный пик в дневные часы, связанный с высокой нагрузкой. Модель эффективно адаптируется к изменяющимся условиям: после резкого роста задержек около 12 : 00 среднее время ожидания быстро снижается за счет перераспределения задач.

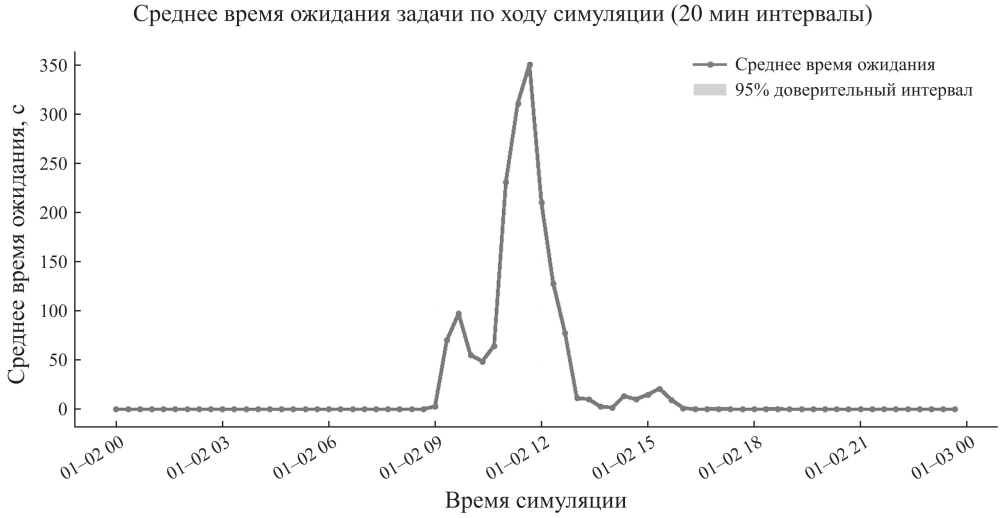


Рис. 6. Среднее время ожидания (20-мин интервалы).

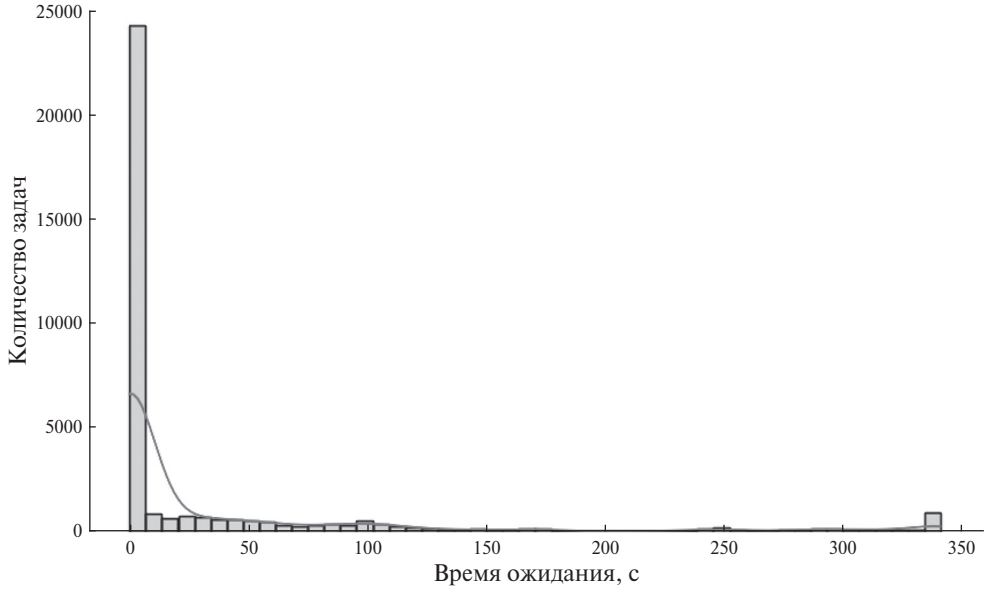


Рис. 7. Распределение времени ожидания (по 98-му перцентилю).

На гистограмме 7 представлено распределение времени ожидания задач, усеченное по 98-му перцентилю. Большинство задач были обслужены менее чем за 50 с, что соответствует целевым SLA-показателям для типичных сценариев.

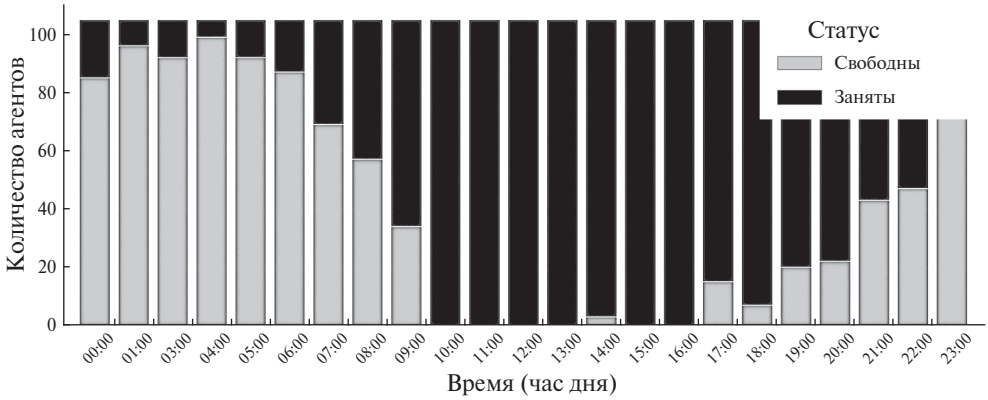


Рис. 8. Почасовая загрузка агентов: сравнение свободных и занятых ресурсов.

Для ключевых кластеров в табл. 4 представлены значения предсказанного времени выполнения задач $z$, количества завершенных обращений $k$, среднего фактического времени $t_{avg}$ и максимальной длительности $t_{\max}$. Кластеры с номерами от 1 до 14 соответствуют тем, что отображены на рис. 5, а кластер с номером 0 объединяет все прочие сегменты, не вошедшие в топ-14. Как видно, величины $z$ хорошо соответствуют эмпирическим средним, несмотря на разную статистику по каждому кластеру. Это подтверждает устойчивость адаптивного предсказания на основе поискового алгоритма.

**Таблица 4**. Параметры категорий

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| $z$ | 143,76 | 163,14 | 3,73 | 174,75 | 147,62 | 196,05 | 159,75 | 151,75 |
| $k$ | 36 858 | 8180 | 6166 | 5764 | 4461 | 3857 | 2523 | 1707 |
| $t_{avg}$ | 124,34 | 158,94 | 3,71 | 163,51 | 135,06 | 174,91 | 161,16 | 157,93 |
| $t_{\max}$ | 200 | 200 | 200 | 200 | 200 | 200 | 200 | 200 |

| | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|---|---|---|---|---|---|---|---|
| $z$ | 219,63 | 151,74 | 321,53 | 144,92 | 147,55 | 191,31 | 49,54 |
| $k$ | 1329 | 943 | 906 | 484 | 265 | 223 | 44 |
| $t_{avg}$ | 233,06 | 153,30 | 330,74 | 158,08 | 156,68 | 198,90 | 87,34 |
| $t_{\max}$ | 200 | 200 | 200 | 200 | 200 | 200 | 44 |

График на рис. 8 иллюстрирует почасовую загрузку операторов в ходе симуляции. В ночное и утреннее время (до 08:00) значительная доля агентов остается свободной, однако начиная с 09:00 и до 15:00 наблюдается полная загрузка всех ресурсов: количество свободных агентов опускается до нуля. Это совпадает с пиком входящего потока задач и подчеркивает необходимость точного предсказания длительности обработки. Вечером и ночью нагрузка постепенно снижается, а система возвращается к сбалансированному состоянию.

Таким образом, модель демонстрирует способность корректно адаптироваться к нагрузке, обеспечивая сдерживание времени ожидания и равномерное распределение задач в течение суток. Предложенный подход позволяет эффективно использовать ресурсы в условиях высокой вариативности обращений и может быть рекомендован для внедрения в распределенные системы поддержки с интенсивной и нерегулярной нагрузкой.

## 7. Заключение

В представленной работе предложен и исследован метод оценивания минимума функционала, изменяющегося во времени, в условиях, когда измерения подвержены помехам. Этот метод, основанный на псевдоградиентном подходе с рандомизацией, не требует знания градиента целевой функции и использует небольшое число измерений на каждой итерации. Сделано предположение об ограниченности скорости изменения (дрейфа) экстремума функционала. Доказано, что асимптотическая ошибка оценивания ограничена величиной $\frac{L}{K}$, где $L$ и $K$ определяются свойствами целевой функции, характеристиками шумов и параметрами алгоритма. Справедливость теоретических выводов была подтверждена результатами численного моделирования. Моделирование адаптации RLHF-агента к зашумленной и динамической обратной связи, в частности, с использованием шума с тяжелыми хвостами и различными скоростями дрейфа предпочтений, продемонстрировало, что поисковый алгоритм обеспечивает сходимость оценки к области целевого значения. Наблюдаемая в моделировании установившаяся ошибка и колебания оценки, обусловленные шумом и дрейфом, согласуются с теоретическими предсказаниями об ограниченности асимптотической ошибки. Кроме того, предложенный метод был протестирован на моделировании, основанном на реальных данных операторского колл-центра. Использование эмпирических характеристик потока обращений и времени обработки задач позволило продемонстрировать применимость алгоритма в задачах динамического распределения нагрузки и предсказания параметров обслуживания в реальных сервисных системах.

## *ПРИЛОЖЕНИЕ*

*Доказательство теоремы 1.*
Обозначим ошибку оценивания $\mathrm{err}_n = \hat{\theta}_n - \theta_n$.

*Шаг 1. Рекуррентное соотношение для ошибки оценивания*

В силу алгоритма (3) имеем

$$\hat{\theta}_{2n} = \hat{\theta}_{2n-2} - \frac{\alpha}{2\beta}\Delta_{2n}(y_{2n} - y_{2n-1}).$$

Следовательно,

$$\mathrm{err}_{2n} = \mathrm{err}_{2n-2} - \underbrace{(\theta_{2n} - \theta_{2n-2})}_{\mathrm{dreif}_n} - \underbrace{\frac{\alpha}{2\beta}\Delta_{2n}(y_{2n} - y_{2n-1})}_{\mathrm{step}_n}.$$

*Шаг 2. Рекуррентное соотношение для оценки функции Ляпунова $V(x)$*

Для векторов $a = \hat{\theta}_{2n-2}$ и $b = \mathrm{dreif}_n + \mathrm{step}_n$ по определению имеем

$$V_{2n}(\hat{\theta}_{2n}) = V_{2n-2}(\hat{\theta}_{2n} - \mathrm{dreif}_n) = V_{2n-2}(a - b) = \|a - b - \theta_{2n-2}\|_{\rho+1}^{\rho+1}.$$

Для функции $V_{2n-2}(a-b)$ применим разложение Тейлора около точки $a$ в направлении $-b$:

$$V_{2n-2}(a-b) = V_{2n-2}(a) - \langle \nabla V_{2n-2}(a - \delta b), b\rangle, \quad \delta \in [0,1], \tag{П.1}$$

где градиент $\nabla V_{2n-2}(a-\delta b)$ вычисляется по формуле:

$$\nabla V_{2n-2}(a-\delta b) = (\rho+1)\mathrm{sign}(\delta) \odot |a - \theta_{2n-2} - \delta b|^{\rho},$$

где $\mathrm{sign}_n^{(i)}(\delta) = 0$ или $\pm 1$ в зависимости от знака выражения $i$-й компоненты вектора $a - \theta_{2n-2} - \delta b$, $|a - \theta_{2n-2} - \delta b|^{\rho}$ – вектор из абсолютных величин компонент вектора $a - \theta_{2n-2} - \delta b$ в степени $\rho$, $\odot$ – символ покомпонентного умножения. Второй член в (П.1) можно оценить следующим образом:

$$\begin{gathered} -\langle \nabla V_{2n-2}(a-\delta b), b\rangle \leqslant -\langle (\rho+1)\mathrm{sign}(0) \odot |a - \theta_{2n-2}|^{\rho}, b\rangle + 2^{1-\rho}\delta^{\rho}\|b\|_{\rho+1}^{\rho+1} \leqslant \\ \leqslant -\langle \nabla V_{2n-2}(a), b\rangle + 2^{1-\rho}\|b\|_{\rho+1}^{\rho+1} \end{gathered}$$

(см. доказательство теоремы 1 в [24], с. 93).

Учитывая вышеизложенное и условие (D.a):

$$\begin{gathered} V_{2n}(\hat{\theta}_{2n}) \leqslant V_{2n-2}(\hat{\theta}_{2n-2}) - \langle \nabla V_{2n-2}(\hat{\theta}_{2n-2}), \mathrm{dreif}_n + \mathrm{step}_n\rangle + \\ +2(A^{\rho+1} + \|\mathrm{step}_n\|_{\rho+1}^{\rho+1}). \end{gathered} \tag{П.2}$$

*Шаг 3. Разложение корректирующего шага*

Из модели наблюдений разложим $\mathrm{step}_n$ на два слагаемых

$$\mathrm{step}_n = \underbrace{\frac{\alpha}{2\beta}\Delta_n(F_{2n}(x_{2n}, w_{2n}) - F_{2n-1}(x_{2n-1}, w_{2n-1}))}_{\text{Почти псевдоградиентный член}} + \underbrace{\frac{\alpha}{2\beta}\Delta_n(v_{2n} - v_{2n-1})}_{\text{Шум}}.$$

**а. Почти псевдоградиентный член.**

Обозначим $n^{\pm} = 2n - \frac{1}{2} \pm \frac{1}{2}$.

Воспользовавшись формулой Тейлора, добавив и вычтя сначала $\sum_{n^\pm}\langle\nabla_x F_{n^\pm}(\hat\theta_{2n-2}, w_{n^\pm}), \beta\Delta_n\rangle$, а потом $\langle\nabla_x F_{2n-2}(\hat\theta_{2n-2}, w_{n^\pm}), \beta\Delta_n\rangle$ и еще раз $\langle\nabla_x F_{2n-2}(\theta_{2n-2}, w_{n^\pm}), \beta\Delta_n\rangle$, выводим

$$\sum_{n^\pm} \pm F_{n^\pm}(x_{n^\pm}, w_{n^\pm}) =$$

$$= \sum_{n^\pm} \pm F_{n^\pm}(\hat\theta_{2n-2}, w_{n^\pm}) + \langle\nabla_x F_{n^\pm}(\hat\theta_{2n-2} \pm \delta_{n^\pm}\beta\Delta_n, w_{n^\pm}), \beta\Delta_n\rangle =$$

$$= \sum_{n^\pm} \pm F_{n^\pm}(\hat\theta_{2n-2}, w_{n^\pm}) + \langle\nabla_x F_{n^\pm}(\hat\theta_{2n-2}, w_{n^\pm}), \beta\Delta_n\rangle +$$

$$+\langle\nabla_x F_{n^\pm}(\hat\theta_{2n-2} \pm \delta_{n^\pm}\beta\Delta_n, w_{n^\pm}), \beta\Delta_n\rangle - \langle\nabla_x F_{n^\pm}(\hat\theta_{2n-2}, w_{n^\pm}), \beta\Delta_n\rangle =$$

$$= \sum_{n^\pm} \pm F_{n^\pm}(\hat\theta_{2n-2}, w_{n^\pm}) + \langle\nabla_x F_{2n-2}(\theta_{2n-2}, w_{n^\pm}), \beta\Delta_n\rangle +$$

$$+\langle\nabla_x F_{2n-2}(\hat\theta_{2n-2}, w_{n^\pm}), \beta\Delta_n\rangle - \langle\nabla_x F_{2n-2}(\theta_{2n-2}, w_{n^\pm}), \beta\Delta_n\rangle +$$

$$+\langle\nabla_x F_{n^\pm}(\hat\theta_{2n-2}, w_{n^\pm}), \beta\Delta_n\rangle - \langle\nabla_x F_{2n-2}(\hat\theta_{2n-2}, w_{n^\pm}), \beta\Delta_n\rangle +$$

$$+\langle\nabla_x F_{n^\pm}(\hat\theta_{2n-2} \pm \delta_{n^\pm}\beta\Delta_n, w_{n^\pm}), \beta\Delta_n\rangle - \langle\nabla_x F_{n^\pm}(\hat\theta_{2n-2}, w_{n^\pm}), \beta\Delta_n\rangle,$$

где $\delta_{n^\pm} \in [0, 1]$.

Применим операцию условного математического ожидания относительно $\sigma$-алгебры $\mathcal{F}_{2n-2}$. Учитывая независимость в силу условия (F) векторов $\Delta_n$ от $w_{n^\pm}$ и $\sigma$-алгебры $\mathcal{F}_{2n-2}$, получаем

$$\frac{\alpha}{2\beta}\mathbb{E}_{\mathcal{F}_{2n-2}}\left\{\Delta_n \sum_{n^\pm} \pm F_{n^\pm}(\hat\theta_{2n-2}, w_{n^\pm})\right\} = 0$$

с силу центрированности $\Delta_n$ и

$$\frac{\alpha}{2\beta}\mathbb{E}_{\mathcal{F}_{2n-2}}\left\{\Delta_n \sum_{n^\pm}\langle\nabla_x F_{2n-2}(\theta_{2n-2}, w_{n^\pm}), \beta\Delta_n\rangle\right\} = 0,$$

так как в силу условия (C) имеем $\mathbb{E}_{\mathcal{F}_{2n-2}}\{\nabla_x F_{2n-2}(\theta_{2n-2}, w_{n^\pm})\} = \nabla_x f_{2n-2}(\theta_{2n-1})$ и градиент функции $f_{2n-2}(\cdot)$ в точке минимума $\theta_{2n-2}$ равен нулю.

В итоге в силу условия (C) выводим для почти псевдоградиентного члена

$$\mathbb{E}_{\mathcal{F}_{2n-2}}\left\{\frac{\alpha}{2\beta}\Delta_n \sum_{n^\pm} \pm F_{n^\pm}(x_{n^\pm}, w_{n^\pm})\right\} = \frac{\alpha}{d}\nabla f_{2n}(\hat\theta_{2n-2}) + \frac{\alpha}{2\beta}\mathbb{E}_{\mathcal{F}_{2n-2}}\mathrm{corr}_n,$$

где

$$\mathrm{corr}_n = \sum_{n^\pm}\langle\nabla_x F_{n^\pm}(\hat\theta_{2n-2} \pm \delta_{n^\pm}\beta\Delta_n, w_{n^\pm}), \beta\Delta_n\rangle - \langle\nabla_x F_{n^\pm}(\hat\theta_{2n-2}, w_{n^\pm}), \beta\Delta_n\rangle +$$

$$+\langle\nabla_x F_{2n-2}(\hat\theta_{2n-2}, w_{n^\pm}), \beta\Delta_n\rangle - \langle\nabla_x F_{2n-2}(\theta_{2n-2}, w_{n^\pm}), \beta\Delta_n\rangle +$$

$$+\langle\nabla_x F_{n^\pm}(\hat\theta_{2n-2}, w_{n^\pm}), \beta\Delta_n\rangle - \langle\nabla_x F_{2n-2}(\hat\theta_{2n-2}, w_{n^\pm}), \beta\Delta_n\rangle.$$

В силу условий (B) и (D.b) имеем оценку

$$\|\mathrm{corr}_n\| \leqslant M\beta^\rho\|\Delta_n\|\left(2\|\Delta_n\|^\rho + 2\|\hat\theta_{2n-2}-\theta_{2n-2}\|^\rho) + 3B\|\hat\theta_{2n-2}-\theta_{2n-2}\|^\rho\right) = \\ = 2M\beta^\rho + (2+3B)\|\hat\theta_{2n-2}-\theta_{2n-2}\|^\rho.$$

**b. Шум.** Применим операцию условного математического ожидания относительно $\sigma$-алгебры $\mathcal{F}_{2n-2}$. Учитывая независимость $\Delta_n$ от $v_{2n}$, $v_{2n-1}$ и $\mathcal{F}_{2n-2}$, получаем

$$\mathbb{E}_{\mathcal{F}_{2n-2}}\left\{\frac{\alpha}{2\beta}\Delta_n(v_{2n}-v_{2n-1})\right\} = 0.$$

**c. Итоговая оценка для второго слагаемого в правой части неравенства** (П.2)**.** В силу условия сильной выпуклости (A) получаем

$$-\mathbb{E}_{\mathcal{F}_{2n-2}}\{\langle\nabla V_{2n-2}(\hat\theta_{2n-2}), \mathrm{dreif}_n + \mathrm{step}_n\rangle\} \leqslant -\frac{\mu\alpha}{d}V_{2n-2}(\hat\theta_{2n-2}) - \\ -\frac{\alpha}{2\beta}\mathbb{E}_{\mathcal{F}_{2n-2}}\langle\nabla V_{2n-2}(\hat\theta_{2n-2}), \mathrm{dreif}_n + \mathrm{corr}_n\rangle \leqslant -\frac{\mu\alpha}{d}V_{2n-2}(\hat\theta_{2n-2}) + \\ +2(A+\alpha M\beta^{\rho-1})^2 + \left(2+\frac{\alpha}{2\beta}(2+3B)\right)\sum_{i=1}^d|\hat\theta^i_{2n-2}-\theta^i_{2n-2}|^{2\rho} \leqslant \\ \leqslant -\frac{\mu\alpha}{d}V_{2n-2}(\hat\theta_{2n-2}) + \varepsilon V_{2n-2}(\hat\theta_{2n-2}) + c_1,$$

где $\varepsilon > 0$ и

$$c_1 = 2(A+\alpha M\beta^{\rho-1})^2 + \varepsilon^{\rho-1}\left(2+\frac{\alpha}{2\beta}(2+3B)\right)^{\frac{1-\rho}{\rho+1}}.$$

*Шаг 4. Оценка третьего слагаемого в правой части неравенства* (П.2)

По аналогии с выкладками на Шаге 4 $\mathrm{step}_n$ можно представить в виде

$$\mathrm{step}_n = \frac{\alpha}{2\beta}\Delta_n\sum_{i=1}^8 a_i,$$

где
$a_1 = \sum_{n^\pm}\pm F_{n^\pm}(\hat\theta_{2n-2}, w_{n^\pm})$,
$a_{2,3} = \langle\nabla_x F_{n^\pm}(\hat\theta_{2n-2}\pm\delta_{n^\pm}\beta\Delta_n, w_{n^\pm}), \beta\Delta_n\rangle - \langle\nabla_x F_{n^\pm}(\hat\theta_{2n-2}, w_{n^\pm}), \beta\Delta_n\rangle$,
$a_{4,5} = \langle\nabla_x F_{n^\pm}(\hat\theta_{2n-2}, w_{n^\pm}), \beta\Delta_n\rangle - \langle\nabla_x F_{n^\pm}(\theta_{n^\pm}, w_{n^\pm}), \beta\Delta_n\rangle$,
$a_{6,7} = \langle\nabla_x F_{n^\pm}(\theta_{n^\pm}, w_{n^\pm}), \beta\Delta_n\rangle$,
$a_8 = v_{2n} - v_{2n-1}$.
Для $a_1$ в силу условия (D.d) имеем $\mathbb{E}_{\mathcal{F}_{2n-2}}|a_1|^{\rho+1} \leqslant DV_{2n-2}(\hat\theta_{2n-2}) + E$,
для $a_2, a_3$ в силу условия (B) имеем $\mathbb{E}_{\mathcal{F}_{2n-2}}|a_i|^{\rho+1} \leqslant M^{\rho+1}\beta^{2\rho+2},\ i = 2,3$,
для $a_4, a_5$ в силу условия (B) и неравенства Йенсена выводим $\mathbb{E}_{\mathcal{F}_{2n-2}}|a_i|^{\rho+1} \leqslant$
$\leqslant (M\beta\|\hat\theta_{2n-2}-\theta_{n^\pm}\|_2^\rho)^{\rho+1} \leqslant M^{\rho+1}\beta^{\rho+1}d^{\frac{\rho-1}{2}}V_{n^\pm}(\hat\theta_{2n-2}),\ i = 4,5$,

для $a_6, a_7$ в силу условия (D.c) имеем $\mathbb{E}_{\mathcal{F}_{2n-2}}|a_i|^{\rho+1} \leqslant C,\ i = 6, 7$,
для $a_8$ в силу условия (E) имеем $\mathbb{E}_{\mathcal{F}_{2n-2}}|a_8|^{\rho+1} \leqslant \sigma_v^{\rho+1}$.
В силу неравенства Йенсена имеем

$$\left(\frac{\sum_{i=1}^{8}|a_i|}{8}\right)^{\rho+1} \leqslant \frac{1}{8}\sum_{i=1}^{8}|a_i|^{\rho+1}$$

и, следовательно, выводим

$$2A^{\rho+1} + 2\mathbb{E}_{\mathcal{F}_{2n-2}}\|\mathrm{step}_n\|_{\rho+1}^{\rho+1} \leqslant 2A^{\rho+1} + 2.8^{\rho}\left(\frac{\alpha}{2\beta}\right)^{\rho+1}\sum_{i=1}^{7}|a_i|^{\rho+1} \leqslant 2A^{\rho+1} +$$
$$+2^{2\rho}\alpha^{\rho+1}\left(2M^{\rho+1}\left(\beta^{\rho+1} + d^{\frac{\rho-1}{2}}\sum_{n^{\pm}}V_{n^{\pm}}(\hat{\theta}_{2n-2})\right) + \right.$$
$$\left. + \frac{2C + DV_{2n-2}(\hat{\theta}_{2n-2}) + E + \sigma_v^{\rho+1}}{\beta^{\rho+1}}\right) \leqslant$$
$$\leqslant c_2\alpha^{\rho+1}V_{2n-2}(\hat{\theta}_{2n-2}) + c_3,$$

где

$$c_2 = 2^{3\rho+1}M^{\rho+1}\left(d^{\frac{\rho-1}{2}} + \frac{D}{\beta^{\rho+1}}\right)$$

и

$$c_3 = 2A^{\rho+1} + 2^{2\rho}\alpha^{\rho+1}\left(2M^{\rho+1}\left(\beta^{\rho+1} + 3{,}2^{\rho}d^{\frac{\rho-1}{2}}\right) + \frac{E + 2C + \sigma_v^{\rho+1}}{\beta^{\rho+1}}\right).$$

*Шаг 5. Формирование рекуррентного неравенства*

Собирая все оценки, получаем

$$V_{2n} \leqslant V_{2n-2} - (\mu\alpha d^{-1} - \varepsilon - c_2\alpha^{\rho+1})V_{2n-2} + c_1 + c_3.$$

Вводя обозначения

$$K = 1 - \mu\alpha d^{-1} + \varepsilon + c_2\alpha^{\rho+1}, \quad L = c_1 + c_3,$$

получаем

$$V_{2n} \leqslant (1 - K)V_{2n-2} + L.$$

Выбрав достаточно малые $\alpha$ и $\varepsilon$, можем обеспечить неравенство $K < 1$, при выполнении которого справедливо заключение теоремы 1. □

## СПИСОК ЛИТЕРАТУРЫ


1. *Поляк Б.Т.* Введение в оптимизацию. М.: Наука, 1983. 384 с.
2. *Поляк Б.Т., Цыпкин Я.З.* Псевдоградиентные алгоритмы адаптации и обучения // АиТ. 1973. № 3. С. 45–68.
3. *Поляк Б.Т., Цыпкин Я.З* Адаптивные алгоритмы оценивания (сходимость, оптимальность, устойчивость) // АиТ. 1979. № 3. С. 71–84.
4. *Поляк Б.Т., Цыпкин Я.З.* Оптимальные псевдоградиентные алгоритмы адаптации // АиТ. 1980. № 8. С. 74–84.
5. *Поляк Б.Т.* О некоторых способах ускорения сходимости итерационных методов // Журн. вычисл. мат. и мат. физики. 1964. V. 4. № 5. С.791–803.
6. *Поляк Б.Т.* Новый метод типа стохастической аппроксимации // АиТ. 1990. № 7. С. 98–108.
7. *Polyak B.T., Yuditskij A.B.* Acceleration of stochastic approximation procedures by averaging // SIAM J. Contr. Optim. 1992. V. 30. No. 4. P. 838–855.
8. *Поляк Б.Т.* Сходимость и скорость сходимости итеративных стохастических алгоритмов. I // АиТ. 1976. № 12. С. 83-94.
9. *Поляк Б.Т.* Сходимость и скорость сходимости итеративных стохастических алгоритмов. II // АиТ. 1977. № 4 С. 101–107.
10. *Поляк Б.Т., Цыбаков А.Б.* Оптимальные порядки точности поисковых алгоритмов стохастической оптимизации // Проблемы передачи информации. 1990. № 26. 2. С. 45–53.
11. *Растригин Л.А.* Статистические методы поиска. М.: Наука, 1968. 376 с.
12. *Граничин О.Н.* Стохастическая аппроксимация с возмущением на входе при зависимых помехах наблюдения // Вестн. ЛГУ. 1989. С. 27–31.
13. *Spall J.C.* Multivariate stochastic approximation using a simultaneous perturbation gradient approximation // IEEE Transact. Autom. Control. 1992. 37(3). С. 332–341.
14. *Spall J.C.* A one-measurement form of simultaneous perturbation stochastic approximation // Automatica. 1997. 33(1). P. 109–112.
15. *Граничин О.Н., Поляк Б.Т.* Рандомизированные алгоритмы оценивания и оптимизации при почти произвольных помехах. М.: Наука, 2003. 291 с.
16. *Granichin O., Volkovich V., Toledano-Kitai D.* Randomized algorithms in automatic control and data mining. Berlin Heldenberg: Springer, 2015. 251 p.
17. *Попков А.Ю.* Градиентные методы для нестационарных задач безусловной оптимизации // АиТ. 2005. № 6. С. 38–46.
18. *Kiefer J., Wolfowitz J.* Stochastic estimation of the maximum of a regression function // The Annals of Mathematical Statistics. 1952. 23(3). P. 462–466.
19. *Вахитов А.Т., Граничин О.Н., Гуревич Л.С.* Алгоритм стохастической аппроксимации с пробным возмущением на входе в нестационарной задаче оптимизации // АиТ. 2009. № 11. С. 70–79.
20. *Granichin O., Amelina N.* Simultaneous perturbation stochastic approximation for tracking under unknown but bounded disturbances // IEEE Transact. Autom. Control. 2015. V. 60. No. 6. P. 1653–1658.
21. *Шибаев И.А.* Безградиентные методы оптимизации для функций с гельдеровым градиентом // Дисс. ... канд. физ.-мат. наук. Долгопрудный: МФТИ, 2024.

22. *Shibaev I., Dvurechensky P., Gasnikov A.* Zeroth-order methods for noisy Holder-gradient functions // Optimization Letters. 2022. V. 16. P. 2123–2143.
23. *Mandelbrot B.* New methods in statistical economics // Journal of Political Economy. 1963. V. 71. No. 5. P. 421–440.
24. *Вахитов А.Т., Граничин О.Н., Сысоев С.С.* Точность оценивания рандомизированного алгоритма стохастической оптимизации // АиТ. 2006. № 4. С. 86–96.
25. *Граничин О.Н.* Поисковые алгоритмы стохастической аппроксимации с рандомизацией на входе // АиТ. 2015. № 5. С. 43–59.
26. *Min T. et al.* Understanding Impact of Human Feedback via Influence Functions. arXiv preprint arXiv:2501.05790. 2025.
27. *Shen W. et al.* Loose lips sink ships: mitigating length bias in reinforcement learning from human feedback // Findings of the Association for Computational Linguistics: EMNLP 2023, 2023. P. 2859–2873.
28. *Christiano P.F. et al.* Deep reinforcement learning from human preferences // Advances in Neural Information Processing Systems. 2017. V. 30. P. 1–9.
29. *Stiennon N. et al.* Learning to summarize with human feedback // Advances in Neural Information Processing Systems. 2020. V. 33. P. 3008–3021.
30. *Ouyang L. et al.* Training language models to follow instructions with human feedback // Advances in Neural Information Processing Systems. 2022. V. 35. P. 27730–27744.
31. *Gans N., Koole G., Mandelbaum A.* Telephone call centers: Tutorial, review, and research prospects // Manufacturing and Service Operations Management. 2003. V. 5. No. 2. P. 79–141.
32. *Anderson. C.* The Long Tail: Why the Future of Business is Selling Less of More, NY.: Hyperion, 2006. 256 p.
33. *Goel S., Broder A., Gabrilovich E., Pang. B.* Anatomy of the long tail: Ordinary People With Extraordinary Tastes // Proceedings of the Third ACM International Conference on Web Search and Data Mining (WSDM'10), ACM, New York, NY, USA, 2010. P. 201–210.
34. *Akinfiev I., Tarasova E.* Cluster-Aware LVP: Enhancing Task Allocation with Growth Dynamics // 15th IFAC Workshop on Adaptive and Learning Control Systems (ALCOS), 2025.

*Статья представлена к публикации членом редколлегии П.С. Щербаковым.*

Поступила в редакцию 23.06.2025

После доработки 30.06.2025

Принята к публикации 04.07.2025