

УДК: 519.85

## Применение градиентных методов оптимизации для решения задачи Коши для уравнения Гельмгольца

Н. В. Плетнев<sup>1,a</sup>, П. Е. Двуреченский<sup>2,b</sup>, А. В. Гасников<sup>1,3,4,c</sup>

<sup>1</sup>Московский физико-технический институт,

Россия, 141701, г. Долгопрудный, Институтский пер., д. 9

<sup>2</sup>Институт прикладного анализа и стохастики им. Вейерштрасса,  
Германия, 10117, г. Берлин, Моренштрассе, д. 39

<sup>3</sup>Институт проблем передачи информации им. А. А. Харкевича Российской академии наук,  
Россия, 127051, г. Москва, Большой Каретный переулок, д. 19, стр. 1

<sup>4</sup>Кавказский математический центр, Адыгейский государственный университет,  
Россия, 385000, Республика Адыгея, г. Майкоп, ул. Первомайская, д. 208

E-mail: <sup>a</sup> nikita.pletnev@phystech.edu, <sup>b</sup> pavel.dvurechensky@wias-berlin.de, <sup>c</sup> gasnikov@yandex.ru

Получено 13.02.2022.

Принято к публикации 13.02.2022.

Статья посвящена изучению применения методов выпуклой оптимизации для решения задачи Коши для уравнения Гельмгольца, которая является некорректной, поскольку уравнение относится к эллиптическому типу. Задача Коши формулируется как обратная задача и сводится к задаче выпуклой оптимизации в гильбертовом пространстве. Оптимизируемый функционал и его градиент вычисляются с помощью решения краевых задач, которые, в свою очередь, корректны и могут быть приближенно решены стандартными численными методами, такими как конечно-разностные схемы и разложения в ряды Фурье. Экспериментально исследуются сходимость применяемого быстрого градиентного метода и качество получаемого таким образом решения. Эксперимент показывает, что ускоренный градиентный метод — метод подобных треугольников — сходится быстрее, чем неускоренный метод. Сформулированы и доказаны теоремы о вычислительной сложности полученных алгоритмов. Установлено, что разложения в ряды Фурье превосходят конечно-разностные схемы по скорости вычислений и улучшают качество получаемого решения. Сделана попытка использовать рестарты метода подобных треугольников после уменьшения невязки функционала вдвое. В этом случае сходимость не улучшается, что подтверждает отсутствие сильной выпуклости. Эксперименты показывают, что неточность вычислений более адекватно описывается аддитивной концепцией шума в оракуле первого порядка. Этот фактор ограничивает достижимое качество решения, но ошибка не накапливается. Полученные результаты показывают, что использование ускоренных градиентных методов оптимизации позволяет эффективно решать обратные задачи.

Ключевые слова: обратные задачи, выпуклая оптимизация, оптимизация в гильбертовом пространстве, методы первого порядка, быстрые градиентные методы, неточный оракул

Исследование выполнено за счет гранта Российского научного фонда (проект № 21-71-30005).

UDC: 519.85

## Application of gradient optimization methods to solve the Cauchy problem for the Helmholtz equation

N. V. Pletnev<sup>1,a</sup>, P. E. Dvurechensky<sup>2,b</sup>, A. V. Gasnikov<sup>1,3,4,c</sup>

<sup>1</sup>Moscow Institute of Physics and Technology,  
9 Institutskiy per., Dolgoprudny, Moscow Region, 141701, Russia

<sup>2</sup>Weierstrass Institute for Applied Analysis and Stochastics,  
39 Mohrenstraße, Berlin, 10117, Germany

<sup>3</sup>Institute for Information Transmission Problems of Russian Academy of Sciences,  
19/1 Bol'shoy Karetnyy per., Moscow, 212705, Russia

<sup>4</sup>Caucasus Mathematical Center, Adyghe State University,  
208 Pervomaysk st., Maikop, Adyghe, 385000, Russia

E-mail: <sup>a</sup> nikita.pletnev@phystech.edu, <sup>b</sup> pavel.dvurechensky@wias-berlin.de, <sup>c</sup> gasnikov@yandex.ru

Received 13.02.2022.

Accepted for publication 13.02.2022.

The article is devoted to studying the application of convex optimization methods to solve the Cauchy problem for the Helmholtz equation, which is ill-posed since the equation belongs to the elliptic type. The Cauchy problem is formulated as an inverse problem and is reduced to a convex optimization problem in a Hilbert space. The functional to be optimized and its gradient are calculated using the solution of boundary value problems, which, in turn, are well-posed and can be approximately solved by standard numerical methods, such as finite-difference schemes and Fourier series expansions. The convergence of the applied fast gradient method and the quality of the solution obtained in this way are experimentally investigated. The experiment shows that the accelerated gradient method — the Similar Triangle Method — converges faster than the non-accelerated method. Theorems on the computational complexity of the resulting algorithms are formulated and proved. It is found that Fourier's series expansions are better than finite-difference schemes in terms of the speed of calculations and improve the quality of the solution obtained. An attempt was made to use restarts of the Similar Triangle Method after halving the residual of the functional. In this case, the convergence does not improve, which confirms the absence of strong convexity. The experiments show that the inaccuracy of the calculations is more adequately described by the additive concept of the noise in the first-order oracle. This factor limits the achievable quality of the solution, but the error does not accumulate. According to the results obtained, the use of accelerated gradient optimization methods can be the way to solve inverse problems effectively.

Keywords: inverse problems, convex optimization, optimization in a Hilbert space, first-order methods, fast gradient method, inexact oracle

Citation: *Computer Research and Modeling*, 2022, vol. 14, no. 2, pp. 417–444 (Russian).

This research was funded by Russian Science Foundation (project No. 21-71-30005).

## 1. Введение

Теория обратных и некорректных задач — обширная и бурно развивающаяся область математики, имеющая многообразные приложения в физике, астрономии, геофизике. Если в прямых задачах ищут описание явлений по начальным условиям и параметрам среды, то в обратных — наоборот. Часто такие задачи оказываются некорректно поставленными и поэтому не могут быть решены обычными методами.

В монографии [Kabanikhin, 2012] предпринята попытка описать все известные на данный момент задачи такого рода. Например, некорректно поставленной является задача Коши для уравнения Гельмгольца, что будет доказано в § 3.

Настоящая работа посвящена ее решению путем сведения к задаче оптимизации в гильбертовом пространстве с дальнейшим применением методов оптимизации первого порядка, то есть методов, использующим лишь значения функции и ее градиента.

Как уже обозначено, эллиптические уравнения, к которым относится и уравнение Гельмгольца, имеют разнообразное применение в физических исследованиях. Однако задача Коши для уравнений данного класса некорректна, то есть ее решение не является непрерывно зависящим от начальных условий (доказано в разделе 3). Поэтому применение для ее решения численных методов затруднено.

Однако существуют достаточно точные методы решения краевых задач. Для получения краевой задачи, имеющей то же решение, что и исходная задача Коши, необходимо решить операторное уравнение. Это, в свою очередь, делается путем решения соответствующей оптимизационной задачи. Такой подход описывается, например, в статье [Kabanikhin et al., 2014] (см. также [Matyukhin, 2021]). Его развитие осуществлено в работе.

Поскольку решение данных задач в любом случае требует применения численных методов, при вычислении градиента возникают погрешности, зависящие от выбора шага сетки. Соответственно, требуется баланс между точностью вычисления градиента и скоростью сходимости градиентного метода. Этот вопрос освещается в монографии [Danskin, 1967]. Соответствующие результаты необходимо распространить на рассматриваемую задачу.

Существуют два основных подхода к описанию шума в градиенте, введенные в работе [Polyak, 1987]:

$$\begin{aligned} \|\tilde{\nabla}J(q) - \nabla J(q)\| &\leq \delta \quad \forall q \in Q \text{ — аддитивный шум;} \\ \|\tilde{\nabla}J(q) - \nabla J(q)\| &\leq \alpha \|\nabla J(q)\| \quad \forall q \in Q, \alpha \in [0; 1) \text{ — относительный шум.} \end{aligned} \quad (1)$$

В статьях [Polyak, 1981; Vasin, Gasnikov, Spokoiny, 2021] показано, что в случае аддитивного шума результаты пессимистичны: с ростом количества итераций ошибка накапливается. Для относительного шума ситуация намного лучше, лишь замедляется в  $\left(\frac{1+\alpha}{1-\alpha}\right)^2$  раз сходимость [Polyak, 1987]. Поэтому данные случаи различимы с помощью эксперимента. В статье [Vasin, Gasnikov, Spokoiny, 2021] описаны результаты, достигаемые в данных концепциях на практике.

Основным содержанием работы являются реализация описанных идей решения некорректной задачи, а также их экспериментальная проверка: проверка применимости к решению градиентных методов, исследование зависимости сходимости от величины шума градиента, выявление модели шума, более адекватно описывающей задачу.

Структура работы: в § 2 приводится постановка задачи и описывается, как она сводится к задаче оптимизации; § 3 обосновывает некорректность исходной задачи. § 4 посвящен исследованию свойств оптимизируемой функции и ее градиента. В § 5 описывается ускоренный метод, применяемый для решения оптимизационной задачи. § 6 содержит описание метода приближенного решения и полученных результатов для вычисления градиента с помощью разностных

схем, а § 7 — то же самое, но с применением рядов Фурье. § 8 посвящен определению адекватной концепции шума. В § 9 приводятся результаты, относящиеся к применению рестартов. В § 10 изложен подход, связанный с дискретизацией исходной задачи, и показано, что он не позволяет улучшить решение. § 11 — заключительный, в нем собраны ключевые выводы.

## 2. Постановка задачи

Исследование стационарных процессов часто приводит к задаче Коши для уравнения Гельмгольца:

$$\begin{cases} \Delta u + \kappa^2 u = 0, & (x, y) \in \Omega = (0, 1) \times (0, 1), \\ u|_{x=0} = f(y), & y \in [0, 1], \\ u_x|_{x=0} = g(y), & y \in [0, 1], \\ u|_{y=0} = u|_{y=1} = 0, & x \in [0, 1]. \end{cases} \quad (2)$$

Эта задача Коши поставлена некорректно, то есть отсутствует непрерывная зависимость решения от начальных условий. Данное утверждение будет доказано в следующем разделе.

Рассмотрим краевую задачу:

$$\begin{cases} \Delta u + \kappa^2 u = 0, & (x, y) \in \Omega = (0, 1) \times (0, 1), \\ u_x|_{x=0} = g(y), & y \in [0, 1], \\ u|_{y=0} = u|_{y=1} = 0, & x \in [0, 1], \\ u|_{x=1} = q(y), & y \in [0, 1]. \end{cases} \quad (3)$$

В отличие от исходной задачи Коши она корректна и допускает эффективное решение численными методами. И если функция  $q(y)$  может быть подобрана так, что решение краевой (прямой) задачи является также решением исходной задачи Коши (которая является обратной), то некорректная задача сводится к корректной. Качество такого подбора может быть оценено с помощью сравнения решения (3) с  $f(y)$ , то есть вычисления нормы разности в гильбертовом пространстве  $C^1[0, 1] \cap L^2[0, 1]$ .

С другой стороны, минимизация данной нормы в гильбертовом пространстве позволяет подобрать требуемую функцию  $q(y)$ . При решении обратной задачи с граничным условием, соответствующим оптимальному элементу, получается решение исходной задачи Коши.

Таким образом, решение некорректно поставленной задачи Коши сводится к оптимизации в гильбертовом пространстве функционала, который вычисляется с использованием решения корректно поставленных задач.

Этот подход и исследуется в работе. Оцениваются вычислительная сложность и погрешность такого решения.

## 3. Некорректность задачи

Запишем определение непрерывности решения по начальным условиям согласно учебному пособию [Тихонов, Самарский, 1999]:

$$\forall \varepsilon > 0 \exists \delta > 0: \forall f, g \in C^1[0, 1] \cap L^2[0, 1] \text{ т. ч. } \|f\|_2 < \delta, \|g\|_2 < \delta \Leftrightarrow \|q\|_2 < \varepsilon,$$

где  $q(y) = u(1, y)$ , а  $u$  — решение соответствующей задачи Коши (2).

Взяв  $\varepsilon = 1$  и зафиксировав  $\delta > 0$ , рассмотрим следующие начальные условия ( $\kappa$  – параметр из уравнения Гельмгольца):

$$f_m(y) = \frac{\delta}{\pi m} \sin \pi m y \quad \text{и} \quad g_m(y) = \sqrt{\pi^2 m^2 - \kappa^2} f_m(y).$$

$m$  – натуральное число, которое будет выбрано позднее. Для них решением задачи Коши будет функция

$$u_m(x, y) = \frac{\delta}{\pi m} e^{\sqrt{\pi^2 m^2 - \kappa^2} x} \sin \pi m y,$$

что легко проверяется по определению.

Норма начальных условий и решения:

$$\|f_m\|_2 = \frac{\delta}{2\pi m} < \delta, \quad \|g_m\|_2 = \frac{\delta}{2\pi m} \sqrt{\pi^2 m^2 - \kappa^2} < \delta, \quad \|q_m\|_2 = \frac{\delta}{2\pi m} e^{\sqrt{\pi^2 m^2 - \kappa^2}} \rightarrow +\infty$$

при  $m \rightarrow +\infty$ , поэтому найдется  $m$  такое, что  $\|q_m\|_2 > \varepsilon$ . Это и обозначает отсутствие непрерывности решения по начальным условиям, то есть некорректность поставленной задачи.

## 4. Функционал и градиент

### 4.1. Постановка задачи оптимизации

Запишем задачу поиска  $q(y)$  по известному  $f(y)$  для решения задачи Коши в операторном виде. Для этого на гильбертовом пространстве со стандартным скалярным произведением

$$H = \{q \in C^1[0, 1] \cap L^2[0, 1] : q(0) = q(1) = 0\},$$

которому принадлежит решение обратной задачи, определим оператор  $A : H \rightarrow H$  следующим образом:

$$(Aq)(y) = u(0, y) \quad \forall y \in (0, 1),$$

где  $u$  – решение краевой задачи (3), определенной в § 2.

Теперь задача принимает вид

$$Aq = f,$$

или

$$J(q) = \frac{1}{2} \|Aq - f\|_2^2 \rightarrow \min_{q \in H}.$$

### 4.2. Вычисление градиента функционала

Из определения следует, что производная Фреше для рассматриваемого функционала равна

$$\nabla J(q) = A^*(Aq - f),$$

где  $A^*$  – сопряженный оператор. Проведем вычисления, следуя методу из пособия [Васильев, 2011].

Итак,

$$J(q) = \frac{1}{2} \int_0^1 |u(0, y) - f(y)|^2 dy,$$

где  $u$  — решение задачи (3):

$$\begin{cases} u_{xx} + u_{yy} + \kappa^2 u = 0, & (x, y) \in (0, 1) \times (0, 1), \\ u|_{x=1} = q(y), & y \in [0, 1], \\ u_x|_{x=0} = g(y), & y \in [0, 1], \\ u|_{y=0} = u|_{y=1} = 0, & x \in [0, 1]. \end{cases}$$

Запишем функцию Лагранжа, взяв множители для первых двух строк, а две последние считая задающими «простое» множество:

$$\begin{aligned} L(q, u; \psi = (\psi(x, y), \lambda(y))) &= \frac{1}{2} \int_0^1 |u(0, y) - f(y)|^2 dy - \\ &- \int_0^1 \int_0^1 (u_{xx} + u_{yy} + \kappa^2 u) \psi(x, y) dx dy + \int_0^1 (q(y) - u(1, y)) \lambda(y) dy. \end{aligned}$$

Для получения сопряженной задачи на  $\psi$  варьируем по  $\delta u(x, y)$ :

$$\delta_u L(q, u; \psi) = \int_0^1 (u(0, y) - f(y)) \delta u(0, y) dy - \int_0^1 \int_0^1 (\delta u_{xx} + \delta u_{yy} + \kappa^2 \delta u) \psi(x, y) dx dy - \int_0^1 \delta u(1, y) \lambda(y) dy.$$

При этом

$$\delta u_x(0, y) = \delta u(x, 0) = \delta u(x, 1) = 0.$$

Интегрируем по частям слагаемые с вторыми производными:

$$\begin{aligned} \int_0^1 \int_0^1 \psi(x, y) \delta u_{xx} dx dy &= \int_0^1 \psi(1, y) \delta u_x(1, y) dy - \int_0^1 (\psi_x(1, y) \delta u(1, y) - \psi_x(0, y) \delta u(0, y)) dy + \\ &+ \int_0^1 \int_0^1 \psi_{xx} \delta u(x, y) dx dy; \\ \int_0^1 \int_0^1 \psi(x, y) \delta u_{yy} dx dy &= \int_0^1 \psi(x, 1) \delta u_y(x, 1) dx - \int_0^1 \psi(x, 0) \delta u_y(x, 0) dx + \int_0^1 \int_0^1 \psi_{yy} \delta u(x, y) dx dy. \end{aligned}$$

Подставляем и вычисляем вариацию функции Лагранжа:

$$\begin{aligned} \delta_u L(q, u; \psi) &= - \int_0^1 \int_0^1 (\psi_{xx} + \psi_{yy} + \kappa^2 \psi) \delta u(x, y) dx dy + \int_0^1 (u(0, y) - f(y) - \psi_x(0, y)) \delta u(0, y) dy + \\ &+ \int_0^1 (\psi_x(1, y) - \lambda(y)) \delta u(1, y) dy - \int_0^1 \psi(1, y) \delta u_x(1, y) dy + \int_0^1 \psi(x, 0) \delta u_y(x, 0) dx - \int_0^1 \psi(x, 1) \delta u_y(x, 1) dx. \end{aligned}$$

Из необходимого условия экстремума получаем систему уравнений:

$$\begin{cases} \psi_{xx} + \psi_{yy} + \kappa^2 \psi = 0, & (x, y) \in (0, 1) \times (0, 1), \\ \psi|_{y=0} = \psi|_{y=1} = 0, & x \in [0, 1], \\ \psi|_{x=1} = 0, & y \in [0, 1], \\ \psi_x|_{x=0} = u(0, y) - f(y), & y \in [0, 1], \\ \psi_x|_{x=1} = \lambda(y), & y \in [0, 1]. \end{cases}$$

Последнее уравнение — связь между множителями Лагранжа. Поскольку

$$\lambda(y) = L_q(y) = \nabla J(q)(y)$$

(формула Демьянова – Данскина – Рубинова [Демьянов, 1974; Danskin, 1967]), вычисление градиента функционала свелось к решению двух краевых задач.

Обозначим через  $\Gamma[a(y), b(y)](x, y)$  решение  $v(x, y)$  следующей краевой задачи:

$$\begin{cases} v_{xx} + v_{yy} + \kappa^2 v = 0, & (x, y) \in (0, 1) \times (0, 1), \\ v|_{x=1} = a(y), & y \in [0, 1], \\ v_x|_{x=0} = b(y), & y \in [0, 1], \\ v|_{y=0} = v|_{y=1} = 0, & x \in [0, 1]. \end{cases}$$

---

**Algorithm 1. Gradient precise oracle**

---

**Input:**  $q(y) \in H$ ;  $f(y)$ ,  $g(y)$ ,  $\kappa$  — параметры задачи

- 1: вычислить  $u(x, y) := \Gamma[q(y), g(y)](x, y)$ ;
- 2: вычислить  $\psi(x, y) := \Gamma[0, u(0, y) - f(y)](x, y)$ ;
- 3: вычислить  $\nabla J(q)(y) := \psi_x(1, y)$ .

**Output:**  $\nabla J(q)(y)$ .

---

**4.3. Константа Липшица для градиента**

Из определения очевидно, что оператор  $A^*$  является линейным;  $A$  — линейный только при  $g(y) \equiv 0$ , такой оператор обозначим как  $A_0$ .

Градиент липшицев:

$$\begin{aligned} \|\nabla J(q_1) - \nabla J(q_2)\|_2 &= \|A^*(Aq_1 - f) - A^*(Aq_2 - f)\|_2 = \|A^*(Aq_1 - Aq_2)\|_2 = \\ &= \|A^*A_0(q_1 - q_2)\|_2 \leq L\|q_1 - q_2\|_2, \end{aligned}$$

где  $L$  — наибольшее собственное число оператора  $A^*A_0$  (поскольку он самосопряженный и неотрицательно определенный, все собственные числа — вещественные и неотрицательные, а наибольшее из них является константой Липшица градиента).

Для нахождения собственных чисел ищем

$$u(x, y) = \Gamma[q(y), 0](x, y) \quad \text{и} \quad (A^*A_0q)(y) = \Gamma[0, u(0, y)]_x(1, y)$$

в виде ряда по собственным функциям стационарного оператора, то есть в виде

$$u(x, y) = \sum_{m=1}^{\infty} X_m(x)Y_m(y).$$



Каждое слагаемое  $X_m(x)Y_m(y)$  удовлетворяет дифференциальному уравнению и условиям

$$Y_m(0) = Y_m(1) = 0.$$

Подставляя в уравнение, получаем

$$X_m''(x)Y_m(y) + X_m(x)Y_m''(y) + \kappa^2 X_m(x)Y_m(y) = 0,$$

откуда

$$\frac{X_m''(x)}{X_m(x)} + \kappa^2 = -\frac{Y_m''(y)}{Y_m(y)} = l.$$

Краевая задача

$$Y''(y) = -lY(y), \quad Y(0) = Y(1) = 0$$

имеет нетривиальные решения только при

$$l_m = \pi^2 m^2, \quad Y_m(y) = \sin \pi m y, \quad m = 1, 2, \dots$$

Как известно из курса математического анализа, система функций  $\{\sin \pi m y\}_{m=1}^{\infty}$  является базисом в пространстве функций

$$\{q(y) \in C^1[0, 1] \cap L^2[0, 1] \mid q(0) = q(1) = 0\}.$$

Решая соответствующие уравнения (подробно это будет описано в подпараграфе 7.1), находим собственные значения оператора  $A^*A_0$ :

$$\lambda_m = \begin{cases} \frac{1}{\cos^2 \sqrt{\kappa^2 - \pi^2 m^2}}, & \pi m < \kappa, \\ 1, & \pi m = \kappa, \\ \frac{1}{\operatorname{ch}^2 \sqrt{\pi^2 m^2 - \kappa^2}}, & \pi m > \kappa. \end{cases}$$

При  $\kappa \leq \pi$  градиент является липшицевым с  $L = 1$ . При  $\kappa^2 = (\pi l)^2 + \left(\frac{\pi}{2} + \pi m\right)^2$  — вырожденный случай, в котором задача не имеет единственного решения. При других  $\kappa > \pi$  градиент липшицев, но  $L > 1$ . Ограничимся случаем  $\kappa \leq \pi$ .

Поскольку  $\lim_{m \rightarrow \infty} \lambda_m = 0$ , функционал не является сильно выпуклым ( $\mu = 0$ ).

Эти оценки объединяет следующая лемма.

**Лемма 1.** При  $\kappa \leq \pi$  задача минимизации функционала  $J(q)$  является выпуклой с липшицевым градиентом, константа Липшица градиента  $L = 1$ , константа сильной выпуклости  $\mu = 0$ .

## 5. Метод подобных треугольников

В статье [Vasin, Gasnikov, Spokoiny, 2021] рассматривается ускоренный метод подобных треугольников (STM), который может применяться для решения выпуклых (не обязательно сильно выпуклых) задач с известными константами Липшица градиента и сильной выпуклости. Выпишем его в применении к решаемой задаче с  $L = 1$ ,  $\mu = 0$ .  $\mathcal{Q}$  в нашей задаче — это множество непрерывно дифференцируемых на  $[0, 1]$  функций, принимающих на концах отрезка значения 0.

Для не сильно выпуклой функции в статье [Vasin, Gasnikov, Spokoiny, 2021] доказана теорема о сходимости данного метода в случае аддитивного шума в градиенте:

$$f(x_N) - f(x^*) \leq \frac{4LR^2}{N^2} + 3\tilde{R}\delta + \frac{N\delta^2}{2L},$$

где  $R = \|x_0 - x^*\|_2$ ,  $\tilde{R} = \max_{0 \leq n \leq N} \{\|x_n - x^*\|_2, \|z_n - x^*\|_2, \|\tilde{x}_n - x^*\|_2\}$ ,  $\delta$  — верхняя оценка аддитивного шума (1),  $\tilde{\nabla} f(x)$  — приближенно вычисляемый градиент  $f(x)$ .



---

**Algorithm 2. STM**

---

**Input:**  $f: Q \rightarrow \mathbb{R}$ ,  $x_{start} \in Q$ ,  $L, N$ .

- 1:  $\tilde{x}_0 := x_{start}$ ;  $A_0 := \frac{1}{L}$ ;  $\alpha_0 := \frac{1}{L}$ ;  $z_0 := \tilde{x}_0 - \alpha_0 \tilde{\nabla} f(\tilde{x}_0)$ ;  $x_0 := z_0$ ;
- 2: **for**  $n = 1 \dots N$  **do**
- 3:    $\alpha_n := \frac{1}{2L} + \sqrt{\frac{1}{4L^2} + A_{n-1}}$ ;
- 4:    $A_n := A_{n-1} + \alpha_n$ ;
- 5:    $\tilde{x}_n := \frac{A_{n-1}x_{n-1} + \alpha_n z_{n-1}}{A_n}$ ;
- 6:    $z_n := z_{n-1} - \alpha_n \tilde{\nabla} f(\tilde{x}_n)$ ;
- 7:    $x_n := \frac{A_{n-1}x_{n-1} + \alpha_n z_n}{A_n}$ ;
- 8: **end for**

**Output:**  $x_N$ .

---

## 6. Приближенное решение с использованием разностных схем

Стандартный подход к решению краевых задач — аппроксимация уравнений в частных производных разностными схемами. Так действовали авторы статьи [Kabanikhin et al., 2014], и мы последуем их примеру.

### 6.1. Разностная схема для краевой задачи

Пусть на  $[0, 1] \times [0, 1]$  задана сетка с шагом  $h$ , то есть

$$x = ih, \quad y = jh, \quad 0 \leq i, j \leq P \quad (Ph = 1).$$

Тогда  $\Gamma_h[a_j, b_j]_{i,j}$  вводится как решение разностной схемы, аналогичной примененной в работе [Kabanikhin et al., 2014]:

$$\begin{cases} \frac{v_{i+1,j} - 2v_{i,j} + v_{i-1,j}}{h^2} + \frac{v_{i,j+1} - 2v_{i,j} + v_{i,j-1}}{h^2} + \kappa^2 v_{i,j} = 0, & 1 \leq i, j \leq P-1, \\ v_{P,j} = a_j, & 1 \leq j \leq P-1, \\ \frac{v_{1,j} - v_{0,j}}{h} = b_j, & 1 \leq j \leq P-1, \\ v_{i,0} = v_{i,P} = 0, & 0 \leq i \leq P. \end{cases}$$

При этом

$$a_j = a(jh), \quad b_j = b(jh), \quad v_{i,j} = v(ih, jh).$$

Поскольку

$$\begin{aligned} v_{i\pm 1,j} &= v(ih, jh) \pm \frac{\partial v}{\partial x} h + \frac{\partial^2 v}{\partial x^2} \frac{h^2}{2} \pm \frac{\partial^3 v}{\partial x^3} \frac{h^3}{6} + \frac{\partial^4 v}{\partial x^4} \frac{h^4}{24} + O(h^5), \quad h \rightarrow 0, \\ \frac{v_{i+1,j} - 2v_{i,j} + v_{i-1,j}}{h^2} &= \frac{\partial^2 v}{\partial x^2} + O(h^2), \end{aligned}$$

аналогично

$$\frac{v_{i,j+1} - 2v_{i,j} + v_{i,j-1}}{h^2} = \frac{\partial^2 v}{\partial y^2} + O(h^2)$$

(из формулы Тейлора; все производные берутся в точке  $(ih, jh)$ ), поэтому первое уравнение системы аппроксимирует уравнение Гельмгольца с невязкой второго порядка.

Аналогично получается, что сеточное приближение для граничного условия имеет первый порядок аппроксимации. Поэтому схема имеет первый порядок аппроксимации.

Поскольку схема устойчива, из теоремы Рябенского – Лакса [Рябенский, 2000] следует сходимость первого порядка к решению дифференциальной краевой задачи при стремлении шага сетки к нулю.

### 6.2. Приближенное вычисление градиента на сетке

При применении градиентного метода в качестве оракула первого порядка используется следующий алгоритм.

---

#### Algorithm 3. Gradient approximate oracle

---

**Input:**  $q(y) \in H$ ;  $f_j = f(jh)$ ,  $g_j = g(jh)$ ,  $\kappa$  — параметры задачи,  $h = \frac{1}{P}$  — шаг сетки.

- 1: вычислить  $u_{i,j} := \Gamma_h[q_j, g_j]_{i,j}$ ;
- 2: вычислить  $\psi_{i,j} := \Gamma_h[0, u_{0,j} - f_j]_{i,j}$ ;
- 3: вычислить  $\tilde{\nabla}J(q)_j := \frac{\psi_{P,j} - \psi_{P-1,j}}{h}$ .

**Output:**  $\tilde{\nabla}J(q)_j$ .

---

### 6.3. Теоретическая оценка погрешности

Поскольку на каждом шаге gradient approximate oracle результат отличается от точного на величину, пропорциональную шагу сетки, то и

$$\max_j |\tilde{\nabla}J(q)_j - \nabla J(q)(jh)| = O(h), \quad h \rightarrow 0.$$

Введем обозначения:  $\delta\varphi$  — погрешность вычисления функции  $\varphi$  (это может быть  $u$ ,  $\psi$ );

$$\delta\Gamma[a, b]_{i,j} := \Gamma_h[a(ih), b(jh)]_{i,j} - \Gamma[a(y), b(y)](ih, jh)$$

— погрешность решения краевой задачи  $\Gamma$  при ее аппроксимации задачей  $\Gamma_h$  на равномерной сетке. Считаем, что  $\phi_{i,j}$  и  $\phi(ih, jh)$  — это одно и то же. Тогда можно вычислить:

$$\begin{aligned} \delta u_{i,j} &= \delta\Gamma[q, g]_{i,j}; \\ \delta\psi_{i,j} &= \delta\Gamma[0, u(0, y) - f(y)]_{i,j} + \Gamma_h[0, \delta u_{0,j}]_{i,j} = \delta\Gamma[0, \Gamma[q, g](0, y) - f(y)]_{i,j} + \Gamma_h[0, \delta\Gamma[q, g]_{0,j}]_{i,j} = \\ &= (\delta\Gamma[0, \Gamma[0, g](0, y) - f(y)]_{i,j} + \Gamma_h[0, \delta\Gamma[0, g]_{0,j}]_{i,j}) + (\delta\Gamma[0, \Gamma[q, 0](0, y)]_{i,j} + \Gamma_h[0, \delta\Gamma[q, 0]_{0,j}]_{i,j}). \end{aligned}$$

Погрешность — это сумма двух слагаемых, одно из которых зависит от  $q$ , а другое — нет. При этом оно не равно тождественно нулю и зависит от мелкости сетки и начальных условий.

Соответственно, градиент не удовлетворяет концепции относительного шума, потому что погрешность такого рода не допускает оценки сверху величиной, пропорциональной норме истинного значения градиента с малым коэффициентом пропорциональности.

С другой стороны, погрешность пропорциональна мелкости разбиения, поэтому концепция аддитивного шума корректно описывает данный способ вычисления.

Обозначая количество точек отрезка  $[0, 1]$  на сетке за  $P$ , получим, что краевая задача сводится к решению линейной системы порядка  $P^2$ . Это требует  $O(P^6)$  арифметических операций, поскольку стандартные методы решения систем линейных уравнений имеют сложность кубического порядка. При этом погрешность вычисления градиента составит  $O(P^{-1})$ .

Объединим эти результаты и получим следующий факт.

**Теорема 1.** *Применение  $N$  итераций STM с приближенным вычислением градиента путем решения разностных задач требует  $O(P^6N)$  арифметических операций и обеспечивает невязку по функционалу  $O\left(\frac{R^2}{N^2} + \frac{\bar{K}}{P} + \frac{N}{P^2}\right)$ .*

*Доказательство* состоит в подстановке оценки для погрешности и  $L = 1$  в теорему о сходимости STM с неточным оракулом, удовлетворяющим аддитивной концепции шума, при  $\mu = 0$ . Лемма 1 гарантирует корректность применения STM.

### 6.4. Эксперименты

Для исследования сходимости используется задача Коши (2) с

$$\kappa = \pi, \quad f(y) = 100(y - y^2), \quad g(y) = 1000((y - 0,5)^2 - 4(y - 0,5)^4),$$

начальное приближение —  $q(y) \equiv 0$ .

Проверяется работа классического метода сопряженных градиентов [Polyak, 1987] и быстрого метода STM, которому посвящен § 6.

#### Метод сопряженных градиентов

Первая серия экспериментов показывает характер сходимости метода сопряженных градиентов при мелкости сетки 0,1, 0,05 и 0,02. На рисунках (а) построены графики зависимости функционала от номера итерации, на рисунках (б) — нормы градиента, а на рисунках (с) — примеры функций  $q_n(y)$ , которые получаются на разных итерациях.

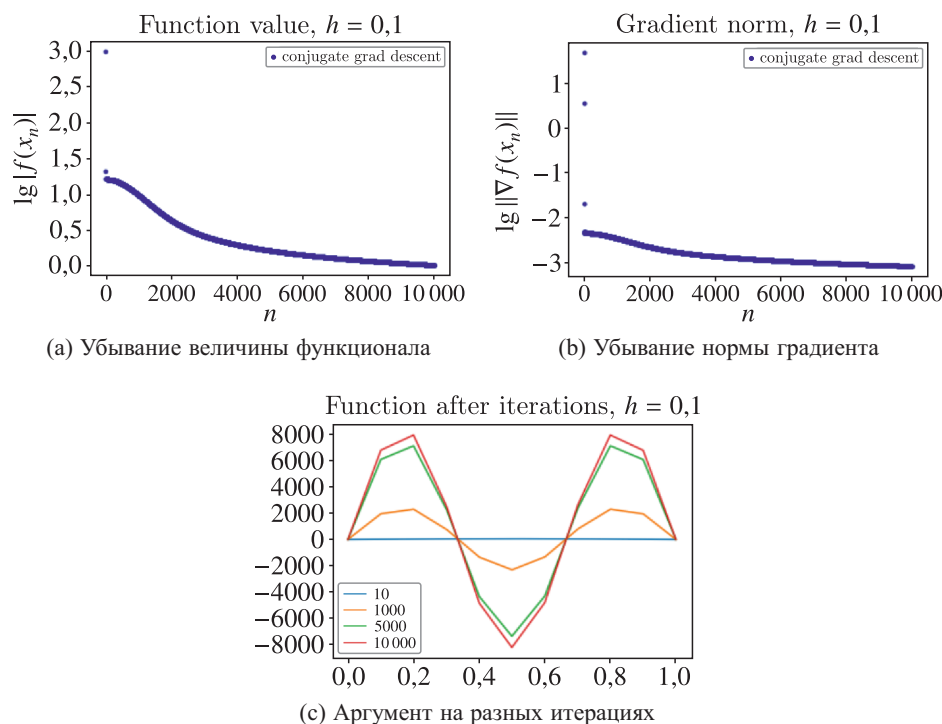


Рис. 1. Работа метода сопряженных градиентов: зависимость функционала и нормы градиента от количества итераций  $n$  и графики аргумента функционала на разных итерациях при мелкости сетки  $h = 0,1$

За первые две итерации значение функционала уменьшается примерно в 100 раз, дальнейшее уменьшение очень медленное: в 10 раз за примерно 5000 итераций (рис. 1).

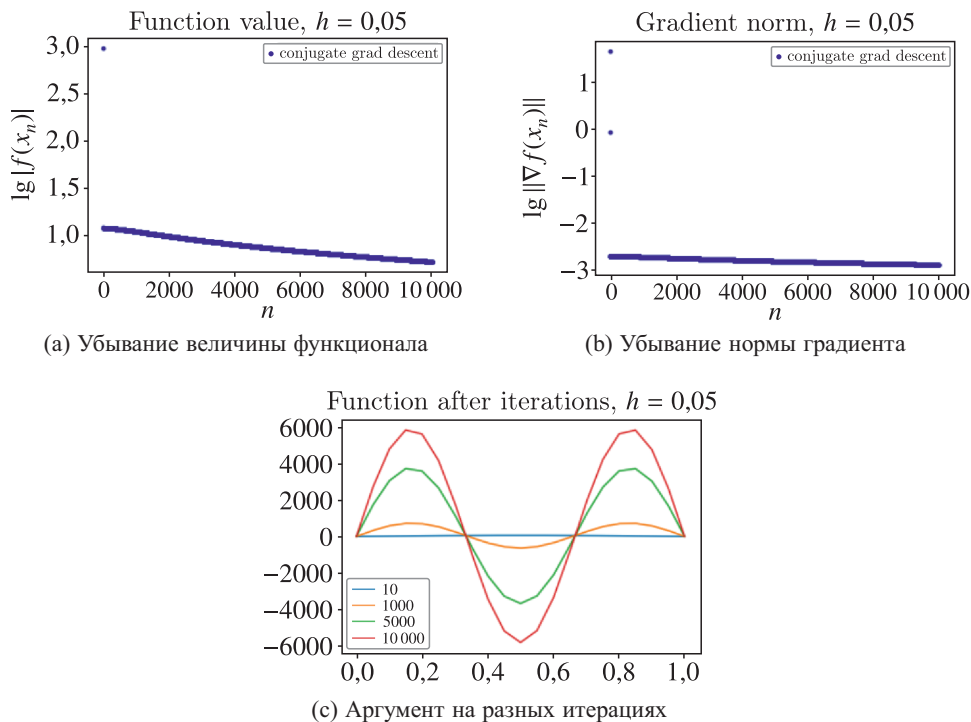


Рис. 2. Работа метода сопряженных градиентов: зависимость функционала и нормы градиента от количества итераций  $n$  и графики аргумента функционала на разных итерациях при мелкости сетки  $h = 0,05$

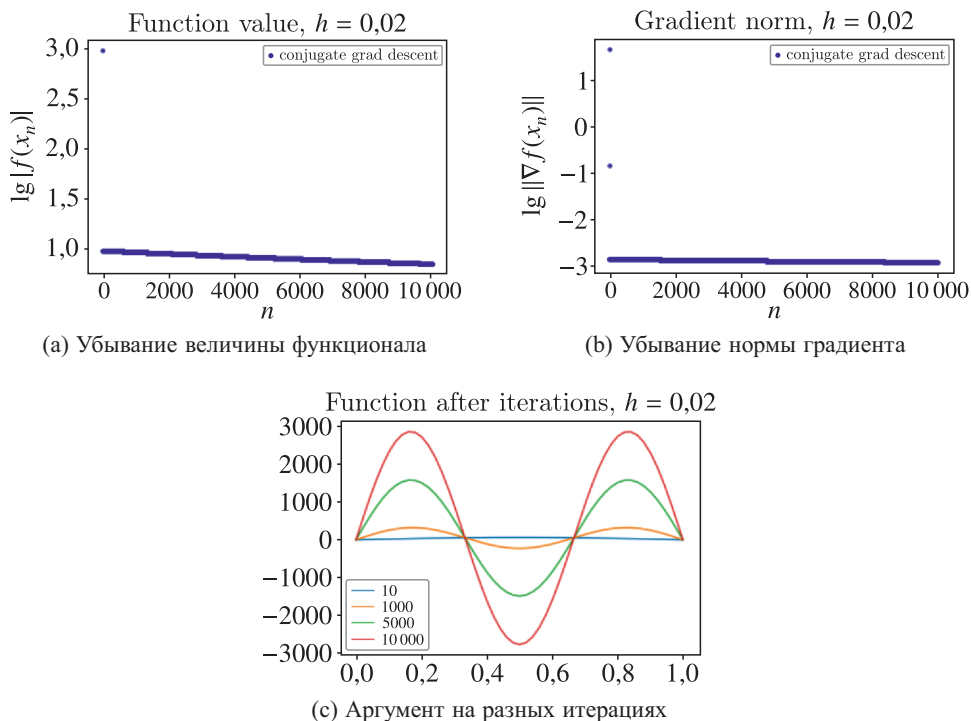


Рис. 3. Работа метода сопряженных градиентов: зависимость функционала и нормы градиента от количества итераций  $n$  и графики аргумента функционала на разных итерациях при мелкости сетки  $h = 0,02$

Уменьшение функционала на первых шагах значительное (рис. 2), а на последующих — медленнее, чем в предыдущем случае.

На первом шаге функционал уменьшается в 100 раз, а на последующих — еще медленнее, чем при  $h = 0,05$  (рис. 3).

$h$	time per iteration, s	function, min
0,1	0,057	1,047
0,05	0,071	5,386
0,02	1,04	7,309

Итак, уменьшение погрешности не ускоряет сходимость, но значительно замедляет вычисления.

### Метод подобных треугольников

Структура эксперимента такая же, как и для метода сопряженных градиентов.

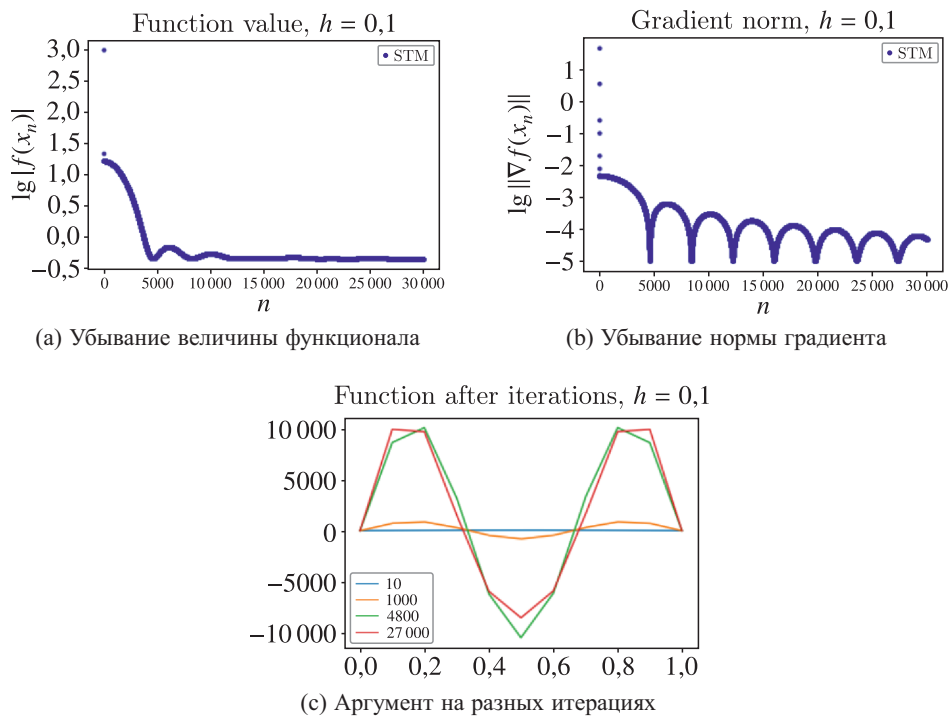


Рис. 4. Работа метода подобных треугольников: зависимость функционала и нормы градиента от количества итераций  $n$  и графики аргумента функционала на разных итерациях при мелкости сетки  $h = 0,1$

Сходимость теряет строго монотонный характер, но достигаются более близкие к минимуму значения, чем для метода сопряженных градиентов (рис. 4).

Картина при  $h = 0,05$  (рис. 5) похожая, причем достигнутый минимум еще ближе к истинному, чем при  $h = 0,1$  (рис. 4).

И снова нет монотонности — более того, наименьшее значение, достигнутое на примерно 12 000-й итерации, больше не достигается. Однако для ускоренного метода уменьшение шага сетки приводит и к уменьшению невязки по функционалу (рис. 6).

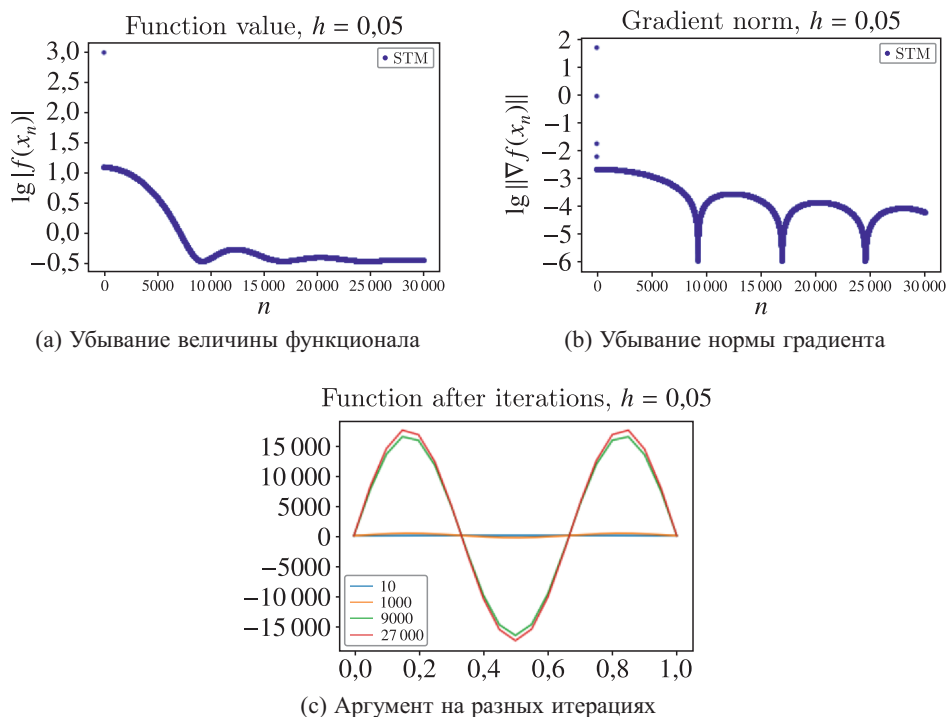


Рис. 5. Работа метода подобных треугольников: зависимость функционала и нормы градиента от количества итераций  $n$  и графики аргумента функционала на разных итерациях при мелкости сетки  $h = 0,05$

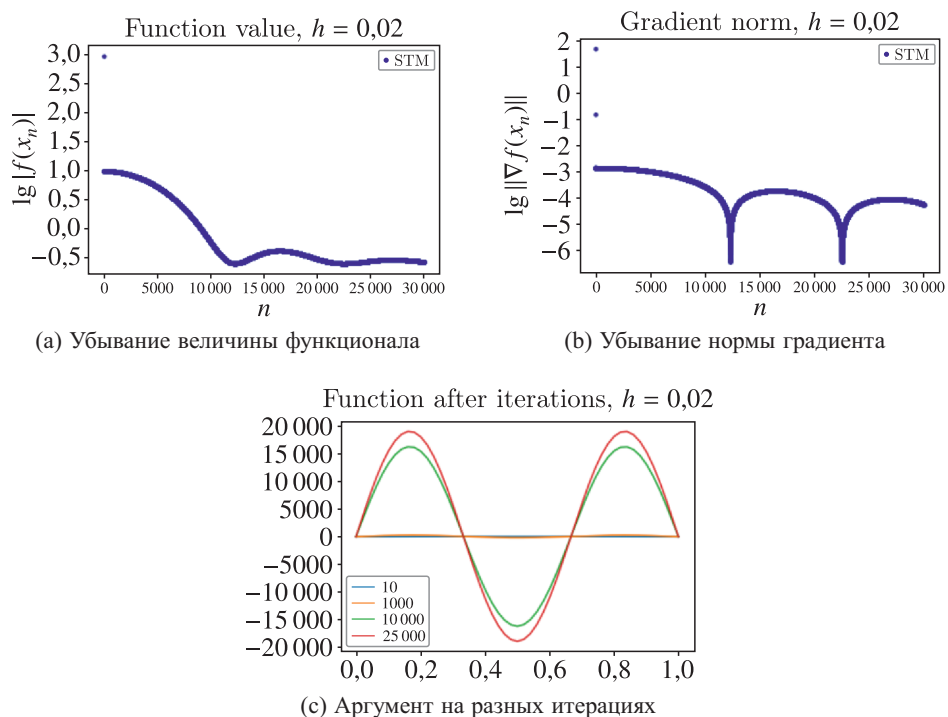


Рис. 6. Работа метода подобных треугольников: зависимость функционала и нормы градиента от количества итераций  $n$  и графики аргумента функционала на разных итерациях при мелкости сетки  $h = 0,02$

$h$	time per iteration, s	function, min
0,1	0,34	0,4465
0,05	0,43	0,3448
0,02	1,74	0,2523

Уменьшение погрешности вычисления градиента уменьшает невязку по функционалу. Ускоренный метод показывает лучшие результаты, чем метод сопряженных градиентов.

Отход от достигнутого минимума позволяет сделать осторожное предположение о том, что относительная концепция шума не является адекватным описанием погрешности, но точный вывод об этом можно сделать лишь после эксперимента с очень большим числом итераций. А это требует вычислять градиент более эффективно. Такой способ обнаружен, и о нем — следующий раздел.

## 7. Приближенное решение с использованием рядов Фурье

В курсе уравнений математической физики, который достаточно подробно изложен в пособии [Тихонов, Самарский, 1999], решения краевых задач на квадрате с нулевыми граничными условиями на двух противоположных сторонах, какой и является задача  $\Gamma$ , часто находятся в виде ряда по собственным функциям стационарного оператора  $-\Delta$ . Этот подход используется в данном разделе.

### 7.1. Решение краевой задачи с помощью разложения в ряды Фурье

Поскольку граничные условия непрерывно дифференцируемы и принимают нулевые значения в концах отрезка, то они могут быть разложены в ряды Фурье по системе  $\{\sin \pi m y\}_{m=1}^{\infty}$ , причем коэффициенты стремятся к нулю не медленнее, чем  $O(m^{-2})$  — стандартная теорема из курса математического анализа.

Вычислим

$$\Gamma[a(y), b(y)](x, y).$$

Пусть

$$a(y) = \sum_{m=1}^{\infty} \alpha_m \sin \pi m y, \quad b(y) = \sum_{m=1}^{\infty} \beta_m \sin \pi m y.$$

Коэффициенты находятся по следующим формулам:

$$\alpha_m = \frac{(a(y), \sin \pi m y)}{(\sin \pi m y, \sin \pi m y)} = \frac{\int_0^1 a(y) \sin \pi m y dy}{\int_0^1 \sin^2 \pi m y dy} = 2 \int_0^1 a(y) \sin \pi m y dy,$$

$$\beta_m = 2 \int_0^1 b(y) \sin \pi m y dy.$$

Тогда

$$\Gamma[a(y), b(y)](x, y) = \sum_{m=1}^{\infty} X_m(x) \sin \pi m y,$$



где  $X_m(x)$  — решение задачи

$$\begin{cases} X_m''(x) + (\kappa^2 - \pi^2 m^2)X_m(x) = 0, & x \in (0, 1), \\ X_m(1) = \alpha_m, \\ X_m'(0) = \beta_m. \end{cases}$$

Поскольку рассматривается только  $\kappa \leq \pi$ , всегда  $\kappa^2 - \pi^2 m^2 \leq 0$ .

Если  $\kappa = \pi$ ,  $m = 1$ , то

$$X_1''(x) = 0 \quad \text{и} \quad X_1(x) = \alpha_1 + \beta_1(x - 1).$$

Иначе, обозначая

$$\gamma_m = \sqrt{\pi^2 m^2 - \kappa^2},$$

получаем уравнение

$$X_m''(x) = \gamma_m^2 X_m(x),$$

его решение —

$$X_m(x) = \left( \frac{\alpha_m}{\operatorname{ch} \gamma_m} - \frac{\beta_m}{\gamma_m} \operatorname{th} \gamma_m \right) \operatorname{ch} \gamma_m x + \frac{\beta_m}{\gamma_m} \operatorname{sh} \gamma_m x.$$

Итак, при  $\kappa < \pi$

$$\Gamma[a(y), b(y)](x, y) = \sum_{m=1}^{\infty} \left( \left( \frac{\alpha_m}{\operatorname{ch} \gamma_m} - \frac{\beta_m}{\gamma_m} \operatorname{th} \gamma_m \right) \operatorname{ch} \gamma_m x + \frac{\beta_m}{\gamma_m} \operatorname{sh} \gamma_m x \right) \sin \pi m y,$$

а при  $\kappa = \pi$

$$\Gamma[a(y), b(y)](x, y) = (\alpha_1 + \beta_1(x - 1)) \sin \pi y + \sum_{n=2}^{\infty} \left( \left( \frac{\alpha_n}{\operatorname{ch} \gamma_n} - \frac{\beta_n}{\gamma_n} \operatorname{th} \gamma_n \right) \operatorname{ch} \gamma_n x + \frac{\beta_n}{\gamma_n} \operatorname{sh} \gamma_n x \right) \sin \pi n y.$$

## 7.2. Обоснование сходимости полученного ряда

Оценим сверху (по модулю) общий член данного ряда. Гиперболические синус и косинус положительного аргумента — возрастающие функции, поэтому  $\left| \frac{\alpha_m}{\operatorname{ch} \gamma_m} \operatorname{ch} \gamma_m x \right| \leq |\alpha_m|$ .

Функция

$$\varphi(x) = \frac{\operatorname{sh} \gamma_m x}{\gamma_m} - \frac{\operatorname{th} \gamma_m}{\gamma_m} \operatorname{ch} \gamma_m x$$

строго возрастает на  $[0, 1]$ , потому что ее производная положительна:

$$\varphi'(x) = \operatorname{ch} \gamma_m x - \operatorname{th} \gamma_m \operatorname{sh} \gamma_m x > (1 - \operatorname{th} \gamma_m) \operatorname{sh} \gamma_m x > 0;$$

$$\varphi(0) = -\frac{\operatorname{th} \gamma_m}{\gamma_m}, \quad \varphi(1) = 0 \Rightarrow |\varphi(x)| \leq \frac{\operatorname{th} \gamma_m}{\gamma_m} < 1.$$

Итак,

$$\left| \left( \left( \frac{\alpha_m}{\operatorname{ch} \gamma_m} - \frac{\beta_m}{\gamma_m} \operatorname{th} \gamma_m \right) \operatorname{ch} \gamma_m x + \frac{\beta_m}{\gamma_m} \operatorname{sh} \gamma_m x \right) \right| \times |\sin \pi m y| \leq \left| \frac{\alpha_m}{\operatorname{ch} \gamma_m} \operatorname{ch} \gamma_m x \right| + |\beta_m \varphi(x)| \leq |\alpha_m| + |\beta_m|.$$

Поскольку, как было указано выше,

$$|\alpha_m| + |\beta_m| = O(m^{-2}),$$

ряд для функции  $\Gamma[a(y), b(y)](x, y)$  мажорируется сходящимся числовым рядом, поэтому сходится абсолютно и равномерно.

### 7.3. Вычисление градиента через ряды Фурье

Пусть

$$f(y) = \sum_{m=1}^{\infty} f_m \sin \pi m y, \quad g(y) = \sum_{m=1}^{\infty} g_m \sin \pi m y, \quad q(y) = \sum_{m=1}^{\infty} q_m \sin \pi m y.$$

Сначала рассмотрим случай  $k < \pi$ . Тогда, выполняя шаги gradient precise oracle, получаем:

$$u(x, y) = \Gamma[q(y), g(y)](x, y) = \sum_{m=1}^{\infty} \left( \left( \frac{q_m}{\operatorname{ch} \gamma_m} - \frac{g_m}{\gamma_m} \operatorname{th} \gamma_m \right) \operatorname{ch} \gamma_m x + \frac{g_m}{\gamma_m} \operatorname{sh} \gamma_m x \right) \sin \pi m y,$$

$$u(0, y) - f(y) = \sum_{m=1}^{\infty} \left( \frac{q_m}{\operatorname{ch} \gamma_m} - \frac{g_m}{\gamma_m} \operatorname{th} \gamma_m - f_m \right) \sin \pi m y,$$

$$\psi(x, y) = \Gamma[0, u(0, y) - f(y)](x, y) = \sum_{m=1}^{\infty} \left( \frac{q_m}{\gamma_m \operatorname{ch} \gamma_m} - \frac{g_m}{\gamma_m^2} \operatorname{th} \gamma_m - \frac{f_m}{\gamma_m} \right) (\operatorname{sh} \gamma_m x - \operatorname{th} \gamma_m \operatorname{ch} \gamma_m x) \sin \pi m y,$$

$$\nabla J(q)(y) = \psi_x(1, y) = \sum_{m=1}^{\infty} \frac{1}{\operatorname{ch} \gamma_m} \left( \frac{q_m}{\operatorname{ch} \gamma_m} - g_m \frac{\operatorname{th} \gamma_m}{\gamma_m} - f_m \right) \sin \pi m y.$$

Теперь рассмотрим случай  $k = \pi$ . Выполняя те же шаги, получаем:

$$u(x, y) = \Gamma[q(y), g(y)](x, y) =$$

$$= (q_1 + g_1(x-1)) \sin \pi y + \sum_{m=2}^{\infty} \left( \left( \frac{q_m}{\operatorname{ch} \gamma_m} - \frac{g_m}{\gamma_m} \operatorname{th} \gamma_m \right) \operatorname{ch} \gamma_m x + \frac{g_m}{\gamma_m} \operatorname{sh} \gamma_m x \right) \sin \pi m y,$$

$$u(0, y) - f(y) = (q_1 - g_1 - f_1) \sin \pi y + \sum_{m=2}^{\infty} \left( \frac{q_m}{\operatorname{ch} \gamma_m} - \frac{g_m}{\gamma_m} \operatorname{th} \gamma_m - f_m \right) \sin \pi m y,$$

$$\psi(x, y) = \Gamma[0, u(0, y) - f(y)](x, y) = (q_1 - g_1 - f_1)(x-1) \sin \pi y +$$

$$+ \sum_{m=2}^{\infty} \left( \frac{q_m}{\gamma_m \operatorname{ch} \gamma_m} - \frac{g_m}{\gamma_m^2} \operatorname{th} \gamma_m - \frac{f_m}{\gamma_m} \right) (\operatorname{sh} \gamma_m x - \operatorname{th} \gamma_m \operatorname{ch} \gamma_m x) \sin \pi m y,$$

$$\nabla J(q)(y) = \psi_x(1, y) = (q_1 - g_1 - f_1) \sin \pi y + \sum_{m=2}^{\infty} \frac{1}{\operatorname{ch} \gamma_m} \left( \frac{q_m}{\operatorname{ch} \gamma_m} - g_m \frac{\operatorname{th} \gamma_m}{\gamma_m} - f_m \right) \sin \pi m y.$$

Поскольку  $\operatorname{ch} 0 = 1$ ,  $\lim_{t \rightarrow 0} \frac{\operatorname{sh} t}{t} = 1$ , формулы для оператора и градиента в случае  $k = \pi$  являются просто предельным случаем формул для  $k < \pi$ .

Необходимо заметить, что ряд для градиента сходится быстрее, чем ряды для  $f(y)$ ,  $g(y)$ ,  $q(y)$ , потому что его коэффициенты убывают экспоненциально.

При численном решении задачи ряд заменяется конечной суммой; коэффициенты  $f_m$ ,  $g_m$  находятся один раз путем приближенного вычисления интегралов; функция  $q(y)$  считается заданной своими коэффициентами Фурье и при необходимости вычисляется в точках сетки.

### 7.4. Теоретическая оценка погрешности

Как уже было сказано, для начальных условий коэффициенты ряда Фурье убывают пропорционально  $O(m^{-2})$ , поэтому остаток ряда ограничивается сверху как  $O(M^{-1})$ , где  $M$  — это количество слагаемых в частичной сумме.

Коэффициенты Фурье вычисляются с погрешностью  $O(S^{-1})$ , где  $S$  — количество точек разбиения отрезка при численном интегрировании.  $S > M$  для точности вычисления коэффициентов с большими номерами.

Погрешность вычисления градиента состоит из слагаемого, зависящего только от  $M$  и начальных условий, и слагаемого, зависящего от функции. Соответственно, концепция относительного шума неприменима.

Поэтому шум в градиенте описывается аддитивной концепцией с

$$\delta = O(M^{-1} + S^{-1}) = O(M^{-1}).$$

На каждой итерации STM требуется  $O(M)$  арифметических операций (вычисление градиента по формулам § 7.2 и его линейных комбинаций с фиксированным количеством ранее найденных точек).

Коэффициенты Фурье для начальных условий вычисляются один раз до начала итераций STM, это требует  $O(MS)$  арифметических операций.

Получаем новый практически важный факт.

**Теорема 2.** *Применение  $N$  итераций STM с использованием рядов Фурье для приближенного вычисления градиента требует  $O((N + S)M)$  арифметических операций и обеспечивает невязку по функционалу  $O\left(\frac{R^2}{N^2} + \frac{\bar{R}}{M} + \frac{N}{M^2}\right)$ .*

*Доказательство* аналогично доказательству теоремы 1.

## 7.5. Эксперименты

Для исследования сходимости используется задача Коши с

$$\kappa = \pi, \quad f(y) = 100(y - y^2), \quad g(y) = 1000((y - 0,5)^2 - 4(y - 0,5)^4),$$

начальное приближение —  $q(y) \equiv 0$ . Выбраны те же начальные условия, что и в эксперименте с сетками, для адекватного сравнения эффективности подходов.

### Метод сопряженных градиентов

Третья серия экспериментов показывает характер сходимости метода сопряженных градиентов для 100, 1000 и 10 000 слагаемых ряда Фурье (интегрирование производилось на 100 000 точках). На левых рисунках построены графики зависимости функционала от номера итерации, на средних — нормы градиента, а на правых — примеры функций  $q_n(y)$ , которые получаются на разных итерациях.

На первом шаге функционал уменьшается в 100 раз, а на последующих — очень медленно, даже за 30 000 итераций не достигая сокращения в 10 раз (рис. 7).

На рис. 8 полностью аналогичная картина.

Видно (рис. 9), что увеличение количества слагаемых в частичной сумме ряда Фурье не влияет на сходимость метода сопряженных градиентов.

Fourier summands	time per iteration, $10^{-3}$ s	function, min
100	0,8	5,416
1000	1,5	5,416
10 000	1,9	5,416

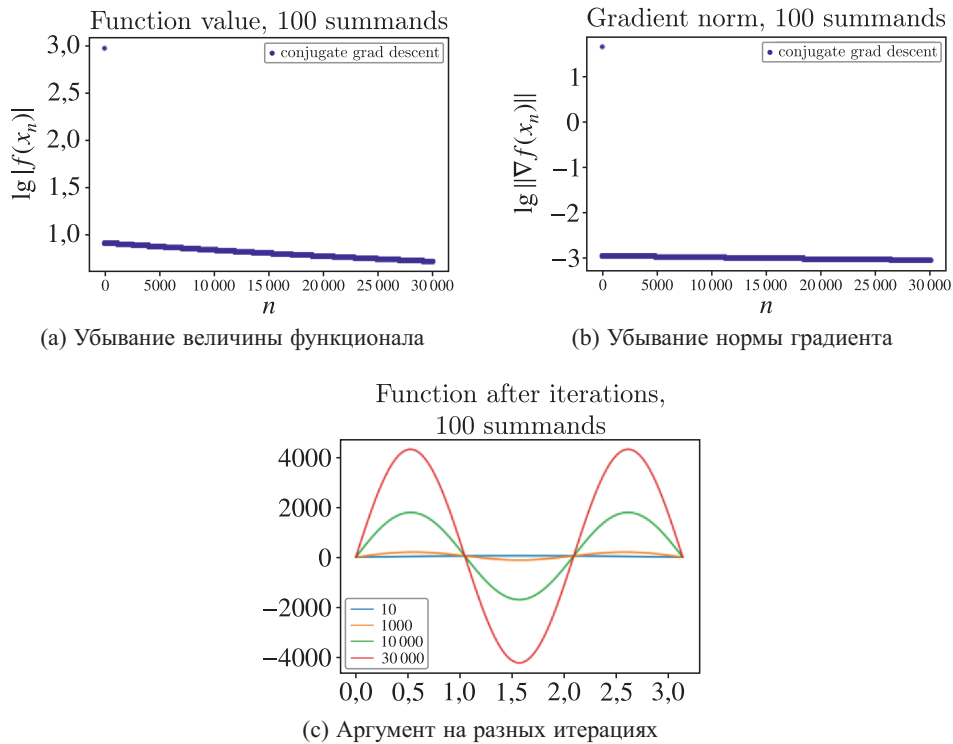


Рис. 7. Работа метода сопряженных градиентов: зависимость функционала и нормы градиента от количества итераций  $n$  и графики аргумента функционала на разных итерациях для 100 слагаемых ряда Фурье

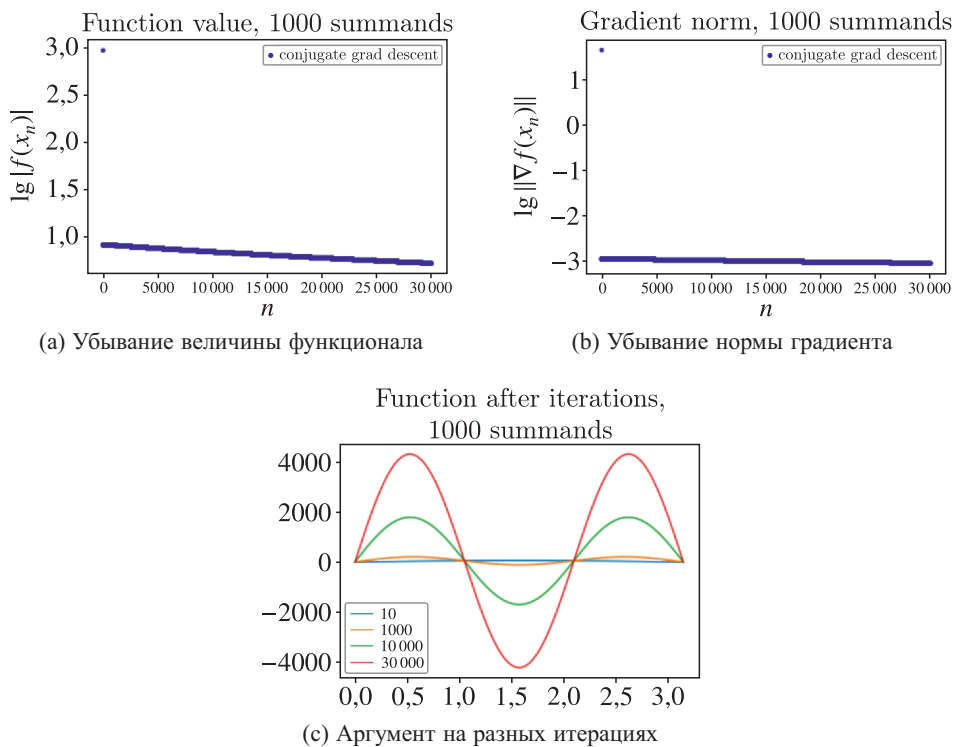


Рис. 8. Работа метода сопряженных градиентов: зависимость функционала и нормы градиента от количества итераций  $n$  и графики аргумента функционала на разных итерациях для 1000 слагаемых ряда Фурье

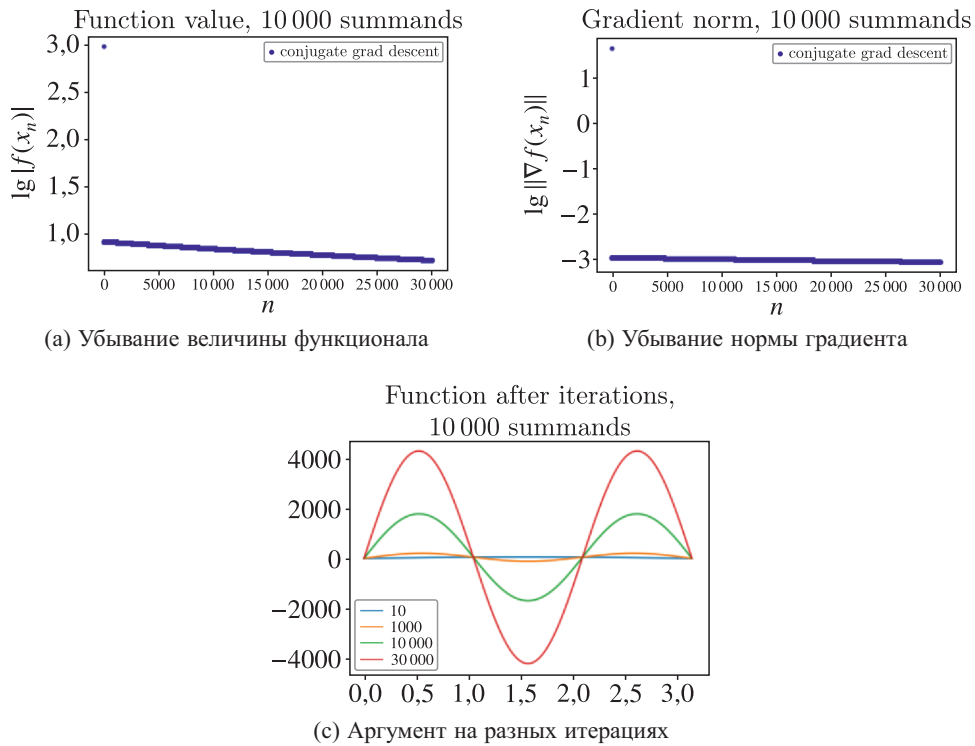


Рис. 9. Работа метода сопряженных градиентов: зависимость функционала и нормы градиента от количества итераций  $n$  и графики аргумента функционала на разных итерациях для 10 000 слагаемых ряда Фурье

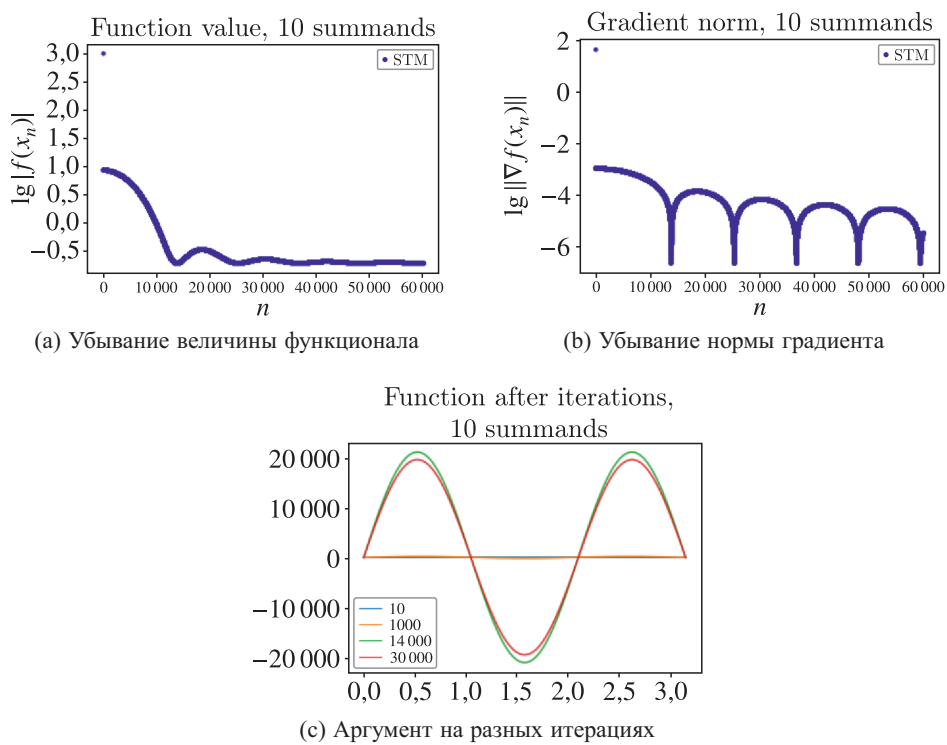


Рис. 10. Работа метода подобных треугольников: зависимость функционала и нормы градиента от количества итераций  $n$  и графики аргумента функционала на разных итерациях для 10 слагаемых ряда Фурье

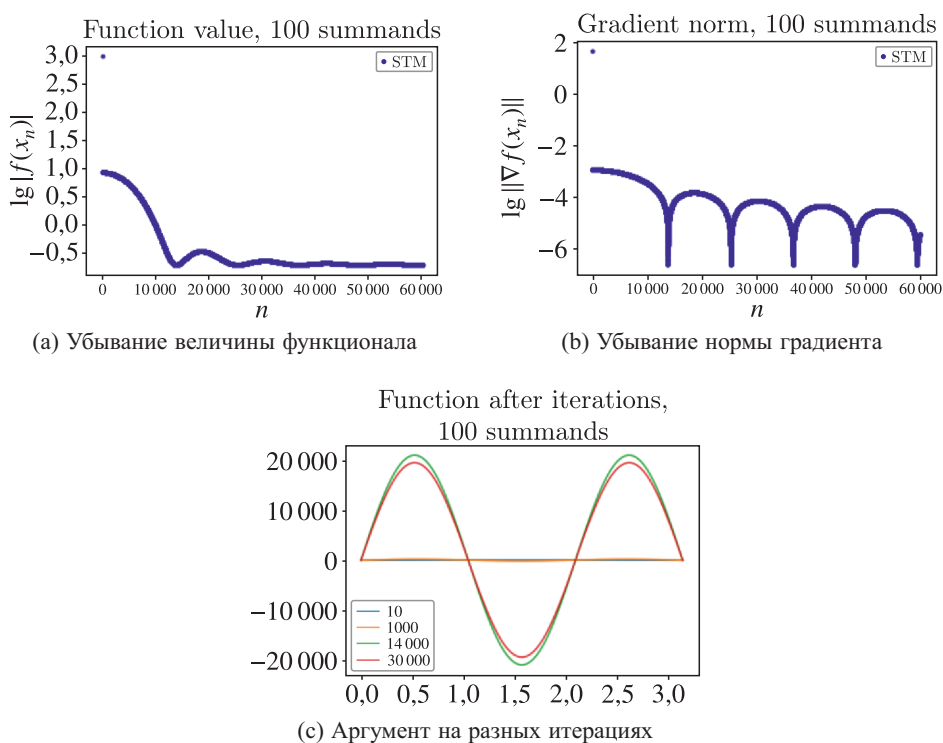


Рис. 11. Работа метода подобных треугольников: зависимость функционала и нормы градиента от количества итераций  $n$  и графики аргумента функционала на разных итерациях для 100 слагаемых ряда Фурье

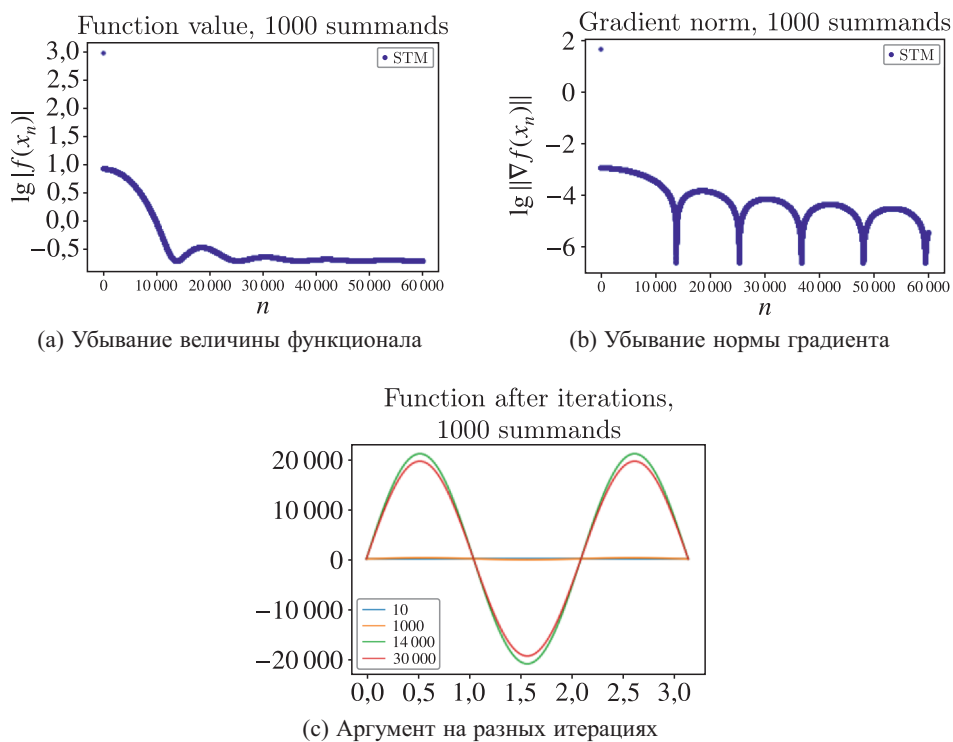


Рис. 12. Работа метода подобных треугольников: зависимость функционала и нормы градиента от количества итераций  $n$  и графики аргумента функционала на разных итерациях для 1000 слагаемых ряда Фурье

### Метод подобных треугольников

Поскольку использование большого количества слагаемых заведомо не имеет смысла, работа STM проверяется на 10, 100 и 1000 слагаемых.

Нет монотонности; рост количества слагаемых не приводит к улучшению сходимости, но замедляет вычисления.

Fourier summands	time per iteration, $10^{-3}$ s	function, min
10	0,87	0,1984
100	0,91	0,1993
1000	1,75	0,1993

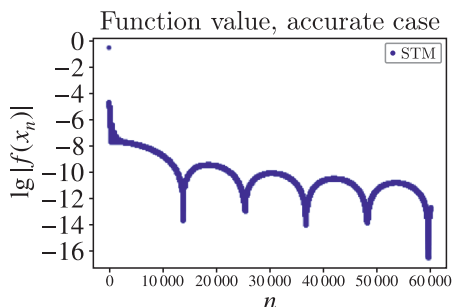
Применение рядов Фурье дает меньшую невязку функционала и является намного более эффективным по времени (каждая итерация быстрее в 1000 раз), чем решение с использованием разностных схем.

Теоремы 1 и 2 показывают, что вычисления с использованием рядов Фурье позволяют решать задачу оптимизации намного эффективнее, чем применение разностных схем. Эксперименты это подтверждают.

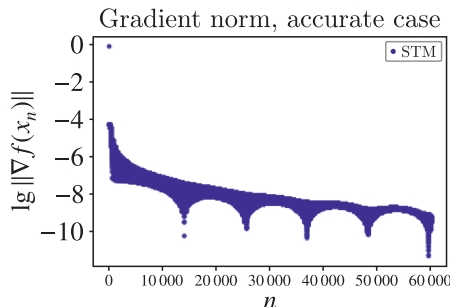
### Модельный случай с точным решением

Проверим работу алгоритма в случае, когда точное решение легко вычислить напрямую:

$$\kappa = \pi, \quad f(y) = \sin \pi y + \frac{\sin 3\pi y}{\operatorname{ch} \pi \sqrt{8}}, \quad g(y) = \frac{\pi \sqrt{3}}{\operatorname{sh} \pi \sqrt{3}} \sin 2\pi y, \quad q^*(y) = \sin \pi y + \sin 2\pi y + \sin 3\pi y.$$

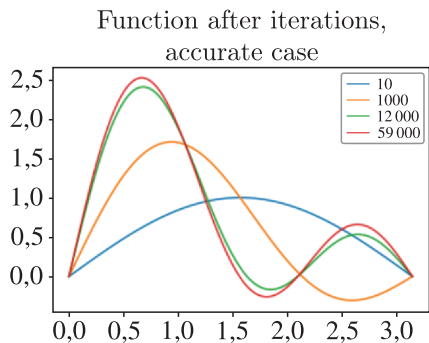


(a) Убывание величины функционала

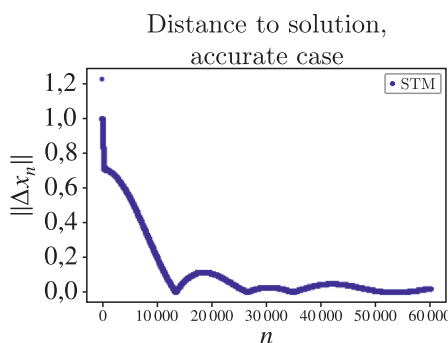


(b) Убывание нормы градиента

Рис. 13. Работа метода подобных треугольников: зависимость функционала и нормы градиента от количества итераций  $n$  с помощью ряда Фурье в случае, когда известно точное решение



(a) Аргумент на разных итерациях



(b) Расстояние до точного решения

Рис. 14. Работа метода подобных треугольников: приближенные решения на разных итерациях и зависимость расстояния до точного решения от количества итераций  $n$



Минимум действительно достигается, что показывает эффективность предложенного метода решения некорректно поставленной задачи. Результаты показаны на рис. 13 и 14.

### 8. Определение концепции шума

Для того случая (применение метода подобных треугольников и простейшего градиентного спуска с 10 слагаемыми в частичных суммах ряда Фурье), в котором это позволяют ресурсы, выполним эксперимент с 100 миллионами итераций и построим график в логарифмическом масштабе по обоим осям.

Цель эксперимента — выяснить, какая из двух концепций шума более адекватно описывает погрешность, возникающую при вычислении градиента.

Для аддитивного шума характерно накопление ошибки и переход к возрастанию невязки; минимум не достигается.

Для относительного шума характерна сходимость, аналогичная отсутствию шума, то есть наклонная прямая на графике в логарифмическом масштабе.

Характерные примеры для обеих концепций приведены в статье [Vasin, Gasnikov, Spokoyny, 2021].

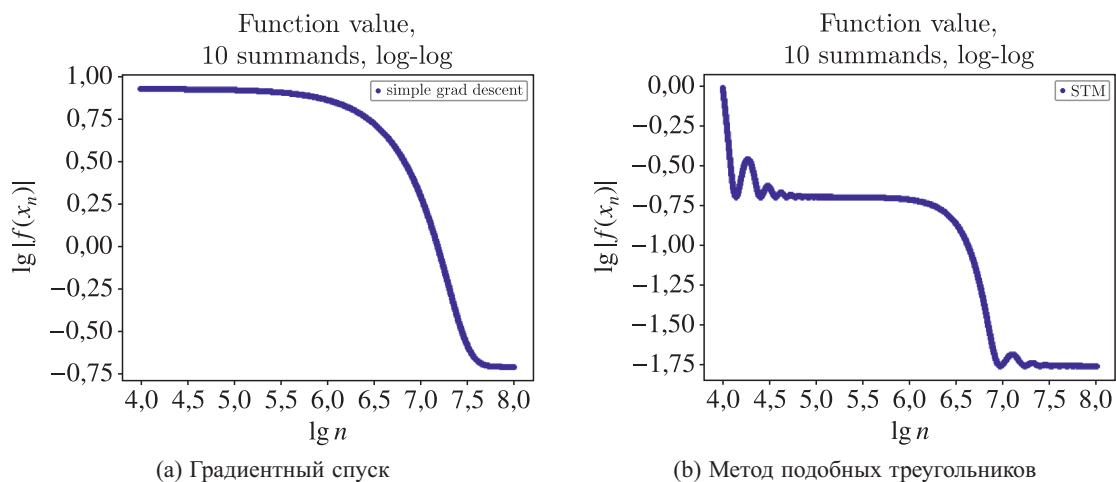


Рис. 15. Функционал в зависимости от количества итераций  $n$  в логарифмическом масштабе для градиентного спуска и метода подобных треугольников

График на рис. 15, *a* построен для простейшего градиентного спуска, график на рис. 15, *b* — для метода подобных треугольников в диапазоне  $10^4 \dots 10^8$ .

Достижимое при применении ускоренного метода значение функционала 0,01734 имеет тот же порядок, что и погрешность аппроксимации частичной суммой из 10 слагаемых ряда Фурье для  $f(y) = 100(y - y^2)$ , которая входит в выражение для функционала.

График в логарифмическом масштабе по обеим осям состоит из горизонтальных участков, участков убывания и небольших нарушений монотонности, что ожидаемо для ускоренных методов. Неускоренный метод сходится монотонно, причем также есть плато.

На участках убывания средний угловой коэффициент приблизительно равен  $-1$  для простейшего градиентного спуска и  $-2$  для метода подобных треугольников. Это соответствует общим оценкам сходимости  $O(\frac{1}{N})$  и  $O(\frac{1}{N^2})$  для обычных и ускоренных градиентных методов.

Наблюдаемый эффект свидетельствует о том, что на разных участках траектории метода неточность градиента имеет разный характер. Поэтому во всем пространстве шум градиента более адекватно описывается аддитивной концепцией, но локально точность аппроксимации

градиента частичной суммой ряда Фурье улучшается, что приводит к локальному ускорению сходимости и выходу графика на наклонную прямую, характерную для относительного шума, с последующей стабилизацией невязки.

## 9. Применение техники рестартов

Для улучшения результатов предпринимается попытка использовать рестарты по невязке функционала. Рестарт — это перезапуск метода с использованием последней точки в качестве новой начальной после достижения определенного условия на невязку (обычно — ее уменьшения вдвое), см., например, [Гасников, 2021].

Техника рестартов дает возможность увеличить скорость сходимости в случае, если функция является (локально) сильно выпуклой. Сформулируем ряд фактов о применении рестартов по невязке функционала в случае, если градиент имеет аддитивный шум.

**Лемма 2.** Если верхняя оценка шума удовлетворяет неравенству

$$\delta \leq \min \left\{ \frac{f(x_0) - f(x^*)}{24\tilde{R}}, \frac{\sqrt[4]{L\mu} \sqrt{f(x_0) - f(x^*)}}{4\sqrt[4]{2}} \right\},$$

то  $N = \sqrt{\frac{32L}{\mu}}$  шагов метода подобных треугольников гарантируют уменьшение невязки функционала вдвое:

$$f(x_N) - f(x^*) \leq \frac{1}{2} (f(x_0) - f(x^*)).$$

Здесь  $x_0$  — начальная точка,  $x^*$  — точка минимума,  $R = \|x_0 - x^*\|_2$ ,  $\tilde{R}$  — наибольшее расстояние до точки минимума среди точек, генерируемых методом подобных треугольников,  $L$  — константа Липшица градиента,  $\mu$  — (локальная) константа сильной выпуклости.

*Доказательство.* Используя следствие определения сильной выпуклости

$$\frac{\mu}{2} \|x - x^*\|_2^2 \leq f(x) - f(x^*)$$

и теорему из статьи [Vasin, Gasnikov, Spokoiny, 2021] о сходимости метода подобных треугольников

$$f(x_N) - f(x^*) \leq \frac{4LR^2}{N^2} + 3\tilde{R}\delta + \frac{N\delta^2}{2L},$$

получаем следующие оценки:

$$\begin{aligned} f(x_N) - f(x^*) &\leq \frac{4L\|x_0 - x^*\|^2}{N^2} + 3\tilde{R}\delta + \frac{N\delta^2}{2L} \leq \frac{8L}{\mu N^2} (f(x_0) - f(x^*)) + 3\tilde{R}\delta + \frac{N\delta^2}{2L} \leq \\ &\leq \frac{1}{4} (f(x_0) - f(x^*)) + \frac{3}{24} (f(x_0) - f(x^*)) + \frac{1}{8} (f(x_0) - f(x^*)) = \frac{1}{2} (f(x_0) - f(x^*)), \end{aligned}$$

что и требовалось.

**ЗАМЕЧАНИЕ.** Если оптимальное значение функции известно, то применение рестартов не требует знания констант или проверки условий, а только вычисления функции в точках  $x_k$  для обнаружения момента рестарта. Более того, глобальная сильная выпуклость тоже не требуется, а только локальная (в окрестности траектории).

**Теорема 3.** Если градиент выпуклой функции  $f(x)$  вычисляется с аддитивной неточностью  $\delta$ ,  $x_0$  — начальная точка,  $x^*$  — точка минимума,  $R = \|x_0 - x^*\|_2$ ,  $\tilde{R}$  — наибольшее расстояние

до точки минимума среди точек, генерируемых методом подобных треугольников,  $L$  – константа Липшица градиента,  $\mu$  – (локальная) константа сильной выпуклости, а

$$\varepsilon \geq \max \left\{ 24\tilde{R}\delta, \frac{16\sqrt{2}\delta^2}{\sqrt{L\mu}} \right\},$$

то достичь выполнения условия  $f(x) - f(x^*) < \varepsilon$  можно за не более чем

$$\sqrt{\frac{32L}{\mu}} \log_2 \frac{f(x_0) - f(x^*)}{\varepsilon}$$

итераций при использовании STM с рестартами при уменьшении функционала вдвое.

*Доказательство.* Поскольку  $f(x_0) - f(x^*) > \varepsilon$  (иначе никаких итераций не требуется), то

$$\delta \leq \frac{\varepsilon}{24\tilde{R}} \quad \text{и} \quad \delta^2 \leq \frac{\sqrt{L\mu}(f(x_0) - f(x^*))}{16\sqrt{2}},$$

то есть выполняются условия леммы 2. Поэтому после не более  $N = \sqrt{\frac{32L}{\mu}}$  итераций условие рестарта будет выполнено.  $\|x_N - x^*\|_2 \leq \tilde{R}$ , поэтому неравенство для  $\varepsilon$  снова будет выполнено. Количество рестартов не превышает  $\log_2 \frac{f(x_0) - f(x^*)}{\varepsilon}$ , что завершает доказательство.

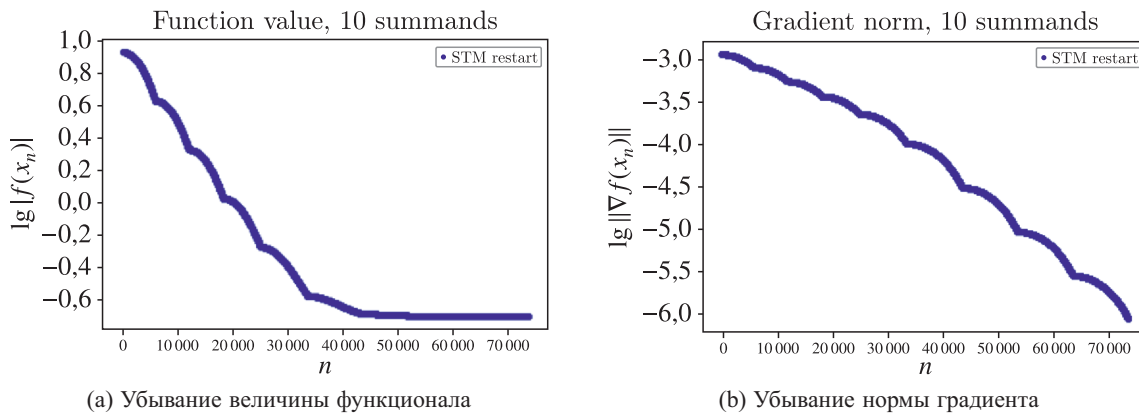


Рис. 16. Работа метода подобных треугольников с рестартами по невязке функционала: зависимость функционала и нормы градиента от количества итераций  $n$

Эксперименты показывают (рисунок 16), что применение рестартов не улучшает сходимость. Это подтверждает тот факт, что задача не является сильно выпуклой.

## 10. Дискретизация задачи

Применим другой подход к решению задачи оптимизации в бесконечномерном пространстве. А именно, вместо приближенного вычисления функционала и градиента на каждой итерации заменим задачу конечномерным (дискретным) приближением.

### 10.1. Дискретизация с использованием разностных схем

Заменим краевую задачу для дифференциального уравнения в частных производных ее разностным приближением:

$$\begin{cases} \frac{u_{i+1,j} - 2u_{i,j} + u_{i-1,j}}{h^2} + \frac{u_{i,j+1} - 2u_{i,j} + u_{i,j-1}}{h^2} + \kappa^2 u_{i,j} = 0, & 1 \leq i, j \leq P-1, \\ u_{P,j} = q_j, & 1 \leq j \leq P-1, \\ \frac{u_{1,j} - u_{0,j}}{h} = g_j, & 1 \leq j \leq P-1, \\ u_{i,0} = u_{i,P} = 0, & 0 \leq i \leq P. \end{cases}$$

При этом

$$q_j = q(jh), \quad g_j = g(jh), \quad u_{i,j} = u(ih, jh), \quad (\tilde{A}q)_j = u_{0,j};$$

$$\tilde{J}(q) = \frac{h}{2} \sum_{j=1}^{P-1} |u_{0,j} - f_j|^2.$$

Поскольку  $u_{i,j}$  — решение системы линейных уравнений с конкатенацией  $q, g$  и нулевого вектора в правой части, данный функционал является квадратичным. Он достигает минимума при условии равенства нулю всех слагаемых, откуда можно найти  $q_j$ .

$$\begin{cases} \frac{u_{i+1,j} - 2u_{i,j} + u_{i-1,j}}{h^2} + \frac{u_{i,j+1} - 2u_{i,j} + u_{i,j-1}}{h^2} + k^2 u_{i,j} = 0, & 1 \leq i, j \leq P-1, \\ u_{0,j} = f_j, & 1 \leq j \leq P-1, \\ \frac{u_{1,j} - u_{0,j}}{h} = g_j, & 1 \leq j \leq P-1, \\ u_{i,0} = u_{i,P} = 0, & 0 \leq i \leq P. \end{cases}$$

Данная система также имеет единственное решение, причем оно может быть найдено точно последовательным вычислением при  $i = 1 \dots P$  без необходимости применения громоздких методов для ранга  $(P+1)^2$ .

Легко заметить, что каждому решению дискретизованной задачи соответствует бесконечное семейство функций из  $H$ . Это семейство имеет вид

$$q(y) + \sum_{m=1}^{\tilde{M}} c_m \sin mPy, \quad c_m \in \mathbb{R}.$$

Соответственно, нельзя говорить о сходимости решения дискретизованной задачи к решению точной задачи.

### 10.2. Дискретизация с использованием рядов Фурье

Воспользуемся теперь методом Фурье, который ранее применялся для приближенных вычислений. Применим формулы из подпараграфа 7.2.  $M$  — количество слагаемых в конечных суммах, аппроксимирующих ряды Фурье.

$$f(y) \rightarrow \tilde{f} = \sum_{m=1}^M f_m \sin \pi m y, \quad g(y) \rightarrow \tilde{g} = \sum_{m=1}^M g_m \sin \pi m y, \quad q(y) \rightarrow \tilde{q} = \sum_{m=1}^M q_m \sin \pi m y.$$

Для данных функций оператор  $A$  вычисляется точно:

$$(A\bar{q})(y) = \sum_{m=1}^M \left( \frac{q_m}{\operatorname{ch} \gamma_m} - \frac{g_m}{\gamma_m} \operatorname{th} \gamma_m \right) \sin \pi m y.$$

Дискретизованная задача минимизации:

$$\tilde{J}_M(\bar{q}) = \frac{1}{2} \|A\bar{q} - \tilde{f}\|^2 = \frac{1}{4} \sum_{m=1}^M \frac{1}{\operatorname{ch}^2 \gamma_m} \left( q_m - g_m \frac{\operatorname{sh} \gamma_m}{\gamma_m} - f_m \operatorname{ch} \gamma_m \right)^2 \rightarrow \min_{\bar{q}}.$$

Эта задача имеет точное решение:

$$\forall m \in \{1, \dots, M\} \quad q_m = f_m \operatorname{ch} \gamma_m + g_m \frac{\operatorname{sh} \gamma_m}{\gamma_m}.$$

Она является квадратичной,

$$\frac{\lambda_{\max}}{\lambda_{\min}} = \frac{\operatorname{ch}^2 \gamma_M}{\operatorname{ch}^2 \gamma_1} \sim \frac{1}{4} e^{2\pi M}.$$

Соответственно, при больших  $M$  задача является плохо обусловленной;  $q_m$  не стремится к 0, поэтому решение задачи минимизации  $J(q)$  не может быть представлено как предел последовательности решений дискретизованных задач при  $M \rightarrow \infty$ .

### 10.3. Применимость к решению задачи

Из формулы точного решения дискретизованной задачи (в подпараграфе 10.2) следует, что оно очень велико по норме. Поэтому при его вычислении в машинной арифметике накапливаются большие погрешности. Например, при подстановке в приближенный функционал решения задачи, дискретизованной с помощью разностных схем при  $h = 0,01$ , погрешность превышает  $10^{100}$ .

Соответственно, дискретизация задачи не помогает получить приемлемое качество решения.

## 11. Заключение

Работа посвящена исследованию применения градиентных методов к решению некорректно поставленной задачи Коши для уравнения Гельмгольца путем сведения ее к задаче оптимизации в гильбертовом пространстве.

Предложены способы решения задачи ускоренным градиентным методом с использованием для вычисления градиента разностных схем или разложения в ряды Фурье, доказаны теоремы 1 и 2 о вычислительной эффективности данных способов: достигается невязка по функционалу

$$O\left(\frac{R^2}{N^2} + \frac{\tilde{R}}{P} + \frac{N}{P^2}\right)$$

после  $O(P^6 N)$  арифметических операций для разностных схем или

$$O\left(\frac{R^2}{N^2} + \frac{\tilde{R}}{M} + \frac{N}{M^2}\right)$$

после  $O((N+S)M)$  арифметических операций для рядов Фурье, где  $P = \frac{1}{h}$  — количество точек на сетке для отрезка,  $M$  — количество слагаемых в суммах Фурье,  $N$  — количество итераций метода подобных треугольников. Проведены эксперименты, сделан вывод о том, что использование рядов Фурье значительно повышает скорость и качество решения задачи.

Выяснено, что погрешность вычисления градиента более адекватно описывается аддитивной концепцией шума. Это подтверждается как теоретическими выкладками, так и экспериментом. При этом погрешность не накапливается.

Доказана теорема 3 о сходимости метода подобных треугольников с рестартами по невязке функционала: если  $\varepsilon \geq \max \left\{ 24\bar{R}\delta, \frac{16\sqrt{2}\delta^2}{\sqrt{L\mu}} \right\}$ , то достичь выполнения условия  $f(x) - f(x^*) < \varepsilon$  можно за не более чем  $\sqrt{\frac{32L}{\mu}} \log_2 \frac{f(x_0) - f(x^*)}{\varepsilon}$  итераций. Проверено, что в исследуемой задаче применение рестартов не повышает скорость сходимости.

Полученные результаты показывают возможность эффективного решения обратных задач математической физики путем применения ускоренных методов оптимизации. Однако аддитивный характер шума ограничивает достижимое качество решения.

Авторы выражают благодарность д. ф.-м. н. Максиму Александровичу Шишленину за помощь в постановке задачи.

## Список литературы (References)

- Васильев Ф. П.* Методы оптимизации. Ч. вторая: Оптимизация в функциональных пространствах. Регуляризация. Аппроксимация. — М.: МЦНМО, 2011.  
*Vasilyev F. P.* Methodi optimizacii. Chast' vtoraya: Optimizaciya v funkcional'nykh prostranstvakh. Regularizaciya. Approssimaciya [Optimization methods, part two: Optimization in function spaces. Regularization. Approximation]. — Moscow: MCCME, 2011 (in Russian).
- Гасников А. В.* Современные численные методы оптимизации. Метод универсального градиентного спуска: учебное пособие. — 2-е изд., испр. — М.: МЦНМО, 2021. — 272 с.  
*Gasnikov A. V.* Sovremenii chislenii metodi optimizacii. Metod universal'nogo gradientnogo spuska [Modern numerical optimization methods. Universal Gradient Descent Method: Textbook]. — Moscow: MCCME, 2021. — 272 p. (in Russian).
- Демьянов В. Ф.* Минимакс: дифференцируемость по направлениям. — Л.: Изд-во Ленингр. ун-та, 1974.  
*Demjanov V. F.* Minimaks: differenciruyemost' po napravleniyam [Minimax: differentiability in directions]. — Leningrad: isd-vo Leningr. un-ta, 1974 (in Russian).
- Рябенский В. С.* Введение в вычислительную математику: учебное пособие. — 2-е изд., испр. — М.: ФИЗМАТЛИТ, 2000. — 296 с.  
*Ryabenkiy V. S.* Vvedeniye v vychislitel'nyuyu matematiku: uchebnoye posobiye [Introduction to Computational Mathematics: Textbook]. — Moscow: FIZMATLIT, 2000. — 296 p. (in Russian).
- Тихонов А. Н., Самарский А. А.* Уравнения математической физики: учебное пособие. — 6-е изд., испр. и доп. — М.: Изд-во МГУ, 1999.  
*Tikhonov A. N., Samarskiy A. A.* Uravneniya matematicheskoy fiziki: uchebnoye posobiye [Equations of Mathematical Physics: Textbook]. — Moscow: Izd-vo MGU, 1999 (in Russian).
- Danskin J. M.* The theory of Max Min. — Berlin: Springer, 1967.
- Kabanikhin S., Shishlenin M., Nurseitov D., Nurseitova A., Kasenov S.* Comparative analysis of methods for regularizing an initial boundary value problem for the Helmholtz equation // Journal of Applied Mathematics. — 2014. — Vol. 2014, Article ID 786326. — 7 p.
- Kabanikhin S. I.* Inverse and ill-posed problems: theory and applications. — Berlin: Walter de Gruyter GmbH & Co, 2012. — 459 p.
- Matyukhin V. et al.* Convex optimization with inexact gradients in Hilbert space and applications to elliptic inverse problems // International Conference on Mathematical Optimization Theory and Operations Research. — Springer, Cham, 2021. — P. 159–175.
- Polyak B. T.* Introduction to Optimization. — New York: Optimization Software, 1987.
- Polyak B. T.* Iterative algorithms for singular minimization problems // Nonlinear Programming. — 1981. — Vol. 4. — P. 147–166.
- Vasin A., Gasnikov A., Spokoiny V.* Stopping rules for accelerated gradient methods with additive noise in gradient // arXiv preprint. — 2021. — <https://arxiv.org/abs/2102.02921>