

УДК 519.853.62

ЧИСЛЕННОЕ ИССЛЕДОВАНИЕ СХОДИМОСТИ МЕТОДА ГРАДИЕНТНОГО СПУСКА НА МОДЕЛЬНЫХ ФУНКЦИЯХ

Ануфриева Е.В., студент

*Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь*

Примичева З.Н. – канд. физ.-мат. наук, доцент

Аннотация. Проведено численное исследование сходимости метода градиентного спуска на модельных функциях с различными свойствами гладкости и выпуклости. Проанализировано влияние числа обусловленности на скорость сходимости и траекторию метода. Построена визуализация итерационных процессов с использованием Python и NumPy.

Ключевые слова. Градиентный спуск, минимизация функций, численные методы оптимизации, сходимость итерационных методов, число обусловленности, выпуклые функции, функция Розенброка, безусловная оптимизация, скорость сходимости, вычислительный эксперимент, машинное обучение, итерационные алгоритмы.

Задачи минимизации функций многих переменных занимают особое место в современном математическом анализе и его приложениях. Они возникают при решении систем нелинейных уравнений, вариационном исчислении, математическом моделировании физических процессов, а также в методах машинного обучения. Среди других способов решения задач метод градиентного спуска является базовым алгоритмом [1, с. 52–58]. Несмотря на простоту реализации и низкие вычислительные затраты на одну итерацию, вопрос о скорости сходимости и устойчивости метода остается критически важным для практического применения [2, с. 15–20].

Целью работы является численное исследование сходимости метода градиентного спуска на модельных функциях с различными свойствами гладкости и выпуклости. Особое внимание уделяется анализу влияния числа обусловленности задачи на скорость убывания градиента и траекторию движения в точку минимума.

Для достижения цели в работе решаются следующие задачи:

1. Анализ теоретических условий сходимости метода градиентного спуска для вогнутых функций.
2. Выбор набора модельных функций, обладающих различными свойствами обусловленности.
3. Проведение вычислительного эксперимента для построения итерационных последовательностей и оценка скорости сходимости на каждом этапе.
4. Сравнение полученных численных результатов с теоретическими оценками скорости сходимости.

Рассматривается задача безусловной минимизации дифференцируемой функции $f: R^n \rightarrow R$:

$$f(x) \rightarrow \min, x \in R^n, \quad (1)$$

где функция f ограничена снизу и обладает непрерывными частными производными первого порядка.

Метод градиентного спуска строит последовательность приближений $\{x^k\}_{k=0}^{\infty}$ по правилу [6, 4:07]:

$$x_i^{(k+1)} = x_i^{(k)} - a_k \frac{\partial f(x^{(k)})}{\partial x_i}, \quad k = 0, 1, 2, \dots, n; \quad i = 1, 2, \dots, n, \quad (2)$$

где $\nabla f(x_k) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n} \right)^T \in R^n$ - градиент функции в точке x^k , указывающий направление наискорейшего роста функции; $a_k > 0$ - параметр шага, определяющий длину перемещения вдоль антиградиента.

Метод градиентного спуска является одним из наиболее распространённых алгоритмов оптимизации. однако его практическая эффективность напрямую зависит от свойств минимизируемой функции и выбора параметров алгоритма. для обоснования применимости метода и оценки скорости его сходимости сформулируем следующую теорему, устанавливающую достаточные условия сходимости для вогнутых функций с липшицевым градиентом

Теорема (о сходимости метода градиентного спуска). Пусть функция $f: R^n \rightarrow R$ удовлетворяет следующим условиям: 1) функция $f(x)$ является выпуклой вниз на некотором интервале (рис. 1), то есть для любых значений аргумента $x, y \in R^n$ и для любого числа $t \in [0, 1]$ имеет место неравенство Йенсена

$$f(tx + (1 - t)y) \leq tf(x) + (1 - t)f(y); \quad (3)$$

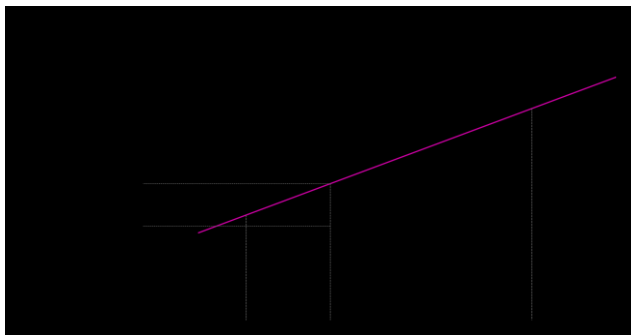


Рисунок 1 – График вогнутой функции, определяемая неравенством Йенсена

2) Градиент функции $f(x)$ удовлетворяет условию липшица, то есть существует такая положительная константа l , что для любых точек x, y из отрезка $[a, b]$ выполняется [1, с. 89] неравенство

$$|f(x) - f(y)| \leq L|x - y|; \quad (4)$$

3) Постоянный шаг a_k выбирается из интервала $(2/L)$. Тогда метод градиентного спуска (2) сходится к стационарной точке x^* , в которой $\nabla f(x^*) = 0$, при этом для значения функции справедлива оценка:

$$f(x_k) - f(x^*) \leq \left(1 - \frac{\mu}{L}\right)^k [f(x_0) - f(x^*)]. \quad (5)$$

Число k , определяемое из условия

$$k = \frac{L}{\mu} \geq 1, \quad (6)$$

называется числом обусловленности задачи и определяет скорость сходимости: чем больше k , тем медленнее сходимость [3, р. 45].

Таблица 1 составлена на основе принципа «от простого к сложному» и включает три класса функций, которые позволяют всесторонне исследовать свойства метода градиентного спуска. Первой выбрана простейшая квадратичная функция (сфера) $f(x) = |x|^2$. Линии уровня представляют собой концентрические окружности. Число обусловленности для неё равно единице. Эта функция служит базовым случаем, на котором градиентный спуск демонстрирует наилучшее поведение, сходясь к минимуму по прямой траектории за минимальное число итераций.

Вторая функция — квадратичный овраг $f(x) = x_1^2 + kx_2^2$ при $k \gg 1$. Оставаясь вогнутой, функция имеет линии уровня в виде сильно вытянутых эллипсов. На такой функции наглядно проявляется главный недостаток градиентного спуска: при большом k метод движется «зигзагом», перпендикулярно дну оврага, что приводит к замедлению сходимости.

Третьей выбрана функция Розенброка. Это невыпуклая функция со сложным ландшафтом, имеющая узкую параболическую долину, для которой x_1 и x_2 — координаты входного вектора $x = (x_1, x_2)$, то есть две независимые переменные, по которым находится минимум функции. Глобальный минимум располагается внутри этого оврага. Данная функция позволяет проверить метод в условиях, близких к реальным задачам оптимизации: алгоритм должен сначала быстро попасть в овраг, а затем медленно двигаться по его дну в точку минимума.

Таблица 1 – Функции и их свойства

Функция	Формула	Свойства	Цель исследования
Квадратичная (сфера)	$f(x) = x ^2$	Сильно вогнута, $\kappa = 1$	Базовый случай быстрой сходимости
Квадратичная (овраг)	$f(x) = x_1^2 + kx_2^2, k \gg 1$	Вогнута	Влияние числа обусловленности на траекторию
Розенброка	$f(x) = (1 - x_1)^2 + 100(x_2 - x_1^2)^2$	Сложный ландшафт	Поведение метода в "овражных" функциях

Рассмотрим задачу поиска минимума функции $f(x, y)$, моделирующей рельеф морского дна, где

- сокровище — точка глобального минимума $x^* = \operatorname{argmin} f(x)$;
 - лодка — текущее приближение $x_k = (x_k, y_k)^T$;
 - туман — отсутствие глобальной информации о функции; лодка «ощущает» только локальный наклон (градиент);
 - компас — вектор антиградиента $\nabla f(x_k)$, указывающий направление наискорейшего спуска;
 - длина шага — параметр a_k , определяющий расстояние перемещения за одну итерацию.
- Движение лодки описывается дискретной динамической системой:

$$\begin{pmatrix} x_{k+1} \\ y_{k+1} \end{pmatrix} = \begin{pmatrix} x_k \\ y_k \end{pmatrix} - a_k \begin{pmatrix} \frac{df}{dx}(x_k, y_k) \\ \frac{df}{dy}(x_k, y_k) \end{pmatrix}. \quad (7)$$

На рисунке 2 представлены линии уровня функции $f(x) = x_1^2 + 10x_2^2$. На рисунке 3 изображена поверхность данной функции. На рисунке 4 представлен график погрешности функции.

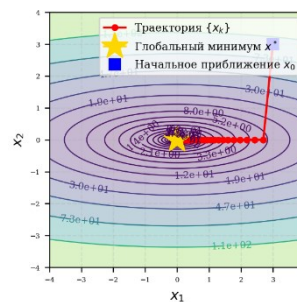


Рисунок 2 – Линии уровня функции $f(x) = x_1^2 + 10x_2^2$

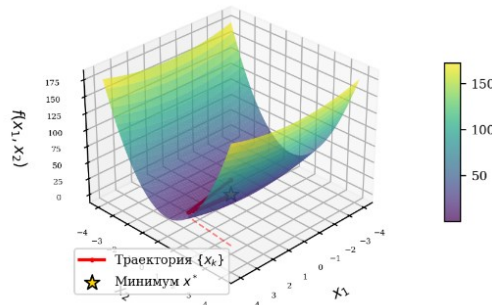


Рисунок 3 – Поверхность $f(x_1, x_2)$ с траекторией спуска

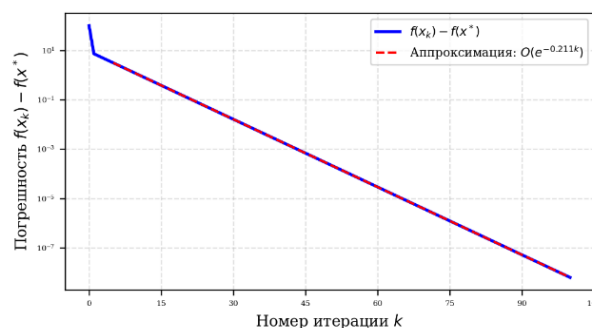


Рисунок 4 – График логарифмической погрешности

Проведем вычислительный эксперимент:

- Критерий остановки: $\|\nabla f(x_k)\| < \varepsilon$ или $|f(x_{k+1}) - f(x_k)| < \varepsilon$, где $\varepsilon = 10^{-6}$.

- Параметры: Фиксированный шаг $a = 0,1$ начальная точка $x_0 = (3,3)^T$.

Рассмотрим функцию $f(x) = x_1^2 + kx_2^2$. Рисунок 5 показывает различные поведения функции при различных числах обусловленности. На рисунке 6 изображены поверхности функции при $k = 1$ и $k = 10$. На рисунке 7 изображена скорость сходимости метода градиентного спуска при различных $k = 1$ и шаге $k = 10$. На рисунке 8 изображены профили функции. В таблице 2 представлены траектории градиентного спуска для различных значений числа обусловленности.

Таблица 2 – Траектории градиентного спуска для различных значений

k (число обусловленности)	Число итераций до сходимости	Наблюдение
1 (сфера)	~15	Прямая траектория к минимуму
10	~45	Слабые колебания
100 (овраг)	>200	Выраженный зигзаг, медленная сходимость

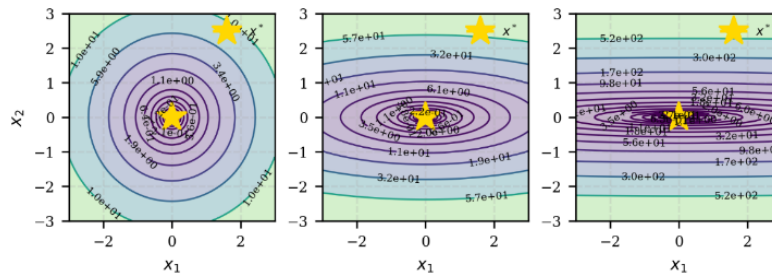


Рисунок 5 – Поверхность $f(x) = x_1^2 + kx_2^2$ при разных k

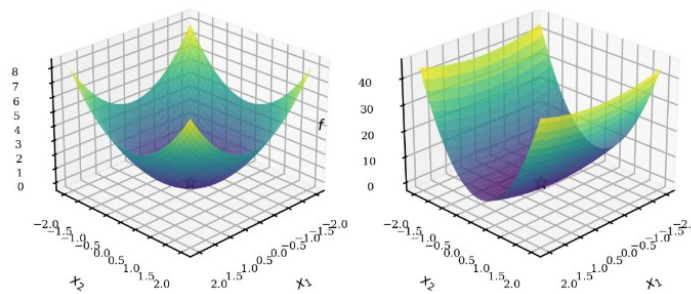


Рисунок 6 – Поверхность $f(x_1, x_2)$ при $k = 1$ и $k = 10$

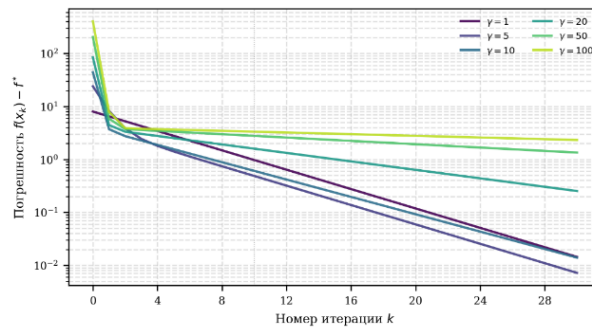


Рисунок 7 – Влияние k на скорость сходимости при $a = 0.050$

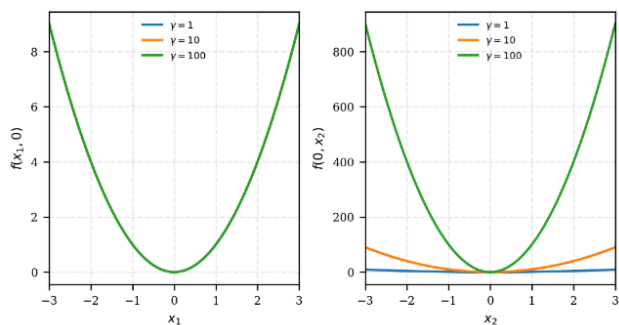


Рисунок 8 – Профили функции $f(x_1, x_2) = x_1^2 + kx_2^2$

Для функции Розенброка метод демонстрирует типичное «застывание» в овраге: после быстрого попадания в узкую долину сходимость вдоль дна оврага становится крайне медленной из-за малого градиента в направлении к минимуму. На рисунке 9 изображен график функции Розенброка. На рисунке 10 отображена сходимость градиентного спуска на функции Розенброка.

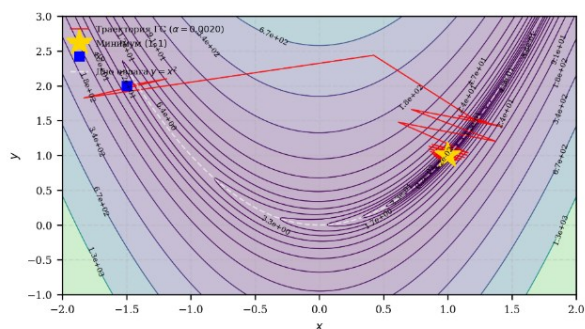


Рисунок 9 – График функции Розенброка

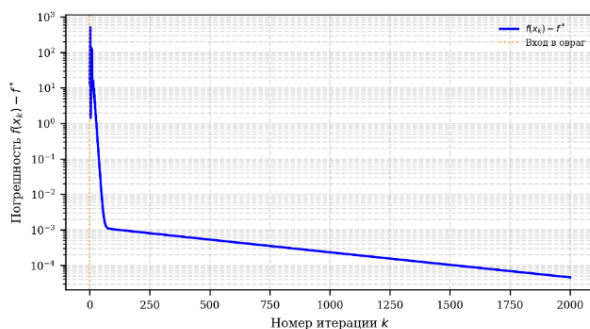


Рисунок 10 – Сходимость градиентного спуска на функции Розенброка

В ходе выполнения работы проведено численное исследование сходимости классического метода градиентного спуска на классе модельных функций. На основе анализа полученных результатов можно сформулировать следующие основные выводы:

1. Подтверждение теоретических оценок сходимости.

Экспериментально доказана зависимость скорости сходимости метода от числа обусловленности к задачи минимизации. Для квадратичной функции $f(x) = x_1^2 + kx_2^2$ установлено, что количество итераций, необходимых для достижения заданной точности ε , растёт пропорционально k , что находится в полном согласии с теоретической асимптотической оценкой $O(k \log(1/\varepsilon))$ для сильно вогнутых функций с липшицевым градиентом [1, с. 95].

Влияние геометрии функции на динамику метода.

Визуализация траекторий на функциях овражного типа (в частности, функции Розенброка) продемонстрировала характерный двухфазный характер сходимости:

- Первый этап: на начальном участке наблюдается быстрое убывание функции, так как метод движется к дну оврага в области с большими значениями градиента.
- Второй этап: при приближении к точке минимума сходимость существенно замедляется. это

связано с тем, что вдоль дна оврага градиент становится очень малым, а сам рельеф характеризуется сильной вытянутостью.

2. Важность выбора параметра шага.

Установлено, что устойчивость итерационного процесса $x^{k+1} = x^k - a_k \nabla f(x^k)$ существенно зависит от величины шага α . Превышение верхнего порога $\alpha > 2/L$ приводит к расходимости или колебательному режиму, тогда как выбор чрезмерно малого α влечёт неоправданное увеличение вычислительных затрат. Это обосновывает необходимость применения адаптивных стратегий выбора шага или методов с перестройкой метрики.

3. Ограничения метода на невыпуклых задачах.

На примере функции Розенброка показано, что в отсутствие глобальной выпуклости метод градиентного спуска не гарантирует попадания в глобальный минимум и может сходиться к локальным экстремумам или седловым точкам, где $\nabla f(x^*) = 0$. Это ограничивает применимость базового алгоритма в задачах глобальной оптимизации без модификаций (стохастизация, методы второго порядка).

Результаты работы систематизируют данные о поведении градиентного спуска на тестовых функциях и предоставляют наглядную иллюстрацию абстрактных понятий выпуклого анализа. Новизна работы заключается в систематизации данных о поведении метода градиентного спуска на конкретных функциях и визуализации зависимости скорости сходимости от геометрических свойств ландшафта функции потерь. Разработанный вычислительный эксперимент может быть использован в учебном процессе для демонстрации связи между геометрией и динамикой итерационных методов.

Проведённое исследование открывает различные направления для развития работы. Во-первых, сравнительный анализ градиентного спуска с методами второго порядка, такими как метод Ньютона или BFGS. Во-вторых, актуальным является исследование стохастических модификаций (SGD) и адаптивных алгоритмов (Adam, RMSprop) при работе с данными, что особенно важно для задач машинного обучения. В-третьих, применение методов преобуславливания может существенно улучшить сходимость на плохо обусловленных задачах, снизив влияние высокого числа обусловленности на скорость работы алгоритма. Реализация этих направлений позволит расширить область применимости градиентных методов и повысить их эффективность в практических задачах.

Список использованных источников:

1. Поляк Б.Т. Введение в оптимизацию [Электронный ресурс]: учеб. пособие / Поляк Б.Т. – Изд-во Моск. ун-та: Наука, 1983. – 386с.
2. Воронцова Е.А. Методы выпуклой оптимизации [Электронный ресурс]: учеб. пособие / Воронцова Е.А. – Изд-во Моск. ун-та: МФТИ, 2021. – 364с.
3. Numerical Optimization [Электронный ресурс] / Jorge Nocedal, Stephen Wright. – Second Edition. – Springer-Verlag New York, 2006. – Электронные данные. – Режим доступа: <https://link.springer.com/content/pdf/10.1007/978-0-387-40065-5.pdf> – Дата доступа: 06.03.2026
4. An automatic method for finding the greatest or least value of a function [Электронный ресурс] / H. H. Rosenbrock // The Computer Journal. – 1960. – Vol. 3, № 3. – P. 175–184. – Электронные данные. – Режим доступа: <https://academic.oup.com/comjnl/article-abstract/3/3/175/345501> – Дата доступа: 12.03.2026
5. Array programming with NumPy [Электронный ресурс] / Charles R. Harris [et al.] // Nature. – 2020. – Электронные данные. – Режим доступа: <https://www.nature.com/articles/s41586-020-2649-2> – Дата доступа: 13.03.2026

UDC 519.853.62

NUMERICAL STUDY OF THE CONVERGENCE OF THE GRADIENT DESCENT METHOD ON MODEL FUNCTIONS

Anufrieva E.V., student

Belarusian State University of Informatics and Radioelectronics, Minsk, Republic of Belarus

Primicheva Z.N. – PhD in Physics and Mathematics

Annotation. A numerical study of the convergence of the gradient descent method on model functions with different properties of smoothness and convexity is carried out. The influence of the conditioning number on the convergence rate and trajectory of the method is analyzed. Visualizations of iterative processes using Python and NumPy are built.

Keywords. Gradient descent, minimization of functions, numerical optimization methods, convergence of iterative methods, conditionality number, convex functions, Rosenbrock function, unconditional optimization, convergence rate, computational experiment, machine learning, iterative algorithms.