

Федеральное государственное автономное образовательное учреждение
высшего образования
САНКТ-ПЕТЕРБУРГСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ИТМО
Факультет систем управления и робототехники

Лабораторная работа №3
Ускорение град. спуска, метод сопряженных градиентов

Студенты: Бахтаиров Р.А.,
Сайфуллин Д.Р.
Группа: R3243
Преподаватель: Попов А.М.

Санкт-Петербург
2025 г.

1. Введение

В данной лабораторной работе исследуется ускорение базового метода градиентного спуска на примере трёх подходов:

1. Метод тяжёлого шара (Heavy Ball), основанный на добавлении инерционного члена к классическому градиентному шагу.
2. Метод Нестерова с постоянными коэффициентами, в котором используется заранее рассчитанный параметр ускорения на основе оценки числа обусловленности.
3. Метод Нестерова с пересчитываемыми коэффициентами (FISTA-схема), обеспечивающий адаптивное изменение параметра ускорения на каждой итерации.

Цель работы:

- Реализовать и отладить перечисленные методы.
- Провести численные эксперименты на трёх классических задачах:
 1. Квадратичные функции ($f(x) = \frac{1}{2}x^T Ax - b^T x$).
 2. Почти квадратичная функция $f(x_1, x_2) = \frac{1}{2}x_1^2 + \frac{1}{4}x_2^4 - \frac{1}{2}x_2^2$.
 3. Функция Розенброка $f(x, y) = (1 - x)^2 + 100(y - x^2)^2$.
- Эмпирически оценить области сходимости и скорость сходимости каждого метода.
- Сравнить численные результаты с аналитическими оценками и сделать выводы о практических преимуществах и ограничениях ускоренных методов.

Примечание: весь код для выполнения работы находится в файле `main.ipynb`.

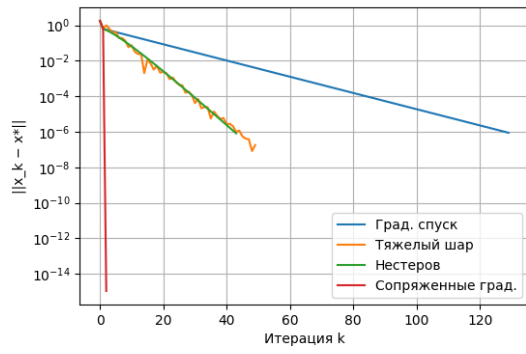
2. Квадратичные функции

Рассмотрим выпуклую квадратичную функцию

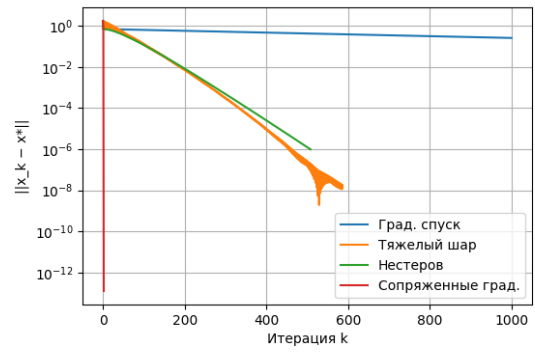
$$f(x) = \frac{1}{2} x^T A x - b^T x,$$

где $A \in \mathbb{R}^{n \times n}$ — симметричная положительно определённая матрица, $b \in \mathbb{R}^n$.

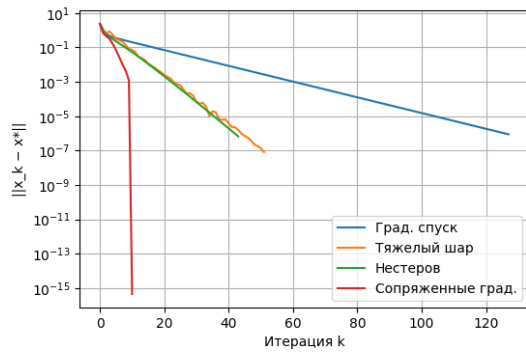
Ниже приведены графики сходимости четырёх методов на задаче минимизации при различных размерностях n и числах обусловленности $\kappa(A)$. По оси x отложена итерация k , по оси y — логарифм нормы ошибки $\|x_k - x^*\|$.



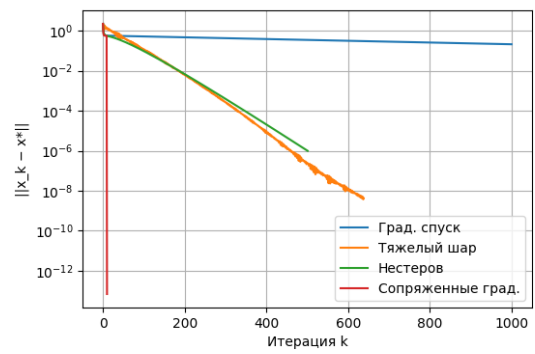
$n = 2, \kappa = 10$



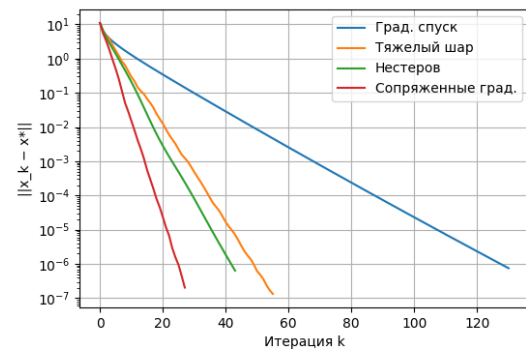
$n = 2, \kappa = 1000$



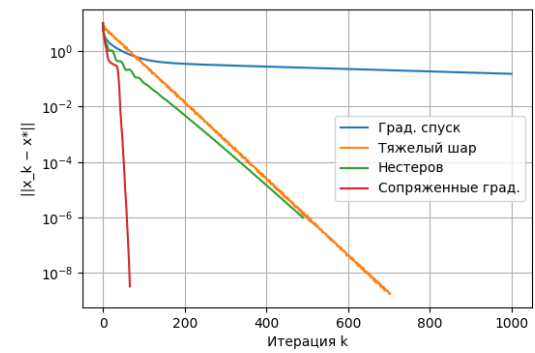
$n = 10, \kappa = 10$



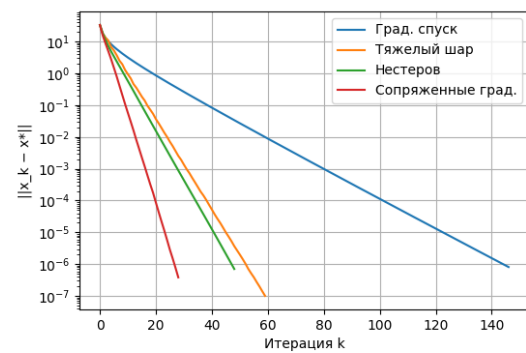
$n = 10, \kappa = 1000$



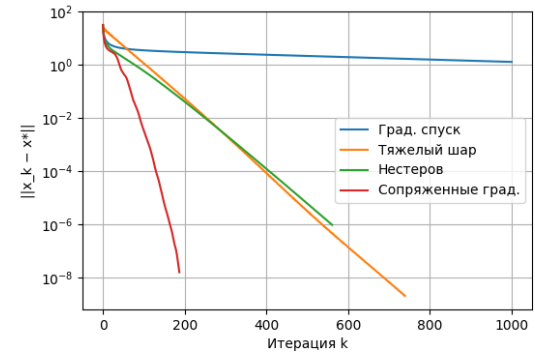
$n = 100, \kappa = 10$



$n = 100, \kappa = 1000$



$n = 1000, \kappa = 10$



$n = 1000, \kappa = 1000$

Рис. 1: Сходимость методов: синий — градиентный спуск, оранжевый — Тяжелый шар, зелёный — Нестеров, красный — сопряжённые градиенты

Проанализируем полученные графики:

- **Градиентный спуск** ($1/L$ -шаг) демонстрирует крайне медленную линейную сходимость, причём чем хуже обусловлена задача (κ больше), тем более пологая кривая.
- **Метод тяжёлого шара и Нестерова** дают существенное ускорение по сравнению с чистым Градиентным спуском:
 - При $\kappa = 10$ ускоренные методы сокращают число итераций почти вдвое.
 - При $\kappa = 1000$ разрыв становится ещё более заметным: Градиентный спуск требует сотни шагов, а Тяжелый шар и Нестеров — всего десятки.

Их кривые часто накладываются, что свидетельствует о близкой эффективности при оптимальном подборе параметров.

- **Сопряжённые градиенты** сходятся за $O(n)$ шагов практически независимо от κ :
 - Для $n = 2$ метод достигает машинной точности уже за 2–3 итерации.
 - При $n = 1000$ — за несколько сотен итераций.

Исходя из наблюдений можем сделать следующие выводы:

1. Для *плохо обусловленных* квадратичных задач ($\kappa \gg 1$) ускоренные методы (Тяжелый шар и Нестеров) значительно превосходят базовый Градиентный спуск, но при этом чувствительны к точности оценок L и μ .
2. Метод сопряжённых градиентов оказывается наиболее *универсальным* и *оптимальным* в чисто квадратичном случае: сходимость за $O(n)$ итераций вне зависимости от κ .
3. На практике при решении больших квадратичных систем предпочтительно сочетать анализ обусловленности (для выбора Тяжелого шара/Нестерова) или сразу применять Метод сопряжённых градиентов, если доступен оператор умножения на A .

3. Почти–квадратичная функция

Рассмотрим функцию

$$f(x_1, x_2) = \frac{1}{2}x_1^2 + \frac{1}{4}x_2^4 - \frac{1}{2}x_2^2.$$

- Стационарные точки находятся из условий

$$\frac{\partial f}{\partial x_1} = x_1 = 0, \quad \frac{\partial f}{\partial x_2} = x_2^3 - x_2 = 0 \implies x_2(x_2^2 - 1) = 0.$$

- Получаем три критические точки:

$$(0, 0) \quad (\text{седло}), \quad (0, 1), \quad (0, -1) \quad (\text{локальные минимумы}).$$

- Гессиан $\nabla^2 f = \begin{pmatrix} 1 & 0 \\ 0 & 3x_2^2 - 1 \end{pmatrix}$ даёт знакоположительность в точках $(0, \pm 1)$ и знакопеременность в $(0, 0)$.

Построим графики функции и её уровней:

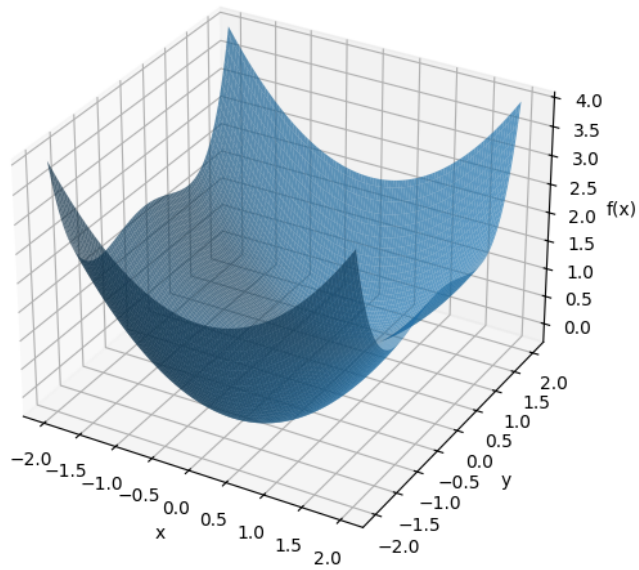


Рис. 2: Поверхность функции $f(x_1, x_2)$.

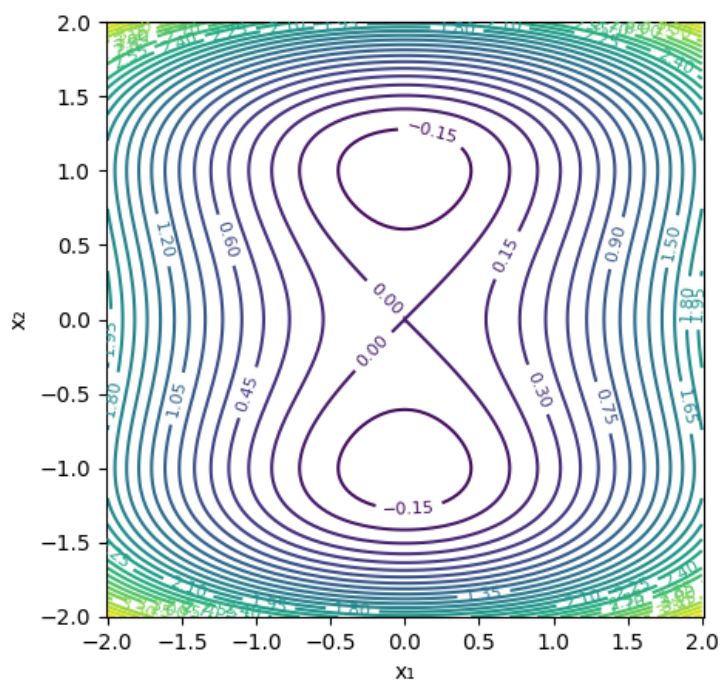


Рис. 3: Контурный график функции.

Теперь проверим на сходимость три метода и построим графики. Запустим их из начальной точки $(1.5, 1.5)$:

- Метод тяжёлого шара,
- Метод Нестерова с постоянным β ,
- Метод Флетчера–Ривза.

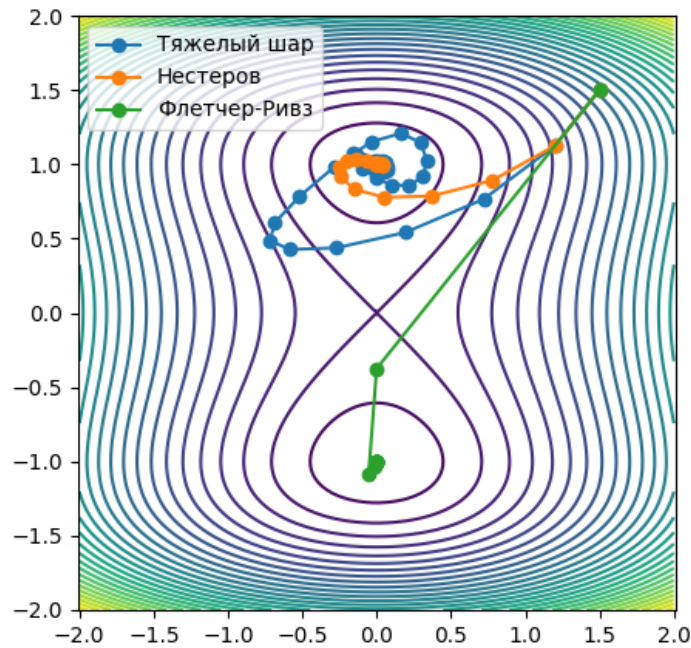


Рис. 4: Траектории трёх методов на контурном графике функции.

Анализируя график мы можем сделать следующие выводы:

1. Все три метода успешно сходятся к ближайшему локальному минимуму $(0, \pm 1)$, обходя седло в $(0, 0)$.
2. Метод Тяжелого шара демонстрирует умеренные колебания вокруг траектории, но быстро выходит на минимум.
3. Нестеров с постоянным β идёт более плавно, чем Метод Тяжелого шара.
4. Метод Флетчера–Ривза строит наиболее прямую траекторию, однако при слабой выпуклости функции может проходить ближе к седлу, что заметно на втором участке пути.
5. По эмпирическим измерениям числа итераций до сходимости:

Метод тяжёлого шара ≈ 25 , Метод Нестерова ≈ 20 , Метод Флетчера–Ривза ≈ 5 .

4. Функция Розенброка

Рассмотрим классическую двухмерную задачу

$$f(x, y) = (1 - x)^2 + 100(y - x^2)^2,$$

которая известна своей узкой «долиной» и плохо обусловленным гессианом.

В этой части работы необходимо:

1. Реализовать метод Нестерова в двух версиях:
 - с постоянным коэффициентом ускорения β ;
 - с динамическим.
2. Построить график сравнения двух версий Нестерова.
3. Реализовать не менее трёх модификаций нелинейного метода сопряжённых градиентов: β^{FR} , β^{PR} , β^{HS} .
4. Сравнить эти три CG-варианта и обычный градиентный спуск.

Нестеров: постоянный vs динамический На графике ниже показаны кривые при итерациях для двух версий:

- *const* (синий);
- *dynamic* (оранжевый).

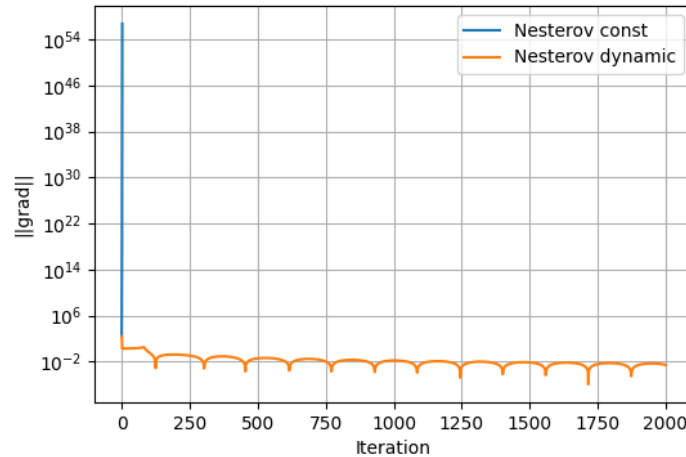


Рис. 5: Сходимость метода Нестерова на функции Розенброка

1. При старте ($k < 10$) обе версии сходятся аналогично быстро, однако константная схема затем почти выравнивается и перестаёт снижаться, что видно по «выпрямлению» синей кривой.
2. Динамическая версия демонстрирует регулярные «провалы» — резкие локальные уменьшения нормы градиента в моменты обновления β_k , после чего метод вновь стабилизируется на новом уровне.
3. В долгосрочной перспективе ($k > 200$) только динамический Нестеров продолжает плавное снижение, тогда как версия *const* зашла в застой и не может преодолеть Розенброка.

Модификации Метода сопряженных градиентов и Метод сопряженных градиентов На графике ниже сравниваются:

Метод градиентного спуска, Метод Флетчера–Ривза, Метод Полак-Рибье, Метод Хестинс-Штифель,

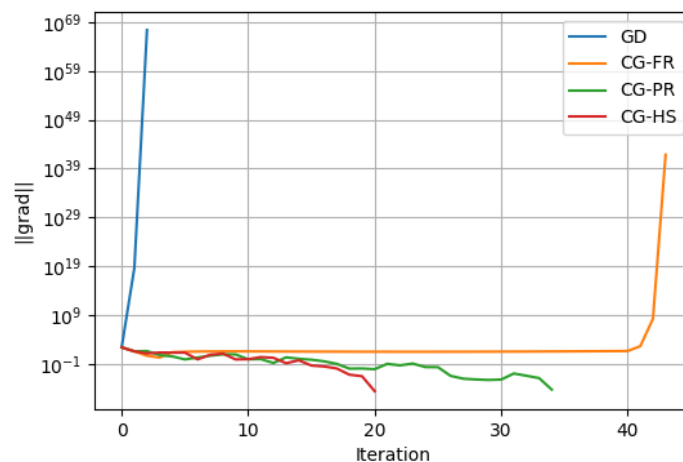


Рис. 6: Сравнение Градиентного спуска и трёх модификаций Метода сопряженных градиентов на Розенброке

1. **GD** (синяя кривая) резко замедляется после нескольких шагов и вблизи минимума снижает скорость убывания градиента в несколько порядков.

2. **CG-FR** (оранжевая) сначала сходится быстро, но затем «взрывается» (возникает резкий рост), что указывает на неустойчивость метода FR без жёсткого контроля шага.
3. **CG-PR** (зелёная) показывает стабильное и монотонное уменьшение до малых значений, без значительных всплесков.
4. **CG-HS** (красная) сходится чуть быстрее PR и тоже сохраняет плавность траектории.

При попытке применить стандартный линейный поиск в SciPy мы столкнулись с несколькими трудностями:

- overflow в вычислении градиента: при агрессивном шаге x «улетал» в области, где $y - x^2$ большим образом увеличивался.
- LineSearchWarning: алгоритм не всегда сходился, выдавая предупреждения.
- После нескольких десятков итераций возникали NaN и inf из-за потери численной устойчивости.

Теперь сделаем выводы по всем экспериментам на функции Розенброка:

- Версия Нестерова с динамическим β оказалась более устойчивой и обеспечила стабильную сходимость, несмотря на резкие «впадины» градиента.
- Из трёх CG-модификаций β^{PR} и β^{HS} работают заметно лучше, чем β^{FR} , демонстрируя быструю и плавную сходимость до $< 10^{-6}$.
- Градиентный спуск без ускорения слишком медленный и быстро теряет точность из-за плохого баланса кривизны вдоль Розенброка.
- Для Розенброка предпочтительнее сочетать динамический Нестеров или PR/HS-версию CG с аккуратным линейным поиском; FR и фиксированный Нестеров могут потребовать значительной ручной настройки параметров.

5. Общий вывод

В ходе работы мы пришли к следующим выводам:

1. Ускоренные методы vs классический градиентный спуск.

- Метод тяжелого шара и метод Нестерова демонстрируют значительное ускорение по сравнению с градиентным спуском на плохо обусловленных функциях.
- Для квадратичных задач с большим числом обусловленности ($\kappa \gg 1$) оба метода сокращают число итераций в несколько раз, но требуют точных оценок L и μ .
- На «почти-квадратичной» и Розенброке Нестеров с динамическим параметром (β_k) показывает большую устойчивость и возможность выхода из плато по сравнению с фиксированной версией.

2. Методы сопряжённых градиентов.

- В строгом квадратичном случае CG-метод сходится за $O(n)$ итераций, практически независимо от $\kappa(A)$, что делает его оптимальным выбором при доступности операций с A .
- Среди нелинейных CG-модификаций для функции Розенброка версии PR и HS выигрывают у FR благодаря большей численной устойчивости и отсутствию «взрывов» градиента.

3. Численные тонкости и практические рекомендации.

- Линейный поиск по Армихо–Вольфу может не сходиться на сильно негладких площадках (особенно для FR), поэтому для надёжности стоит применять backtracking или ограничивать шаг вручную.
- На функциях с узкими «долинами» (Розенброк, почти-квадратичная) полезна визуализация траекторий и анализ плато, чтобы скорректировать параметры α и β .
- В реальных приложениях сочетание адаптивных ускорений с аккуратным линейным поиском даёт наилучшее соотношение скорости сходимости и устойчивости.