Министерство науки и высшего образования
Московский Авиационный институт
(Национальный исследовательский университет)

Институт №8
«Компьютерные науки и прикладная математика»

Отчёт о выполнении курсовой работы
по дисциплине «Численные методы».
Тема: "Нахождение собственных значений и собственных векторов симметричных разреженных матриц большой размерности. Метод Ланцоша"

Студент	Свистельников И.В.
Вариант	3
Группа	М8О-405Б-21
Преподаватель	Демидова О.Л.
Дата	14 декабря 2024 г.

Оценка: ____________

Подпись преподавателя: ____________

Подпись студента: ____________

Москва 2024

Содержание

1 Сравнение алгоритма Ланцоша с другими методами и области применения

Алгоритм Ланцоша является важным инструментом для нахождения нескольких (k) собственных значений и собственных векторов разреженных симметричных матриц. Его эффективность и популярность обусловлены рядом свойств, которые делают его предпочтительным для некоторых задач, однако в других случаях альтернативы, такие как степенной метод или метод Арнольди, могут оказаться более подходящими. Рассмотрим области применения и ситуации, в которых алгоритм Ланцоша находит оптимальное применение, а также ситуации, когда стоит рассмотреть другие методы.

1.1 Когда использовать алгоритм Ланцоша

Алгоритм Ланцоша лучше всего подходит для следующих задач:

- **Анализ больших разреженных матриц.** Ланцош специально разработан для работы с разреженными симметричными матрицами. Это делает его идеальным для задач, возникающих в физике, графовой теории и анализе больших сетей, где плотность матрицы мала.
- **Поиск нескольких крайних собственных значений.** Если требуется найти только первые k минимальных или максимальных собственных значений, Ланцош позволяет избежать вычисления полного спектра, значительно снижая вычислительные затраты.
- **Квантовая механика и моделирование физических систем.** Многие операторы, такие как гамильтонианы, имеют разреженные симметричные представления. Алгоритм Ланцоша помогает находить энергетические уровни систем.
- **Спектральное кластеризование и машинное обучение.** В задачах кластеризации графов или снижения размерности (например, Principal Component Analysis) нахождение нескольких главных собственных значений матрицы смежности или ковариации может быть эффективно выполнено с помощью Ланцоша.
- **Ограниченные вычислительные ресурсы.** Благодаря своей итеративной природе, Ланцош экономит память и подходит для суперкомпьютеров и распределённых вычислений.

1.2 Когда отдать предпочтение другим методам

Несмотря на свои преимущества, алгоритм Ланцоша имеет ограничения, из-за которых другие методы могут быть более эффективными:

- **Невозмущённые симметричные матрицы.** Степенной метод (Power Iteration) может быть более простым и подходящим для нахождения одного доминирующего собственного значения (или доминирующего собственного вектора), как в задаче Google PageRank.
- **Неограниченные матрицы.** Ланцош работает только с симметричными матрицами. Для работы с несимметричными матрицами лучше подойдёт метод Арнольди, который является обобщением Ланцоша и может обрабатывать широкий класс задач.

- **Проблемы ортогонализации.** Ланцош подвержен эффекту «запоминания» (loss of orthogonality), что может привести к снижению точности для некоторых матриц. В таких случаях метод Арнольди может быть более устойчивым.

- **Точность.** Если требуется высокая точность и нахождение всех собственных значений, следует использовать прямые методы, такие как метод Якоби или QR-разложение, несмотря на их высокую вычислительную стоимость.

- **Маленькие плотные матрицы.** Для матриц небольшого размера или с высокой плотностью данных Ланцош не предоставляет явных преимуществ. Прямые методы или SVD могут быть предпочтительнее из-за их детерминированной точности.

1.3 Заключение

Алгоритм Ланцоша является мощным методом, предназначенным для узкого спектра задач, связанных с разреженными симметричными матрицами. Он обеспечивает высокую вычислительную эффективность при решении задач физики, анализа данных и машинного обучения. Однако его применение имеет ограничения, и для задач с несимметричными матрицами, высокими требованиями к точности или особой устойчивостью стоит рассматривать другие методы, такие как степенной метод или метод Арнольди.

2 Теоретические выкладки

2.1 Основные определения

Разреженная матрица — матрица с преимущественно нулевыми элементами. В противном случае, если большая часть элементов матрицы ненулевая, матрица считается плотной или заполненной.

Разреженные матрицы имеют широкий спектр применения в задачах, где данные представлены в виде графов, сетей или других дискретных структур. Примеры включают матрицы смежности графов, матрицы коэффициентов для систем линейных уравнений в численных методах, а также матрицы в задачах машинного обучения, например, для представления текстовых данных в виде мешка слов.

2.2 Ранжирование разреженных матриц

Для классификации разреженности матрицы используется понятие плотности, которое определяется как отношение числа ненулевых элементов к общему числу элементов:

$$\text{Плотность} = \frac{\text{Количество ненулевых элементов}}{\text{Общее количество элементов}}.$$

На основе плотности матрицы можно разделить на следующие категории:

- **Сверхразреженные (ultrasparse):** Плотность меньше 0.1%. Такие матрицы встречаются, например, в рекомендационных системах (пользователь-продукт), где большинство пользователей взаимодействуют с небольшим числом продуктов.
- **Сильно разреженные:** Плотность от 0.1% до 1%. Примером служат матрицы смежности графов в социальных сетях.
- **Разреженные:** Плотность от 1% до 10%. Часто встречаются в задачах численных методов, например, при моделировании физических процессов.
- **Умеренно плотные:** Плотность от 10% до 50%. Такие матрицы иногда требуют перехода к плотным представлениям из-за высокой вычислительной стоимости операций.
- **Плотные:** Плотность более 50%. Типичны для задач машинного обучения и анализа данных, таких как матрицы ковариации или корреляции.

2.3 Особенности работы с разреженными матрицами

Работа с разреженными матрицами требует специализированных методов и структур данных для эффективного хранения и выполнения операций. Преимущества разреженных матриц включают:

- **Экономия памяти.** Вместо хранения всех элементов используется только информация о ненулевых значениях и их позициях. Для этого применяются такие форматы, как CSR (compressed sparse row), CSC (compressed sparse column), COO (coordinate format) и др.
- **Ускорение вычислений.** Многие операции, такие как умножение на вектор или вычисление произведений, можно выполнять только над ненулевыми элементами, избегая обработки нулей.

- **Масштабируемость.** Разреженные матрицы позволяют работать с огромными размерами данных, что делает их незаменимыми в задачах анализа графов и обработки больших данных.

Однако работа с разреженными матрицами также имеет ограничения. Например, алгоритмы, которые не учитывают разреженность (такие как стандартные LU- или QR-разложения), могут значительно ухудшать производительность и приводить к «заполнению» (fill-in), то есть увеличению числа ненулевых элементов при выполнении преобразований.

2.4 Применение разреженных матриц

Разреженные матрицы встречаются в широком спектре задач:

- **Рекомендационные системы.** Матрицы пользователь-продукт, где каждое значение представляет взаимодействие, например, оценку.
- **Анализ графов.** Матрицы смежности или матрицы инцидентности графов.
- **Численные методы.** Решение больших систем линейных уравнений, например, в задачах конечно-элементного анализа (Finite Element Analysis).
- **Обработка текстов.** Представление текстовых данных через матрицу «мешка слов» или TF-IDF, где строки соответствуют документам, а столбцы — словам.
- **Обработка изображений.** В сжатых представлениях изображений и видео, где большая часть коэффициентов сжимающих преобразований является нулевой.

Таким образом, разреженные матрицы являются важным инструментом для решения задач, связанных с большими размерами данных, где экономия ресурсов и производительность имеют ключевое значение.

Симметричная квадратная матрица - такая квадратная матрица, для которой выполняется соотношение $A = A^T$.
Симметричную матрицу можно так же получить преобразованием из произвольной квадратной матрицы преобразованием:

$$A^* = \frac{1}{2}(A + A^T).$$

Доказательство: Возьмем произвольную матрицу A размерности n на n:

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & a_{24} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & a_{34} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & a_{n4} & \cdots & a_{nn} \end{pmatrix}$$

Найдем A^T:

$$A = \begin{pmatrix} a_{11} & a_{21} & a_{31} & a_{41} & \cdots & a_{n1} \\ a_{12} & a_{22} & a_{32} & a_{42} & \cdots & a_{n2} \\ a_{13} & a_{23} & a_{33} & a_{43} & \cdots & a_{n3} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & a_{3n} & a_{4n} & \cdots & a_{nn} \end{pmatrix}$$

Теперь запишем преобразование:

$$\frac{1}{2}(A + A^T) = \begin{pmatrix} a_{11} & \frac{a_{12}+a_{21}}{2} & \frac{a_{13}+a_{31}}{2} & \frac{a_{14}+a_{41}}{2} & \cdots & \frac{a_{1n}+a_{n1}}{2} \\ \frac{a_{21}+a_{12}}{2} & a_{22} & \frac{a_{23}+a_{32}}{2} & \frac{a_{24}+a_{42}}{2} & \cdots & \frac{a_{2n}+a_{n2}}{2} \\ \frac{a_{31}+a_{13}}{2} & \frac{a_{32}+a_{23}}{2} & a_{33} & \frac{a_{34}+a_{43}}{2} & \cdots & \frac{a_{3n}+a_{n3}}{2} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{a_{n1}+a_{1n}}{2} & \frac{a_{n2}+a_{2n}}{2} & \frac{a_{n3}+a_{3n}}{2} & \frac{a_{n4}+a_{4n}}{2} & \cdots & a_{nn} \end{pmatrix}$$

Данная матрица удовлетворяет условиям симметричности, что и требовалось доказать.

Подпространство Крылова - такое подпространство размерности m, порожденное вектором $v \in \mathbb{C}^n$ и матрицей $A \in \mathbb{C}^{n\times n}$ и имеющее следующий вид:

$$\mathbb{K}_m(v, A) = span\{v, Av, A^2v, \cdots, A^{m-1}v\}.$$

Свойства подпространства Крылова:

- $\mathbb{K}_p$ инвариантно относительно A и $\mathbb{K}_m = \mathbb{K}_p$ для любого $m \geq p$.
- $\dim \mathbb{K}_m = \min m, p$.

2.5 Алгоритм Ланцоша

- Инициализация:
  - Выбрирается случайный вектор q_1 размера n (такой же, как и у матрицы A).
  - Этот вектор нормализуется:
  $$q_1^* = \frac{q_1}{||q_1||}.$$
  - Создается матрица Q размерности $n \times k$ для хранения ортогональных векторов (базиса подпространства Крылова), тридиагональная матрица T размерности $k \times k$, в которую записываются коэффициенты α_i и β_i.
  - Задаются начальные значения: первый вектор q_1 записывается как первый столбец матрицы Q, скаляр $\beta = 0$ и вектор $q_0 = 0$ (нулевой вектор).
- Итерационный процесс: Для каждой итерации $j = 1, 2, \cdots, k$ выполняются следующие действия:
  - Умножение матрицы A на текущий вектор q_j:
  $$z = Aq_j.$$
  - Ортогонализация вектора z:
    * Убирается вклад предыдущего вектора q_{j-1} с учетом коэффициента β_{j-1}:
    $$z \longleftarrow z - \beta_{j-1}q_{j-1}$$
    * Убирается вклад текущего вектора q_j с учетом коэффициента α_j:
    $$\alpha_j = q_j^T z, z \longleftarrow z - \alpha_j q_j$$

– Нормализация вектора z:
  * Норма β_j определяется как:
$$\beta_j = ||z||.$$
  * Если $\beta_j = 0$, то процесс завершается (получено полное подпространство).
– Запись нового вектора q_{j+1}:
$$q_{j+1} = \frac{z}{\beta_j}.$$
  Этот вектор добавляется в Q.
– Запись коэффициентов в тридиагональную матрицу:
  * Диагональный элемент $T[j, j] = \alpha_j$.
  * Соседние элементы $T[j, j-1] = T[j-1, j] = \beta_{j-1}$.

- Завершение процесса:
  – После выполнения k итераций строится тридиагональная матрица T_k размера $k \times k$.
  – Вычисляются собственные значения и собственные тридиагональной векторы матрицы T_k:
$$\lambda_i, v_i, i = 1, 2, \cdots, k.$$
  – Собственные векторы исходной матрицы A восстанавливаются как линейные комбинации векторов q_j из матрицы Q:
$$u_i = Qv_i.$$

Доказательство:
Метод Ланцоша основан на использовании **Крыловского подпространства** и свойствах симметричных матриц. Разберем ключевые этапы доказательства.
Основные понятия и идеи:

- **Сжатие пространства:** Вместо того чтобы работать в пространстве размерности n, метод Ланцоша приближает спектральные свойства A в $K_m(A, q_1)$, где $m \ll n$.
- **Тридиагональная форма:** Векторы $q_1, q_2, \ldots, q_m$, построенные в ходе алгоритма, образуют ортонормальный базис $K_m(A, q_1)$. Матрица A в этом базисе имеет тридиагональную форму T_m.

Шаги доказательства

- Построение ортонормального базиса в $K_m(A, q_1)$: Каждый вектор q_i строится рекурсивно:
$$Aq_j = \beta_{j-1} q_{j-1} + \alpha_j q_j + \beta_j q_{j+1},$$
  где:
$$\alpha_j = q_j^T A q_j, \quad \beta_j = \|z - \alpha_j q_j - \beta_{j-1} q_{j-1}\|.$$
  Утверждение: Векторы $q_1, q_2, \ldots, q_m$ ортонормальны. Это следует из конструкции: каждый новый вектор q_{j+1} ортогонализируется по отношению к предыдущим.

- Тридиагональная форма в базисе $K_m(A, q_1)$: Пусть $Q_m = [q_1, q_2, \dots, q_m]$ — матрица, содержащая столбцами ортонормальный базис $K_m(A, q_1)$. Тогда:

  $$T_m = Q_m^T A Q_m.$$

  Поскольку q_i ортонормальны, $Q_m^T Q_m = I_m$ (единичная матрица). В этом базисе: - T_m — тридиагональная симметричная матрица с диагональными элементами α_i и внедиагональными β_i.

- Собственные значения T_m приближают собственные значения A: Собственные значения тридиагональной матрицы T_m совпадают с собственными значениями матрицы A, ограниченной на $K_m(A, q_1)$.

  Ключевая идея: Пространство $K_m(A, q_1)$ стремится "уловить" основные направления (собственные векторы), соответствующие наибольшим собственным значениям A. Собственные значения T_m сходятся к k-крупнейшим собственным значениям A при $m \to n$.

- Приближение собственных векторов: Собственные векторы v_i тридиагональной матрицы T_m связаны с собственными векторами u_i исходной матрицы A через:

  $$u_i = Q_m v_i.$$

  Это означает, что собственные векторы u_i являются линейными комбинациями базисных векторов $q_1, q_2, \dots, q_m$.

Корректность алгоритма

- **Конечность подпространства:** Метод Ланцоша работает в подпространстве $K_m(A, q_1)$, что позволяет сжать матрицу A до тридиагональной T_m и эффективно найти её собственные значения.

- **Симметричность T_m:** Благодаря симметрии A, матрица T_m также является симметричной, что гарантирует вещественность собственных значений.

- **Ортогональность:** Ортонормальность векторов q_i предотвращает накопление ошибок и сохраняет устойчивость метода.

Итог

- Собственные значения и собственные векторы тридиагональной матрицы T_m сходятся к соответствующим значениям исходной матрицы A.

- При $m \to n$, подпространство $K_m(A, q_1)$ приближается к полному спектральному пространству A.

3 Реализация алгоритма Ланцоша для случайной разреженной матрицы

Для тестирования работы алгоритма была взята случайная разреженная матрица размерности $N \times N$ с плотностью элементов 0.02 (2% элементов являются ненулевыми). Так же было взято следующие значения для k: $50, 500$. Результат работы алгоритма приведен ниже на графиках. График отражает следующую информацию:

- Синим цветом показаны истинные (теоретические) значения определенного количества собственных значений матрицы A.
- Красными точками обозначены вычисленные приближенные собственные значения матрицы A.

Так же для каждого рассчета была посчитана L_2-норма вектора ошибок. График ошибок в зависимости от значения k так же приведен ниже.

Численное нахождение собственных векторов

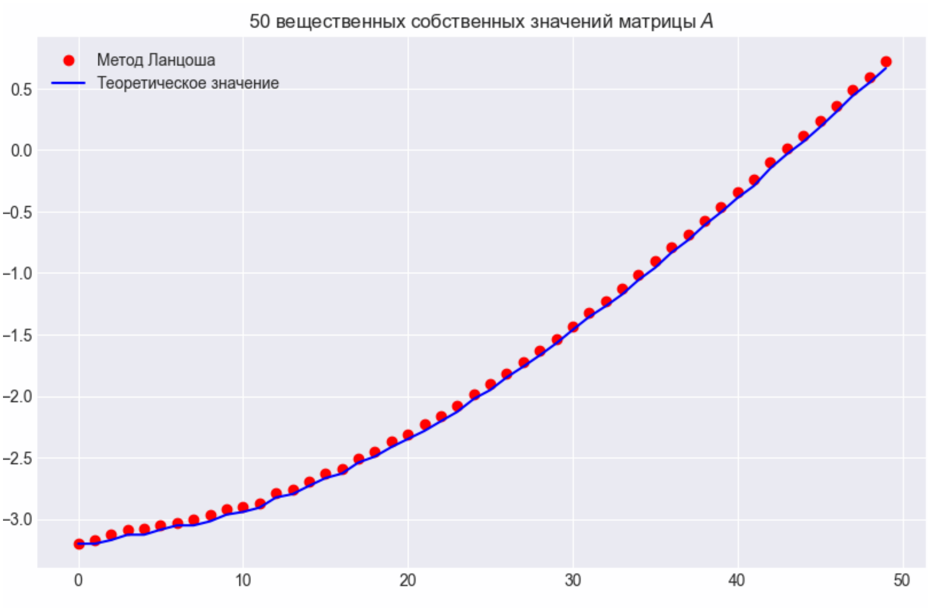

Рис. 1: Численное нахождение собственных векторов для $k = 50$.

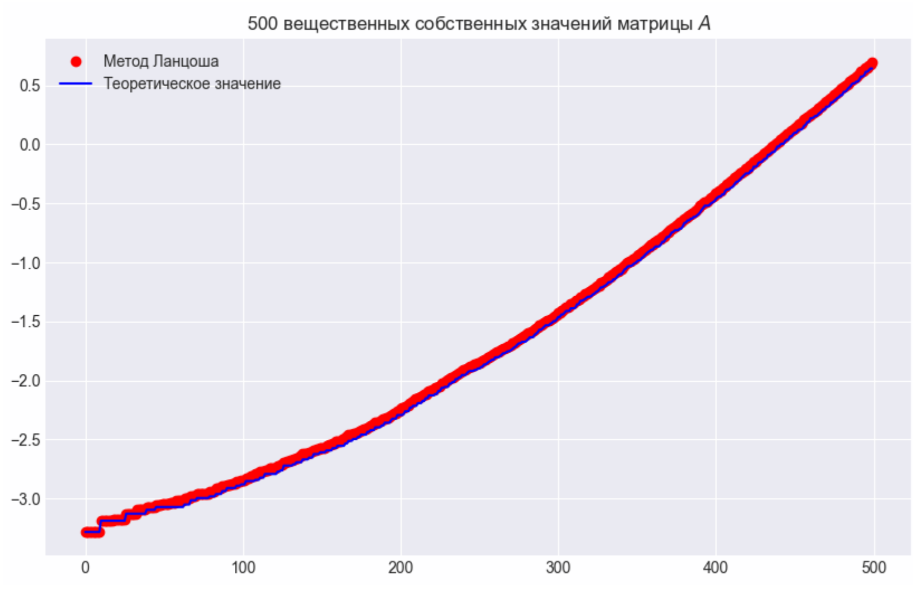

Рис. 2: Численное нахождение собственных векторов для $k = 500$.

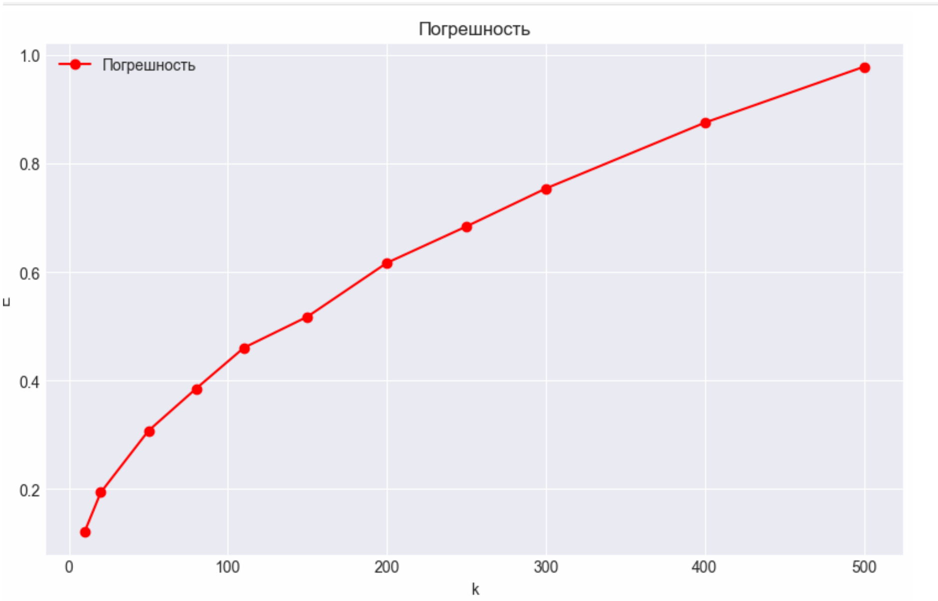

Рис. 3: Норма погрешности в зависимости от значения k.

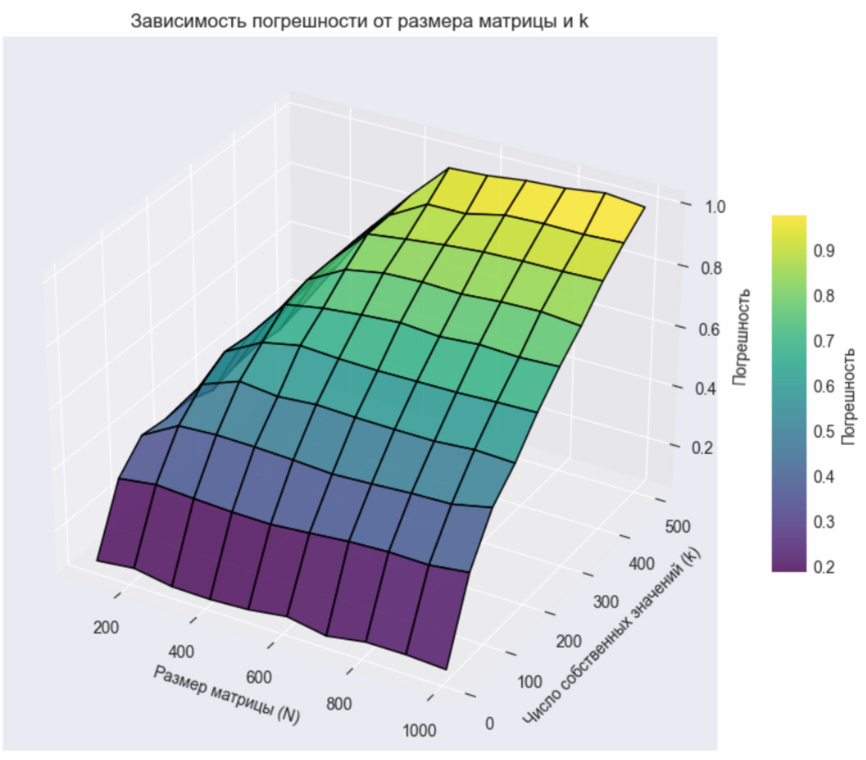

Рис. 4: Поверхность ошибок в зависимости от N и k k.

4 Выводы

- Был реализован работоспособный алгоритм Ланцоша для разреженных матриц большой размерности.
- Была выявлена зависимость для погрешности ошибки данного метода, которая показала что значения увеличиваются при увеличении N и k.
- Была выявлена зависимость для погрешности ошибки данного метода, которая показывает, что алгоритм реализован качественно и на больших количествах показывает небольшую погрешность.
- Были найдены собственные вектора и собственные значения матрицы размерности $N \times N$.

Список литературы

[1] Пирумов У.Г. *Численные методы.* - М.: Дрофа, 2006.

[2] Шарый, С. П. *Курс вычислительных методов.* Новосибирский государственный университет, 2024.