МИНОБРНАУКИ РОССИИ

САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ

ЭЛЕКТРОТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

«ЛЭТИ» ИМ. В. И. УЛЬЯНОВА (ЛЕНИНА)

Кафедра МО ЭВМ

отчет

**по лабораторной работе**

**по дисциплине «Компьютерная математика»**

**Тема: Повороты графа в пространстве**

|  |  |  |
| --- | --- | --- |
| Студент гр. 0382 |  | Афанасьев Н. С.  Крючков А. М. |
| Преподаватели |  | Коптелов Я. Ю. |

Санкт-Петербург

2022

**Оглавление**

[Задание. 3](#_Toc104407299)

[Выбранный метод выполнения задачи. 3](#_Toc104407300)

[Построение графа. 5](#_Toc104407301)

[Выбор «правильных» рёбер. 5](#_Toc104407302)

[Матрицы и координатные векторы. 5](#_Toc104407303)

[Поворот графа. 6](#_Toc104407304)

[Решение обратной задачи. 6](#_Toc104407305)

[Алгоритм Кабша. 7](#_Toc104407306)

[Сингулярное разложение. 9](#_Toc104407307)

[Вывод угла из матрицы поворота. 11](#_Toc104407308)

[Вывод и визуализация результатов. 11](#_Toc104407309)

[Проверка корректности восстановления графа. 15](#_Toc104407310)

[Выводы. 16](#_Toc104407311)

[Список использованных источников. 18](#_Toc104407312)

## Задание.

Пусть дана информация о структуре молекулы в формате JSON из базы данных PubChem, например, молекула аспирина.

Программа должна распарсить JSON (только атомы с координатами и связи с кратностями, игнорируя заряды и прочую дополнительную информацию), построить в памяти граф молекулы (вершины = атомы, рёбра = связи) и определить рёбра, которые одновременно:

- одинарные (т. е. соответствуют одинарным, а не двойным и не тройным связям)

- не принадлежат циклам

- не являются "висячими", т. е. продолжаются с обоих концов

В вышеупомянутой молекуле аспирина таких рёбер четыре (если смотреть на двумерную картинку, в которой не показаны водороды); в трёхмерной картинке с водородами таких рёбер можно насчитать пять.

Далее, программа должна повернуть молекулу в трёхмерном пространстве вокруг каждого из найденных рёбер на случайный угол так, чтобы вращалась не вся молекула, а лишь "половина", лежащая по одну сторону от ребра. Координаты атомов при этом, конечно, изменятся, изменится форма молекулы.

Далее, программа должна забыть случайные углы, проанализировать получившиеся координаты атомов и найти такие углы вращения вокруг найденных связей, чтобы атомы встали на исходные места.

## Выбранный метод выполнения задачи.

Существуют несколько возможных способов решения поставленной задачи, был выбран способ, включающий в себя алгоритм Кабша.

При решении прямой задачи (для получения входных данных), после получения графа и нахождения “правильных” рёбер, случайным образом выберем одну вершину. От этой вершины произведём обход графа для того, чтобы установить направление “правильных” рёбер от взятой вершины.

После того, как мы выбрали направление рёбер, перемешиваем список этих рёбер, а затем устанавливаем случайные углы от -π до π и поворачиваем граф. Значение углов забываем, однако запоминаем выбранную вершину. В итоге мы получаем набор изменённых вершин, где, как минимум, одна вершина не изменит свои координаты.

Для решения обратной задачи начнём обход графа в ширину, начиная с запомненной вершины. Таким образом, на каждом шаге обхода мы имеем возможность сравнивать тройки вершин исходного и изменённого графа (по одной вершине на 3-х последовательных уровнях BFS), причём координаты двух вершин совпадают с изначальными, а координаты третьей вершины отличаются от оригинала. Если мы сможем восстановить поворот проходимого ребра так, чтобы третья вершина встала на исходное положение, то мы сможем произвести обратный поворот графа вокруг этой вершины и продолжить обход уже изменённого графа. По окончании обхода, мы должны получить приближение к исходному графу.

Для определения угла поворота был использован алгоритм Кабша, для получения однозначного ответа, которому необходимо как раз минимум три точки. Сам алгоритм будет описан ниже.

Так как в данном решении на каждом шаге рассматриваются три точки, две из которых находятся на изначальном положении, то избавиться от условия как минимум одной стационарной точки не получиться. Если такой точки не будет, то получить её уже после поворота не получится, так как для восстановления координат случайной точки графа уже понадобится искать нужный поворот для конкретных рёбер. Если же восстанавливать координаты этой точки не через вращение вокруг ребра, а через трансляцию, то мы нарушим структуру графа (длина рёбер изменится). Восстанавливать координаты точки, используя равенство длин всех рёбер, не получиться, так как этот факт неверен (длина химической связи обратно пропорциональна энергии связи).

## Построение графа.

Структура молекулы получается из сервиса PubChem путём отправки HTTP GET запроса. Полученный JSON файл парсится и из него достаются необходимые данные. На основе этих данных в памяти строится граф: список координат вершин, словарь смежности вершин, словарь кратностей рёбер.

## Выбор «правильных» рёбер.

Для выбора “правильных” рёбер ребро удовлетворяет следующим свойствам:

1. Вершины на концах не являются висячими
2. Порядок ребра равен 1
3. Ребро не принадлежит циклам

Для этого мы итерируемся по графу, представленному в виде списка смежности. Для проверки первого условия мы смотрим количество элементов в списке смежности для каждой из вершины на концах. Порядок ребра (1 или более) берётся из списка смежности *order,* который получен из JSON файла. Для того, чтобы проверить нахождения ребра в цикле ребро удаляется из графа. Если путь между вершинами можно найти, то ребро принадлежит какому-то циклу. Затем ребро возвращается на место.

Такие рёбра заносятся в список и записываются в поле графа для дальнейшего условия.

## Матрицы и координатные векторы.

Для выполнения работы были реализованы простые классы Matrix и CVector для работы с матрицами и координатными векторами. Для обоих классов определены операции сложения и умножения. У матриц есть такие необходимые для работы методы, как транспонирование и поиск определителя, а у векторов – получение нормы вектора.

## Поворот графа.

Один из способов описания поворотов в пространстве – матрицы поворота, которые представляют из себя ортогональные матрицы с единичным определителем. При умножении вектора на такую матрицу изменяются координаты, однако длина вектора остаётся прежней.

Для выполнения задачи необходимо уметь поворачивать точки в графе относительно одного из рёбер. Для этого необходимо рассчитать единичный вектор из вектора ребра и вычислить матрицу поворота, затем необходимо переместить вершины так, чтобы конечная точка ребра лежала в начале системы координат, после чего необходимо применить к вершинам матрицу поворота и переместить вершины обратно:

*,* гдеи *–* конечные и начальные координаты точки*, –* трансляция*,* а *M –* матрица поворота*.*

Сама матрица поворота вокруг произвольной оси составляется следующим образом:

Где – это произвольный единичный вектор, а – угол поворота.

Для выполнения работы были реализованы две функции: функция поворота конкретной точки вокруг конкретного ребра и функция поворота графа, в которую передаются список рёбер, вокруг которых необходимо произвести вращение, список углов, список вершин, и словарь, содержащий информацию о рёбрах.

## Решение обратной задачи.

В начале решения обратной задачи нам известна структура графа, начальные и изменённые в результате поворота координаты вершин, а также вершина, с которой будут начинаться обход графа (и, следовательно, направление рёбер, вокруг которых могло происходить вращение). Изначально нам неизвестно, вокруг каких именно рёбер и на какой угол происходило вращение, а также порядок вращений.

Всё решение основано на обходе графа в ширину, о котором велась речь выше, в пункте «Выбранный метод выполнения задачи». На каждом этапе рассматриваются три точки, координаты, одной из которых не совпадают с оригинальными координатами. Для этих трёх точек применяется алгоритм Кабша, результатом которого является матрица поворота. Зная единичный вектор для поворота (вычисляется из координат ребра) и матрицу поворота, можно вычислить угол вращения. Угол и ребро запоминаются, а сам граф вращается вокруг этого ребра на обратный угол, чтобы рассматриваемая точка встала на своё место, а остальные вершины изменили своё положение.

Нужно заметить, что, если в тройке вершин все вершины стоят на своих местах, то алгоритм Кабша и поворот не выполняются. Это важно, например, когда у одной из вершин несколько дочерних, и нет необходимости проворачивать все описанные действия ля каждой из них, так как поворот относительно ребра для них осуществлялся одинаковый.

## Алгоритм Кабша.

Данный алгоритм предназначен для расчёта оптимальной матрицы поворота с минимизацией среднеквадратического отклонения для двух наборов точек. Сам алгоритм состоит из следующих этапов:

1. Получение входных матриц P и Q, которые представляют из себя координаты набора точек. Так как количество точек и размерность должны совпадать, то алгоритм должен проверять равенство размеров матриц. Так как в рассматриваемой задаче точки находятся в трёхмерном пространстве, то матрицы должны быть размера 3xN. Таким образом матрица должна иметь вид:

2. Трансляция координат. Для поворота обычно высчитываются центроиды координат точек и вычитаются из исходных координат, чтобы в результате центроиды обоих наборов совпадали с началом координат. Однако в случае данного решения задачи можно заранее передавать необходимый сдвиг, который будет совпадать с координатами второй точки. Данное суждение можно описать в виде графика (см. рис. 1).

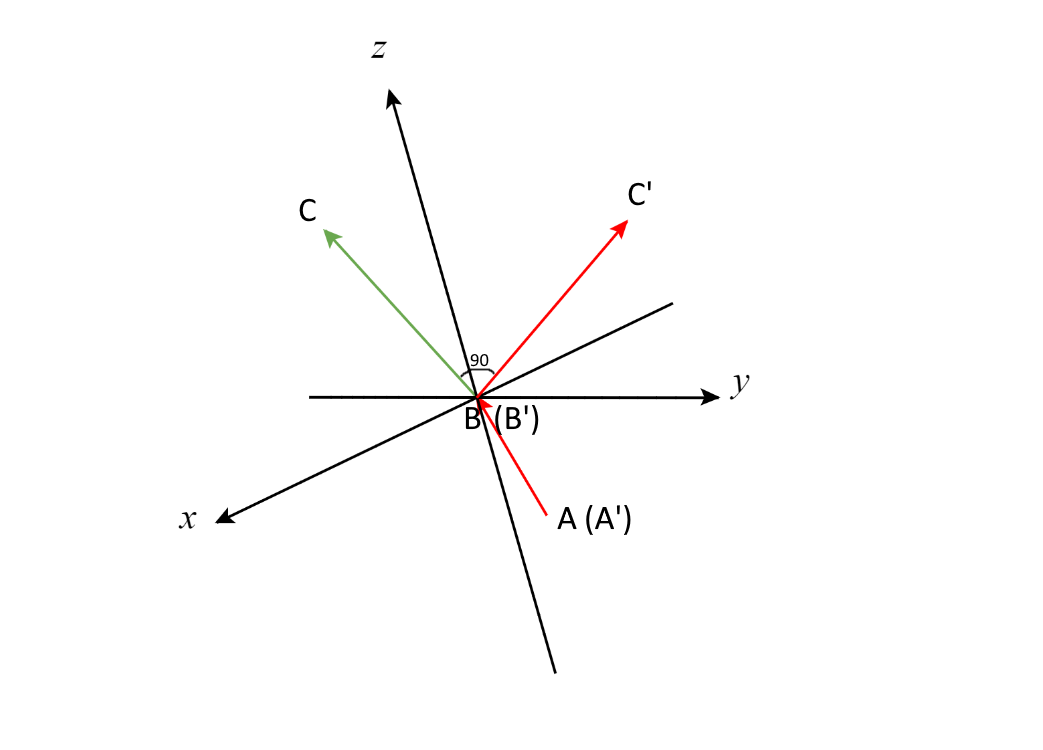


Рисунок 1 – Расположение рассматриваемой тройки точек

3. Вычисление ковариационной матрицы:

4. Вычисление матрицы поворота. Для начала высчитывается сингулярное разложение ковариационной матрицы: (описание сингулярного разложения приведено в следующем пункте). Тогда матрица поворота представима как .

5. Обеспечение правосторонней систему координат. Если определитель полученной матрицы меньше нуля, то необходимо поменять знак у последнего столбца матрицы и посчитать матрицу поворота сначала:

## Сингулярное разложение.

Сингулярное разложение (SVD) – это разложение вида , где Am\*n – входная матрица, Um\*k – матрица левых сингулярных векторов, Vn\*k – матрица правых сингулярных векторов, а k\*k – матрица сингулярных чисел. Геометрический смысл разложения представлен на рис. 2.

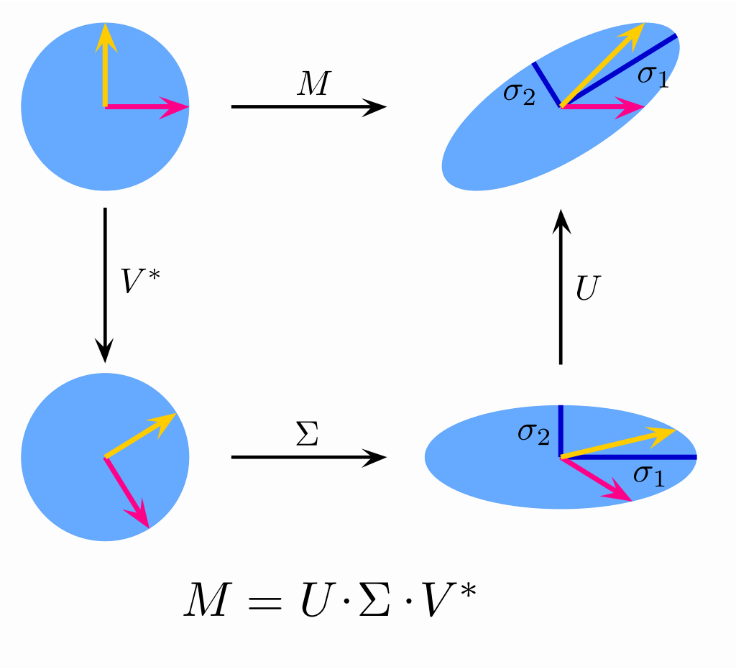


Рисунок 2 – Сингулярное разложение

Алгоритм поиска разложения состоит из следующих этапов:

1. Найти собственные числа матрицы A. Так как эта матрица может быть треугольной нужно рассматривать матрицу . Необходимо получить характеристический многочлен: , корни которого и будут являться собственными числами.

2. Составить матрицу . На диагоналях этой матрица расположены сингулярные числа в убывающем порядке. Сингулярные числа представляют из себя корни собственных чисел.

3. Найти собственные вектора. Получаются вектора путём решения системы вида , где – рассматриваемое собственное число.

4. Составить матрицу V из собственных векторов.

5. Найти матрицу U. Из начальной формулы можно вывести формулу для матрицы U: .

6. Транспонировать матрицу V, и разложение найдено.

Реализация такого алгоритма весьма затруднительна из-за расчёта собственных чисел и векторов, поэтому существуют другие реализации поиска. В рамках курса упоминался алгоритм простой итерации для поиска сингулярного разложения (power iteration method). Этот метод основан на итерации вида , где А – исходная матрица, b0 – случайный вектор. Итерация продолжается пока результат не сойдётся (разница меньше эпсилона; существует вариация алгоритма, которая рассчитывает необходимое число итераций). Найденный вектор будет являться доминирующим собственным вектором, из которого можно получить собственное и, следовательно, сингулярное число. Для получения следующего вектора мы должны удалить доминирующее направление из матрицы и повторить поиск наиболее доминирующего сингулярного значения: .

Этот метод проще для реализации, однако на практике он обычно не используется, так как имеет ряд ограничений. Во-первых, исходная матрица должна быть диагонализируемой. Во-вторых, каждый следующий собственный вектор для доминирующего собственного числа должен быть строго меньше предыдущего. В-третьих, выбранный случайным методом вектор должен иметь ненулевую компоненту в направлении собственного вектора, связанного с доминирующим собственным числом. Если эти ограничения не выполняются, то метод может не сойтись. В виду возможной недетерминированности и численной нестабильности, в работе использовалась готовая реализация сингулярного разложения (хотя описанный метод был также реализован).

## Вывод угла из матрицы поворота.

Зная матрицу поворота и координаты границ ребра, можно высчитать угол поворота. Для начала приведём ребро в вид нормированного вектора: . Далее, из формулы матрицы вокруг произвольной оси, можно высчитать косинус и синус: . Стоит заметить, что знаменатели могут оказаться нулями, поэтому для корректного вычисления нужно будет видоизменять формулу (брать другие коэффициенты R и, соответственно, другие переменные). Значение угла можно получить через арккосинус с учётом знака полученного синуса.

## Вывод и визуализация результатов.

Вывод осуществляется в консоль и файл. Сначала записывается таблица координат: начальный, изменённые и восстановленные. Затем приводится исходный список рёбер, вокруг которых происходил поворот, и углов, а затем приводится полученный список рёбер и углов. Выводятся только те найденные рёбра, значение угла поворота которых не меньше 0.01°. Пример вывод можно увидеть в следующем пункте.

Визуализация осуществляется в файл html. Была необходима библиотека, которая умеет создавать простейшую картинку в 3d. Была выбрана библиотека *plotly* за удобство просмотра результата. Сначала в 3d пространства заносятся координаты молекул (красные точки). Затем рисуются рёбра между(серые линии).

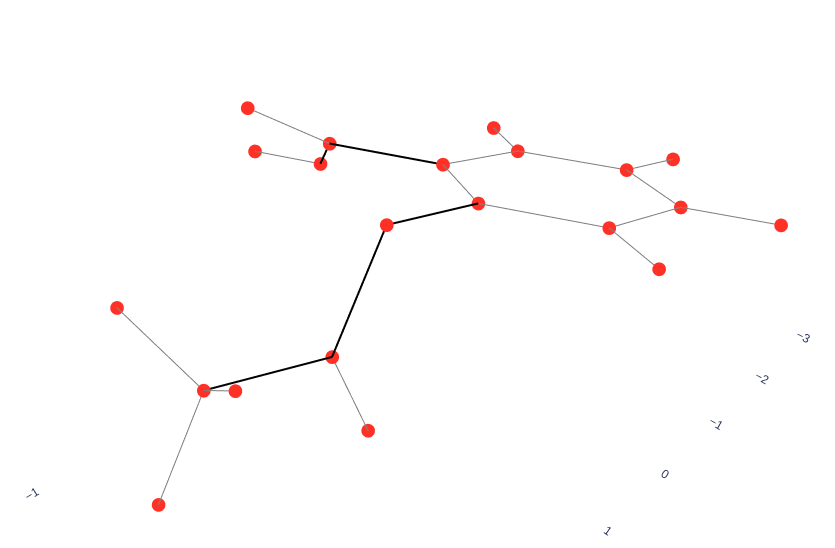


Рисунок 3 - Начальное состояние молекулы Аспирина

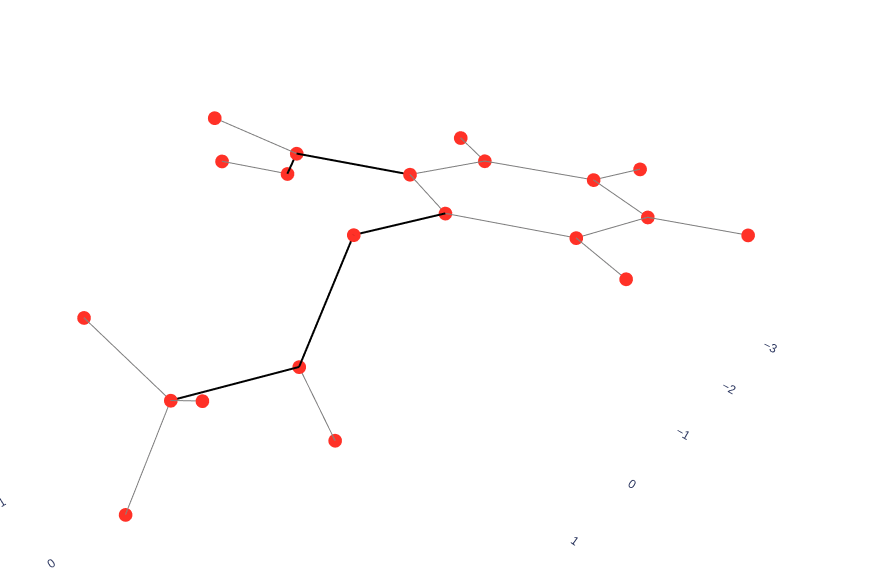
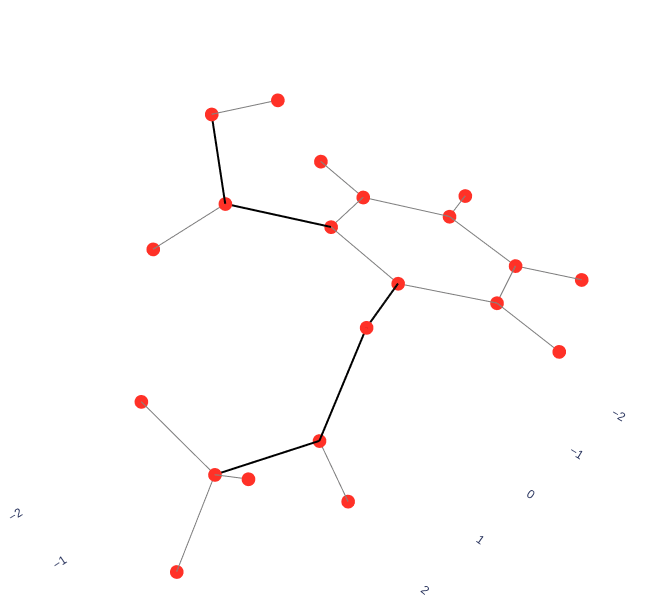
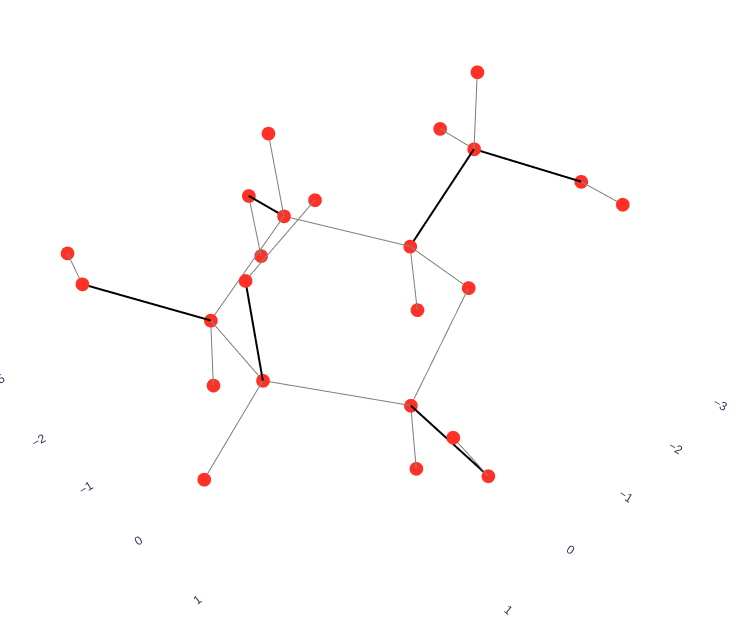
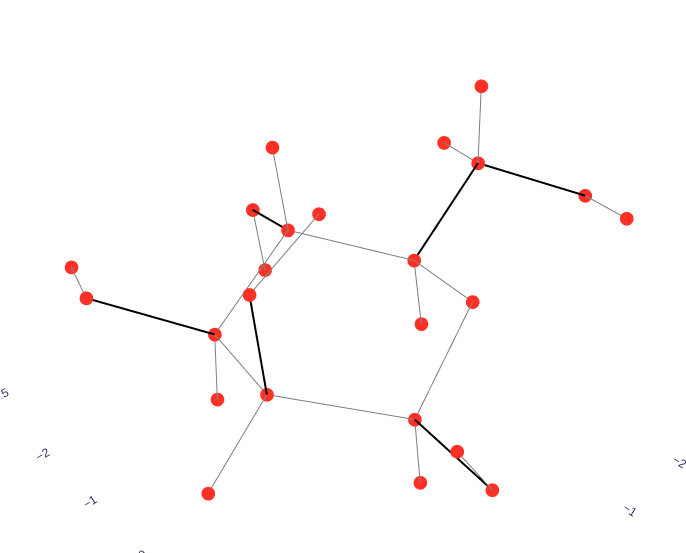
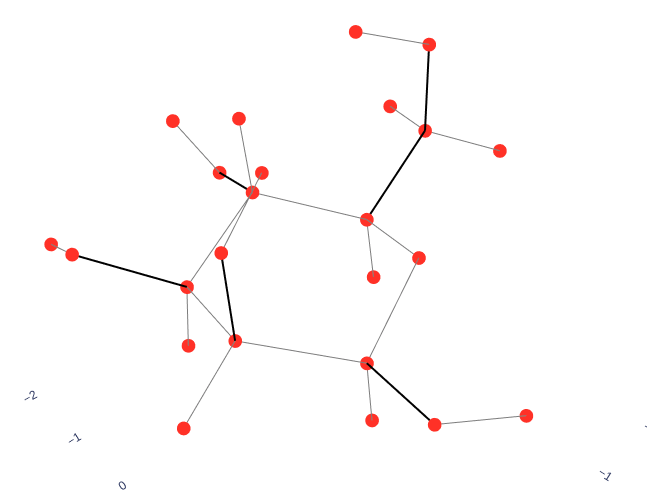


Рисунок 4 – восстановленное состояние молекулы аспирина

Рисунок 5 – Повернутое состояние молекулы аспирина

Рисунок 7 - d-mannose до поворота

Рисунок 9 — d-mannose после восстановления

Рисунок 10 — d-mannose повернутое состояние

## Проверка корректности восстановления графа.

Для проверки достаточно посмотреть на результаты выполнения программы и сравнить полученные координаты и углы.

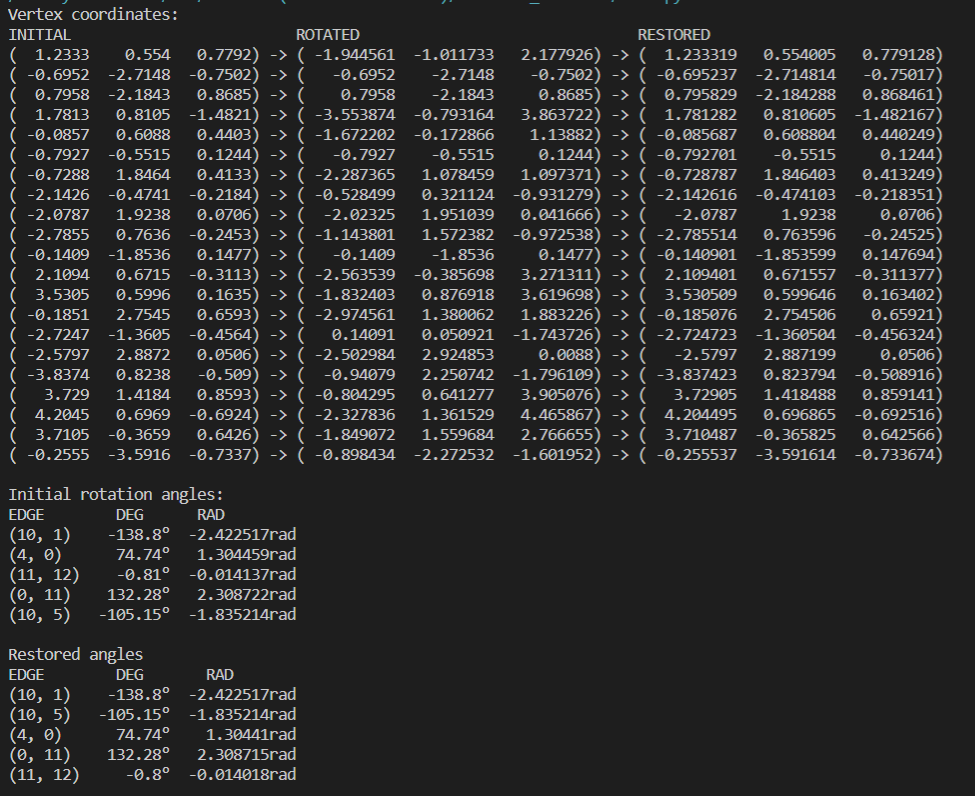


Рисунок 11 - Пример 1: молекула аспирина



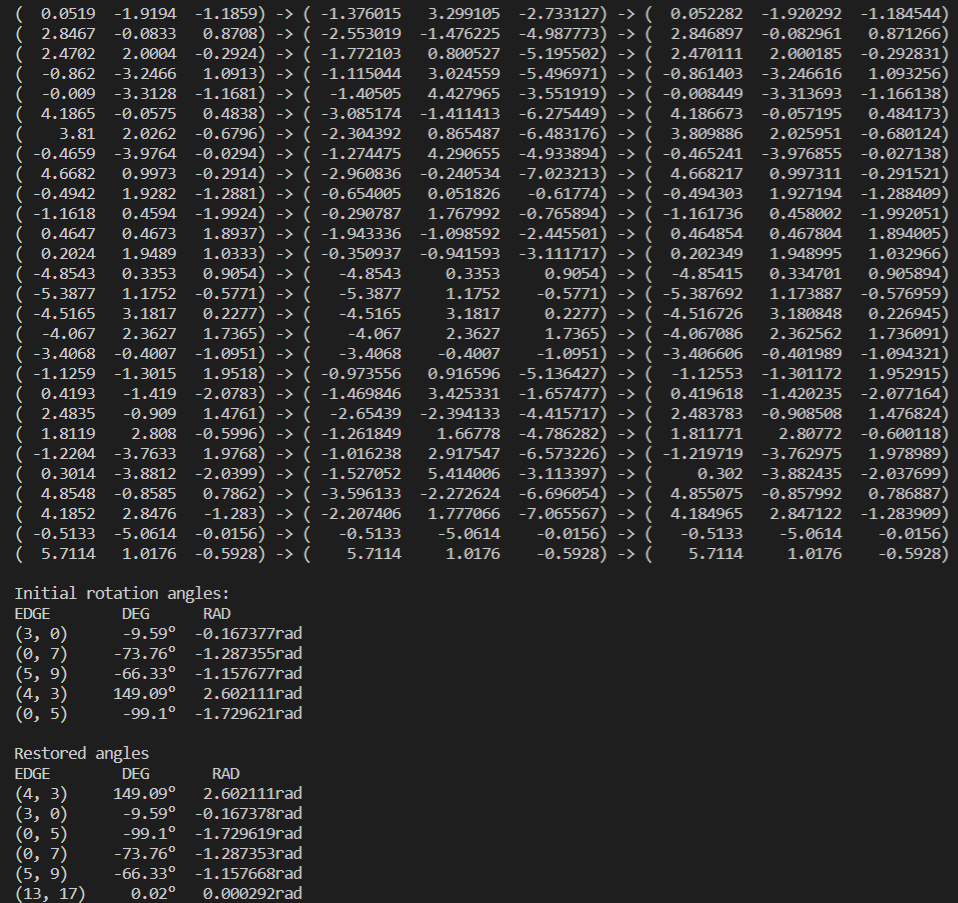


Рисунок 12 - Пример 2: молекула антазолина

Как видно из результатов, рёбра находятся правильно. Координаты восстанавливаются с очень маленькой погрешностью, соответственно углы тоже.

## Выводы.

Был изучен и применён алгоритм Кабша. Было изучено применение сингулярного разложения на практике. Придумано и реализовано решение обратного поворота молекулы. Была осуществлена работа с парсингом json файла. Были применены алгоритмы на графах. Были реализованы функции поворота векторов в пространстве с помощью матриц поворота.

## Список использованных источников.

1. <https://www.bsuir.by/m/12_113415_1_70397.pdf> - описание матриц поворота
2. <https://ru.wikipedia.org/wiki/Матрица_поворота> - ещё одно описание матриц поворота
3. <https://drive.google.com/drive/folders/1MhRtUpfdauweNcWX5NtfEnyewFd1ofcP> - конспекты лекций по алгоритму Кабша
4. <https://en.wikipedia.org/wiki/Kabsch_algorithm> - алгоритм Кабша
5. <https://en.wikipedia.org/wiki/Singular_value_decomposition> - SVD
6. <https://towardsdatascience.com/simple-svd-algorithms-13291ad2eef2> - SVD Power iteration method
7. <https://jeremykun.com/2016/05/16/singular-value-decomposition-part-2-theorem-proof-algorithm/> - SVD
8. <https://en.wikipedia.org/wiki/Power_iteration> - power iteration method
9. <https://c3d.libretexts.org/CalcPlot3D/index.html> - отрисовка графов в 3D