LINEAR ALGEBRA, MOLECULES AND MATLAB

MÅRTEN LEVENSTAM
KLAS SAMUELSSON
CHALMERS FINITE ELEMENT CENTER
CHALMERS UNIVERSITY OF TECHNOLOGY

1. Introduction

Molecule visualization and computations on molecules is a fast growing field of research. To visualize a three-dimensional molecule on the computer screen is an interesting and good example of a combination of programming and analytical geometry.

The structure of molecules can be represented in various ways, one popular file format is the pdb format. Primarily used for proteins it is widespread and a large number of molecules can be found in this format. One such database is the Protein Data Bank (PDB) operated by the Research Collaboratory for Structural Bioinformatics. This database can be found at:
`www.rcsb.org/pdb`.

Other databases with molecules in the pdb format can be found on the web page for this project, see below.

The aim of the project is to write a suite of matlab function for visualization and computation on molecules in this format. This is both a training in mathematics and programming.

The pdb file-format is a rather complex format so in order to make things simpler we have created a simpler format which we call mpdb. A molecule in pdb format can be converted to a molecule in mpdb format by a small unix command called pdb2matlab.

2. Programming

To help you get started we have written a couple of matlab functions which you can use as a starting point for your project. These programs are documented on the web-page for the project. These programs can be used as a draft for your own functions and m-files.

3. Mathematics

The project is an application of vectors in three space dimensions and matrix algebra. The idea is to use your knowledge of analytical geometry and apply this into a working program.

Written 1999, modified by Stig Larsson, 2000.

4. Additional information

Additional information on this project and links to useful places on the web can be found at:
`http://www.md.chalmers.se/Centres/Phi/education/courses/2000/ala-b/project1`
Here you can also find the source code for the matlab functions we have written.

5. Presentation

The project should be presented in a written report made by the group. This report should contain the theoretical work done and a documentation of the matlab programs together with examples of program output. The matlab documentation could preferable be written as comments inside the m-files as in the example codes given to you. The report should be written as a technical report with title, abstract, introduction and references. The software documentation can be put in an appendix to the report.

6. Project assignments

The project is divided into two parts: One set of compulsory questions to be done by all groups and one part with more freedom of choice. A good starting point is to modify and extend the set of routines for vectors in the plane, given earlier in the course, to a set of function for vectors in three-dimensions.

6.1. **Compulsory assignments.** The following list of assignments should be made by all groups.

1. Bond lengths and angles
   (a) Show how the bond lengths and bond angles can be computed. Prove that these are invariant under rotation, translation and reflection. Show that the dihedral angle is not invariant under reflection. These measures are defined in [1].
   (b) Write functions in matlab that computes bond length and bond angles.
2. Transformations
   (a) Show how a molecule is modified if it is rotated an angle α around a given vector r.
   (b) Show how a molecule is modified if it is reflected in a plane with normal n passing through the point r.
   (c) Write functions in matlab that performs the above rotations and reflections producing a new transformed molecule which can be visualized.
   (d) Use these functions to experimentally verify the properties of the angles and bond lengths obtained above.

6.2. **Free assignments.** The following list is an example of other things that can be computed within this project. Every group should do at least one of these assignments. You are also free to come up with your own questions, the list below is just a suggestion for things to do. Try out your programs with different molecules.

1. Make a better visualization. Showing for instance names of atoms, different colors or sizes for different atoms, a nicer representation of bonds.
2. Find the closest atom to a given point.
3. Find the closest atom to a given plane.

4. Compute the maximum distance between all atoms in the molecule.
5. Make a two-dimensional plot of the molecule by projecting the molecule to a plane.
6. Compute the minimum size of a rectangular box containing the molecule.
7. Sort the bonds by their length.
8. Compute the total mass of the molecule.
9. Compute the centre of gravity of the molecule.
10. Find all atoms which have n number of bonds.
11. Find all bonds forming a closed loop.

If you divide these assignments between the groups then maybe you can put together the efforts of all groups into one working package where all groups can share the other groups work.

7. Matlab programs

The matlab functions you can obtain from the web page are:

- `moleculeRead` - Read a molecule from file
- `moleculeWrite` - Save a molecule to file
- `moleculePlot` - Plot a molecule
- `moleculeNeighbours` - Find atoms linked through a bond

in addition to this you can also use the set of routines for vectors that you have seen earlier in the course.

References

[1] A. Neumaier, *Molecular Modeling of Proteins and Mathematical Prediction of Protein Structure*, SIAM Review **39** (1997), 407–460.