Bioinformatics III

Prof. Dr. Volkhard Helms — Saarland University
Maryam Nazarieh, Daria Gaidar, Duy Nguyen, Thorsten Will — Chair for Computational Biology
Winter Semester 2015/2016

Exercise Sheet 10

Due: January 22, 2015 13:15

Submit your solutions on paper, hand-written or printed at the beginning of the lecture or in building E2.1, Room 3.02. Alternatively you may send an email with a single PDF attachment. If possible, please include source code listings. Additionally hand in all source code via mail to maryam.nazarieh@bioinformatik.uni-saarland.de.

Mapping of Crystal Structures into EM Maps

For some protein complexes one has available both a low resolution image like, e.g., an AFM image or an EM density map of the whole complex and the atomic structure of the individual constituents. Then, one is interested in where these constituents are located in the cluster. The task is consequently to fit the position and orientation of a given structure with atomic resolution into a blurred density map such that the correlation is maximized. To acchieve a maximal overlap, the high resolution structure has to be blurred, i.e., convoluted with the experimental resolution. An efficient way to calculate the convolution of the atomic structure data with the experimental resolution is via the Fourier theorem. Therefore you will look at various properties of the Fourier transform in the first exercise.
However, you will not perform a full 3D reconstruction of multiple fragments into a blurred complex but try to fit a 2D structure into a smeared image of itself.

At first, remember the following mathematical definitions.

The continuous Fourier transform of a function $f(x)$ is defined as:

$$FT[f(x)] = F(k) = \int f(x)e^{-ikx}dx$$

and its inverse is, consequently:

$$f(x) = FT^{-1}[F(k)] = \frac{1}{2\pi}\int F(k)e^{ikx}dk.$$

The convolution of two functions $f(x)$ and $g(x)$ is defined as:

$$(f \star g)(x) = \int f(x-y)g(y)dy.$$

The delta distribution is defined as:

$$\delta(x_1 - x_2) = \frac{1}{2\pi}\int e^{ik(x_1-x_2)}dk.$$

The Gaussian distribution is defined as:

$$g(x, x_0) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-x_0)^2}{2\sigma^2}}$$

Exercise 10.1: Properties of the Fourier transform (FT) (30 points)

(a) **FT and its inverse (10)**
Show that $FT^{-1}[FT[f(x)]] = f(x)$ using the delta distribution.

(b) **Shift of the argument (6)**
Show that a shift of the argument x to $x + \Delta x$ of the function f shows up as a phase factor $e^{ik\Delta x}$ in $FT[f]$

(c) **Linearity (5)**
Show that $FT[f(x) + a \cdot g(x)] = F(k) + a \cdot G(k)$ whereby $a \in \mathbb{C}$.

(d) **Convolution Theorem (9)**
Show that $FT[(f \star g)](k) = F(k)G(k)$.

Exercise 10.2: Blurring the Structure (20 points)

Calculate the convolution of a model molecule with an experimental uncertainty, which is described by a Gaussian distribution of width σ, centered around x_0. The density $p(x)$ of the model molecule is given by a sum of delta peaks with masses m_i at the atom positions x_i.

$$p(x) = \sum m_i \delta(x - x_i)$$

Give the formula to compute the blurred density.

Exercise 10.3: Reconstruction of Low Resolution Images (50 points)

For this 2D fit you are given a file `hello.txt` with the atomic structure of the hypothetical protein and various smeared images. Implement a reconstruction program with which you perform the tasks given below.

The objective is to minimize the difference between the given experimental maps and the blurred map from the structure. For this, use the sum of the squared differences between the two maps at each grid point.

As the center of mass is the same for the original structure and for the resulting blurred image, you deal with the displacement by first determining the center of mass of the blurred map and then shifting the structure to have the same center of mass. This will give you different offsets for different rotation angles.

In the structure file each line holds, in this order, the x- and y-positions of an atom and its mass. This mass determines, how much a given atom contributes to the image, i.e., how visible this atom is to the imaging.

<u>Hint</u>: You can start from the supplied Python script, which was used to generate the blurred maps. Note that the shifting and rotation parameters saved in this script are not the ones used to generate the given density maps.

(a) **Resolution Calibration (25)**
To calibrate the resolution to be used for the reconstruction, minimize the difference between the given map `hello_shift.dat` and the map generated from the atomic structure by varying the width for the Gaussian used to smear the high resolution structure.

Give the offset and plot the sum of the squared differences against the width σ. Create a 2D plot of the smoothed image with the optimal σ. Try to include the atom positions, too.

Keep this optimal σ for the subsequent reconstructions.

(b) **Angular Correlation (25)**
In the next experimental map, `hello_rotshift.dat`, the protein is rotated and displaced. Calculate the difference between the given map and the blurred known structure for rotation angles between 0 and 2 π in at least 100 angular steps. Plot the difference vs. the rotation angle and determine the best fit rotation angle. Plot the reconstructed image. In the second plot show the angle-dependent x- and y-offsets.

Hint: Determine the required shift after performing the rotation.

Recommended functions to answer parts (a) and (b) in exercise 10.3:

- read the atomic structure of the given molecule data using 'readData' function. This function reads the input molecule data, here 'hello.txt' line by line storing the x- and y-position plus size of each atom.
- use the 'computeGrid' function to convolute and place the molecule onto the image grid based on a user given sigma.
- 'readGrid' function helps to read the grid.
- You should write the 'compare' function to get the distance between the convoluted image grid and the low resolution images 'hello_shift.dat' and 'hello_rotshift.dat'.
- You should write the 'shift' function to find the offset between two images.
- 'rotate' function will help to rotate the given molecule data.

Have fun!