

Homework 5. MDS with Uncertainty

*Instructor: Yuan Yao**Due: Tuesday April 19, 2016*

The problem below marked by * is optional with bonus credits.

1. *Protein Folding*: Consider the 3D structure reconstruction based on incomplete MDS with uncertainty. Data file:

<http://math.stanford.edu/~yuany/course/data/protein3D.zip>

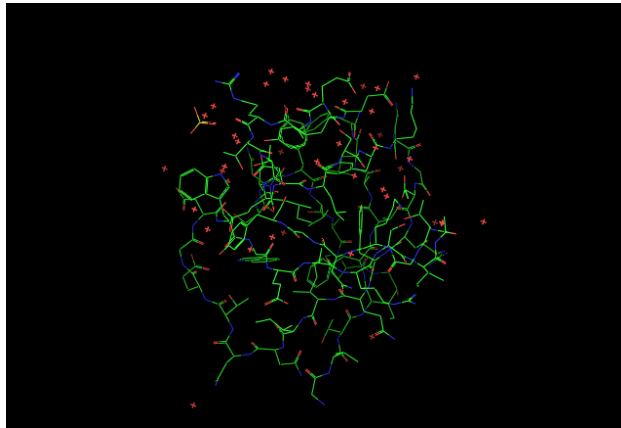


Figure 1: 3D graphs of file PF00018_2HDA.pdb (YES_HUMAN/97-144, PDB 2HDA)

In the file, you will find 3D coordinates for the following three protein families:

PF00013 (PCBP1_HUMAN/281-343, PDB 1WVN),

PF00018 (YES_HUMAN/97-144, PDB 2HDA), and

PF00254 (O45418_CAEEL/24-118, PDB 1R9H).

For example, the file PF00018_2HDA.pdb contains the 3D coordinates of alpha-carbons for a particular amino acid sequence in the family, YES_HUMAN/97-144, read as

VALYDYEAR TTEDLSFKKGERFQIINNTEGDWWEARSIATGKNGYIPS

where the first line in the file is

97 V 0.967 18.470 4.342

Here

- ‘97’: start position 97 in the sequence
- ‘V’: first character in the sequence
- $[x, y, z]$: 3D coordinates in unit \AA .

Figure 1 gives a 3D representation of its structure.

Given the 3D coordinates of the amino acids in the sequence, one can compute pairwise distance between amino acids, $[d_{ij}]^{l \times l}$ where l is the sequence length. A *contact map* is defined to be a graph $G_\theta = (V, E)$ consisting of l vertices for amino acids such that an edge $(i, j) \in E$ if $d_{ij} \leq \theta$, where the threshold is typically $\theta = 5\text{\AA}$ or 8\AA here.

Can you recover the 3D structure of such proteins, up to an Euclidean transformation (rotation and translation), given noisy pairwise distances restricted on the contact map graph G_θ , i.e. given noisy pairwise distances between vertex pairs whose true distances are no more than θ ? Design a noise model (e.g. Gaussian or uniformly bounded) for your experiments.

When $\theta = \infty$ without noise, classical MDS will work; but for a finite θ with noisy measurements, SDP approach can be useful. You may try the matlab package SNLSDP by Kim-Chuan Toh, Pratik Biswas, and Yinyu Ye, downloadable at <http://www.math.nus.edu.sg/~mattokkc/SNLSDP.html>.