Math 420, Spring 2021
First Team Homework

I. (8pts) Consider the text files $kn57Nodes1to57\_exactdist.txt$ and $kn57Nodes1to57\_dist.txt$, both attached to this homework. They are based on the KN57 dataset described here:

`https://people.sc.fsu.edu/~jburkardt/datasets/cities/cities.html`

They contain pairwise distances between $n = 57$ cities. In the first file distances are floating point numbers; in the second file distances are integers. The files have the following format:

```
line 1: n
line 2: d11 d12 d13 ... d1n
line 3: d21 d22 d23 ... d2n
...
line n+1: dn1 dn2 dn3 ... dnn
```

where n denotes the number of vertices of this geometric graph, $d11$,...,dnn represents the pairwise distances between these n points. Note the following: the file $kn57\_exactdist.txt$ contains the noiseless distances (in particular, $dii = 0$); the file $kn57\_dist.txt$ contains approximated measurements of these distances (no guarantee of symmetry or positivity).

Write a Matlab script that performs the following tasks, and apply separately on these two files

1. Read-in the file and create the matrix R of pairwise distances and S of squared-pairwise distances ($S_{k,j} = R^2_{k,j}$);
2. Apply Algorithm 1 to compute the estimated Gramm matrix G;
3. Plot the eigenvalues of G; Print out the first 10 largest eigenvalues;
4. for d=2 and d=3 perform:
   (a) Apply Algorithm 2 to determine a d-dimensional embedding of this geometric graph; call Y the $d \times n$ matrix of coordinates; plot the point cloud and print out the figure;
   (b) Compute the pairwise distances between the d-dimensional points contained in Y: Let $\hat{R}$ be the $n \times n$ matrix whose (k, j) entry is the Euclidean norm

$$\hat{R}_{k,j} = \|Y(1:d,k) - Y(1:d,j)\|_2$$

   Detemine and print the norm $\|R - \hat{R}\|_F$;
   (c) Compute $\varepsilon = \|G - Y^T Y\|_F$, the approximation error; print the result on screen;

(d) Compute $\sigma = \sqrt{\sum_{k=d+1}^{n} \lambda_k^2}$ and print out the result; here, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ are the ordered eigenvalues of G;

(e) Compare ε with σ.

2. (2pts) Denote by Y_{clean} and Y_{noisy} the two estimates matrices of ccordinates obtained by your code at part 1 when run respectively on *kn57_exactdist.txt* and *kn57_dist.txt*. Compute the Frobenius norm $\|Y_{clean} - Y_{noisy}\|_F$.