Math 420, Spring 2021
Second Team Homework

Consider the text files $kn57Nodes1to57\_exactdist.txt$ and $kn57Nodes1to57\_dist.txt$ assigned to the last homework, as well as the new data files entitled $SparseXkn57Nodes1to57\_exactdist.txt$ and $SparseNoisyXNodes1to57kn57\_dist.txt$, with X an integer. You find these files attached to this homework in the archive $SparseNoisy10Nodes1to57kn57\_dist.zip$. The new files (i.e., those starting with 'Sparse..') have the following format:

```
line 1: n m
line 2: i1 j1 d(i1,j1)
line 3: i2 j2 d(i2,j2)
line 4: i3 j3 d(i3,j3)
...
line m+1: im jm d(im,jm)
```

where n denotes the number of vertices of a geometric graph, m denotes the number of available edge distances. Lines 2 to $m+1$ list these distances: $i1, j1, i2, j2, ..., im, jm$ denote vertices (from 1 to n), and $d(i1, j1)$, $d(i2, j2)$, ... , $d(im, jm)$ represent the distances respectively between these pairs of vertices, $(i1, j1)$, $(i2, j2)$, and so on.

The digit X from the file name contains approximatly the fraction of available number of pairwise distances. Roughly, $m = (X/10)n(n-1)/2$ (i.e., $10X$ %).

Files whose names contain 'Noisy' include noisy measurements of these pairwise distances.

Your homework is to develop and implement a Matlab code, that estimates the Gram matrix of the full graph, as well as a 2- and 3-dimensional embeddings of this graph, for each of the 18 data files assigned to this homework. Compare your results to the data embedding using the full set of pairwise distances (i.e., the results from previous homework).

Specifically, write a Matlab script and modularize various computations into functions, and apply on the "Sparse..." named files:

1. Apply the SDP based Algorithm to compute the estimated Gramm matrix G; For this you need to choose a tolerance for the inequality constraints; use two values: $\varepsilon = 0.1$ and $\varepsilon = 1$.
2. Apply Algorithm 1 from previous homework on the full data files (Pair_psb420.dist and MeasuredPair_psb420.dist, respectively) to obtain G_{true} and G_{noisy} as the "target" Gram matrices.
3. Compute the Frobenius norm of the difference between your estimated G and the "target" Gram matrix. Let $Error(X, \varepsilon) = \|G - G_{true}\|_F$ and $ErrorNoisy(X, \varepsilon) = \|G - G_{noisy}\|_F$ denote the respective errors. The two matrices are indexed by X, that runs from 1 to 9, and ε, the two values (or more if needed) indicated above.

4. For each value of ε and on same figure, plot $Error$ and $ErrorNoisy$ as function of X, the percentage of available distances. You should obtain two figures, one for each value of ε.

In addition to the plots requested above, your homework should contain answers to the following questions:

1. Can you determine the decay rate of error vs. number ofdistances m? Is it of the form $Error \sim \frac{1}{m^\alpha}$ or $Error \sim e^{-\beta m}$? Can you estimate α and β that best fit in some sense (e.g., least-squares but in log-log plot, for polynomial fitting, and semi-log for the exponential fitting).
2. What difference does ε make?
3. It is possible that CVX might not find a feasible solution if ε is too small. Comment on the trade-off between precision (feasibility) and feasibility of finding solutions.