Can We Recover a Map from Distances Alone?

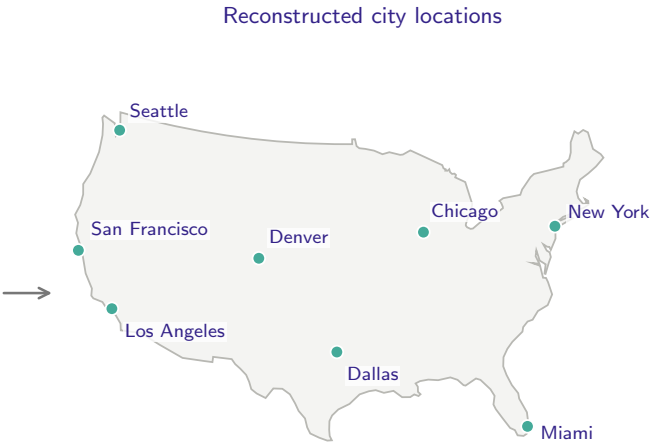
Due: Wednesday, November 11 at 11:59pm

Instructions: Please submit your *handwritten* solutions as a single PDF on Gradescope.

We know the distances between cities, but their coordinates are missing. Let's reconstruct a map using those distances.

	Seattle	San Francisco	Los Angeles	Denver	Dallas	Chicago	New York	Miami
Seattle	0	1.2	2.4	2.7	7.3	7.8	15.1	19.3
San Francisco	1.2	0	0.3	2.5	5.9	9.1	17.3	17.7
Los Angeles	2.4	0.3	0	1.8	4.0	7.8	15.4	14.2
Denver	2.7	2.5	1.8	0	1.1	2.1	6.7	7.6
Dallas	7.3	5.9	4.0	1.1	0	1.7	4.8	3.2
Chicago	7.8	9.1	7.8	2.1	1.7	0	1.3	3.7
New York	15.1	17.3	15.4	6.7	4.8	1.3	0	3.1
Miami	19.3	17.7	14.2	7.6	3.2	3.7	3.1	0

Squared map distances (10^6 km^2)



Two coordinates per city

Entries are rounded squared distances in a flat map; reconstruction uses unrounded values. The recovered points are oriented and positioned on the outline for comparison. City locations: U.S. Census Bureau; outline: Natural Earth.

More generally, let $\mathbf{X} \in \mathbb{R}^{n \times d}$ have rows $\mathbf{x}_i^\top$ and satisfy $\sum_{i=1}^n \mathbf{x}_i = \mathbf{0}$. We only observe $\mathbf{D} \in \mathbb{R}^{n \times n}$, where

$$D_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|_2^2 = \|\mathbf{x}_i\|_2^2 + \|\mathbf{x}_j\|_2^2 - 2\mathbf{x}_i^\top \mathbf{x}_j.$$

Define the row means and overall mean by

$$\bar{D}_i = \frac{1}{n}\sum_{j=1}^n D_{ij}, \qquad \bar{D} = \frac{1}{n^2}\sum_{i=1}^n \sum_{j=1}^n D_{ij}.$$

1. Since the coordinates are centered, the average of $\mathbf{x}_i^\top \mathbf{x}_j$ over j is zero. How can we express $\|\mathbf{x}_i\|_2^2$ using only $\bar{D}_i$ and $\bar{D}$?

2. The distance formula relates the squared norms to the inner products. How can we recover entry (i, j) of $\mathbf{G} = \mathbf{X}\mathbf{X}^\top$ from the distances?

3. Write $\mathbf{G} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top$, with eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n \geq 0$. For $1 \leq k \leq \text{rank}(\mathbf{X})$, which coordinates $\mathbf{Z} \in \mathbb{R}^{n \times k}$ minimize $\|\mathbf{G} - \mathbf{Z}\mathbf{Z}^\top\|_F^2$? Why do these coordinates recover every original distance when we keep all the positive eigenvalues?

Takeaway: *Pairwise distances can reveal the geometry of data even when the original features are unavailable. A few eigenvectors can turn those relationships back into coordinates.*