

# Homework #6

Eric Tao  
Math 123: Homework #6

March 11, 2023

**Question 1.** Let  $L = D - W$  be the unnormalized graph Laplacian associated to a graph  $\mathcal{G} = (V, W)$  on points  $V = \{x_i\}_{i=1}^n$  with symmetric weight matrix  $W$  and diagonal degree matrix  $D$ . Let  $\{C, \bar{C}\}$  be any partition of  $V$ , and let:

$$f_i^C = \begin{cases} -\sqrt{\text{Vol}(\bar{C})/\text{Vol}(C)} & \text{if } x_i \in C \\ \sqrt{\text{Vol}(C)/\text{Vol}(\bar{C})} & \text{if } x_i \in \bar{C} \end{cases}$$

- (a) Prove that  $\langle Df^C, \mathbf{1} \rangle = 0$ .
- (b) Prove that  $(f^C)^T Df^C = \text{Vol}(V)$ .
- (c) Prove that  $(f^C)^T Lf^C = \text{Vol}(V) \cdot \text{Ncut}(C, \bar{C})$ .

*Solution.* (a)

We notice, that this amounts to summing up the components of  $Df^C$ . First, consider some  $x_i$  such that  $x_i \in C$ . Then, we have that, with the notation  $W(x_i, x_j) = w_{ij}$ :

$$(Df^C)_i = \sum_{x_j \in \mathcal{G}} W(x_i, x_j) \cdot -\sqrt{\text{Vol}(\bar{C})/\text{Vol}(C)}$$

Summing over  $x_i \in C$ , we find that:

$$\sum_{x_i \in C} (Df^C)_i = \sum_{x_i \in C} \sum_{x_j \in \mathcal{G}} W(x_i, x_j) \cdot -\sqrt{\text{Vol}(\bar{C})/\text{Vol}(C)}$$

Here, we notice though, that, by definition, the double summation is exactly equal to  $\text{Vol}(C)$ ! Thus, we have that:

$$\sum_{x_i \in C} (Df^C)_i = -\text{Vol}(C) \cdot \sqrt{\text{Vol}(\bar{C})/\text{Vol}(C)} = -(\text{Vol}(C)\text{Vol}(\bar{C}))^{1/2}$$

It is not hard to see that for  $x_i \in \bar{C}$ , that we get a similar result:

$$\sum_{x_i \in \bar{C}} (Df^C)_i = \text{Vol}(\bar{C}) \cdot \sqrt{\text{Vol}(C)/\text{Vol}(\bar{C})} = (\text{Vol}(C)\text{Vol}(\bar{C}))^{1/2}$$

Thus, summing over all  $x_i$  in  $(Df^C)_i$ , we find that:

$$\sum_{x_i \in \mathcal{G}} (Df^C)_i = \sum_{x_i \in C} (Df^C)_i + \sum_{x_i \in \bar{C}} (Df^C)_i = -(\text{Vol}(C)\text{Vol}(\bar{C}))^{1/2} + (\text{Vol}(C)\text{Vol}(\bar{C}))^{1/2} = 0$$

as desired.

(b)

In a similar fashion to (a), we will look at this term by term. We already found that, for  $x_i \in C$ , that  $(Df^C)_i = \sum_{x_j \in \mathcal{G}} W(x_i, x_j) \cdot -\sqrt{\text{Vol}(\overline{C})/\text{Vol}(C)}$ . Then, we have that in the inner product with  $(f^C)^T$ , the component that corresponds to this component will multiply to find:

$$-\sqrt{\text{Vol}(\overline{C})/\text{Vol}(C)} \cdot \left( \sum_{x_j \in \mathcal{G}} W(x_i, x_j) \cdot -\sqrt{\text{Vol}(\overline{C})/\text{Vol}(C)} \right) = \frac{\text{Vol}(\overline{C})}{\text{Vol}(C)} \cdot \sum_{x_j \in \mathcal{G}} W(x_i, x_j)$$

Then, summing over the  $x_i \in C$ , we find that:

$$\sum_{x_i \in C} \frac{\text{Vol}(\overline{C})}{\text{Vol}(C)} \cdot \sum_{x_j \in \mathcal{G}} W(x_i, x_j) = \frac{\text{Vol}(\overline{C})}{\text{Vol}(C)} \sum_{x_i \in C} \sum_{x_j \in \mathcal{G}} W(x_i, x_j) = \frac{\text{Vol}(\overline{C})}{\text{Vol}(C)} \text{Vol}(C) = \text{Vol}(\overline{C})$$

Without writing it out, it should be clear that in the same way, if  $x_i \in \overline{C}$ , then summing over those, we retrieve  $\text{Vol}C$ .

Thus, we have that:

$$\begin{aligned} (f^C)^T Df^C &= \sum_{x_i \in C} \left( \frac{\text{Vol}(\overline{C})}{\text{Vol}(C)} \cdot \sum_{x_j \in \mathcal{G}} W(x_i, x_j) \right) + \sum_{x_i \in \overline{C}} \left( \frac{\text{Vol}(C)}{\text{Vol}(\overline{C})} \cdot \sum_{x_j \in \mathcal{G}} W(x_i, x_j) \right) = \text{Vol}(\overline{C}) + \text{Vol}(C) = \\ &\sum_{x_i \in \overline{C}, x_j \in \mathcal{G}} W(x_i, x_j) + \sum_{x_i \in C, x_j \in \mathcal{G}} W(x_i, x_j) = \sum_{x_i, x_j \in \mathcal{G}} W(x_i, x_j) = \text{Vol}(G) \end{aligned}$$

(c)

We recall from homework 5, question 2, we found that for any vector  $y$ , that  $y^T L y = \frac{1}{2} \sum_i \sum_j w_{ij} (y_i - y_j)^2$ . Applying this for  $y = f^C$ , we have that:

$$(f^C)^T L f^C = \frac{1}{2} \sum_i \sum_j w_{ij} ((f^C)_i - (f^C)_j)^2$$

Clearly, if  $i = j$ , this value is 0. Further, if  $x_i, x_j \in C$ , or if  $x_i, x_j \in \overline{C}$ , then of course we have that  $(f^C)_i = (f^C)_j$ , and thus disappears. Lastly, we notice that this double counts everything, and that under interchange of  $x_i, x_j$ , the value is unchanged due to for  $x_i \in C, x_j \in \overline{C}$ ,  $((f^C)_i - (f^C)_j)^2 = ((f^C)_j - (f^C)_i)^2$ . So, we can say that:

$$\begin{aligned} (f^C)^T L f^C &= \frac{1}{2} \sum_i \sum_j w_{ij} ((f^C)_i - (f^C)_j)^2 = \sum_{x_i \in C, x_j \in \overline{C}} w_{ij} \left( \sqrt{\frac{\text{Vol}(\overline{C})}{\text{Vol}(C)}} + \sqrt{\frac{\text{Vol}(C)}{\text{Vol}(\overline{C})}} \right)^2 = \\ &\sum_{x_i \in C, x_j \in \overline{C}} w_{ij} \left( \frac{\text{Vol}(\overline{C}) + \text{Vol}(C)}{\sqrt{\text{Vol}(C)\text{Vol}(\overline{C})}} \right)^2 = \sum_{x_i \in C, x_j \in \overline{C}} w_{ij} \text{Vol}(G) \frac{\text{Vol}(\overline{C}) + \text{Vol}(C)}{\text{Vol}(C)\text{Vol}(\overline{C})} = \\ &\text{Vol}(G) \sum_{x_i \in C, x_j \in \overline{C}} w_{ij} \left( \frac{1}{\text{Vol}(C)} + \frac{1}{\text{Vol}(\overline{C})} \right) = \text{Vol}(G) \cdot \text{Ncut}(C, \overline{C}) \end{aligned}$$

as desired. □

**Question 2.** Recall that one construction of the weight matrix for a graph on data points  $\{x_i\}_{i=1}^n$  is to use the Gaussian kernel

$$W_{ij} = \begin{cases} \exp(-\|x_i - x_j\|_2^2 / \sigma^2) & i \neq j \\ 0 & i = j \end{cases}$$

for some choice of  $\sigma > 0$ .

(a) What happens to the resulting Laplacian matrix as  $\sigma \rightarrow 0$ ?

(b) What happens to the resulting Laplacian matrix as  $\sigma \rightarrow \infty$ ?

*Solution.* (a)

Qualitatively, we notice that as  $\sigma \rightarrow 0$ , then this amounts to narrowing the Gaussian, and making it sharper around  $\|x_i - x_j\|_2^2 = 0$ . We can imagine in the limiting case that it has the same effect as the Dirac measure, where we assign the value to be 1 if and only if the values are the same. In such a limiting case, we would see that the Laplacian matrix would converge towards a matrix of all 0s (assuming each data point takes on distinct values), as we may pick a  $\sigma$  such that  $\|x_i - x_j\|_2^2 < \epsilon$ .

Formally: Let  $\epsilon > 0$  be given. Since we know that as  $t \rightarrow \infty$ ,  $e^{-t} \rightarrow 0$ , we can choose  $\delta$  such that so long as  $t > \delta$ ,  $e^{-t} < \epsilon$ .

WLOG, choose  $x_i = 0$ , and fix some  $x_j$  such that  $x_j \neq x_i$ . Of course the sequence  $\|x_j\|_2^2 n \rightarrow \infty$  for  $n \in \mathbb{N}$ ,  $n \rightarrow \infty$ . Thus, we may choose  $N \in \mathbb{N}$  such that for all  $m > N$ ,  $\|x_j\|_2^2 m > \delta$ . Further, since we know that as  $x \rightarrow 0$ ,  $1/x \rightarrow \infty$ , we may choose  $L$  such that for all  $l > L$ ,  $1/l > m$ .

Choose  $\sigma$  such that  $\sigma > L + 1$ . Then, we have that:

$$\frac{\|x_j\|_2^2}{\sigma^2} > \frac{\|x_j\|_2^2}{(L+1)^2} > \|x_j\|_2^2 m > \delta$$

Thus, we have that for this choice of  $\sigma$ ,  $e^{-\|x_j\|_2^2 / \sigma^2} < \epsilon$ . Since this can be done for any  $x_i \neq x_j$ , we have that these weights go to 0 for all distinct points.

We can then conclude that all edges between distinct points go to a weight of 0, which implies that, assuming our points are distinct, and that we do not consider any edges from  $x_i \rightarrow x_i$ , that is, any loops, that  $L$  converges to the 0 matrix.

(b)

In the opposite case, where we take  $\sigma \rightarrow \infty$ , we expect instead for edge weights to converge to the value 1. WLOG, choose  $x_i = 0$ , and fix some  $x_j$ . Let  $\epsilon > 0$  be given. Since the exponential is continuous, there exists  $\delta > 0$  such that  $\|1 - e^t\| < \epsilon$  for all  $|t| < \delta$ .

Now, by the Archimidean principle, there exists an  $N \in \mathbb{N}$  such that  $\frac{\|x_j\|_2^2}{N} < \delta$ . Consider  $\sigma > N$ . Then, we have that:

$$\frac{\|x_j\|_2^2}{\sigma^2} \leq \frac{\|x_j\|_2^2}{N^2} < \frac{\|x_j\|_2^2}{N} < \delta$$

Thus, for any  $\sigma > N$ , we have that  $\|1 - e^{\|x_j\|_2^2 / \sigma^2}\| < \epsilon$ , and we are done. Since this can be done for all  $x_i, x_j$  in the limit, this works for every point in our data set, where we're not concerned about issues on the supremum being equal to  $\infty$ , since we are in the limiting cases.

Thus, in terms of the Laplacian matrix, we expect things to converge to  $n$  on the diagonal, and  $-1$  on every other entry, where  $n$  is the dimensionality of the Laplacian, or, equivalently, the number of data points.  $\square$

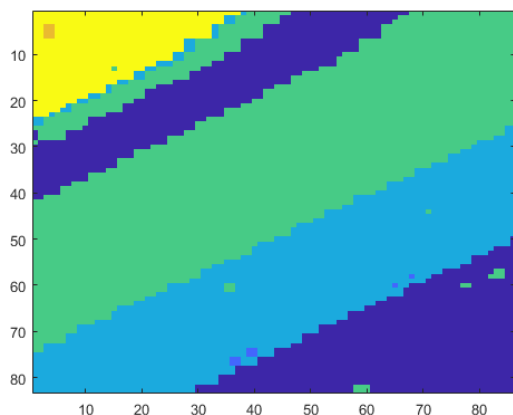
**Question 3.** Load the dataset "SalinasA\_corrected.mat" and "Salinas-S-groundtruth".

(a) Run spectral clustering on this data, using a sparse Laplacian with different numbers of nearest neighbors and  $K = 6$  clusters. How do the results compare to the ground truth data?

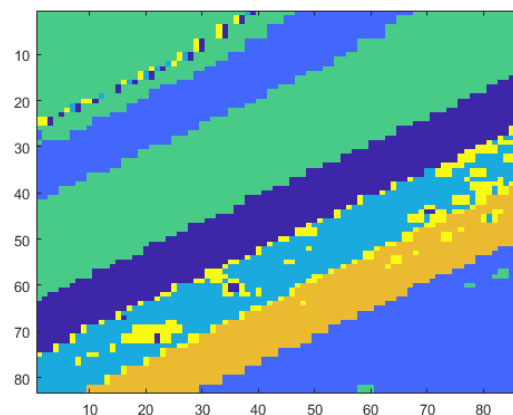
(b) Plot the first 10 eigenvalues of the data for different choices of  $\sigma$ . What does the eigengap estimate as the number of cluster for these choices of  $\sigma$ ?

(c) Compare the projections onto the first three principal components with the first three Laplacian eigenvectors by plotting both sets in different figures using 'scatter3'. How do the representations differ qualitatively?

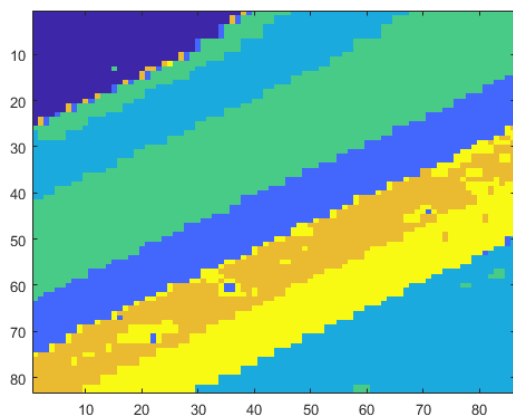
*Solution.* Here follows some spectral clustering on this data, with a sparse Laplacian, 6 clusters, and various numbers of nearest neighbors, with a fixed value of  $\sigma = 10000$ :



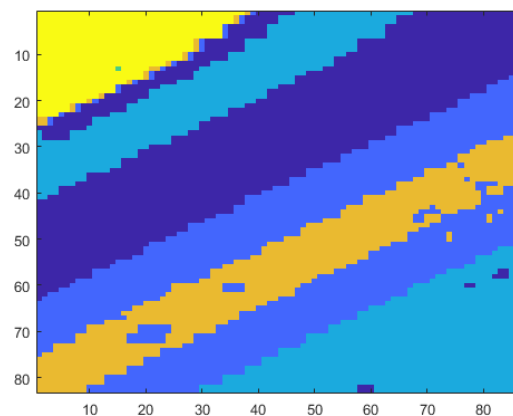
**Figure 1:** 10 Nearest Neighbors



**Figure 2:** 50 Nearest Neighbors

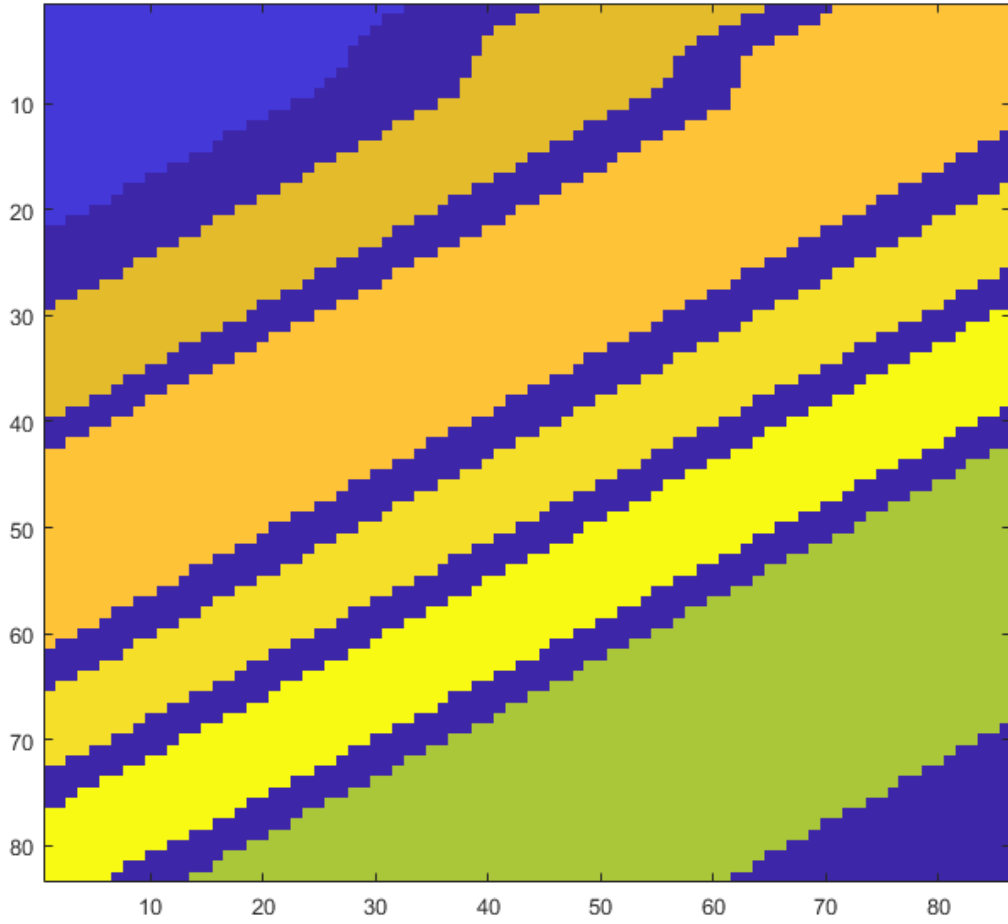


**Figure 3:** 100 Nearest Neighbors



**Figure 4:** 500 Nearest Neighbors

In comparison, this is the groundtruth:

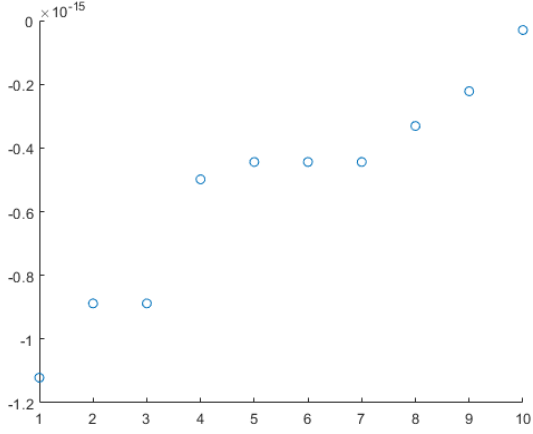


**Figure 5:** Salinas A Groundtruth

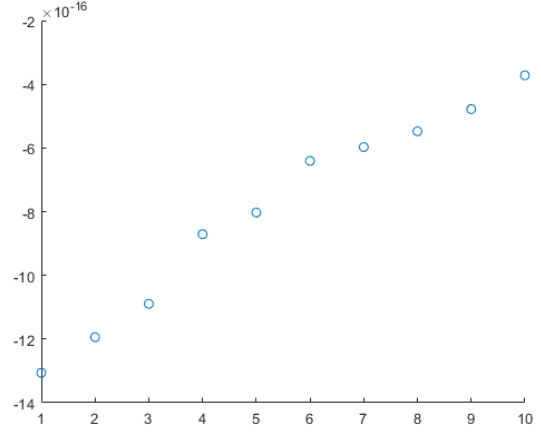
We can see the structures pretty easily, and qualitatively, we have a decent match. There is definitely some tuning that could be done here, the boundary can be messy, and perhaps we need better separation of points with a higher  $\sigma$  value.

(b)

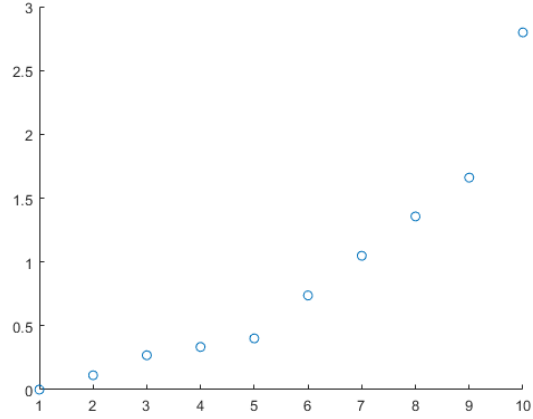
Here follows the first 10 eigenvalues for specific choices of  $\sigma$ , and a fixed 500 nearest neighbors:



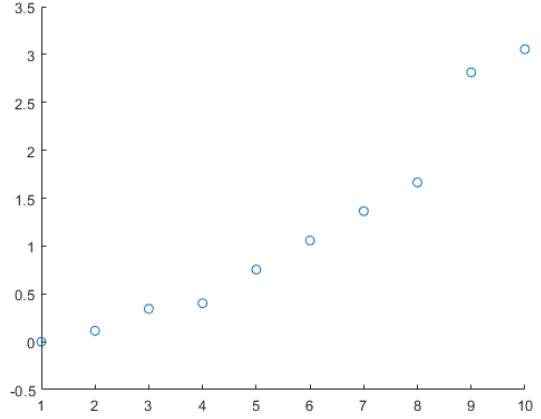
**Figure 6:**  $\sigma = 1$



**Figure 7:**  $\sigma = 100$



**Figure 8:**  $\sigma = 10000$



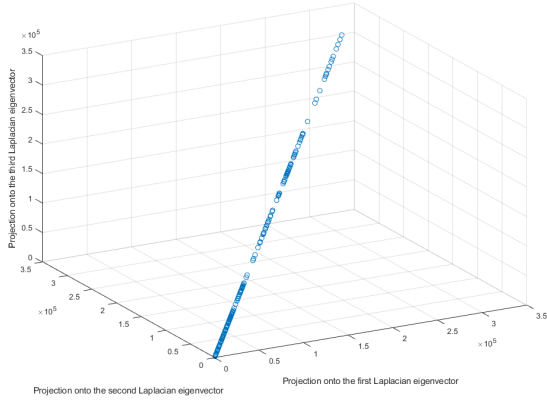
**Figure 9:**  $\sigma = 1000000$

At the scale of  $\sigma = 1, 100$ , the values are close to float rounding errors so probably should be discarded, but they seem to suggest 3 clusters, looking at the jump from the 3rd to the 4th eigenvalue.

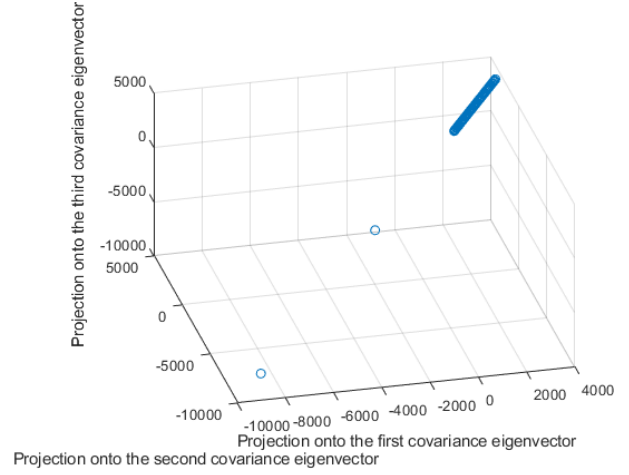
At the scale of  $\sigma = 10000, 1000000$ , we see a clear jump at  $9 \rightarrow 10$  for 10k, and  $8 \rightarrow 9$  for  $\sigma = 1000000$ . This seems to suggest that as we increase  $\sigma$ , we get more separation in the data, and thus more reasonable distinct clustering.

(c)

Here, we compare projection onto the Laplacian eigenvectors for 100 nearest neighbors,  $\sigma = 10000$ , and PCA:



**Figure 10:** Projection onto the 3 Laplacian eigenvectors with smallest eigenvalues



**Figure 11:** Projection onto the 3 principal components of largest variance

Qualitatively, we notice that the Laplacian projections are strictly positive, whereas the projections onto the eigenvalues of the covariance matrix can take on both positive or negative values. Structurally, they seem to be very similar, but it is fairly hard to tell.

□