1

1.a

We have the starting form:

$$min\frac{1}{n}\sum_{i=1}^{n}l_i(y_i, w^Tx_i+b)+\lambda||w||^2 \tag{1}$$

Since only one side of the support vector is active at a time we can use the constraints to show that:

$$\xi^+ + \xi^- = max(|y_i - w^Tx_i - b| - \epsilon, 0) \tag{2}$$

$$min\frac{1}{n}\sum_{i=1}^{n}\xi^+ + \xi^- + \lambda||w||^2 \tag{3}$$

Next we pull the $\lambda||w||^2$ out:

$$min\frac{\lambda}{n}||w||^2 + \frac{1}{n}\sum_{i=1}^{n}\xi^+ + \xi^- \tag{4}$$

$$min.\frac{C\lambda}{n}||w||^2 + \frac{C}{n}\sum_{i=1}^{n}\xi^+ + \xi^- \tag{5}$$

$$min\frac{1}{2}||w||^2 + \frac{C}{n}\sum_{i=1}^{n}\xi^+ + \xi^- \tag{6}$$

Therefore:

$$\frac{C\lambda}{n} = \frac{1}{2} \tag{7}$$

$$\lambda = \frac{n}{2C} \tag{8}$$

1.b

Set up the Legrangian:

$$L = \frac{C}{n}\sum_{i=1}^{n}\xi^+ + \xi^- + \frac{1}{2}||w||^2 - \sum_{i=1}^{n}\alpha_i^+[y_i - w^Tx_i - b + \epsilon + \xi_i^+] - \sum_{i=1}^{n}\alpha_i^+[-y_i + w^Tx_i + b - \epsilon + \xi_i^-] - \sum_{i=1}^{n}\beta_i^+\xi_i^+ + \beta_i^-\xi_i^- \tag{9}$$

Apply KKT conditions:

$$\frac{\partial L}{\partial \xi_i^+} = \frac{C}{n} - \alpha_i^+ - \beta_i^+ \tag{10}$$

$$0 = \frac{C}{n} - \alpha_i^+ - \beta_i^+ \tag{11}$$

Since $\beta_i^+ > 0$

$$\alpha_i^+ \leq \frac{C}{n} \tag{12}$$

Following the same logic, we get:

$$\frac{\partial L}{\partial \xi_i^-} = \frac{C}{n} - \alpha_i^- - \beta_i^- \tag{13}$$

$$\alpha_i^- \leq \frac{C}{n} \tag{14}$$

$$\frac{\partial L}{\partial b} = \sum_{i=1}^{n} (\alpha_i^+ - \alpha_i^-) = 0 \tag{15}$$

$$\frac{\partial L}{\partial w} = w + \sum_{i=1}^{n} (\alpha_i^+ - \alpha_i^-) x_i = 0 \tag{16}$$

Using these constraints:

$$||w||^2 = ||\sum_{i=1}^{n} (\alpha_i^+ - \alpha_i^-) x_i||^2 \tag{17}$$

$$(\sum_{i=1}^{n} (\alpha_i^+ - \alpha_i^-) x_i)^T (\sum_{i=1}^{n} (\alpha_i^+ - \alpha_i^-) x_i) \tag{18}$$

Plugging back in:

$$L = \sum_{i=1}^{n} \alpha_i^+ + \alpha_i^- + \frac{1}{2} (\sum_{i=1}^{n} (\alpha_i^+ - \alpha_i^-) x_i)^T (\sum_{i=1}^{n} (\alpha_i^+ - \alpha_i^-) x_i)$$
$$- \sum_{i=1, j=1}^{n} \alpha_i^+ \alpha_j^+ y_i y_j < x_i, x_j >$$
$$- \sum_{i=1, j=1}^{n} \alpha_i^- \alpha_j^- y_i y_j < x_i, x_j >$$

Combining:

$$L = \sum_{i=1}^{n} \alpha_i^+ + \alpha_i^-$$

$$+\frac{1}{2} \sum_{i=1,j=1}^{n} \alpha_i^+ \alpha_j^+ y_i y_j < x_i, x_j >$$
$$+\frac{1}{2} \sum_{i=1,j=1}^{n} \alpha_i^- \alpha_j^- y_i y_j < x_i, x_j >$$
$$- \sum_{i=1,j=1}^{n} \alpha_i^+ \alpha_j^+ y_i y_j < x_i, x_j >$$
$$- \sum_{i=1,j=1}^{n} \alpha_i^- \alpha_j^- y_i y_j < x_i, x_j >$$

Simplifying:

$$max(\alpha^+, \alpha^-) = \sum_{i=1}^{n} \alpha_i^+ + \alpha_i^-$$

$$-\frac{1}{2} \sum_{i=1,j=1}^{n} \alpha_i^+ \alpha_j^+ y_i y_j < x_i, x_j >$$
$$-\frac{1}{2} \sum_{i=1,j=1}^{n} \alpha_i^- \alpha_j^- y_i y_j < x_i, x_j >$$

Once More:

$$max(\alpha^+, \alpha^-) = \sum_{i=1}^{n} \alpha_i^+ + \alpha_i^-$$

$$-\frac{1}{2} \sum_{i=1,j=1}^{n} (\alpha_i^+ \alpha_j^+ + \alpha_i^- \alpha_j^-) y_i y_j < x_i, x_j >$$

1.c

We kernelize an objective function by turning it into an inner product. For SVR using equation 17:

$$||w^2|| = || \sum_{i=1}^{n} (\alpha_i^+ - \alpha_i^-) x_i^2 || \tag{19}$$

$$||w^2|| = ||(\alpha^+ - \alpha^-)x^T x|| \tag{20}$$

$$||w^2|| = ||(\alpha^+ - \alpha^-) < x, x > || \tag{21}$$

We can recover w* using the kkt condition:

$$w + \sum_{i=1}^{n}(\alpha_i^+ - \alpha_i^-)x_i = 0 \tag{22}$$

$$w* = -\sum_{i=1}^{n}(\alpha_i^{*+} - \alpha_i^{*-})x_i \tag{23}$$

For b*:

$$b* = y_i - < w^*, x_i >= y_i + \sum_{j=1}^{n}(\alpha_i^{*+} - \alpha_i^{*-}) < x_j, x_i > \tag{24}$$

The final SVR equation in terms of w* and b*:

$$f(x) = -\sum_{i=1}^{n}(\alpha_i^{*+} - \alpha_i^{*-}) < x_j, x > +b* \tag{25}$$

1.d

As only α^+ or $\alpha-$ will be active at a time, that means that looking at either independently will yield the same boundary line. Or, only α^+ or $\alpha-$ exclusively will be needed.

2

2.a

Show that:

$$min \sum_{i=1}^{n} ||x_i - \mu - A\theta_i||^2 = n \sum_{j=k+1}^{d} \lambda_j \tag{26}$$

From slides:

$$min \sum_{i=1}^{n}(x_i - \hat{x}_i)^2 \tag{27}$$

We can then replace this with the reconstruction error:

$$err_k = \sum_{i=1}^{n}[(\bar{x} + \sum_{j=k+1}^{d} A_{ji}u_j) - (\bar{x} + \sum_{j=k+1}^{d} A_{ji}u_j)]^2 \tag{28}$$

$$\sum_{i=1}^{n}[(\sum_{j=1}^{d} A_{ji}u;_j) - (\sum_{j=1}^{d} A_{ji}u_j)]^2 \tag{29}$$

$$\sum_{i=1}^{n}[\sum_{j=k+1}^{d} A_{ji}u_j]^2 \tag{30}$$

$$n[\sum_{j=k+1}^{d} A_j]^2 U^T U \tag{31}$$

Because A is PSD:

$$n\sum_{j=k+1}^{d} A_j \tag{32}$$

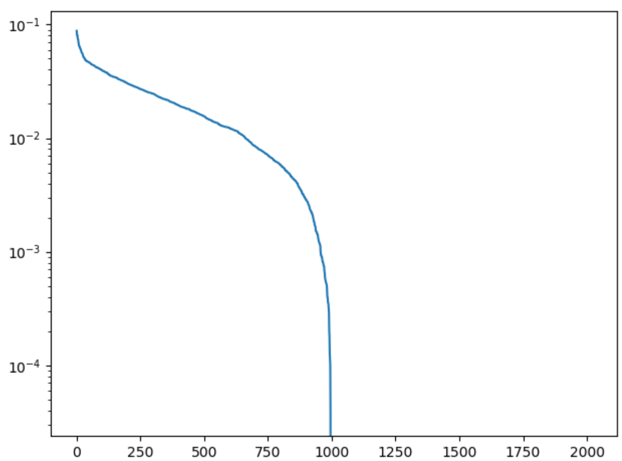
By definition Au = λu:

$$n\sum_{j=k+1}^{d} \lambda_j \tag{33}$$

2.b

3

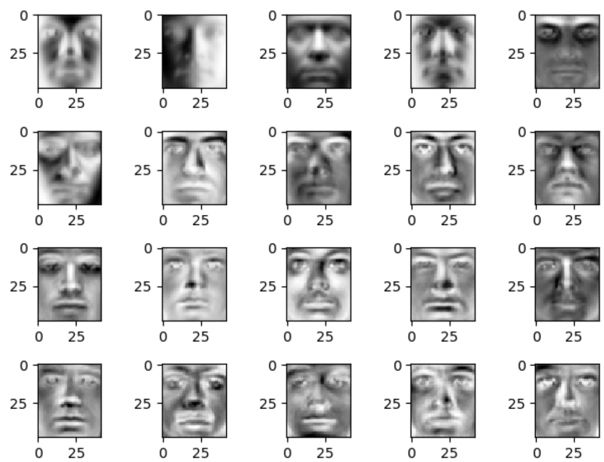
3.a



Build PCA done Plot cm done 59 to hit 95% variance 201 to hit 99% variance 1 - 59/2414 = 97.56% dimension reduction 1 - 201/2414 = 91.67% dimension reduction

3.b

The light seems to come from different sides of the person. It also seems to on



some be more spotlighted.

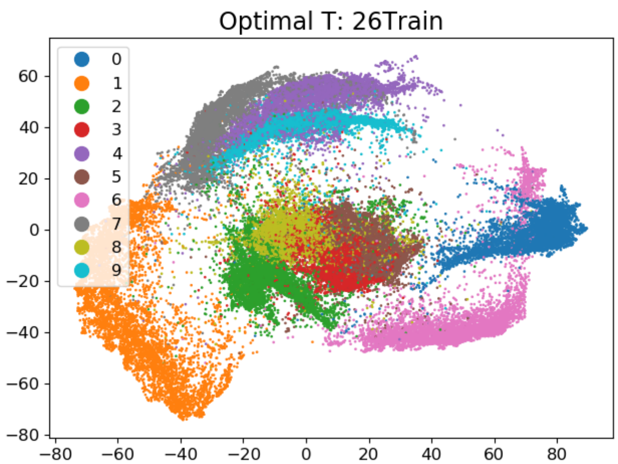
3.c

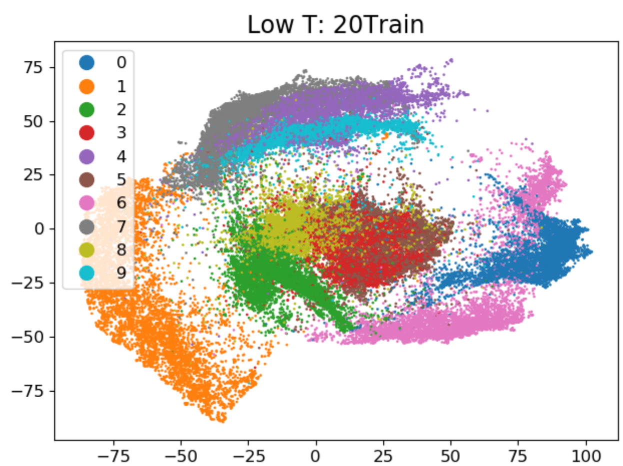
Code Turned In

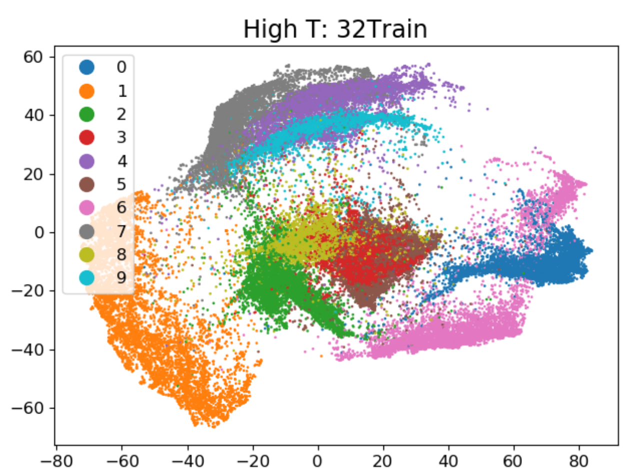
4

4.a

For training: t = 26, chosen t_lower = 20, chosen t_higher = 32
The optimal t seems to give the best representation based on the images. The relative position does make sense at least to me. 0 and 1 for example are as different as can be, and they are the furthest apart digits. As well, 9, 7, and 4 are fairly close which also makes sense as they have similar characteristics.

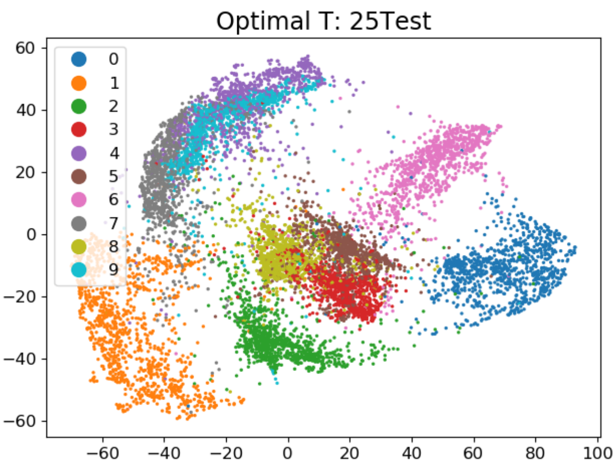

Optimal T: 26Train
0
1
2
3
4
5
6
7
8
9
60
40
20
0
−20
−40
−60
−80
−80 −60 −40 −20 0 20 40 60 80

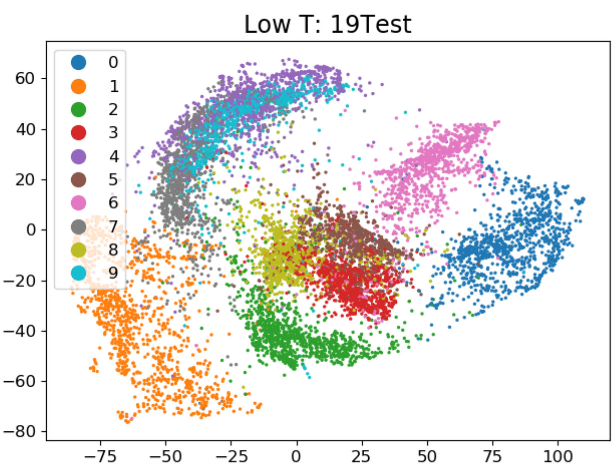

Low T: 20Train
0
1
2
3
4
5
6
7
8
9
75
50
25
0
−25
−50
−75
−75 −50 −25 0 25 50 75 100

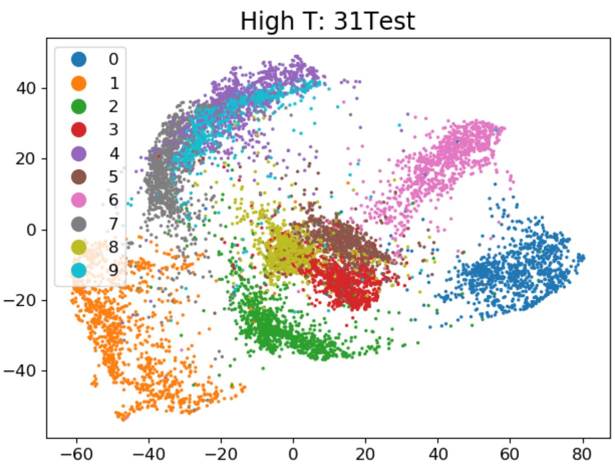

High T: 32Train
0
1
2
3
4
5
6
7
8
9
60
40
20
0
−20
−40
−60
−80 −60 −40 −20 0 20 40 60 80

4.b

For testing: t = 25, chosen t_lower = 19, chosen t_higher = 31
The images look fairly similar. Also, the t value was fairly similar. The training data is simply giving a more accurate representation as it has more samples to pull from. The location of digit 6 did move quite a bit though. The boundaries of 9,7, and 4 are also less defined.

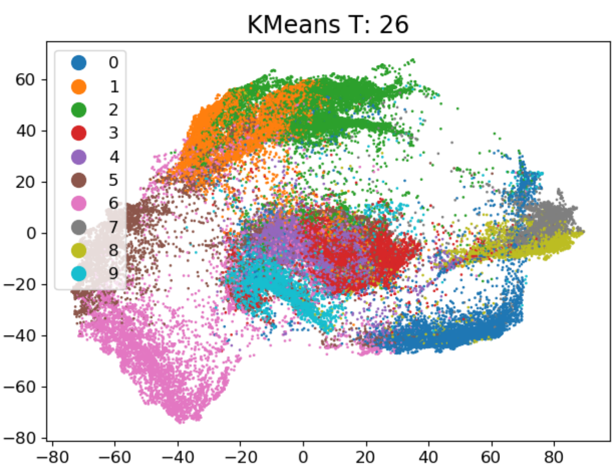






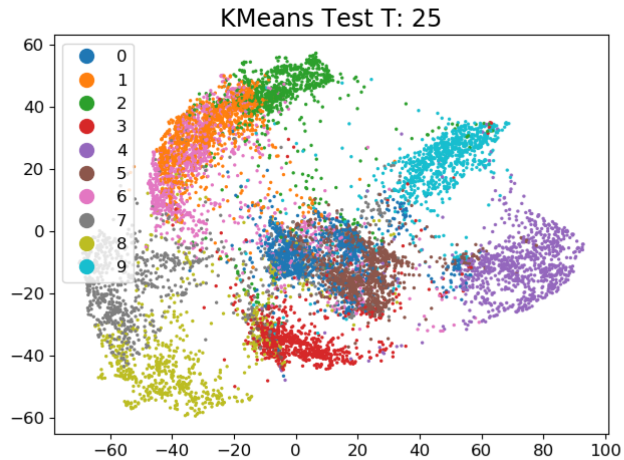
4.c

KMeans ARI: 0.36. 0.36 is not a great score for ARI. The image places 7 and 8 as almost overlapping which is bad because 7 and 8 really do not look alike feature wise.



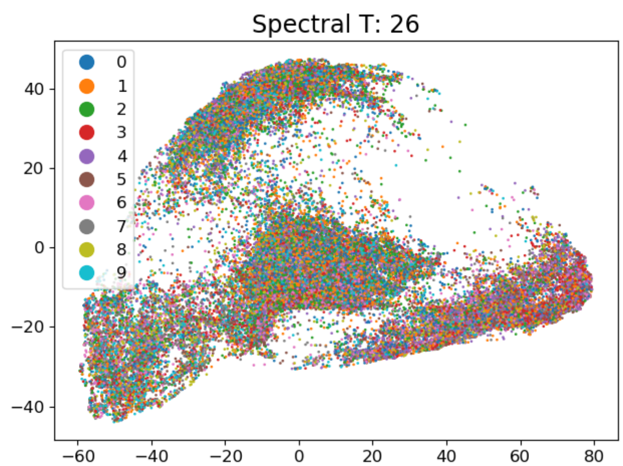
4.d

KMeans ARI: 0.38. Same as part c - 0.38 is not a great score for ARI. The image again places 7 and 8 next to each other. It also places 1 and 6 next to each other which tend to not look alike. So, Kmeans seems to not be good for clustering this particular data set.



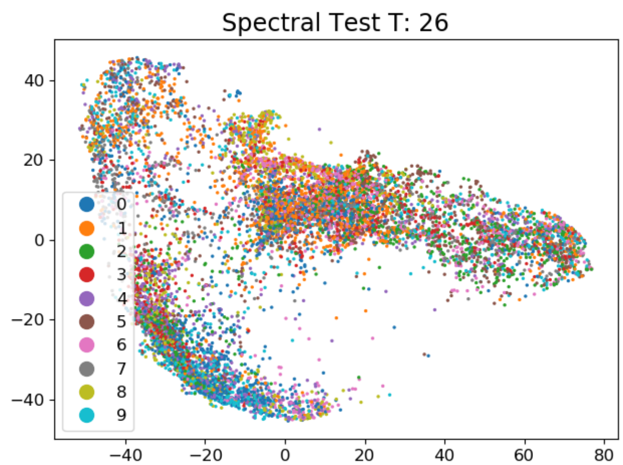
4.e

ARI: almost 0. -7e-7
ARI with bandwidth selection: .426
It performed better than kmeans, but still not great. Ideally a score a bit closer to 1 or -1 would be ideal.



4.f

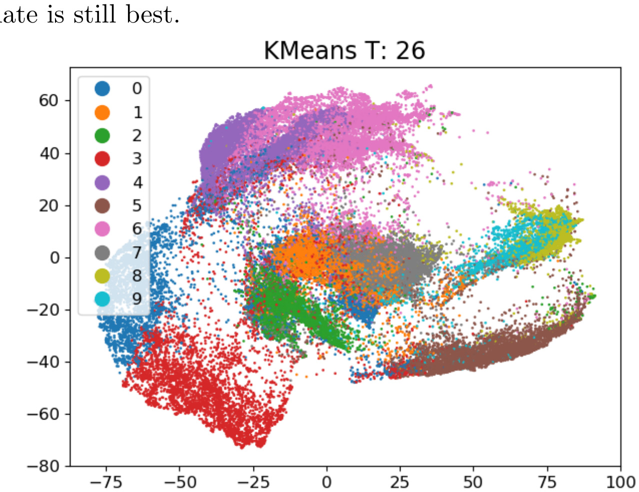
ARI: almost 0. -5e-5
ARI with bandwidth selection: .502
It performed better than kmeans, but still not great. Ideally a score a bit closer to 1 or -1 would be ideal. This is still not as good as phate though.



4.g

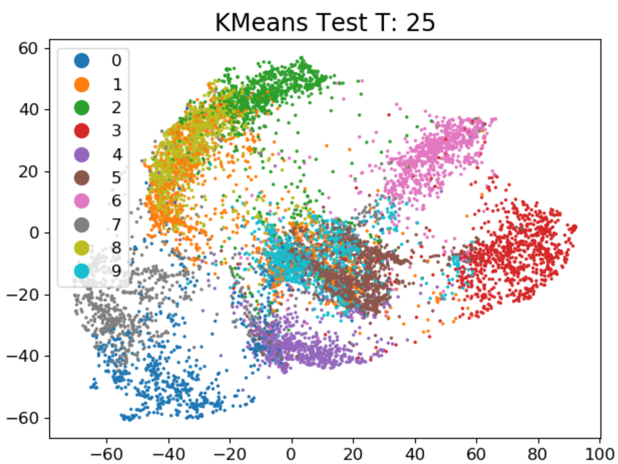
For training: t = 26
KMeans ARI: 0.3605577808903296 Spectral ARI: almost 0. -1e-6
Spectral ARI with bandwidth selection: .429

The Phate representation is still the best. Graphically, there is a lot more overlap with the this 10 dim kmeans though it performed better than 2 dim kmeans. Spectral Clustering had a better ARI still then both of the kmeans. Phate is still best.



4.h

For test: t = 25
KMeans ARI: 0.37795051682516406
Spectral ARI: almost 0. -5e-5
Spectral ARI with bandwidth selection: .502



It is odd that the position of almost every number changed, but it wasn't too different in terms of accuracy from using the training data.

4.i

You could tune it graphically by checking to see if the clusters are more or less tightly coupled. This of course will have some challenges. You could also run multiple clustering strategies and use the most common value in an ensemble.

4.j

Turn in code

5

Prove a relaxed NCut is solved by normalized spectral decomposition.

Original Laplacian:

$$f^{'}Lf \tag{34}$$

With constraints via Raleigh Ritz Theorem:

$$Df \perp 1 \tag{35}$$

Also:

$$f'Df = vol(v) \tag{36}$$

Relaxed Laplacian Using:

$$g = D^{\frac{1}{2}}f \tag{37}$$

$$g'D^{\frac{1}{2}}LD^{\frac{1}{2}}g \tag{38}$$

With constraints applying Raleigh Ritz Theorem on transformed Matrix:

$$g \perp D^{\frac{1}{2}}1 \tag{39}$$

Also:

$$||g||^2 = vol(v) = f^{'}Df \tag{40}$$

With:

$$L_{sym} = D^{-\frac{1}{2}}LD^{-\frac{1}{2}} = D^{-1}L \tag{41}$$

$$g^T L_{sym} g = \sum_i^n = 1\lambda_i |(U^T x)_i|^2 = \sum_i^n = 2\lambda_i |U_i^T x|^2 = \lambda_2 g^T g \tag{42}$$

At $g = u_2$ with the constraints from earlier:

$$min \frac{g^T L_{sym} g}{g^T g} = \frac{f^T L f}{f^T D f} \tag{43}$$

$$f_j(v_i) = f_j(i) = \frac{1}{\sqrt{vol(A_j)}} if v_i \in A_j, else 0 \tag{44}$$

Therefore:

$$f^T L f = \frac{cut(A_i, \bar{A}_i)}{vol(A_i)} \tag{45}$$