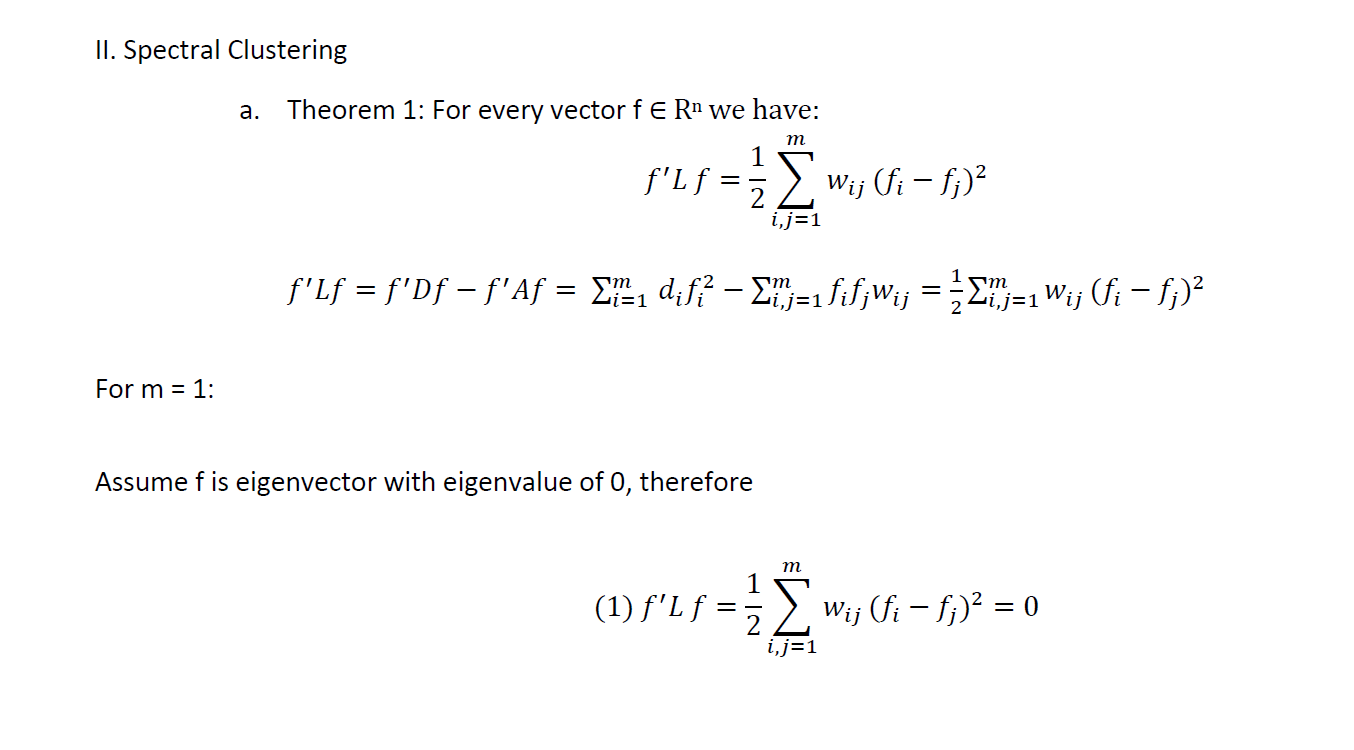
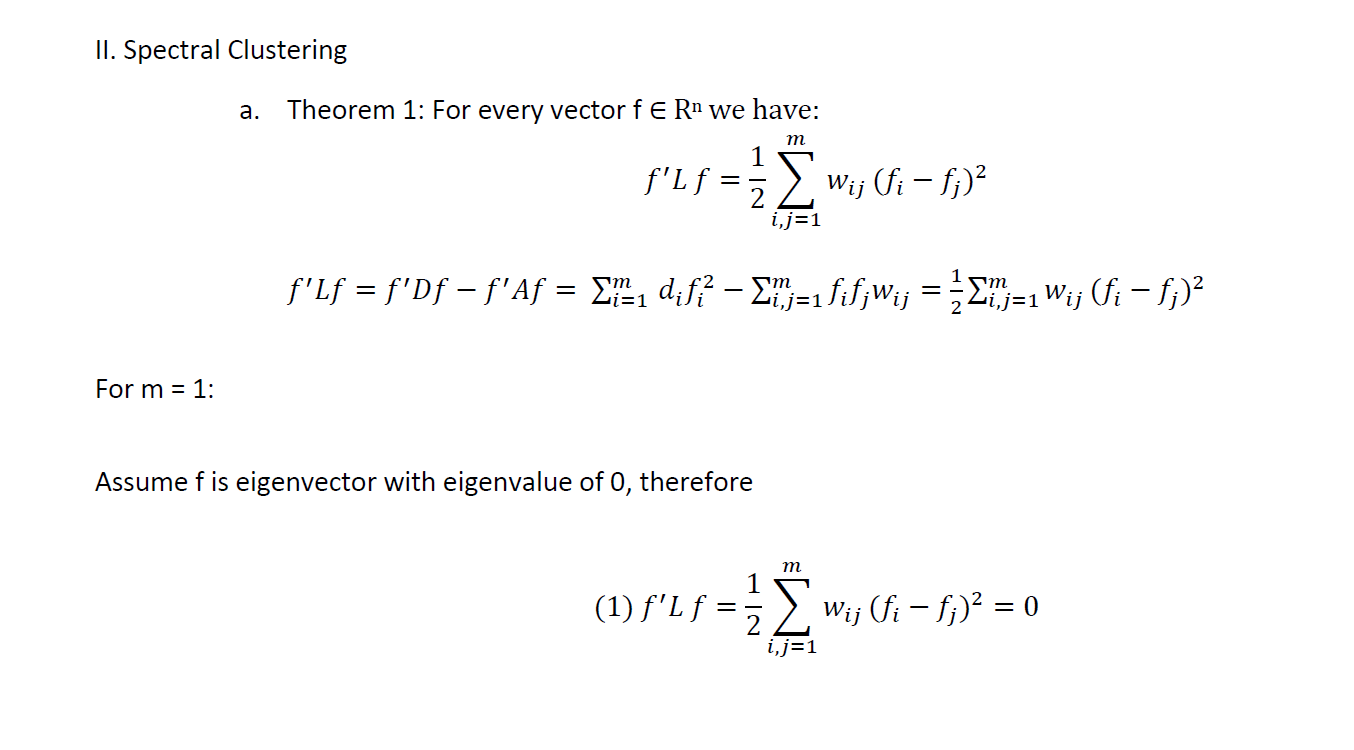
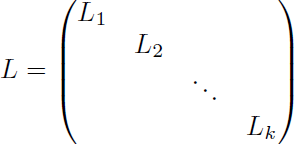
Kien Duc Vu GTID: 903552166

Homework 2

**Q1a. Spectral Clustering**



(1) Happens if and only if 𝑓𝑖 = 𝑓𝑗 ( 𝑠𝑖𝑛𝑐𝑒 𝑤𝑖𝑗 > 0). Thus, f is an indicator vector of the connected component. Now, considering Laplacian matrix L with m > 1,



Each of the Li is a proper graph Laplacian on its own corresponding to the subgraph of the i-th connected component, with m=1, that has an indicator vector of eigenvalue 0. Therefore, L has m eigenvectors corresponding to eigenvalue and the indicator vectors of these components span the zero eigenspace.

**Q2. PCA**

1. Matrix data has shape of (13, 20) with 13 rows representing 13 countries, and 20 column features representing 20 kinds of food being analyzed in this study. Step by step PCA:

* Rescale and normalize dataset by mean and standard deviation
* Calculate mean, and covariance matrix
* Calculate eigenvectors and eigenvalues
* Choose k eigenvectors that have largest eigenvalues
* Find k principal component directions corresponding to k eigenvectos
* Project data onto k-principal component directions
* Plot PCA

1. X ∈ R nxd

{ui} is basic orthonormal vector representing d-dimensional space. Xn can be represented as linear combination of ui.

Equivalently,

We approximate each data point Xn by

Where is fixed for all data point and {zi} {ui} are datapoint-dependent. Distortion function J:

Taking derivative of J with respect of z and b gives

Therefore, J can be expressed as

General solution is obtained by choosing the {ui} to be eigenvectos of the covariance matrix given by

And