EE 381V: Large Scale Learning **Spring 2013**

Lecture 6 — January 31

Lecturer: Caramanis & Sanghavi *Scribe: Rajiv Khanna*

6.1 Last time

In the previous lecture, we introduced the spectral clustering algorithm. We showed how running a distanced based clustering on rows formed from bottom-k eigenvectors of the Laplacian stacked as columns in the matrix results on a toy data with non-trivial visible clustering gives perfect clusters. To explain the goodness of the result, we proved that the smallest eigenvalue of the Laplacian is 0; this helped us in proving that running spectral clustering on a disconnected graph with k connected components would yield perfect clusters. We then moved onto matrix perturbation to extend usability of the spectral clustering algorithm on general graphs which are "close to" ones with k-connected components. In this class, we continue our discussion on perturbation results, and then apply it to our spectral clustering algorithm.

6.2 Perturbation of Symmetric Matrices

[Defn] **Distance between Subspaces:** Let $M \in \mathbb{R}^{n\times n}$ be the original symmetric matrix. Let Δ be the symmetric perturbation applied to it, so that $M + \Delta$ is another symmetric matrix. Further, let $E_0 \in \mathbb{R}^{n\times k}$ be the matrix formed by stacking the bottom-k eigenvectors (i.e. eigenvectors corresponding to least-k eigenvalues). Let, F_0 be the corresponding matrix for $M + \Delta$. Then, the distance between subspaces spanned by E_0 and F_0 is given by:

$$d_p(E_0, F_0) = ||E_0E_0^* - F_0F_0^*||_2. \tag{6.1}$$

[Defn] **Principal angles:** With E_0 and F_0 defined as above (the definition is true for general subspaces), there exists a set of k angles $\boldsymbol{\theta} = \{\theta_1, \ldots, \theta_k\}$, defined recursively as follows:

$$\theta_1 := \min\left[\arccos\left(\frac{\langle u, v\rangle}{||u||||v||}\right)|u \in E_0, v \in F_0\right] = \angle(u_1, v_1),$$

$$\theta_j := \min\left[\arccos\left(\frac{\langle u, v\rangle}{||u||||v||}\right)|u \in E_0; v \in F_0; u \perp u_i, v \perp v_i, \forall i\, \text{s.t.}\, 1 \leq i \leq (j-1)\right] = \angle(u_j, v_j).$$

If $\Theta = diag(\boldsymbol{\theta})$, it can be shown that :

$$\begin{aligned} E_0^*F_0 &= U\cos\Theta V^*;\ \ U, V \in O(k), \\ O(k) &= \{Q \in \mathbb{R}^{k\times k} : Q^*Q = \mathbb{I}_k\}. \end{aligned} \tag{6.2}$$

Theorem 6.1. $\sin\Theta$ ***theorem***
By Spectral theorem, both M and $M+\Delta$ have eigenvalue decomposition.

$$M = [E_0 \mid E_1] \begin{bmatrix} M_0 & \mathbf{0} \\ \mathbf{0} & M_1 \end{bmatrix} [E_0 \mid E_1]^*,$$

$$M+\Delta = [F_0 \mid F_1] \begin{bmatrix} \hat{M}_0 & \mathbf{0} \\ \mathbf{0} & \hat{M}_1 \end{bmatrix} [F_0 \mid F_1]^*,$$

where $M_0, M_1, \hat{M}_0, \hat{M}_1$ are diagonal matrices of respective eigenvalues.
If $\exists a, b, \delta$, such that $M_0(i,i) \in [a,b]$, $\forall i$ and $\hat{M}_1(i,i) \in (-\infty, a-\delta) \cup (b+\delta, \infty)$ $\forall i$ then, $d_p(E_0, F_0) \leq \frac{1}{\delta}\|\Delta\|_2$.

Before we can prove Theorem 6.1, we need several results that would be useful. These results are presented and proved as lemmas.

Lemma 6.2. *With $F = [F_0|F_1]$, $E = [E_0|E_1]$, and Θ as diagonal matrix of principal angles, we have*

$$||F_1^* E_0||_2 = ||F_0^* E_1||_2 = ||\sin\Theta||_2.$$

Proof:

$$\begin{aligned} ||E_0^* F_1||_2 &= ||E_0^* F_1 F_1^* E_0||_2^{0.5} \\ &= ||E_0^*(I - F_0 F_0^*)E_0||_2^{0.5} \\ &= ||I - U\cos\Theta V^* V \cos\Theta U^*||_2^{0.5} && (6.3) \\ &= ||U(I - \cos^2\Theta)U^*||_2^{0.5} \\ &= ||I - \cos^2\Theta||_2^{0.5} && (6.4) \\ &= ||\sin\Theta||_2. \end{aligned}$$

Here, step 6.3 above comes from Equation 6.2 above. Step 6.4 comes from the properties of 2$-$norm.

The proof from $||F_0^* E_1||_2$ is similar.

□

Lemma 6.3. *If*

$$E_0^* F_0 = A = U\cos\Theta V^*,$$

then,

$$\exists \hat{U}, E_1^* F_0 = B = \hat{U}\cos\Theta V^*.$$

Proof:

$$\begin{aligned} B^*B &= F_0^*E_1E_1^*F_0 \\ &= F_0^*(I - E_0E_0^*)F_0 \\ &= I - F_0^*E_0E_0^*F_0 \\ &= I - A^*A \\ &= I - V\cos\Theta U^*U\cos\Theta V^* \\ &= I - V\cos^2\Theta V^* \\ &= V(\sin^2\Theta)V^*. \end{aligned}$$

□

Lemma 6.4.

$$d_p(E_0, F_0) = ||\sin\Theta||_2.$$

Proof: From A and B as defined in Lemma 6.3, note that we can write $F_0 = E_0A + E_1B$. This helps us to write in $[E_0|E_1]$-basis :

$E_0E_0^* = \begin{pmatrix} I & | & 0 \\ -- & | & -- \\ 0 & | & 0 \end{pmatrix}$, and $F_0F_0^* = \begin{pmatrix} AA^* & | & AB^* \\ -- & | & -- \\ BA^* & | & BB^* \end{pmatrix}$.

This gives,

$$\begin{aligned} E_0E_0^* - F_0F_0^* &= \begin{pmatrix} I & | & 0 \\ -- & | & -- \\ 0 & | & 0 \end{pmatrix} - \begin{pmatrix} AA^* & | & AB^* \\ -- & | & -- \\ BA^* & | & BB^* \end{pmatrix} \\ &= \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix} - \begin{pmatrix} U\cos^2\Theta U & U\sin\Theta\cos\Theta\hat{U} \\ \hat{U}\sin\Theta\cos\Theta U^* & \hat{U}\sin^2\Theta\hat{U}^* \end{pmatrix} \\ &= \begin{pmatrix} U & 0 \\ 0 & \hat{U} \end{pmatrix} \begin{bmatrix} \sin^2\Theta & -\sin\Theta\cos\Theta \\ -\sin\Theta\cos\Theta & -\sin^2\Theta \end{bmatrix} \begin{pmatrix} U & 0 \\ 0 & \hat{U} \end{pmatrix}. \\ \implies ||E_0E_0^* - F_0F_0^*||_2 &= \left|\left| \begin{bmatrix} \sin^2\Theta & -\sin\Theta\cos\Theta \\ -\sin\Theta\cos\Theta & -\sin^2\Theta \end{bmatrix} \right|\right|_2 \\ &= \max_i \left|\left| \begin{bmatrix} \sin^2\theta_i & -\sin\theta_i\cos\theta_i \\ -\sin\theta_i\cos\theta_i & -\sin^2\theta_i \end{bmatrix} \right|\right|_2 \qquad (6.5) \\ &= \max_i |\sin\theta_i| \left|\left| \begin{bmatrix} \sin\theta_i & -\cos\theta_i \\ \cos\theta_i & -\sin\theta_i \end{bmatrix} \right|\right|_2 \\ &= \max_i |\sin\theta_i| \qquad (6.6) \\ &= ||\sin\Theta||_2. \qquad (6.7) \end{aligned}$$

□

Above, Step 6.5 follows from properties of diagonal matrices, Step 6.6 follows from the fact that the matrix in the previou step is a rotational matrix (with only real eigenvalue being $= 1$), Step 6.7 follows from $\sin \Theta$ being a diagonal matrix again.

Lemma 6.5.

$$d_p(E_0, F_0) = ||F_1^* E_0||_2 = ||F_0^* E_1||_2 = ||\sin \Theta||_2.$$

Proof: This follows immediately from Lemma 6.2 and 6.4. $\square$

Lemma 6.6. *Given:*

$$A \in \mathbb{R}^{n \times n}, B \in \mathbb{R}^{m \times m},$$
$$||A^{-1}||_2 \leq (\alpha + \delta)^{-1}, ||B||_2 \leq \alpha,$$
$$X \in \mathbb{R}^{n \times m}, C \in \mathbb{R}^{n \times m}.$$

Define $C = AX - XB$, then $||C||_2 \geq \delta ||X||_2$.

Proof: Note that

$$||XB||_2 \leq ||X||_2 . ||B||_2 \leq \alpha ||X||_2,$$

and,

$$||X||_2 = ||A^{-1} AX||_2 \leq (\alpha + \delta)^{-1} ||AX||_2.$$

Consider,

$$||C||_2 = ||AX - XB||_2 \geq ||AX||_2 - ||XB||_2 \geq (\alpha + \delta)||X||_2 - \alpha ||X||_2 = \delta ||X||_2.$$

$\square$

We are now ready to prove the $\sin \Theta$ theorem - Theorem 6.1.

Proof: Let $R = \Delta E_0 = (M + \Delta)E_0 - M E_0$.
Consider,

$$\begin{aligned}
E_0^* \Delta F_1 &= R^* F_1 \\
&= E_0^* (M + \Delta) F_1 - M_0^* E_0^* F_1 \\
&= (E_0^* F_1) \hat{M}_1 - M_0^* (E_0^* F_1). \\
\implies ||E_0^* \Delta F_1||_2 &= ||(E_0^* F_1) \hat{M}_1 - M_0^* (E_0^* F_1)||_2 \\
&\geq \delta ||E_0^* F_1||_2 && (6.8) \\
&= \delta d_p(E_0, F_0). && (6.9) \\
\implies d_p(E_0, F_0) &\leq \frac{1}{\delta} ||E_0^* \Delta F_1||_2 \\
&\leq \frac{1}{\delta} ||\Delta||_2.
\end{aligned}$$

Above, Step 6.8 follows from Lemma 6.6, and Step 6.9 follows from Lemma 6.5.

$\square$

6.3 Application of $\sin\Theta$ theorem

Having proved the theorem, we shall now apply it to matrices which are "close to" matrices of graphs of k connected components. The idea is that if such matrices are sufficiently close, we should be able to reproduce the clusters using a distance based clustering.

For the graph with k connected components, the smallest k eigenvalues of the Laplacian are 0. Let L be the laplacian matrix of the graph. Let Y be the matrix formed by stacking eigenvectors corresponding to smallest k eigenvectors. Let $\hat{L}$ be the perturbed Laplacian, and $\hat{Y}$ be corresponding eigenvector matrix. By $\sin\Theta$ theorem,

$$d_p(Y,\hat{Y}) \leq \frac{1}{\hat{\lambda}_{k+1}}||L-\hat{L}||_2. \tag{6.10}$$

Note that δ in the $\sin\Theta$ theorem is $(k+1)^{\text{th}}$ eigvenvalue of the perturbed Laplacian $\hat{L}$. This is because the smallest k eigenvalues of L are 0.

However, notice that the Equation 6.10 gives the distance between subspaces spanned by *columns* of Y and $\hat{Y}$, while we used *rows* in our algorithm. Hence, we need a different distance function. Define:

$$d_c(Y,\hat{Y}) = \min_{Q,R\in O(k)} ||YQ-\hat{Y}R||_2 = \min_{R\in O(k)} ||Y-\hat{Y}R||_2.$$

The relationship between the two distance measures is as follows.

Lemma 6.7.

$$d_p(Y,\hat{Y}) \leq d_c(Y,\hat{Y}) \leq \sqrt{2}d_p(Y,\hat{Y}).$$

Proof:

$$\begin{aligned}
[d_c(E_0,F_0)]^2 &= \min_Q ||E_0 - F_0Q||_2^2 \\
&= \min_Q ||E_0^*E_0 + Q^*F_0^*F_0Q - Q^*F_0^*E_0 - E_0^*F_0Q||_2 \\
&= \min_Q ||2 - Q^*V\cos\Theta U^* - U\cos\Theta V^*Q||_2 \\
&= \min_Q ||2 - Q^*\cos\Theta - \cos\Theta Q||_2 \\
&= 2\min_Q \max_{||x||=1}\{1-\langle x, Q^*\cos\Theta x\rangle\} \\
&= 2\{1-\max_Q \min_{||x||=1}\langle Qx,\cos\Theta x\rangle\}.
\end{aligned}$$

Consider

$$\begin{aligned}
\max_Q \min_{||x||=1}\langle Qx,\cos\Theta x\rangle &\geq \min_{||x||=1}\langle Qx,\cos\Theta x\rangle \\
&= \min_i \cos\theta_i.
\end{aligned}$$

Also,

$$\begin{aligned}\max_Q \min_{||x||=1} \langle Qx, \cos\Theta x\rangle &\leq \max_Q \langle Qe_j, \cos\Theta e_j\rangle \\ &= \min_i \cos\theta_i,\end{aligned}$$

where e_j is a vector of all 0s except at j^{th} position where it is 1, and $j = \arg\min_i \cos\theta_i$

Hence, $\max_Q \min_{||x||=1}\langle Qx, \cos\Theta x\rangle\} = \min_i \cos\theta_i$.

This gives,

$$\begin{aligned}d_c(E_0, F_0) &= \sqrt{2||1-\cos\Theta||_2} \\ &= ||2\sin(\Theta/2)||_2.\end{aligned}$$

Comparing with $d_p(E_0, F_0) = ||\sin\Theta||_2 = ||2\sin(\Theta/2)\cos(\Theta/2)||_2$ proves the lemma. □

Using the lemma, we redefine the matrix if eigenvectors for perturbed Laplacian.

Theorem 6.8.

$$\exists Q \in O(k)\ s.t.\ \ Y' = \hat{Y}Q,$$

and

$$d_c(Y, Y') = ||Y - Y'||_2,$$

and,

$$||Y - Y'||_2 \leq \frac{\sqrt{2}}{\hat{\lambda}_{k+1}}||L - \hat{L}||_2,$$

then,

$$\frac{1}{n}\sum_{i=1}^{n}||y_i - y_i'||_2^2 \leq \frac{2k}{n\hat{\lambda}_{k+1}^2}||L - \hat{L}||_2^2. \tag{6.11}$$

Proof:

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}||y_i - y_i'||_2^2 &= \frac{1}{n}||Y - Y'||_F^2 \\ &= \frac{1}{n}\sum_j \sigma_j^2 \\ &\leq \frac{1}{n}.k.\max_j \sigma_j^2 \\ &= \frac{k}{n}||Y - Y'||_2^2 \\ &\leq \frac{2k}{n\hat{\lambda}_{k+1}^2}||L - \hat{L}||_2^2.\end{aligned}$$

□

Equation 6.11 gives a bound on average euclidean distance of points generated in our clustering algorithm in terms of perturbation in the Laplacian. However, RHS contains $\hat{\lambda}$. It is desirable to minimize the dependence of the bound on perturbed matrix. The next lemma helps replacing $\hat{\lambda}$ with λ.

Lemma 6.9.

$$\hat{\lambda}_{k+1} \geq \lambda_{k+1} - ||L - \hat{L}||_2.$$

Proof: Let $V = \text{span}(u_1, u_2, \ldots, u_k)$, where u_i is eigenvector corresponding to $\hat{\lambda}_k$, then,

$$\begin{aligned}
\hat{\lambda}_{k+1} &= \max_{dim(V)=k} \min\{\langle x, \hat{L}x\rangle : x \in V^{\perp}, ||x|| = 1\} \\
&\geq \min_{||x||=1} \{\langle x, \hat{L}x\rangle : x \in V^{\perp}\} \\
&\geq \min_{||x||=1, x\in V^{\perp}} \{\langle x, Lx\rangle\} - \max_{||x||=1} \{\langle x, (L - \hat{L})x\rangle\} \\
&= \lambda_{k+1} - ||L - \hat{L}||_2.
\end{aligned}$$

□

Lemma 6.9 helps us modify Theorem 6.8 :

Theorem 6.10.

$$\exists Q \in O(k) \ s.t.\ Y' = \hat{Y}Q,$$

and

$$d_c(Y, Y') = ||Y - Y'||_2,$$

then,

$$\frac{1}{n}\sum_{i=1}^{n} ||y_i - y_i'||_2^2 \leq \frac{2k}{n(\lambda - ||L - \hat{L}||_2)^2} ||L - \hat{L}||_2^2. \tag{6.12}$$

6.4 Next time

Next time we shall discuss about planted models.